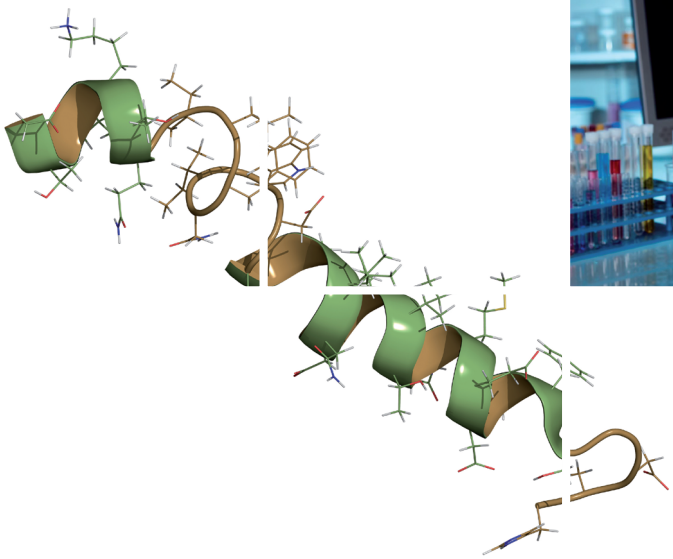
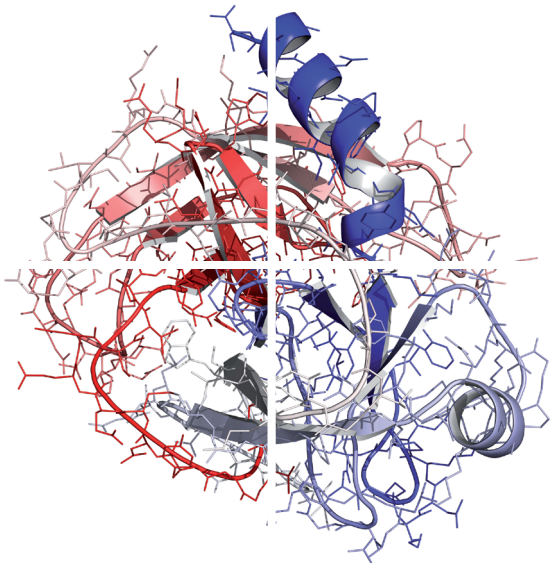
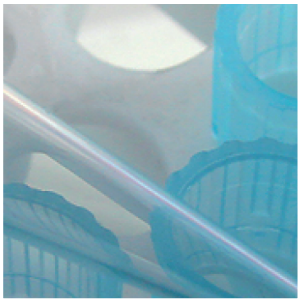


Manual de Proteómica





Sociedad Española de Proteómica

Manual de Proteómica

Editado por Fernando Corrales y Juan J. Calvete.
Edición digital.

D.L.: NA 1462-2014
ISBN: 978-84-697-1281-8

Enmaquetado e impreso por 23 pares Ilustraciones Científicas S.L.
Diseño de portada, Joanna Armendáriz.

Dedicado a
Juan Pablo Albar

Prólogo

La proteómica es una joven disciplina en pleno desarrollo tecnológico que aporta medios impensables hace tan solo unos años para analizar con una capacidad sin precedentes las proteínas integrantes de un sistema biológico, su proteoma. Integra múltiples estrategias analíticas que se pueden combinar en flujos de trabajo relativamente complejos que exigen cierto nivel de especialización. La Sociedad Española de Proteómica, en su afán por diseminar el conocimiento en proteómica y fomentar la formación de nuevos investigadores, ha promovido la edición del presente manual. Esperamos que la edición en español contribuya a consolidar los puentes entre las comunidades proteómicas hispanoparlantes a ambos lados del océano Atlántico y a facilitar la formación académica de nuestros estudiantes con inquietudes en Biología.

Es de obligado cumplimiento agradecer a los investigadores, amigos, que han hecho posible este proyecto con sus magníficas contribuciones sobre los fundamentos de las técnicas que constituyen nuestro arsenal proteómico, desde los procedimientos de separación a los más sofisticados análisis basados en la espectrometría de masas, métodos estadísticos, etc.

Fernando J. Corrales.
Presidente de la SEProt

ÍNDICE

13 01. De la Genómica a la Proteómica

21 02. Estructura química y propiedades ácido-base de péptidos y proteínas

33 03. Cromatografía. Análisis y separación de Péptidos y Proteínas

57 04. Técnicas electroforéticas

73 05. Métodos de análisis estadístico en proteómica

111 06. Historia de la espectrometría de masas

127 07. Métodos y mecanismos de ionización

173 08. Consideraciones sobre la interpretación de espectros de masas

187 09. Analizadores de masa en proteómica

203 10. Espectrómetros de masa basados en tiempo de vuelo

289 11. FTMS Espectrometría de Masas por Transformada de Fourier

313 12. Espectrómetro de masas de trampa iónica cuadrupolar

329 13. El universo de configuraciones de analizadores de masas

345 14. HPLC-Chip/MS. Flexible y robusta tecnología microfluídica para nanocromatografía/MS

363 15. Preparación de la muestra. Fraccionamiento de subproteomas e inmunodepleción de proteínas mayoritarias

381 16. Aislamiento selectivo de Péptidos. Scape

395 17. Mecanismos de fragmentación de iones peptídicos mediante disociación inducida por colisión (CID)

409 18. Secuenciación de novo de péptidos mediante espectrometría de masas: ¿cómo interpretar un espectro MS/MS “a mano”?

431 19. Identificación de proteínas mediante huella peptídica y datos de $(MS)^n$

481 20. Determinación del número de grupos sulfidrilo y de enlaces disulfuro mediante espectrometría de masas

489 21. Glicómica

501 22. Fosfoproteómica. Técnicas de enriquecimiento de fosfopéptidos. Métodos espectrométricos para la identificación de fosfopéptidos

521 23. Proteómica redox: análisis de modificaciones postraduccionales oxidativas

537 24. Proteómica de Organismos Huérfanos

561 25. Proteómica en hongos fitopatógenos

583 26. Proteómica clínica. Biomarcadores y dianas terapeúticas

603 27. MS Imaging

615 28. Estrategias en proteómica de alto rendimiento (high hroughput). Caracterización de proteomas. LC-MS/MS. FDR

637 29. Aplicaciones de las herramientas proteómicas en Tecnología de la Alimentación

653 30. METALÓMICA. Aspectos medioambientales

673 31. Intensity Fading MALDI-TOF MS en Proteómica Funcional e Interactómica

687 32. Microarrays de proteínas

De la Genómica a la Proteómica

01

Juan Jose Cazzulo

Instituto de Investigaciones Biotecnológicas Dr. Rodolfo Ugalde - Instituto Tecnológico de Chascomús (IIB-INTECH, UNSAM-CONICET).

1. El Genoma y la genómica.

El término "genoma" fue introducido por el botánico alemán Hans Winkler en 1920, a partir de las palabras "GENes" y "cromosOMAs" (1), y puede definirse en consecuencia como el juego completo de cromosomas y sus genes en una especie biológica determinada.

Wilhelm Johanssen introdujo en 1911 los conceptos de "genotipo" y "fenotipo". El genotipo consiste en la información hereditaria completa de un individuo, se exprese o no en un momento dado. Puede considerárselo como el plano maestro o juego de instrucciones para construir y mantener funcional a una criatura viviente. El fenotipo, en cambio, es la manifestación visible del genotipo; son las propiedades observables de la célula u organismo, como su morfología, su desarrollo, su comportamiento. El fenotipo depende no sólo del genotipo, sino, aunque en menor medida, de la interacción del organismo con su ambiente. Dos organismos que difieran aún en un solo locus (posición en el genoma) tienen distintos genotipos, aunque eventualmente pueden tener fenotipos indistinguibles. Inversamente, dos individuos con idéntico genotipo (por ejemplo, los hermanos gemelos) pueden tener diferencias menores en su fenotipo.

La "genómica" (genomics, en Inglés) es el estudio de los genomas completos de los diferentes organismos, y consiste en la determinación de la secuencia completa de su ADN y la ubicación de los genes secuenciados en los diferentes cromosomas. El estudio intensivo de un único gen y su(s) producto(s) no entra en esta definición de genómica. La secuenciación completa de un genoma implica el conocimiento completo del genotipo del individuo del cual se extrajo el ADN. La secuenciación de los genomas se hizo posible con la introducción de los métodos de secuenciación del ADN, a partir del establecimiento del método utilizando dideoxinucleótidos por Frederick Sanger, en 1975 (2), que le permitió secuenciar el genoma completo del bacteriófago ϕX174, con 5,375 pares de bases, en 1977. La primera secuenciación completa del genoma de un organismo de vida libre fue la de Haemophilus influenzae (1.8 Mb), en 1995. El Human Genome Project (http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml) comenzó en 1990 y se completó en 2003; la estimación del número total de genes en el genoma humano no es aún definitiva, y la estimación es de 20,000 a 25,000. En Noviembre de 2013 la Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) registraba ya 200 genomas eucarióticos completos, y 18 en "borrador", además de 2,529 genomas de bacterias y 161 de Archaea.

El tener acceso a la información genómica de un organismo es de fundamental importancia en la actualidad para los científicos interesados en el mismo; debe destacarse, sin embargo, que

la información con que se cuenta es estrictamente sobre “marcos abiertos de lectura” (ORFs, de la sigla inglesa), es decir segmentos de ADN con un codón de iniciación y uno de terminación en el mismo marco de lectura, que por lo tanto pueden ser transcriptos dando un ARNm y eventualmente este puede traducirse a proteína. Los ORFs encontrados se “anotan”, es decir, por similitud con genes ya conocidos, se los identifica tentativamente como el gen codificante de una determinada proteína. La identificación definitiva del gen se logra en realidad sólo cuando se lo expresa y la proteína resultante tiene efectivamente la actividad biológica correspondiente. Todos los genomas secuenciados hasta hoy tienen seguramente genes mal anotados, y además en todos los casos hay un porcentaje elevado de “proteínas hipotéticas”, producto de ORFs para los cuales no se ha identificado un gen similar en otro organismo y en consecuencia se ignora completamente la naturaleza de la proteína codificada.

2. Fenotipo celular: definición de proteoma.

El genotipo es, como vimos, el conjunto de la información genética de un organismo, uni o pluricelular. El ADN conserva la información genética y la transmite de generación en generación, a través de la replicación, y la expresa, al transcribirse al ARN mensajero. Los ARNm serán entonces traducidos a proteínas, y estas serán las que se encargarán de llevar a cabo todas las funciones de la célula, directa o indirectamente. En efecto, muchas de esas funciones serán ejercidas directamente por proteínas, pero otras serán cumplidas por otro tipo de compuestos, por ejemplo lípidos o carbohidratos. La información para su síntesis no está contenida de modo directo en el genoma, sino que están allí codificados los sistemas enzimáticos que se encargarán de ella. La expresión de los genes llevará al fenotipo de la célula u organismo en cuestión, a través de la acción de las proteínas. Por analogía con el concepto de genoma, uniendo los términos “PROTEína” y ”genOMA”, Marc Wilkins acuñó en 1994 el término “proteoma”, para indicar el conjunto completo de las proteínas que una célula u organismo puede expresar (3). En el caso de los organismos pluricelulares, el genoma de todas sus células es idéntico, pero su proteoma es característico de cada tipo celular. Hay proteínas que se expresan en todas o la mayoría de las células de un organismo, pero hay otras que son características y se expresan sólo en un tipo celular o en unos pocos. El organismo humano tiene más de 200 tipos celulares diferentes, y lo que origina la diferencia es la variación en la expresión de sus proteínas. La hemoglobina se expresa en los glóbulos rojos, pero no en las neuronas o en las células musculares; estas últimas contienen en cambio mioglobina. Todo esto hace que, para conocer realmente la función de los genes, deberemos conocer todas las características químicas, bioquímicas y biológicas de las proteínas que codifican. Esto implica la secuencia y estructura tridimensional de las proteínas; la presencia de modificaciones post-traduccionales y procesamiento proteolítico; la capacidad de unir ligandos, incluyendo las interacciones con otras proteínas; la especificidad de tejido, célula y organela; la velocidad de síntesis y degradación, de las cuales dependerá la concentración de la proteína en un momento determinado; la especificidad de sexo, en los organismos que lo presentan; la presencia durante el desarrollo embrionario, el período post-natal y/o el envejecimiento (1). Es evidente entonces que la dilucidación completa del proteoma de un organismo es una tarea mucho más compleja y difícil que la dilucidación de su genoma.

El desarrollo histórico de las metodologías aplicables a lo que ahora llamamos proteómica (4) se inició en 1964, con la introducción de la “disc electrophoresis” en geles de poliacrilamida por L. Ornstein y B.J. Davis. En 1970 Laemmli introdujo el uso del dodecilsulfato de sodio (SDS) para separar las proteínas en función de su peso molecular (SDS-PAGE). A mediados de la década de 1970 varios autores, entre ellos J. Klose y P.H. O’Farrell, introdujeron la electroforesis bidimensional (2D-PAGE), combinando el isoelectroenfocado (IEF) con la SDS-PAGE; O’Farrell pudo detectar con esta técnica, usando las proteínas totales de Escherichia coli marcadas ra-

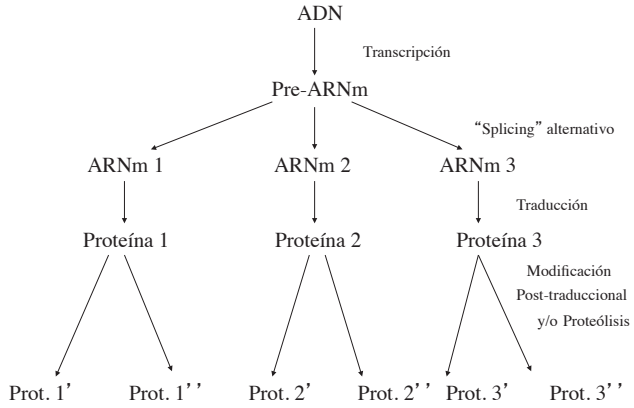
dioactivamente, 1100 manchas en un único gel de 2D-PAGE, de las cuales alrededor de 400 fueron detectadas tambien por tinción con Coomassie Blue. En 1982 Norman y Leigh Anderson presentaron un plan para catalogar todas las proteínas producidas en el cuerpo humano, al que llamaron "The Human Protein Index". La introducción de la espectrometría de masa para la identificación de las proteínas separadas en los geles 2D dió un enorme impulso a estos estudios. La Human Proteome Organization (HUPO, http://www.hupo.org/) fue fundada en Febrero de 2001 y el "Human Proteome Project", diseñado a partir de 2008, tiene un Grupo de Tareas establecido desde principios de 2010.

Para comprender la complejidad del estudio del proteoma, hay algunas preguntas fundamentales que debemos contestar.

3. ¿El número de proteínas que produce un organismo es igual al número de genes?.

La respuesta instintiva es que así debería ser, y de hecho la teoría original de George Beadle y Edward Tatum en 1941 se basaba en el postulado "un gen, una enzima", luego extendido a "un gen, una proteína", y finalmente, al observarse la existencia de proteínas heterooligoméricas, a "un gen, un polipéptido". Sin embargo, la respuesta es claramente negativa. Los organismos pluricelulares en general son capaces de producir un número de proteínas apreciablemente superior al número de sus genes. Hay dos razones principales para este aparente contrasentido. La gran mayoría de los genes de organismos eucarióticos tienen en su secuencia zonas que se expresan, a través del ARNm, en la proteína codificada, y se denominan "exones", y otras zonas que se eliminan durante el procesamiento del pre-ARNm para llegar al ARNm maduro, no codifican en consecuencia residuos de aminoácidos presentes en la proteína expresada, y se denominan "introns". Ahora bien, hay numerosos genes, sobre todo en los eucariotas superiores, que constan de un número considerable de introns y exones. En tales casos, frecuentemente existe lo que se denomina, empleando el término inglés, el corte y empalme ("splicing") alternativo, que implica que, durante la maduración del ARNm, se pueden generar diferentes combinaciones de exones (2). Un ejemplo de esta situación es el gen único que codifica a los dos kininógenos humanos, el de alto peso molecular y el de bajo peso molecular, cuya diferencia está dada por la expresión o no de un largo exón en la zona C-terminal de la proteína. Pero el entrecortado alternativo no es la única razón de la discrepancia numérica; muchas proteínas contienen residuos de aminoácidos que no son idénticos a los codificados en su gen estructural,

Figura 1.



debido a que han sufrido modificaciones químicas covalentes después de la traducción a nivel de los ribosomas. Se conocen mas de 200 modificaciones post-traduccionales covalentes, que en general son esenciales para la funcionalidad de una determinada proteína. Más aún, numerosas proteínas requieren, después de terminada la traducción, un procesamiento proteolítico parcial, que requiere proteasas altamente específicas. Un ejemplo son las proteínas que deben seguir la vía secretoria, todas las cuales requieren la presencia de un "péptido señal" en su extremo N-terminal, el cual es removido proteolíticamente al ingresar la proteína naciente al lumen del retículo endoplásmico. Este procesado es absolutamente esencial: se ha comprobado por experimentos de mutagénesis sitio-dirigida que alterando el sitio de clivaje de la proteína naciente, la misma permanece unida a la membrana del retículo y no sigue la vía normal, perdiendo en consecuencia su funcionalidad. Lo que necesita la célula es la proteína "madura", es decir bien sintetizada y plegada, con las modificaciones post-traduccionales y/o procesado proteolítico correcto, localizada en el compartimento sub-celular correcto, para que sea funcional. El resultado de los diversos procesos que hemos descripto es que se estima que la mayoría de los genes en un organismo superior pueden dar origen a seis o mas proteínas diferentes (Fig. 1). Si consideramos un número de 20 a 25,000 genes en el genoma humano, es altamente posible que las proteínas que nuestro organismo es capaz de producir sean mas de 150,000, talvez muchas más.

4. ¿Una célula produce en todo momento todas las proteínas que sería capaz de producir?

Nuevamente, la respuesta a esta pregunta es negativa. En primer término, cada célula será capaz de producir sólo ciertas proteínas y otras no, debido a las modificaciones en la cromatina, específicas para cada tipo celular, que llevan a la falta de transcripción de un número grande de genes. En segundo término, el proteoma tiene un carácter dinámico, lo que implica que está en continuo cambio ajustándose a las necesidades de la célula en ese momento y en esas condiciones ambientales. El proteoma determinado para una célula en una dada condición experimental podemos considerarlo como una fotografía que nos muestra la situación en ese momento, y que será diferente en algunos aspectos de otra tomada bajo diferentes condiciones experimentales. Un ejemplo clásico y muy claro es la inducción y represión de enzimas en las bacterias. Escherichia coli es capaz de producir todas las enzimas necesarias para la síntesis de los 20 aminácidos proteicos, y lo hace cuando crece en un medio simple, con glucosa como fuente carbonada. Sin embargo, si se le suministra en el medio de cultivo uno o más aminoácidos, rápidamente la bacteria reprimirá la síntesis de las enzimas que llevan a la producción de ese o esos aminoácidos, y los tomará del medio. Es obvio que si determináramos el proteoma de E. coli crecida en uno u otro medio de cultivo, encontraremos diferencias, debidas a la presencia o ausencia de una serie de proteínas enzimáticas. Por cierto, este fenómeno no es privativo de los procariotas; todas nuestras células deben ajustar la expresión de muchas proteínas, en relación con las variaciones del medio y con los estímulos de diferentes clases que reciben.

5. ¿Basta la integridad del gen estructural de una proteína para su correcta expresión y funcionalidad?

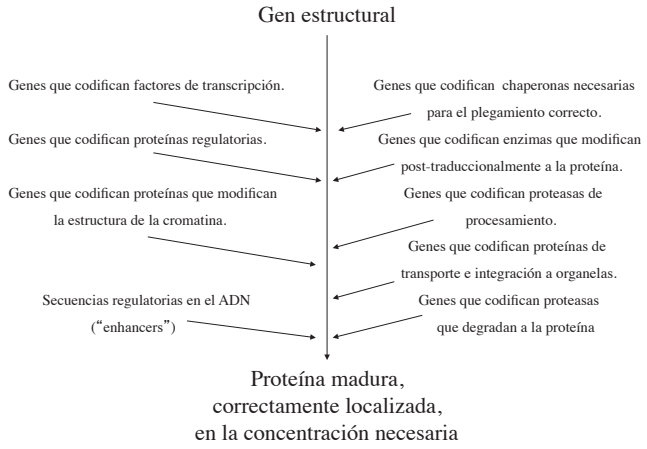
Nuevamente, la respuesta es no. Por cierto, si el gen estructural que codifica a la proteína está mutado, de tal manera que la mutación, por ejemplo la de residuos catalíticos en una enzima, impide su función, la proteína mutante no será funcional. Pero la integridad y correcta función

de otros genes es esencial para que la proteína en cuestión sea completamente funcional. Es necesaria la integridad de los genes que codifican a las proteínas responsables de las modificaciones post-traduccionales necesarias; a las proteinasas específicas responsables del procesado proteolítico; a las proteinasas que intervienen en su degradación, y por lo tanto regulan, tanto como la síntesis, el nivel de una proteína en un determinado momento; a las proteínas requeridas para el transporte y la localización de la proteína en cuestión en el compartimento subcelular correcto; a los factores de transcripción requeridos para la expresión de la proteína; a proteínas regulatorias que deben interactuar con la proteína en cuestión para que sea funcional; a proteínas involucradas en la metilación del ADN y en la estructura de la cromatina. Incluso, se requiere la integridad de secuencias regulatorias en el ADN, como las secuencias potenciadoras (enhancers) (Fig. 2). Mutaciones en genes como los antes mencionados que no anulen la actividad de la proteína que codifican, pero introduzcan cambios, pueden llevar a cambios en la concentración celular y a diferencias en propiedades de la proteína en cuestión. Esto ha llevado a Klose (1) a concluir que las proteínas tienen un carácter poligénico, en el sentido de que su nivel celular y funcionalidad no depende exclusivamente del gen estructural.

6. Los desafíos de la proteómica.

Kettman y col. (5) han calculado que un linfocito humano podría contener en cualquier momento alrededor de 40,000 moléculas de ARNm, pertenecientes a 5,000 especies diferentes, 4,000 de ellas presentes en bajo nivel, las cuales codificarían alrededor de 1,000,000,000 de moléculas de proteína. De estas proteínas las 100 mas abundantes representarán un 90 % de la masa total proteica de la célula, y las otras 4,900 especies moleculares proteicas el 10 % restante. El rango de abundancia de las proteínas celulares es de alrededor de 10,000,000; una proteína del citoesqueleto como la actina estará presente con una abundancia 10 millones de veces mayor que una proteína regulatoria. Estas cifras, aunque sean una estimación grosera, nos dan una primera idea de las dificultades con que se enfrenta la proteómica; las proteínas mas abundantes tenderán a obliterar completamente a las de menor abundancia, y su detección requerirá enfoques experimentales especiales, como por ejemplo la previa purificación de la organela que contiene a las proteínas minoritarias. Esto es particularmente evidente cuando se utiliza una de las técnicas mas empleadas en proteómica, la 2D-PAGE, pues la carga proteica en el gel es limitada y las proteínas mas abundantes darán manchas tan extensas que cubrirán completamente a las minoritarias.

Figura 2.



Las dificultades quedan muy evidentes cuando se comparan en una serie de aspectos las proteínas con los ácidos nucleicos. Ya dijimos que el número de proteínas en una célula dada es considerablemente mayor que el número de sus genes (al menos, de los que se expresan en ese tipo celular). Además, como indicaron claramente Hochstrasser y col. (6), la heterogeneidad de las proteínas es mucho mayor que la de los ácidos nucleicos. La suma de las masas de los residuos de aminoácidos que componen a la proteína, su naturaleza y su secuencia, van a determinar sus propiedades: a) la masa de la proteína, que puede variar entre unos pocos miles y varios millones de Daltons, y no puede ser inferida de la secuencia del gen que la codifica, por la posible presencia de modificaciones post-traduccionales y/o procesamiento proteolítico. b) su carga total a diferentes valores de pH, y su punto isoeléctrico (pI), dependientes de la abundancia relativa de grupos ácidos y básicos; en tanto que los ácidos nucleicos presentan un rango estrecho de valores de pI, en la zona ácida, las proteínas presentan valores de pI entre menos de 3 y más de 12, con dos zonas de mayor abundancia, una alrededor de pH 5.5 y la otra alrededor de 8.5. c) su hidrofobicidad, que dependerá de la abundancia relativa de residuos de aminoácidos con cadenas laterales hidrofóbicas; en tanto que los ácidos nucleicos son muy solubles en agua, muchas proteínas son muy hidrofóbicas e insolubles en agua, pudiendose calcular que alrededor de un 15 % de las proteínas de una célula eucariótica no entrarán en un gel de 2D-PAGE debido a su gran hidrofobicidad. Por todo esto, el análisis en gran escala de un amplio rango de proteínas requiere una cuidadosa planificación experimental y la aplicación de múltiples tecnologías, que incluyen, utilizadas en paralelo o secuencialmente, centrifugación diferencial; diversos métodos cromatográficos, como cromatografía de afinidad, de intercambio iónico y en fase reversa; electroforesis capilar, y electroforesis en gel (mono o bidimensional), terminando siempre en la identificación de las proteínas o los péptidos obtenidos de ellas por hidrólisis parcial, por alguna de las diferentes formas de espectrometría de masa.

7. Diferentes enfoques en los estudios proteómicos.

Si bien los métodos a aplicar y las dificultades experimentales serán similares en todos los casos, el enfoque de un estudio proteómico puede variar considerablemente. Puede ir desde intentar identificar el proteoma completo de una célula, en diversas circunstancias de su vida, un objetivo de máxima y muy dificultoso, hasta obtener información mucho mas restrigida, pero mas factible, por ejemplo determinar el proteoma de una determinada organela previamente purificada, o hacer una proteómica diferencial. Esta última consistirá en la comparación del proteoma de una misma célula sometida o no a un determinado estímulo, buscando identificar las proteínas que se expresan o no, o en una cantidad muy diferente, en un estado u otro de la célula. En organismos unicelulares con ciclos de vida complejos, es muy importante determinar el proteoma diferencial de los distintos estadíos de ese ciclo; es decir, identificar a las proteínas que se expresan o dejan de expresarse, o lo hacen con diferencias cuantitativas muy marcadas, en el proceso de diferenciación de un estadío al siguiente. Otra variante de la proteómica diferencial puede ser la búsqueda de factores de virulencia, a través de estudios proteómicos comparados de un organismo patógeno y una mutante avirulenta del mismo.

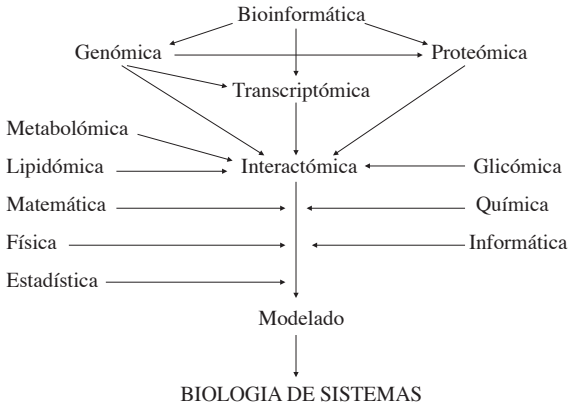
Otro enfoque directamente relacionado con la proteómica consiste en la llamada genómica estructural, que busca determinar la estructura tridimensional de cada proteína codificada por un determinado genoma, usando una estrategia mixta, experimental (cristalografía de rayos X; resonancia magnética nuclear, RMN) y de modelado computacional. Dado que la estructura está estrechamente ligada a la función de la proteína, y el plegamiento está mas conservado que la secuencia, este enfoque podría permitir orientarse hacia la función, algo muy importante en el caso de las muchas proteínas hipotéticas anotadas como tales en los diferentes geno-

mas. El enfoque experimental de la genómica estructural tiene los problemas generales de la cristalografía, la obtención de cristales de la proteína que difracten satisfactoriamente y permitan alcanzar una resolución adecuada, preferentemente de 2Å o menor, y de la RMN, el tamaño de la molécula en estudio. Para ello, las proteínas recombinantes deben obtenerse como proteínas solubles, puras y en cantidad suficiente. Todo esto limita este enfoque, pero no le quita validez y en los casos exitosos aporta información muy importante. Se han creado organizaciones destinadas a llevar adelante ambiciosos proyectos de genómica estructural, como la Protein Structural Initiative (PSI) de los National Institutes of Health (NIH) de los Estados Unidos de Norteamérica (http://www.nigms.nih.gov/psi/), que ha obtenido, desde su iniciación en 2000 hasta Noviembre de 2013, las estructuras de mas de 6,300 proteínas. El National Institute of General Medical Sciences (NIGMS) ha anunciado que no planea continuar PSI después de completar la PSI:Biology phase en 2015. Como ejemplo de un organismo específico, el Mycobacterium tuberculosis, el TB Structural Genomics Consortium, que agrupa investigadores de varios países, ha aportado ya (a Noviembre de 2013) 199 nuevas estructuras proteicas de este patógeno, sobre 1349 estructuras totales, a las cuales se puede acceder desde su página web (http://www.webtb.org/Structures/).

8. Hacia una biología de sistemas.

El conocimiento del genoma completo de un organismo, y aún el del proteoma completo que es capaz de producir, no nos da aún una idea completa de la célula o del organismo en cuestión. Está muy claro actualmente que, si bien las proteínas llevan a cabo, directa o indirectamente, todas las funciones esenciales de una célula, no actúan individualmente. Lo hacen interactuando con otras moléculas, proteicas o de otra composición química, y esas interacciones son esenciales para su función. El conjunto completo de interacciones entre las moléculas que constituyen una célula se ha definido como el "interactoma", término acuñado por Bernard Jacq (7). Aparte de las interacciones entre moléculas de diversa composición química, desde el punto de vista de la proteómica nos interesan particularmente las interacciones entre proteínas (el interactoma de proteínas), y tambien entre proteínas y ADN (la red de interacciones entre factores de transcripción y otras proteínas capaces de interaccionar con el ADN, y los genes, el interactoma proteína ADN). El conjunto de los ARNm celulares se ha descripto como el "transcriptoma". Mas aún, la interacción de los diversos componentes macromoleculares lleva en general a la producción, modificación o utilización de los distintos metabolitos presentes en la célula en un momento dado; el conjunto de metabolitos se ha definido como el "metaboloma".

Figura 3.



Estos, y otros nuevos conceptos han dado origen, del mismo modo que de genoma y proteoma se derivaron las disciplinas de genómica y proteómica, a transcriptómica, metabolómica, glicómica (estudio del complemento de carbohidratos de la célula), lipidómica (similar para los lípidos), etc. Todas estas disciplinas y los resultados obtenidos a través de ellas confluyen, junto con la Bioinformática y otras ciencias, en lo que se denomina actualmente Biología de Sistemas (8) (Fig. 3). Puede considerarse su precursor a Ludwig von Bertalanffy, que a partir de la década de 1940 desarrolló la teoría general de sistemas. La Biología de Sistemas es, por lo tanto, un estudio interdisciplinario centrado en la Biología, que intenta estudiar todas las interacciones entre los componentes de un sistema biológico, y comprender cómo esas interacciones dan origen al funcionamiento y comportamiento del sistema. Sus proponentes consideran que está basada en un concepto diferente al seguido generalmente en Ciencia, al que consideran "reduccionista", por tratarse del análisis de los componentes individuales de un sistema para tratar de comprenderlo. Su propuesta sería, en cambio, "holística", tratando de comprender el "todo"que constituye el sistema, y que, por las interacciones entre las partes, es mayor que su simple suma. Obviamente, el conocimiento de los partes es fundamental para comprender el sistema en su conjunto, y eso hace a las diversas "ómicas" herramientas esenciales para la Biología de Sistemas. El conocimiento de los componentes, sus propiedades e interacciones llevará al desarrollo de modelos que permitan predecir el comportamiento de los sistemas y sus respuestas a las variaciones del ambiente, y tendrá aplicaciones en diversos campos del conocimiento biológico y médico.

REFERENCIAS.

1. **Klose J.** Genotypes and phenotypes. Electrophoresis 1999:20, 643 – 52.

2. **Stryer L, Berg JM, Tymoczko JL,** Bioquímica. 5 ed. Barcelona: Editorial Reverté, S.A., 2003.

3. **Wilkins MR.** From proteins to proteomes: largescale protein identification by two-dimensional electrophoresis and amino acid analysis. Bio/Technology1996: 14, 61 – 5.

4. **Klose J.** From 2-D electrophoresis to proteomics. Electrophoresis. 2009: 30, S142-49.

5. **Kettman JR, Coleclough C, Frey JR, Lefkovits I.** Clonal proteomics: One gene – family of proteins. Proteomics 2002: 2, 624 – 31.

6. **Hochstrasser DF, Sanchez J-C, Appel RD.** Proteomics and its trends facing nature's complexity. Proteomics 2002: 2, 807 – 12.

7. **Sanchez C, Lachaize C, Janody F, et al.** Grasping at molecular interactions and genetic networks in Drosophila melanogaster using FlyNets, and Internet database. Nucel.Acids Res. 1999: 27, 89 – 94.

8. **Kitano H.** Systems biology: a brief overview. Science 2002: 295, 1662 – 4.

Estructura química y propiedades ácido-base de péptidos y proteínas

02

Daniel Pérez-Hernández[1,2], Raquel Mesa Carrasco[1,2], Jesús Vázquez[1,2]

[1] Laboratorio de Proteómica Cardiovascular. Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), E-28049, Madrid, Spain;

[2] Centro Nacional de Investigaciones Cardiovasculares, E-28029, Madrid, Spain.

1. Introducción.

Desde el punto de vista químico, las proteínas son poliamidas de gran tamaño. Son macromoléculas formadas por la polimerización de ácidos α-aminocarboxílicos. Las propiedades físicas y químicas de las proteínas dependen no sólo de su composición aminoacídica, sino también de su estructura tridimensional, la cual es tremendamente compleja. La proteómica moderna, sin embargo, se lleva a cabo a nivel de péptidos, que resultan normalmente de la digestión tríptica de las proteínas. Los péptidos con los que se trabaja habitualmente no tienen la complejidad estructural de las proteínas, y sus propiedades están definidas en gran medida por las características físico-químicas de las cadenas laterales de los aminoácidos, mostrando un comportamiento mucho más homogéneo que el de las proteínas desde el punto de vista químico. El conocimiento de las propiedades físico-químicas de los péptidos reviste en la actualidad una importancia creciente debido al desarrollo de nuevos métodos que se basan en su modificación química, incluyendo las técnicas de marcaje isotópico y los procedimientos de derivatización química empleados para la purificación selectiva, incluyendo el estudio a gran escala de modificaciones post-traduccionales.

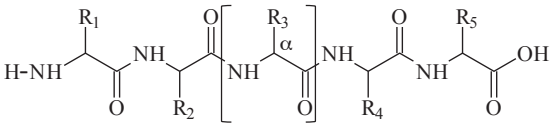
Es también muy importante tener en cuenta que el estudio de la química de los aminoácidos aislados no refleja en absoluto las propiedades químicas de los péptidos, puesto que, exceptuando los extremos N- y C-terminal, los grupos amino y ácido de los aminoácidos, que tienen un carácter químico muy acusado, desaparecen en el polipéptido, siendo los grupos amida de los enlaces peptídicos relativamente estables y muy poco reactivos. Conviene por tanto considerar la estructura de los aminoácidos en su forma polimérica y no en su forma monomérica. En este artículo, en lo que sigue, y por simplicidad, nos referiremos a estas unidades estructurales como "aminoácidos".

2. Estructura química de los Péptidos y las Proteínas.

La estructura general de las proteínas y los péptidos derivados de ellas (por ejemplo a consecuencia de proteólisis o de digestión tríptica) se presenta en la figura 1 utilizando la notación simplificada de varillas. Recuérdese que en esta notación todos los puntos de unión de

dos o más varillas y los extremos de éstas representan átomos de carbono unidos a los átomos de hidrógeno necesarios para formar un total de cuatro enlaces en cada uno de ellos. La ventaja fundamental de esta notación, además de la simplicidad que supone, es que la presencia de heteroátomos (es decir, átomos diferentes al C y al H) se realza en la estructura, haciéndose así más patentes los puntos donde se concentra la reactividad química de la cadena.

Figura 1. Estructura química de los péptidos y las proteínas.



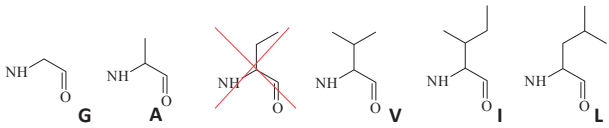
La estructura de un polipéptido típico se forma por la condensación de aminoácidos formándose puentes amida (o enlaces peptídicos), separados por el carbono α de cada aminoácido, de manera que al dibujar la cadena de forma horizontal las cadenas laterales de los aminoácidos se proyectan, alternativamente, hacia arriba y hacia abajo (Figura 1). Como hemos comentado anteriormente, es importante recordar que la unidad estructural de los péptidos no son los aminoácidos, sino el conjunto de átomos comprendidos entre el grupo amino y el grupo carbonilo de los puentes amídicos consecutivos (indicada por un corchete cuadrado en el Figura 1). Por esta razón, en espectrometría de masas no se utilizan los pesos moleculares de los aminoácidos, sino los pesos moleculares incrementales, correspondientes a esas unidades estructurales. La masa molecular del péptido se puede entonces calcular sumando los pesos incrementales más el peso de una molécula de agua. Los pesos incrementales resultan, además, extraordinariamente útiles para calcular las series de fragmentación que se generan al fragmentar los péptidos por espectrometría de masas en tándem.

La clasificación más habitual de los aminoácidos se basa en sus propiedades estructurales. En este capítulo usaremos una clasificación algo diferente, orientada a organizar su reactividad química, la cual depende en gran medida de sus propiedades iónicas, es decir de sus propiedades ácido-base en disoluciones acuosas, que están íntimamente relacionadas con su carácter nucleófilo y electrófilo. Consideraremos, por un lado, los aminoácidos no ionizables y, por otro los ionizables, que clasificaremos de acuerdo a su carácter ácido o básico.

3. Aminoácidos no ionizables.

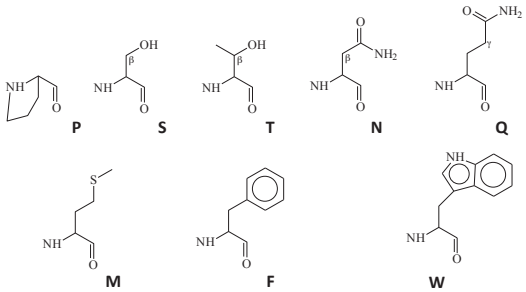
En este grupo se incluyen, entre otros, los aminoácidos glicina (G), alanina (A), valina (V), isoleucina (I) y leucina (L) (figura 2). Estos aminoácidos se construyen añadiendo sucesivamente grupos metileno ($-CH_2-$) a la estructura más sencilla de la glicina. Estos aminoácidos no poseen grupos ionizables ni heteroátomos y por tanto no son reactivos. Desempeñan un papel fundamentalmente estructural, permitiendo a la proteína conformar una variedad de superficies de carácter hidrofóbico. Conviene indicar que en este proceso se genera un aminoácido que no está codificado por el código genético (tachado en la figura 2). Podría especularse que tal vez este aminoácido existiera originalmente pero desapareciera en el curso de la evolución debido a la redundancia estructural que ofrecen los otros aminoácidos de este grupo. Los aminoácidos isoleucina y leucina son isobáricos, es decir tienen exactamente la misma masa molecular y, por tanto, son indistinguibles en un espectrómetro de masas, excepto en condiciones de fragmentación de alta energía.

Figura 2. Estructura de los grupos no ionizables (I).



A este grupo también pertenecen los aminoácidos prolina (P), serina (S), treonina (T), asparagina (N), glutamina (Q), metionina (M), fenilalanina (F) y triptófano (W) (figura 3). Estos aminoácidos no poseen grupos ionizables ni son reactivos, pero tienen heteroátomos o anillos aromáticos que les confieren ciertas particularidades estructurales o químicas.

Figura 3. Estructura de los grupos no ionizables (II).



La prolina es un caso estructural especial. En este aminoácido la cadena lateral se une al nitrógeno del enlace peptídico, formando una amida secundaria (el nitrógeno está unido a dos grupos) en un anillo de cinco átomos, siendo en realidad un iminoácido (figura 3). Es el único donde la cadena lateral se une a alguno de los componentes del enlace peptídico. En solución acuosa esta estructura tiene las mismas propiedades que el resto de los aminoácidos de este grupo, siendo un aminoácido sin reactividad química apreciable. Sin embargo en fase gaseosa la presencia de una amida secundaria, que tiene un carácter más básico que las primarias, tiene un efecto estructural sobre las series de fragmentación cuando los péptidos son fragmentados en un espectrómetro de masas, dando lugar a las series de prolina. Esta propiedad se discute en más detalle en el capítulo de fragmentación de péptidos.

La serina y la treonina son aminoácidos polares que contienen un grupo hidroxilo o alcohol acoplado a su estructura alifática (-OH). Este grupo confiere una mayor polaridad a estos aminoácidos, capaces de establecer puentes de hidrógeno con sus moléculas compañeras o con el agua, siendo aminoácidos muy solubles. El ángulo de enlace H-O-C es de 107°, disminuyendo un poco la polaridad de esta molécula con respecto a la del H_2O. La presencia de un átomo de oxígeno más electronegativo que el carbono y el hidrógeno polariza la molécula y la hace potencialmente susceptible de reactividad con otras moléculas. Otro aspecto de esta molécula es que el oxígeno dispone al menos de un par de electrones expuestos y sin compartir, constituyéndose en una base de Lewis (agente nucleofílico). El hidrógeno de la serina, al ser un alcohol primario, está menos firmemente unido al oxígeno y la salida de protones de la molécula está por ello más favorecida, siendo por tanto su acidez ligeramente será mayor que la de la treonina. Sin embargo, la acidez del grupo alcohol de serina y treonina es semejante al de las moléculas de agua que lo rodean y, por tanto, su reactividad en medio acuoso es prácticamente nula. Por esa razón estos grupos se consideran no reactivos en química de proteínas.

Una notable excepción es el grupo hidroxilo de la serina en el centro activo de un conjunto muy importante de proteasas, las serin-proteasas (entre las que se encuentran la tripsina, la quimotripsina y la elastasa), que por la influencia de los aminoácidos adyacentes está ionizado, constituyendo un potentísimo nucleófilo. En espectrometría de masas de péptidos la ionización se produce en fase gaseosa, y estos grupos manifiestan una tendencia a perder una molécula de H_2O durante el proceso de fragmentación, producida por la ruptura del enlace entre el C y el OH^- y recaptación de un H^+ del medio, con la consiguiente pérdida neutra por parte del fragmento.

La asparagina y la glutamina son las amidas derivadas del ácido aspártico y del glutámico, respectivamente, y su reactividad y propiedades químicas son comunes a todos los enlaces peptídicos (amidicos) de la cadena. Las amidas contienen un átomo de nitrógeno unido a un grupo carbonilo, los cuales comparten un par de electrones π. La resonancia en las amidas es una potente fuerza de estabilización y da lugar a numerosos efectos estructurales. A diferencia de la disposición piramidal de los enlaces en el amoniaco y en las aminas, los tres enlaces formados por el nitrógeno de una amida quedan en el mismo plano, de manera que los electrones

π se proyectan por encima y por debajo del plano, pudiendo ser compartidos por los átomos de carbono y oxigeno del grupo CO (carbonilo) adyacente. Son aminoácidos bastante polares que muestran una gran asociación intermolecular con las moléculas de agua del medio formando enlaces de hidrógeno a través del átomo de oxígeno negativo y los enlaces N—H, mucho más polarizados que en las aminas. Las amidas son sólo muy débilmente básicas, debido a la estabilización por resonancia producida por la interacción mesómera entre el doble enlace carbonílico y el par de electrones del átomo de nitrógeno. La protonación de las amidas sólo tiene lugar de manera significativa en condiciones fuertemente ácidas y ocurre sobre el átomo de oxígeno, ya que así la carga del catión resultante queda deslocalizada. En estas condiciones extremas las amidas experimentan reacciones de hidrólisis, en las que la cadena se descompone en sus aminoácidos componentes y la asparagina y glutamina dan lugar a sus derivados ácidos aspártico y glutámico. Por otra parte, la cesión de electrones por parte del nitrógeno estabiliza el grupo carbonilo de las amidas y disminuye la velocidad a la que los nucleófilos atacan al grupo carbonilo. Por todo ello, y a diferencia de sus homólogos carboxílicos, estos aminoácidos no son reactivos en condiciones convencionales y se consideran prácticamente inertes en química de proteínas. En fase gaseosa, no obstante, estos aminoácidos tienen una marcada tendencia a experimentar una pérdida neutra de amonio cuando están situados en el extremo N-terminal. Esta reacción se discute en el capítulo dedicado a la fragmentación de péptidos.

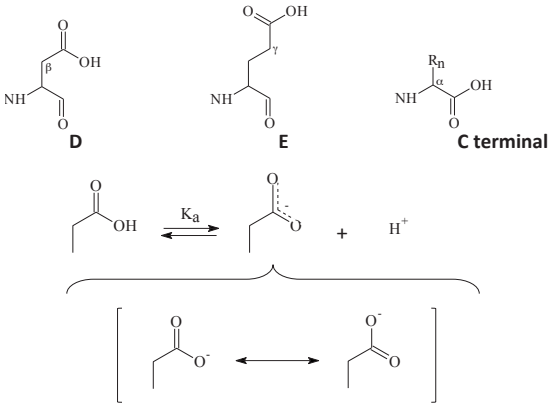
En este grupo se encuadran dos de los aminoácidos cuya cadena lateral posee un anillo aromático. La fenilalanina es una alanina que lleva unida un grupo fenilo y el triptófano tiene un grupo indol como cadena lateral. Ambos tienen en su estructura anillos aromáticos poco solubles en agua. El grupo benceno de la fenilalanina es bien conocido por su escasa reactividad química. El grupo indol del triptófano es un compuesto aromático heterocíclico que presenta una estructura bicíclica con un anillo de benceno unido a un anillo pirrólico. De acuerdo a las leyes de Hückell sobre la aromaticidad, el número de electrones π deslocalizados en el anillo debe ser 4n+2, siendo n un número entero. En el grupo indol cada átomo de carbono en el heterociclo aporta un electrón, de manera que el átomo de nitrógeno tiene que aportar un par de electrones para reunir los 10 electrones que necesita la estructura para ser aromática. Por tanto el átomo de nitrógeno presente en el anillo de pirrol no tiene electrones suficientes para dar lugar a un centro básico como en el caso de las aminas. Finalmente, la posición más reactiva del grupo, que es la posición 3, está ocupada por el enlace que le une al esqueleto del péptido, y por tanto no es accesible. El carácter aromático y la escasa basicidad del nitrógeno hacen al grupo indol muy estable y muy poco reactivo. Por estas razones la fenilalanina y el triptófano se consideran también aminoácidos escasamente reactivos desde el punto de vista de la química de proteínas.

La metionina pertenece a los aminoácidos no polares y tiene en la cadena lateral un grupo tioéter. Los tioéteres son similares a los éteres, conteniendo un átomo de azufre en vez del átomo de oxígeno. Debido a que el oxígeno y el azufre pertenecen al grupo de los anfígenos en la tabla periódica, las propiedades químicas y reactividad de los éteres y tioéteres tienen algunos puntos en común pero en general la reactividad de los primeros es mayor al ser mayor su acidez y capacidad de oxidación. La oxidación de la metionina en el entorno acuoso se produce con relativa facilidad; los tioéteres se oxidan formando sulfóxidos (R-S(=O)-R) en condiciones prácticamente espontáneas. A su vez los sulfóxidos pueden ser oxidados formando sulfonas ($R\text{-}S(=O)_2\text{-}R$) en presencia de oxidantes. La oxidación espontánea de la metionina es una reacción muy importante que hay que considerar a la hora de identificar el péptido por espectrometría de masas o de diseñar péptidos proteotípicos, ya que la masa del péptido incrementa en la masa correspondiente a un átomo de oxígeno. Ésta es prácticamente la única reacción que produce la cadena lateral de éste aminoácido, ya que aunque los tioéteres son potencialmente nucleófilos, esta propiedad no se manifiesta en disolución acuosa por la competencia de las moléculas de agua, tal y como sucede con la serina y la treonina.

4. Aminoácidos ácidos.

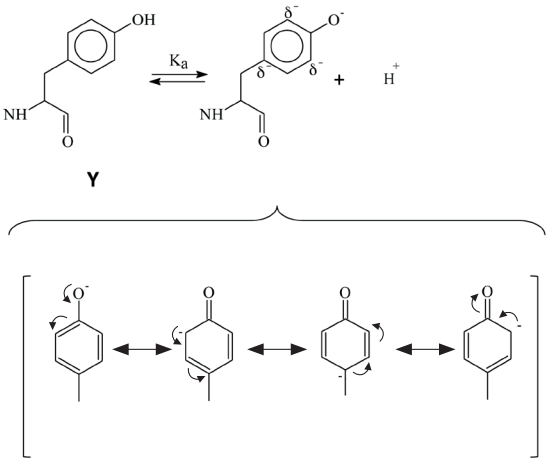
En este conjunto se incluyen los grupos β– y γ–carboxilo del ácido aspártico (D) y del glutámico (E), respectivamente, el grupo α–carboxilo del extremo C-terminal (CT), que debe incluirse como grupo reactivo con personalidad propia, y los grupos fenol de la tirosina (Y) y tiol de la cisteína (C) (Figuras 4-6). El grupo carboxilo se compone de un grupo carbonilo (C=O) y un grupo hidroxilo (-OH). El átomo de carbono usa tres orbitales híbridos sp^2, situados en el mismo plano, para unirse al oxígeno del grupo OH, al oxígeno carbonilo y al resto de la cadena. El orbital *p* restante del carbono forma un enlace π con un orbital *p* del oxígeno carbonilo. En el grupo hidroxilo, el oxígeno es más electronegativo que el hidrógeno, polarizándose el enlace, que se debilita y tiende a perder un protón. La carga negativa resultante se deslocaliza simétricamente al compartirse 4 electrones π entre los dos átomos de oxígeno y el átomo de carbono. Esta estructura puede también explicarse en términos de formas resonantes donde se sitúa alternativamente la carga (figura 4). Esta estabilización por resonancia o deslocalización electrónica provoca que la base conjugada del ácido sea muy estable, lo que explica el carácter ácido de este grupo. Aunque los tres grupos tienen una acidez parecida, existe una ligera diferencia de acidez entre el C-terminal, el aspártico y el glutámico, debida al efecto inductivo del grupo alquílico, que es más acusado al aumentar el número de carbonos de la cadena. Por esa razón, el grupo α–carboxilo (CT) es ligeramente más ácido que el β (aspártico) y éste que el γ (glutámico). Los tres grupos tienen por ello carácter electrófilo, y pueden reaccionar con grupos nucleófilos existentes en el entorno.

Figura 4. Estructura de los grupos ácidos: D, E y CT.



La tirosina tiene un grupo lateral fenol, que es un alcohol aromático con un marcado carácter ácido. Los fenoles son mucho más ácidos que los alcoholes de cadena abierta (serina y treonina) porque cuando se libera el protón el anillo aromático deslocaliza de manera efectiva la carga negativa extra del oxígeno (figura 5). Por esta razón la constante de ionización del grupo ácido (K_a) es del orden de 10^{-10}, correspondiente a un ácido considerablemente más fuerte que el agua, aunque más débil que los ácidos carboxílicos. Por esta razón la tirosina comienza a ionizarse, adquiriendo carga negativa, cuando el pH del medio aumenta por encima de 9. La tirosina en forma ionizada actúa como nucleófilo, compitiendo con la cisteína, la lisina y el grupo N-terminal; por ello la reactividad de la tirosina puede controlarse ajustando adecuadamente el pH del medio.

Figura 5. Grupo ácido Y.

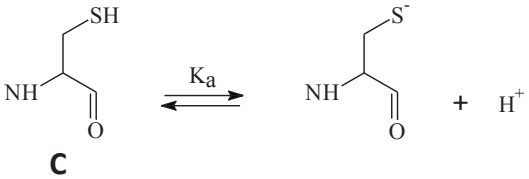


El anillo aromático de la tirosina tiene también tendencia a dar reacciones de sustitución aromática electrófila,

en las que un electrófilo desplaza a uno o más de los átomos de hidrógenos del anillo aromático. La estructura resonante de la forma ionizada (figura 5) ayuda a entender por qué los carbonos situados en posición orto y para tienen una mayor densidad de carga negativa, siendo más reactivos al ataque electrófilo que en posición meta. Un ejemplo típico de esta reacción es la iodinación de la tirosina que tiene lugar para formar la hormona tiroidea (tetraiodotironina).

La cisteína contiene un átomo de azufre formando un grupo tiol (-SH). Éste es un grupo análogo al grupo alcohol (-OH), pero al ser el azufre menos electronegativo que el oxigeno, sus electrones externos ocupan orbitales más extendidos en el espacio y por consiguiente los átomos de azufre forman enlaces de hidrogeno más débiles que los átomos de oxigeno. Por la misma razón el grupo tiol es mucho más ácido que el grupo alcohol, siendo su pK_a alrededor de 8.5 (figura 6). Por tanto, a pH neutro la cisteína está protonada, pero comienza a ionizarse cuando el pH se hace ligeramente básico, aumentando por encima de 8. Los iones tiolato así formados son nucleófilos extremadamente potentes, siendo la cisteína el más reactivo de todos los aminoácidos. Por la misma razón la reactividad de la cisteína depende críticamente del pH, que debe ser controlado cuidadosamente en reacciones que impliquen a la cadena lateral de este aminoácido. Debido a la alta reactividad del grupo tiol, conviene protegerlo o bloquearlo utilizando agentes alquilantes para que no haya reacciones secundarias no controladas. La cisteína es un aminoácido que se oxida con facilidad. Tal vez la reacción de oxidación más conocida es la formación de puentes disulfuro entre dos cisteínas, formando un grupo cistina, que se favorece a pH ligeramente básico. En presencia de oxidantes, la cisteína da lugar, sucesivamente, a los ácidos sulfénico (-SOH), sufónico ($-SO_2H$) y sulfínico ($-SO_3H$).

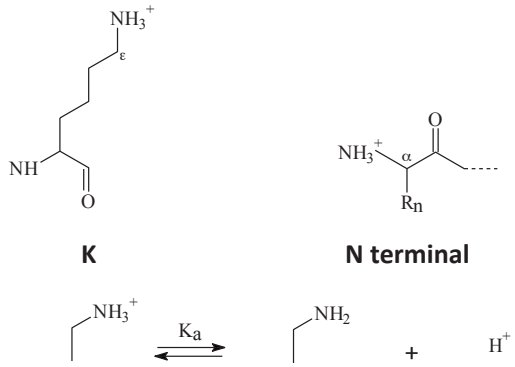
Figura 6. Estructura de la cisteína.



5. Aminoácidos básicos.

Este conjunto de aminoácidos incluye el grupo ε-amino de la lisina (K), el grupo α-amino del extremo N-terminal (NT), que de nuevo debe incluirse aquí como grupo reactivo con personalidad propia, el grupo guanidinio de la arginina (R) y el grupo imidazol de la histidina (H). La lisina y el extremo NT contienen aminas primarias (figura 7), en las cuales los cinco electrones de la última capa del átomo de nitrógeno se distribuyen en cuatro orbitales *sp*^{3} idénticos formando una estructura tetraédrica en la que uno de los orbitales contiene dos electrones sin compartir. Este orbital confiere al grupo amino un marcado carácter nucleófilo, y una tendencia a captar protones de la disolución, comportándose así como una base de Lewis (desde el punto de vista termodinámico), cuyo pK_a se sitúa en torno a 11 para la lisina y en torno a 8 para el extremo NT (Figura 7). La diferencia de basicidad se explica por el efecto inductivo de los grupos alquílicos, ya

Figura 7. Grupos básicos: K, NT.

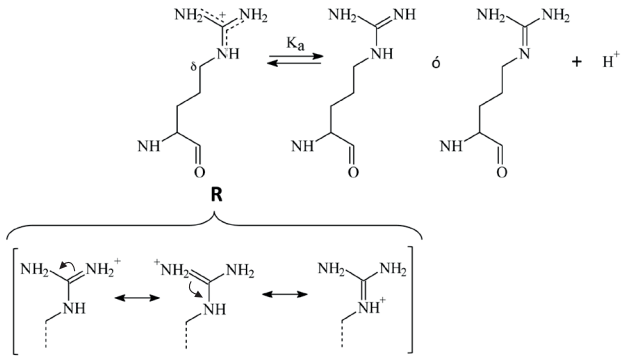


que al ser éstos donadores de electrones, tienden a incrementar la densidad electrónica en el átomo de nitrógeno haciendo más asequible el par electrónico. Este efecto aumenta cuando aumenta el número de eslabones del grupo alquílico. El extremo NT está unido al carbono α del aminoácido y que a su vez tiene un grupo carbonilo contiguo, mientras que el grupo alquilo de la lisina tiene cinco carbonos alifáticos.

Es importante resaltar aquí que la protonación de estos grupos amino destruye su carácter nucleófilo, de manera que la reactividad de estos grupos puede controlarse ajustando cuidadosamente el pH. A pHs por encima de 11 los dos grupos tienen carácter nucleófilo y por lo tanto tienden a reaccionar con electrófilos; sin embargo, cuando el pH se disminuye por debajo de 7.5 esta reactividad desaparece. A pHs intermedios es posible atenuar parcialmente la nucleofilia del extremo N terminal, permitiendo así que la lisina reaccione preferentemente.

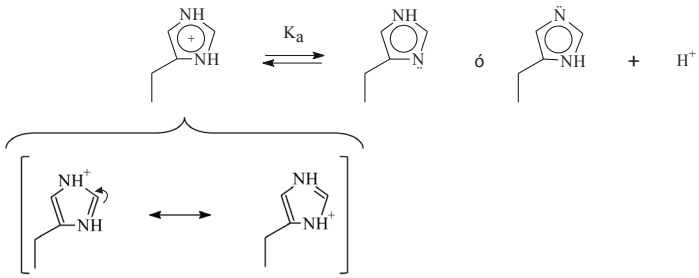
La arginina es un caso aparte entre los aminoácidos. Está formada por un grupo guanidinio unido al carbono δ. El grupo guanidinio contiene tres grupos amino, y tiene la particularidad de que en su forma protonada los tres grupos forman un triángulo plano donde tres de los orbitales de los nitrógenos tienen hibridación sp^2 y se sitúan en torno a un único átomo de carbono central. Los orbitales *p* proyectados verticalmente en relación a este plano permiten una considerable estabilización de la carga del protón, que es compartida por los tres núcleos de nitrógeno y el núcleo de carbono. En la figura 8 se muestran las formas canónicas del híbrido de resonancia resultante. Debido a esta conformación estructural, el grupo guanidino de la arginina se distingue de los demás a causa de su pronunciada basicidad ($pK_a \approx 12,5$), que es tan extrema que el grupo permanece completamente protonado no sólo a pH neutro, sino también en condiciones en las cuales el grupo α-amino (pK 8-11) se comporta como nucleófilo.

Figura 8. Grupo básico R.



El aminoácido histidina contiene un grupo imidazol en el carbono β (Figura 9). El imidazol es un anillo aromático de cinco átomos que contiene dos átomos de nitrógeno. Los electrones de los nitrógenos y de los carbonos del anillo se sitúan en orbitales sp^2, formando triángulos planos, y en orbitales p_z verticales donde se sitúan los electrones π. La hibridación genera orbitales deslocalizados por encima y por debajo del anillo. De acuerdo a la regla de Hückell, para preservar la aromaticidad el anillo debe tener 6 electrones π; tres de ellos son suministrados por los átomos de carbono, mientras que dos de ellos son suministrados por uno de los átomos de nitrógeno, el cual utiliza el electrón sobrante en el orbital sp^2 para formar un enlace con un átomo de hidrógeno y, por tanto, no tiene carácter básico. Sin embargo, el

Figura 9. Grupo básico H.



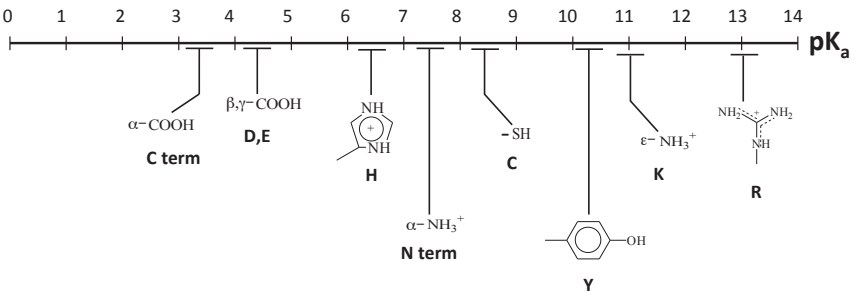
segundo átomo de nitrógeno sólo tiene que aportar un electrón **π,** y por tanto dispone de dos electrones sobrantes que forman un orbital *sp*^{2} sin compartir y que se proyecta hacia el exterior en el plano del anillo. De forma análoga a los grupos amino, esta pareja de electrones confiere a este aminoácido un carácter nucleófilo, mostrando avidez por atacar a reactivos electrófilos por los protones del medio, actuando así como un grupo básico. Sin embargo, el poder de estabilización de la carga positiva del grupo imidazol, debido a la deslocalización de la carga por el anillo (figura 9), es mayor que en el caso de las aminas y por tanto este grupo tiene un carácter más ácido (es decir, menos básico) que la lisina. Debido a su menor pK_a, que se sitúa en torno a 6.5, este grupo permanece en forma protonada a pHs ligeramente básicos, en los que la lisina mantiene su carácter nucleófilo, siendo así posible dirigir la reactividad selectiva de esta última.

Es importante indicar que al haber dos átomos de nitrógeno no situados simétricamente respecto al grupo metileno que une el grupo imidazol al carbono **α** del aminoácido, la histidina puede configurarse en dos estructuras estables posibles (Figura 9), existiendo un equilibrio dinámico entre ellas en disolución acuosa. La presencia de dos átomos de nitrógeno hace también posible que la histidina capte un protón a través de uno de ellos, cediéndolo inmediatamente a través del átomo de nitrógeno en el extremo opuesto de la molécula actuando, así, como un "transbordador de protones".

6. PH y Nucleofilia.

Las especies protonada y no protonada de los aminoácidos que contienen grupos ionizables están en equilibrio en solución acuosa. La desprotonación de los grupos nucleófilos hace que los electrones en la capa de valencia del átomo más electronegativo se distancien de los núcleos que están apantallados por electrones en orbitales más internos, siendo así más susceptibles de ser intercambiados a centros donde la densidad de carga negativa sea menor (electrófilos) y, por tanto aumenta su carácter nucleófilo. Por esta razón el carácter nucleófilo aumenta, en general, al aumentar el pH del medio. Sin embargo, el efecto del pH sobre la protonación depende críticamente del pK_a de cada una de las cadenas laterales, siendo necesario que el pH sea cercano o inferior al pK_a para que el grupo se protone. En la figura 10 se esquematizan los valores normales de pK_a de cada uno de los aminoácidos; este esquema es muy útil para comprender e interpretar la reactividad nucleófila de las cadenas laterales. Por ejemplo, a pH neutros o ácidos los grupos amino, la cisteína y la tirosina están protonados y por tanto son poco reactivos; sin embargo la histidina comienza a ser nucleófila a pH ligeramente ácidos, a los cuales todavía no está totalmente protonada. De la misma manera, a pHs ligeramente básicos se comienza a notar el carácter nucleófilo de la cisteína antes que el de la tirosina, y sólo se hace patente el de la lisina cuando el pH se hace marcadamente básico. Y debido a su pK_a extremo la arginina puede considerarse como un grupo no nucleófilo.

Figura 10. Distribución de pK_as de los aminoácidos ionizables a lo largo del rango de pH.



7. El punto isoeléctrico.

El punto isoeléctrico (pI) es un parámetro que evalúa la influencia del pH sobre la carga del péptido o de la proteína. Se define como el pH en el que la carga neta de la proteína o el péptido es cero. El punto isoeléctrico depende de los pK_as de las cadenas laterales de los aminoácidos que componen el péptido y la proteína. Sin embargo el efecto de los pK_as sobre el punto isoeléctrico no es trivial y hay que considerar la contribución de cada grupo a la carga de forma separada.

Supongamos una proteína que tiene *N* grupos aspártico, de los cuales N_{A^-} y N_{AH} son el número de ellos que están ionizados y sin ionizar, respectivamente. Sean $[AH]$ y $[A^-]$ las concentraciones de dichos grupos. La proporción relativa de estas especies depende del *pH* del medio de acuerdo al equilibrio ácido-base correspondiente:

$$AH \rightleftharpoons A^- + H^+$$

De forma que

$$K_a = \frac{[A^-] \cdot [H^+]}{[AH]} = \frac{N_{A^-} \cdot [H^+]}{N_{AH}}$$

Siendo K_a la constante de disociación del grupo ácido del aspártico. Puesto que $N = N_{A^-} + N_{AH}$, despejando N_{A^-} se obtiene

$$N_{A^-} = N \frac{K_a}{K_a + [H^+]}$$

Dividiendo entre K_a y puesto que $pK_a = -log\ (K_a)$, se obtiene finalmente

$$N_{A^-} = \frac{N}{1+10^{(pK_a - pH)}} \quad [2]$$

La misma expresión se obtiene para el resto de los aminoácidos que al ionizarse se cargan negativamente, es decir, glutámico, cisteína, tirosina y el grupo ácido del extremo C-terminal.

En el caso de los aminoácidos básicos, cuyo equilibrio ácido-base tiene lugar mediante protonación (adquieren carga positiva):

$$B + H^+ \rightleftharpoons BH^+$$

$$K_a = \frac{[B] \cdot [H^+]}{[BH^+]} = \frac{N_B \cdot [H^+]}{N_{BH^+}} \qquad \text{y} \qquad N = N_B + N_{BH+}$$

De donde

$$N_{BH^+} = N \frac{[H^+]}{K_a + [H^+]}$$

Con lo que el número de cargas positivas queda finalmente expresado como sigue

$$N_{BH^+} = \frac{N}{1+10^{(pH - pK_a)}} \quad [3]$$

Para evaluar el efecto de modificaciones post-traduccionales sobre la carga de las cadenas laterales tomaremos como caso tipo las producidas por fosforilación. En el caso de aminoácidos fosforilados el equilibrio ácido-base, a *pH* inferiores a 11 y superiores a 3, viene dado por la segunda reacción de ionización del grupo fosfato:

$$-OPO_3^{2-} + H^+ \rightleftharpoons -OPO_3H^-$$

De forma que

$$N_{-OPO_3H^-} = \frac{N}{1+10^{(pH - pK_{ap})}} \qquad \text{y} \qquad N_{-OPO_3^{2-}} = N_B - N_{-OPO_3H^-}$$

El número total de cargas negativas debidas al grupo fosfato de los grupos fosforilados viene entonces dado por:

$$z_p = N_{-OPO_3H^-} + 2N_{-OPO_3^{2-}} = 2N - N_{-OPO_3H^-}$$

$$z_p = N\left[2 - \frac{1}{1+10^{(pH - pK_{ap})}}\right] \qquad [4]$$

De forma análoga puede procederse con otras modificaciones postraduccionales.

Integrando las contribuciones de los grupos ácidos, básicos y fosforilados (o en general, modificados), la carga neta de la proteína z_Q viene entonces dada por la siguiente expresión

$$z_Q = \sum_i \frac{N_i}{1+10^{(pH - pK_{ai})}} - \sum_j \frac{N_j}{1+10^{(pK_{aj} - pH)}} - N_p\left[2 - \frac{1}{1+10^{(pH - pK_{ap})}}\right] \qquad [2]\text{-}[3]\text{-}[4]$$

Donde el sumatorio en *i* se extiende a los grupos que adquieren carga positiva al ionizarse (grupos básicos: R, K, H y N-terminal, de acuerdo con al ecuación [2]), el sumatorio en *j* se extiende a los grupos que adquieren carga negativa (grupos ácidos: D, E, Y, C y C-terminal de acuerdo con la ecuación [3]), el subíndice *p* hace referencia a los grupos fosforilados y pK_{ap} corresponde al segundo equilibrio de ionización del grupo fosfato (de acuerdo con la ecuación [3]).

A modo de ejemplo, se ha calculado el punto isoeléctrico del péptido TDMEAEHKVLKHRLIHCDWECSKTYPRK que contiene 2 aspárticos, 3 glutámicos, 3 histidinas, 2 cisteínas, 4 lisinas, 1 tirosina y 2 argininas. Obviamente, para el cómputo se han de tener también en cuenta los extremos NT y CT, siempre que el péptido no sea cíclico. El resto de los aminoácidos que componen el péptido no es importante para el cálculo del pI, ya que no son ionizables.

En la tabla 1, las cargas aportadas por cada grupo se han calculado de acuerdo a las ecuaciones anteriores en función del pH (primera columna) y los pK_as aproximados de los aminoácidos, indicados en la primera fila (en rojo). La carga neta se indica en la última columna. Podemos observar cómo al aumentar el pH se produce una progresiva desprotonación del péptido. Según estos cálculos, el pI de este péptido está aproximadamente entre 8 y 8.25 (resaltado en amarillo). En la figura 11 se representa la suma de todas las cargas positivas (línea superior, en negro) y todas las cargas negativas (línea inferior, en negro) en función del pH, así como la carga neta del péptido (línea roja).

Los puntos de inflexión de las curvas son debidos a los cambios de carga de los grupos ionizables que, al desprotonarse al aumentar el pH, pierden su carga positiva y corresponden a los pKa respectivos. Así, a pH en torno a 3 y 4 se observa el primer salto por aparición de cargas negativas en el grupo carboxilo del extremo C-terminal y en los residuos de aspártico y glutámico. A un pH entre 6 y 9 se observa la pérdida de protones en la histidina, el grupo amino del extremo N-terminal y la cisteína. Y a pH 10 se comienza a observar la ionización de la tirosina y la pérdida de un protón por parte de la lisina. Los residuos de arginina contribuyen con una carga positiva que sólo se comienza a perder cuando el pH se aproxima a 10.

Figura 11. Variación de la carga del péptido en función del pH.

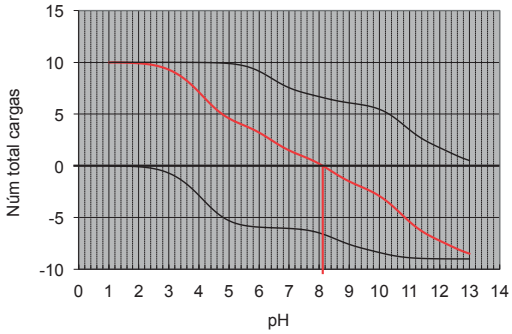


Tabla 1. Cálculo teórico del pI.

pKa	3.2	4.0	4.3	6.4	7.5	8.5	10.1	10.8	12.5	Suma de Cargas Negativas	Suma de Cargas Positivas	Total
C. Lateral	Cterminal	E	D	H	Nterminal	C	Y	K	R			
Nº de aa	1	2	3	3	1	2	1	4	2			
Rango de pH												
1	-0.01	0	0	3	1	0	0	4	2	-0.01	10	9.99
1.25	-0.01	0	0	3	1	0	0	4	2	-0.02	10	9.98
1.5	-0.02	-0.01	0	3	1	0	0	4	2	-0.03	10	9.97
1.75	-0.03	-0.01	-0.01	3	1	0	0	4	2	-0.05	10	9.95
2	-0.06	-0.02	-0.01	3	1	0	0	4	2	-0.09	10	9.91
2.25	-0.10	-0.03	-0.03	3	1	0	0	4	2	-0.16	10	9.84
2.5	-0.17	-0.06	-0.05	3	1	0	0	4	2	-0.27	10	9.73
2.75	-0.26	-0.11	-0.08	3	1	0	0	4	2	-0.45	10	9.55
3	-0.39	-0.18	-0.14	3	1	0	0	4	2	-0.71	10	9.29
3.25	-0.53	-0.30	-0.25	3	1	0	0	4	2	-1.08	10	8.92
3.5	-0.67	-0.48	-0.41	3	1	0	0	4	2	-1.56	10	8.44
3.75	-0.78	-0.72	-0.66	2.99	1	0	0	4	2	-2.16	10	7.84
4	-0.86	-1.00	-1.00	2.99	1	0	0	4	2	-2.86	10	7.13
4.25	-0.92	-1.28	-1.41	2.98	1	0	0	4	2	-3.61	10	6.38
4.5	-0.95	-1.52	-1.84	2.96	1	0	0	4	2	-4.31	10	5.68
4.75	-0.97	-1.70	-2.21	2.93	1	0	0	4	2	-4.89	10	5.11
5	-0.98	-1.82	-2.50	2.89	1	0	0	4	2	-5.30	10	4.69
5.25	-0.99	-1.89	-2.70	2.80	0.99	0	0	4	2	-5.58	9.79	4.21
5.5	-1	-1.94	-2.82	2.66	0.99	0	0	4	2	-5.76	9.65	3.89
5.75	-1	-1.97	-2.90	2.45	0.98	0	0	4	2	-5.86	9.43	3.57
6	-1	-1.98	-2.94	2.15	0.97	-0.01	0	4	2	-5.93	9.11	3.19
6.25	-1	-1.99	-2.97	1.76	0.95	-0.01	0	4	2	-5.97	8.70	2.74
6.5	-1	-1.99	-2.98	1.33	0.91	-0.02	0	4	2	-5.99	8.24	2.24
6.75	-1	-2	-2.99	0.93	0.85	-0.03	0	4	2	-6.02	7.77	1.75
7	-1	-2	-2.99	0.60	0.76	-0.06	0	4	2	-6.05	7.36	1.31
7.25	-1	-2	-3	0.37	0.64	-0.11	0	4	2	-6.10	7.01	0.91
7.5	-1	-2	-3	0.22	0.50	-0.18	0	4	2	-6.18	6.72	0.53
7.75	-1	-2	-3	0.13	0.36	-0.30	0	4	2	-6.30	6.48	0.18
8	*-1*	*-2*	*-3*	*0.07*	0.24	*-0.48*	*-0.01*	*3.99*	*2*	*-6.49*	6.31	-0.18
8.25	*-1*	*-2*	*-3*	*0.04*	0.15	*-0.72*	*-0.01*	*3.99*	*2*	*-6.73*	6.18	-0.55
8.5	-1	-2	-3	0.02	0.09	-1.00	-0.02	3.98	2	-7.02	6.09	-0.93
8.75	-1	-2	-3	0.01	0.05	-1.28	-0.04	3.96	2	-7.32	6.03	-1.29
9	-1	-2	-3	0.01	0.03	-1.52	-0.07	3.94	2	-7.59	5.98	-1.62
9.25	-1	-2	-3	0	0.02	-1.70	-0.12	3.89	2	-7.82	5.91	-1.91
9.5	-1	-2	-3	0	0.01	-1.82	-0.20	3.81	2	-8.02	5.82	-2.20
9.75	-1	-2	-3	0	0.01	-1.89	-0.31	3.67	2	-8.20	5.68	-2.53
10	-1	-2	-3	0	0	-1.94	-0.44	3.45	1.99	-8.38	5.45	-2.93
10.25	-1	-2	-3	0	0	-1.97	-0.59	3.12	1.99	-8.55	5.11	-3.44
10.5	-1	-2	-3	0	0	-1.98	-0.72	2.66	1.98	-8.70	4.65	-4.05
10.75	-1	-2	-3	0	0	-1.99	-0.82	2.12	1.97	-8.81	4.08	-4.73
11	-1	-2	-3	0	0	-1.99	-0.89	1.55	1.94	-8.88	3.49	-5.40
11.25	-1	-2	-3	0	0	-2	-0.93	1.05	1.89	-8.93	2.94	-5.99
11.5	-1	-2	-3	0	0	-2	-0.96	0.67	1.82	-8.96	2.48	-6.48
11.75	-1	-2	-3	0	0	-2	-0.98	0.40	1.70	-8.98	2.10	-6.88
12	-1	-2	-3	0	0	-2	-0.99	0.24	1.52	-8.99	1.76	-7.23
12.25	-1	-2	-3	0	0	-2	-0.99	0.14	1.28	-8.99	1.42	-7.58
12.5	-1	-2	-3	0	0	-2	-1	0.08	1.00	-9	1.08	-7.92
12.75	-1	-2	-3	0	0	-2	-1	0.04	0.72	-9	0.76	-8.23
13	-1	-2	-3	0	0	-2	-1	0.03	0.48	-9	0.51	-8.49

8. Estimación aproximada del punto isoeléctrico.

El punto isoeléctrico puede también calcularse de forma aproximada partiendo de la ecuación ácido/base de Henderson-Hasselbach, expresada de la siguiente forma

$$pK_a - pH = log_{10} \frac{[A^-]}{[AH]}$$

De esta ecuación se deduce que, cuando el pH iguala al pK_a, el logaritmo de la derecha se hace nulo y, por tanto, la proporción entre las especies desprotonada y protonada debe ser igual a la unidad. Es decir, su concentración debe ser idéntica y podemos afirmar que la mitad de las moléculas están protonadas y la otra mitad desprotonadas (50% de protonación).

De forma similar, cuando el pH es una unidad superior al pK_a, el logaritmo debe valer uno y por tanto la proporción de la especie protonada debe tener una concentración diez veces inferior que la de la especie desprotonada (10% de protonación). Análogamente, cuando el pH es una unidad inferior al pKa, la especie protonada debe ser unas diez veces más abundante que la especie desprotonada (90% de protonación). De la misma manera se puede estimar el porcentaje de protonación cuando la diferencia entre el pH y el pK_a es de dos, tres o más unidades. Esta intuitiva relación entre el pH, el pK_a y el porcentaje de protonación del grupo se representa en la figura 12.

Este esquema permite estimar de forma aproximada la contribución de cada grupo a la carga total, en función del pH y, por tanto, llevar a cabo un cálculo aproximado del punto isoeléctrico del péptido o de la proteína. Para ilustrar este método, en la tabla 2 se ha calculado el punto isoeléctrico del péptido del ejemplo anterior. Se observa que el punto isoeléctrico estimado queda finalmente entre 8 y 8,5, lo que supone una excelente concordancia con el cálculo analítico.

Figura 12. Contribución aproximada de especies protonadas.

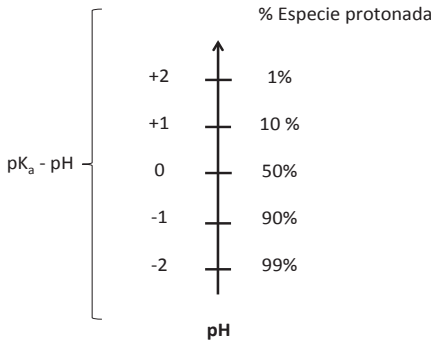


Tabla 2. Cálculo aproximado del pI.

pKa	3'5		4'5		6.5		8	8.5		10		11		13		Suma de cargas Negativas	Suma de cargas Positivas	Total
C. Lateral	Cterminal		D,E		H		Nterminal	C		Y		K		R				
Nº de aa	1		5		3		1	2		1		4		2				
Rango de pH																		
0	0	0		1	x3	1	0		0		1	x4	1	x2		0	10	10
1	0	0		1	x3	1	0		0		1	x4	1	x2		0	10	10
1.5	-0.01	0		1	x3	1	0		0		1	x4	1	x2		-0.01	10	9.99
2.5	-0.1	-0.01	x5	1	x3	1	0		0		1	x4	1	x2		-0.15	10	9.85
3.5	-0.5	-0.1	x5	1	x3	1	0		0		1	x4	1	x2		-1	10	9
4.5	-0.9	-0.5	x5	0.99	x3	1	0		0		1	x4	1	x2		-3.4	10	6.6
5.5	-0.99	-0.9	x5	0.9	x3	1	0		0		1	x4	1	x2		-5.49	9.97	4.48
6.5	-1	-0.99	x5	0.5	x3	0.9	-0.01	x2	0		1	x4	1	x2		-5.97	9.7	3.73
7.5	-1	-1	x5	0.1	x3	0.75	-0.1	x2	0		1	x4	1	x2		-6.2	8.4	2.2
8	*-1*	*-1*	*x5*	*0.025*	*x3*	*0.5*	*-0.25*	*x2*	*0*		*1*	*x4*	*1*	*x2*		*-6.5*	*7.05*	*0.55*
8.5	*-1*	*-1*	*x5*	*0.01*	*x3*	*0.25*	*-0.5*	*x2*	*0*		*1*	*x4*	*1*	*x2*		*-7*	*6.575*	*-0.425*
9	-1	-1	x5	0		0.1	-0.75	x2	-0.1	x1	0.99	x4	1	x2		-7.6	6.28	-1.32
9.5	-1	-1	x5	0		0.025	-0.9	x2	-0.25	x1	0.95	x4	1	x2		-8.05	6.06	-1.99
10	-1	-1	x5	0		0.001	-0.95	x2	-0.5	x1	0.9	x4	1	x2		-8.4	5.825	-2.575
11	-1	-1	x5	0		0	-1	x2	-0.9	x1	0.5	x4	0.99	x2		-8.9	5.601	-3.299
12	-1	-1	x5	0		0	-1	x2	-0.99	x1	0.1	x4	0.9	x2		-8.99	3.98	-5.01
13	-1	-1	x5	0		0	-1	x2	-1	x1	0.01	x4	0.5	x2		-9	2.2	-6.8
14	-1	-1	x5	0		0	-1	x2	-1	x1	0		0.1	x2		-9	1.04	-7.96

Cromatografía. Análisis y separación de Péptidos y Proteínas

03

Gabriel Padrón Palomares[1]

[1] Centro de Ingeniería Genética y Biotecnología, La Habana, Cuba.

1. Introducción.

La cromatografía es en estos momentos una de las técnicas más poderosas y versátiles con que cuentan la química, la biología y la bioquímica para la separación, análisis y purificación de mezclas complejas, tanto en el laboratorio como en la industria.

El creador de la cromatografía fue el botánico ruso Mijaíl Tsvet, nacido en Italia en 1872. Tsvet presentó los resultados que dieron origen a esta técnica en 1901 en un congreso en San Petersburgo y la palabra "cromatografía" apareció escrita por primera vez en 1906. Este nombre respondió a su primera aplicación, que fue la separación de pigmentos en una matriz de carbonato de calcio(1). El desarrollo de la cromatografía recibió un impulso decisivo con el trabajo de los científicos ingleses Archer Martin y Richard Synge durante los años 40 y 50. Ambos recibieron el premio Nobel de Química en 1952.

Con el nombre de cromatografía se agrupan varias técnicas que pueden ser clasificadas atendiendo a diferentes criterios. Entre estas tenemos, sin ningún intento de clasificación: cromatografía de papel, cromatografía de columna, cromatografía de capa fina, cromatografía gaseosa, cromatografía de fluidos supercríticos. En todos los casos los componentes de una mezcla se separan por su interacción con una fase estacionaria y una fase móvil.

En este capítulo solo se tratará la cromatografía líquida, donde la fase estacionaria es un sólido contenido en una columna y la fase móvil es un líquido. En particular nos referiremos al análisis y separación de péptidos y proteínas con el uso de la Cromatografía Líquida de Alta Eficacia (HPLC, del inglés *High Performance Liquid Chromatography*).

HPLC es el resultado de la optimización de los parámetros que determinan la eficiencia de la separación cromatográfica, especialmente los relacionados con la matriz, o gel cromatográfico, y la columna. Es muy importante destacar que la "P" de HPLC no debe relacionarse con la presión, como se hace erróneamente en numerosos artículos científicos. La presión no es un parámetro que contribuya a la eficiencia de la separación, sino el resultado de la optimización de los parámetros que la gobiernan, fundamentalmente el tamaño de las partículas del material cromatográfico.

El surgimiento de la HPLC fue una verdadera revolución en el campo de la cromatografía. En el caso de los péptidos y proteínas, los geles cromatográficos existentes eran muy ineficientes y

para tratar de mejorar la eficiencia era necesario utilizar columnas muy largas y flujos muy lentos, con lo cual las corridas cromatográficas podían durar varias horas. Esta situación cambió radicalmente con el advenimiento de la HPLC. En la actualidad muchos de los análisis por HPLC duran solo algunos minutos, aunque la separación de mezclas complejas puede durar más de una hora.

Los geles cromatográficos que no son considerados para HPLC también se han desarrollado ampliamente y ya no existe una diferencia tan grande como en los primeros tiempos. Esta tendencia también se observa en los equipos empleados para ambos tipos de cromatografía, donde apenas se notan diferencias en la actualidad, incluido el precio.

Desde el punto de vista de los mecanismos de separación no hay diferencia alguna entre HPLC y cromatografía "tradicional". Así en HPLC también podemos tener cromatografía de intercambio iónico, de separación por talla (filtración en gel), de fase reversa, de interacción hidrofóbica, de afinidad, etc.

2. Fase estacionaria en HPLC.

Como se verá más adelante, uno de los parámetros que más influyen en la eficiencia de una columna es el diámetro de la partícula. Actualmente, la mayoría de las columnas de HPLC tienen partículas con diámetro promedio entre 5 µm y 1.7 µm. Las de 5 µm requieren presiones de más de 100 bar y las de 1.7 µm tienen que ser eluídas a presiones próximas a 1 000 bar.

Por tanto, la matriz cromatográfica tiene que estar basada en materiales resistentes a esas presiones. Los más usados son el gel de sílice y los copolímeros de poliestireno y divinilbenceno.

El gel de sílice ha sido el material más usado como fase estacionaria en la cromatografía desde sus inicios. Está constituido por un enrejado de óxido de silicio y en su superficie se encuentran los grupos silanoles (Si-OH). Es muy resistente y tiene excelentes propiedades como material adsorbente. Desde el punto de vista químico, los grupos silanoles permiten la modificación de la superficie con variadas alternativas.

El divinilbenceno permite el entrecruzamiento entre las cadenas de poliestireno, lo que incrementa la resistencia del polímero resultante. También los polímeros de acrilamida se usan en la preparación de matrices para HPLC.

El gel de sílice puede ser usado a pH entre 2 y 7. A mayor pH puede solubilizarse y dañar seriamente la columna. Sin embargo, se encuentran disponibles en el mercado algunas columnas de HPLC basadas en sílice que son resistentes a pH básico. Las columnas basadas en poliestireno pueden ser usadas en un rango amplio de pH.

El desarrollo de estos materiales, junto con el avance en los procedimientos químicos para la derivatización de la superficie, han posibilitado la disponibilidad de materiales cromatográficos para todos los mecanismos de separación en HPLC.

Actualmente existe en el mercado otro material basado en partículas híbridas que combinan componentes orgánicos e inorgánicos. Estas partículas están constituidas por un enrejado de polietoxisilano y grupos de metilsiloxano, estos últimos distribuidos tanto en el núcleo de la partícula como en la superficie[2]. Las partículas resultantes son muy resistentes y pueden tener diámetros inferiores a los 2 µm. Además pueden ser usadas a pH básicos.

3. Principios y Parámetros que determinan la separación[3].

3.1. Ensanchamiento de la banda

La separación de una mezcla en una columna ocurre por la diferente interacción de los componentes con las fases estacionaria y móvil. En cada momento, cada componente estará en equilibrio y distribuido entre ambas fases. Aquel que se encuentre en mayor proporción en la fase móvil avanzará más rápidamente. Idealmente, todas las moléculas de un componente dado deberían avanzar juntas en una banda muy estrecha. Sin embargo, esto no ocurre así y algunas moléculas avanzan más rápido que otras, lo que causa un ensanchamiento de la banda.

La capacidad de separación de una columna cromatográfica depende del ensanchamiento de las bandas que corresponden a cada uno de los componentes de la mezcla que se intenta separar, a medida que avanzan a lo largo de la columna, e incluso después de la salida de la columna (figura 1). Mientras más estrechas se mantengan las bandas, mayor será la capacidad de separación de la columna.

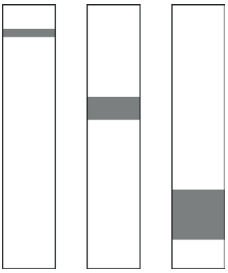


Figura 1. Ensanchamiento de la banda. Debido al efecto de la dispersión y de la resistencia a la transferencia de masas, las bandas correspondientes a los diferentes componentes de la mezcla se van ensanchando a medida que avanzan en la columna. Mientras más estrechas sean estas bandas, más eficiente será una columna y por tanto podrá separar componentes con tiempos de retención más próximos.

A este ensanchamiento contribuyen dos tipos de factores: los intrínsecos de la columna y los efectos externos a la columna. Entre estos últimos están el volumen de inyección, el diámetro y el largo de la tubería desde el inyector hasta la columna y de ahí hasta el detector, el volumen de la celda del detector y el diámetro y el largo de la tubería desde la salida del detector hasta el punto de recolección, en caso de que se deseen colectar los componentes separados. En general, la tubería debe ser tan corta como sea posible y tanto el diámetro de la tubería como el volumen de la celda deben ser muy pequeños, sin que esto llegue a causar un incremento indeseado de la presión. Si se emplean tuberías de diámetros muy grandes o tramos muy largos se mezclarán nuevamente los componentes previamente separados en la columna. Las tuberías más usadas en HPLC analítico tienen un diámetro interno de 0.25 mm o menos y el volumen de la celda oscila entre 1 y 12 µL, en dependencia de la columna analítica que se esté usando.

Entre los factores propios del material cromatográfico y del empaquetamiento de la columna que causan el ensanchamiento de la banda tenemos: la difusión (difusión de Eddy y difusión longitudinal) y la transferencia de masa.

Figura 2. Difusión de Eddy. Cuando una columna está mal empaquetada hay diferentes caminos y la muestra avanza más por las zonas donde el empaquetamiento es menos compacto, lo que provoca el ensanchamiento de la banda.

3.2. Fenómenos de difusión

La difusión de Eddy o difusión radial está relacionada con la diferencia de velocidad de la fase móvil en distintas zonas de la columna, debida fundamentalmente a un empaquetamiento irregular de la columna (figura 2). En regiones más densamente empaquetadas la velocidad del flujo será menor que en regio-

nes de menor densidad, lo que causa un ensanchamiento de la banda. Un empaquetamiento uniforme eliminará la influencia de la difusión de Eddy al ensanchamiento de la banda. Con partículas esféricas y de menor diámetro se logrará un empaquetamiento más compacto y uniforme. En HPLC, la mayoría de los resultados se han obtenido con el uso de partículas esféricas de 3 o 5 µm de diámetro promedio.

La difusión longitudinal se debe a la migración de un componente dado desde zonas de mayor concentración a zonas de menor concentración. La difusión en líquidos no es tan efectiva como en el caso de los gases, por lo que este factor no es nocivo, excepto a flujos muy bajos, donde hay tiempo suficiente para que se manifieste.

3.3. Transferencia de masa

Los geles o materiales cromatográficos más usados están compuestos por partículas porosas (figura 3). La existencia de poros incrementa considerablemente la superficie de interacción de la muestra con la matriz cromatográfica, pero crea problemas debidos a la transferencia de masa.

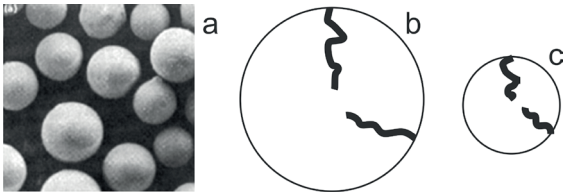


Figura 3. Partículas porosas. En a) se muestra una fotografía de microscopía electrónica de partículas esféricas porosas (cortesía de Merck KGaA, Alemania). En b) y c) se comparan esquemáticamente dos partículas esféricas de diferente diámetro. A mayor diámetro los poros son más profundos y la fase móvil estancada causa mayor resistencia a la transferencia de masas.

La resistencia a la transferencia de masa es uno de los factores que más contribuye al ensanchamiento de las bandas. Por una parte, los componentes de la mezcla requieren de un tiempo mínimo para establecer el equilibrio de distribución entre la fase estacionaria y la fase móvil. Por otra, no todas las moléculas de un componente de la mezcla pueden penetrar de la misma forma en todos los poros de todas las partículas. Estos fenómenos son dependientes de la velocidad de la fase móvil y aunque no pueden eliminarse totalmente sí pueden reducirse.

Un componente importante de la resistencia a la transferencia de masa es la fase móvil estancada en los poros de la partícula. La mayor prueba está en las partículas peliculares que no poseen poros, donde la fase estacionaria se encuentra fijada a la superficie de las partículas como una fina película. La eficiencia de una columna empaquetada con partículas peliculares es mayor que la de una columna empaquetada con partículas porosas de igual diámetro. Sin embargo, las partículas peliculares no han sido muy usadas debido a su baja capacidad, pues los poros incrementan extensamente la superficie de la fase estacionaria, lo que permite la separación de mayor cantidad de muestra.

El uso de partículas porosas de pequeño diámetro también reduce el efecto de la resistencia a la transferencia de masa, por cuanto los poros y el volumen de la fase móvil estancada son mucho menores (figura 3), además de que reducen el volumen de la fase móvil que queda entre las partículas.

Las llamadas columnas monolíticas que se encuentran actualmente en el mercado no están empaquetadas con partículas de sílice. Estas columnas están constituidas por un cilindro continuo de sílice de alta pureza con poros de dos diámetros diferentes: mesoporos (13 nm) y macroporos (2 µm), como se muestra en la figura 4. Al no existir partículas porosas se mejora considerablemente la transferencia de masa y las columnas tienen una gran eficiencia[4,5]. Se pueden fabricar columnas monolíticas de diferentes diámetros, desde capilar hasta preparativas.

También se ha logrado preparar columnas similares con copolímeros de poliestireno y divinilbenceno.

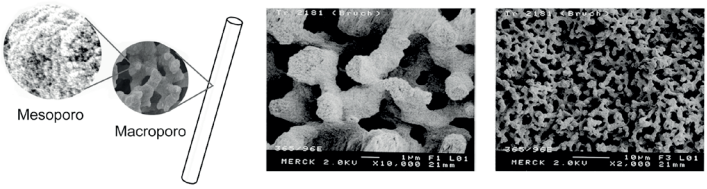


Figura 4. Columnas monolíticas de gel de sílice. El cilindro de la columna monolítica posee macroporos y mesoporos. Las fotografías de microscopía electrónica muestran los macroporos (centro) y mesoporos (derecha) en las columnas monolíticas (cortesía de Merck KGaA, Alemania).

3.4. Medición de la eficiencia de una columna

La eficiencia de una columna se expresa en términos del número de platos teóricos (N) y por la altura de un plato teórico (H).

Para definir N y H deben especificarse primeramente algunos parámetros experimentales como los que se indican en la figura 5.

El tiempo de retención t_R es el tiempo transcurrido desde el inicio de la corrida cromatográfica hasta el centro del pico.

El tiempo t_0 es el tiempo transcurrido para la elución de un componente que no se retiene en la columna (también conocido como frente del solvente).

El ancho del pico, medido al nivel de la línea base, es w_b.

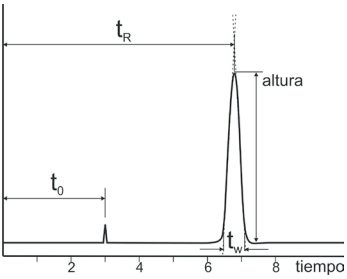


Figura 5. Cromatograma y parámetros experimentales. Los parámetros que pueden medirse directamente del cromatograma son: tiempo de elución de un componente no retenido t_0, tiempo de retención de un componente t_R, ancho del pico t_W, altura y área de un pico. Estos dos últimos son proporcionales a la cantidad de cada componente de la mezcla. En las abscisas se representa el tiempo de la corrida y en las ordenadas la respuesta del detector.

Por último tenemos la altura y el área bajo la curva del pico, que dan una medida de la cantidad en que está presente un componente en la mezcla. Ambos pueden ser usados para fines cuantitativos.

El número de platos teóricos N se define en función del tiempo de retención (t_R) y del ancho del pico en su base (w_b).

$$N = 16\,(t_R/w_b)^2$$

Mientras mayor sea el valor de N, mayor será la eficiencia de la columna. El ancho del pico indica que moléculas de un mismo componente están eluyendo a diferentes tiempos. Mientras más estrecho sea un pico, la columna será más eficiente y capaz de separar componentes con tiempos de retención muy próximos.

Medir el ancho del pico al nivel de la línea base puede ser difícil, por cuanto esta no siempre es una línea horizontal y puede tener una tendencia a subir o bajar. Por esta razón, para calcular N, muchas veces se prefiere otra expresión, donde el ancho del pico se mide a la mitad de su altura (w_h):

$N = 5.54 (t_R/w_h)^2$

En muchos libros y revistas se puede encontrar $w_{1/2}$ en lugar de w_h, pero ese término no es recomendado por la IUPAC[3].

La altura de un plato se define como:

H = L/N, donde L es la longitud de la columna. Por tanto, mientras más pequeño sea el valor de H, más eficiente será la columna.

Ya se ha visto que uno de los parámetros más importantes que influyen en la eficiencia es el diámetro de la partícula.

Los conocidos gráfico y ecuación de Van Deemter muestran la dependencia de H con la velocidad de la fase móvil y cómo esa dependencia está muy relacionada con el diámetro de la partícula.

La ecuación de Van Deemter es: H = A + B/u + C/u, donde u es la velocidad lineal promedio de la fase móvil. A, B y C son constantes para una columna y unas condiciones experimentales dadas, aunque dependen de la muestra, de la fase móvil y de la temperatura.

La constante A está relacionada con la difusión de Eddy, por lo que en una columna bien empaquetada no es necesario considerarla.

La constante B se relaciona con la difusión longitudinal y solo es de importancia a flujos muy bajos.

C está relacionada con la resistencia a la transferencia de masa y es el factor más importante que influye en el ensanchamiento de la banda dentro de la columna.

Como puede observarse en la figura 6, H se incrementa con la velocidad de la fase móvil, lo cual reduce la eficiencia de la columna. Al usar un flujo muy alto, no es posible que se establezca el equilibrio de la muestra entre la fase estacionaria y la fase móvil; la muestra no tiene tiempo suficiente para interactuar con la fase estacionaria, por lo que se reduce la eficiencia de la columna.

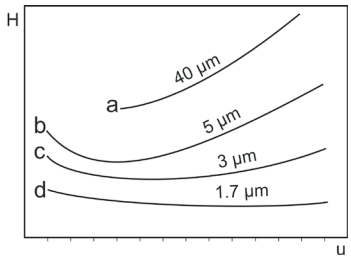


Figura 6. Gráfico de van Deemter. Dependencia de H con la velocidad lineal de la fase móvil. H se incrementa (y por tanto disminuye la eficiencia de la columna) con la velocidad lineal y para cada columna existe una velocidad óptima. H también se incrementa a flujos muy bajos debido a la influencia de la difusión longitudinal. La dependencia de H con la velocidad lineal de la fase móvil se reduce a medida que disminuye el diámetro de las partículas. Las columnas monolíticas mejoran la transferencia de masas y se comportan de forma similar a las columnas con partículas de pequeño diámetro (3-5 µm).

También H se incrementa a flujos muy pequeños, debido a que la difusión longitudinal afecta en esas condiciones. De este gráfico se deduce que para una columna determinada existe una velocidad óptima, que es la que se debe utilizar.

Sin embargo, puede observarse que esa dependencia de H con la velocidad de la fase móvil se reduce al disminuir el tamaño de las partículas, y para un diámetro de 1.7 µm la curva es casi una línea horizontal (curva e en la figura 6). Columnas empaquetadas con partículas de 1.7 µm pueden utilizarse con un alto flujo sin que se afecte la eficiencia, mientras el tiempo de análisis se reduce al orden de los segundos o muy pocos minutos. Las columnas monolíticas se com-

portan de una manera muy similar a las de partículas de 3 µm y 5 µm de diámetro y tienen la ventaja de no necesitar una presión elevada aun a flujos relativamente altos.

Para reducir el efecto negativo de la resistencia a la transferencia de masa es aconsejable usar partículas de pequeño diámetro (o columnas monolíticas), evitar el uso de una fase móvil muy viscosa y flujos muy altos (excepto cuando se usan columnas con partículas de muy pequeño diámetro). El incremento de la temperatura mejora la eficiencia de la columna. En una misma columna se obtendrá mayor eficiencia al separar moléculas pequeñas que moléculas de gran talla, como péptidos y proteínas. En este último caso es necesario usar materiales cromatográficos con diámetro de poro de 250 Å o 300 Å, para así facilitar la transferencia de masa de estas macromoléculas.

H será menor, y por tanto se tendrá mayor eficiencia, cuando sea menor:

- El diámetro de las partículas del material cromatográfico.
- El flujo y la viscosidad de la fase móvil.
- El tamaño de las moléculas del analito.

3.5. Medición de la resolución

Factor de retención (factor de capacidad) k

Este parámetro es una medida del tiempo que un componente está retenido en la columna.

El factor de retención se puede expresar como:

$$k = t_R - t_0 / t_0$$

Factor de separación (selectividad) α

El factor de separación α viene dado por la siguiente expresión:
$\alpha = k_2/k_1$, donde k_1 y k_2 son los factores de retención de dos picos adyacentes en el cromatograma y k_2 corresponde al componente de mayor tiempo de retención. Por tanto, α será siempre mayor o igual que 1 y es una medida de la selectividad de un sistema cromatográfico (fase estacionaria y fase móvil) que se emplea para separar dos componentes. Cuando sea igual a 1 significa que no hay separación y que ambos componentes coeluyen.

3.6. Resolución R

Este parámetro indica si una columna es capaz o no de separar dos componentes con t_R muy próximos y se define como la diferencia de los tiempos de retención de los dos picos adyacentes entre el promedio del ancho de los dos picos.

$$R = t_{R2} - t_{R1} / (½) (w_{b1} + w_{b2}) = 2 \Delta t / (w_{b1} + w_{b2})$$

La resolución puede expresarse en función de las variables cromatográficas ya definidas:

$R = ¼ [(\alpha - 1)/\alpha] N^{1/2} [k/(k+ 1)]$, donde k es el promedio de las k_1 y k_2 de dos picos adyacentes.

Por tanto, la resolución depende de estos tres parámetros: N, k y α. Para mejorar R se pueden optimizar uno o varios de estos parámetros.

Es posible incrementar N aumentando el largo de la columna, reduciendo la velocidad de la fase móvil o cambiando la columna por otra que tenga un empaquetamiento más eficiente, como por ejemplo, con partículas de menor diámetro. Sin embargo, las dos primeras opciones implican un incremento del tiempo de análisis y un ensanchamiento de los picos. Debe notarse además que R es proporcional a la raíz cuadrada de N, por lo que para lograr un incremento sensible de R, N deberá tener un incremento considerable.

Para incrementar α es necesario cambiar la naturaleza de la fase estacionaria o de la fase móvil. En intercambio iónico debe sustituirse un intercambiador catiónico por uno aniónico, o variar el pH de la fase móvil, por ejemplo.

Por último, un incremento de k debe mejorar la resolución. Sin embargo, R no depende directamente de k, sino del cociente k/(k + 1). Este cociente toma los siguientes valores al variar k (Tabla 1).

Tabla1. Valores de k/(k + 1) en función de k.

k	0	1	2	3	5	10
k/(k + 1)	0	0.50	0.67	0.75	0.83	0.91

Para valores de k por encima de 5 no se logrará un incremento sustancial de R, pero el tiempo de retención se incrementa demasiado, al igual que el ancho de los picos, por lo que k no debe ser mayor de 5, en general. Por otra parte, este parámetro es el que más fácilmente puede modificarse al variar la fortaleza de la fase móvil.

Para incrementar o reducir k bastaría con reducir o incrementar la fortaleza de la fase móvil, respectivamente. En una columna de intercambio iónico esto se logra con la reducción o el incremento de la concentración iónica de la fase móvil (figura 7).

Si dos componentes quedan mal separados, pero con tiempos de retención muy cercanos a t_0 (es decir, poco retenidos en la columna), la mejor opción para optimizar la resolución es reducir la fortaleza de la fase móvil para incrementar k (figuras 7 a y b). Por el contrario, si dos componentes quedan mal separados, incluso con tiempos de retención relativamente altos ($k \geq 5$, figura 7c), la mejor opción será modificar α, cambiando la naturaleza de la fase estacionaria o de la fase móvil (figura 7e).

Cuando se tengan dos componentes con una separación insuficiente, y que no sea posible mejorar al modificar α o k, será imprescindible optimizar N y cambiar a una columna con partículas de menor diámetro (de 5 µm a 1.7 µm por ejemplo, figura 7d).

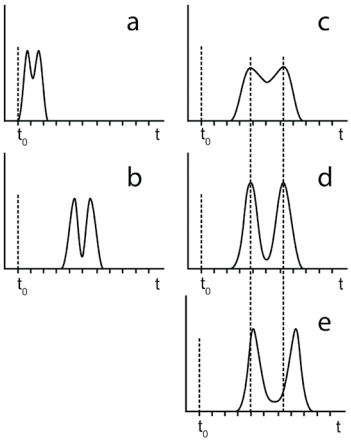


Figura 7. Efecto de modificaciones en N, k y α sobre la separación. En a), los dos componentes no se separan por estar muy poco retenidos. Al incrementar k por disminución de la fortaleza de la fase móvil se logra la separación b). Si el valor de k es alrededor de 5 y no ocurre una buena separación c), es necesario incrementar N para hacer los picos más estrechos y permitir la separación d) o modificar α cambiando la naturaleza de la fase estacionario o de la fase móvil e).

3.7. Selectividad y eficiencia

Es importante resaltar la diferencia entre selectividad y eficiencia. En la figura 8 se muestra la diferencia entre estos conceptos y su relación con la resolución.

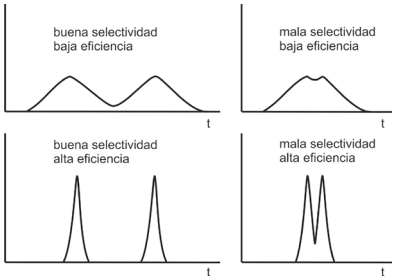


Figura 8. Selectividad y eficiencia. La selectividad se relaciona con la separación entre los picos y la eficiencia con el ancho de los picos.

La selectividad está relacionada con la separación de dos picos adyacentes. Cuando los picos están muy separados significa que la selectividad es buena, mientras que cuando están muy próximos es mala, independientemente del ancho de los picos.

Por otra parte, la eficiencia se relaciona con el ancho de los picos. La eficiencia es alta cuando los picos son muy estrechos, es decir, cuando todas las moléculas de un componente eluyen en un rango de tiempo muy pequeño. La eficiencia es baja cuando los picos son muy anchos, independientemente de la separación entre los picos.

De esto se puede inferir que cuando se tiene muy buena selectividad no es necesario optimizar los parámetros relacionados con la eficiencia. Esto explica porqué en general no se usan columnas de HPLC en la cromatografía de afinidad. Por el contrario, cuando no se puede lograr una buena selectividad para separar dos componentes es imprescindible tener la mayor eficiencia posible. Esto último también se aplica si intentamos separar mezclas muy complejas, como las de los péptidos que deben ser separadas en estudios de proteómica.

3.8. El problema general de elución. Uso de gradientes

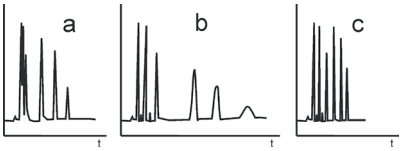


Figura 9. Necesidad del uso de gradientes. En mezclas complejas no es posible encontrar condiciones de separación isocrática de todos los componentes. En a) se logran separar solo los componentes de mayor tiempo de retención, mientras que en b) solo se pueden separar los de menor tiempo de retención pues el resto de los picos se ensanchan considerablemente. Solo se puede lograr la separación de todos los componentes con un gradiente (c).

Una separación cromatográfica en condiciones isocráticas, es decir, sin que se modifique la composición de la fase móvil, es muy ventajosa, porque requiere un sistema cromatográfico mucho más sencillo. Sin embargo, cuando se tienen mezclas con múltiples componentes, estos eluirán en un rango muy amplio de factores de retención. Por tanto, si se trata de optimizar la separación de los componentes de mayor retención aumentando la fortaleza de la fase móvil, los componentes de menor k no serán separados o ni siquiera retenidos en la columna (figura 9a). Por el contrario si optimizamos la separación de los componentes con poca retención disminuyendo la fortaleza de la fase móvil, se incrementará considerablemente la retención de los componentes de mayor k, con el consiguiente ensanchamiento de los picos y pérdida de sensibilidad (figura 9b).

La solución es usar un gradiente para la elución de la mezcla, y cambiar la composición de la fase móvil en el tiempo (figura 9c): se iniciará con una composición con baja fortaleza que se irá incrementando en el tiempo. Para este fin se usan dos soluciones, A y B, la última con mayor fortaleza. Se puede comenzar el gradiente con 100 % de la solución A y la proporción de B se irá incrementando gradualmente hasta alcanzar el 100 %.

Es muy frecuente el uso de gradientes lineales, donde el incremento de la proporción de B es constante en el tiempo (figura 10a, curvas 1 y 2). En estos casos es muy importante la pendiente del gradiente. En general se tendrán mejores separaciones con el uso de gradientes con pendientes suaves, aunque esto implica un incremento del tiempo de análisis. En el caso de

mezclas muy complejas también pueden usarse gradientes lineales donde la pendiente cambia en el tiempo (figura 10a, curva 3). Por ejemplo, se puede iniciar el gradiente con una pendiente muy suave, y cuando eluyan los picos de menor retención, incrementarla para acelerar la elución de los componentes de mayor retención. También pueden usarse gradientes de otros tipos, como convexos o cóncavos.

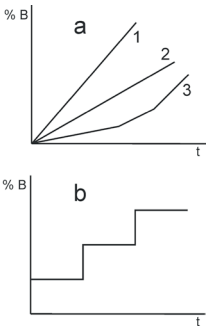


Figura 10. Diferentes tipos de gradientes lineales. La pendiente del gradiente tiene gran influencia en la separación. En general gradientes con menor pendiente, a), dan mejores resultados, aunque incrementan el tiempo de análisis. En algunas mezclas puede utilizarse un gradiente lineal que incremente la pendiente durante la corrida, como se muestra en la curva 3. Una alternativa al gradiente es la elución por pasos, donde se incrementa la fortaleza de la fase móvil una o dos veces, pero cada etapa es una elución isocrática (b).

Otra alternativa es la llamada elución por pasos (figura 10b). En este caso se inicia la corrida con una composición fija de A y B, y a un tiempo dado se incrementa la proporción de B y se continúa sin variar la nueva composición. Esto puede repetirse una o dos veces más. Cada paso es en definitiva una corrida isocrática. Este procedimiento se usa con frecuencia en cromatografía preparativa, donde el objetivo es aislar uno o dos componentes de una mezcla, pero no es eficaz cuando se requiere la separación de todos los componentes, como en proteómica.

4. Mecanismos de separación.

Ya hemos visto que en HPLC se tienen los mismos mecanismos de separación que en la cromatografía "tradicional". En lo adelante se analizarán cada uno de los mecanismos usados con más frecuencia en el análisis y separación de péptidos y proteínas: cromatografía de intercambio iónico, de exclusión por talla (filtración en gel), de fase reversa, de interacción hidrofóbica, de interacción hidrofílica y de afinidad.

5. Cromatografía de exclusión por talla molecular (filtración en gel).

La separación ocurre de acuerdo al tamaño de las moléculas en la mezcla. Las partículas del gel son porosas. Las moléculas más pequeñas podrán penetrar en los poros, no así las moléculas de mayor talla, que tendrán un acceso restringido. Por tanto, estas últimas eluyen más rápidamente que las moléculas más pequeñas (figura 11).

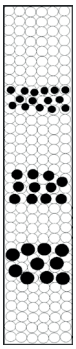


Figura 11. Cromatografía de filtración en gel. Las moléculas más grandes tienen restricción para penetrar en los poros de las partículas y por tanto avanzan más rápidamente.

Este es el único mecanismo de separación cromatográfico en el que no ocurre interacción de la muestra con la fase estacionaria. Incluso, la fase móvil debe contener algunos componentes para evitar posibles interacciones. Por ejemplo, a la fase móvil se le adicionan usualmente entre 0.2 y 0.3 mol/L de cloruro de sodio para evitar posibles interacciones electrostáticas entre las proteínas y la fase estacionaria. Por la misma razón, en este mecanismo de separación no tiene sentido el uso de gradientes y las corridas son siempre isocráticas.

La resolución solo puede ser controlada por N, por lo que en general las columnas son más largas que en otros tipos de cromatografía y usan flujos más pequeños.

Desde el momento en que se aplica la muestra en la columna comienza la separación, por lo que tanto el volumen como la masa de la muestra tienen una gran influencia y deben mantenerse dentro de valores limitados, de acuerdo a las dimensiones de la columna. Otra característica de esta técnica es que causa la dilución de la muestra.

5.1. Definiciones y terminología específica de la cromatografía de filtración en gel

V_e: volumen de elución de un compuesto

V_o: volumen que ocupa la fase móvil entre las partículas del gel. Es igual al volumen de elución de un compuesto que no puede penetrar en los poros de las partículas y por tanto no es retenido en la columna.

V_i: volumen de la fase móvil que ocupa el interior de los poros de las partículas.

V_t: volumen total de la columna. $V_t = V_o + V_i + V_m$, donde V_m es el volumen del material sólido de la matriz cromatográfica.

Aunque el volumen de elución V_e pudiera ser usado para caracterizar el comportamiento de un compuesto en una columna de filtración en gel, es más apropiado usar el coeficiente de distribución K_d, que es la fracción del volumen en el interior de los poros (volumen de los poros) que está disponible para un componente dado para la difusión.

$K_d = V_e - V_o / V_i$

Sin embargo V_i es muy difícil de medir o estimar, por lo que se prefiere sustituir V_i por $V_t - V_o$ y usar la siguiente expresión $K_{av} = V_e - V_o / V_t - V_o$

V_o puede ser fácilmente determinado experimentalmente, midiendo el volumen de elución de un compuesto que no pueda penetrar en los poros de la matriz cromatográfica, como por ejemplo dextrana azul 2000.

K_{av} no es un verdadero coeficiente de distribución por cuanto V_t incluye el volumen del gel V_m, sin embargo K_d y K_{av} son proporcionales para una matriz cromatográfica dada.

Este tipo de cromatografía tiene una resolución relativamente baja, por lo que una separación completa de dos componentes solo se puede lograr si hay una diferencia de talla suficientemente grande. De ahí que una de sus principales aplicaciones sea la identificación y separación de posibles dímeros, trímeros y agregados de una proteína (figura 12).

Otra aplicación muy generalizada es el desalado o cambio de tampón de una preparación de proteínas. Las sales tienen tallas muy pequeñas comparadas con las proteínas, por lo que aquellas tendrán una retención mucho mayor (figura 13).

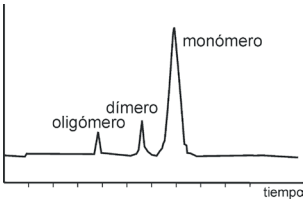


Figura 12. Separación de dímeros, trímeros y oligómeros de proteínas en columnas de filtración en gel.

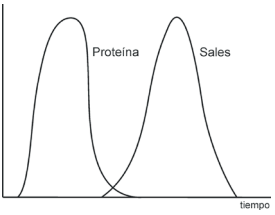


Figura 13. Desalado de proteínas en columnas de filtración en gel. Las sales son mucho más pequeñas que las proteínas y pueden ser fácilmente separadas con este método cromatográfico.

Por último, este tipo de cromatografía se puede utilizar para determinar el peso molecular aproximado de una proteína a partir de su volumen de elución o mas específicamente de K_{av}. Existe una relación lineal entre K_{av} y el logaritmo del peso molecular en un rango amplio de pesos moleculares y que es específico para cada matriz.

Para determinar el peso molecular aproximado de una proteína, se construye el gráfico de calibración con proteínas de pesos moleculares conocidos y el peso molecular de la proteína en cuestión se estima por interpolación (figura 14).

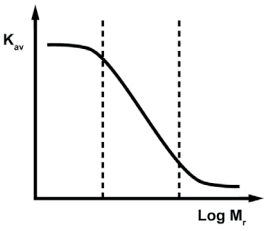


Figura 14. Determinación del peso molecular aproximado de una proteína en una columna de filtración en gel. Se construye la curva de calibración con proteínas de pesos moleculares conocidos en un gráfico de K_{av} contra el logaritmo del peso molecular Mr. La estimación del peso molecular de una proteína se logra por interpolación.

El rango óptimo de separación de una columna de exclusión por talla está relacionado con el diámetro de los poros del material. Esto implica que se necesita tener disponible un grupo de columnas, cada una con un rango óptimo de separación, y se usará aquella que nos pueda dar el mejor resultado de acuerdo a la mezcla de proteínas que se desee separar.

6. Cromatografía de intercambio iónico.

En este tipo de cromatografía los grupos funcionales de la fase estacionaria están cargados positiva o negativamente y la separación ocurre por la interacción de los componentes de la muestra con carga opuesta. La elución ocurre por la adición de sales a la fase móvil, donde el contraión de la sal desplaza al soluto. El incremento de la concentración iónica va desplazando progresivamente a los componentes de la mezcla en dependencia de la carga de cada uno de ellos (figura 16).

Existen dos tipos de cromatografía de intercambio iónico, de acuerdo a la carga de la fase estacionaria: intercambio catiónico e intercambio aniónico. En el primer caso la matriz cromatográfica se encuentra cargada negativamente y, por tanto, intercambia cationes. En el segundo caso, la matriz está cargada positivamente e intercambia aniones. En la figura 15 se muestra esquemáticamente la separación de una mezcla en una matriz de intercambio aniónico.

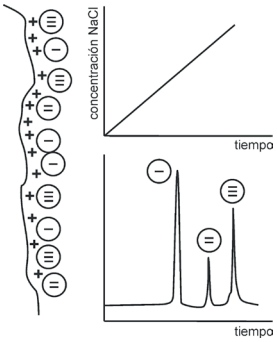


Figura 15. Cromatografía de intercambio iónico. En la figura se muestra esquemáticamente una matriz de intercambio aniónico. La elución tiene lugar al incrementar la fortaleza de la fase móvil (incremento de la concentración de NaCl) y eluyen primeramente los compuestos con menor carga.

Una mezcla de péptidos o proteínas puede ser separada en cualquiera de estas dos matrices, en dependencia del pH de trabajo y de los puntos isoeléctricos de los componentes de la mezcla. Esto a su vez dependerá de la proporción relativa de aminoácidos básicos y ácidos en la composición de cada péptido o proteína. A pH ácido predominarán los cationes, mientras que a pH alcalino predominarán los aniones (figura 16).

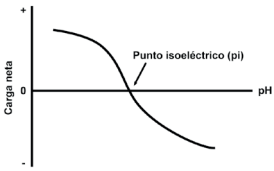


Figura 16. Dependencia de la carga de las proteínas con el pH. A valores de pH por encima del punto isoeléctrico, las proteínas estarán cargadas negativamente, mientras que a valores inferiores de pH, las proteínas estarán cargadas positivamente.

Las matrices intercambiadoras de cationes o aniones pueden ser fuertes o débiles en dependencia del rango de pH en que se mantengan cargadas.

Las matrices intercambiadoras fuertes se mantienen ionizadas en un amplio rango de pH, mientras que las débiles solo estarán cargadas, y activas como intercambiadoras iónicas, a pH ácido (aniónicas) o a pH básico (catiónicas).

Los grupos químicos funcionales en las matrices intercambiadoras catiónicas son ácidos sulfónicos o carboxílicos. Los primeros son ácidos fuertes y estarán ionizados en un rango de pH amplio, en tanto los segundos solo estarán ionizados a pH básico.

En las matrices intercambiadoras aniónicas los grupos químicos funcionales son sales de amonio cuaternarias o aminas. Estas últimas solo estarán cargadas a pH ácidos.

En la Tabla 2 se muestran algunos de los grupos funcionales usados en las matrices de intercambio iónico para la separación de péptidos y proteínas.

Tabla 2: Algunos grupos funcionales usados en las matrices de intercambio iónico

Grupo funcional		Tipo de intercambiador
$-OCH_2CH_2N(CH_2CH_3)_2$	DEAE	Aniónico débil
$-OCH_2CH_2N^+(C_2H_5)_3$	Q amina cuaternaria	Aniónico fuerte
$-OCH_2CHOHCH_2OCH_2CHOHCH_2N^+(CH3)_3$	Q amina cuaternaria	Aniónico fuerte
$-OCH_2COOH$	CM carboximetilo	Catiónico débil
$-OCH_2CHOHCH_2OCH_2CH_2CH_2SO_3^-$	SP sulfopropilo	Catiónico fuerte

La cromatografía de intercambio iónico es muy versátil y poderosa para la separación de péptidos y proteínas. En principio una mezcla de proteínas siempre pudiera separarse por intercambio iónico. Sin embargo, la optimización de esa separación puede ser tan extremadamente compleja que en la práctica no se pueda lograr. Esto es debido a los numerosos parámetros que intervienen en la separación de proteínas por intercambio iónico:

- Matriz: Puede ser catiónica o aniónica, y en ambos casos fuerte o débil.

- pH: El cambio de pH modifica de forma diferente la ionización de cada proteína en la mezcla. Incluso un pequeño cambio en el pH puede provocar un comportamiento muy diferente de la mezcla de proteínas. Además, el pH puede modificar el nivel de ionización de las matrices débiles.

- Naturaleza del tampón y del contraión utilizados: La separación puede variar según la naturaleza del tampón empleado. En cuanto al contraión se tiene el siguiente orden en relación a la capacidad de desplazamiento:

- Aniones: citrato^{3-} (fuerte) >SO_4^{2-} >oxalato^{2-} >I^- >NO_3^- >SCN^- >Cl^- >formiato$^-$ >acetato$^-$ >OH^- >F^- (débil).

- Cationes: Ba^{2+} (fuerte) >Ca^{2+} >Cu^{2+} >Mg^{2+} >Cs^+ >K^+ >NH_4^+ >Na^+ >H^+ >Li^+ (débil).

- El cloruro de sodio es la sal más frecuentemente usada para la elución en la purificación de proteínas, tanto en intercambio catiónico como aniónico.

- Adición de solvente orgánico: La adición de una cierta cantidad de solvente orgánico puede tener un efecto muy positivo en la separación de una mezcla de proteínas. Sin embargo, la adición debe controlarse para no provocar la precipitación de las sales y de la muestra.

- Temperatura: En general, un incremento de la temperatura puede mejorar la separación.

En proteómica, el intercambio catiónico ha sido frecuentemente empleado para la separación de mezclas complejas de péptidos. A pH entre 2 y 3, prácticamente todos los péptidos están cargados positivamente, por lo que el intercambio catiónico puede ser muy efectivo.

7. Cromatografía de Fase Reversa.

En cromatografía, varios conceptos han sido nombrados con vocablos que por su etimología no es posible asociarlos con el fenómeno que describen. El propio nombre de "cromatografía" está dado por la primera aplicación de esta técnica, usada en la separación de colorantes. El término "platos teóricos" (N) está tomado de la teoría de la destilación. Así ocurre también con el término de "fase reversa".

La aparición de la HPLC estuvo en gran medida determinada por el desarrollo tecnológico en la preparación de las matrices cromatográficas con las propiedades y reproducibilidad requeridas, fundamentalmente de partículas de pequeño diámetro (3-10 µm). El uso de estas partículas en las columnas de HPLC implicaba el empleo de presiones elevadas, de más de 100 bar. El gel de sílice era uno de los pocos materiales disponibles capaces de soportar esas presiones, pero su empleo dependía de la posibilidad de modificar químicamente su superficie, de manera que se tuvieran disponibles matrices cromatográficas para todos los tipos de cromatografía de acuerdo a los diferentes mecanismos de separación. Así, el segundo desarrollo tecnológico que permitió la consolidación de la HPLC fue el perfeccionamiento de los métodos de síntesis para modificar la superficie del gel de sílice con diferentes fases estacionarias químicamente enlazadas a la superficie de las partículas de sílice. Una de estas fases está constituida por cadenas de hidrocarburos, o sea, una superficie muy hidrofóbica.

Por otra parte, las primeras matrices usadas en cromatografía fueron el carbonato de calcio (en el trabajo original de Tsvet) y el gel de sílice sin modificación. Ambas son matrices muy polares. Consecuentemente, los compuestos polares eran más retenidos y la elución se lograba partiendo de solventes muy poco polares, tales como el hexano. La fortaleza de la fase móvil se incrementaba por adición de solventes más polares como cloroformo, tetrahidrofurano o metanol.

En las matrices con superficies hidrofóbicas ocurre todo lo contrario, pues los compuestos más retenidos en la columna son los muy hidrofóbicos. La elución se inicia con el uso de solventes muy polares (usualmente agua) y el incremento de la fortaleza de la fase móvil se logra con la adición de solventes orgánicos menos polares como propanol, acetonitrilo o tetrahidrofurano. Este proceder "reverso" con respecto al conocido hasta ese momento originó el nombre de "cromatografía de fase reversa" y la cromatografía existente hasta entonces, con fases estacionarias polares, comenzó a llamarse "cromatografía de fase normal".

Los principales grupos funcionales que aparecen en la superficie de matrices de fase reversa (RP, del inglés *reversed phase*) son hidrocarburos de 4, 8 y 18 átomos de carbono, conocidos como C4, C8 y C18, este último también llamado ODS, por octadecilsilano. También se usa con frecuencia el grupo fenilo.

La naturaleza de la retención en RP no está totalmente conocida[6].

En los estudios de proteómica sería muy importante predecir el tiempo de retención en RP de los péptidos, ya que sería un parámetro adicional que ayudaría considerablemente a la identificación confiable de las proteínas. Sin embargo, se ha podido comprobar que es muy difícil hacer esa predicción, aun cuando se han publicado algunas alternativas que tienen en cuenta varios

factores, incluyendo los coeficientes de partición octanol/agua[7,8]. No obstante, la interacción hidrofóbica que ocurre en RP permite racionalizar la separación en este tipo de cromatografía. Esta interacción no es en absoluto semejante a lo que ocurre en una interacción electrostática y se puede considerar como un proceso dirigido por la entropía[9]. Las moléculas de agua están usualmente formando agrupaciones unidas entre sí por los enlaces de hidrógeno. Sin embargo, en los alrededores de la superficie hidrofóbica de la matriz las moléculas de agua están ordenadas formando una lámina o capa, pues no pueden formar puentes de hidrógeno con el ligando de la matriz, de forma similar a lo que sucede en la superficie de un líquido y que da lugar a la tensión superficial. Al unirse una molécula hidrofóbica del analito (o una porción hidrofóbica de una proteína) a la fase estacionaria, desplaza moléculas de agua, se reduce el número de moléculas de agua ordenadas y se logra un incremento de la entropía del sistema. Es un fenómeno similar al que se observa cuando se colocan dos gotas de aceite en agua. Al poco tiempo, se podrá observar que ambas gotas se unen para reducir el área total de la superficie hidrofóbica y por tanto disminuye el número de moléculas ordenadas alrededor de esa superficie. De esta forma, la interacción del analito con la matriz hidrofóbica se ve favorecida, y mientras más hidrofóbico sea el analito, mayor será la retención.

Al preparar las matrices de RP es importante lograr que no queden grupos silanoles libres, pues las separaciones ocurrirían por mecanismos mixtos, lo cual complica la optimización de los parámetros de la separación y reduce la eficiencia de la columna. Los silanoles libres residuales se reducen haciéndolos reaccionar con hidrocarburos ramificados de menor tamaño de cadena. Este procedimiento se conoce como bloqueo o cobertura final (*endcapping*).

La cromatografía de RP es una de las más usadas en la actualidad para el análisis y purificación de compuestos orgánicos. En la industria farmacéutica es ampliamente utilizada para el control de calidad de los ingredientes activos y de las formas terminadas de los medicamentos.

Su auge se debe a su elevada capacidad de separación (factor de selectividad), que se manifiesta incluso en moléculas estructuralmente muy similares. Probablemente no exista ningún otro método con semejante capacidad de separación y que no implique la necesidad de modificar químicamente los componentes de una mezcla compleja. La cromatografía de gases también tiene una elevada capacidad de separación, pero es necesario derivatizar químicamente los compuestos polares para incrementar su volatilidad.

La temperatura puede tener un efecto marcado en la separación en columnas de RP. Generalmente se trabaja alrededor de 40 °C y es muy recomendable que siempre se mantenga una temperatura uniforme.

7.1. RP de péptidos y proteínas

El RP también ha tenido una amplia aplicación en el campo de las proteínas y es prácticamente el método obligado para el análisis y purificación de péptidos.

A diferencia de las moléculas pequeñas, los péptidos y proteínas eluyen de una columna de RP en un rango estrecho de la proporción del solvente orgánico. Un pequeño incremento del solvente orgánico puede causar la elución de una proteína. Esto último explica porqué una columna de RP tiene una alta eficacia para separar una mezcla de péptidos y proteínas a pesar de que estos eluyan en un rango estrecho de la proporción del solvente orgánico. Por esta misma razón la elución isocrática no es práctica y se prefiere usar la elución con gradiente.

Por otra parte, la optimización de las condiciones de elución es relativamente sencilla. El solvente A es usualmente agua con alrededor de 0.1 % de ácido trifluoroacético (TFA) y el solvente B acetonitrilo con una cantidad similar de TFA. La optimización de la separación, muchas veces consiste solamente en ajustar la pendiente del gradiente.

El propanol (o el isopropanol) es un mejor solvente que el acetonitrilo para proteínas, sin embargo, tiene dos grandes desventajas. La primera es su elevada viscosidad, que reduce la eficiencia de la columna al incrementar la resistencia a la transferencia de masa y eleva la presión del sistema. La segunda desventaja se relaciona con la detección en UV aproximadamente a 220 nm, que es la más usada en la cromatografía de péptidos y proteínas, pues el propanol no es apropiado para ser utilizado en ese rango, debido a la llamada "absorción final de los alcoholes" en esta zona del espectro UV.

El TFA funciona como un par iónico que se une a los grupos básicos ionizados de los péptidos y proteínas y permite modular la hidrofobicidad y mejorar la eficiencia de la separación, pues tiene un efecto considerable en la selectividad. Además, el pH ácido contribuye a suprimir la ionización de los posibles grupos silanoles libres que afectan la separación.

El TFA es un par iónico relativamente débil. Otros ácidos tienen mejores capacidades, como el ácido heptafluorobutírico, pero el TFA es preferido por tener menor absorción en UV y ser más volátil.

En la literatura científica se pueden encontrar numerosas aplicaciones de RP para el análisis y purificación de proteínas. La elevada capacidad de separación de RP puede ejemplificarse en estos casos que se muestran a continuación.

El interferón alfa 2b (IFN) recombinante es una proteína de 19 KDa que se obtiene en *E.Coli* en el Centro de Ingeniería Genética y Biotecnología (CIGB) de La Habana. Una pequeña proporción de la proteína (entre 3 y 7 %) está acetilada en el N-terminal. En la figura 17a se muestra el cromatograma en una de las fases finales del proceso de purificación. El pico principal corresponde al IFN y el pico a su derecha, con mayor tiempo de retención, es el interferón acetilado. La columna de C8 es capaz de separar estas dos especies de una proteína de 19 KDa, que solo difieren en un grupo acetilo en el N-terminal. La etapa final de la purificación del interferón es precisamente RP. En la figura 17b se muestra el interferón después de este paso de purificación en RP.

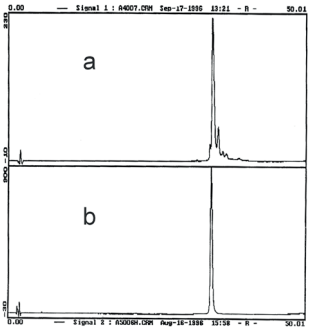


Figura 17. Cromatograma de RP del interferón alfa 2b (IFN). En a) se muestra el cromatograma de una de las etapas finales de la purificación del IFN. El pico principal corresponde al IFN y el pico minoritario al IFN acetilado en el N-terminal. Este componente se elimina por una cromatografía de RP preparativa y el resultado se muestra en b).

Las condiciones que se usan en la cromatografía de fase reversa son propicias para la desnaturalización de las proteínas: pH entre 2 y 3 y altas proporciones de solvente orgánico. Sin embargo, muchas proteínas pueden ser purificadas por RP y recobrar su estructura tridimensional y función biológica. En nuestro laboratorio del CIGB tenemos varios ejemplos positivos como el interferón, la interleucina 2 y la tripsina. Las proteínas muy hidrofóbicas tienen muy bajo recobrado en RP.

En el caso de los péptidos, el factor de selectividad es también muy alto. Una pequeña modificación en la secuencia de un péptido se traduce en un claro cambio del tiempo de retención. Esta característica se aprovecha en el llamado mapa peptídico de una proteína. El análisis consiste en la digestión de la proteína con una enzima específica, casi siempre tripsina, y la mezcla de péptidos obtenida se separa en una columna de RP de C18.

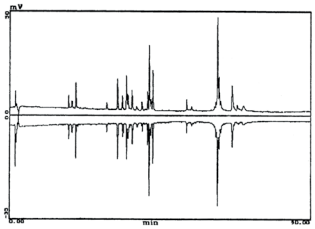


Figura 18. Mapa peptídico del interferón alfa 2b. Para el control de la producción del IFN, cada nuevo lote (arriba) se compara con el control. Tiene que existir una correspondencia exacta en el número de picos y en los tiempos de retención de cada pico.

El mapa peptídico está incluido en la Farmacopea como un análisis obligatorio para todos los lotes de producción de una proteína recombinante. El mapa de cada nuevo lote se compara con el de un patrón (figura 18). Tiene que existir total coincidencia en el número de picos y en el tiempo de retención de cada pico.

Un pequeño cambio en el tiempo de retención de algún pico o la aparición o desaparición de algún pico demuestra una alteración en la secuencia de la proteína producida, como puede apreciarse en la comparación del interferón y la especie acetilada.

En los dos cromatogramas se observa una gran similitud entre el interferón (figura 19a) y el interferón acetilado (figura 19b), excepto en la región enmarcada, que es donde se encuentran los picos correspondientes al N-terminal de la proteína[10].

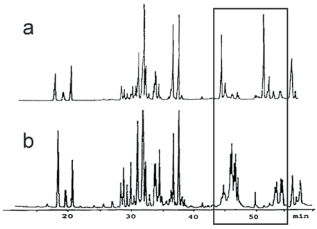


Figura 19. Comparación del mapa peptídico del IFN (a) y del IFN acetilado (b). Se observa una gran similitud entre los dos mapas, excepto en la zona enmarcada, donde se encuentran los péptidos que contienen el N-terminal de la molécula.

La cromatografía en fase reversa es un método excelente para el desalado de péptidos. Las sales son muy polares, por lo que no serán retenidas en una pequeña columna de RP, mientras que los péptidos serán retenidos. En este caso, el desalado con columnas de filtración en gel no es eficaz, pues los péptidos son moléculas mucho más pequeñas que las proteínas y no se logra una buena separación de las sales.

7.2. Efecto del pH

Aunque la gran mayoría de las separaciones en RP se realizan en medio ácido, es posible tener muy buenas separaciones de mezclas de péptidos a pH básico[11]. Por supuesto, es necesario usar columnas resistentes a ese pH. La hidrofobicidad de cada péptido puede cambiar radicalmente al pasar de pH 2-3 a pH 10-12, por lo que los tiempos de retención relativos de una mezcla de péptidos pueden variar totalmente y se comportan como si fueran sistemas ortogonales. En la figura 20 se muestran los cromatogramas del mapa peptídico de la estreptoquinasa, proteína de 47 KDa, a pH ácido y básico. La posición de algunos de los péptidos se señala en cada cromatograma, que cambian considerablemente al cambiar el pH. En particular es interesante el caso de los dos péptidos que coeluyen a pH ácido (aproximadamente a los 38 minutos) y que tienen tiempos de retención muy diferentes a pH básico. Las secuencias de ambos péptidos se muestran en el cromatograma.

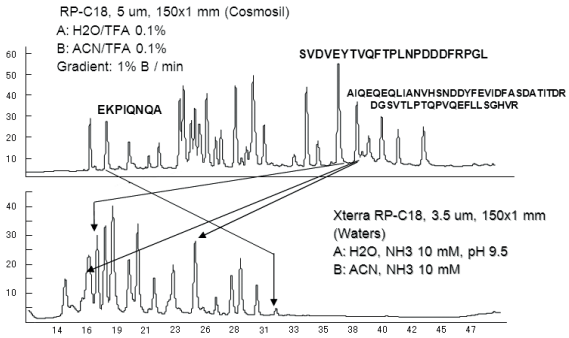


Figura 20. Comparación de la separación de una mezcla de péptidos a pH ácido y básico por RP. En la figura superior se muestra el cromatograma de los péptidos tripticos de la estreptocinasa a pH 2.3 y en la inferior a pH 9.5. Se señalan los cambios en los tiempos de retención de algunos de los péptidos. En particular los péptidos AIQEQEQLIANVHSNDDYFEVIDFASDATITDR y DGSVTLPTQPVQEFLLSGHVR coeluyen a pH ácido (aproximadamente a los 38 minutos), sin embargo a pH básico tienen tiempos de retención muy diferentes (alrededor de 16 y 25 minutos en el cromatograma inferior).

8. Cromatografía de Interacción Hidrofóbica.

El principio de separación en la cromatografía de interacción hidrofóbica (HIC), así como los principales grupos funcionales (C4, C8 y fenilo), son similares a la fase reversa. La diferencia entre ambas radica en que en la HIC el grado de sustitución (micro moles de ligando/mL de gel) de los ligandos hidrofóbicos es mucho menor que en RP.

Con una menor densidad de ligandos las proteínas no tendrán varios puntos de anclaje a la matriz, y por tanto, la interacción hidrofóbica en HIC es mucho más débil. Para lograr la asociación de las moléculas de proteínas a la fase estacionaria es necesario inducir la interacción hidrofóbica, para lo cual se utiliza una alta concentración de sales en el tampón inicial. El efecto es similar al que se logra cuando se tiene una solución de proteínas y estas precipitan por adición de sales (*salting out*).

La capacidad de varias sales para promover la interacción hidrofóbica es la siguiente:

$Na_2SO_4 > K_2SO_4 > (NH_4)_2SO_4 > Na_2HPO_4 > NaCl > LiCl$

Aunque los sulfatos de sodio y potasio tienen un mayor efecto, el sulfato de amonio es la sal más usada, debido a su elevada solubilidad en agua. Esto, además, es una ventaja, pues en muchas ocasiones una de las primeras etapas en la purificación de proteínas es la precipitación fraccionada con sulfato de amonio, por lo cual el siguiente paso de purificación puede ser con una columna de cromatografía de interacción hidrofóbica, sin necesidad de eliminar el sulfato de amonio residual.

La elución de las proteínas se logra con un gradiente que va disminuyendo la concentración de sales hasta terminar en 100 % de agua. Algunas proteínas no logran eluir en estas condiciones y es necesario usar una mezcla de agua con un solvente orgánico.

La resolución que se logra con esta cromatografía es mucho menor que en RP y no se usa con fines analíticos. Sin embargo, se aplica con mucha frecuencia para la purificación de proteínas, pues las condiciones de trabajo no son desnaturalizantes como en RP y en principio se puede usar para cualquier proteína.

9. Cromatografía de Afinidad.

Este tipo de cromatografía está caracterizada por una elevada especificidad de la interacción entre el compuesto que se desea separar y el ligando de la fase estacionaria. La más representativa y específica de estas interacciones es la que ocurre entre un anticuerpo y el antígeno correspondiente, que en este caso se conoce también como inmunocromatografía. Otras interacciones muy específicas que se aprovechan en la cromatografía de afinidad son la de una enzima con su inhibidor y la de una proteína con su receptor.

De acuerdo con la expresión de la resolución, $R = ¼ [(\alpha - 1)/\alpha] N^{1/2} [k/(k+ 1)]$, aquí el factor de selectividad es el mayor que se puede esperar en cualquier tipo de cromatografía, pues en una mezcla solo se unirá a la matriz el componente deseado. Los restantes componentes de la mezcla no se retienen y con un paso de lavado se elimina cualquier compuesto retenido inespecíficamente. La elución finalmente se logra con un tampón capaz de romper la interacción, sin necesidad de gradientes. En la figura 21 se muestra un cromatograma típico de una purificación por afinidad.

Esta elevada selectividad justifica que en la cromatografía de afinidad no sea imprescindible usar matrices cromatográficas de HPLC que optimicen la eficiencia. Tampoco es frecuente que esta cromatografía se use con fines analíticos.

Su principal limitación es que se necesita tener disponible el compuesto específico que se fijará a la fase estacionaria. Una vez que se tiene, se puede adquirir la matriz cromatográfica activada y lista para preparar la fase estacionaria. Adicionalmente, existen comercialmente varias matrices para cromatografía de afinidad para varios tipos de proteínas, como por ejemplo: proteína A para purificar inmunoglobulinas; calmodulina para la purificación de proteínas con dominios de unión a esa proteína y las lectinas que han sido ampliamente utilizadas para el aislamiento de glicoproteínas y otros glicoconjugados[12].

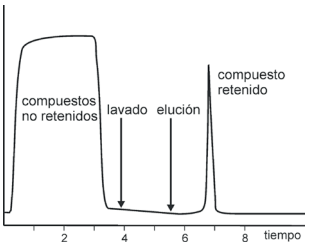


Figura 21. Cromatografía de afinidad. Cromatograma típico de una separación por cromatografía de afinidad. Al aplicar la mezcla a la columna se detectan todos los componentes que no son retenidos. Al terminar la aplicación se la lava durante algunos minutos para eliminar posibles componentes unidos inespecíficamente y finalmente se eluye el componente deseado, unido a la columna.

En el mercado están disponibles varios sistemas que se basan en la cromatografía de afinidad para la obtención y purificación de una proteína de forma relativamente sencilla y con alta pureza. Estos sistemas suministran todo lo necesario para clonar y obtener una proteína de fusión, donde una parte es la proteína deseada y la otra es una proteína (o un fragmento de una proteína), de la que también se suministra un gel de afinidad para la purificación. Uno de estos sistemas es el que está basado en la enzima glutatión S transferasa (GST). Esta enzima dimérica tiene una elevada afinidad con el glutatión. Una vez expresada la proteína de interés fusionada con la GST, se purifica en una matriz de afinidad con glutatión como ligando. La purificación se logra en un solo paso. Además el sistema incluye una proteasa que corta específicamente el sitio de unión de la proteína con la GST, la mezcla resultante se pasa nuevamente por la columna de afinidad, donde queda retenida la GST, y se puede obtener la proteína de interés pura.

Otro tipo de cromatografía de afinidad es la basada en la interacción con iones metálicos (IMAC, del inglés *Immobilized Metal Affinity Chromatography*). El ligando de la matriz es un compuesto químico capaz de formar quelatos con iones metálicos divalentes (Fe^{2+}, Ni^{2+}, Co^{2+}, Cu^{2+}, Zn^{2+}, Ga^{2+} y otros). Los ligandos más usados son el ácido nitrilotriacético (NTA) y el ácido iminodiacético (IDA).

IDA: $HN\text{-}(CH2\text{-}COOH)_2$ NTA: $N\text{-}(CH2\text{-}COOH)_3$

Las proteínas, fundamentalmente aquellas que poseen varias histidinas en su secuencia, pueden unirse a esta matriz con cierta afinidad y un paso de purificación con IMAC puede ser muy eficaz en esos casos. Para aprovechar las ventajas de esta cromatografía en la aplicación al campo de las proteínas se ideó añadir una cola de 4 o 6 histidinas en el N-terminal o en el C-terminal de una proteína recombinante[13]. De esta forma se incrementa considerablemente la afinidad, y por tanto, la eficacia de esta cromatografía, que puede lograr una alta pureza en un solo paso.

La pequeña cola de histidinas no interfiere en la actividad de muchas proteínas, por lo que estas pueden ser usadas para diferentes fines. Si no fuera así, es posible diseñar la construcción genética de la proteína, de manera que se incluya además de la cola de histidinas, un sitio de corte de una proteasa específica que pueda eliminarla. Una de las proteasas más usadas es la TEV (*tobacco etch virus*), que tiene el sitio de reconocimiento: Glu-Asn-Leu-Tyr-Phe-Gln/Gly y corta entre la glutamina y la glicina.

En proteómica, la IMAC se emplea con mucha frecuencia para el aislamiento de fosfopéptidos, pues el grupo fosfato se une con gran afinidad, aunque no muy selectivamente, por cuanto los péptidos con varios residuos ácidos (glutámico y aspártico) en la secuencia también se unen a la matriz de IMAC[14,15].

10. Cromatografía de Interacción Hidrofílica.

La cromatografía de interacción hidrofílica (HILIC, del inglés *Hydrophilic Interaction Chromatography*) es una variante de la cromatografía de fase normal. En ambos casos la fase estacionaria es hidrofílica y la elución ocurre con el uso inicial de solventes poco polares y el incremento posterior de la proporción de solventes polares. La principal diferencia radica en el uso del agua. En la fase normal "tradicional" no se emplea el agua y varios de los solventes orgánicos utilizados son inmiscibles en agua (hexano, diclorometano), mientras en la HILIC el solvente B es agua o una mezcla de agua con un solvente orgánico, y los solventes orgánicos usados tienen que ser miscibles en agua.

El mecanismo de separación, al igual que en RP, no es simple. Pudiera considerarse una cromatografía de partición entre la fase estacionaria polar y la fase móvil orgánica, pero además existen enlaces de hidrógeno, interacciones electrostáticas y otras. En dependencia de los solventes y las condiciones usadas predominarán algunos de estos factores.

En definitiva, la interacción hidrofílica nos permite racionalizar la retención en HILIC, por lo que los compuestos más polares serán más retenidos por la matriz.

En HILIC se han usado fases estacionarias con varios grupos funcionales como amida (TSKgel amide 80), intercambiadores catiónicos y aniónicos (Polysulfoethyl A y PolyWAX) e incluso gel de sílice sin derivatizar. Los solventes usados son similares a RP, pero en orden inverso. Es decir, A es un solvente orgánico (acetonitrilo para péptidos y proteínas) en alta proporción y el solvente B es fundamentalmente agua. El TFA también puede usarse en proporciones similares a RP. De hecho, con algunas matrices como la TSKgel Amide-80 se puede usar el mismo sistema de soluciones que en RP, pero con un gradiente inverso que parta, por ejemplo, de 90 % de B hasta 100 % de A.

Dependiendo del tipo de la fase estacionaria se utilizan algunos aditivos, fundamentalmente sales que deben tener alta solubilidad en los tampones con alto contenido de solvente orgánico: acetato de amonio, formiato de amonio, fosfato de trietilamina, bicarbonato de amonio y otras.

La HILIC se ha comenzado a utilizar con frecuencia para el análisis y purificación de péptidos y proteínas[16].

Al igual que en la cromatografía de RP, el pH puede tener una marcada influencia, particularmente a pH ácidos (entre 2 y 3) o alcalinos (entre 10 y 12), pues la hidrofilicidad de un péptido puede cambiar radicalmente.

Entre las principales aplicaciones de HILIC en proteómica tenemos la separación de péptidos con modificaciones postraduccionales muy polares, como la glicosilación y la fosforilación, dos de las modificaciones más frecuentes e importantes[17, 18].

11. Cromatografía Bidimensional.

La electroforesis bidimensional tiene varias limitaciones para la separación de proteínas en los estudios de proteómica. Una alternativa, presentada por primera vez por John Yates III y conocida como MudPIT (*multidimensional protein identification technology*) o *shotgun proteomics*[19], se ha convertido en la opción más utilizada en proteómica en estos momentos y consiste en

digerir la mezcla de proteínas, sin separación previa, con una enzima, generalmente tripsina. Los péptidos obtenidos son analizados por LC-MS-MS y las proteínas son identificadas a partir de los péptidos secuenciados.

La mezcla de péptidos que se obtiene es enorme y muy compleja (más de 100 000 péptidos). Una separación por RP no será suficiente y se hace necesaria la utilización de la cromatografía multidimensional, que no es más que el uso de dos (bidimensional) o más métodos cromatográficos que se basen en mecanismos de separación diferentes, es decir, que sean ortogonales.

En la cromatografía bidimensional, la separación por RP, debido a su elevada eficacia y a los solventes usados en la elución, es siempre el método final que se acopla directamente al espectrómetro de masas. Se ha sugerido que la HILIC pudiera ser más eficaz, pues los péptidos eluyen con mayor proporción del solvente orgánico que en RP, por lo que se incrementaría la eficiencia de la ionización en electrospray[20]. Sin embargo, esta práctica no ha sido seguida por otros grupos.

Para la primera dimensión se han usado fundamentalmente tres métodos: intercambio catiónico, interacción hidrofílica y RP a pH básico. La cromatografía de filtración en gel se ha usado también, aunque muy poco debido a su baja resolución. El método más usado hasta el momento ha sido la cromatografía de intercambio catiónico. En muchas ocasiones es necesario utilizar tres métodos cromatográficos, pues mientras más se pueda fraccionar la mezcla de péptidos es posible esperar que se identifique un mayor número de proteínas.

La combinación de dos métodos cromatográficos puede hacerse en línea (incluso con el uso de columnas donde se empaquetan las dos fases estacionarias) o hacer la primera dimensión separadamente, colectar un grupo de fracciones y posteriormente analizar cada fracción por LC (RP)-MS-MS. Ambas opciones son utilizadas en los estudios de proteomica aunque la última predomina. La principal ventaja de usar los dos métodos en línea es que se reduce la manipulación de la muestra y el procedimiento total es más sencillo. Por otra parte, los solventes que se usen en la primera dimensión tienen que ser compatibles con la segunda dimensión (RP), lo que muchas veces limita el uso de los solventes más apropiados. Por ejemplo, si se utiliza intercambio catiónico en la primera dimensión no podría adicionarse acetonitrilo u otro solvente orgánico a los tampones, algo que ha demostrado ser eficaz en la separación de péptidos, pues interferiría en la unión de los péptidos a la columna de RP de la segunda dimensión.

Al hacer la primera dimensión separadamente, hay mayor flexibilidad y se pueden optimizar más fácilmente las condiciones de separación y sin limitaciones.

11.1. Estudio de modificaciones postraduccionales

Otras combinaciones han sido utilizadas para el estudio de modificaciones postraduccionales. En estos casos se trata inicialmente de aislar un grupo determinado de péptidos como fosfopéptidos o glicopéptidos, que se pueden encontrar en muy bajas concentraciones.

Para el aislamiento de fosfopéptidos se han usado varias combinaciones, como son: intercambio catiónico e IMAC; intercambio catiónico y óxido de titanio; IMAC y óxido de titanio, o HILIC e IMAC. En los cuatro casos las fracciones aisladas son analizadas finalmente por RP-MS-MS. También se ha usado la combinación de inmunoprecipitación y RP, como por ejemplo, para el estudio de péptidos fosforilados en tirosinas.

Para el aislamiento de glicopéptidos se ha usado cromatografía de afinidad con lectinas y RP o HILIC y RP[21]. Es frecuente que previamente se lleve a cabo una cromatografía de afinidad con lectinas a nivel de proteína. Antes de analizar los glicopéptidos por RP se procede a la desglicosilación (con N-glycosidase F o Endoglycosidase H) para eliminar la heterogeneidad causada por las cadenas de carbohidratos.

Finalmente nos referiremos a las dimensiones de las columnas que se utilizan para la cromatografía bidimensional en proteómica. En la primera dimensión, si no se hace en línea, pueden utilizarse columnas de 1, 2 o 4.6 mm de diámetro en dependencia de la cantidad de muestra que se quiere analizar.

Cuando se usan columnas de 1 o 2 mm de diámetro es necesario tener mucho cuidado con los efectos externos a la columna, particularmente la tubería y el volumen de la celda del detector, que pueden destruir las ventajas de estas columnas que permiten manipular menores cantidades de muestra. Las tuberías para columnas de 1 o 2 mm deben ser de 0.12 mm de diámetro y el volumen de la celda de 1 µL.

Las columnas de RP que se conectan directamente al espectrómetro de masas son capilares de 50, 75 o 100 µm de diámetro.

ENLACES ÚTILES.

http://www.chemlin.net/chemistry/high_performance_liquid_chromatography.htm
http://www.separationsnow.com/coi/cda/list.cda?catId=2737&type=Link&sort=az&chld=4
http://teaching.shu.ac.uk/hwb/chemistry/tutorials/chrom/chrom1.htm
http://hplccourse.com/hplclinks.php
http://www.hplcweb.com/About_Chromatography/Lectures_and_Tutorials/
www.lcresources.com/

REFERENCIAS.

1. **Ettre LS. M.S.** Tswett and the investigation of Chromatography. LCGC North America, 2003; 21: 458-67.

2. **Cheng YF, Walter TH, Lu Z, et al.** Hybrid Organic–Inorganic Particle Technology: Breaking Through Traditional Barriers of HPLC Separations. LCGC. 2000; 18: 1162-1172.

3. **Es útil consultar las recomendaciones de la IUPAC para la nomenclatura en cromatográfica:** Commision on chromatography and other analytical separations and Commision on analytical Nomenclature. Nomenclature for Chromatography, Pure & Appl Chem. 1993; 65: 818-872.

4. **Guiochon Gs. Monolithic** columns in high-performance liquid chromatography J Chromatogr A. 2007; 1168: 101-168.

5. **Unger KK, Skudas R, Schulte MM.** Particle packed columns and monolithic columns in high-performance liquid chromatography-comparison and critical appraisal. J Chromatogr A. 2008; 1184: 393-415.

6. **Rafferty JL, Zhang L, Siepmann JI, Schure MR.** Retention Mechanism in Reversed-Phase Liquid Chromatography: A Molecular Perspective. Anal Chem. 2007; 79: 6551-6558.

7. **Baczek T, Kaliszan R.** Predictions of peptides' retention times in reversed-phase liquid chromatography as a new supportive tool to improve protein identification in proteomics. Proteomics. 2009; 9: 835-847.

8. **Petritis K, Kangas LJ, Yan B, et al.** Improved Peptide Elution Time Prediction for Reversed-Phase Liquid Chromatography-MS by Incorporating Peptide Sequence Information. Anal Chem. 2006; *78*: 5026–5039.

9. **Porath J, Sundberg L, Fornstedt N, Olson, I.** Salting-out in Amphiphilic Gels as a New Approach to Hydrophobia Adsorption. Nature. 1973; 245: 465-466.

10. **Padrón G, Besada V, Agraz A, et al.** Mass spectrometric analysis of recombinant human α-2 interferon Anal Chim Acta. 1989; 223: 361-369.

11. Liu H, Xu B, Ray MK, Shahrokh Z. Peptide mapping with liquid chromatography using a basic mobile phase. J Chromatogr A. 2008; 1210: 76-83.

12. Monzo A, Bonn GK, Guttman A. Lectin-immobilization strategies for affinity purification and separation of glycoconjugates. Trends Anal Chem. 26; 2007: 423-432.

13. Hochuli E, Bannwarth W, Döbeli H, Gentz R, Stüber D. Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. Bio/Technology. 1988; 6:1321-1325.

14. Cantin GT, Yi W, Lu B, et al. Combining Protein-Based IMAC, Peptide-Based IMAC, and MudPIT for Efficient Phosphoproteomic Analysis. J Proteome Res. 2008; 7: 1346-1351.

15. Villén J, Gygi SP. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc. 2008; 3: 1630-1638.

16. Boersema PJ, Mohammed S, Heck AJR. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal Bioanal Chem. 2008; 391:151-159.

17. Wuhrer M, de Boer AR, Deelder AM. Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. Mass Spectrom Rev. 2009; 28:192-206.

18. McNulty DE, Annan RS. Hydrophilic Interaction Chromatography Reduces the Complexity of the Phosphoproteome and Improves Global Phosphopeptide Isolation and Detection. Mol Cell Proteomics. 2008; 7: 971-980.

19. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001; 19: 242-247.

20. Naidong W. Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases. J Chromatogr B. 2003; 796: 209-224.

21. Calvano CD, Zamboninb CG, Jensen ON. Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. J Proteomics. 2008; 71: 304-317.

Técnicas electroforéticas

04

Fernando J. Corrales[2], Enrique Santamaría[1] y Joaquín Fernández-Irigoyen[1]

[1] Navarrabiomed-Fundación Miguel Servet;

[2] CIMA, Universidad de Navarra.

La electroforesis, entendida como la técnica que permite estudiar la movilidad de los biopolímeros cargados en presencia de un campo eléctrico, engloba un conjunto de técnicas analíticas que han influido de forma decisiva en el exponencial avance que han experimentado la Bioquímica y la Biología Molecular en las últimas décadas. Desde los primeros trabajos que describían la observación de fenómenos electrocinéticos a principios del siglo XIX, las técnicas electroforéticas han evolucionado hasta convertirse en uno de los pilares de disciplinas emergentes como la proteómica, en base a su capacidad para resolver de forma robusta miles de proteínas en un único experimento, facilitando así la comparación del componente proteico de dos sistemas biológicos. En este capítulo se revisarán los principios generales de la electroforesis, así como las técnicas más ampliamente utilizadas en el estudio de las proteínas.

1. Ontogénesis. Que lindo viaje electrocinético.

La historia de la electroforesis cuenta con más de doscientos años y ha sido revisada de manera brillante por diversos autores (1,2,3,4), incluyendo artículos de prosa indiscutiblemente maravillosa, de muy recomendable lectura para los amantes del tema. Los primeros fenómenos electrocinéticos fueron descritos a principios del siglo XIX por F.F. Reuss quién comprobó la migración de un soluto (arcilla en aquel caso) y el flujo de solvente que se filtraba a través de un complejo sistema de microcanales (simplemente arcilla) en presencia de un campo eléctrico. Así se describen por primera vez un proceso electroforético y electroosmótico respectivamente. Algunos años más tarde, los estudios de G.H. Wiedemann, quién formuló sus teorías para explicar el flujo electroosmótico, supusieron una nueva contribución al avance de esta naciente disciplina. Es, sin embargo, en 1937 cuando la mayoría de los autores sitúan el origen de la electroforesis moderna. Fue entonces cuando el sueco Arne Tiselius desarrolló la electroforesis de frente móvil (*Moving Boundary Electrophoresis*). Construyó un sofisticado instrumento que permitía el seguimiento de la migración de las proteínas en un campo eléctrico por un método análogo al utilizado en ultracentrifugación. El instrumento de Tiselius proporcionó un método único para medir con precisión la movilidad electroforética de las proteínas, ya que su técnica no implicaba el uso de las sustancias poliméricas típicas de la electroforesis zonal desarrollada posteriormente, que obstaculiza el movimiento de los macroiones. Sin embargo, la electroforesis zonal facilitó la separación de los componentes de una muestra, como se demostró en primera instancia aislando pigmentos amarillos a partir de veneno de serpiente utilizando papel como soporte (5). La separación electroforética en papel de biomoléculas cargadas como los

núcleotidos, núcleosidos, aminoácidos, entre otras, se llevaba a cabo en grandes láminas de papel que permitían la resolución de 50 muestras simultáneamente en 10 minutos, por lo que alcanzó gran popularidad en los años 40 y 50 del siglo pasado. En 1950 se introdujo la electroforesis en gel de agarosa (6) que inicialmente presentaba serias limitaciones derivadas del severo fenómeno electroosmótico inherente al agar crudo que se utilizaba en la preparación de los geles. La utilización de agarosa con un alto grado de pureza permitió el desarrollo de la primera técnica ortogonal que combinaba la separación de proteínas y su posterior detección inmunoespecífica en el mismo gel (7). El método tuvo cierto impacto en el ámbito de la química clínica y podría ser considerado como el embrión del actual *Western blot*. Durante estos años se utilizaron diversas matrices como el almidón (8), el acetato de celulosa (9), sílice (10) e incluso finas láminas de sephadex (11) y se comenzaron a utilizar los sistemas de tampones con pHs discontinuos (12). Indudablemente, todos estos intentos supusieron pasos adelante en la evolución de la electroforesis, pero quizás el momento culminante fue a principios de los 60 cuando Raymond introdujo los geles de poliacrilamida, polímero de notables propiedades que permitía modular el tamaño de poro del tamiz simplemente modificando su concentración. Se demostró entonces, que las proteínas se separaban en los geles de poliacrilamida de acuerdo a dos parámetros, la carga de la superficie y la masa molecular y que ambos parámetros se podían distinguir estudiando el comportamiento de las moléculas en geles de diferente porosidad (13). Unos años más adelante Shapiro y cols (14) introdujeron la utilización del SDS. Este surfactante aniónico recubre la superficie de los polipéptidos en una proporción de 1.4 mg SDS/mg proteína enjugando su carga anfotérica original. Así, las proteínas recubiertas con SDS se resuelven eficazmente en el tamiz de poliacrilamida, sobre todo si se utilizan tampones discontinuos, en estrechas bandas cuya movilidad guarda una relación lineal con el logaritmo de la masa molecular (15).

Otra de las claves en el desarrollo posterior de la electroforesis bidimensional, fue la idea de separar las proteínas en gradientes de pH, concepto que impulsó la aparición de diversas técnicas, entre ellas el isoelectroenfoque (IEF). A pesar de que los principios teóricos del IEF fueron ya establecidos en 1961 (16,17,18), se necesitó alrededor de una década para consolidar la técnica en base al desarrollo de los anfolitos portadores o anfolinas (*carrier ampholytes*) (19). Estas moléculas son iones anfotéricos capaces de alcanzar condiciones de estado estacionario a lo largo del espacio de separación formando un gradiente de pH continuo. Sin embargo, el método presentaba serias limitaciones debido a la reducida estabilidad en el tiempo de los gradientes de pH basados en la utilización de tampones anfotéricos solubles. La robustez y reproducibilidad necesarias para consolidar el método se consiguieron con el empleo de gradientes de pH inmovilizados (IPGs) en los que se empleaban reactivos bifuncionales (Inmobilinas) derivados de la acrilamida (20), posteriormente desarrollados por el grupo de Righetti (21,22,23).

La combinación de dos métodos ortogonales como el IEF y el SDS-PAGE permitió la óptima separación de cientos de proteínas según dos coordenadas: el pI y la masa molecular. Las primeras separaciones mediante electroforesis bidimensional se realizaron en 1970 utilizando condiciones no desnaturalizantes en el IEF (24). El método, tal y como se utiliza hoy en día (IEF en condiciones desnaturalizantes seguido de SDS-PAGE), fue descrito por 3 grupos de forma independiente en 1975 (25,26,27), aunque el trabajo de O'Farrell fue el más reconocido al resolver más de 1000 proteínas de un extracto de *E. coli* en un único mapa bidimensional. La técnica tal y como la empleó O'Farrell no consiguió pleno auge hasta unos años más tarde debido a la dificultad para reproducir los resultados y a limitaciones técnicas como la reducida capacidad de carga en los gradientes de pH. El desarrollo de los IPG en geles con un soporte plástico en sustitución de los tradicionales gradientes de pH con anfolinas en formato tubular (20,28) supuso la definitiva consolidación de la electroforesis bidimensional como un método robusto, reproducible y versátil, permitiendo construir gradientes de pH a la medida de la resolución requerida por el estudio.

La detección de las especies proteicas tras su separación electroforética es otro de los aspectos fundamentales, cuya evolución desde los primeros métodos, aún hoy utilizados, ha permitido, no solo incrementar en órdenes de magnitud la sensibilidad, sino detectar específicamente proteínas portadoras de modificaciones postraduccionales concretas. Dejando aparte el marcaje metabólico de las proteínas utilizando precursores radioactivos, muy habitual en los primeros tiempos, el método de tinción pionero fue el basado en la utilización de azul de Coomassie (29), que, a pesar de las limitaciones en cuanto a sensibilidad, sigue siendo ampliamente utilizado. La introducción a principios de los años 80 de la tinción basada en la utilización de iones de plata (30,31,32) consiguió incrementar la sensibilidad entre 1 y dos órdenes de magnitud con respecto a la tinción de Coomassie. A pesar de ello, el estrecho rango dinámico y la reducida linealidad de los protocolos iniciales suponían limitaciones importantes, sobre todo a la hora de cuantificar. El desarrollo de métodos de tinción basados en la utilización de compuestos fluorescentes ha contribuido a minimizar estos inconvenientes. La fluorescamina o el más sensible y estable MDPF (2-methoxy-2,4-diphenyl-3(2H)furanone) se utilizaron a finales de los 70 para visualizar proteínas no histonas asociadas a cromatina y para teñir geles bidimensionales (33). Hoy día disponemos de sondas fluorescentes para tinción de geles como las SYPRO y otras que proporcionan un nivel de sensibilidad al menos igual que los iones de plata, pero mediante un procedimiento mucho más sencillo, reproducible y con una linealidad mayor, de hasta 3 órdenes de magnitud; incluso se han desarrollado métodos de tinción fluorescente específicos para proteínas portadoras de modificaciones postraduccionales concretas, como formas fosforiladas o glicosiladas. Además de éstos y otros métodos de tinción post-electroforesis quizás menos utilizados, disponemos de procedimientos que se aplican antes de realizar el proceso de separación, que generalmente consisten en la incorporación de compuestos fluorescentes a las proteínas en estudio. Estos fluorocromos son similares estructuralmente, pero difieren en cuanto a sus propiedades espectroscópicas, lo que hace posible su combinación en un único gel y el correspondiente análisis de imagen diferencial lo que representa una indudable ventaja en términos de semicuantificación (34). El desarrollo de métodos de tinción más sensibles y cuantitativos, de métodos informáticos de análisis de imagen cada vez más precisos y automatizados (otro de los campos que se ha desarrollado de forma espectacular durante la última década) y la combinación con espectrometría de masas para la identificación de las proteínas de interés, hacen de la electroforesis bidimensional un método ampliamente utilizado en estudios de proteómica comparativa, a pesar de la aparición de estrategias alternativas basadas en separaciones cromatográficas en combinación con espectrometría de masas.

2. Principios generales de la electroforesis.

La mayoría de los polímeros biológicos poseen carga eléctrica y, por lo tanto pueden migrar bajo la influencia de un campo eléctrico. Estos polielectrolitos se pueden clasificar de forma arbitraria como débiles o fuertes, dependiendo de las constantes de ionización de sus grupos de carácter ácido y base, que, en gran medida van a condicionar su movilidad en un campo eléctrico. Sin embargo, éste no es el único factor que influye en el comportamiento electroforético de la molécula, existen otros parámetros que es necesario considerar y que complican enormemente la elaboración de una teoría de la electroforesis exacta. Las dificultades se deben a que, en general, casi todos los estudios se realizan en soluciones acuosas en presencia de iones tamponantes y sales, es decir, sustancias cargadas, cuya presencia, además de las macromoléculas en estudio, condiciona la aparición de fenómenos cuya influencia en el proceso global es difícil de cuantificar. Si consideramos una simplificación ideal, una partícula de carga q, aislada en un medio no conductor, la fuerza que experimenta en un campo eléctrico E es, de acuerdo a la ley de Coulomb:

$$F = qE$$

La partícula migrará a velocidad constante, v, que estará determinada por el balance entre la fuerza eléctrica, qE, y la resultante del rozamiento con el medio, -fv, donde f es el coeficiente de fricción.

$$Eq = fv$$

A partir de esta expresión se puede definir el parámetro movilidad electroforética, U, que corresponde a la velocidad que toma la partícula por unidad de campo:

$$U = v/E = q/f$$

Si consideramos que la partícula es esférica, la expresión anterior se puede transformar aplicando la ley de Stoke:

$$U = q/6\pi\eta R$$

donde R es el radio de la partícula y η la viscosidad del medio. En principio, dado que la movilidad depende del coeficiente de fricción que, a su vez, es función de algunos parámetros físicos de las moléculas, su estimación podría proporcionar información acerca del tamaño y forma de la partícula. Sin embargo, hemos de tener en cuenta que estas consideraciones son de aplicación en un caso ideal y que, a pesar de que parámetros como los correspondientes a la hidratación o la irregularidad de la molécula se podrían introducir para generalizar el principio enunciado en la ecuación anterior, otros factores que condicionan en gran medida el comportamiento de los macroiones en un campo eléctrico, introducen una gran complejidad en el tratamiento teórico. En solución acuosa, una macromolécula está recubierta por una película de contraiones que el campo eléctrico empujará en sentido opuesto al macroión, lo que induce una fuerza electrostática sobre el mismo, opuesta a la inducida por el campo externo. Otro de los fenómenos a tener en cuenta es el flujo electroosmótico, o migración de solvente hacia el ánodo o el cátodo, especialmente trascendente en la electroforesis capilar y gracias al cual, las partículas con carga positiva, neutras y las negativamente cargadas migrarán hacia un extremo del capilar, aunque a velocidades distintas. Se han introducido correcciones en la ecuación anterior tratando de tener en cuenta estos y otros factores, pero los resultados son siempre aproximaciones y por lo tanto, no disponemos de una expresión que defina de forma exacta la movilidad de un macroión en condiciones reales. En consecuencia, si bien la electroforesis es un excelente método de separación de macromoléculas, presenta serias limitaciones a la hora de determinar con precisión sus propiedades fisicoquímicas, tales como la forma, el tamaño o incluso la carga. Esta aproximación teórica simplificada es, sin embargo, válida para establecer los principios generales de esta técnica: la movilidad electroforética se incrementa con la carga, decrece frente al coeficiente de fricción y se anula para moléculas sin carga neta.

3. Electroforesis en geles de poliacrilamida en condiciones desnaturalizantes, SDS-PAGE.

Las técnicas electroforéticas son herramientas ampliamente utilizadas en el laboratorio para la separación, aislamiento, caracterización e identificación de biomoléculas. Actualmente, la modalidad más extendida es la electroforesis zonal cuyo principio general es la utilización de una matriz a través de la cual migrará una fina capa de la molécula de interés. A lo largo de la historia se han desarrollado distintas variantes pero en la actualidad el método de separación más ampliamente utilizado en química de proteínas y, por lo tanto en proteómica, es la SDS-PAGE.

3.1. Los geles y sistemas discontinuos de tampones

Los geles de poliacrilamida son polímeros de acrilamida y un agente entrecruzante, generalmente N,N'-metilenbisacrilamida. El persulfato amónico y la N,N,N',N'-tetrametiletilendiamina (TEMED) son los responsables de catalizar la reacción de polimerización. Sustituyeron a otros soportes como el almidón, la celulosa, etc, por dos razones: presentar una adsorción del material proteico prácticamente despreciable y, sobre todo, por proporcionar el fácil control del poro del tamiz, simplemente modificando la concentración total de acrilamida A y el grado de entrecruzamiento B:

$$A = (x + y) * 100/V\ [\%], \quad B = y * 100/(x + y)\ [\%]$$

donde x e y corresponden a las masas de acrilamida y metilenbisacrilamida respectivamente (en g) y V al volumen (en ml). Si B permanece constante el aumento de A supone la disminución del tamaño del poro. Variaciones de B para un valor de A constante inducen variaciones del tamaño de poro de acuerdo a una función parabólica con valores máximos para B alto o bajo y mínimo alrededor del 4%. La concentración de los catalizadores es otro parámetro a tener en cuenta para asegurar la reproducibilidad de los geles. Es importante utilizar una concentración de TEMED superior a la de persulfato amónico ya que este último puede reaccionar con el Tris a pH básico, efecto que minimiza el exceso de TEMED.

Las proporciones de A y B generalmente utilizadas son del 12.5% y 3% respectivamente para generar un tamiz que permite un nivel de resolución adecuado para las proteínas entre 15 y 100 kDa. Aunque no es fácil predecir la concentración de acrilamida correcta (35), en términos generales la disminución de A mejorará la resolución de proteínas de elevado peso molecular y viceversa, el aumento de A proporcionará mayor resolución en el rango de bajo peso molecular. Los geles en gradiente de acrilamida proporcionan un intervalo de separación y de estimación de masa molecular más amplios, así como un nivel mayor de resolución debido a que el tamaño del poro decrece continuamente.

La resolución que se alcanza en un gel típico de acrilamida se obtiene por el uso de un gel de concentración y de un sistema de tampones discontinuo (Figura 1). Sobre el gel de resolución, preparado según los principios anteriormente mencionados, se polimeriza un gel de poliacrilamida de menor concentración y, por lo tanto, de mayor tamaño de poro, en el que se aplicará la muestra. Además, éste último se prepara en un tampón Tris/HCl a pH 6,7, mientras que el corespondiente al gel de resolución es un Tris/HCl a pH 8,8. Otro de los aspectos clave es la presencia de glicina en el tampón de carrera. La glicina tiene un pK para el grupo amino de 9,6, por lo que su protón amino estará débilmente disociado a pH por debajo de 9, por lo tanto en su forma zwitterionica sin carga con una movilidad electroforética muy reducida. La eficiencia del sistema para producir bandas estrechas depende de que la movilidad de la glicina sea menor que la de los iones Cl^-. Cuando aplicamos el voltaje, los iones Cl^- migran hacia el ánodo a mayor velocidad que la glicina, lo que podría producir una discontinuidad en la concentración de iones y la consecuente caída de la conductividad en el frente. Sin embargo, como la corriente debe de mantenerse constante en todo el gel, este efecto es compensado mediante un incremento del gradiente de voltaje de forma que se produce un fino frente de glicina-Cl^- que migra de forma homogénea. En estas condiciones, debido al bajo pH del tampón de concentración, las moléculas de proteína tendrán movilidades intermedias entre la glicina y el Cl^- y formarán estrechos apilamientos entre ambos. Como en este caso la concentración de acrilamida es baja, no se produce un tamizado molecular efectivo, ni la consecuente separación por tamaño. La entrada en el gel de resolución, se producen dos importantes cambios: un incremento de pH que favorece la ionización de la glicina y el consecuente aumento de su movilidad y el aumento de la concentración de acrilamida que dificulta la migración de las proteínas. Ambos factores facilitan que la glicina migre en un frente común con el Cl^- por delante a las proteínas que se separarán por filtrado molecular en un tampón glicina uniforme. Así, cada proteína entrará en el gel de resolución como una fina banda concentrada, aunque ésta fuera más difusa cuando la

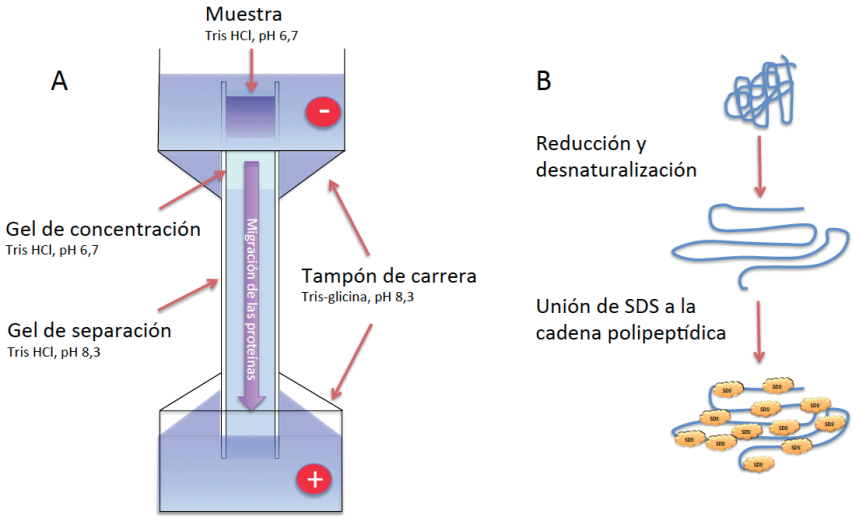


Figura 1. Electroforesis de proteínas en geles de poliacrilamida, en condiciones desnaturalizantes, SDS-PAGE. Esquema representativo de un sistema de electroforesis vertical (A) en gel de poliacrilamida en un sistema de tampones discontinuo: el gel de concentración de menor porcentaje de acrilamida, a pH 6,7 y el gel de resolución, en la parte inferior, con un porcentaje de acrilamida mayor, generalmente 10-12%, y pH 8,3. La muestra, representada en azul oscuro en la parte superior del gel, se prepara en tampón de carga y se desnaturaliza por calentamiento y con SDS (B). El SDS se une en una proporción de 1,4g SDS/g de proteína, mediante interacciones iónicas y elimina la carga intrínseca de los polipéptidos que quedan con la misma densidad de carga negativa y, en consecuencia, migrarán en el gel en relación directa a su tamaño molecular.

muestra se cargó en el gel. Este es el sistema descrito por Laemmli (36) que se ha modificado para fines específicos como el incremento de resolución por debajo de 10 kDa sustituyendo la glicina por tricina como ion de terminación, o la mejora de la separación de glicoproteínas utilizando un tampón Tris-borato (37).

3.2 Preparación de la muestra, separación de las proteínas y aplicaciones

El método más popular para estimar el tamaño de un polipéptido es midiendo su movilidad electroforética (14). La movilidad de una macromolécula a través de un tamiz al aplicar un campo eléctrico depende no solo de su tamaño sino también de su carga neta y de su forma (38). Estos dos últimos parámetros pueden ser normalizados mediante calentamiento, el tratamiento con agentes reductores como el 2-mercaptoetanol y con el detergente dodecil sulfato sódico (SDS) (39). De esta forma se consigue eliminar la estructura terciaria/cuaternaria de las proteínas mediante la liberación de los grupos sulfidrilo implicados en puentes disulfuro intra e intercatenarios y el efecto desnaturaliazante del SDS que impone formas y densidades de carga neta comparables. El principio físico que fundamenta el efecto del SDS en las proteínas se desconoce, el método se basa en observaciones empíricas. El SDS al 1% desnaturaliza la mayoría de las proteínas y se une a ellas manteniendo una relación constante de 1,4 g SDS/g proteína (40). El tipo de interacción que se establece entre la proteína y el detergente es predominantemente iónica y elimina la carga intrínseca del polipéptido. Como la tasa de unión de los distintos complejos SDS/proteína es constante, su densidad de carga negativa será la misma. Además, estos complejos tienen forma similar; la unión con el SDS induce una cierta desnaturalización que resulta en partículas de proteína elongadas y flexibles, de diámetro constante y longitud proporcional al número de residuos (Figura 1). Las proteínas con pI extremos, de carácter muy ácido y básico, son excepciones ya que sus complejos con el detergente pueden mantener contribuciones de la carga intrínseca de la proteína.

Como consecuencia de las mencionadas propiedades, los complejos SDS-proteína presentan una movilidad únicamente relacionada con el peso molecular del polipéptido debido a la acción del gel como tamiz molecular. Así, las proteínas se separan en bandas y se cumple que la distancia recorrida es inversamente proporcional al logaritmo de su peso molecular. Esta relación permite la estimación del tamaño de polipéptidos, con una desviación del 10%, en base a la migración de proteínas patrón de peso molecular conocido, mediante interpolación en la recta de regresión, como se muestra en la figura 2. Desviaciones del comportamiento ideal, es decir el cálculo de pesos moleculares aparentes distintos del esperado, pueden proporcionar información sobre propiedades del polipéptido en estudio. Por ejemplo, se ha propuesto que el incremento de carga negativa superficial puede dificultar la interacción de la proteína con el SDS resultando en una menor migración electroforética y, en consecuencia, en un peso molecular aparente mayor de lo esperable de acuerdo a la secuencia. Así se ha demostrado para algunas proteínas fosforiladas (41) y en el caso de polimorfismos que resultan en la introducción de residuos acídicos en la secuencia (42). Por otra parte, modificaciones en el protocolo de preparación de la muestra pueden aportar información adicional relativa a la conformación de la proteína. Por ejemplo, es posible determinar la presencia de puentes disulfuro comparando la movilidad de una proteína en condiciones reductoras y no reductoras, sin tratamiento con mercaptoetanol o DTT en el proceso de preparación (43). Asimismo, la separación electroforética en condiciones nativas, sin tratamiento con SDS ni desnaturalización por calor, puede resultar de utilidad para caracterizaciones preliminares de estructuras cuaternarias, aunque es preciso tener en cuenta que en este caso la migración electroforética no es únicamente función del peso molecular de la longitud de la cadena de aminoácidos, la carga intrínseca y la forma del polipéptido también contribuyen y complican la interpretación del resultado. En este caso, no se tiene porqué cumplir la relación inversa entre la migración electroforética y el logaritmo del peso molecular.

4. Isoelectroenfoque.

Las proteínas son moléculas anfotéricas, es decir poseen tanto grupos con carga positiva como negativa cuyo balance, dependiente del pH, determina la carga neta del polipéptido. En un en-

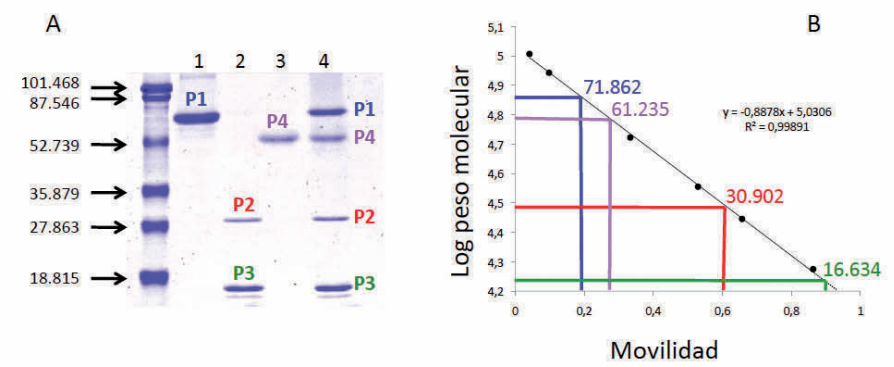


Figura 2. Cálculo de la masa molecular de polipéptidos a partir de su migración electroforética. Ejemplo de un gel SDS PAGE teñido con azul de Coomassie (A) en el que se muestran los marcadores de peso molecular (1), cuatro proteínas distintas analizadas de forma independiente (2-4) y una mezcla de las mismas (5). La estimación de la masa molecular de las proteínas analizadas se calculó de acuerdo a su movilidad electroforética relativa al frente de azul de bromofenol (distancia migrada/distancia migrada por el frente) (B) por interpolación en la recta de regresión construida a partir de las movilidades de los marcadores de peso molecular conocido.

torno ácido la proteína presentará carga positiva mientras que si el pH es básico estará cargada negativamente. Como para cualquier anfolito, existirá un valor de pH en el que la carga neta sea nula, valor al que se denomina punto isoeléctrico (pI). De acuerdo a esta propiedad, si una proteína se somete a un campo eléctrico en presencia de un gradiente de pH, migrará hacia el ánodo o el cátodo, según su carga neta, hasta que encuentre el valor de pH que iguala su pI, momento en el que pierde su carga y cesará su migración (Figura 3). Cualquier difusión fuera del pI supondría la adquisición de carga y, por lo tanto, el retorno del polipéptido a la posición de carga cero. Este es el conocido fenómeno de enfoque, que explica la resolución del método y permite la separación de mezclas de proteínas con razonable eficacia.

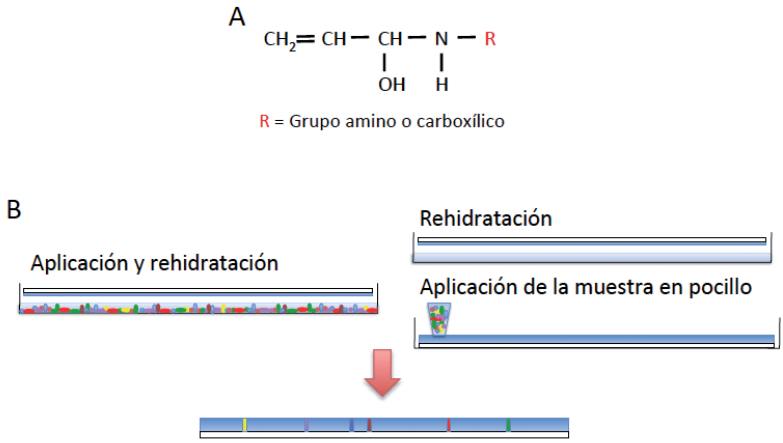


Figura 3. Enfoque isoeléctrico de proteínas. Fórmula general de una inmobilina (A). Esquema representativo de los dos métodos posibles para aplicar la muestra a la tira IPG, aplicación y rehidratación simultánea y carga con pocillo (*cup loading*) (B).

4.1 Generación de los gradientes de pH

Uno de los aspectos fundamentales del enfoque isoeléctrico es la generación de gradientes de pH estables. Si el gradiente se formase mediante la simple difusión de dos soluciones de distinto pH, o mezclando dos tampones de forma análoga a como se preparan los gradientes de concentración, su estabilidad sería limitada porque los iones que lo forman migrarían en el campo eléctrico y, por lo tanto, únicamente se podrían separar macromoléculas que migrasen a mayor velocidad que los iones del gradiente. La solución viene dada por la utilización de polianfolitos, que son moléculas multicargadas sintéticas de bajo peso molecular que cubren un amplio rango de pI. Tradicionalmente se han seguido dos métodos para generar los gradientes de pH: utilizando anfolitos portadores en solución que generan el gradiente al aplicar el campo eléctrico, o con gradientes de pH inmovilizados, en cuyo caso los grupos tamponantes están unidos al propio gel.

El primer método fue desarrollado en los años 60 (17,18,44) gracias a la síntesis de una mezcla heterogénea de isómeros de ácidos alifáticos oligoamino-oligocarboxílicos, de bajo peso molecular, buenos conductores de la electricidad, con valores próximos de pI, donde presentan una alta capacidad tamponante y que no presentan interacciones biológicas. Inicialmente, el gel con los anfolitos portadores presenta un valor promedio de pH, pero al aplicar un campo eléctrico, los anfolitos de carga negativa migrarán hacia el ánodo y los positivos hacia el cátodo hasta alinearse entre ambos electrodos de acuerdo a su pI, determinando el pH de su entorno y, en consecuencia, generando el gradiente en el rango de pH deseado, por ejemplo, entre pH 3 y 10. Para mantener la estabilidad del gradiente, se emplea una solución ácida en el ánodo y básica en el cátodo. De esta forma, si un anfolito acídico llegase al ánodo, su grupo básico

tamponante se protonaría y la molécula sería atraída de nuevo hacia el cátodo. Sin embargo, el enfoque isoeléctrico es un proceso relativamente lento por la reducida movilidad de las proteínas en el entorno de su pI. Durante este tiempo prolongado se produce una deriva del gradiente de pH debido a que los anfolitos portadores están en solución, lo que genera patrones dependientes del tiempo, se pierden proteínas básicas y se inducen adaptaciones del gradiente dependientes de la muestra debido al carácter anfotérico de las proteínas que pueden comportarse como anfolitos portadores adicionales.

Estos problemas se resolvieron mediante la utilización de gradientes de pH inmovilizados (IPG) que se generan con las inmobilinas, ácidos o bases débiles, derivadas de acrilamida, con capacidad tamponante que copolimerizan con la matriz del gel (Figura 3) (45). Las constantes de disociación de las inmobilinas acídicas están en el rango de pK entre 0,8 y 4,6 y para las básicas, entre 6,2 y 12 y son necesarias al menos una ácida y una básica para tamponar un valor de pH específico. El gradiente de pH se obtiene variando de forma continua la proporción de inmobilinas, resultando en los valores de pH definidos por la ecuación de Henderson-Hasselbach:

$$pH = pK_B + log\ (C_B - C_A) / C_A$$

$$pH = pK_A + log\ C_B / (C_A + C_B)$$

si las inmobilinas tamponantes son una base o un ácido, respectivamente, donde C_A es la concentración de inmobilina ácida y C_B la correspondiente a inmobilina básica. Los primeros intentos de los grupos de Righetti (46) y Hochstrasser (47) utilizando IPGs en capilares o tubos de gel no obtuvieron resultados óptimos, pero los geles en un soporte plástico o tiras IPG, introducidos en 1982 (20) y desarrollados posteriormente por el grupo de Görg (48) hasta los dispositivos comerciales disponibles hoy día, de diferentes longitudes y rangos de pH, aseguran mediciones altamente reproducibles de forma relativamente sencilla.

4.2 Aplicación de la muestra y separación de las proteínas

Las tiras IPG comerciales contienen un gel seco que ha de ser rehidratado antes de realizar el análisis. En este proceso se utilizan tampones con altas concentraciones de agentes caotrópicos como la urea, detergentes no iónicos, agentes reductores y una pequeña proporción de anfolitos (49). En cuanto a la muestra, preparada en un tampón desnaturalizante de composición similar al de rehidratación, su aplicación en la tira de separación puede realizarse simultáneamente a la rehidratación (*sample in-gel rehydration*) o una vez realizada ésta, en un punto específico de la tira utilizando un pocillo de plástico que se coloca sobre el gel (*cup-loading*) (Figura 3). La primera opción es un proceso lento porque los poros del gel son pequeños inicialmente por lo que las moléculas como la urea, el detergente, etc, penetran en la matriz del gel más rápidamente que las proteínas, muchas de las cuales difunden posteriormente en el gel completamente rehidratado. Este método presenta complicaciones en el caso de proteínas muy hidrofóbicas, alcalinas y de elevado peso molecular, si bien la entrada en el gel de proteínas de gran tamaño se mejora aplicando un voltaje bajo (30-50 V) durante el proceso de carga/hidratación. En el segundo caso, es recomendable aplicar la muestra en las zonas de pH extremos, es decir, cerca de alguno de los electrodos. Se suelen obtener mejores resultados cuando la muestra se aplica cerca del ánodo, algo que es imprescindible si se utilizan gradientes de pH básicos.

Las condiciones de separación en el IEF se deben ajustar a la naturaleza y composición de la muestra. Factores como la sobrecarga, un tiempo de enfoque inadecuado, oxidaciones de cisteínas o la presencia de sales, ácidos nucleicos, fosfolípidos, etc, pueden producir rayas (streaking) horizontales en las bandas de proteína. Factores como la temperatura y el campo

eléctrico han de ser ajustados con precisión. A pesar de que en el IEF las proteínas a analizar están desnaturalizadas, es importante trabajar a una temperatura definida ya que se ha comprobado que la posición de enfoque de algunas proteínas, principalmente las de pI extremos, puede variar de forma temperatura dependiente (50). En general es recomendable trabajar utilizando dispositivos que permiten mantener de forma precisa una temperatura de 20º C, incluso en las zonas de la tira en las que se produce un calentamiento local debido a la formación de los frentes de iones (sales y otras sustancias iónicas de la muestra). A esta temperatura no se forman cristales de urea y se previene la carbamilación de las proteínas que podría producirse por calentamiento durante el IEF. En cuanto a las condiciones eléctricas, la corriente se suele limitar a 50 μA por tira para evitar el sobrecalentamiento de las tiras, sobre todo en ciertas partes del gradiente. El proceso se controla programando el voltaje en varias etapas. En una primera etapa (1-2 h o hasta 5 si el contenido en sales es alto), dependiendo de la cantidad de proteína, volumen a aplicar y conductividad, se utiliza un voltaje bajo (150 V) para asegurar la correcta entrada de la muestra (si se utiliza *cup loading*) evitando fenómenos de agregación y precipitación de proteína, así como el calentamiento que podrían producir las sales presentes en la preparación. Los iones salinos migran rápidamente hacia los electrodos permitiendo utilizar gradientes de voltaje de hasta 12.000 V. En una tercera etapa se mantiene el valor máximo de voltaje hasta alcanzar el valor de Vh adecuado, en función del gradiente de pH, la distancia de separación o la cantidad de proteína cargada en la tira. Si el valor Vh es bajo, el enfoque no ha sido suficiente y se observarán deformaciones horizontales en las bandas. Si por el contrario si el valor Vh es demasiado alto, se pueden producir oxidaciones de cisteínas, ruptura de algunas proteínas que son inestables en su pI y deformaciones en la zona de pH superior a 10.

5. Electroforesis bidimensional.

La electroforesis bidimensional (2DE) consiste en la combinación de dos métodos electroforéticos ortogonales para incrementar la resolución en los análisis de mezclas complejas de proteínas. Desde los experimentos pioneros de O'Farrell, que consiguió resolver varios cientos de proteínas de un extracto de *E. coli* en un único análisis, la 2DE combina, de forma general, una primera separación de las proteínas mediante enfoque isoeléctrico y una segunda dimensión basada en SDS-PAGE. Las proteínas extraídas en tampones de lisis con composiciones estandarizadas (urea/tiourea, agentes reductores, detergentes no iónicos) se separan de acuerdo a su pI en tiras IPG, éstas se equilibran en tampones que contienen agentes reductores y yodocaetamida para evitar posteriores oxidaciones en residuos de cisteína y SDS; una vez equilibradas, las tiras se cargan en la parte superior del gel de acrilamida, en el que se realizará la separación por tamaño molecular, de acuerdo a los principios descritos en el apartado previo.

La 2DE ha sido y es, uno de los procedimientos base del análisis comparativo del proteoma de sistemas biológicos. Ello es debido a su capacidad de proporcionar una instantánea del proteoma de la muestra en estudio basada en la detección simultanea de miles de proteínas (Figura 4). Sin embargo, la información obtenida está lejos de reflejar la complejidad del proteoma en toda su extensión, existen limitaciones inherentes a la técnica en sí y a la propia complejidad de la muestra biológica (48). Las proteínas con valores de pI por encima de 11 y por debajo de 3 exceden la capacidad de resolución de las procedimientos estándar, aunque se han conseguido resolver proteínas con pI de 12 (51). Los geles habitualmente utilizados en la segunda dimensión, incluso los geles en gradiente de acrilamida, no son eficaces con proteínas por debajo de 7.000 o por encima de 120.000 Da. Asimismo, la heterogeneidad de las especies proteicas en cuanto a sus propiedades físico-químicas, así como el amplio rango dinámico del proteoma, de más de 10 órdenes de magnitud, son factores que condicionan de forma decisiva la capacidad de análisis de la 2DE. En un experimento convencional se pueden obtener

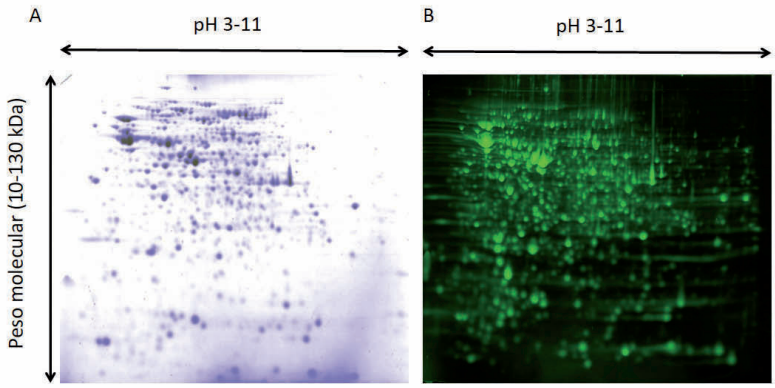


Figura 4. Separación de proteínas mediante electroforesis bidimensional y visualización. Resolución de un extracto de E. coli y visualización mediante azul de Coomassie (A) y Sypro Ruby (B).

mapas de unas 2.500 proteínas con un claro sesgo hacia la detección de proteínas hidrofílicas, abundantes y con valores de pI y tamaño molecular en los rangos señalados anteriormente. La cobertura se puede ampliar de varias maneras: aumentando la resolución del análisis mediante la utilización combinada de tiras IPG de rango estrecho de pH, analizando diferentes fracciones generadas mediante una extracción seriada de las proteínas constituyentes, en función de su diferente solubilidad, por ejemplo, eliminando proteínas mayoritarias para detectar las de menor representación (ver detalles en el capítulo de preparación de muestra), analizar de forma independiente las distintas fracciones subcelulares, etc. En general todas estas aproximaciones procuran un aumento de la resolución o una disminución de la complejidad de la muestra en cada análisis individual, o ambas, pero suponen un incremento en el tiempo de ejecución y coste del experimento.

A pesar de que los protocolos están muy optimizados, en ocasiones se pueden producir deformaciones de los mapas bidimensionales que pueden tener diversos orígenes. Las deformaciones localizadas en zonas concretas del gel suelen deberse a problemas en la polimerización de la acrilamida (incompleta o demasiado rápida). Las rayas horizontales pueden resultar de una mala solubilización de las proteínas de la muestra, de la presencia de sustancias ionicas contaminantes que interfieran en el enfoque, de un exceso de carga de la tira, o de un enfoque defectuoso por el empleo de un tiempo escaso o excesivo. Las rayas verticales se pueden producir por un deficiente equilibrado de las tiras IPG tras el enfoque, demasiado corto o con poco SDS, o por la utilización de tampones para la segunda dimensión de composición incorrecta; la precipitación de la proteína durante el proceso de *cup loading*, que ocurre si se utilizan voltajes bajos o altos, también puede producir esta distorsión. En ocasiones pueden aparecer en el gel 2D franjas verticales en blanco que suelen ser originadas por impurezas en la muestra o en el tampón de rehidratación o a la presencia de una burbuja en la interfase entre la tira IPG y el gel de acrilamida (52).

6. Métodos de visualización.

La visualización de las manchas proteicas en los geles bidimensionales ha de realizarse mediante métodos que permitan un análisis robusto y eficaz, tanto desde una perspectiva cualitativa como cuantitativa. Los métodos de tinción deberían, idealmente, ser suficientemente sensibles como para detectar proteínas de bajo número de copias, permitir análisis cuantitativos, ofrecer

una amplia linealidad y rango dinámico, ser compatible con posteriores análisis de espectrometría de masas, no ser tóxico, ser respetuoso con el medio ambiente y ser barato (53,54). A pesar de que ninguno de los métodos disponibles en la actualidad cumple todos estos requisitos, se han desarrollado reactivos que han supuesto indudables mejoras en muchos de los mencionados aspectos. Los métodos de tinción más comúnmente utilizados se pueden clasificar de acuerdo a que el tipo de colorante utilizado sea visible o fluorescente.

6.1. Visualización de proteínas con colorantes visibles.

Uno de los reactivos más utilizados para la detección de proteínas, ampliamente utilizado aún hoy en día en electroforesis es el azul de Coomassie (CBB, *Coomassie Brilliant Blue*). Existen dos variantes, la R-250 y la forma dimetilada G-250, que en condiciones acídicas interaccionan con los grupos amino de las proteínas mediante interacciones electrostáticas e hidrofóbicas. El CBB (Figura 4) es compatible con análisis de espectrometría de masas posteriores, ofrece un rango lineal de un orden de magnitud aproximadamente y una capacidad de detección de 10-100 ng de proteína. La sensibilidad de esta tinción se ve incrementada hasta 10-30 ng con CBB coloidal. El método fue introducido por Diezel y cols (55) aunque el protocolo más ampliamente utilizado es el desarrollado por Neuhoff y cols (56) en el que se utilizan ácido fosfórico y CBB G en presencia de persulfato amónico. La utilización del colorante en estado coloidal estabiliza su unión a la proteína y reduce la señal de fondo. Modificaciones de este método han permitido incrementar la sensibilidad (*blue silver* o CBB fluorescente) y desarrollar aplicaciones como el estudio de complejos de proteínas permitiendo su elución posterior del gel (*blue native electrophorsis*). Otros colorantes utilizados, aunque menos frecuentemente, son el amido Black y el Fast Green FCF, éste último utilizado para teñir geles de IEF ya que no interacciona con los anfolitos.

Las tinciones de plata se comenzaron a usar a finales de los 70 (30,31,32) y consisten en saturar los geles con iones de plata, eliminar de la matriz aquellos unidos más débilmente y reducir los unidos a las proteínas para formar plata metálica. Existen un gran número de variantes para esta técnica que se pueden agrupar en dos categorías, las tinciones de plata alcalina y acídica. En el caso de los protocolos alcalinos, se utiliza un complejo diamínico de nitrato de plata que se disuelve en un entorno básico y el revelado se realiza en soluciones acídicas diluidas de formaldehído (57). Para los métodos acídicos, se emplea una solución acuosa de nitrato de plata y para el revelado, una solución de formaldehído a pH básico (58,59). La utilización de uno u otro método de tinción con plata dependerá de las necesidades específicas del estudio ya que cada variante presenta ventajas e inconvenientes concretos (60). En general, debido a que no es un procedimiento a punto final, las intensidades observadas pueden variar de gel a gel, aunque, dependiendo del protocolo, se pueden llegar a detectar 100 pg - 1 ng de proteína por mancha en el gel, sensibilidad muy superior a la ofrecida por el método de tinción con Coomassie. Sin embargo, su rango dinámico sigue siendo solo de uno o dos órdenes de magnitud, por lo que no es ideal para cuantificar; si la tinción no se controla cuidadosamente se alcanza la saturación y se produce un efecto de tinción negativa (efecto donut) en las manchas más abundantes. Una de las limitaciones para la aplicación de este tipo de tinción en proteómica era su incompatibilidad con los ulteriores análisis de espectrometría de masas para identificar las proteínas de interés. Por ello se desarrollaron protocolos acídicos que mejoran este aspecto, eliminando el glutaraldehído en las soluciones de sensibilización y el formaldehído en las soluciones de nitrato de plata (61).

Las tinciones negativas en las que las proteínas quedan desteñidas sobre un fondo blanquecino semi opaco, representan una tercera alternativa que consiste en sumergir el gel SDS en una solución de una sal de cobre, cobalto níquel o zinc. Éste último es el más ampliamente utilizado y ofrece una sensibilidad similar a la de la tinción de plata. La tinción negativa con zinc, permite la visualización de proteínas y otros biopolímeros como oligonucleótidos y glicopéptidos, a través de la asociación del metal con grupos cargados o polarizables como His, Cys, Glu and Asp en proteínas, fosfatos y grupos amino en ácidos nucleicos, polihidroxilos y carboxilatos y fosfatos en

lipopolisacáridos (62). El metal presente en la matriz del gel, no unido a las biomoléculas, reaccionará con imidazol en una posterior incubación para formar el imidazolato correspondiente que proporcionará el color de fondo característico. El método es compatible con análisis de espectrometría de masas y, al ser reversible, permite la elución de las proteínas con razonable eficacia.

6.2 Visualización de proteínas con colorantes fluorescentes.

Históricamente se han utilizado distintos colorantes fluorescentes para teñir proteínas (33), como la fluorescamina. En electroforesis, hoy día se utilizan principalmente las sondas SYPRO cuyas primeras versiones, Rojo y Naranja, se unen a la película de detergente asociada a las moléculas de proteína procesadas en geles de SDS en condiciones desnaturalizantes. Más recientemente se desarrolló el SYPRO Ruby, basado en la utilización de un reactivo sintético formado por complejos orgánicos de metales tipo rutenio que se une a los residuos de aminoácidos básicos de las proteínas de acuerdo a un mecanismo similar al CBB coloidal (Figura 4). Se puede visualizar con luz UV o con láser, ofrece un rango dinámico de 3 logaritmos, es tan sensible como la mejor tinción con plata, pero es más sencillo de utilizar, más reproducible y es totalmente compatible con estudios de espectrometría de masas posteriores. Más recientemente se ha desarrollado el método *Deep Purple Total Protein Stain* que se basa en la utilización de una sustancia aislada del hongo *Epicoccum nigrum* (63), la epicoconona, una azofilona que emite fluorescencia al interaccionar con las proteínas, más concretamente con las colas lipofílicas del SDS unido a ellas. Con respecto al SYPRO Ruby, proporciona una sensibilidad y rango lineal mayores, pero su fotoestabilidad es menor, por lo que se pierde la señal más rápidamente tras recibir radiación UV para su visualización.

Además de los referidos métodos para teñir de manera general todas las proteínas resueltas en el gel 2D, existen tinciones fluorescentes para determinadas modificaciones postraduccionales como la fosforilación y la glicosilación. En muchas ocasiones las desviaciones de los pI o masa molecular experimentales frente a los teóricamente calculados en base a la secuencia de la proteína analizada, son indicativas de modificaciones postraduccionales. Sin embargo, para asignar correctamente las modificaciones, sobre todo en el contexto de un mapa 2D, es necesario contar con métodos de tinción específicos. En el caso de la fosforilación los anticuerpos específicos para pSer/pThr y para pTyr son ampliamente utilizados para reconocer dichos residuos en réplicas de los geles 2D en soportes de nitrocelulosa o PVDF. Uno de los métodos pioneros para teñir residuos de pSer y pThr, pero no pTyr, consiste en teñir con verde metilo los complejos nitrofosfo-molibdato generados a partir de los residuos fosforilados (64). El método, a pesar de haber demostrado su eficacia para algunas proteínas modelo, presenta serias limitaciones relacionadas con su distinto comportamiento en diferentes polipéptidos. Más recientemente se ha desarrollado el colorante fluorescente Pro-Q Diamond® que proporciona una especificidad más amplia, un rango lineal de 3 ordenes de magnitud, una capacidad de detección de 1 a 16 ng de fosfoproteína por mancha, un protocolo sencillo y reacción con pSer, pThr y pTyr sin condicionantes asociados a la secuencia o contexto molecular del residuo. A pesar de las mencionadas ventajas, es necesario tener en consideración que no hay una correlación perfecta entre las fosfoproteínas teñidas con Pro-Q Diamond y las marcadas con ^{32}P (65). El procedimiento clásico para teñir glicoproteínas es la tinción PAS (*periodic acid-Schiff*) que se desarrolló inicialmente como una técnica para tinción inmunohistoquímica y se adaptó posteriormente para la detección de glicoproteínas (66) aunque su sensibilidad es realmente reducida, 1-10 μg de proteína con alto grado de glicosilación. Alternativamente, la tinción con Azul de Alcian, una ftalocianina de cobre que se utilizó inicialmente en microscopía electrónica, con una capaciadad de detección de unos 50 ng de polisacáridos ácidos. Hoy día disponemos de la tinción fluorescente Pro-Q Emerald, cuyo límite de detección llega a los 300 pg y, como otras tinciones fluorescentes tiene un rango lineal de 2-3 logaritmos. Existen dos variedades con distinta longitud de onda de excitación máxima, que son, en ambos casos, hidracinas fluorescentes y que se conjugan con las proteínas de acuerdo a un mecanismo PAS pero siguiendo un procedimiento más sencillo.

REFERENCIAS.

1.Penque, D. (2009). Two-dimensional gel electrophoresis and mass spectrometry for biomarker discovery. *Proteomics Clinical Applications* 3, 155-172.

2. Righetti, P. G. (2004). Bioanalysis: its past, present, and some future. *Electrophoresis* 25, 2111-2127.

3.Righetti, P. G. (2005). Electrophoresis: the march of pennies, the march of dimes. *J Chromatogr A* 1079, 24-40.

4. Vesterberg, O. (1989). History of electrophoretic methods. *J Chromatogr* 480, 3-19.

5. von Klobusitzky, D. & König, P. (1939). Biochemical study of snake venom from the genus *Bothrops*. *Arch. Exptl. Pathol. Pharmakol.* 192, 271-275.

6. Gordon, A. H., Keil, B., Sebasta, K., Knessel, O. & Sorm, F. (1950). Electrophoresis of proteins in agar jelly. *Coll. Czech. Chem. Com.* 15, 1-16.

7. Grabar, P. & Williams, C. A. (1953). [Method permitting the combined study of the electrophoretic and the immunochemical properties of protein mixtures; application to blood serum]. *Biochim Biophys Acta* 10, 193-194.

8. Smithies, O. (1955). Zone electrophoresis in starch gels: group variations in the serum proteins of normal human adults. *Biochem J* 61, 629-641.

9. Kohn, J. (1957). A cellulose acetate supporting medium for zone electrophoresis. *Clin Chim Acta* 2, 297-303.

10. Honegger, C. G. (1961). Dünnschicht-ionophorese und dünnschicht-ionophorese-chromatographie. *Helv. Chim. Acta* 44, 173-179.

11. Fasella, P., Giartosio, C. & Turand, C. (1964). In *Thin layer chromatography* (Marini-Bettolo, G. B., ed.), pp. 105. Elsevier, Amsterdam.

12. Poulik, M. D. (1957). Starch gel electrophoresis in a discontinous system of buffers. *Nature* 180, 1477-1479.

13. Ferguson, K. A. (1964). Starch-gel electrophoresis. Application to the classification of pituitary proteins and polypeptides. *Metab. Clin. Exp.* 13, 985.

14. Shapiro, A. L., Vinuela, E. & Maizel, J. V., Jr. (1967). Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochem Biophys Res Commun* 28, 815-820.

15. Chrambach, A. & Rodbard, D. (1971). Polyacrylamide gel electrophoresis. *Science* 172, 440-451.

16. Svensson, H. (1961). Isoelectric fractionation,analysis, and characgterization of ampholytes yn natural pH gradients.I. The differential equation of solute concentrations at a steady state and its solution for simple cases. *Acta Chemica Scandinavica* 15, 325-341.

17. Svensson, H. (1961). Isoelectric fractionation, analysis, and characterization of ampholytes in natural pH gradients. I. The differential equation of solute concentrations at a steady state and its solution for simple cases. *Acta Chemica Scandinavica* 15, 325-341.

18. Svensson, H. (1962). Isoelectric fractionation, analysis, and characterization of ampholytes in natural pH gradients.II. Buffering capacity and conductance of ioionic ampholytes. *Acta Chemica Scandinavica* 16, 456-466.

19. Righetti, P. G. (1983). *Isoelectric focusing: theory, methodology and applications.*, Elsevier, Amsterdam.

20. Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Gorg, A., Westermeier, R. & Postel, W. (1982). Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J Biochem Biophys Methods* 6, 317-339.

21. Dossi, G., Celentano, F., Gianazza, E. & Righetti, P. G. (1983). Isoelectric focusing in immobilized pH gradients: generation of extended pH intervals. *J Biochem Biophys Methods* 7, 123-142.

22. Gianazza, E., Astrua-Testori, S. & Righetti, P. G. (1985). Some more formulations for immobilized pH gradients. *Electrophoresis* 6, 113-117.

23. Gianazza, E., Giacon, P., Sahlin, B. & GRighetti, P. G. (1985). Non-linear pH courses with immobilized pH gradients. *Electrophoresis* 6, 53-56.

24. Kenrick, K. G. & Margolis, J. (1970). Isoelectric focusing and gradient gel electrophoresis: a two-dimensional technique. *Anal Biochem* 33, 204-207.

25. Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* 26, 231-243.

26. O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250, 4007-4021.

27. Scheele, G. A. (1975). Two-dimensional gel analysis of soluble proteins. Charaterization of guinea pig exocrine pancreatic proteins. *J Biol Chem* 250, 5375-5385.

28. Gorg, A., Postel, W. & Gunther, S. (1988). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 9, 531-546.

29. Fazekas de St Groth, S., Webster, R. G. & Datyner, A. (1963). Two new staining procedures for quantitative estimation of proteins on electrophoretic strips. *Biochim Biophys Acta* 71, 377-391.

30. Merril, C. R., Switzer, R. C. & Van Keuren, M. L. (1979). Trace polypeptides in cellular extracts and human body fluids detected by two-dimensional electrophoresis and a highly sensitive silver stain. *Proc Natl Acad Sci U S A* 76, 4335-4339.

31. Oakley, B. R., Kirsch, D. R. & Morris, N. R. (1980). A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels. *Anal Biochem* 105, 361-363.

32. Switzer, R. C., 3rd, Merril, C. R. & Shifrin, S. (1979). A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Anal Biochem* 98, 231-237.

33. Jackson, P., Urwin, V. E. & Mackay, C. D. (1988). Rapid imaging, using a cooled charge-coupled-device, of fluorescent two-dimensional polyacrylamide gels produced by labelling proteins in the first-dimensional isoelectric focusing gel with the fluorophore 2-methoxy-2,4-diphenyl-3(2H)furanone. *Electrophoresis* 9, 330-339.

34. Unlu, M., Morgan, M. E. & Minden, J. S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18, 2071-2077.

35. Langen, H., Roder, D., Juranville, J. F. & Fountoulakis, M. (1997). Effect of protein application mode and acrylamide concentration on the resolution of protein spots separated by two-dimensional gel electrophoresis. *Electrophoresis* 18, 2085-2090.

36. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

37. Poduslo, J. F. (1981). Glycoprotein molecular-weight estimation using sodium dodecyl sulfate-pore gradient electrophoresis: comparison of tris-glycine and tris-borate-EDTA buffer systems. *Anal Biochem* 114, 131-139.

38. Blattler, D. P. & Reithel, F. J. (1970). Molecular weight determinations and the influence of gel density, protein charges and protein shape in polyacrylamide gel electrophoresis. *J Chromatogr* 46, 286-292.

39. Weber, K. & Osborn, M. (1975). Proteins and SDS: molecular weight determination on polyacrilamide gels and related procedures. In *The Proteins* (Neurath, H., Hill, R. L. & Border, C., eds.). Academic Press, New York, USA.

40. Cereighton, T. E. (1993). Chemical properties of polypeptides. In *Proteins. Structures and molecular properties*, pp. 1-48. Freeman.

41. Kuroda, F., Moss, J. & Vaughan, M. (2007). Regulation of brefeldin A-inhibited guanine nucleotide-exchange protein 1 (BIG1) and BIG2 activity via PKA and protein phosphatase 1gamma. *Proc Natl Acad Sci U S A* 104, 3201-3206.

42. Munoz, J., Fernandez-Irigoyen, J., Santamaria, E., Parbel, A., Obeso, J. & Corrales, F. J. (2008). Mass spectrometric characterization of mitochondrial complex I NDUFA10 variants. *Proteomics* 8, 1898-1908.

43. Freifelder, D. (1981). Electroforesis. In *Técnicas de Bioquímica y Biología Molecular*, pp. 234-260. Editorial Reverté, Barcelona.

44. Vesterberg, O. (1969). Synthesis and isoelectric fractionation of carrier ampholytes. *Acta Chemica Scandinavica* 23, 2653-2666.

45. Righetti, P. G. (1990). *Immobilized pH gradients: Theory and Methodology*, Elsevier, Amsterdam.

46. Bossi, M. L., Bossi, O., Gelfi, C. & Righetti, P. G. (1988). Casting immobilized pH gradients into cylindrical polyacrylamide gels. *J Biochem Biophys Methods* 16, 171-183.

47. Hochstrasser, D., Augsburger, V., Funk, M., Appel, R., Pellegrini, C. & Muller, A. F. (1986). Immobilized pH gradients in capillary tubes and two dimensional gel electrophoresis. *Electrophoresis* 7, 505-511.

48. Gorg, A., Drews, O., Luck, C., Weiland, F. & Weiss, W. (2009). 2-DE with IPGs. *Electrophoresis* 30 Suppl 1, S122-132.

49. Altland, K. & Rossmann, U. (1985). Hybridoisoelectri focusing in rehydrated immobilized pH gradients with added carriers amplolytes: demonstration of human globins. *Electrophoresis* 6, 314-325.

50. Gorg, A., Postel, W., Friedrich, C., Kuick, R., Strahler, J. R. & Hanash, S. M. (1991). Temperature-dependent spot positional variability in two-dimensional polypeptide patterns. *Electrophoresis* 12, 653-658.

51. Gorg, A., Obermaier, C., Boguth, G., Csordas, A., Diaz, J. J. & Madjar, J. J. (1997). Very alkaline immobilized pH gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. *Electrophoresis* 18, 328-337.

52. Simpson, R. J. (2003). Preparative two-dimensional gel electrophoresis woth immobilized pH gradients. In *Proteins and proteomics: a laboratory manual* (Simpson, R. J., ed.). Cold Spring Harbor Laboratory Press.

53. Harris, L. R., Churchward, M. A., Butt, R. H. & Coorssen, J. R. (2007). Assessing detection methods for gel-based proteomic analyses. *J Proteome Res* 6, 1418-1425.

54. Miller, I., Crawford, J. & Gianazza, E. (2006). Protein stains for proteomic applications: which, when, why? *Proteomics* 6, 5385-5408.

55. Diezel, W., Kopperschlager, G. & Hofmann, E. (1972). An improved procedure for protein staining in polyacrylamide gels with a new type of Coomassie Brilliant Blue. *Anal Biochem* 48, 617-620.

56. Neuhoff, V., Stamm, R. & Hansjorg, E. (1985). Clear Background and higly sensitive protein staining with Coomassie Blue dyes in polyacrylamide gels: a systematic analysis. *Electrophoresis* 6, 427-448.

57. Hochstrasser, D. F., Patchornik, A. & Merril, C. R. (1988). Development of polyacrylamide gels that improve the separation of proteins and their detection by silver staining. *Anal Biochem* 173, 412-423.

58. Blum, H., Beier, H. & Gross, J. H. (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* 8, 93-99.

59. Heukeshoven, J. & Dernick, R. (1985). Simplified method for silver staining of proteins in polyacrylamide and the mechanism of silver staining. *Electrophoresis* 6, 103-112.

60. Rabilloud, T. (1990). Mechanisms of protein silver staining in polyacrylamide gels: a 10-year synthesis. *Electrophoresis* 11, 785-794.

61. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68, 850-858.

62. Hardy, E. & Castellanos-Serra, L. R. (2004). "Reverse-staining" of biomolecules in electrophoresis gels: analytical and micropreparative applications. *Anal Biochem* 328, 1-13.

63. Bell, P. J. & Karuso, P. (2003). Epicocconone, a novel fluorescent compound from the fungus epicoccumnigrum. *J Am Chem Soc* 125, 9304-9305.

64. Cutting, J. A. & Roth, T. F. (1973). Staining of phospho-proteins on acrylamide gel electropherograms. *Anal Biochem* 54, 386-394.

65. Chen, Z., Southwick, K. & Thulin, C. D. (2004). Initial analysis of the phosphoproteome of Chinese hamster ovary cells using electrophoresis. *J Biomol Tech* 15, 249-256.

66. Zacharius, R. M., Zell, T. E., Morrison, J. H. & Woodlock, J. J. (1969). Glycoprotein staining following electrophoresis on acrylamide gels. *Anal Biochem* 30, 148-152.

Métodos de análisis estadístico en proteómica

05

Luis Valledor[1,2], Jesús V. Jorrín-Novo[3], y Mónica Meijón[4]*

[1] Departamento de Biología Universidad de Aveiro, Aveiro, Portugal.

[2] Global Change Research Center, Academia de las Ciencias de la República Checa, Brno, República Checa.

[3] Bioquímica y Proteómica Vegetal y Agrícola. Dpto. de Bioquímica y Biología Molecular, Universidad de Córdoba, España.

[4] Programa Forestal. Servicio Regional de Investigación y Desarrollo Agroalimentario (SERIDA), Finca "La Mata", Asturias, España.

[*] Autor para correspondencia: Luis Valledor. Website: http://www.valledor.info E-mail: luis@valledor.info

1. Introducción.

Una de las principales disciplinas proteómicas es la proteómica cuantitativa, cuyo objeto es la determinación de la abundancia de las distintas especies proteicas, así como la comparación de las mismas entre distintos orígenes, individuos, poblaciones, tratamientos,… En definitiva tratar de responder a la preguntas ¿Hay (o no hay) diferencias entre mis grupos muestrales? ¿Qué proteínas se acumulan de forma diferencial? ¿Existe alguna relación entre los cambios de las distintas proteínas? Siendo estas cuestiones clave para posteriormente poder analizar el significado biológico de las diferencias encontradas en nuestro sistema biológico.

En la actualidad para realizar proteómica cuantitativa se pueden emplear distintas plataformas que, pese a diferir tanto en los protocolos empleados para la separación de proteínas (solubilización diferencial, punto isoeléctrico, electroforesis en gel o cromatografía líquida) como en las técnicas para la cuantificación de las mismas (análisis de imagen o espectrometría de masas), poseen el objetivo común de establecer perfiles proteicos correspondientes a individuos o poblaciones, tanto desde una perspectiva cualitativa (presencia o ausencia) como cuantitativa. A lo largo de este manual de proteómica el lector encontrará detallada información referente a las distintas técnicas de análisis proteómico.

A pesar de las mejoras técnicas alcanzadas en los últimos años, sobre todo en lo referente al enriquecimiento y validación de bases de datos (Mead et al., 2009), al aumento de la potencia y facilidad de uso de los espectrómetros de masas (Han et al., 2008; Mueller et al., 2008) y a la reproducibilidad en electroforesis bidimensional y los sistemas cromatográficos y de espectrometría de masas; el diseño experimental y el análisis de datos se perfilan como dos de los puntos más débiles a ser considerados para la mejora de los estudios proteómicos (Taylor et al., 2007; Valledor y Jorrín, 2011).

El diseño del experimento es un aspecto crucial, puesto que determinará no solo el análisis estadístico sino también el tipo de preguntas que podamos realizar a nuestro sistema (Valledor y Jorrin, 2011). Si queremos obtener la máxima información de nuestro experimento deberemos, además de planificar un cuidado diseño, aplicar correctamente las distintas técnicas estadísticas, teniendo en cuenta las particularidades de los datos proteómicos.

Los datos obtenidos en proteómica son generalmente analizados empleando exclusivamente estadística univariable (t-test, Mann-Whitney, ANOVA, Kruskal-Wallis), a pesar de que estos mé-

todos no son los ideales, pues aumentan la posibilidad de falsos positivos y no pueden predecir tendencias o detectar relaciones entre proteínas (Chich et al., 2007). Por otro lado los análisis multivariables (por ejemplo Análisis de Componentes Principales, Independientes, Regresión de Mínimos Cuadrados Parciales, Mapas auto organizativos) están descritos como más efectivos en estos casos, debido a su capacidad para reducir la complejidad de los datos y predecir tendencias y relaciones entre variables, siendo estos tests más tolerantes a la estructura de los datos proteómicos (Rodríguez-Piñeiro et al., 2007; Grove et al., 2008).

Actualmente muchos investigadores consideran suficientemente resolutivas las herramientas estadísticas incluidas en el software comercial destinado a analizar imágenes de geles bidimensionales o cromatogramas. Sin embargo, pese a que estos programas son fáciles e intuitivos en su manejo suelen estar limitados en el apartado estadístico, con un bajo número de pruebas estadísticas disponibles (sobre todo multivariables) y muestran un bajo rendimiento en el análisis de muestras imperfectas (geles que muestran irregularidades, artefactos o valores perdidos, entre otros) (Berth et al. 2007). Además, en la mayoría de las ocasiones, no tienen en cuenta la variación inherente a la digitalización o al software empleado (Stessl et al., 2009).

En este capítulo se presenta una guía enfocada a maximizar el poder de los estudios basados en electroforesis bidimensional de proteínas y proteómica libre de geles y etiquetas (gel-free, label free), mediante el empleo de un adecuado diseño experimental y de herramientas estadísticas adecuadas a cada situación. Los distintos apartados en los que se divide el capítulo (diseño experimental, pre-procesado de datos, análisis univariable y análisis multivariable) se desarrollan desde un punto de vista teórico y práctico, mediante el empleo del software libre, principalmente R (véase cuadro 1). En las distintas secciones además se proponen numerosos ejercicios con el fin de guiar al alumno mientras profundiza en su aprendizaje.

2. Diseño Experimental.

El diseño experimental se puede considerar la base sobre los que se va a asentar el desarrollo de cualquier ensayo así como el análisis de los datos obtenidos. En consecuencia un buen diseño experimental considerará los siguientes factores:

- Las fuentes de variación existentes. Esta variación puede ser técnica, biológica o dependiente de los distintos tratamientos aplicados (véase cuadro 2).
- La capacidad operativa del laboratorio, es decir, cuantas muestras como máximo podrían procesarse en paralelo.
- Los test estadísticos que se deseen aplicar, así como el nivel de diferencia mínimo que se desee detectar entre muestras.
- El empleo del mínimo número de réplicas que permita satisfacer el apartado anterior, a fin de no malgastar recursos y esfuerzo.

Existen distintos tipos de réplicas que podremos asociar con el tipo de ruido-variación dentro del sistema. Las réplicas técnicas (o analíticas) pueden definirse como medidas repetidas de la misma muestra biológica, y su objetivo es obtener un valor de medición verdadero. Este tipo de réplicas se recomienda para los casos en los que existe mucha variación asociada a la metodología de análisis. Las réplicas biológicas pueden definirse como muestras procedentes de individuos diferentes sometidos al mismo tratamiento, siendo su función recoger la variación

biológica existente en campo. En algunas situaciones, por ejemplo cuando el interés del estudio recae sobre la población y no el individuo, cuando no es posible extrae suficiente proteína de un individuo aislado o cuando la variación intrapoblacional es muy alta, se pueden combinar muestras procedentes de varios individuos formando "pooles". Este procedimiento se basa en asumir como correcto el promedio biológico de las distintas muestras, por lo que si es necesario hacer pooles es preferible disponer de varios pooles pequeños a uno único de mayor tamaño. De esta forma es posible estimar la varianza entre los mismos intra-tratamiento y los valores atípicos pueden ser identificados (Kendziorski et al., 2005).

Una pregunta muy frecuente cuando se plantea un diseño experimental es la siguiente ¿Cuántas réplicas analíticas y biológicas debo realizar? Como norma general se recomienda que, siempre que el número de geles a realizar sea el factor limitante, la replicación biológica es siempre preferible a la replicación analítica y, sobre todo, si esta última ocurre en detrimento del número de réplicas biológicas. La replicación analítica debería reservarse solo para aquellos casos en los que la metodología aplicada muestra demasiado ruido. Desde aquí recomendamos la lectura de algunos trabajos que discuten este tema en profundidad (Karp et al., 2005; Rowell et al., 2005; Horgan, 2007; Valledor y Jorrin, 2011).

El número de réplicas biológicas a realizar dependerá en gran medida de la potencia estadística (o poder estadístico) que deseamos conseguir. La potencia puede ser definida como la habilidad de un test estadístico para detectar la variación (1-β; donde β es la tasa de falsos negativos) y depende del número de réplicas, de la varianza intra-tratamiento, de la magnitud de la variación a estudiar y del nivel de significación deseado. En consecuencia, para estimar la potencia estadística necesitaremos conocer la varianza intra-tratamiento o, al menos, una aproximación a dicho valor. Estos datos pueden obtenerse de estudios preliminares o bien de estudios similares ya publicados. Existen numerosas herramientas que nos permiten calcular la potencia de un ensayo, tanto gratuitos (módulos para R, Lenth's power tool; Lenth, 2001) como comerciales (nQuery Advisor, PASS, Power and Precission). Una potencia estadística superior a 0.8 es óptima para los estudios en proteómica (Karp et al., 2005) puesto que significa que si el experimento se repitiese 5 veces la variación se detectaría al menos en 4 de ellas.

- **Ejercicio 1**

  En este ejercicio el lector deberá calcular empleando una herramienta de cálculo de potencia estadística, por ejemplo la contribuida por Russ Lenth (Lenth, 2001; apéndice 1), el número de réplicas necesario para obtener una potencia de 0.8 en la aplicación de una prueba t (α=0,05) a dos muestras con una desviación estándar (SD) de 0,5 en los siguientes supuestos:

  a] La diferencia entre medias que queremos detectar es de 0,1.
  b] La diferencia entre medias que queremos detectar es igual a la SD.
  c] La diferencia entre medias que queremos detectar es el doble de la SD.
  d] Manteniendo fijo el número de réplicas en 5, observar el cambio en la potencia estadística a medida que se incrementa o reduce la diferencia mínima que se desea ser capaz de detectar.

  RESULTADO: Consultar apéndice 2.

Como el lector habrá podido percibir mediante la realización del ejercicio, los experimentos con un bajo número de réplicas no tendrán la capacidad de catalogar como estadísticamente significativas ciertas variaciones importantes, por otro lado el "gigantismo experimental", debido al aumento innecesario del número de réplicas no tendría ningún efecto sobre la potencia estadística. En ambos casos se estarían malgastando recursos y esfuerzo lo que justificaría, pues,

el análisis previo de la potencia estadística. La diferencia mínima a detectar es dependiente de cada caso experimental, por lo que desde aquí no se realizará ninguna recomendación específica en cuanto al número de réplicas a realizar. Estudios recientes han analizado la potencia estadística empleando distintos métodos, muestras y tamaños (Karp et al., 2007; Fuxius et al., 2008; Eravci et al., 2009).

Durante la realización de un experimento pueden existir distintas fuentes de variación o tendencias ocultas, en ocasiones difícilmente detectables, que pueden conducir a un error sistemático en la interpretación de los datos. Incluso para los investigadores más experimentados la identificación de estas fuentes de variación es difícil, siendo necesario el empleo de la aleatorización para controlar estas variaciones. La aleatorización es otro de los pilares del diseño experimental, puesto que reduce la variación en el estudio ecualizando otros factores obviados como la deriva instrumental.

Una causa importante de variación es la debida a los artefactos o heterogeneidades técnicas tales como diferencias en la extracción de proteínas, isoelectroenfoque (IEF), polimerización y carrera de geles. La aleatorización de cada una de las etapas del desarrollo experimental permitirá reducir el impacto de estas variaciones, aumentando la precisión del experimento. Además se debe tener en cuenta que la mayoría de las ocasiones no podremos procesar a la vez todas las muestras que forman parte de un experimento, bien sea por no disponer de los recursos técnicos o suficiente personal, siendo entonces necesario realizar distintas tandas o bloques. Los distintos bloques, que podemos definirlos como un conjunto de muestras para las que todas las etapas técnicas se han desarrollado en paralelo minimizando las derivas instrumentales y manipulatorias, deberán también aleatorizarse e incluir muestras pertenecientes a todos los tratamientos. Si existiese una heterogeneidad debida a las distintas etapas técnicas esta podría ser identificada comprando entre si las muestras correspondientes al mismo tratamiento pero analizadas en distintos bloques. Además esta variación podría ser cuantificada, empleando este dato para mejorar la precisión general del análisis. Este tipo de diseños se ha aplicado con efectividad en 2D-PAGE (Karp et al., 2005; Chich et al., 2007).

El último punto fundamental relativo al diseño experimental que trataremos en este apartado es la importancia del muestreo y procesado de muestras. Este aspecto es determinante sobre todo cuando el objeto del estudio son individuos complejos o de gran tamaño. Por ejemplo Jorge et al. (2005) demostraron variaciones en el proteoma de la hoja de la encina debidas al muestreo a diferentes horas del día, o diferentes orientaciones geográficas. Por otro lado, pese a ser un hecho evidente conviene recalcar que el mismo protocolo de extracción de proteínas deberá ser mantenido a lo largo de todo el desarrollo del experimento. Por ejemplo se ha determinado que la diferente polimerización entre geles es la mayor causa de variación, influyendo en la determinación de hasta un 16% de los *spots* considerados como diferenciales. Tandas independientes de isoelectroenfoque y electroforesis pueden ocasionar variaciones próximas al 50%, pero un sistema experimental adecuado, incluyendo un correcto diseño por bloques y aleatorización puede reducir este efecto hasta una variación próxima al 5% (Eravci et al., 2007; 2009). Este resultado ilustra la importancia práctica del diseño experimental para mejorar la fiabilidad de los estudios cuantitativos basados en 2-DE.

3. Pre-procesado y descripción de datos.

Los conjuntos de datos obtenidos tras el análisis de imagen de geles bidimensionales o la cuantificación de espectros suelen caracterizarse por no ajustarse a una distribución normal, presentando un porcentaje elevado de valores perdidos y una relación intensidad de un *spot*/desviación estándar elevada (Valledor y Jorrín, 2011). Estas peculiaridades en los datos afecta-

rán a la elección de los test estadísticos que queramos aplicar, limitándonos en principio al uso de pruebas no paramétricas. En el caso de la estadística multivariable los resultados se verían distorsionados por el hecho de que los *spots* mayores agruparían entre si puesto que presentan la mayor relación intensidad/desviación estándar (Grove et al., 2006; Meunier et al., 2007). Un adecuado pre-procesado de datos mitigará estos efectos en la mayor parte de las situaciones, por lo que es necesaria su aplicación antes de realizar cualquier análisis estadístico.

En algunos trabajos de los últimos años se analizan y explican las causas que originan que los datos obtenidos a partir de geles bidimensionales presenten estas. Por ejemplo, la dependencia existente entre la intensidad de un *spot* y su desviación estándar está causada por el proceso de digitalización de los geles (Grove et al., 2006). Los *spots* más grandes necesitarán un mayor número de pixeles para ser digitalizados. A mayor número de pixeles mayor probabilidad de varianza técnica, puesto que la intensidad de un *spot* se calcula a partir de la fórmula:

$$IntensidadSpot = \sum_{i=1}^{n} Px_i$$

(Siendo *Px* la intensidad del pixel *i* y *n* el número de pixeles correspondientes al *spot*).

El caso de los valores perdidos es más difícil de explicar, el por qué desaparece un *spot* en un gel mientras que en el resto de réplicas sí aparece. Se han propuesto distintas inconsistencias técnicas o de manipulación de la muestra (IEF imperfecto, fallo de transferencia de proteínas de la primera a la segunda dimensión, degradación de la muestra, variaciones cromatográficas, artefactos de la espectrometría, …) como causa de este fenómeno, pero en la actualidad no existe una demostración práctica consistente.

A lo largo de esta sección se tratará de describir, de manera ordenada, los pasos necesarios para reducir la influencia de estos fenómenos sobre nuestros conjuntos de datos. Además se presentarán las herramientas básicas para caracterizar nuestros datos que definirán su normalidad, homocedasticidad y dispersión, facilitando la identificación de los aquellos valores atípicos para su posterior eliminación del análisis.

3.1. Imputación de valores perdidos:

El manejo de los valores perdidos es una de las fases más controvertidas del pre-procesado de datos, debido a la gran frecuencia con la que aparecen y a la influencia que tienen en el resultado final de los análisis.

¿Qué hacer cuando tenemos un valor perdido? En la actualidad se recomiendan dos aproximaciones que son dependientes de la intensidad media que presente *spot* o pico a imputar. Para el caso de variables muy débiles, al borde del límite de detección, se aconseja sustituir el valor perdido por el valor más bajo de los obtenidos dentro del grupo de réplicas correspondientes a ese tratamiento y variable.

Para el resto de valores perdidos se recomienda emplear un algoritmo de imputación, siendo una de las aproximaciones más extendidas el algoritmo de los k-vecinos más cercanos (KNN, k-nearest neighbor) (Albrecht et al., 2010). Este algoritmo analiza en primer lugar todas las variables de nuestra muestra, buscando aquellas que se comportan de forma más próxima a la variable que queremos imputar para definir los vecinos más cercanos. A partir de estos y de los valores del *spot* a imputar disponibles en el resto de réplicas se determinará su valor. Como regla general podemos decir que se deberá evitar la imputación de valores perdidos cuando estos supongan más del 20-25% del total y que ésta deberá hacerse tratando de forma independiente los distintos grupos experimentales (i.e. control, tratamiento1, tratamiento2,…). El empleo de los

4-6 vecinos más cercanos en el cálculo puede considerarse correcto para set de datos grandes, con más de 400 variables.

Dentro del entorno estadístico R existen numerosos paquetes para la imputación de valores perdidos, nosotros utilizaremos el paquete SeqKNN. Una vez descargado e instalado R, deberemos instalar el paquete SeqKNN y posteriormente cargarlo. Ambas acciones están disponibles en el menú "Paquetes" dentro del entorno R.

A continuación cargaremos nuestros datos en el programa. El proceso de importación y exportación de datos en R ha sido abordado por numerosos tutoriales (Cuadro 2). Para cargar automáticamente los datos emplearemos el siguiente comando:

```
> tabla1 <- dget("http://www.valledor.info/wp-content/uploads/2013/12/
tabla1.txt")
```

Para ver la tabla escribiremos:

```
> tabla1
```

En caso de existir cualquier problema con la importación de la tabla 1, al final del capítulo se encuentra un apéndice con todos los datos empleados durante el mismo (apéndice 1). Esta tabla se construyó definiendo las variables 1-7 con los siguientes valores (0,5; 0,8; 0,5; 1,0; 1,1; 1,2; 1,5) y aleatorizándose ± 10% para cada uno de los geles (G1-G6).

Una vez cargados los datos el uso del comando SeqKNN es sencillo, necesitando emplear solamente dos argumentos que en R se separan mediante el empleo de una coma (","): el primer argumento es el nombre objeto que contiene la tabla de datos (tabla1 en nuestro ejemplo) y el segundo el número de vecinos más próximos a emplear en el cálculo (2 en nuestro ejemplo):

```
> tabla1.imp <- SeqKNN(tabla1,2)
```

1

Mediante el comando anterior se crea un nuevo objeto "tabla1.imp" en el cual se guardará el resultado de ejecutar SeqKNN (esto se indica con "<-"). El programa devuelve el valor 1, indicando que solo se imputó un valor perdido y se generó "tabla1.imp", que incluye los valores imputados. Para ver estos valores escribiremos:

```
> tabla1.imp
        G1       G2        G3        G4        G5        G6
[1,] 0.5400   0.4500  0.500000    0.4500    0.5400    0.4800
[2,] 0.8720   0.8480  0.880144    0.8480    0.8720    0.8240
[3,] 0.8636   0.8090  0.836300    0.9546    0.9091    0.9546
[4,] 0.9184   0.9302  1.080200    1.0306    1.0714    1.0714
[5,] 1.1460   1.0540  1.008000    1.2150    1.1115    1.0540
[6,] 1.2889   1.3270  1.149200    1.0857    1.0857    1.2508
[7,] 1.6600   1.6440  1.356000    1.5640    1.4040    1.3880
```

Empleando los 2 vecinos más próximos el valor imputado es de 0,880144 Como ejercicio final proponemos al lector que halle los valores que se obtendrían empleando 1, 3 o 4 vecinos. En este caso, al disponer de solo 7 variables, vemos como perdemos precisión cuando empleamos a los 3 o 4 vecinos más próximos.

3.2 Normalización de los datos:

La normalización podemos definirla como un procedimiento encaminado a corregir las diferencias entre las distintas muestras que no son debidas a causas biológicas. Estas diferencias pueden ser causadas, por ejemplo, por cargar distintas cantidades de proteína, por una digestión parcial de las mismas en el caso de LC-MS, por tener distintas eficiencias en la tinción de los geles o incluso por diferencias entre las distintas digitalizaciones de los mismos.

Estudios basados en 2D-PAGE

El proceso más simple de normalización, dentro de cada gel, consiste en dividir las intensidades individuales de cada spot respecto a la suma total de intensidades de spots para cada gel.

Este procedimiento controla los supuestos descritos anteriormente pero asume que la misma cantidad de proteína ha sido cargada en todos los geles. En algunos casos se ha descrito que la suma total de intensidades de proteína se ve distorsionada por unos pocos *spots* demasiado intensos. Una posible solución para este problema consiste en considerar estos *spots* como valores atípicos, eliminándose del estudio (Berth et al., 2007).

Estudios basados en (gel-free) label-gel free LC-MS/MS

La Proteómica libre de geles y libre de marcaje (gel free, label free) se basa en la cuantificación de la intensidad o del número de espectros de masas que están asignados a los pépidos empleados para identificar cada proteína particular. Se parte de la base que las proteínas más abundantes generarán también los péptidos más abundantes en la muestra. Un aumento del número de péptidos generará un aumento del spectral counting y también de su intensidad. La aproximación más sencilla para el procesado de estos datos consiste en el empleo del factor normalizado de abundancia espectral (NSAF, normalized spectral abundance factor, Paoletti et al. 2007) la cual es calculada de la siguiente forma:

$$(NSAF)_k = (PSM/L)_k \Big/ \sum_{i=1}^{N} (PSM/L)_i$$

dónde el número total de espectros asignados a la proteína k (PSM) es dividido por la longitud de la proteína en aminoácidos (L), y posteriormente este valor es dividido por la suma de los PSM/L para todas las proteínas.

Esta aproximación es altamente dependiente de las bases de datos empleadas, ya que solo se cuantificarán aquellos iones que se hayan asignado a un péptido conocido que forme parte de una proteína (Valledor et al. 2012). El análisis ideal requeriría de la disposición del genoma del individuo a analizar, asi como de todos sus modelos génicos. En caso de no estar disponible una base suficientemente amplia del genoma, o en caso de buscar una cuantificación más precisa y pese a que existen formas más complejas y sofisticadas de solventar este problema, una aproximación sencilla consiste en analizar la intensidad de los distintos iones presentes en la muestra. Una vez estos iones estén cuantificados, podrán asignarse posteriormente a una proteína específica. Protmax es un software sencillo que permite alinear los espectros correspondientes a múltiples muestras y realizar tanto contabilizar los espectros como determinar su intensidad. Esta aproximación se aleja del objeto de este capítulo, pero desde aquí podemos recomendar su descarga y la realización de los distintos tutoriales si se desea explorar esta alternativa de cuantificación (ver Egelhofer et al. 2013).

Independientemente de la aproximación empleada, los valores numéricos obtenidos tras la aplicación de cualquiera de estas dos aproximaciones podrán ser considerados para su análisis estadístico, directamente o tras una transformación, del mismo modo que los ejemplos que ilustran el resto del capítulo.

3.3. Transformación de los datos

En caso de que exista una correlación entre la intensidad media de los variables y su desviación, esta puede corregirse de manera sencilla transformando nuestros datos. En primer lugar deberá realizarse una representación gráfica de la intensidad y la desviación para cada una de las variables y mediante una regresión estimar la correlación existente. La tabla1 fue diseñada simulando este efecto. La representación de la media de intensidad de cada variable frente a su desviación muestra una cierta correlación (R^2>0,7; Figura 1a). Para reducir esta dependencia se aplicó una transformación logarítmica a los datos (Figura 1b). En este caso se puede apreciar una notable reducción de la misma.

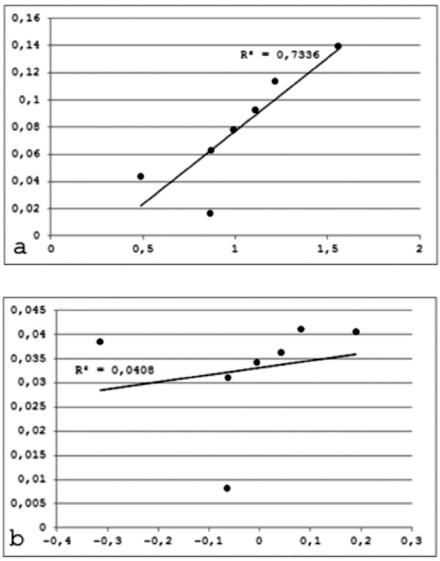


Figura 1: (a) Correlación entre intensidad media de spots y su desviación. (b) Correlación entre intensidad media y los spots y su desviación después de aplicar transformación logarítmica a los datos

A priori, no puede determinarse que transformación funcionará mejor para unos datos concretos, por lo que desde aquí se recomienda aplicar dos o tres transformaciones distintas y comparar los resultados obtenidos. No obstante, las dos transformaciones empleadas con mayor frecuencia en el tratamiento de datos 2-DE y LC-MS son la raíz cúbica y el logaritmo (Meleth et al., 2005; Valledor et al., 2008; Curto et al., 2010; Valledor et al. 2013).

3.4. Eliminación de ruido

En la mayor parte de las ocasiones nos encontramos ante spots o picos muy débiles, muy próximos al límite de detección y que solo aparecen en una o dos muestras. En estos casos puede resultar conveniente eliminar estas variables antes de continuar. De forma rutinaria, para considerar una variable se debe cumplir al menos una de estas dos condiciones: a) que esté presente en todas las réplicas de un tratamiento (o en >80% cuando se realizan 5 o más réplicas); b) que esté presente en, al menos, un 25% del total de muestras analizadas. De esta forma eliminaremos variables que no nos van a aportar ninguna información biológica, puesto que no están presentes en un número significativo de muestras, y al mismo tiempo reducimos el ruido estadístico que tendremos al realizar un análisis multivariable. Obviamente estos "thresholds" podrán ser modificados en función de las condiciones particulares de cada estudio. Una alternativa semi-automatizada para realizar la imputación, normalización, transformación y filtrado de nuestras tablas de datos se encuentra disponible en la colección de funciones pReOmics (disponible en http://www.valledor.info)

3.5. Conociendo nuestros datos

Medidas de Dispersión

Las medidas de dispersión se emplean para mostrar la variabilidad de una distribución o grupo de muestras, indicando por medio de un número o de una representación gráfica si las variables están próximas o alejadas a la media. Cuanto mayor sea el valor de la dispersión, mayor será la variabilidad de la muestra, y cuanto menor, mayor homogeneidad muestral. El cálculo de la dispersión nos permitirá saber si todos los casos bajo estudio son parecidos, o varían entre ellos. En último término podríamos definir casos o geles que presenten valores atípicos para eliminarlos del estudio.

Además de las medidas clásicas de dispersión tales como la desviación media y la varianza, en Proteómica es habitual estudiar la variación mediante el coeficiente de variación (o de Pearson) (CV%):

$$CV\% = \frac{\sigma}{\bar{x}} \times 100$$

Esta medida de dispersión es útil para comparar la variación entre los distintos spots, ya que no está influenciada por los cambios de escala. Para emplear este coeficiente es importante que todos los valores bajo estudio sean positivos. Los CV% medios en los estudios en Proteómica oscilan entre el 30-40% para los datos sin transformar, reduciéndose este coeficiente a valores inferiores al 15% cuando los datos son transformados adecuadamente.

Por ejemplo, calculando el CV% de los 7 *spots* incluidos en "tabla1.imp" se obtiene un valor de 7,55%. Si previamente se transforma la tabla aplicando el log (previamente se multiplicaron todos los valores por 10 para evitar valores negativos tras la aplicación del logaritmo) este coeficiente se reduce a 3,42%.

- **Ejercicio 2:**
  Se propone el cálculo del coeficiente de variación promedio de los 7 spots tras aplicar las transformaciones raíz cúbica y seno.

  RESULTADO: Consultar apéndice 2.

Mediante una representación gráfica, el diagrama de caja, es posible visualizar nuestro conjunto de datos. El gráfico está basado en cuartiles y compuesto por un rectángulo llamado caja (cuyo valor inferior es el cuartil 1, Q1, y el superior el 3, Q3) y dos brazos, los bigotes, que se extienden hasta el último valor no atípico. La línea media a la caja representa el valor de la mediana. Los valores atípicos los podremos definir como aquellos inferiores a Q1-1,5*IQR o superiores a Q3+1,5*IQR. Siendo el rango inter-cuartil (IQR) Q3-Q1. Todos los valores que queden fuera podrán considerarse como atípicos, y su aparición deberá ser examinada minuciosamente para tratar de determinar si esta aparición se debe a algún error técnico, de manipulación (suelen aparecer cuando se introduce una muestra proveniente de otra población) o resultado de la aplicación de un tratamiento en particular.

A modo de ejemplo generaremos un diagrama de caja empleando R. Antes de continuar deberemos en primer lugar transponer la tabla, puesto que todos los test estadísticos que realizaremos a lo largo de este capítulo consideran que las diferentes variables están dispuestas en las columnas, mientras que los individuos/replicas ocupan las filas:

```
> tabla1.imp <- t(tabla1.imp)
```

Para trazar el diagrama en caja sencillo solo tenemos que emplear el comando boxplot, e indicar como único argumento el nombre de la tabla a representar.

```
> boxplot(tabla1.imp)
```

El resultado obtenido se muestra en la figura 2. Como puede observarse la muestra analizada no contiene valores atípicos. Como ejercicio para el lector se propone crear un nuevo objeto (`>tabla2 <- edit(tabla1.imp)`), que incluya el cambio de dos valores

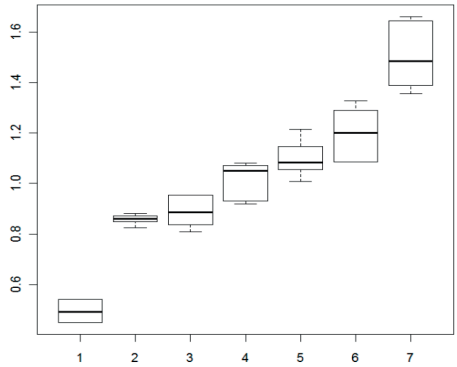


Figura 2: Diagrama de cajas de los datos "tabla1.imp". 1-7 indican las 7 variables analizadas.

correspondientes a dos variables cualquiera dentro de tabla1, por ejemplo 10 y 15 y re-trazar el gráfico. Nótese la aparición de los puntos correspondientes a los valores atípicos.

Normalidad

Para comprobar si los datos se ajustan a una distribución gaussiana recomendamos los test de normalidad de Lilliefors o de Shapiro-Wilk. Ambos comprueban la hipótesis nula que los datos provienen de una distribución normal. Deberemos estudiar todas las variables de manera independiente.

Ejemplo

Se empleará el objeto tabla1.imp, previamente transpuesto, para ejemplificar el uso del test de Shapiro-Wilk en R. Para realizar esta prueba no hay que instalar ni cargar ningún nuevo paquete, puesto que esta función, shapiro.test(), está incluida en el entorno base. Se programará además un bucle que permita analizar sucesivamente todas las variables:

```
> pvalores <- c()     #creamos el objeto vacío pvalores
> i=1     #creamos el contador i, con valor inicial 1.
> while(i <=ncol(tabla1.imp)){
                      #abrimos el bucle, que continuara
       #mientras i sea menor o igual al número de columnas
                      #de nuestra tabla
+ a<-shapiro.test(tabla1.imp [,i])
                      #creamos el objeto intermedio a,
          #que contiene el resultado del shapiro.test para la
                      #columna i de la tabla imputadot.
+ pvalores <- c(pvalores,a$p.value)
                      #añadimos al objeto pvalores el
          #valor de a$p.value, que contiene el p-valor del
          #test de Shapiro para la correspondiente columna.
+ i=i+1               #aumentamos en 1 unidad el contador
+ }                   #cerramos el bucle.
```

Los p-valores resultantes del test los encontraremos de forma ordenada dentro del objeto pvalores:

```
 > pvalores

[1] 0.20659041 0.39712347 0.43222572 0.05066252 0.78440219 0.29046334 0.16577492
```

En aquellos casos en los que el p-valor se encuentre por encima del nivel de significación se podrá aceptar la hipótesis nula y considerar que la distribución de dicha variable se ajusta a una distribución normal.

Homocedasticidad:

El último criterio que nuestras muestras y tratamientos deben cumplir para poder emplear test paramétricos es la homocedasticidad u homogeneidad de varianzas. Para comprobar esta hipótesis se recomienda el uso del test de Levene o de Bartlett. La hipótesis nula de ambos test considera la homogeneidad de varianzas. El nivel de significación empleado, tanto para probar la homocedasticidad como la normalidad, suele establecerse en 0,05.

- **Ejercicio 3:**
  Programar un bucle que permita el cálculo de los p-valores resultantes de la aplicación del test de Bartlett para la tabla1.imp. Se considerarán dos tratamientos con tres réplicas (Tratamiento 1: Geles1-3; Tratamiento 2: Geles 4-6), creándose una nueva columna

para definir los distintos tratamientos. La ayuda en línea disponible para cada uno de los comandos de R contiene una descripción de cada comando y sus argumentos, así como numerosos ejemplos de sintaxis. El lector debería empezar el ejercicio llamando a la ayuda para el test de Bartlett:

```
> help(bartlett.test)
```

RESULTADO: Ver apéndice 2.

4. Análisis univariable.

Dentro del análisis univariable, o conducido por una hipótesis, agrupamos a aquellos test estadísticos cuyo objeto es comparar las distintas variables, o muestras, mediante el contraste de una hipótesis nula previamente definida. La hipótesis nula se presume verdadera hasta que una evidencia estadística indique lo contrario. Dentro de estos test se agrupan la prueba t, los ANOVA y sus correspondientes test no paramétricas (prueba U, Kruskal-Wallis).

Estas pruebas se emplean, de forma general, para estudiar las posibles diferencias existentes entre los distintos tratamientos debido a su facilidad, sensibilidad y robustez. Pero, como contrapartida, no permiten el análisis conjunto de variables, agruparlas o diferenciarlas así como predecir tendencias ocultas (Esta aproximación será tratada en el siguiente apartado).

Su utilidad es indiscutible para establecer aquellos *spots* que presentan una acumulación diferencial, por ejemplo entre nuestro individuo control y el tratado. No obstante, no debe olvidarse que, en función del diseño experimental (número de factores, bloques, etc.) y de las características de los datos obtenidos, variará el tipo y la forma de test a emplear. A continuación se describen las situaciones experimentales más comunes en los ensayos de Proteómica, así como alguno de los test recomendados para cada caso.

4.1. Comparando dos tratamientos

Esta es la situación más simple, en la que solamente se compararán dos muestras (por ejemplo control frente a tratamiento, mutante frente a silvestre) y no se tiene en cuenta ningún diseño por bloques. La "tarea" se reduce a encontrar aquellos *spots* que muestran una abundancia diferencial entre las muestras. La prueba más popular es la t de Student, o su variante no paramétrica descrita por Welch que no presupone la homogeneidad de varianzas. La hipótesis nula considera que las medias en los niveles de acumulación de los *spots* son las mismas. El rechazo de la hipótesis nula supondrá que el nivel del *spot* es diferencial entre tratamientos. En caso de no disponer de muestras que se ajusten a una distribución normal, la alternativa no paramétrica es la prueba U de Mann-Whitney-Wilcoxon.

En la actualidad el empleo de alternativas no paramétricas está ampliamente extendido. Estudios recientes desarrollados por Fay y Poschan (2009) demuestran que el empleo del test Mann-Whitney-Wilcoxon es más atractivo que el empleo de la prueba t cuando se comparan muestras con más de 7 individuos/replicas por grupo puesto que este test es menos sensible a los valores atípicos y presenta una sensibilidad similar. Esta robustez radica en el empleo de la mediana para su cálculo, y no de la media como en el caso de la prueba t.

Algunos autores además del criterio estadístico emplean un criterio basado en el cambio de abundancia de cada *spot* para determinar una expresión diferencial. Puesto que pequeños

cambios de expresión, pese a ser matemáticamente significativos, no tendrían por qué serlo biológicamente. Desde aquí consideramos particularmente peligrosa esta práctica, sobre todo cuando se realizan análisis ciegos. Ya que por ejemplo cambios mínimos en el nivel de expresión de un factor de transcripción pueden ser biológicamente determinantes, mientras que duplicar el nivel de expresión de ciertas proteínas constitutivas puede no tener efecto biológico alguno. En caso de conocer perfectamente las proteínas estudiadas, y su función, este criterio podría emplearse en una segunda etapa de análisis para retener o descartar aquellos *spots* con una variación estadísticamente significativa.

Esta sección empleará la tabla 2 para ejemplificar el uso de los distintos test. Esta tabla simula los resultados de un hipotético experimento en el que se desea estudiar el efecto del frío (cultivo a 4°C) sobre tres variedades de la misma especie vegetal. Los valores presentes en esta tabla están adecuadamente pre-procesados.

Para poder realizar estos test en R la tabla de datos deberá tener una estructura ordenada, tal y como se describe a continuación:

Casos	Fact[1]	Fact[2]	...	Bloque	Var[1]	Var[2]	Var[3]	...
1								
2								
...								

Los distintos casos están en las filas, y las columnas corresponden a los distintos factores a considerar y a las distintas variables. En nuestro caso el primer factor es el tratamiento (Trat) con los niveles: control (c) y tratado (t); y el segundo las variedades (Variedad) con los niveles: siempreverde, anaranjada y morada. En caso de realizar el experimento en tandas o bloques, deberá indicarse esa información en la tabla.

Supongamos que quisiéramos comparar la respuesta de esta especie al tratamiento con frío, para simplificar no se considerará, de momento, la presencia de las distintas variedades. Antes de comenzar se podría visualizar los distintos datos, por ejemplo los *spots* 3, 4, 8 y 9) empleando diagramas de caja (Figura 3):

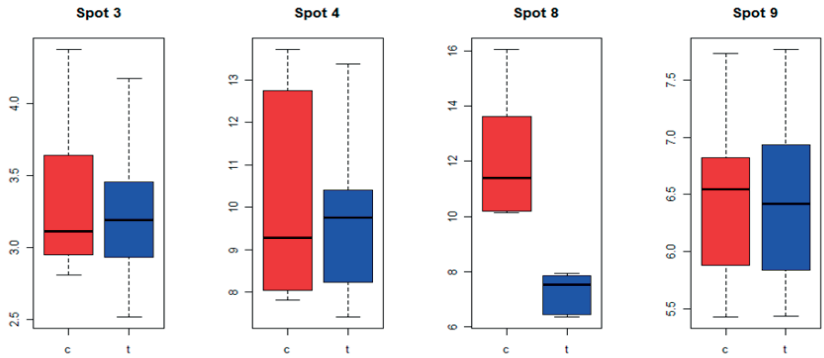


Figura 3: Diagrama de cajas para la respuesta a frío de esta especie en relación a los spots 3, 4, 8, 9. En rojo se muestra el control y en azul los valores para los spots después del tratamiento con frío.

```
> tabla2 <- dget("http://www.valledor.info/wp-content/uploads/2013/12/
tabla2.txt")
                    # Se cargan los datos, en primer lugar
```

```
> par(mfrow=c(1,4))
                  # Se divide el espacio gráfico en 1 fila y 4
                  # columnas
> boxplot(tabla2[,'Spot3']~Trat,data=tabla2, main='Spot 3',
+ col=c("red","blue"))
             # Se traza el gráfico correspondiente al Spot3,
                  # con los datos agrupados por tratamiento
                  # (Trat), coloreando los mismos en rojo y azul
                  # (col=c("red","blue") y con título principal
                  # (main) "Spot 3"
> boxplot(tabla2[,'Spot4']~Trat,data=tabla2, main='Spot 4',
+ col=c("red","blue"))
             #Se repite el mismo procedimiento para Spot 4
> boxplot(tabla2[,'Spot8']~Trat,data=tabla2, main='Spot 8',
+ col=c("red","blue"))
             #Se repite el mismo procedimiento para Spot 8
> boxplot(tabla2[,'Spot9']~Trat,data=tabla2, main='Spot 9',
+ col=c("red","blue"))
             #Se repite el mismo procedimiento para Spot 9
```

El resultado se muestra a continuación. En este caso podemos ver como todos los valores de intensidad presentan una distribución similar entre tratamientos, salvo para el *spot* 8 que muestra valores superiores en el grupo control (color rojo).

La representación gráfica es útil, pero en la mayoría de las ocasiones el investigador necesitará además el resultado del estadístico correspondiente. En R la realización de una prueba t es sencilla, no necesitando cargar ningún paquete adicional. Pongamos como ejemplo la descripción del código necesario para efectuar una prueba t sobre el *Spot* 4:

```
> res.pruebat=t.test(tabla2[,"Spot4"]~Trat,data=tabla2)
       # se crea un
       # objeto, res.prueba, que contiene el resultado
       # de aplicar una prueba t, a la tabla2,
       # seleccionando el Spot4 [,"Spot4"] y
       # diferenciando tratamientos
```

El lector debe percibir que, tal y como se empleó en el ejemplo anterior para trazar los diagramas de cajas, tabla2[,"Spot4"] indica al programa la selección de la columna llamada Spot4 dentro del objeto tabla2. En nuestro ejemplo sería equivalente indicar tabla2[,6], puesto que Spot4 es la sexta columna de la tabla. Si quisiera seleccionarse una fila esta se indicaría por el valor a la izquierda de la coma dentro del corchete. Un valor en blanco indica que se aplica a toda la tabla. El resultado de la prueba t, y sus tablas, están guardadas en res.pruebat:

```
> res.pruebat

        Welch Two Sample t-test

data:  tabla2[, "Spot4"] by Trat
t = 0.25, df = 15.689, p-value = 0.8058
alternative hypothesis: true difference in means is not equal to 0 95
percent confidence interval:  -2.007765  2.543615
sample estimates: mean in group c mean in group t
      10.179493        9.911568
```

En este caso apreciamos que se trata de una prueba de Welch, que considera que no existe homogeneidad de varianzas. Esta es la opción por defecto en R. Si se analiza el p-valor puede comprobarse que es superior al nivel de significación por lo que no se tiene suficiente evidencia para rechazar la hipótesis nula.

De entre todos los datos generados al realizar una prueba estadística en R al investigador, como norma general, le interesará aislar el resultado del p-valor, por ejemplo en un bucle para analizar todas las variables de la muestra, para ello hay que seleccionar el componente "p.value":

```
> res.pruebat$p.value   # Se selecciona el p.valor dentro de
[1] 0.8058416           # res.pruebat.
```

Acudiendo a la ayuda disponible podremos obtener una información completa, con todos los componentes y argumentos de cualquier comando. De manera análoga a cuando se consultó el test de Bartlett, para la prueba t el comando sería necesario teclear:

```
> help(t.test)
```

El *Spot* 8 se estudiará empleando la prueba t clásica, para ello se debe añadir el argumento "var.equal=T":

```
> res.pruebat=t.test(tabla2[,"Spot8"]~Trat,data=tabla2, var.equal=T)
> res.pruebat

        Two Sample t-test

data:  tabla2[, "Spot8"] by Trat
t = 6.478, df = 16, p-value = 7.623e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:   3.398754 6.705274
sample estimates:  mean in group c mean in group t
      12.322838        7.270823
```

En este caso puede apreciarse como el p-valor se encuentra dentro de la región crítica, rechazando la hipótesis nula y considerando diferentes los valores del control y el tratamiento.

Como se comentó anteriormente en algunos casos deberemos recurrir al empleo de test no paramétricos. La prueba U de Mann-Whitney-Wilcoxon, la alternativa no paramétrica a la prueba t, se ejecuta en R de manera similar, solo variando el nombre del comando:

```
> res.pruebau=wilcox.test(tabla2[,"Spot8"]~Trat,data=tabla2)
> res.pruebau

        Wilcoxon rank sum test

data:  tabla2[, "Spot8"] by Trat
W = 81, p-value = 4.114e-05
alternative hypothesis: true location shift is not equal to 0
```

- **Ejercicio 4:**
  Programar dos bucles que permitan la aplicación de una prueba t y una prueba U para contrastar los posibles efectos causados por el tratamiento con frío a la especie estudia-

da en la tabla2. No se considerarán las diferencias entre variedades. ¿Existe diferencias entre ambos grupos de p-valores? En caso de existir, ¿podría darse una explicación?

RESULTADO: Ver apéndice 2.

4.2. Comparando más de más de dos tratamientos

La comparación de más de dos tratamientos, cuando las muestras son homocedásticas y su distribución se ajusta a una normal, se realizan de forma clásica empleando un conjunto de técnicas denominadas ANOVA (Daniel, 2009). En resumen, estos test pueden considerarse como una extensión de la prueba t para más de dos tratamientos, en la cual se comparara la varianza entre los distintos tratamientos (j) para establecer si las diferencias observadas pueden ser atribuidas a variaciones experimentales relacionadas con el tratamiento aplicado, en lugar que por la variación biológica y/o técnica (ε). El valor para un *spot* Xij se puede modelar de la siguiente forma:

$$X_{ij} = \mu_i + \alpha_j + \varepsilon$$

Siendo μ la intensidad media del *spot* i, α el efecto del tratamiento j sobre el *spot* y ε el error aleatorio.

Intuitivamente se puede advertir que si el tratamiento aplicado tuviera algún efecto sobre la intensidad del *spot* (Xij), este debería ser causado por un αj diferencial. Si las muestras no presentaran una distribución normal o fueran homocedásticas debe emplearse el test de Kruskal-Wallis, la extensión de la prueba Mann-Whitney-Wilcoxon. Estas aproximaciones poseen las mismas ventajas y desventajas previamente descritas para la pruebas t y U. Esta breve introducción al ANOVA debería ampliarse con lecturas más específicas (ver Chich et al., 2007; Daniel, 2009).

Estudiando un factor sobre más de dos tratamientos:

Se denomina modelo factorial con un factor o ANOVA con un factor al modelo (lineal) en el que la variable analizada depende de un sólo factor. Por consiguiente las causas de su variabilidad son debidas al efecto exclusivo de ese único factor sin olvidar el debido al error experimental (error aleatorio). Este modelo se aplicaría cuando solo exista una fuente de variabilidad no aleatoria, por ejemplo el efecto de distintos tratamientos aplicados a una población homogénea, o la resistencia-susceptibilidad de distintos tipos de variedades bajo una situación experimental determinada.

En el apartado anterior se estudió si el tratamiento de frío tuvo efecto sobre los *spots* mediante una prueba t. Por el contrario, en este momento deseamos comprobar si, bajo una temperatura no alterada las distintas variedades poseen el mismo perfil proteómico, es decir, si los *spots* presentan la misma intensidad en las variedades anaranjada, morada y siempreverde. Este planteamiento es muy común en los experimentos en Proteómica, en los cuales se desean comparar 3 o más tratamientos distintos (drogas, especies, localizaciones, etc.) en busca de *spots* diferenciales.

Para este ejemplo se creará en primer lugar nueva tabla, tabla3, que incluya solamente los casos control de la tabla2:

```
> tabla3 <- subset(tabla2, Trat=="c")
```

A modo de ejemplo se estudiará el *Spot* 5, comenzando con la su distribución entre las distintas variedades (Figura 4). Nótese que en este caso se han añadido nuevos argumentos al co-

mando boxplot para dotar de color a las cajas. Además incluimos el código para generar una figura similar en ggplot2, una librería que genera unos gráficos más estéticos (Figura 5).

```
> boxplot(tabla3[,"Spot5"]~Variedad,data=tabla3,
+ col=c("orange","purple","green"))

> ggplot(tabla3, aes(Variedad,Spot5, fill=Variedad))+geom_boxplot()+ geom_jitter(position=position_jitter(w=0.1, h=0.1), size=3)+scale_fill_manual(values=c("orange", "purple", "green"))
```

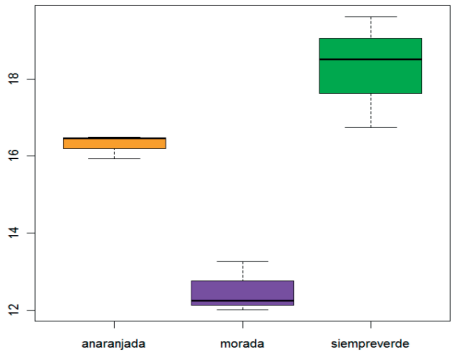


Figura 4: Diagrama de cajas de valores *spot 5*. En distintos colores se muestran los valores del *spot 5* para las 3 variedades analizadas.

Aparentemente las distintas variedades presentan intensidades diferentes para el *Spot* 5. Sin embargo, puesto que en algunas ocasiones la interpretación de la representación gráfica puede ser equívoca, se ejecutará un análisis de la varianza empleando el comando aov(), creando en nuestro ejemplo el objeto res.anova. Para ver la tabla del análisis de la varianza sobre el objeto aov, emplearemos el comando anova():

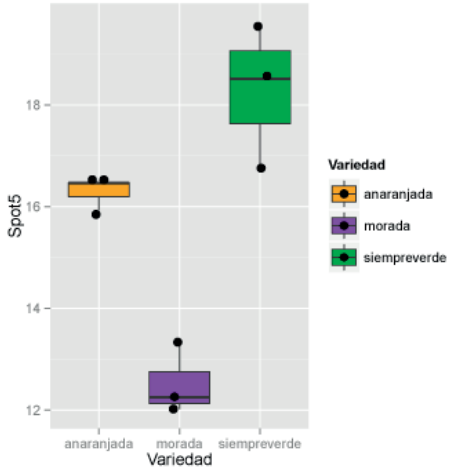


Figura 5: Diagrama de cajas de valores *spot 5* bajo el paquete ggplot2. En distintos colores se muestran los valores del *spot 5* para las 3 variedades analizadas.

```
> res.anova=aov(tabla3[,"Spot5"]~Variedad,data=tabla3)
> anova(res.anova)
Analysis of Variance Table

Response: tabla3[, "Spot5"]
          Df Sum Sq Mean Sq F value
Pr(>F)
Variedad   2 51.770 25.8850  29.598 0.0007795 ***
Residuals  6  5.247  0.8746
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para recuperar únicamente el p-valor tendríamos que seleccionarlo del siguiente modo:

```
> anova(res.anova)$'Pr(>F)'[1]
[1] 0.000779487
```

El p-valor resultado del ANOVA para el *Spot* 5 es inferior al nivel de significación, por lo tanto se rechaza la hipótesis nula y se demuestra estadísticamente que existen diferencias entre variedades para el *Spot* 5. Una vez determinada la existencia de diferencias entre variedades inmediatamente surge la siguiente pregunta, ¿Qué variedades difieren entre sí? Para responder esta pregunta es necesario emplear un test post hoc. Estos test se ejecutan tras el ANOVA y nos permitirán definir las posibles diferencias entre grupos (las variedades de nuestro caso). Uno de los test más empleados es el de Tukey. En R se realiza mediante el comando TukeyHSD(), necesitando un objeto del tipo aov para poder ejecutarse:

```
> TukeyHSD(res.anova)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = tabla3[, "Spot5"] ~ Variedad, data = tabla3)

$Variedad
                             diff        lwr       upr     p adj
morada-anaranjada       -3.783655 -6.1265084 -1.440801 0.0061420
siempreverde-anaranjada  2.000217 -0.3426366  4.343070 0.0876837
siempreverde-morada      5.783872  3.4410183  8.126725 0.0006731
```

Como puede observarse existen diferencias significativas entre las variedades morada-anaranjada (p-valor ajustado de 0,00614) y morada-siempreverde (p-valor 0,00067). Sin embargo no existe suficiente evidencia para considerar diferencias entre las variedades siempreverde y anaranjada para el *Spot* 5.

El empleo de un ANOVA seguido de un test de Tukey (en los casos en que el ANOVA sea significativo) permite un adecuado y rápido análisis de nuestras muestras. Al igual que en los test anteriores se puede programar un bucle para analizar automáticamente todas las variables. Una vez realizado el test recomendamos seleccionar aquellas variables cuyo p-valor sea significativo y realizar individualmente cada test post hoc.

```
> pvalores <- c()
        #creamos el objeto vacio pvalores
> i=3
        #contador i con valor 3, primera columna con datos
> while(i<=ncol(tabla3)){
        #se abre el bucle
+ res.anova=aov(tabla3[,i]~Variedad,data=tabla3)
        #se crea
        #res.anova, conteniendo el análisis de la varianza
+ pvalores <- c(pvalores,anova(res.anova)$'Pr(>F)'[1])
        #añadimos al
        #objeto pvalores el pvalor correspondiente.
+ i=i+1
        #aumentamos en 1 unidad el contador
+ }
        #cerramos el bucle.

> pvalores
[1] 1.225136e-03 2.161879e-05 4.462270e-01 6.773244e-06 7.794870e-04
[6] 1.764243e-04 9.203035e-07 9.791763e-01 7.641075e-01 4.563216e-02
```

- **Ejercicio 5:**
  Realizar un de Tukey para aquellos *spots* que presenten unas diferencias significativas entre variedades. Además se recomienda la programación de un bucle para realizar el test equivalente no paramétrico, el de Kruskal Wallis, sobre la tabla3 empleando el comando kruskal.test().

  RESULTADO: Ver apéndice 2.

Estudiando más de un factor y diseño por bloques

Cuando existe más de un factor no aleatorio causante de variabilidad se deberá emplear el análisis ANOVA multifactorial, el cual considerará ambos factores y además introduce un nuevo concepto, la posible interacción entre ellos. Desde un punto de vista no formal, puede decirse que existe interacción entre dos factores cuando el efecto de uno de ellos sobre la variable no es el mismo en todos los niveles del otro factor. Esto equivale a afirmar que existe interacción cuando el resultado de la combinación de dos factores difiere de la suma de los efectos principales de esos factores (Pardo et al., 2007).

Si durante la planificación y ejecución de un experimento se han tenido en cuenta más de un factor, tratamiento y variedad en nuestro ejemplo, el modelo de ANOVA ha de ser adaptado, para considerar las posibles diferencias debidas a cada uno de los factores así como y a una posible interacción entre los estos. Del mismo modo, si el experimento se diseña con diferentes bloques experimentales el modelo general del ANOVA debe ser adaptado (Kerr et al., 2001; Brien et al., 2006; Daniel, 2009:

$$X_{ikj} = \mu_i + \beta_k + \alpha_j + \gamma + \varepsilon$$

Siendo μ la intensidad media del *spot* i, βk el efecto del bloque k (variación técnica), α el efecto del tratamiento j sobre el *spot*, γ la interacción bloque-tratamiento y ε el error aleatorio. Al igual que en el resto de ANOVAs, la hipótesis nula supone que los distintos tratamientos no poseen ningún efecto sobre la media.

La mayoría de software de análisis de imagen, y todos los paquetes estadísticos, incluyen la opción de análisis multifactoriales y diseños en bloque pero, una vez más, nosotros recomendamos la realización manual de los análisis estadísticos para procesar adecuadamente los datos a la vez que se controla el error estadístico.

En la tabla2 se incluyen los resultados de un experimento en el cual se ha estudiado la respuesta de 10 *spots* al tratamiento con frío en tres variedades de la misma especie vegetal. Hasta este momento hemos obviado el experimento en su conjunto, pero ahora vamos a analizar todos los datos disponibles empleando un ANOVA bifactorial. Además, vamos a suponer que como nuestro laboratorio es pequeño para poder procesar todas las muestras fue necesario el empleo de 3 tandas (bloques) de 6 geles cada una, por lo que el efecto de las distintas tandas también deberá ser considerado.

En primer lugar crearemos una columna nueva en la tabla2, denominada Tanda, en la cual se incluirá la información correspondiente a que geles formaban parte de cada uno de los bloques. Este diseño consideró aleatoriedad durante todas las etapas analíticas. Obviamente en un experimento real esta columna deberá formar parte de la tabla original, puesto que dispondremos de dicha información antes de comenzar los experimentos.

```
> tabla2$Tanda <- c(1,2,3)
```

El comando para realizar este tipo de ANOVA es el aov(), ampliamente descrito a lo largo de este capítulo. Simplemente se añade el siguiente factor a considerar (los dos factores se separan por un *) y también se indica la columna en la que se encuentra la información de bloques tras un +:

```
> res.anova=aov(tabla2[,'Spot4']~Trat*Variedad+Tanda,data=tabla2)
> anova(res.anova)
Analysis of Variance Table

Response: tabla4[, "Spot4"]
```

```
              Df Sum Sq Mean Sq F value    Pr(>F)
Trat           1  0.323   0.323  0.5402    0.4777
Variedad       2 72.464  36.232 60.5899 1.151e-06 ***
Tanda          1  1.434   1.434  2.3979    0.1498
Trat:Variedad  2  2.228   1.114  1.8631    0.2010
Residuals     11  6.578   0.598
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(res.anova)$'Pr(>F)'[1:4]
[1] 4.777310e-01 1.151470e-06 1.497732e-01 2.009904e-01
```

En este caso no obtenemos un solo p-valor, sino 4. En el ejemplo puede verse como solo existen diferencias significativas entre variedades (Variedad), pero no entre tratamientos (Trat). No existe un efecto de bloque, puesto que el p-valor para Tanda es superior al nivel de significación, tampoco existe una interacción entre tratamiento y variedad. Si se deseara seguir profundizando en el estudio de las diferencias particulares, podría aplicarse el test de Tukey tal y como se aplicó con anterioridad en este capítulo.

- **Ejercicio 6:**
  Realizar un análisis estadístico completo de la tabla 2, teniendo en cuenta el diseño por bloques. En primer lugar se representarán, en un mismo gráfico, los resultados obtenidos para los *spots* 3, 4, 8 y 9 mediante un diagrama de caja, el cual presente dos colores diferenciales para cada tratamiento. A continuación se programará un bucle que realice automáticamente el ANOVA bifactorial con efecto de bloque a todos los *spots* de la muestra y guarde los 4 p-valores en un nuevo objeto. Finalmente se creará una nueva tabla que presente 10 filas, correspondientes a cada variable, y 4 columnas correspondientes a los distintos p-valores y se copiarán los datos al portapapeles.

  RESULTADO: Ver apéndice 2.

Controlando el error estadístico

Hasta este momento se han descrito una serie de test estadísticos encaminados a comprobar la probabilidad de hipótesis nula. Como norma general consideramos un nivel de significación de 0,05. Este valor presupone una especificidad para el test del 95% o, es decir, tenemos una probabilidad del 5% de rechazar la hipótesis nula siendo esta verdadera (debido al azar). En otras palabras, si realizamos un test 100 veces, en 5 de ellas nos encontraríamos ante un error de tipo I. La probabilidad de obtener error del tipo I es pequeña cuando se analiza un pequeño número de variables, pero cuando se desean comparar todos los *spots* de un gel esta probabilidad se dispara debido al elevado número de test a realizar. Por ejemplo para un análisis proteómico estándar con 500 spots diferentes esperaríamos encontrar 25 falsos positivos (500 x 0,05). Teniendo en cuenta que la mayoría de estudios proteómicos están enfocados a la determinación de spots diferenciales, es conveniente emplear algún sistema que permita controlar la aparición de este tipo de error.

El error de tipo I puede controlarse mediante el empleo de distintas reglas de decisión. Podríamos decir que en la actualidad existen dos aproximaciones principales. La primera de ellas, descrita por Holm-Bonferroni (Strassburger y Bretz, 2008), está basada en controlar la tasa de error por familia (Family Wise Error Rate, FWER). Su base es sencilla, ajustar el nivel de significación inicial (α) en función del número de variables a testar (m) según la fórmula $\alpha'=\alpha/m$. Para nuestro ejemplo de 500 *spots* el nivel de significación corregido sería 0,0001. Como puede verse esta regla es muy estricta, reduciendo al máximo el error de tipo I, pero a costa de elevar

el error de tipo II a niveles inaceptables en la mayoría de las situaciones. Solo se recomienda cuando se estén estudiando un pequeño número de variables o se desee establecer de manera robusta algún tipo de biomarcador. Por el contrario este procedimiento debe evitarse si el propósito del estudio es comprobar la igualdad (no diferencia) entre muestras.

La tasa de descubrimientos falsos (False Discovery Rate, FDR), descrita por Benjamini y Hochberg (1995), es una aproximación que corrige α para múltiples comparaciones controlando la proporción esperada de error de tipo I (por ejemplo 5%) mediante el análisis de la distribución de los p-valores (Storey, 2002) obtenidos tras aplicar el test estadístico, estimando a partir de ella una nueva distribución ajustada (q-valores). Los q-valores podrían considerarse como los análogos a los p-valores tras el ajuste FDR. Desde aquí recomendamos evitar tomar un valor de corte arbitrario, como pudiera ser 0,05, puesto que la interpretación del q-valor es sencilla y mucho más informativa. Para interpretar el q-valor en primer lugar se deben ordenar los valores de menor a mayor con el objeto de saber la posición relativa que ocupan en la lista. Supongamos que un q-valor de 0,0125 ocupa el puesto 75 de la lista, esto significa que podemos esperar que el 1,25% de todos los *spots* que poseen q-valor inferior a 0,0125 sean falsos positivos. En nuestro caso serían 75 x 0,0125 = 0,93 falsos positivos, es decir, menos de uno. Jugando con esta relación podremos estimar el número de falsos positivos que esperamos en nuestro análisis, pudiendo adaptar la regla de decisión (q-valor de corte) en función de las características particulares de nuestro experimento, gracias a la mayor información que nos proporcionan estos valores.

- **Ejercicio 7:**
  Calcular los q-valores resultantes de aplicar FDR empleando el paquete fdrtool (debe ser instalado y cargado) sobre los p-valores obtenidos en el ejercicio 5 y sobre la lista de p-valores incluida como ejemplo dentro del paquete fdrtool.

  RESULTADO: Ver apéndice 2.

5. Análisis Multivariable.

Los métodos de análisis multivariable están orientados a reducir la complejidad de la muestra mediante el establecimiento de los distintos patrones subyacentes en los datos. Estos métodos son útiles para determinar con exactitud los *spots* que son relevantes en el nuestro ensayo, empleando para ello no solo las diferencias individuales entre *spots*, sino también la relación de covarianza entre todos los *spots* analizados (Nedenskov et al., 2008). Los *spots* seleccionados mediante esta aproximación, que generalmente se organizarán en torno a varios grupos diferenciados, pueden emplearse para definir, por ejemplo, los distintos procesos biológicos que se encuentran tras el cambio de expresión de un gran número de proteínas (Martens et al., 2001). Estos métodos pueden agruparse en tres categorías diferentes: estadística multivariable, agrupamiento (clustering) y aquellos basados en aprendizaje automático, una rama de la inteligencia artificial.

5.1. Estadística multivariable

Los métodos de estadística multivariable como por ejemplo el análisis de componentes principales (PCA), el análisis de componentes independientes (ICA), la regresión mediante mínimos cuadrados parciales (PLS), o el análisis discriminante (DA) ofrecen una aproximación robusta para obtener una visión general de la variación y de las posibles interacciones ente los distintos *spots* (Nedenskov et al., 2008). Estos métodos también son utilizados para la selección de *spots* significativos, así como para la clasificación de las muestras en diferentes grupos desde una perspectiva diferente a la estadística univariable (Schultz et al., 2004).

Análisis de componentes principales (PCA) y análisis de componentes independientes (ICA)
El propósito general de estos métodos es reducir la complejidad inicial de la muestra, generalmente compuesta por varios cientos de *spots*, a unos pocos componentes que recojan la mayor parte de la variación original de la muestra. De esta forma se podrán valorar las interacciones y relaciones existentes entre los distintos *spots*, así como simplificar los datos.

El PCA está basado en la transformación del total de variables observadas en un nuevo conjunto de variables independientes denominados componentes principales (PC). Los distintos PC no se correlacionan matemáticamente entre sí, por lo que cada componente recogería una fuente distinta de variación, que posteriormente podría asociarse a un proceso biológico subyacente. La importancia de cada componente podemos definirla basándonos tanto en el porcentaje de variación que recoge, como en su capacidad para diferenciar tratamientos entre sí. Una gran parte de la información proporcionada por un PCA se puede representar mediante gráficos, que permiten tanto clasificar las muestras como visualizar las distintas estructuras de datos y grupos de observaciones similares además de identificar valores atípicos (Nedenskov et al., 2008). Una vez realizado el PCA obtendremos una matriz de correlaciones (o matriz de carga) la cual nos muestra la correspondencia de cada *spot* con cada componente principal.

Los *spots* que tienen mayor carga en cada PC pueden ser empleados para explicar o "nombrar" la fuente de variación recogida por dicha PC y también para agrupar a las mismas según su sobre- o sub-acumulación en relación a cada componente. A continuación se empleará un ejemplo para ilustrar estos conceptos.

Utilizaremos la tabla 2. El comando para realizar el análisis de componentes principales es el prcomp(),incluido en la distribución base de R. Los argumentos scale y center de dicho comando permitirán escalar y centrar los valores de las distintas variables, permitiendo compararlas independientemente de las distintas magnitudes que puedan presentar, estandarizando de esta forma los análisis. En este caso seleccionaremos para realizar el análisis de componentes principales las columnas 3 a 12, puesto que son las que corresponden a los valore numéricos de los distintos *spots*.

```
> tabla2 <- dget("http://www.valledor.info/wp-content/uploads/2013/12/
tabla2.txt")
> row.names(tabla2)<-c("SV-C1","SV-C2","SV-C3","A-C1","A-C2",
+ "A-C3","M-C1","M-C2","M-C3","SV-T1","SV-T2","SV-T3","A-T1",
+ "A-T2","A-T3","M-T1","M-T2","M-T3")
                    #Nombramos los casos
> tabla2.pca <- prcomp(tabla2[,3:12], scale=TRUE, center=TRUE)
                    #Realizamos el PCA
> summary(tabla2.pca, scale=TRUE)
                    #Obtenemos el resumen de la informacion contenida
en el objeto tabla2.pca

Importance of components:
                          PC1      PC2      PC3       PC4     PC5     PC6
Standard deviation      1.9205   1.5928   1.2351    1.0138 0.76061 0.62272
Proportion of Variance  0.3689   0.2537   0.1525    0.1028 0.05785 0.03878
Cumulative Proportion   0.3689   0.6226   0.7751    0.8779 0.93572 0.97450
                          PC7      PC8      PC9      PC10
Standard deviation     0.43778  0.2121  0.13547 1.910e-10
Proportion of Variance 0.01917  0.0045  0.00184 0.000e+00
Cumulative Proportion  0.99367  0.9982  1.00000 1.000e+00
```

Este comando nos permite ver la relevancia de los distintos componentes, por ejemplo PC1, PC2 y PC3 recogen, respectivamente, un 36,89%, 25,37% y 15,25% de la variación original. Con los primeros 5 componentes tendríamos recogida prácticamente un 94% de la variación total de la muestra. En nuestro ejemplo, con solo 10 *spots*, los resultados obtenidos pueden no parecer muy poderosos, pero generalmente 5-9 componentes suelen ser suficientes para explicar el 99% de la variación en muestras proteómicas reales, con más de 500 *spots* cada una de ellas. Para saber si los distintos componentes principales nos permiten diferenciar las distintas muestras entre si podemos emplear una representación gráfica (Figura 6):

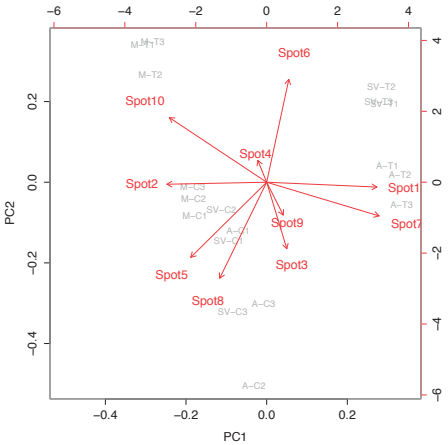


Figura 6: Representación gráfica de análisis de componentes independientes (ICA), componentes 1 y 2. Abreviaturas: A-C = Anaranjada-Control, A-T = Anaranjada-Tratamiento Frío, M-C = Morada-Control, M-T = Morada Tratamiento Frío, SV-C = Siempreverde-Control, SV-T = Siempreverde Tratamiento Frío.

```
> biplot(tabla2.pca,choices=c(1,2),col=c("gray","red"), cex=c(0.75,1))
#Obtenemos la representación gráfica de los componentes 1 y 2, definidos
en el vector choices, col y cex nos permiten definir los colores y el ta-
maño de la fuente respectivamente.
```

En el gráfico podemos ver la correlación de los distintos *spots* con los componentes principales 1 y 2, así como la distribución de los casos en función de su puntuación (score) para cada uno de los componentes. Podemos ver como los casos correspondientes a los individuos tratados se encuentran perfectamente diferenciados de los casos control. Además las distintas variedades se pueden diferenciar perfectamente. Se propone como ejercicio la representación del resto de componentes principales que se consideren significativos. Una primera aproximación a este problema, la selección de componentes significativos, puede realizarse en base a la cantidad de variación que recoge cada uno de ellos, por ejemplo aquellos que recogen más de un 1% de variación.

La siguiente cuestión que se plantea es cuales son aquellos *spots* más correlacionados con los distintos componentes principales. En nuestro ejemplo los PC1 y PC2 nos permiten diferenciar las muestras aceptablemente bien pero, ¿cuales son los *spots* cuya variación está recogida en cada uno de esos dos componentes? Para responder a esta pregunta se deberá consultar la matriz de cargas (loadings) la cual recoge la correlación existente entre cada *spot* y cada uno de los componentes principales:

```
> print(tabla2.pca$rotation)
```

	PC1	PC2	PC3	PC4	PC5	PC6
Spot1	0.47682620	-0.02524725	-0.07714321	-0.19797247	-0.248208369	-0.37803620
Spot2	-0.43144552	-0.01047750	0.34856256	-0.09160758	-0.363216367	-0.24435217
Spot3	0.08723687	-0.34699933	0.45153196	0.08728740	0.729636291	-0.28517135
Spot4	-0.03827876	0.11374918	-0.74370558	-0.01203074	0.351069915	-0.32337740
Spot5	-0.32848387	-0.39186856	-0.19626965	0.16878963	0.122291089	0.51309074
Spot6	0.09586423	0.53638234	0.23099080	0.22492598	0.189224490	-0.06492072
Spot7	0.48659313	-0.17583800	-0.03003558	-0.05944260	-0.008445108	0.31516610
Spot8	-0.20306586	-0.50083758	-0.09852069	-0.22738318	-0.104183307	-0.41577228

```
Spot9     0.07174754  -0.17114523  -0.08276131    0.90018078  -0.255166124  -0.24856885
Spot10   -0.42091337   0.33703179  -0.07897817    0.00273991   0.155083925  -0.10012720

              PC7            PC8            PC9            PC10
Spot1     0.18983879     0.38073803     0.5674188682   -0.13918341
Spot2     0.23034456     0.39209608    -0.3771392256   -0.37840601
Spot3     0.20265541     0.07005047     0.0110134413   -0.02629832
Spot4     0.07174710     0.25031259    -0.3506812365   -0.11725501
Spot5    -0.09065101     0.46641686     0.3524516413   -0.20814418
Spot6    -0.63598159     0.25239261     0.0638275464   -0.29476502
Spot7     0.06226954    -0.24067046    -0.2672366377   -0.70374454
Spot8    -0.58862647    -0.28819092     0.0836079728   -0.16208134
Spot9     0.07689322    -0.11766762     0.0007107358   -0.04161438
Spot10    0.30939357    -0.44459767     0.4538731905   -0.41173731
```

De la tabla anterior se deduce que los *spots* 1 y 7, y 2 y 10 son los que más correlacionan directa e inversamente con el PC1. Siendo el *Spot* 6 el que lo hace en mayor medida con el PC2. Por ejemplo si los *spots* 1 y 7 correspondieran a proteínas relacionadas con estrés térmico por bajas temperaturas podríamos denominar a este componente como "Resistencia al Frío", si bien el empleo de solo 2 *spots* podría considerarse especulativo para dicho propósito. En un estudio proteómico normal se estudian varios cientos de proteínas, y se suele correlacionar con cada componente de un modo intenso (>0.8) un buen número de *spots*, lo que otorga más poder a la definición funcional de cada componente. Como el lector habrá podido observar la pericia y conocimientos del investigador a la hora de relacionar los *spots* que muestran más correlación con cada uno de los componentes será definitorio a la hora de inferir el proceso biológico que se oculta tras ellos.

Hasta ahora no se ha descrito el procedimiento para graficar las distintas muestras en relación a cada componente principal. Para ello hay que conocer el score o puntuación de cada muestra con cada componente:

```
> print(tabla2.pca$x, center=TRUE)
```

Las puntuaciones obtenidas nos permitirán emplear estadística común para procesarlas. Mediante un análisis univariable buscaremos aquellos scores que sean diferenciales para los conjuntos de muestras correspondientes a cada una de las poblaciones. De esta forma, podremos definir de manera alternativa a la representación gráfica o a la variación recogida, aquellos componentes principales que nos permiten discriminar las distintas muestras.

En contraste con el PCA, el ICA descompone el conjunto de datos originales para establecer componentes independientes entre sí. Esta proyección parece captar la estructura esencial de los datos (Hyvärinen y Oja, 1997), por lo cual permite la extracción ciega de características así como separar fuentes de variación mixtas. Además el ICA también se puede emplear para reducir los efectos del ruido o los artefactos de la señal (Hyvärinen et al., 2001) ya que, por lo general, el ruido es generado por fuentes independientes. El ICA demostró ser más rápido y más robusto que el análisis de varianza cuando se trata de datos proteómicos, ya que reduce el ruido y no dependen del umbral de selección, mostrando además una mayor capacidad para distinguir entre grupos (Hyvärinen y Oja, 2000).

Empleando la tabla2 y el algoritmo fastICA bajo R al hallar 3 componentes independientes podemos observar como las muestras se separan perfectamente tanto empleando los componentes independientes 1 y 2 como 2 y 3 (Figura 7). La matriz de cargas (*mixing matrix*) nos indicará, de modo similar al anterior, la importancia de cada variable para cada componente independiente.

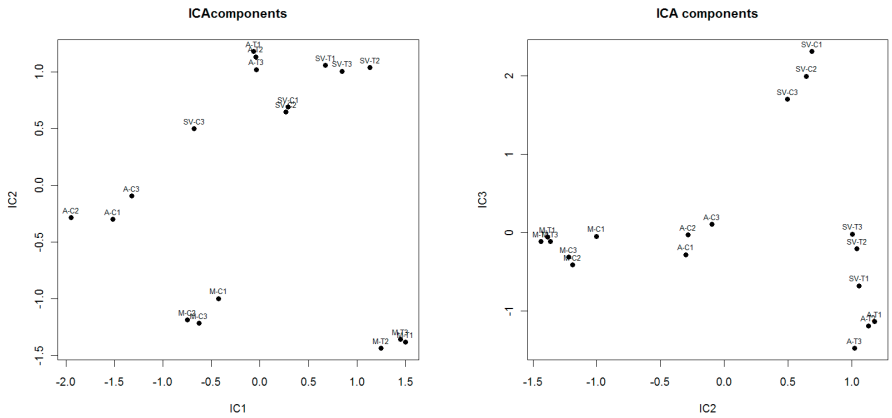


Figura 7: Representación gráfica de análisis de componentes principales (PCA. Componentes 1 y 2, componentes 2 y 3. Abreviaturas: A-C = Anaranjada-Control, A-T = Anaranjada-Tratamiento Frío, M-C = Morada-Control, M-T = Morada Tratamiento Frío, SV-C = Siempreverde-Control, SV-T = Siempreverde Tratamiento Frío.

- **Ejercicio 8:**
  Empleando el algoritmo fastICA, su ayuda on-line, y la tabla 2 generar los dos gráficos anteriores y seleccionar aquellos spots con más carga para cada uno de los componentes independientes extraídos. Instalar el paquete PLS y, tras leer la ayuda disponible, realizar un análisis PLS de la tabla2.

  RESULTADO: Ver apéndice 2.

5.2. Árboles de agrupamiento y heatmaps.

El agrupamiento jerárquico de las muestras y variables es una técnica exploratoria que puede ser empleada para definir conjuntos y estructuras en nuestros datos así como para detectar muestras atípicas (Cross et al, 1995). El agrupamiento trata de identificar *spots* que se comportan de forma similar en las distintas muestras, lo cual implica en cierta medida que están correlacionados. El empleo de mapas de color (heatmaps) nos permitirá visualizar de forma rápida las distintas variaciones en los niveles de acumulación de los *spots* en las distintas situaciones experimentales.

El cálculo de las distancias entre las distintas variables puede ser hallado mediante el empleo de numerosos algoritmos si bien Meunier et al. (2007) concluyeron que el coeficiente de correlación de Pearson o la distancia Manhattan y el método de agregación de Ward son la mejor estrategia de agrupamiento para datos proteómicos, seguidos muy de cerca por el empleo de la distancia Euclidea y la agregación de datos mediante UPGMA.

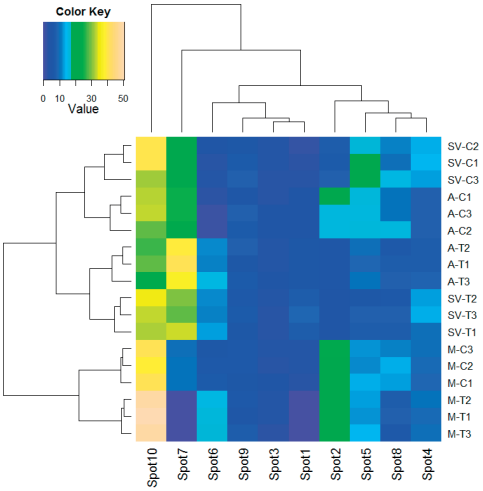


Figura 8: Heatmap. Abreviaturas: A-C = Anaranjada-Control, A-T = Anaranjada-Tratamiento Frío, M-C = Morada-Control, M-T = Morada Tratamiento Frío, SV-C = Siempreverde-Control, SV-T = Siempreverde Tratamiento Frío.

Desde aquí recomendamos el empleo del software Cluster 3.0 y JavaTreeView para la realización de los distintos clusters y

heatmaps debido a su potencia y sencillez, si bien el entorno R ofrece también paquetes específicos para ello como por ejemplo heatmap.2() que se encuentra dentro del paquete gplots, pheatmap(), o emplear el potente ggplot2. Este comando posee numerosos argumentos que nos permitirán decidir sobre todas las variables posibles a la hora de trazar el mapa, como por ejemplo el tipo de distancia a calcular (distfun), el método de agregación (htclust), la reordenación de filas y columnas (reorderfun), o los colores (col). Por defecto el Heatmap.2 emplea la distancia Euclídea y una agregación completa (Figure 8):

```
> tabla3 <- tabla2[,3:12]
      #Seleccionamos las columnas numéricas

> heatmap.2(tabla3, col=topo.colors(100), density.info="none",
trace="none", scale=c("none"), symm=FALSE, symkey=FALSE)
```

Se puede observar como los distintos *spots* agrupan entre si formando dos grupos diferenciados (6, 9, 3, 1 y 2, 5, 8, 4). La relación de proximidad entre los distintos tipos de muestras también puede verse en los distintos árboles formados. Por un lado la variedad morada se diferencia del resto, principalmente por la sobreacumulación del *spot* 2 y poca presencia del *spot* 7 comparado con el resto. El resto de variedades se muestran más próximas entre sí, difiriendo fundamentalmente las tratadas con frío de las variedades control.

A la hora de realizar los distintos árboles se recomienda emplear el set de datos completo, y no solo aquellos *spots* diferenciales, puesto que se pierde mucha información sobre la estructura original de los datos.

- **Ejercicio 9:**
  Se plantea como ejercicio la repetición de este árbol empleando la distancia Manhattan, el método de agregación de Ward y mostrando el valor de cada celda empleando pheatmap().

5.3. Introducción a las redes neuronales. Mapas auto organizativos.

Las redes neuronales artificiales pueden definirse como modelos matemáticos que intentan simular la estructura y aspectos funcionales de las redes neuronales biológicas. Consisten en una interconexión de nodos (neuronas artificiales) y distintas etapas iterativas de procesado de la información. Al procesado de la información se le denomina aprendizaje que puede estar basada en información interna, inherente a nuestros datos, o externa proporcionada por el investigador (Cross et al., 1995).

Una herramienta útil para la proteómica es el empleo de mapas auto-organizativos, SOM (Kohonen, 2006), que pueden ser empleados para reducir la dimensionalidad de la muestra. De forma resumida este método permite agrupar de forma iterativa aquellos patrones que varían de forma más próxima entre sí, separándolos de aquellos que son más diferentes. Estos algoritmos son más robustos que los citados anteriormente puesto que están basados en redes neuronales que, por definición, son capaces de procesar datos con alta proporción de señales diferentes y de ruido. Wehrens y Buydens (2007) han desarrollado un excelente tutorial para entender la base matemática de este algoritmo y como realizar estos mapas, por lo que aquí se omitirá la descripción teórica del mismo.

A la hora de realizar un SOM debe tenerse en cuenta que el investigador es el que escoge la topología de la red, es decir, el número de nodos y la distribución inicial de los mismos. Generalmente han de probarse distintas topologías buscando la que mejor se adapte a nuestros datos. A continuación ejemplificaremos un análisis SOM sobre los datos de tabla2.

En R existen distintos paquetes para hacer mapas auto-organizativos, aquí se empleará el paquete kohonen que debe ser instalado y cargado antes de continuar con el ejemplo. En primer lugar cargaremos la tabla2 y le asignaremos los nombres a cada caso:

```
> tabla2 <- dget("http://www.valledor.info/wp-content/uploads/2013/12/
tabla2.txt")
> row.names(tabla2)<-c("SV-C1","SV-C2","SV-C3","A-C1","A-C2",
+ "A-C3","M-C1","M-C2","M-C3","SV-T1","SV-T2","SV-T3","A-T1",
+ "A-T2","A-T3","M-T1","M-T2","M-T3")
```

Al igual que ocurría en el caso de análisis PCA es necesario escalar nuestras variables para que puedan ser comparadas entre sí sin que existan desviaciones debidas a la magnitud de las mismas. A continuación sembraremos una semilla aleatoria que definirá la posición de las distintas muestras dentro de la red neuronal al inicio de nuestro análisis SOM:

```
> tabla2.sc <- scale(tabla2[,3:12])
> set.seed(32)
```

El comando som() posee una serie de argumentos que es necesario conocer, el primero de ellos es grid, que indica la topología de la red tanto en número de nodos como en su estructura hexagonal o toroidal. De forma "natural" se emplea la forma hexagonal. El segundo argumento es rlen, que indica el número de iteraciones que se realizarán. En nuestro ejemplo realizaremos un análisis con una topología hexagonal de 3 x 2 y 100 iteraciones (Figura 9):

```
> tabla2.som <- som(tabla2.sc, grid=somgrid(3,2,"hexagonal"),rlen=100)
> par(mfrow=c(2,2)) # definimos la matriz de graficos
> plot(tabla2.som)  # dibujamos distintos parámetros del análisis
> plot(tabla2.som, "mapping", label=row.names(tabla2), cex=0.75)
> plot(tabla2.som, "counts")
> plot(tabla2.som, "distance")
> plot(tabla2.som, "dist.neighbours")
```

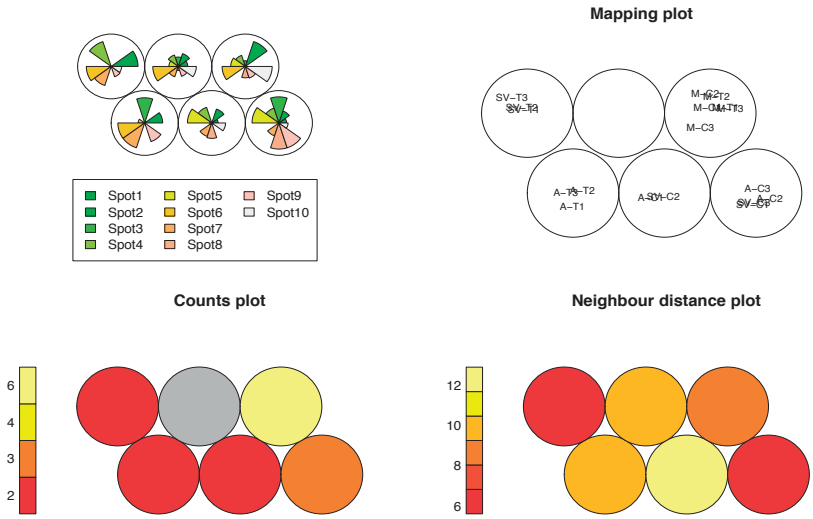


Figura 9: Mapas autoorganizativos. Ejemplos de análisis hexagonal 3x2 100 interacciones: "Mapping plot", "Counts plot", "Neighbour distance plot". Abreviaturas: A-C = Anaranjada-Control, A-T = Anaranjada-Tratamiento Frío, M-C = Morada-Control, M-T = Morada Tratamiento Frío, SV-C = Siempreverde-Control, SV-T = Siempreverde Tratamiento Frío.

La imagen anterior contiene numerosa información sobre nuestra muestra. En primer lugar en el "Mapping plot" podemos observar la topología de la red y cuáles son las muestras más próximas a cada uno de los nodos. En el ejemplo puede verse que en cada nodo se encuentran agrupadas un tipo de muestra diferente, con la excepción de A-C1, clasificada junto con las muestras moradas control. En el gráfico con los distintos *spots* se indica la relación de cada uno de ellos con el nodo, es decir, las muestras clasificadas en ese nodo se caracterizan por la presencia de los *spots* indicados y en unos niveles de expresión proporcionales al tamaño de su área. Por ejemplo, el primer nodo se caracterizaría por una fuerte presencia de *spot* 1 y media de 4,6 y 7. Las muestras moradas se caracterizarían por el *spot* 2, presente tanto en la muestra tratada como en la control. El *spot* 8 solo presente en los controles, se perdería con el tratamiento al frío. El "counts plot" indica el número de muestras clasificadas en cada nodo. El "Neighbour distance plot" indica la distancia de los distintos nodos entre sí. En realidad los nodos no se encuentran tangentes unos con otros como se puede ver en los gráficos, representándose así por convención, sino que la distancia entre cada uno de ellos dependerá del proceso de aprendizaje determinado por las distintas muestras. La información contenida en esta imagen puede ser utilizada para clasificar las distintas muestras lo cual permitiría mostrar la distancia real entre las mismas.

- **Ejercicio 10:**
  Repetir el análisis anterior cambiando la topología de la red a 2x2 y 1x6. Observar las diferencias. Obtener la matriz de distancias en cada uno de los dos casos.

  RESULTADO: Ver apéndice 2.

6. Conclusiones.

Para extraer toda la información de un experimento basado en 2-DE, además de prevenir artefactos o falsos positivos, es necesario partir de un diseño experimental correcto y un posterior análisis de datos riguroso. El establecimiento del adecuado número de réplicas, que se base en el poder estadístico que deseemos, así como el establecimiento de bloques experimentales homogéneos, precede a cualquier análisis de datos proteómicos. Puesto que los datos proteómicos es sabido que se ajustan mal a las pruebas estadísticas, es necesario un paso previo de procesado de los datos para mejorar el análisis, lo cual incluye la imputación de valores perdidos, normalización y transformación de los datos.

Decir que una diferencia es estadísticamente significativa desde una aproximación univariable (prueba-t, ANOVA) no es lo mismo que decir que es significativa basándonos en un método multivariable, tipo PCA, ICA o SOM. Sin embargo, dependiendo del objetivo del experimento, ambos tipos de pruebas son útiles. En un experimento de 2-DE es esencial el empleo de métodos multivariables para asegurarnos de que se ha extraído toda la información pertinente al experimento. El uso de PCA, ICA, PLS y métodos SOM ayuda en la definición de algunos procesos biológicos putativos o tendencias subyacentes en base de los diferentes grupos de proteínas establecidos por estos métodos estadísticos. No obstante, aunque los métodos multivariables nos dan una visión global de nuestros datos, a menudo es necesario emplear una estrategia univariable en la definición de marcadores biológicos. En ese caso se recomienda calcular el nivel de significancia para controlar la tasa de falsos positivos, usado FDR, o Bonferroni cuando es necesario un p-valor elevado. Por último, puesto que diferentes aproximaciones estadísticas proporcionan diferentes puntos de vista de nuestros datos, se recomienda llevar acabo diferentes pruebas estadísticas, ya que esto aumentará en gran medida la información obtenida de los geles 2-DE.

REFERENCIAS.

1. Albrecht D, Kniemeyer O, Brakhage AA, Guthke R. Missing values in gel-based proteomics. Proteomics 2010; 10: 1202-1211.

2. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 1995; 57: 289-300.

3. Berth M, Moser FM, Kolbe M, Bernhardt J. The State of the art in the analysis of two dimensional gel electrophoresis images. Applied Microbiology and Biotechnology 2007; 76: 1223-1243.

4. Brien CJ, Bailey RA. Multiple randomizations (with discussion). Journal of The Royal Statistical Society, Series B 2006; 68: 571–609.

5. Castillejo MA, Maldonado AM, Dumas-Gaudot E, Fernández-Aparicio M, et al. Differential expression proteomics to investigate responses and resistance to Orobanche crenata in Medicago truncatula. BMC Genomics 2009; 10: 294.

6. Chich JF, David O, Villers F, Schaeffer B, et al. Statistics for proteomics: experimental design and 2-DE differential analysis. Journal of Chromatography B 2007. 849:261-272.

7. Corder GW, Foreman DI. Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach. Wiley 2009. ISBN 978-0-470-45461-9.

8. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC. The dynamic range of protein expression: a challenge for proteomic research. Electrophoresis 2000; 21:1104-1115.

9. Cross SS, Harrison RF, Kennedt RL. Introduction to Neural Networks. The Lancet 1995; 346: 1075-1079.

10. Curto M, Valledor L, Navarrete C, Gutierrez D, et al. 2-DE based proteomic analysis of Saccharomyces cerevisiae wild and K+ transport-affected mutant (trk1,2) strains at the growth exponential and stationary phases. Journal of Proteomics, in press.

11. Daniel WW. Biostatistics: Basic Concepts and Methodology for the Health Sciences. Wiley Series in Probability and Statistics 2009. ISBN 978-0470413333. de Hoon MJL, Imoto S, Nolan S, Miyano S. Open Source Clustering Software. Bioinformatics 2004; 20: 1453-1454.

12. Kang Y, Techanukul T, Mantalaris A, Nagy JM. Comparison of three different commercially available DIGE Analysis Software Packages: Minimal User Intervention in Gel-Based Proteomics. Journal of Proteome Research 2009; 8:1077-1084.

13. Dowsey AW, Dunn MJ, Yang G-Z. Automated image alignment for 2D gel electrophoresis in a high-throughput proteomics pipeline. Bioinformatics 2008; 24: 950-957.

14. Dudoit S, van der Laan MJ, Pollard KS. Multiple Testing. Part I. Single-Step Procedures for Control of General Type I Error Rates. Statistical Applications in Genetics and Molecular Biology 2004; 3: article 13.

15. Egelhofer, V, Hoehenwarter, W, Lyon, D, Weckwerth, W, & Wienkoop, S (2013). Using ProtMAX to create high-mass-accuracy precursor alignments from label-free quantitative mass spectrometry data generated in shotgun proteomics experiments. Nature protocols, 8(3), 595-601.

16. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proceedings of the Natural Academy of Sciences of the USA 1998; 95: 14863-14868.

17. Eravci M, Fuxius S, Broedel O, Weist S, et al. Improved comparative proteome analysis based on two-dimensional gel electrophoresis. Proteomics 2007; 7: 513-523.

18. Eravci M, Mansmann U, Broedel O, Weist S, et al. Strategies for a Reliable Bioestadistical Analysis of Differentially Expressed Spots from Two-Dimensional Electrophoresis Gels. Journal of Proteome Research 2009; 8: 2601-2607.

19. Fay MP, Proschan MA. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretation of decision rules. Statistical surveys 2009; 4:1-39.

20. Fuxius S, Eravci M, Broedel O, Weist S, et al. Technical Strategies to reduce the amount of “false significant” results in quantitative proteomics. Proteomics 2008; 8: 1780-1784.

21. Gentleman R, Carey VJ, Bates DM, Boldstad B, et al. Bioconductor: open software development of computational biology and bioinformatics. Genome Biology 2004; 5: R80.

22. Germano T. Self Organizing Maps. http://davis.wpi.edu/~matt/courses/soms/ (Cited May 25th, 2010).

23. Gold DL, Miecznikowski JC, Liu S. Error control variability in pathway-based microarray analysis. Bioinformatics 2009; 25: 2216-2221.

24. Grove H, Hollung K, Uhlen A, Martens H, et al. Challenges related to analysis of protein spot volume from two-dimensional gel electrophoresis revealed by replicate gels. Journal of Proteome Research 2006; 5: 3399-3410.

25. Grove H, Jorgensen BM, Jessen F, Sondergaard I, et al. Combination of Statistical Approaches for Analysis of 2-DE Data gives complementary results. Journal of Proteome Research 2008; 7:5119-5124.

26. Han X, Aslanian A, Yates JR. Mass spectrometry for Proteomics. Current Opinion in Chemical Biology 2008; 12:483-490.

27. Horgan GW. Sample Size and Replication in 2D Gel Electrophoresis Studies. Journal of Proteome Research 2007; 6: 2884-2887.

28. Hunt SMN, Thomas MR, Sebastian LT. Pedersen SK, et al. Optimal replication and the importance of Experimental Design for Gel-Based Quantitative Proteomics. Journal of Proteome Research 2005; 4: 809-819.

29. HUPO Views. A Gene-centric Human Proteome Project. Molecular and Cellular Proteomics 2010; 9:427-429.

30. Hyvärinen A, Karhunen J, Oja E.Independent Component Analysis. New York: John Wiley & Sons 2001.

31. Hyvärinen A, Oja E. A Fast Fixed-Point Algorithm for Independent Component Analysis. Neural Computation 1997; 9: 1483-1492.

32. Hyvärinen A, Oja E. Independent Component Analysis: Algorithms and Applications. Neural Networks 2000; 13: 411-430.

33. Independent Component Analysis: A Tutorial Introduction. James V. Stone 2004. The MIT Press ISBN-13:978-0-262-69315-8

34. Jacobsen S, Grove H, Nedenkov Jensen K, et al. Multivariate of 2-DE protein patterns-Practical approaches. Electrophoresis 2007; 28: 1289-1299.

35. Jorge I, Navarro RM, Lenz C, Ariza D, et al. The holm oak leaf proteome: analytical and biological variability in the protein expression level assessed by 2-DE and protein identification tandem mass spectrometry de novo sequencing and sequence similarity searching. Proteomics 2005; 5:222-234.

36. Jorrín-Novo JV, Maldonado AM, Echevarría-Zomeño S, Valledor L, et al. Plant proteomics update (2007-2008): Second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge. Journal of Proteomics 2009; 72: 285-314.

37. Karp NA, Lilley K. Maximizing sensitivity for detecting changes in protein expression: Experimental design using minimal CyDyes. Proteomics 2005; 5: 3105-3115.

38. Karp NA, Griffin JL, Lilley KS. Application of partial least squares discriminant analysis to two-dimensional difference gel studies in expression proteomics. Proteomics 2005; 5: 81-90.

39. Karp NA, Lilley KS. Design and Analysis issues in Quantitative Proteomics Studies. Practical Proteomics 2007; 7 (S1): 43-50.

40. Karp NA, McCormick PS, Russel MR, Lilley KS. Experimental and statistical considerations to avoid false conclusions in proteomic studies using differential in-gel electrophoresis. Molecular and Cellular Proteomics 2007; 6: 1354-1364.

41. Karp NA, Spencer M, Lindsay H, O'Dell K, et al. Impact of replicate types on proteomic expression analysis. Journal of Proteome Research 2005; 4: 1867-1871.

42. Kendziorski C, Irizarry R, Chen K, Haag J, et al. On the utility of pooling biological samples in microarray experiments. Proceedings of the Natural Academy of Sciences USA 2005; 102: 4252-4257.

43. Kerr MK, Churchill GA. Experimental design for gene expression microarrays. Biostatistics 2001; 2:183-301.

44. Kim KY, Yi GS (2008) Sequential KNN imputation method v. 1.0.1. CRAN R project. http://cran.r-project.org/web/packages/SeqKnn/index.html

45. Klose J, Kobalz U. Two dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. Electrophoresis 1995; 16:1034-1059.

46. Klukas C, Schreiber F. Integration of –omics data and networks for biomedical research with VANTED. Journal of Integrative Bioinformatics 2010; 7: 112.

47. Kohonen T. Self-organizing neural projections. Neural Networks 2006; 19: 723-733.

48. Kreutz C, Timmers J. Systems Biology: Experimental design. FEBS J 2008; 276: 923-942.

49. Laboratory of Computer and Information Science. University of Helsinki. http://www.cis.hut.fi/research/ (Cited May 25th, 2010)

50. Lenth R. Some practical guidelines for effective sample size determination. The American Statician 2001; 55: 187-193.

51. Little RJA, Rubin DB. Statistical Analysis with Missing Data, 2nd Edition. Willey Interscience. New Jersey, USA. ISBN 0-471-18386-5.

52. Lopez JL. Two-dimensional electrophoresis in proteome expression analysis. Journal of Chromatography B 2007. 849:190-202.

53. Luhn S, Berth M, Hecker M, Bernhard J. Using standard positions and image fusion to create proteome maps from collections of two-dimensional gel electrophoresis images. Proteomics 2003; 3: 1117-1127.

54. Maldonado AM, Echevarría-Zomeño S, Jean-Baptiste S, Hernández M, Jorrín-Novo JV. Evaluation of three different precipitation protocols of protein extraction for Arabidopsis thaliana leaf proteome analysis by two-dimensional electrophoresis. Journal of Proteomics 2008; 71:461-472.

55. Martens H, Martens M. Multivariate Analysis of Quality-An introduction. 2001. John Willey & Sons.

56. Mead JA, Bianco L, Bessant C. Recent developments in public proteomic MS repositories and pipelines. Proteomics 2009; 9:861-881.

57. Meleth S, Deshane J, Kim H. The case of well conducted experiments to validate statistical protocols for 2D gels: different pre-processing = different list of significant proteins. BMC Biotechnology 2005; 5:7.

58. Meunier B, Dumas E, Piec I, Béchet D, et al. Assessment of Hierarchical Clustering Methodologies for Protemic Data mining. Journal of Proteome Research 2007; 6: 358-366.

59. Möller B, Posch S. Robust features for 2-DE gel image registration. Electrophoresis 2009; 30:4137-4148.

60. Molloy MP, Brzezinski EE, Hang J, McDowell MT, VanBogelen, RA. Overcoming technical variation and biological variation in quantitative proteomics. Proteomics 2003; 3: 1912-1919.

61. Morgan BJT, Ray APG. Non-uniqueness and Inversions in Cluster Analysis. Applied statistics 1995; 44: 117-134.

62. Morgenthal K, Wienkoop S, Scholz M, Selbig J, Weckwerth W. Correlative GC-TOF-MS-based metabolite profiling and LC-MS-based protein profiling reveal time-related systemic regulation of metabolite-protein networks and improve pattern recognition for multiple biomarker selection. Metabolomics 2005; 1: 109-121.

63. Mueller LN, Brusniak MY, Mani DR, Aebersold R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. Journal of Proteome Research 2008; 7:51-61.

64. Nedenskov Jensen K, Jessen F, JØrgensen M. Multivariate Data Analysis of Two-Dimensional Gel Electrophoresis Protein Patterns from Few Samples. Journal of Proteome Research 2008; 7: 1288-1296.

65. Noo ME, Tollenar R, Deelder AM, Bouwman LH. Current status and prospects of clinical proteomics studies on detection of colorectal cancer: Hopes and fears. World Journal of Gastroenterology 2006; 12: 6594-6601.

66. Oberg AL, Vitek O. Statistical design of quantitative mass spectrometry-based proteomic experiments. Journal of Proteome Research 2009; 8: 2144-2156.

67. Pardo A, Garrido J, Ruiz MA, San Martín R. La interacción entre factores en el análisis de varianza: errores de interpretación. Psicothema 2007; 19(2):343-349

68. Paoletti, A C, Parmely, T J, Tomomori-Sato, C, Sato, S, Zhu, D, Conaway, R C, Conaway, J W, Florens, L, and Washburn, M P (Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. Proceedings of the National Academy of Sciences 2006; 103: 18928-18933.

69. Patton WF. A thousand points of light: The application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. Electrophoresis 2000; 6: 1123-1144.

70. Pedreschi R, Hertog MLATM, Carpentier SC, Lammertyn J, et al. Treatment of missing values for multivariate statistical analysis of gel-based proteomics data. Proteomics 2008; 8: 1371-1383.

71. R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

72. Rodríguez-Piñeiro AM, Rodríguez-Berrocal FJ, Páez de la Cadena, M. Improvements in the search for potential biomarkers by proteomics: Application of principal component and discriminant analyses for two-dimensional maps evaluation. Journal of Chromatography B; 849: 251-260.

73. Rowell C, Carpenter M, Lamartiniere CA. Modeling biological variability in 2-DE gel proteomic carcinogenesis experiments. Journal of Proteome Research 2005; 4: 1619-1627.

74. Safavi H, Correa N, Xiong W, Roy A, et al. Independent component analysis of 2-D electrophoresis gels. Electrophoresis 2008; 29: 4017-4026.

75. Saldanha AJ. Java Treeview-extensible visualization of microarray data. Bioinformatics 2004; 20: 3246-3248.

76. Schultz J, Gottlieb DM, Petersen M, Nesic L, et al. Explorative data analysis of two-dimensional electrophoresis gels. Electrophoresis 2004; 25: 502-511.

77. Seiler M, Huang CC, Szalma S, Bhanot G. ConsensusCluster: A Software Tool for Unsupervised Cluster Discovery in Numerical Data. OMICS: A Journal of Integrative Biology 2010; 14: 109-113.

78. Sghaier-Hammami B, Valledor L, Drira N, Jorrin-Novo JV. Proteomic analysis of date palm (Phoenix dactylifera L.) zygotic embryos development and germination. Proteomics 2009; 9: 2543-2554.

79. Shaffer J P. Multiple Hypothesis Testing. Annual Review Psychology 1995; 46: 561-584.

80. Sigal A, Milo R, Cohen A, geva-Zatorsky N, et al. Variability and memory of protein levels in human cells. Nature 2006;444:643-645.

81. Smit S, Hoefsloot HCJ, Smilde AK. Statistical data processing in clinical proteomics. J Chromatogr B 2008; 866: 77-88.

82. Stainberg TH. Protein gel staining methods: An introduction and overview. Methods in Enzymology 2009; 463: 541-563.

83. Stessl M, Noe CR, Lachmann B. Influence of image-analysis software on quantitation of two-dimensional gel electrophoresis data. Electrophoresis 2009; 30:325-328.

84. Storey JD. A direct approach to false discovery rates. of the Royal Statistical Society, Series B 2002; 64: 479-498.

85. Strassburger K, Bretz F. Compatible simultaneous lower confidence bounds for the Holm procedure and other Bonferroni-based closed tests. Statistics in Medicine 2008;27: 4914-4927.

86. Tamayo P, Slonim D, Mesirov J, Zhu Q, et al. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. Proceedings of the Natural Academy of Sciences of the USA 1999; 96: 2907-2912.

87. Taylor CF, Paton NW, Lilley KS, Binz PA, et al. The minimum information about a proteomics experiment (MIAPE). Nature Biotechnology 2007; 25:887-893.

88. Tyers M, Mann M. From Genomics to Proteomics. Nature 2003; 422:193-197.

89. Valero J, Echevarría-Zomeño S, Ariza D, Valledor L, et al. Proteomics of Holm oak. Plant Proteomics in Europe FA0603. Nitra Meeting 2009.

90. Valledor L, Castillejo MA, Lenz C, Rodríguez R, et al. Proteomic analysis of Pinus radiata needles: 2-DE map and protein identification by LC/MS/MS and substitution-tolerant database searching. Journal of Proteome Research 2008; 7: 2616-2631.

91. Valledor L, Jorrín JV. Back to the basics: Maximizing the information obtained by quantitative two dimensional gel electrophoresis analysis by an appropriate experimental design and statistical analyses. Journal of proteomics 2011; 74: 1-18.

92. Valledor L, Jorrin-Novo JV, Rodríguez JL , Lenz C. Combined Proteomic and Transcriptomic analysis identifies differentially expressed pathways and stress responses associated to Pinus radiata needle maturation. Journal of Proteome Research 2010; 9: 3954–3979.

93. Valledor, L., Furuhashi, T., Hanak, A. M., & Weckwerth, W. Systemic cold stress adaptation of Chlamydomonas reinhardtii. Molecular & Cellular Proteomics 2013; 12(8):2032-2047.

94. Valledor L, Recuenco-Muñoz L, Egelhofer V, Wienkoop S, Weckwerth W. The impact of the different genome annotation releases over the identification of proteins in proteogenomic studies: same spectra but different results. Journal of Proteomics 2012; 75(18):5883-5887.

95. Verhoeven KJF, Simonsen KL, McIntyre LM. Implementing False Discovery Rate control: increasing your power. Oikos 2005; 108: 643-647.

96. Wasinger V, Cordwell S, Cerpa-Poljak A, Yan J, et al. Progress with gene-product mapping of Mollicutes: *Mycoplasma genitalium*. Electrophoresis 2005; 16:1090-1094.

97. Weherns R, Buydens LMC. Self- and Super-organizing Maps in R: The Kohonen Package. Journal of Statistical Software 2007; 21:1-19

98. Wheelock AM, Buckpitt AR. Software-induced variance in two-dimensional gel electrophoresis image analysis. Electrophoresis 2005; 26: 4508-4520.

99. Wheelock AM, Goto S. Effects of post-electrophoretic analysis on variance in gel-based proteomics. Expert Reviews on proteomics 2006; 3:129-142.

100. Wilkins MR, Apple RD, Van Eyk JE, Chung MCM. Guidelines for the next 10 years of proteomics. Proteomics 2006; 6:4-8.

101. Zhao Y, Lee W-NP, Xiao GG. Quantitative Proteomics and Biomarker Discovery in Human Cancer. Expert Review in Proteomics 2009; 6: 115-118.

INFORMACIÓN SUPLEMENTARIA Y APÉNDICES.

Cuadro 1

Herramientas y Recursos para el análisis estadístico disponibles on-line que se emplearán durante el desarrollo del capítulo. Todas ellas son gratuitas y multiplataforma.

Paquete estadístico R y distintos módulos de interés:

Distribuciones binarias de R: http://cran.r-project.org/
SeqKNN tool para R: http://cran.r-project.org/web/packages/SeqKnn/
FDR tool para R: http://cran.r-project.org/web/packages/fdrtool/
ICA package para R: http://cran.r-project.org/web/packages/fastICA/
PLS package para R: http://cran.r-project.org/web/packages/pls/
SOM package para R: http://cran.r-project.org/web/packages/som/
Gplots package para R: http://cran.r-project.org/web/packages/gplots/
Ggplot2 package para R: http://cran.r-project.org/web/packages/ggplot2/
pheatmap package para R: http://cran.r-project.org/web/packages/pheatmap/

Manuales introductorios a R:

En inglés: http://cran.r-project.org/manuals.html
En castellano: http://cran.r-project.org/other-docs.html#nenglish

Otras herramientas:

Herramienta de Russ Lenth para el cálculo de la potencia estadística y el tamaño de muestra: http://www.stat.uiowa.edu/~rlenth/Power/

Cluster 3.0: http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm

JavaTreeView: http://jtreeview.sourceforge.net/

Protmax: http://www.univie.ac.at/mosys/software.html

Un espejo de todas las herramientas empleadas en este capítulo se encuentra disponible en la dirección: http://www.unioviedo.es/epiphysage/mp/

Cuadro 2:

Varianza: (σ^2) En teoría de probabilidad, la varianza o coeficiente de variación de una variable aleatoria es una medida de su dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media. La varianza poblacional (con n-individuos) se calcula con la siguiente fórmula:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(\chi_i - \bar{\chi})^2$$

Desviación típica: (σ) se define como la raíz cuadrada de la varianza.

Coeficiente de Variación: (CV) es una medida de dispersión normalizada que se define como la relación entre la desviación estándar y la media. Normalmente se expresa en %. La media de CV% de una muestra con k variables se define mediante la siguiente fórmula:

$$CV\% = \frac{1}{k}\sum_{i=1}^{\kappa}\left(\frac{\sigma_k}{\bar{\chi}_k}\right)$$

Covarianza: es la medida de la variación común de dos variables y, por tanto, una medida del grado y tipo de relación entre ambas.

Homocedasticidad: dos o más muestras son homocedásticas si tienen la misma varianza.

Mediana: (χ ó $\mu_{1/2(\chi)}$) es el valor de la variable que deja el mismo número de datos antes y después que él, una vez ordenados estos. De acuerdo con esta definición el conjunto de datos menores o iguales que la mediana representarán el 50% del total de datos de la muestra, y el otro 50%, los mayores.

Valores perdidos: se produce un valor perdido cuando se pierde el valor de un dato para una variable en una observación determinada.

Test no paramétricos: conjunto de métodos de análisis estadístico que no se basan en la asunción de que los datos sean homocedásticos y/o normales.

Normalidad: la distribución de una variable se denomina "normal" si sigue una distribución normal o Gaussiana, la cual se caracteriza por presentar el mayor número de valores concentrados cerca de la media disminuyendo la frecuencia de los mismos a medida que se incrementa la distancia con la media.

Test paramétricos: conjunto de test estadísticos que asumen que los datos siguen una distribución normal y son homocedásticos e independientes. Como ejemplo de test paramétricos tenemos la prueba t y el ANOVA.

Hipótesis nula: (H_0) la hipótesis nula es la hipótesis que se contrasta mediante la aplicación de una prueba estadística concreta. La hipótesis nula no se considera como resultado, a pesar de que puede ser rechaza por los datos.

Hipótesis alternativa: alternativa a la hipótesis nula. Cuando no se define, se puede considerar H_1 como "H_0 es falsa".

Valores atípicos: Aquellos datos cuyos valores son extremos o atípicamente diferentes del resto.

p-valor: es la probabilidad de obtener un resultado al menos tan extremo como el que realmente se ha obtenido (valor del estadístico calculado), suponiendo que la hipótesis nula es cierta. El p-valor es convencionalmente considerado significativo si $p < 0{,}05$ ó 0,01.

Nivel de significancia: (α) Es la máxima probabilidad de error que estamos dispuestos aceptar para dar como válida nuestra hipótesis Un p-valor más bajo que el nivel de significancia exige rechazar la hipótesis nula.

Poder estadístico: es la probabilidad de que mediante prueba estadística determinada se rechace una hipótesis nula falsa. Se calcula como $1 - \beta$.

Apéndice 1: Tablas de datos empleadas en el capitulo

Tabla 1: Para cargar la tabla en R empleando una conexión a internet:

```
> tabla1 <- dget("http://www.valledor.info/wp-content/uploads/2013/12/
tabla1.txt")

      G1       G2       G3       G4       G5       G6
1   0.5400   0.4500   0.5000   0.4500   0.5400   0.4800
2   0.8720   0.8480     NA     0.8480   0.8720   0.8240
3   0.8636   0.8090   0.8363   0.9546   0.9091   0.9546
4   0.9184   0.9302   1.0802   1.0306   1.0714   1.0714
5   1.1460   1.0540   1.0080   1.2150   1.1115   1.0540
6   1.2889   1.3270   1.1492   1.0857   1.0857   1.2508
7   1.6600   1.6440   1.3560   1.5640   1.4040   1.3880
```

Tabla 2: Para cargar la tabla en R empleando una conexión a internet:

```
> tabla2 <- dget("http://www.valledor.info/wp-content/uploads/2013/12/tabla2.txt")

   Trat Variedad   Spot1    Spot2    Spot3    Spot4    Spot5    Spot6    Spot7    Spot8    Spot9  Spot10
1    c  siempreverde 2.373711  6.700789  3.399847 13.723018 18.514947  2.771076 20.37435 10.147177  6.820759 40.17432
2    c  siempreverde 1.962564  7.286019  3.054240 13.369968 16.750485  3.627909 21.58820 11.398085  5.699728 40.26280
3    c  siempreverde 2.749107  7.947417  2.936546 12.745858 19.613626  3.444222 20.99318 15.234579  7.729940 31.60553
4    c   anaranjada 3.891381 18.159777  2.812462  8.041782 16.486847  2.490281 24.97272 10.207177  5.423980 32.51359
5    c   anaranjada 4.172354 16.225822  4.378396  7.805819 16.462771  1.507278 23.17975 16.061291  6.700759 28.50576
6    c   anaranjada 3.829937 15.741391  3.841649  7.908165 15.928789  1.382270 25.06442 10.167177  7.699940 33.43626
7    c      morada  2.921974 19.479824  3.639088  8.941782 13.261265  6.880160 10.59617 12.765402  5.879728 40.63461
8    c      morada  3.335619 21.613983  2.951140  9.278116 12.016543  6.422280 10.49525 13.616566  6.224803 39.04570
9    c      morada  3.144846 22.390490  3.114314  9.800928 12.249634  5.662989 10.07572 11.308085  6.543952 40.70904
10   t  siempreverde 7.579709  4.564126  2.935711 10.075721  7.428115 12.857602 33.74193  7.540021  5.839728 32.43734
11   t  siempreverde 7.585627  4.386471  3.190910 12.862585  5.962962 12.069769 30.27655  6.451821  6.414051 35.79926
12   t  siempreverde 8.776142  4.265536  2.657681 13.372854  8.087298 11.496264 28.71289  7.862293  6.613952 33.15509
13   t   anaranjada 5.672307  5.395609  3.458724  7.405819  8.973312 11.496264 40.67459  7.590021  5.829728 28.50363
14   t   anaranjada 6.029112  5.454911  3.301657  7.608165  9.706870 12.069769 39.39020  6.341821  7.769940 27.32756
15   t   anaranjada 5.759456  5.183511  4.175606  8.241782 10.431815 15.462311 37.09090  7.892293  6.931348 23.83098
16   t      morada  0.000000 19.181052  3.127346  9.486281 12.642815 16.440116  0.00000  7.933996  5.433980 50.75441
17   t      morada  0.000000 20.184309  3.465992 10.397977 12.975045 15.462311  0.00000  7.453325  6.254803 48.80624
18   t      morada  0.000000 19.086371  2.516884  9.752926 13.977096 16.935353  0.00000  6.371821  7.599940 48.75961
```

Apendice 2: Respuesta a los ejercicios

Ejercicio 1: Respuestas: a, 394; b, 17; c, 6

Ejercicio 2: Resultados: para la raíz cúbica 2,52%; para el seno 4,29%.

Ejercicio 3: En primer lugar se crea un objeto conteniendo la información de los distintos grupos:

```
> g<-factor(paste("Group",rep(1:2, each=3)))
```

Este objeto es necesario para poder realizar el test. El bucle para realizar el test podría programarse como se indica a continuación:

```
> i=1
       #ponemos el contador en 3, primera columna numérica
> pvalores <- c()  #creamos el objeto pvalores
> while(i <=ncol(tablaimputada)){
       #se abre bucle, que durara hasta que el contador i sea igual al
numero de columnas de la tablaimputada.
+ a <-bartlett.test(tablaimputada[,i], g)
       #Test de Bartlett
+ pvalores <- c(pvalores,a$p.value)
#se añade el p-valor a al objeto con los distintos p-valores.
+ i=i+1}
```

Los valores resultantes de la aplicación del Test de Bartlett se encuentran ordenados dentro del objeto pvalores.

Ejercicio 4:
Bucle para realizar una prueba T:

```
> i=3
#ponemos el contador en 3, primera columna numérica
> pvalores <- c()
#creamos el objeto pvalores
> while(i <=ncol(tabla2)){
#se abre bucle, que durara hasta que el contador i sea igual al
numero de columnas de la table 2.
+ a <-t.test(tabla2[,i]~Trat,data=tabla2)
#PruebaT, agrupando por columna Trat
+ pvalores <- c(pvalores,a$p.value)
       #se añade el pvalor a al objeto con los distintos p-valores
+ i=i+1}
       Bucle para realizar una prueba U:
> i=3
       #ponemos el contador en 3, primera columna numérica
> pvaloreswt <- c()
       #creamos el objeto pvalores
> while(i <=ncol(tabla2)){
       #se abre bucle, que durara hasta que el contador i sea
igual al numero de columnas de la table 2.
+ a <-wilcox.test(tabla2[,i]~Trat,data=tabla2)
       #PruebaU, agrupando por columna Trat
```

```
+ pvaloreswt <- c(pvaloreswt,a$p.value)
      #se añade el pvalor a al objeto con los distintos p-valores
+ i=i+1}
```

Los pvalores se encuentran en los objetos pvalores y pvaloreswt. Comparandolos podremos ver las diferencias existentes entre los mismos.

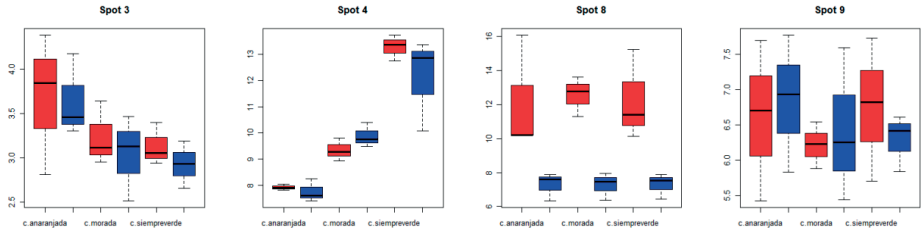
Ejercicio 5:

Se realizará un test ANOVA seguido de Tukey para todos los spots que muestran diferencias significativas (Todos los spots menosel 3, 8 y 9) tal y como se indica en el texto del capítulo. El bucle para realizar el test de Kruskal Wallis se puede programar como se indica a continuación:

```
> i=3
      #ponemos el contador en 3, primera columna numérica
> pvaloreskw <- c()
      #creamos el objeto pvalores
> while(i <=ncol(tabla2)){
      #se abre bucle, que durara hasta que el contador i sea
igual al numero de columnas de la table 2.
+ a <-kruskal.test(tabla2[,i]~Trat,data=tabla2)
      #PruebaKruskal
+ pvaloreskw <- c(pvaloreskw,a$p.value)
      #se añade el pvalor a al objeto con los distintos p-valores
+ i=i+1}
```

Ejercicio 6:

Representación Gráfica:




```
> par(mfrow=c(1,4))
> boxplot(tabla2[,'Spot3']~Trat*Variedad,data=tabla2, main='Spot 3',
col=c("red","blue"))
> boxplot(tabla2[,'Spot4']~Trat*Variedad,data=tabla2, main='Spot 4',
col=c("red","blue"))
> boxplot(tabla2[,'Spot8']~Trat*Variedad,data=tabla2, main='Spot 8',
col=c("red","blue"))
> boxplot(tabla2[,'Spot9']~Trat*Variedad,data=tabla2, main='Spot 9',
col=c("red","blue"))
```

Programación del bucle, presentación de resultados:

```
> pvalores <- c()
      #creamos el objeto vacio pvalores
```

```
> i=3
        #contador i con valor 3, primera columna con datos
> while(i<ncol(tabla2)){
        #se abre el bucle hasta que i sea menor que
        # ncol. El cambio se explica por añadir al final de
        # la tabla la columna Tanda, sobre la que no
        # deseamos realizar el test.
+ res.anova=aov(tabla2[,i]~Trat*Variedad+Tanda,data=tabla2) #se crea
        #res.anova, conteniendo el análisis de la varianza
        #aov. Se indican la interacción y los datos.
+ pvalores <- c(pvalores,anova(res.anova)$'Pr(>F)'[1:4])
        #añadimos al
        #objeto pvalores los pvalores del ANOVA
+ i=i+1
        #aumentamos en 1 unidad el contador
+ }
        #cerramos el bucle.

> pvalores
 [1] 1.561042e-06  1.337975e-09  1.405877e-01  5.049753e-10  9.372288e-08
 [6] 6.994867e-11  7.288686e-01  3.236066e-06  5.481457e-01  1.024418e-01
[11] 9.405187e-01  9.439996e-01  4.777310e-01  1.151470e-06  1.497732e-01
[16] 2.009904e-01  3.742342e-08  8.730045e-01  3.592859e-01  7.505821e-07
[21] 6.797569e-10  9.912412e-05  7.076525e-01  2.158970e-01  6.166060e-06
[26] 1.713667e-12  7.693409e-02  5.191261e-09  2.084963e-04  9.843477e-01
[31] 7.056904e-01  9.799528e-01  9.893933e-01  5.547455e-01  3.896326e-03
[36] 5.659893e-01  8.210605e-01  1.339283e-06  1.511363e-01  8.347489e-04

> pval.ordenado <- t(matrix(pvalores,4))
> pval.ordenado
             [,1]             [,2]            [,3]            [,4]
 [1,]    1.561042e-06     1.337975e-09     0.140587658     5.049753e-10
 [2,]    9.372288e-08     6.994867e-11     0.728868596     3.236066e-06
 [3,]    5.481457e-01     1.024418e-01     0.940518688     9.439996e-01
 [4,]    4.777310e-01     1.151470e-06     0.149773219     2.009904e-01
 [5,]    3.742342e-08     8.730045e-01     0.359285925     7.505821e-07
 [6,]    6.797569e-10     9.912412e-05     0.707652484     2.158970e-01
 [7,]    6.166060e-06     1.713667e-12     0.076934087     5.191261e-09
 [8,]    2.084963e-04     9.843477e-01     0.705690441     9.799528e-01
 [9,]    9.893933e-01     5.547455e-01     0.003896326     5.659893e-01
[10,]    8.210605e-01     1.339283e-06     0.151136332     8.347489e-04
```

Para copiar al portapapeles:

```
> write.table(pval.ordenado,file="clipboard")
```

Ejercicio 7:

Instalar y cargar el paquete fdrtool, posteriormente ejecutar el comando indicando el objeto con los pvalores. En este ejercicio emplearemos las opciones por defecto:

```
> fdrpvalores <- fdrtool(pvalores)
```

Una vez realizado el test veremos una representación gráfica de la densidad de probabilidades de pvalores, asi mismo en el objeto creado se almacenarán los p- y q-valores así como el FDR local. Para obtener los q-valores deberemos emplear:

```
> fdrpvalores$qval
```

Este listado lo podremos exportar u ordenar en el propio R para poder definir el valor de corte, según la distribución de los mismos tal y como se explica en el capítulo.

Ejercicio 8:

```
> tabla2 <- dget("http://www.valledor.info/wp-content/uploads/2013/12/tabla2.txt")
> row.names(tabla2)<-c("SV-C1","SV-C2","SV-C3","A-C1","A-C2",
+ "A-C3","M-C1","M-C2","M-C3","SV-T1","SV-T2","SV-T3","A-T1",
+ "A-T2","A-T3","M-T1","M-T2","M-T3") #Nombramos los casos
> tabla2.ica <- fastICA(tabla2[,3:12], 3)
       #Se hace el ICA y un nuevo objeto, tabla2.ica
```

Gráfico 1:

```
> plot(tabla2.ica$S,main="ICAcomponents", ylab="IC2", xlab="IC1", pch=16)
> text(tabla2.ica$S, row.names(tabla2), cex=0.7, adj=c(0.5,-0.75))
```

Gráfico 2:

```
> plot(tabla2.ica$S[,2],tabla2.ica$S[,3],main="ICA components", pch=16,
ylab="IC3", xlab="IC2")
> text(tabla2.ica$S[,2],tabla2.ica$S[,3], row.names(tabla2), cex=0.7,
adj=c(0.5,-0.75))
Matriz de Carga (Stimated Mixing Matrix):
> tabla2.ica$A
           [,1]       [,2]        [,3]      [,4]      [,5]
[1,] -0.2856388 -1.5288050 -0.21017580 0.9513070 -1.737716
[2,]  1.9433935 -6.7130018  0.04574451 0.5027478 -1.168813
[3,] -1.0194792 -0.4304683 -0.09113522 1.5834990  2.714446
           [,6]      [,7]       [,8]        [,9]     [,10]
[1,]  4.3563171 -3.424737 -2.1590861 -0.09180634  4.130341
[2,] -0.2244336 11.846094 -0.7215237  0.11730954 -5.627518
[3,] -2.9172505 -3.626990  1.2929665  0.04568762  2.448724
```

Ejercicio 9:

Ejemplo empleando la distancia Manhattan, también descrita como de utilidad para la clasificación de muestras proteómicas y el método de agregación de Ward. Antes de continuar es necesario tener instalado el paquete pheatmap.

```
> pheatmap(tabla2[,3:12], clustering_distance_rows=dist(tabla2[,3:12],
method = "manhattan"), clustering_distance_cols=dist(t(tabla2[,3:12]),
method = "manhattan"), clustering_method="ward",display_numbers=TRUE)
```

Ejercicio 10:

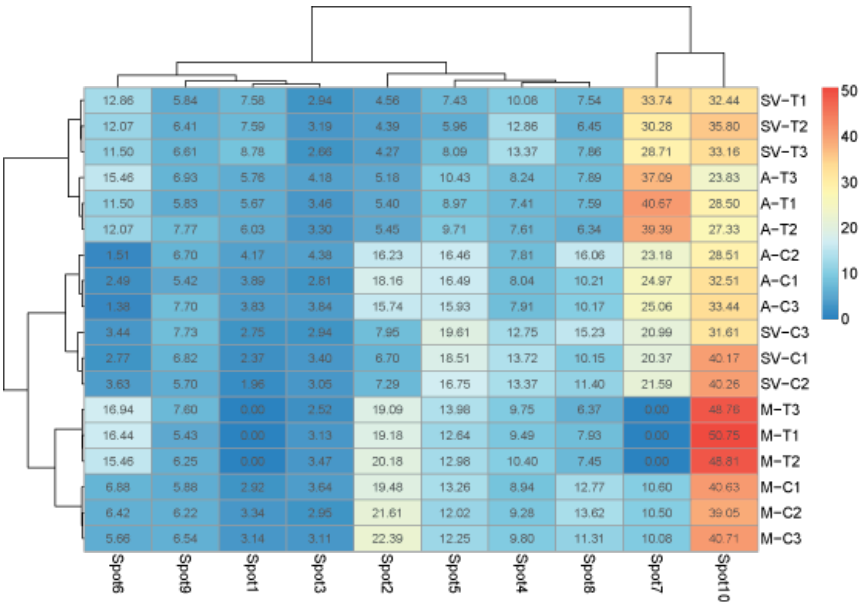
Realización del SOM con las distintas topologías:

```
> tabla2.sc <- scale(tabla2[,3:12])
> set.seed(7)
```

```
> tabla2.som <- som(tabla2.sc, grid=somgrid(2,2,"hexagonal"),rlen=100)
> par(mfrow=c(2,2))
      # definimos la matriz de graficos
> plot(tabla2.som)
      # dibujamos distintos parámetros del análisis
> plot(tabla2.som, "mapping", label=row.names(tabla2), cex=0.75)
> plot(tabla2.som, "counts")
> plot(tabla2.som, "distance")
> plot(tabla2.som, "dist.neighbours")

> set.seed(23)
> tabla2.som <- som(tabla2.sc, grid=somgrid(2,2,"hexagonal"),rlen=100)
> par(mfrow=c(2,2))
      # definimos la matriz de graficos
> plot(tabla2.som)
      # dibujamos distintos parámetros del análisis
> plot(tabla2.som, "mapping", label=row.names(tabla2), cex=0.75)
> plot(tabla2.som, "counts")
> plot(tabla2.som, "distance")
> plot(tabla2.som, "dist.neighbours")
```

Obtención de la matriz de distancias:

```
> tabla2.som$distances
      #Obtenemos la distancia de cada muestra a cada uno de los nodos.

> dists <- unit.distances(tabla2.som$grid, toroidal=FALSE)
      #Obtenemos la distancia de cada uno de los nodos entre si.

> plot(tabla2.som, type="property", property=dists[1,], main= "Distances
to unit 1", zlim=c(0,4),palette = rainbow, ncolors = 2, contin = TRUE)
```



Historia de la espectrometría de masas. Principios básicos. Partes y tipos de espectrómetros de masas

06

Emilio Gelpí

Emilio Gelpí, Instituto de Investigaciones Biomédicas de Barcelona (IIBB-CSIC-IDIBAPS) Laboratorio de Proteómica CSIC/UAB. Facultad de Medicina (UAB).

1. Introducción.

Como el titulo indica, este capitulo repasa brevemente los aspectos históricos del desarrollo de la espectrometría de masas, sus principios básicos y los diferentes tipos y geometrías de estos instrumentos, algunos de estos ya en desuso y otros hoy en día totalmente imprescindibles en los estudios de Proteómica. La cronología completa de los avances tanto instrumentales como metodológicos de esta técnica y su correspondiente bibliografía detallada se aborda en el capitulo 9. Existe una gran variedad de publicaciones, capítulos de libro o libros enteros sobre la temática objeto de este capitulo por lo que resulta difícil dar recomendaciones concretas. No obstante, para quienes deseen un tratamiento en más profundidad, tanto de los aspectos históricos como de los Instrumentales, se recomienda la consulta de los tres textos siguientes.

- Measuring Mass. From Positive Rays to Proteins. MA, Grayson, Editor. 2002, American Society for Mass Spectrometry. 1-149.

- Watson JT, Sparkman OD, Introduction to Mass Spectrometry. Instrumentation, Applications and Strategies for Data Interpretation. John Wiley and Sons, Ltd. 4th ed. 2007. Las figuras 9, 10, 12, 13, 14 y 15 han sido reproducidas a partir de esta referencia con el permiso de John Wiley and Sons.

- An Introduction to Mass Spectrometry in the Health Sciences. Tony Mallet. Published by IM Publications LLP 6 Charlton Mill, Charlton, Chichester, West Sussex PO18 0HY, UK. T: +44-1243-811334, F: +44-1243-811711, E: info@impublications.com, www.impublications.com

2. El Inicio Histórico.

En 1886 el físico alemán Eugene Goldstein describió, en la que podríamos considerar la primera publicación de Espectrometría de Masas (EM), que los rayos de luz de electricidad positiva (Kanalstrahlen) en el interior de un tubo de descarga con un gas a baja presión se expandían en líneas rectas a partir de los orificios de un disco de metal perforado que se utilizaba como

cátodo en su experimento (Figura 1). Estas observaciones fueron confirmadas y ampliadas en 1898 por otro físico alemán, Wilhelm Carl Wien, quien demostró que estos rayos eran desviados por un campo magnético y que su comportamiento podía estudiarse mediante los efectos combinados de campos eléctricos y magnéticos. El aparato utilizado por Wien en sus experimentos consistía en un sistema de campos eléctrico y magnético perpendiculares entre si que podía utilizarse como filtro de velocidad de partículas con carga eléctrica. Al mismo tiempo en 1897, el físico ingles Sir John Joseph Thomson, quien en 1906 recibió el Premio Nóbel por su descubrimiento del electrón, desarrolló un sistema para determinar la diferencia entra la relación masa/carga entre un electrón y un átomo de hidrógeno y fue el primero en demostrar que los rayos catódicos podían desviarse de su trayectoria por el efecto de un campo eléctrico.

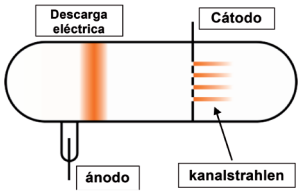


Figura 1. Esquema del aparato utilizado por JJ Thomson para la producción de rayos de electricidad positiva (Kanalstrahlen) entre el ánodo y un cátodo en un tubo lleno de gas a baja presión.

En la figura 2 se ilustra el esquema del segundo de los aparatos utilizados por Thomson en estos estudios. Mediante ajustes de los campos eléctrico y magnético obtuvo el valor de la relación de la carga ***e*** a la masa ***m*** de los rayos catódicos. Repitiendo el experimento para gases diferentes determinó que el valor de ***e/m*** era independiente de la naturaleza del gas y del material del eléctrodo y que por tanto estos "corpúsculos con carga negativa" eran lo que hoy conocemos como electrón. Un aspecto a destacar de estas primeras investigaciones y que naturalmente es aplicable a todo el desarrollo posterior de la espectrometría de masas hasta nuestros días es la necesidad de trabajar con gases a presiones reducidas, lo cual hace 100 años no dejaba de ser un importante desafío tecnológico. Para poder calcular con precisión las relaciones masa/carga había que asegurarse de minimizar el efecto de posibles colisiones entre las partículas cargadas y las moléculas de gas, mucho más masivas y por tanto con posibilidades de alterar la trayectoria de la partícula que se observa. Otro aspecto importante fue el de la creación de "haces de electrones" con "velocidad conocida". Calentando un hilo de metal a alta temperatura se obtiene una nube de electrones pero para obtener un haz de electrones o de cualquier tipo de partícula con carga eléctrica, que se muevan en la dirección deseada y con la velocidad necesaria, es preciso aplicar otro campo eléctrico para acelerar las partículas o electrones lejos del hilo emisor. Este principio todavía se usa hoy para crear haces de iones moleculares en EM.

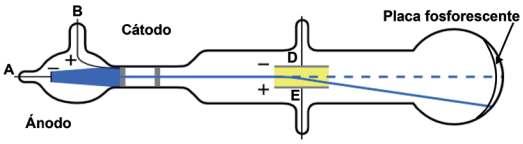


Figura 2. Aparato de Thomson para el estudio de los rayos catódicos. A. Ánodo, C. Cátodo. Para provocar una descarga eléctrica a través del gas se usaban los electrodos indicados por los signos + y – en la figura. Los rayos catódicos resultantes se aceleraban hacia el bulbo derecho del aparato incidiendo sobre una placa fosforescente en donde producían una imagen con la que podía medirse la deflexión de estos rayos. Esta desviación se conseguía aplicando una diferencia de potencial a través del par de placas paralelas D y E. El cambio de trayectoria también se conseguía colocando un campo magnético a cualquier lado del tubo en la región de las placas metálicas (no incluido en la figura).

En 1907 Thomson empezó sus investigaciones sobre los rayos de electricidad positiva (Kanalstrahlen) con la idea de identificar la partícula portadora de carga eléctrica de estos rayos. Estas partículas diferían mucho de los haces de electrones y se comportaban de forma diferente para cada gas siendo necesario recurrir a campos mucho mas potentes para desviarlas. Cuando sometió los "Kanalstrahlen" a campos eléctricos y magnéticos combinados en un aparato similar al utilizado para el estudio de los rayos catódicos observó varias "líneas" sobre una pantalla fosforescente cuya posición dependía del gas utilizado para generar estos rayos. Por ello su aparato se considera el primer espectrógrafo de masas.

Dos años después, en 1909 Francis W. Aston empezó a trabajar en el laboratorio de Thomson en Cavendish en donde construyo un nuevo espectrógrafo de masas (Figura 3), modificando para ello el experimento de Thomson, en el que los campos eléctrico y magnético actuaban sobre la misma posición de modo que las partículas cargadas atravesaban los dos simultáneamente. En el equipo de Aston las partículas pasaban primero por el campo eléctrico antes de penetrar en el magnético. Con esta disposición la resolución de la imagen de la partícula cargada (ion) solo dependía de su masa y no de su velocidad. Aston registro las imágenes sobre una tira de papel fotográfico y acuño el término "espectrógrafo de masas" que utilizó para sus experimentos en "espectroscopia de masas". La posición de las imágenes le indico a las masas de los diferentes iones en el haz. De esta forma Aston pudo separar primero los dos isótopos del Neon y del Cloro y después los de otros muchos elementos, por lo que recibió el premio Nóbel en 1922.

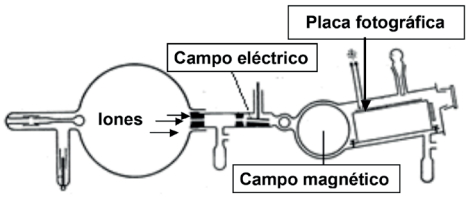


Figura 3. Espectrógrafo de Masas de FW Aston.

En 1918 Arthur Dempster, un físico de la Universidad de Chicago diseñó un espectrómetro de masas (1) con enfoque magnético direccional en el que un haz de iones monoenergéticos, al atravesar un campo magnético de 180° se dividía en diversas trayectorias semicirculares de radio determinado por su masa y velocidad para focalizarse sobre el plano del detector (Figura 4). Iba equipado con fuentes de ionizacion por impacto de electrones y de ionizacion térmica. Dos años antes el mismo Dempster ya había propuesto que el mecanismo de ionizacion del hidrogeno por impacto de electrones "parecía implicar solo el desprendimiento de un electrón" de la molécula de hidrogeno. Después de esto tuvieron que pasar varios años sin acontecimientos destacables en el campo del diseño e introducción de nuevos instrumentos hasta que en 1934 Joseph Mattauch y Richard Herzog desarrollaron el concepto del doble enfoque.

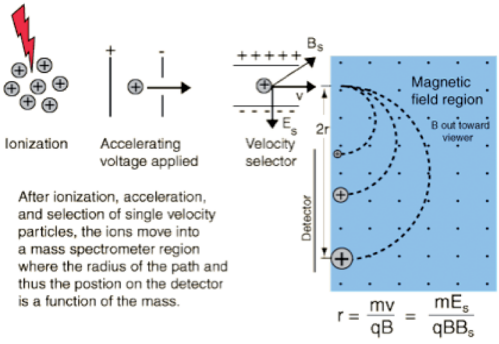


Figura 4. EM diseñado por A Dempster.

Los instrumentos de doble enfoque se basan en la incorporación de un sector eléctrico (*E*), después del acelerador de iones en la fuente de iones, para la selección de iones monoenergéticos y un sector magnético (B) para analizar y separar los iones según su momento. Las determinaciones de masas con una exactitud de unas pocas partes por millón requieren haces de iones monoenergéticos en contra de lo que sucede a la salida de la fuente de iones donde el haz de iones presenta una dispersión inicial de energías cinéticas. Esta dispersión es uno de los factores limitantes en las prestaciones de un espectrómetro de enfoque simple. Si por ejemplo esta distribución de energías cinéticas de los iones tiene un margen de 0 a 3 voltios y el haz sufre una aceleración de extracción de la fuente de iones de 3000 voltios, la distribución de energía cinética estará entre 3000 ± 3 V, lo cual quiere decir que, en estas condiciones, la resolución de las masas de los iones no podrá superar 1 parte por 1000. Por el contrario la adición del sector eléctrico en un equipo de doble enfoque permitirá seleccionar aquellos iones cuya energía cinética este comprendida en el margen de 3000 ± 0.03 V lo que se traduce en una resolución de 1 parte en 100.000, primero por la corrección de la dispersión angular del haz de iones por enfoque de energía en el campo eléctrico y después enfoque por momento en el campo magnético.

El esquema del diseño óptico del espectrómetro de masas de doble enfoque Mattauch-Herzog (2) se ilustra en la figura 5. Como se indica en el esquema todos los iones se separan y enfocan simultáneamente sobre un plano focal en el que se coloca un detector de matricial que registra todos los iones al mismo tiempo.

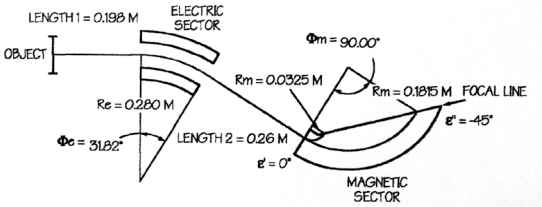


Figura 5. Espectrómetro de masas de doble enfoque de geometría Mattauch-Herzog [Reproducido de la referencia 2 con permiso, John Wiley & Sons, Ltd.:, Copyright (1996)].

Sin embargo este tipo de instrumentos eran de gran tamaño, muy pesados por el gran imán de 180° que utilizaban y por tanto difíciles de construir y manejar hasta que en 1940 otro físico americano Alfred Nier graduado de la Universidad de Minesota diseño un nuevo y revolucionario instrumento (3) basado en un sector de campo magnético de 60° (Figura 6). Como se aprecia en el esquema de la figura 6, el diseño óptico del instrumento es simétrico, por lo que al estar la fuente de iones y el detector equidistantes del imán se minimizaba la influencia del campo magnético y en lugar de las técnicas de detección fotográfica de los anteriores espectrógrafos de masas utilizaba medidas electrónicas de los haces de iones. Este instrumento, por su gran versatilidad tuvo aplicación inmediata en múltiples campos y fue el precursor de los instrumentos de doble enfoque (velocidad y dirección).

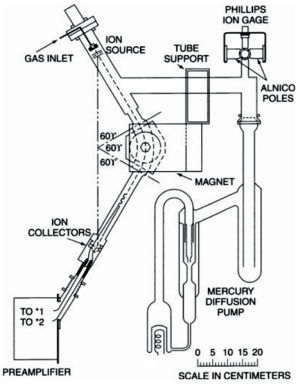


Figura 6. Diagrama esquemático del espectrómetro de masas de sector magnético de 60° diseñado por A Nier. [Reproducido de la referencia 3 con permiso, John Wiley & Sons, Ltd.:, Copyright (2006)].

Llegados a este punto del desarrollo histórico de la Espectrometría de Masas y antes de proseguir conviene detenerse en una descripción de los diferentes componentes de un espectrómetro de masas (sistema de producción de iones, analizador de masas y sistema de detección) y sus principios básicos de operación.

3. Partes de un espectrómetro de masas.

En su esencia más básica un espectrómetro de masas manipula haces definidos de átomos y moléculas mediante la aplicación de campos electromagnéticos por lo que estos átomos o moléculas necesariamente han de ser portadores de una o de varias carga(s) iónicas (positivas o negativas) para poder ser manipulados por estos campos. Como se desprende del análisis de las diferentes partes en el esquema de la Figura 6, en su concepto mas simplificado un espectrómetro de masas se configura alrededor de los componentes necesarios para a) producir un haz de iones a partir de una muestra de un liquido, un sólido o un gas, b) proceder al análisis de la masa individual de estos iones (relación masa/carga o *m/z*) y c) finalmente detectar su presencia y abundancia, como se esquematiza en la Figura 7. Primero es necesario generar un haz de iones a partir de la muestra que se quiere analizar. Los iones en dicho haz han de presentar una dispersión energética y divergencia angular reducida. Para ello se necesita una fuente de iones capaz de transformar en iones con carga positiva o negativa las moléculas neutras de la muestra en estudio. Una vez generados, estos iones tienen que formar un haz que ha de colimarse para reducir su divergencia espacial y ha de acelerarse fuera de la fuente (2). Si la muestra a analizar está en fase gas o es un líquido volátil se utilizan fuentes de impacto de

electrones (EI), ionizacion química (CI) o desorción por campo (FI). Si se trata de sólidos o disoluciones en fase liquida se pueden utilizar las modernas fuentes de electrospray (ESI), ionización química a presión atmosférica (APCI), ionización por láser asistida por matriz (MALDI), ionización por iones secundarios (LSIMS) o incluso las mas recientes técnicas de ionización directa (DESI o DART). Ver capitulo 9 para una descripción de la cronología y detalles de la introducción de estas técnicas de ionización así como su impacto en el desarrollo de la Química de Proteínas. La secuencia temporal de su introducción puede resumirse como se indica en Tabla 1.

Tabla 1. Cronología de la introducción en el tiempo de los diversos métodos de ionización utilizados en Espectrometría de Masas. Se mantienen las abreviaturas comúnmente utilizadas en la terminología inglesa.

Método	Año de introducción
Descarga gaseosa	1898
Ionización en Superficie	1918
Impacto Electrónico (EI)	1929
Ionización Química (CI)	1965
Desorción por Láser (LD)	1968
Ionizacion Electrohidrodinámica (EHD)	1973
Ionización Química por Desorción (DCI)	1974
Ionización por Thermospray (TS)	1980
Ionización por Electrospray (ESI)	1984
MALDI	1985

Una vez generado el haz de iones, compuesto por especies con diferentes relaciones de masa/carga (cada ion viene definido por su masa molecular y la carga o cargas positivas o negativas que soporta; relación *m/z*) este se dirige hacia el analizador de masas que básicamente es un filtro que separa los iones en el espacio o en el tiempo en función de sus respectivas masas mediante una combinación de campos magnéticos, eléctricos o de radiofrecuencia (2). Dependiendo de la energía cinética de cada ion, estos son desviados por efecto de la fuerza del campo en diferentes trayectorias divergentes que inciden sobre una placa fotográfica o electrónica (detección electro óptica) para la detección simultanea de todos los iones o bien, por efecto de barrido del campo eléctrico o magnético, los diferentes valores de *m/z* se van enfocando secuencialmente sobre la rendija de entrada de un detector del tipo de multiplicador de electrones. Este sistema se basa en el principio de amplificación de la señal por emisión secundaria de electrones de modo que el haz de iones procedente del analizador de masas se focaliza sobre un dinodo de conversión que emite electrones en proporción directa a la cantidad de iones que inciden sobre el. El proceso se repite en varias etapas dirigiendo los electrones hacia otros dinodos en cascada y así se amplifica la señal.

Aunque seria deseable poder trabajar con eficiencias de transmisión de iones del 100% en la práctica esto no es posible debido a las pérdidas de iones en su recorrido hacia el detector, por ejemplo por colisiones con otras moléculas de gas y otras especies neutras (radicales) o también con otros iones. Es por ello que el instrumento, desde la fuente de iones (con la excepción de las modernas fuentes que operan a presión atmosférica) hasta el detector debe mantenerse en un estado de alto vacío de entre 10^{-4} a 10^{-7} Pa, como se indica en el esquema de la figura 7. Por otra parte el haz de iones tiene tendencia a abrirse por divergencia o repulsión electrostática entre los iones del mismo signo. Este efecto es mas señalado en haces de iones de baja energía cinética. Por ello otro componente de gran importancia, que a efectos de simplificación se suele omitir de los esquemas como el de la figura 7, son las guías de iones (4, 5) como las guías de radiofrecuencia, las

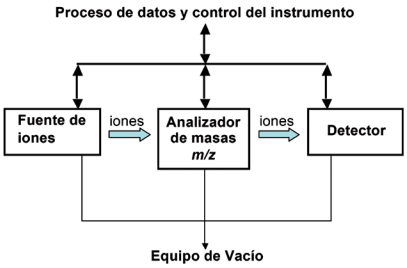


Figura 7. Partes básicas de un espectrómetro de masas.

lentes focalizadoras electrostáticas o los embudos de iones desarrolladas para aumentar la tasa de transferencia de iones. Hoy en día todos estos componentes están bajo control computerizado para la operación del equipo y para la obtención y procesado de los datos (Figura 7).

3.1. Principios básicos de operación.

El principio en el que se basaba la operación de un analizador de masas en los primeros instrumentos diseñados por Aston, Dempster o Nier residía en la separación de todo el haz de iones de diversas masas moleculares, producidos por la fuente de iones, en haces individuales para cada valor de m/z según la ecuación:

$$M/z = er^2 H^2 / 2V$$

Los iones son acelerados al exterior de la fuente de iones por un campo eléctrico *V*. *m* es la masa del ion, *z* es el numero de cargas del ion, *e* es la magnitud de la carga eléctrica y *H* es la fuerza del campo magnético (B). El haz de iones total, al entrar en el campo magnético es desviado de su trayectoria por la fuerza centrípeta ortogonal al campo H en un grado que dependerá de la fuerza centrifuga derivada de su momento (velocidad) y que se plasmara por tanto en un radio de curvatura proporcional a su momento, como indica la siguiente ecuación

$$r = mv / zeH$$

Así el haz de iones total se dispersara en diferentes haces para cada valor de *m* y solo una de las trayectorias, aquella en que ambas fuerzas (centrífuga y centrípeta) están equilibradas, alcanzara el detector. Como el radio de curvatura del instrumento es fijo para obtener un espectro de masas (enfoque secuencial de los diferentes valores de *m/z* sobre el detector) será necesario barrer el campo magnético *B* o bien el campo eléctrico de aceleración *E*.

A partir de aquí y en función del desarrollo de los diversos tipos de espectrómetros de masas en los que el análisis de masas se efectúa mediante otro tipo de campos como por ejemplo los de radiofrecuencia, será necesario valorar las ecuaciones que rigen en cada caso la trayectoria de los iones hacia el detector. Aunque los instrumentos que surgieron después de la introducción de los sectores magnético y eléctrico difieran de estos en su configuración, conceptualmente siempre se trata de generar haces de iones cuanto mas monoenergéticos mejor y que se aceleran hacia un detector. En el camino sus trayectorias axiales se desvían de forma controlada en función de su masa y momento al someterlas a la acción de fuerzas centrípetas o radiales debidas a la aplicación de otros campos.

4. Diferentes tipos de espectrómetros de masas.

Después de la introducción de la idea del doble enfoque por Mattauch y Herzog en 1934 y de la simplificación del sector magnético por Nier en 1940, durante un largo periodo de tiempo la espectrometría de masas se practico en instrumentos de sector magnético o en instrumentos que combinaban sectores magnéticos y eléctricos en diferentes combinaciones. En 1937 y con el incentivo del análisis de hidrocarburos en petróleos y gasolinas Herbert Hubert Jr, fundó en EEUU la Consolidated Engineering Company (CEC) que a principios de la década de los 40 empezó a fabricar el modelo CEC 21-101. La primera unidad comercial de este modelo pionero que utilizaba la geometría de Dempster con un analizador magnético de 180 grados y un peso de dos toneladas pero que era capaz de analizar muestras de hidrocarburos ligeros a una velocidad ocho veces mayor que la de los métodos convencionales de destilación se instaló en

1943 en la Atlantic Refining Company de Filadelfia. En 1955 tras varios años de dominio de los instrumentos de sectores eléctrico y/o magnético, dos investigadores de los laboratorios de la Bendix Aviation Corporation en Detroit publicaron la descripción de un espectrómetro de tiempo de vuelo (TOF-MS) (6,7) que la propia Bendix Corporation comercializó en competencia con los instrumentos de la compañía CEC. Se estima que a mediados de los años 60 CEC había vendido unas 700 unidades y Bendix unas 300.

A continuación se presenta una breve relación de los conceptos y geometrías básicas de los diferentes tipos de espectrómetros de masas ya que en el capitulo 9 se describe en detalle la cronología de su introducción en el mercado y de sus mejoras operativas así como sus aplicaciones en Proteómica. En la Tabla 2 se especifican los diferentes modelos, año de introducción y características técnicas de rango de masas, exactitud y resolución para los instrumentos de diversas casas comerciales a partir de 1981. Hasta este año el campo de la espectrometría de masas había estado dominado por los instrumentos magnéticos de enfoque simple o doble y por los primeros cuadrupolos de transmisión desarrollados por Finnigan (ver capitulo 9).

4.1. Instrumentos de doble enfoque

Como ya se ha indicado los instrumentos de doble enfoque constan de un sector eléctrico (E) para la selección de haces monoenergéticos y de un sector magnético (B) para el análisis del momento de los iones en cada haz preseleccionado (2). Mediante esta geometría es posible obtener suficiente resolución como para separar especies iónicas con la misma masa nominal (isobáricas) y a la vez determinar la masa exacta de los iones.

En este tipo de instrumentos el sector eléctrico puede preceder al sector magnético (geometría convencional) o al revés (geometría reversa). El primer ejemplo de un instrumento de geometría normal es el de Mattauch-Herzog, antes citado, el cual separa y enfoca todos los iones sobre un plano focal de modo que todos ellos cumplen simultáneamente las condiciones del doble enfoque (Figura 5).

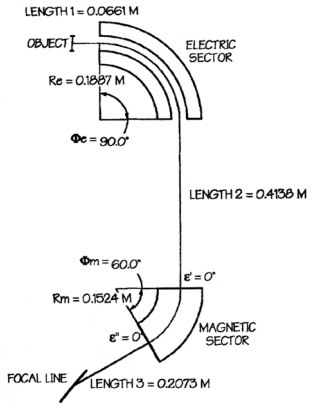


Figura 8. Esquema de un espectrómetro de masas de geometría Nier - Johnson [Reproducido de la referencia 2 con permiso, John Wiley & Sons, Ltd.:, Copyright (1996)].

Otro instrumento bien conocido es el basado en el diseño de Nier-Johnson realizado en 1952 y considerado como un espectrómetro de punto focal ya que en este caso los iones seleccionados se focalizan sobre un punto en lugar de sobre un plano focal (Figura 8).

4.2. Tiempo de vuelo

Los instrumentos de tiempo de vuelo lineal miden el tiempo necesario para que un ion viaje desde la fuente de iones al detector situado a 1 – 2 m de distancia (Figura 9). En la fuente de iones, a todos los iones se les imparte la misma energía cinética durante su aceleración instantánea. Sin embargo, como tienen masas (valores de *m/z*) diferentes viajan a velocidades diferentes según su masa y por lo tanto cada especie iónica alcanza el detector como un paquete de iones separado de los demás en función de su masa. Los de menor masa llegan primero seguidos de los más pesados. La ventaja inherente de este sistema es que no tiene limite de masa pero para reducir al mínimo la posibilidad de las colisiones de los iones con especies neutras en su largo recorrido hasta el detector hay que mantener el valor de la distancia

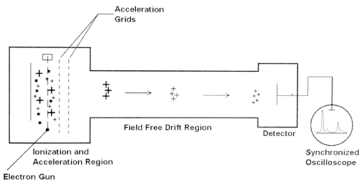


Figura 9. Espectrómetro de masas de Tiempo de Vuelo (Time-of-Flight, TOF) [Reproducido con el permiso de John Wiley & Sons, Ltd.:, Copyright (2008) Ver Introducción].

media ínter-colisión igual o mayor que 1 metro. Para ello se trabaja a un vacío mas elevado que en otros tipos de espectrómetros de masas. Sin embargo desde sus inicios estos equipos proporcionaban una resolución limitada. Por ello, la aparición en la década de los 60 de los instrumentos de tipo cuadrupolar, por su mayor resolución y escala lineal de *m/z*, represento el ocaso del Tiempo de Vuelo hasta que la aparición del MALDI (ver capitulo 9) lo rescató del olvido y lo convirtió en el instrumento mas práctico para el uso de estas técnicas de iotización.

La resolución de un instrumento de Tiempo de Vuelo Lineal depende de la longitud del tubo, del voltaje de aceleración de los iones y de las distribuciones espaciales y de velocidad del paquete de iones inicial (8). Aunque el voltaje del pulso de extracción y aceleración fuera de la fuente se aplica para que los iones reciban exactamente la misma energía cinética y por lo tanto produzcan un paquete inicial de iones mono- energético, en la practica se produce una dispersión espacial y temporal, por un lado como consecuencia de las pequeñas diferencias en las posiciones de partida de los iones y por otro en los tiempos de vuelo por las pequeñas diferencias energéticas de los iones de la misma masa. Todo ello hace que el paquete inicial de iones no tenga una anchura infinitesimal y que no sea mono-energético con la consiguiente perdida de resolución. Para solucionar esta problemática del TOF se utilizan espejos de iones que reflejan los iones del mismo valor de m/z pero de forma que se compensan sus pequeñas diferencias de energía cinética (9). Estos sistemas reciben el nombre de espectrómetros de masas tipo reflectron TOF (*re*TOF) provistos de un campo eléctrico que hace las funciones de un espejo de iones. El principio de operación se ilustra esquemáticamente en la figura 10. Como puede verse en el esquema, el espejo compensa las diferencias en la distribución de energías cinéticas y velocidades para los tres iones de la misma masa que de no hacer nada no llegarían al detector en el mismo instante (perdida de resolución). Este efecto de compensación se produce porque aunque los tres iones indicados salen del espejo con las mismas diferencias de velocidad y energías cinéticas que cuando entraron, no obstante se han modificado las longitudes de sus recorridos por la diferente penetración en el campo eléctrico del espejo.

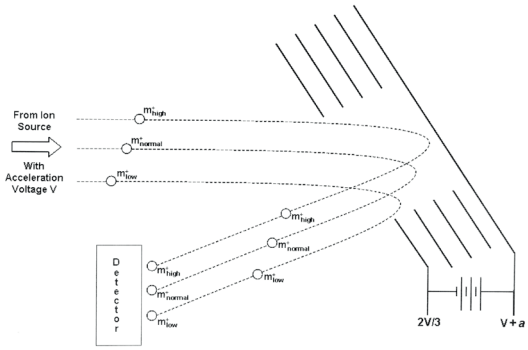


Figura 10. Esquema de un reflectron. [Reproducido con el permiso de John Wiley & Sons, Ltd.:, Copyright (2007) Ver Introducción].

4.3. Cuadrupolos de transmisión

Este tipo de analizador de masas deriva de la tecnología desarrollada por Wolfgang Paul a principios de la década de los 50 en la Universidad de Bonn, quien en 1989 compartió el premio Nóbel de física con Georg Dehmelt por "el desarrollo de las técnicas de la trampa de iones (10). Este tipo de analizador utiliza como filtro de masas una combinación de campos eléctricos de corriente directa (DC) y radiofrecuencia (RF) de amplitud variable. Estos campos se aplican sobre cuatro superficies paralelas opuestas, de sección hiperbólica y conectadas entre si como se ilustra en la figura 11. Los iones extraídos de la fuente se aceleran con un potencial de entre 5 y 15 voltios hacia el espacio central entre las superficies cuadrupolares, de modo que las relaciones de m/z seleccionadas viajan a lo largo del eje longitudinal hacia el detector. En este caso el campo eléctrico que mueve los iones es bidimensional y a este tipo de instrumentos se les conoce como filtros de masas cuadrupolares. En el capitulo 9 se describen sus orígenes y desarrollo. El espectro de masas se obtiene incrementando o disminuyendo la magnitud de los potenciales DC y de la amplitud del campo RF en una relación fija entre ambos campos que define la resolución del instrumento. Todo el espectro de masas se obtiene barriendo los potenciales a relación constante de DC/RF. Este tipo de instrumento representó con su introducción

comercial, a mediados de la década de los años 60, el principio de una nueva era de popularización de la técnica de la espectrometría de masas y de declive de los instrumentos magnéticos y de tiempo de vuelo (ver Capitulo 9), aunque, como ya se ha indicado, estos últimos en cuestión de unos 20 años volverían a adquirir protagonismo gracias a la introducción de las nuevas técnicas de ionización por MALDI.

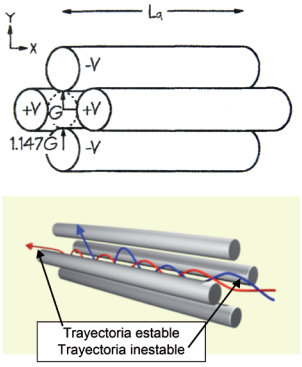


Figura 11. Quadrupolo de transmision

El quadrupolo de transmisión es la base de los instrumentos diseñados para espectrometría de masas en tándem (MS/MS) en su configuración de triple quadrupolo. Estos instrumentos fueron comercializados por primera vez en 1981 por Sciex Inc. (Ontario, Canada) y por Finnigan Corporation (Sunnyvale, California (ver capitulo 9). Como su nombre indica los cuadrupolos están montados en tándem de modo que el primero (Q1) y el tercero (Q3) tienen la capacidad de seleccionan masas como lo hacen los cuadrupolos simples mientras que el segundo en línea (q2) se opera solo en el modo RF y sirve como célula de colisión. De este modo el Ion precursor seleccionado en Q1 se transmite a q2 en donde se somete a procesos de colisión y los iones producto resultantes se analizan en Q3 (11). Según el tipo de la configuración de estos instrumentos los experimentos de espectrometría de masas en tándem pueden realizarse en el espacio o en el tiempo. En la configuración de tándem en el espacio los distintos elementos de selección de precursores, colisión y análisis de iones producto están físicamente separados, como por ejemplo en los instrumentos de sectores, de cuadrupolos de transmisión o de tiempo de vuelo. Por el contrario, los experimentos de tándem en el tiempo implican que los iones se encuentran confinados en un solo espacio, realizándose los diferentes pasos de separación en el tiempo pero dentro de un mismo recinto. Este es el caso de las trampas de iones o de los instrumentos de FTMS.

4.4. Trampas de iones

El espectrómetro cuadrupolar de trampa de iones (3D QIT) utiliza un campo de cuadrupolo tridimensional para almacenar iones en orbitas concéntricas tridimensionales de modo que el sistema almacena todos iones a todos los valores de *m/z*. En sus primeras versiones, a diferencia de todos los demás diseños en los que la fuente de iones esta físicamente separada del analizador de masas, los iones se generaban y almacenaban en la propia trampa (equipo de ionizacion interna). Como se ilustra en la Figura 12 este analizador esta construido a partir de dos electrodos tapón (end cap) eléctricamente aislados de ambos lados de un electrodo anular. Para ionizar la muestra se aplica un campo de RF fijo al electrodo anular a fin de acelerar un haz de electrones a 70 eV hacia el interior de la trampa. Todos los iones formados, irrespectivamente de su valor de m/z por encima de un limite inferior predeterminado, se acumulan en orbitas tridimensionales (una para cada valor de m/z) que posteriormente, mediante la aplicación de una variación de la amplitud de RF se desestabilizan secuencialmente y son atraídos hacia los dos electrodos tapón (barrido del espectro de masas), uno de los cuales tiene un orificio para permitir la salida de los iones hacia el detector. Este barrido se efectúa según la técnica del método de eyección selectiva de masas o el método de inestabilidad selectiva de masas (10). Una de las características diferenciales de la trampa de iones es que opera a presiones mucho mas elevadas que los demás equipos (unos 10^{-1} PA frente a los 10^{-4} PA

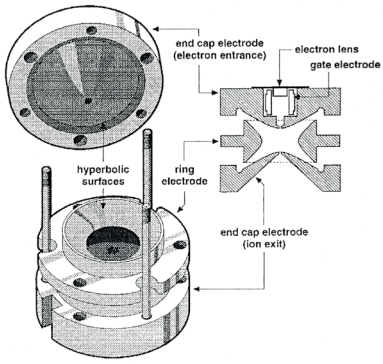


Figura 12. Trampa de iones. [Reproducido con el permiso de John Wiley & Sons, Ltd.:, Copyright (2007) Ver Introducción].

del cuadrupolo de transmisión o los 10^{-7} del TOF. Ello se debe a la necesidad de utilizar un gas tampón para reducir las energías de vibración y rotación de los iones y atenuar su movimiento y posibilidades de colisión.

La primera comercialización de una Trampa de iones corrió a cargo de Finnigan Corporation que en 1983 presentó en la Conferencia de Pittsburg un equipo de cromatografía de gases-espectrometría de masas comercializado como el ITD 700 (ver Capitulo 9). Uno de los problemas fundamentales que tuvieron que afrontar los primeros usuarios de este tipo de espectrómetro es el efecto de carga en el espacio que limita el numero máximo de iones que pueden alojarse en un espacio finito sin excesivas repulsiones entre si. Por ello Finnigan Corporation introdujo primero el AGC o control automático de ganancia que ajusta el periodo de ionizacion para no sobrepasar el número máximo de iones tolerado por la trampa y posteriormente la técnica de modulación axial para aumentar el número de iones en la trampa. Ello le permitió a Finnigan concentrarse en el desarrollo de equipos 3D QIT de segunda generación con sistemas de ionizacion externa como se ilustra en la Figura 13 y necesarios para el acoplamiento a la cromatografía de líquidos.

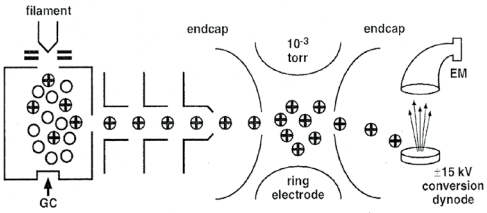


Figura 13. Esquema de una Trampa de iones (3D QIT) de segunda generación con sistema de ionización externa. [Reproducido con el permiso de John Wiley & Sons, Ltd.:, Copyright (2007) Ver Introducción].

Finalmente es importante citar que estos equipos de trampa de iones son únicos en su capacidad para trabajar en la modalidad de Tandem en el tiempo en contraposición a los equipos a base de analizadores múltiples para Tandem en el espacio (ver capitulo 9).

La Trampa lineal de iones es una reciente adición de una importante variación en este tipo de instrumentos. Fue introducida comercialmente por primera vez en PittCon 2002 por Applied Biosystems con el nombre de QTrap 2000. La Trampa Lineal de Iones (LIT) (12) conocida también como trampa de iones bidimensional (2D QIT) es, aparte del Orbitrap y del Synapt, uno de los analizadores de m/z mas nuevos en el mercado. Consta de la misma estructura que un cuadrupolo de transmisión pero en lugar de filtrar los iones de un haz no resuelto a todos los valores de m/z, excepto los de m/z seleccionados, se utiliza para capturar iones, modificar sus trayectorias y efectuar una eyección selectiva de la masa seleccionada. Los sistemas LIT proporcionan eficiencias de inyección y de almacenamiento superiores a los de las trampas convencionales. El LTQ comercial de Thermo Scientific utiliza la eyección radial en lugar de axial (13) para obtener un espectro de masas y la trampa lineal también se ha utilizado como Q3 en un triple cuadrupolo con eyección axial de iones.

4.5. Orbitrap

El Orbitrap es un nuevo analizador de masas desarrollado por Alexander Makarov (14) que Thermo Electron introdujo en 2005 en la reunión de la ASMS de San Antonio y que comercializa como parte de un espectrómetro de Tandem en el espacio (ver Figura 14). Básicamente el analizador consiste en dos electrodos coaxiales asimétricos; un electrodo externo en forma de barril y un electrodo coaxial interno en forma de huso. Los iones se inyectan tangencialmente en un campo eléctrico entre los electrodos y quedan atrapados porque su atracción electrostática hacia el electrodo interior es balanceada por fuerzas centrífugas. Así los iones dan vueltas alrededor del electrodo central formando paquetes de anillos. Adicionalmente, los paquetes de iones también se mueven hacia delante y hacia atrás a través del eje del electrodo central. Así, los iones con una relación específica de masa-carga se mueven en anillos que oscilan alrededor del huso central. La frecuencia de estas oscilaciones armónicas es independiente de la velocidad del ion y es inversamente proporcional a la raíz cuadrada de la relación masa-carga.

Como puede apreciarse en el esquema de la Figura 14 los iones se almacenan en la trampa lineal (LIT) y tras su eyección axial de la misma se atrapan en un quadrupolo curvado de RF en forma de letra C en donde se enfrían y confinan en un menor volumen y mediante un voltaje pulsado se inyectan en el Orbitrap en donde se acumulan como nubes en forma de anillos y se detectan. En estas condiciones el Orbitrap es un potente analizador de masas con resoluciones del orden de 1.500.000 y un rango de masas de 6000 a un nivel exactitud de 2-5 ppm.

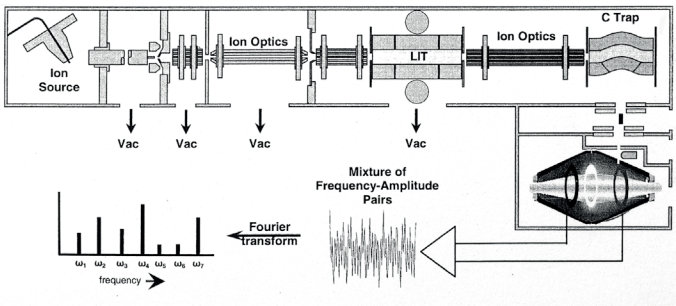


Figura 14. Orbitrap. [Reproducido con el permiso de John Wiley & Sons, Ltd.:, Copyright (2008) Ver Introducción].

4.6. FTICR-MS

La espectrometría de masas de transformada de Fourier se basa en la técnica original de la espectrometría de masas de resonancia de ciclotrón en la que una partícula con carga eléctrica tendrá un movimiento de precesion en un campo magnético a una frecuencia relacionada con su valor de m/z (15). En la condición de resonancia se puede transferir la energía disponible a la frecuencia del ciclotrón al ion oscilante en el campo magnético. En la célula del ICR el campo magnético se orienta desde el panel frontal a través del panel trasero de la célula cúbica. En ICRMS cuando se transmite energía a una frecuencia especifica al interior de la célula de ICR, un ion de una masa determinada absorberá esta energía por la condición de resonancia y pasara a una orbita de radio creciente, al tiempo que mantiene su frecuencia de precesion característica. Cuando este radio supera las dimensiones de la célula los iones colisionan con las paredes provocando una señal eléctrica que puede medirse y cuya potencia esta relacionada con la abundancia de estos iones. Cuando operamos la célula de ICR en el modo de FT se transmiten al ICR una amplia variedad de frecuencias por lo que los iones de diferentes valores de m/z pueden absorber energía al mismo tiempo pero al ser la irradiación del ICR muy breve, los iones no alcanzan una orbita que supere las dimensiones de la célula aunque si entran en fase coherente y oscilan como un grupo en fase con el campo de excitación, por lo que inducen una carga oscilante sobre las paredes del ICR. Esta frecuencia de la oscilación inducida sobre la pared del ICR puede detectarse y amplificarse para relacionarla con el valor de m/z de cada ion en precesion. El enfoque de FTICRMS permite la detección simultánea de todos los iones con una elevada sensibilidad y resolución superior a 1.500.000.

4.7. Espectrómetro de movilidad iónica

Aunque el espectrómetro de movilidad iónica (IMS) no es propiamente un espectrómetro de masas se ha afianzado como un complemento de la espectrometría de masas. En la Figura 15 se ilustra el esquema conceptual de este instrumento, que aunque lo parezca no tiene nada que ver con un instrumento de tiempo de vuelo. Mientras que en este ultimo los iones se desplazan a través de un tubo libre de la interacción de cualquier campo en el IMS y sometidos a un alto vacío para evitar las colisiones con moléculas neutras, en el IMS los iones están sometidos a la acción de un campo eléctrico y se mueven a través de un gas de colisión a presión atmosférica que fluye en dirección contraria a la del avance de los iones. En el IMS los iones se separan en paquetes definidos por su tamaño o área cros-seccional y conceptualmente el proceso podría

visualizarse como una electroforesis en fase gas. A nivel comercial la empresa Waters Corporation incorpora el IMS en su nuevo SYNAPT HDMS (Tabla 2), un hibrido de cuadrupolo-IMS-TOF.

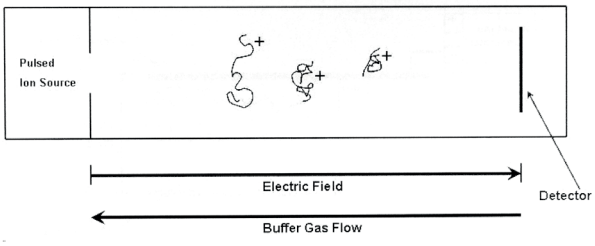


Figura 15. Espectrómetro de movilidad iónica (IMS). [Reproducido con el permiso de John Wiley & Sons, Ltd.: ver Ref. [1], Copyright (2007) Ver Introducción].

Tabla 2. Resumen cronológico y técnico del desarrollo instrumental a cargo de diversos fabricantes de instrumentación para espectrometría de masas. La información contenida en estas tablas ha sido facilitada por los representantes de cada una de estas empresas comerciales, como se indica en los agradecimientos al final del capitulo y ha sido reproducida tal como se recibió.

a) Instrumentación de Agilent

Espectrómetro de Masas	Modelo	Año de introducción	Rango de Masas	Exactitud de Masas	Resolución
Cuadrupolo simple	1100 serie LC/MD	1997	50-3000	+/- 0.13 amu	
Trampa de Iones	1100 serie LC/MSD	Marzo, 2000	50-2200		
	Trap		200-4000		
TOF	LC/MSD TOF	June, 2003	15-3000	3 ppm	10,000 FWHM @1,000
Cuadrupolo Triple	6410	Jan, 2006	15-1650		
QTOF	6510	Jan, 2006	20-12000	MS 2 ppm RMS MS/MS 5 ppm RMS sobre m/z 397	>13,000

b) Instrumentación de Applied Biosystems

Espectrómetro de Masas	Modelo	Año de introducción	Rango de Masas	Exactitud de Masas	Resolución
Cuadrupolo Triple	API 2000	1998	5-1800	0.1 amu	Unit
	API 3000	1998	5-3000	0.1 amu	Unit
	API 3200	2005	5-1800	0.1 amu	Unit
	API 4000	2001	5-3000	0.1 amu	Unit
	API 5000	2005	5-1250	0.1 amu	Unit
	FlashQuant WorkStation (API4000/ 4000QTRAP)	2007	5-3000/ 5- 2800	0.1 amu	Unit/ 7,000
MALDI-TOF	Voyager DE	No info			
	Voyager DE PRO	1999	Linear 100-350,000 Da Reflector 100-4000 Da	50 ppm	Linear 2,500 Reflector 12,000
	Voyager DE-STR		Linear 100-350,000 Da Reflector 100-4000 Da	20 ppm	Linear 3,500 Reflector 14,000
Trampas Híbridas	QTRAP 3200	2005	5 – 1700	0.1 amu	7,000
	QTRAP 4000	2003	5 – 2800	0.1 amu	7,000
Hybrid Quad-TOF	QSTAR XL	2002	5 – 40,000	5 ppm	10,000
	QSTAR Elite	2006	5 – 40,000	2 ppm	15,000
MALDI TOF-TOF	4700 MALDI	2002	Linear 100-150,000 Da Reflector 100-7000 Da MSMS 100-4000 Da	5 ppm	Linear 7,000 Reflector 20,000 MSMS 5,000
MALDI TOF-TOF	4800 MALDI	2005	Linear 100-500,000 Da Reflector 100-10000 Da MSMS 100-4000 Da	5 ppm	Linear 7,000 Reflector 25,000 MSMS 6,000

c) Instrumentación de Bruker Daltonics

Espectrómetro de Masas	Modelo	Año de introducción	Rango de Masas	Exactitud de Masas	Resolución
MALDI-TOF	Reflex	1991-1998			
	Biflex	1995-2000			
	Autoflex	2000			
	Microfelx	2003	0 - 300.000	< 15ppm	>15,000
	Autoflex III SmartBeam	2007	0 - 300.000	< 5ppm	> 20,000
	Ultraflex III SmartBeam	2007	0 - 300.000	< 3ppm	>25,000
Trampas de iones	Esquire-LC	1997-1999			
	Esquire 3000	1999-2002			
	Esquire 6000	2002	50-4000u	± 0.2u	≤ 10,000
	HCT	2003	50-4000u	± 0.15u	≤ 12,000
	HCTultra	2004	50-4000u	± 0.15u	≤ 12,000
	HCT Ultra ETD II	2008	50-4000u	± 0.15u	≤ 12,000
Hybrid Quad-TOF	Micro TOF-Q II	2008	50-20,000u	< 2ppm	17,500
	Maxis	2008	50-20,000u	<1ppm	>40,000
ESI-TOF	MicroOTOF	2003	50-40,000u	< 2ppm	16,500
TOF-TOF	Ultrafelx	2004			
	autoflex III SmartBeam	2007	0 - 300.000	< 5ppm	> 20,000
	ultraflex III SmartBeam	2007	0 - 300.000	< 3ppm	> 25,000
FT-MS	BioApex	1995-2003			
	apex-Q	2003-2007			
	apex ultra	2007	50-10,000	< 0.3ppm	>2,000,000

d) Instrumentación de Thermo Fisher Scientific

Espectrómetro de Masas	Modelo	Año de introducción	Rango de Masas	Exactitud de Masas	Resolución
Cuadrupolos triples	Finnigan TSQ 45	1981	800		Unit
	Finnigan TSQ 70	1896	4000		Unit
	Finnigan TSQ 700	1990	4000		Unit
	Finnigan TSQ 7000	1994	4000		Unit
	TSQ Quantum	2000	1500	5ppm	0.1 Da
	TSQ Quantum Discovery	2003	1500	0.1 amu	0.1 Da
	TSQ Quantum Ultra	2003	3000	5ppm	0.1 Da
	TSQ Vantage	2008	3000	5ppm	0.1 Da
Trampas de Iones	Finnigan ITD	1983	650	± 0.1 amu	Unit
	Finnigan ITMS	1985	650	± 0.1 amu	Unit
	Finnigan LCQ	1996	2000	± 0.1 amu	Unit
	LCQ Duo	1997	2000	± 0.1 amu	Unit
	LCQ Deca	1998	2000 or 4000	± 0.1 amu	Unit - zona scan
	LCQ Deca XP Plus	2002	15-200,50-2000,100-4000,1000-20000		
	LCQ Advantage MAX	2003	15-200,50-2000,100-4000,1000-20000		
	LCQ Deca XP MAX	2003	15-200 and 50-2000	± 0.1 amu	Unit
	LCQ Advantage	2001	15-200 and 50-2000	± 0.1 u	Unit
	LCQ Deca XP	2001	2000 or 4000	± 0.1 amu	Unit + zs + ultra zs
	LTQ	2003	2000 or 4000	± 0.1 amu	Unit + zs + ultra zs
	LXQ	2005	2000 or 4000	± 0.1 amu	Unit + zs + ultra zs
	LTQ XL	2006	2000 or 4000	± 0.1 amu	Unit + zs + ultra zs
	LCQ Fleet	2007	2000 or 4000	± 0.1 amu	Unit + zs
	LTQ FT	2004			
	LTQ Orbitrap	2005			

e) Instrumentación de Waters Corporation

Espectrómetro de Masas	Modelo	20Año de introducción	Rango de Masas	Exactitud de Masas	Resolución
Cuadruplos simples	Platform	1993	3200	nominal	unit
	Platform II	1993	3200/4000	nominal	unit
	Platform LC	1996	1600	nominal	unit
	ZMD	1999		nominal	unit
	EMD 1000	2003	1000	nominal	unit
	ZQ	2000	2000/4000	nominal	unit
	AQUITY SQD	2006	2000	nominal	unit
	3100	2006	2000	nominal	unit
Cuadruplos triples	Bio Q	1990	4000	nominal	unit
	Quatro	1989	4000	nominal	unit
	Quatro II	1994	4000	nominal	unit
	Quatro LC	1997	1600/3200	nominal	unit
	Quatro Premier	2003	3000	nominal	unit
	Quatro Premier XE	2005	3000	nominal	unit
	Quatro Ultima	1999	3000	nominal	unit
	Quatro Micro API	2001	2000	nominal	unit
	Quatro Micro GC	2002	1500	nominal	unit
	AQUITY TQD	2006	2000	nominal	unit
	Xevo TQ MS	2008	2000	nominal	unit
MALDI-TOF	TQF Spec	1992	60K	0,02%	1500
	MALDI Micro	2000	100K	10ppm	10000
	Q-Tod Ultima MALDI	2001	100K	5ppm	17500
	Q-tof Global	2001	100K	5ppm	17500
	MALDI SYNAPT MS	2008	100K	3ppm	13000
	MALDI SYNAPT HDMS	2007	100K	3ppm	13000
Quad-IMS-TOF	SYNAPT HDMS	2006	100K	2ppm	17500
TOF	LCT Premier XE	2006		3ppm	12000
	LCT	1997		5ppm	500
Hybrid TOF	Q-TOF	1996	100K	5ppm	5000
	Q-TOF II	1999	100K	5ppm	10000
	Q-TOF Micro	2001	100K	5ppm	5000
	Q-TOF Ultima	2001	100K	5ppm	17500
	Q-TOF Premier	2004	100K	3ppm	17500
	SYNAPT MS	2008	100K	2ppm	17500
Sector	AutoSpec Premier	2004	2300	1,5mDa	80000
	AutoSpec Ultima	1990	2300	1,5mDa	80000

f) Instrumentación de Shimadzu / Kratos Analytical

Espectrómetro de Masas	Modelo	20Año de introducción	Rango de Masas	Exactitud de Masas	Resolución
MALDI-TOF	Kompact MALDI II	1992	1-500,000		
	Kompact MALDI III	1992	1-500,000		
	Kompact MALDI I	1994	1-500,000		
	Kompact MALDI IV	1995	1-500,000		
	Kompact Alpha	1997	1-500,000		
	Kompact Discovery	1997	1-500,000		
	Kompact Probe	1998	1-500,000		
	Kompact SEQ	1998	1-500,000		
	AXIMA LNR	2000	1-500,000	<50ppm	>4,000
	AXIMA CFR	2000	1-500,000	<20ppm	>10,000
	AXIMA CFRplus	2002	1-500,000	<20ppm	>20,000
	AXIMA QIT (ion trap-TOF)	2002	50-12,000	<5ppm	>6,500
	AXIMA Assurance	2008	1-500,000	<30ppm	>5,000
	AXIMA Confidence	2008	1-500,000	<10ppm	>15,000
	AXIMA Perfomance	2008	1-500,000	<5ppm	>20,000
Trampa de iones	TOF LC-IT-TOF	2005	50-5,000	<3ppm	>10,000
Cuadruplos simples	LCMS-2010EV	2007	10-2,000		
	LCMS-2020	2008	10-2,000		

4.8 Instrumentos Híbridos para espectrometría de masas en Tandem o MS/MS

Los espectrómetros de masas híbridos son instrumentos idóneos para la espectrometría de masas en Tandem (Figura 16) y se basan en la combinación de dos o más analizadores de masas diferentes y complementarios. En la operación combinada de dos analizadores de masas, se selecciona en el primer analizador (MS1) un determinado ion según su masa, a continuación se provoca su fragmentación adicional en una cámara de reacción y finalmente en la segunda etapa del proceso los iones producto de la reacción se analizan en el segundo analizador (MS2) (16). En la Tabla 3 se relacionan los diferentes tipos de instrumentos y sus prestaciones según actúen como MS1 o MS2. Todas las combinaciones son posibles y tienen una traducción comercial como se ilustra en la tabla 2, excepto para el sector eléctrico y el orbitrap que no han sido utilizados como primer sector de un Tandem. Todos estos instrumentos miden la relación m/z de un ion bien por su dispersión según su momento o energía cinética (instrumentos de sectores eléctricos y magnéticos), por su separación en el tiempo según su velocidad (instrumentos de tiempo de vuelo), por su transmisión a través de un campo elctrohydrodinamico (filtros de masas cuadrupolares) o por su movimiento periódico en un campo magnético o hidrodinámico (trampas de iones). Los equipos híbridos utilizan varias combinaciones de estos analizadores para construir un espectrómetro de masas capaz de operar en la modalidad de MS/MS (ver capitulo 9).

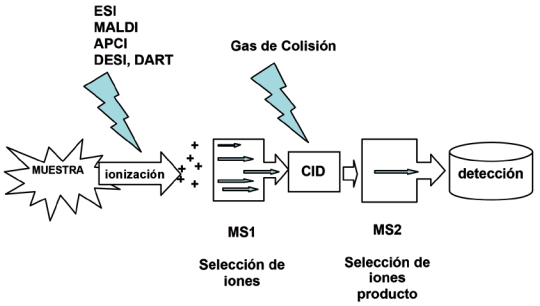


Figura 16. Experimento de MS/MS

Los primeros instrumentos híbridos utilizaban la tecnología existente en la década de los años 80 que se basaba en los instrumentos de sector magnético y cuadrupolar en configuraciones, por ejemplo de BqQ, BEqQ, QqQ, EBqQ, QEB, B/TOF, QqTOF. A este tipo de configuraciones le siguieron las que incorporaron la trampa de iones o el ICR como el Q/IT o el Q/FTICR. Después vinieron los instrumentos en los que la trampa se situaba en la posición de MS1 como en el QIT/TOF o LIT/TOF. Finalmente las combinaciones mas recientes se basan en el uso de dos trampas a fin de conseguir mejor exactitud de medidas de masas y poder de resolución. En estos casos el MS2 es un FTICR o un orbitrap (LIT/FT-ICR o LIT/orbitrap (Ver tabla 3 y referencia 16).

REFERENCIAS.

1. Dempster AJ. A new method of positive ray analysis. *Phys. Rev.*1918; 11: 316.

2. Burgoyne TW, Hiefte GM, An introduction to Ion optics for the mass spectrograph, Mass Spectrom. Rev. 1996;15: 241-259

3. De Laeter J, Kurz MD, Alferd Nier and the sector field mass spectrometer, J.Mass Spectrom. 2006; 41: 847-854

4. Guan S, Marshall AG. Stacked-Ring Electrostatic Ion Guide. J. Am. Soc. Mass Spectrom. 1996, 7: 101-106

5. Berkout VD, Doroshenko VM. Improving the quality of the ion beam exiting a quadrupole ion guide. J. Am. Soc. Mass Spectrom. 2006, 17: 335-340

6. Wiley WC, McLaren IH, TOF-MS. Rev. Sci. Instrum. 1955; 26: 1150-1157

7. Karas M, Historical Comment on reprint of “Time of Flight Mass Spectrometer with Improved Resolution. J. Mass Spectrometry 1997; 32: 1-3

8. Guilhaus M. Principles and Instrumentation for TOF-MS. J. Mass Spectrom. 1995; 30: 1519-1532.

9. Mamyrin BA, Reflectron TOF-MS. Int. J. Mass Spectrom. Ion Processes 1994;131: 1-19

10. March R, Todd J. Quadrupole Ion Trap Mass Spectrometry, 2nd Ed. Wiley- Interscience: New York, 2005.

11. Yost RE, Fetterolf DD, MS/MS Instrumentation. Mass Spectrom. Rev. 2: 1-46, 1963

12. Douglas D, Frank A, Mao D, Linear Ion Traps in MS. Mass spectrum. Rev. 2005; 24: 1-29.

13. Londry FA, Hager JW, Mass selective axial ion ejection from a linear quadrupole ion trap. J. Am. Soc. Mass Spectrom. 2003; 14: 1130-1147.

14. Makarov A. Electrostatic axially harmonic orbital trapping: a high performance technique of mass analysis. Anal. Chem. 2000; 72: 1156-1162

15. Marshall AG, Hendrickson CL, Jackson GS, FT-ICR MS: a primer. Mass Spectrom. Rev. 1998; 17: 1-35

16. Glish GL, Burinsky DJ, Hybrid Mass Spectrometers for Tandem Mass Spectrometry, J. Am. Soc. Mass Spectrom. 2008; 19: 161-172

Métodos y mecanismos de ionización.

Desarrollo cronológico y conceptual de la espectrometría de masas: de la química de proteínas a la proteómica

Emilio Gelpí

Instituto de Investigaciones Biomédicas de Barcelona (IIBB-CSIC-IDIBAPS), Laboratorio de Proteómica CSIC/UAB. Facultad de Medicina (UAB).

Como el titulo del capitulo indica se exponen aquí, de forma estrictamente cronológica, los continuos avances de la espectrometría de masas (EM), centrando esta descripción en el desarrollo de los diferentes conceptos metodológicos e instrumentales que eventualmente han popularizado el estudio de péptidos y proteínas por EM. La versión en Ingles de esta información se encuentra en dos publicaciones recientes en el Journal of Mass Spectrometry (1,2). En concreto se trata de visualizar la secuencia de conceptos y eventos que, desde los primeros trabajos del grupo de Klaus Biemann en el MIT en 1959 hasta los principios del nuevo siglo han colocado a la EM en una posición claramente dominante y ventajosa respecto a las técnicas clásicas de secuenciación por métodos químicos automatizados, como el de Edman (3).

1. Impacto de Electrones (IE). Las primeras incursiones de la EM en la Química de Proteínas.

Dos acontecimientos importantes marcaron en 1965 el nacimiento de la química bioanalítica en el campo del estudio y determinación de compuestos biológicos por EM. Ambos podrían considerarse como los inicios de la tremenda revolución, de curso insospechado en aquellos años, en relación a la determinación directa por EM de macromoléculas biológicas, como los péptidos y las proteínas.

El primero de estos eventos fue la llegada al mercado del LKB 9000 (Figura 1), el primer equipo comercial de Cromatografía de Gases (CG) acoplada a un Espectrómetro de Masas (EM) de sector magnético (CG-EM) con una fuente de ionización de IE y una interfaz de jet de "separación molecular". El control del instrumento y el manejo de datos eran totalmente manuales, pero el nuevo equipo fue un éxito inmediato en un sin fin de aplicaciones en muestras biológicas. La sencillez y la eficiencia de esta interfaz estimularon el desarrollo de otras interfaces, pero ninguna fue tan eficiente (4).

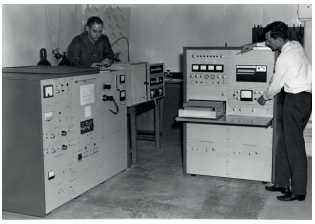


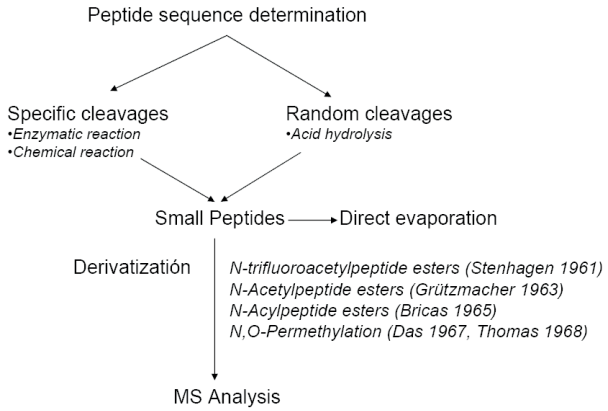
Figura 1. Darryl Nooner y Emilio Gelpí con el LKB 9000 en el Departamento de Química de la Universidad de Houston en 1966.

El otro fue la identificación por EM de la Fortuitina y la Peptidolipina NA por Edgar Lederer en el CNRS en Francia con la ayuda de Mike Barber (AEI, Manchester). La Fortuitina es un peptidolipido natural (éster metílico del eicosanoil nonapéptido) aislado del *Mycobacterium fortuitum* y que contiene aproximadamente cantidades equimoleculares de ácidos eicosánico y docosanoico, unidos a un oligopéptido de 9 aminoácidos. Esta fue la primera determinación por EM de una secuencia de AAs parcialmente desconocida con un Mr de 1359 (5). La muestra (menos de 10 µg) se introdujo en un espectrómetro de masas MS9 de doble enfoque mediante la técnica de evaporación directa a 200 º C, descrita por Mike Barber. Esto fue posible porque el péptidolípido estaba acilado de modo natural con ácidos grasos de cadena larga, O-acetilado en sus dos residuos de treonina y, además metilado en su carboxilo terminal, siendo por tanto suficientemente volátil para evaporación directa. Los mismos grupos confirmaron la estructura de la peptidolipina NA, heptapéptido macrocíclico natural N-acilado de Mr 963 (6). Estas identificaciones son un ejemplo de la utilización de la técnica de evaporación directa sin derivatización de la muestra (Figura 2). Por tanto, podría considerarse que ambos eventos marcaron el final de la dominación de la química del petróleo en la EM y el comienzo de la EM biomolecular y bioorgánica y su aplicación al análisis de trazas de péptidos y proteínas.

Los fundamentos de la secuenciación de péptidos y proteínas mediante reacciones químicas fueron introducidos a principios de los años 50 por Frederick Sanger en Cambridge (7) y Pehr Edman en Melbourne (8) y después a finales de la década Klaus Biemann en el MIT desarrolló nuevas estrategias por EM (9, 10). El método de Sanger se basaba en el marcado N-terminal de di-y tripéptidos con un grupo dinitrofenilo, seguido de la hidrólisis y la identificación de los aminoácidos marcados y no marcados por cromatografía de papel mientras que Edman se basó en la degradación progresiva de proteínas a partir de su grupo N-terminal. (Ver capítulo 6). La principal limitación en el uso de técnicas de EM para esta aplicación fue la necesidad de la preparación de derivados volátiles de los péptidos, susceptibles al mismo tiempo de ionización por IE y lo suficientemente estables para simplificar los espectros de masas resultantes.

Los estudios preliminares de Biemann sobre el uso de la EM para la determinación de secuencias de AAs se basaron en la reducción con $LiAlH_4$ o $LiAlD_4$ de los enlaces peptídicos y los grupos carboxilo terminales a poliaminoalcoholes o poliaminas. El cambio de enlaces poliamida a poliamina facilita la rotura de los enlaces peptídicos, siendo los poliaminoalcoholes más volátiles que los poliamidas ácidas. Más tarde, los péptidos pequeños se transformaron en ésteres N-trifluoroacetilados, ésteres de N-Acetilpéptido o N-Acilpéptido o también derivados N, O-Permetilados (11, 12) (Figura 2).

Figura 2. Primeras estrategias para la determinación de secuencias peptídicas por espectrometría de masas. [Reproducido con el permiso de John Wiley & Sons, Ltd.: ver Ref. [1], Copyright (2008)].



2. Los primeros ordenadores en química de proteínas.

El uso de ordenadores, para tratar datos de EM, debutó en 1966 cuando Klaus Biemann y cols. utilizaron un ordenador para determinar secuencias de AAs en oligopéptidos (13). Del mismo modo, Barber y cols. y simultáneamente Senn y McLafferty en Purdue describieron el uso de ordenadores para la interpretación de espectros (14, 15). Los tres grupos utilizaron un CEC 21-110 con introducción directa de la muestra en la fuente de iones. Sin embargo, el uso de la EM para la determinación de secuencias de AAs en péptidos, en aquel momento no fue más que una alternativa prometedora, restringida a oligopéptidos puros y de no más de 10 residuos (16). Por este tiempo también debutaron los primeros espectrómetros de masas controlados por ordenadores, como el Atlas CH5 (ver a continuación) y el Finnigan 1015. Aunque, los ordenadores han llegado a controlar todas las funciones de un espectrómetro de masas podemos afirmar que el desarrollo de la Proteómica moderna no hubiese sido posible sin el establecimiento y uso de las bases de datos de proteínas, como veremos mas adelante.

3. Ionización química.

La introducción en 1966 de la EM de ionización química por B. Munson y F. Field en la Esso de Baytown, Texas (17), marcó un hito importante. Esta técnica, comercializada con mucho éxito, representó otro intento de ionización suave de moléculas biológicas, después de la temprana introducción de la Ionización por Campo (18). Como «técnica de ionización suave", tenía la capacidad de dar información sobre el peso molecular ya que la muestra se ionizaba por reacciones iónicas menos energéticas que las del IE.

En 1967, la empresa alemana MAT comercializó el CH5, un instrumento de sector magnético de 90º de enfoque simple con circuitos transistorizados para mejorar la estabilidad, lo que permitía un buen nivel de resolución. El instrumento se fabricó hasta 1975 y representó un pequeño pero significativo avance en la carrera desde los circuitos totalmente analógicos en EM hacia los digitales. Más tarde, en 1971, el analizador del CH 5 se utilizó en el MAT 111, un CG/EM compacto que ganó el premio como uno de los 100 Nuevos Productos Técnicos más importantes del año.

4. Primeras secuencias proteicas por métodos no espectrométricos y primera publicación sobre "Haces Moleculares de Macroiones".

Edman y Begg en 1967 describieron la secuenciación automatizada de proteínas por métodos no espectrométricos (19) (Ver capítulo 6) e inventaron un secuenciador automático para la eliminación de los reactivos y disolventes de cada ciclo de reacción, que Beckman comercializó en 1970. El método rápidamente sustituyó al reactivo de Sanger (fluorodinitrobenzeno) para romper secuencialmente los residuos amino terminales de las cadenas polipeptídicas. Los diversos fragmentos se aislaban y purificaban y el proceso se repetía hasta que todos los fragmentos tenían menos de 25 residuos. En este procedimiento, el péptido que se quiere secuenciar se adsorbe

sobre una superficie dura. El reactivo de Edman, fenilisotiocianato, se añade al péptido adsorbido y reacciona con el grupo amino del aminoácido N-terminal. El derivado resultante se separa selectivamente y luego se isomeriza a una feniltiohidantoina más estable que puede recuperarse por lavado e identificarse por cromatografía. La cantidad mínima de proteína necesaria, a causa de las pérdidas de la muestra en el sistema, era 10-100 nmol. Inicialmente, se utilizó la cromatografía en capa fina hasta la descripción de las primeras separaciones por HPLC (20). Todo el ciclo se puede repetir con una eficiencia del 98%. Esto permitía determinar de forma segura y automática cerca de 50 AAs, lo cual era mucho más de lo que las técnicas de EM podían ofrecer en aquel momento. Sin embargo, la degradación de Edman no estuvo exenta de limitaciones, ya que no funcionaba en AAs bloqueados por el extremo N-terminal. Esto proporcionó una oportunidad única para la EM, como se demostraría con el TRF. Además, las regiones hidrofóbicas de los AAs también eran una limitación del método ya que originaban importantes pérdidas en los recipientes de reacción, lo cual representaba otro incentivo para el uso de la EM.

Un año después, en 1968, Malcolm Dole publicó un documento, entonces prácticamente ignorado, sobre "Haces Moleculares de Macroiones" (21) que tras un largo periodo de germinación culminó en el desarrollo de la ionización por electrospray. En ese trabajo, Dole estudió lo que sucede durante la evaporación de un disolvente en una gota provista de carga eléctrica. Dole infundía una disolución de moléculas a través de un tubo de poliestireno de pequeño diámetro, mantenido a un alto potencial en relación a un contra-electrodo pero los espectrómetros de masas de la época eran incapaces de ver los iones monocargados de más de 10000 Daltons (22).

5. Los primeros cuadrupolos y la Disociación Inducida por Colisión.

En septiembre de 1968 Robert Finnigan, que el año anterior había fundado FINNIGAN INSTRUMENTS, inició la introducción comercial del modelo 1015, el primer CG/EM de cuadrupolo con sistema de datos integrado (Figura 3). Aunque, el cuadrupolo usado como filtro de masas en el modelo 1015 era un dispositivo de baja resolución, limitado a la determinación de iones de energía cinética relativamente baja, rápidamente se convirtió, igual que el LKB 9000, en un instrumento analítico extensamente aceptado, especialmente para usos ambientales. También es interesante señalar que la utilización de la técnica de la Disociación Inducida por Colisión a energías de colisión elevadas (rango de keV), que con el tiempo se convirtió en una herramienta indispensable para elucidación estructural, realmente se inició aquel año con los trabajos de K. Jennings sobre la CID de iones moleculares aromáticos (23) y los de F. McLafferty sobre iones metastables inducidos por colisión (24). Durante algún tiempo esta fue "una técnica en búsqueda de un problema" hasta que la aparición del CID de baja energía (hasta 200 eV) y su incorporación a los cuadrupolos triples, junto con la introducción del bombardeo por átomos pesados, contribuyo a la rápida popularización de los protocolos de CID (25).

Figura 3. Modelo 1015 de Finnigan. El primer GC/EM de cuadrupolo computerizado con sistema de datos integrado. (Courtesy of D. Cho). [Reproducido con el permiso de John Wiley & Sons, Ltd.: ver Ref. [1], Copyright (2008)].

6. Espectrometría de Masas de Desorción por Campo.

En aquellos tiempos, la búsqueda de métodos para el estudio de sustancias de baja volatilidad y térmicamente inestables condujo a la introducción de la EM de Desorción por Campo por H.D. Beckey (26). Con esta técnica el autor demostró que un hidrato de carbono no derivatizado podía ionizarse sobre la superficie de un ánodo de ionización por campo, generando iones moleculares sin descomposición. Además, también mostró que los péptidos libres muy polares podían ionizarse por Desorción por Campo, incluyendo aquellos que contenían arginina o histidina no protegidas, y que no producían espectros útiles por IE o Ionización Química (27).

7. EM de Alta Resolución. El salto del análisis de péptidos a las proteínas.

En 1969, los métodos químicos y enzimáticos clásicos para la determinación de secuencias de polipéptidos se habían sofisticado y extendido a moléculas de gran tamaño y complejidad, como la γG inmunoglobulina humana con dos cadenas de 214 y dos de 446 aminoácidos. Fue entonces cuando Howard Morris y Mike Barber, quienes mas tarde se convertirían en personajes clave en la búsqueda de nuevos métodos de análisis de macromoléculas por EM, demostraron por primera vez en un AEI MS902 el potencial de la EM de alta resolución para la determinación de secuencias de AAs en péptidos obtenidos de hidrolizados de una proteína de la fibroina de la seda (28). No obstante, aunque la EM ya apuntaba un futuro en este campo y con esta publicación había realizado el salto al análisis de proteínas, todavía no podía competir con los métodos de secuenciación química en términos de sensibilidad y por ello resultaba mas adecuada para casos especiales relacionados con AAs inusuales o péptidos con el N-terminal bloqueado. Un claro ejemplo en esta dirección fue el uso de la EM para descifrar la estructura de un elusivo y pequeño péptido natural de baja volatilidad. Este es un caso histórico único en el que Dominic Desiderio, quien trabajaba en el grupo de Roger Guillemin en el Baylor College of Medicine (29), consiguió determinar en 1970 la estructura primaria del Factor Liberador de Tirotropina (TRF). Guillemin hacia años que estaba interesado en la identificación de hormonas peptídicas en el cerebro, por lo que recibió el Premio Nobel en 1977.

Partiendo de 300.000 hipotálamos bovinos aislaron un total de 10 mg de TRF. El análisis químico tras la hidrólisis había proporcionado la composición en AAs (His, Pro, Gln/Glu) pero la secuencia no era conocida y además resultaba imposible de establecer debido al bloqueo de los grupos N- y C-terminales. Cabía la posibilidad de seis posibles tripéptidos isoméricos por lo que la contribución de la EM resultaba necesaria. No obstante, como más tarde describió Desiderio (30) los primeros intentos a alta resolución en un CEC 21-110 no proporcionaron un espectro por las dificultades para desorber térmicamente el TRF natural. Finalmente, la ionización química en un Finnigan 1015 proporcionó un peso molecular para el TRF y posteriores modificaciones de la sonda de introducción de muestras en el CEC 21-110 permitieron obtener un espectro de alta resolución que contenía los fragmentos determinantes de la secuencia de AAs dando finalmente la estructura.

8. Espectrometría de Masas de Alta Resolución y análisis directo de mezclas de oligopéptidos.

Aproximadamente en la misma época Fred McLafferty y sus colaboradores demostraron la posibilidad de obtener información directa de secuencia en mezclas de oligopéptidos a través de la medida exacta de las masas de 396 picos en el espectro de una mezcla de oligopéptidos esterificados y acetilados (31), confirmando así el potencial de la EM de alta resolución para la secuenciación de péptidos. No obstante, la base de los procedimientos de secuenciación todavía residía en el uso de métodos químicos y enzimáticos y en 1971 Richard Laursen (32) construyó un secuenciador de péptidos en fase sólida que utilizaba un soporte de poliestireno empaquetado en una columna a través de la cual fluían los reactivos y disolventes de Edman. Esta técnica contribuyó a limitar las perdidas de muestra relativas a los anteriores secuenciadotes y, desde su introducción en 1970-1971, estas maquinas fueron muy utilizadas hasta su sustitución por las técnicas modernas de EM. En aquellos primeros tiempos, estos métodos de secuenciación química manual y automatizada permitieron ya en 1978 la colección y anotación de aproximadamente 1100 secuencias proteicas con unos 120.000 residuos de AAs en el primer "Atlas of Protein Sequence and Structure" editado por Margaret Dayhoff y cols. (1965) en la National Biomedical Research Foundation, Georgetown University Medical Center (33).

9. Primeras secuencias de aminoácidos por CG-EM y técnicas de derivatización.

En línea con el interés por el incremento de la volatilidad de un péptido el grupo de Klaus Biemann describió en 1972, un método general de CG-EM controlado por ordenador para la determinación de secuencias de AAs en mezclas complejas de oligopéptidos que, después de su conversión a poliamino alcoholes, se trimetilsililaban (34) y así producían espectros de masas fácilmente interpretables. Como prueba de la idoneidad del enfoque de secuenciación por CG-EM, el grupo publicó la determinación inequívoca de la secuencia de 20 aminoácidos del fragmento C-terminal de la actina (35). El péptido se hidrolizaba a una mezcla de di-, tri- y tetrapéptidos. Más tarde, en 1979 este grupo demostró la eficacia de sus métodos de secuenciación de proteínas por CG-EM mediante la identificación de los 248 aminoácidos de una larga cadena proteica de la membrana de la bacteriorodopsina (36). Casi al mismo tiempo, Caprioli y sus colaboradores informaron sobre un nuevo método de secuenciación de péptidos, también por CG-EM, a partir de la hidrólisis enzimática con dipeptidil aminopeptidasa I y conversión de los dipéptidos resultantes a ésteres de pentafluoropropionilo (PFP) (37). Tal como Biemann describió más tarde (38), hasta la llegada del FAB en 1981 (véase más adelante), el análisis de secuencias de proteínas y oligopéptidos mediante EM se realizaba por degradación química y/o enzimática de la muestra obteniéndose una compleja mezcla de péptidos pequeños de entre dos y ocho residuos. A esto le seguía la derivatización química de los componentes de la mezcla y la aplicación de la combinación de CG con EM para la separación y caracterización de la estructura primaria de cada péptido en la mezcla. Así, el análisis de mezclas en los primeros tiempos de la EM se abordaba principalmente por la separación cromatográfica previa en línea y, más concretamente, separación por CG de alta eficacia y con largas columnas capilares de acero inoxidable. El uso del espectrómetro de masas para ambas tareas (separación de componentes seguida por la caracterización individual de los distintos componentes) no era todavía una posibilidad generalizada en la práctica. Sin embargo, en 1973 John Beynon y cols. en el Departamento de Química de Purdue describieron el desarrollo instrumental del Espectrómetro MIKE (39). Este diseño se basó en trabajos anteriores del mismo grupo sobre los espectros de energía cinética de iones (IKE) de moléculas orgánicas

y, aunque inicialmente definía la energética de las reacciones de descomposición de iones, fue con este instrumento combinado con una técnica de ionización suave, como la ionización química, con el que el equipo de Purdue demostraría unos años más tarde la viabilidad del análisis directo de mezclas (40, 41).

10. Ionización Química a presión atmosférica.

En 1973 el grupo de Evan Horning en el Institute for Lipid Research en el Baylor College of Medicine de Houston, Texas, describió una alternativa a la ionización química basada en el desarrollo de sistemas de ionización química a presión atmosférica (APCI). Esto se conseguía con los electrones primarios emitidos por una fuente de radiación de níquel-63 de 12.5 mCi sobre un gas portador muestreado en continuo en una cámara de reacción a presión atmosférica (42). Los iones formados en la cámara API mediante una serie de reacciones ion-molécula, por transferencia de protón o de carga, eran arrastrados a través de un orificio por el flujo de gas portador al interior del cuadrupolo Finnigan 1015. Así se obtenían límites de detección mucho menores que en los sistemas convencionales de aquel tiempo, por lo que este trabajo pionero abrió las puertas al diseño y uso de sistemas API para CG-EM o HPLC-EM bajo control de ordenador (43). El grupo de Horning reconoció las ventajas de vaporizar el disolvente y los analitos a presión atmosférica en lugar de en la cámara de vacío, utilizando para ello el disolvente como reactivo de ionización química. No obstante, la idea quedo arrinconada hasta que este trabajo condujo al desarrollo e introducción por SCIEX en 1977 del primer instrumento comercial basado en la APCI. Este instrumento era un sistema de vigilancia ambiental comercializado bajo el nombre de TAGA-3000 para el análisis de trazas de gases en la atmósfera.

11. Espectrometría de Masas por Desorción de Plasma.

Durante estos primeros tiempos de la historia de la “espectrometría de masas pre-biológica” varios grupos intentaban mejorar la tasa de volatilización de los productos naturales “pesados”, minimizando al mismo tiempo su descomposición, para alcanzar así buenos niveles de detección. En esta línea, en 1974 tuvo lugar un importante avance gracias al trabajo de McFarlane y cols., en la Texas A & M University, quienes publicaron los primeros resultados obtenidos por desorción simultánea/ionización química de la arginina y la cistina. Para ello utilizaron una fuente de partículas de fisión de 252Cf de alta energía a fin de generar partículas de 100-MeV para bombardear la muestra depositada sobre una delgada lámina de aluminio. Sin embargo, como el bajo nivel de eventos de fisión por segundo en la fuente de 252Cf, dio lugar a corrientes de iones de baja intensidad, se necesitaban instrumentos capaces de registrar todas las masas simultáneamente por lo que el análisis de masas se llevó a cabo en un instrumento TOF (44).

12. Nuevos métodos de separación y secuenciación de proteínas.

En 1975 sucedieron dos eventos importantes relacionados con avances significativos en el estudio de proteínas y la química de péptidos. Por un lado, J. Hughes y cols. en la Universidad de Aberdeen, describieron en la revista Nature la identificación de dos pentapéptidos de cerebro con una potente actividad agonista opioide (45). Por otra parte, ese mismo año O'Farrell describió el desarrollo de un método basado en la electroforesis bidimensional en gel de poliacrilamida (2D-PAGE) para la separación de proteínas (46) (véase el capítulo 6). Previamente Hughes había descrito una sustancia que competía por los sitios de unión de los receptores opiáceos en el cerebro y que se denominaba Encefalina. En colaboración con Howard Morris del Imperial College de Londres, y con la ayuda de la EM, las incertidumbres sobre longitud de la cadena y asignación de secuencia se resolvieron por análisis del derivado permetilado de la N-acetil Encefalina natural. Así, la EM ya estaba demostrando ser un prometedor método para el análisis de aminoácidos alternativo al método de degradación de Edman, especialmente en casos conflictivos.

La separación por 2D-PAGE se realizaba por isoelectroenfoque (IEF) según la carga (pI) en la primera dimensión y según el tamaño (Mr) por SDS-PAGE en la segunda dimensión. La técnica demostró una capacidad única para la resolución de mezclas complejas de proteínas, lo que permitía el análisis simultáneo de cientos o incluso miles de productos derivados de los genes y por lo tanto se convirtió en un hito importante en el camino hacia la Proteómica, como demostrarían más adelante en 1977 Anderson y Anderson con la adaptación de la técnica de PAGE para el análisis de proteínas del plasma humano (47).

Entre 1972 y 1978 se desarrollaron varios métodos y enfoques nuevos de secuenciación, tales como la degradación de Edman-Dansilo (Cap. 6), en la que se hidroliza el péptido dansilado y el aminoácido N-terminal dansilado se extrae e identifica por cromatografía en capa fina bidimensional. Este método fue el primero de secuenciación manual de elevada sensibilidad utilizado en los primeros análisis de secuencias de proteínas. Otro método era el de rotura de las proteínas con bromuro de cianógeno (CNBr) que dividía las cadenas polipeptídicas por sus residuos de metionina (Cap. 6). Así, gracias a una combinación de métodos, se pudo establecer, por ejemplo, la estructura primaria completa de proteínas ribosómicas en cantidades picomolares, de modo que 36 de las 55 proteínas ribosomales de *E. coli* y / o sus mutantes se secuenciaron por completo en este período de tiempo. Además, la llegada de los modernos equipos de HPLC proporciono la posibilidad de resolver en un solo análisis AAs derivatizados (48).

13. Los primeros intentos de acoplamiento de la CL con la EM.

A pesar de las dificultades técnicas, en 1977 Finnigan logro acoplar la CL a la EM comercializando el modelo 4000 que fue el primer sistema CL-EM de esta firma. El equipo incorporaba la interfaz de cinta transportadora de Kapton-desarrollada por los grupos de Scott y McFadden en 1974-1976 (4) y patentada por Finnigan en 1976. La denominada cinta móvil era capaz de proporcionar espectros de masas que podían buscarse en bibliotecas de impacto electrónico pero era mecánicamente complicada. En el mismo año, la empresa de Espectrometría de Masas de Bremen presentó el MAT 44, un cuadrupolo para CG-EM que, según Curt Brunnée (49) fue el primer espectrómetro de masas analítico completamente controlado por ordenador en el mercado.

14. Análisis directo de mezclas y Espectrometría de Masas MIKE.

En los años 60 y 70, el interés por el acoplamiento de la cromatografía al EM, ya fuese CG o CL, era alto, ya que esto claramente podía abrir el camino hacia el estudio de complejas mezclas naturales. La idea era que a pesar de que la EM podía proporcionar una cantidad inusual de información sobre los distintos compuestos en una muestra biológica, de por sí no era lo suficientemente específica como para dar respuestas totalmente selectivas para determinados compuestos. Por ello, la identificación tenía que ir precedida por la separación previa de los componentes de la mezcla. Sin embargo, también hubo quien reconoció el potencial de la EM para el análisis directo de mezclas y en esta línea, el concepto de análisis de directo mezclas ya se menciona en la literatura en 1967 con la separación e identificación de isómeros en mezclas complejas de hidrocarburos. En 1970, como se ha citado anteriormente, Fred McLafferty en un estudio sobre secuenciación de mezclas de oligopéptidos utilizó la vaporización parcial de la muestra desde una sonda de inserción directa, las masas exactas y los iones metastables generados en un espectrómetro de masas AEI MS-902. Así obtuvo la especificidad necesaria para determinar cuales eran los picos en el espectro de una mezcla de oligopéptidos que procedían del mismo oligopéptido (31). Más tarde, en 1972, el mismo grupo amplió su método a una variedad de mezclas complejas, señalando también que el uso de la activación por colisión produciría más picos y por lo tanto más información que los espectros de iones metastables (50). La capacidad de análisis de mezclas convertía a la EM en una herramienta atractiva para la secuenciación polipeptídica, reduciendo el número de operaciones de derivatización química y las cantidades individuales de oligopéptidos (50).

Al mismo tiempo, John Beynon, Graham Cooks y sus colaboradores estaban desarrollando el concepto de desacoplamiento del analizador de doble enfoque e implementaron la técnica de espectrometría de doble sector con geometría inversa (configuración B/E) o espectrometría de masas de energía cinética de iones (MIKES) (39). La técnica MIKES, desarrollada por Cooks y colaboradores, se utilizó en varios estudios sobre química de iones en fase gas hasta 1976-1978. Este período de dos años quedo marcado por un frenesí de actividad adicional plasmada en publicaciones sobre prometedoras aplicaciones de la nueva técnica de MS/MS. De todos estos informes, merecen ser destacados los del grupo de Cooks en Purdue (40, 41), quienes utilizando un espectrómetro de masas MIKE (es decir, un espectrómetro de doble sector de geometría inversa), demostraron por primera vez el análisis (no cromatográfico) directo total de mezclas via EM, tanto para iones positivos como negativos. La colocación del imán (análisis de masas) antes del sector eléctrico (análisis de energía cinética) permitió la selección de un ion, generado preferentemente por ionización química. El uso de una "técnica de ionización suave" facilita el análisis de las mezclas, ya que cada componente de la mezcla se ioniza a su ion molecular con poca o ninguna fragmentación. Esto simplifica enormemente el espectro de masas de la mezcla y como los iones moleculares se separan por sus relaciones de m/z en la primera etapa (analizador magnético) pueden seleccionarse para su caracterización en la segunda etapa (sector eléctrico) mediante procesos de fragmentación inducidos via transiciones metastables o por disociación inducida por colisión (CID). Este trabajo fue revisado por Kondrat y Cooks en su clásica publicación sobre Instrumentación en Analytical Chemistry (51).

15. Activación de iones por Colisión en Cuadrupolos Triples.

La evolución de la instrumentación hacia nuevas configuraciones del analizador que facilitaban el modo de operación por activación colisional, junto con los métodos de ionización suave, contribuyeron a convertir el análisis de mezclas en una posibilidad real. Según Graham Cooks (RG. Cooks, comunicación personal), la práctica del análisis directo de mezclas mediante EM realmente "comenzó con la combinación de la ionización suave y las posibilidades que ofrecía la geometría inversa" y pronto fue incorporada y simplificada por los cuadrupolos triples. Estos fueron el resultado del desarrollo de la EM de triple cuadrupolo controlado por ordenador realizado por Yost y Enke en 1978 en Michigan State (52).

Esta configuración instrumental facilitó el uso de la activación colisional de iones precursores seleccionados en el análisis de mezclas y sin duda fue un acontecimiento importante en el camino hacia una identificación rápida y directa de mezclas complejas por EM en Tandem y, eventualmente, hacia la Proteómica. En la práctica y para trabajar a alta sensibilidad por MS/MS, el número y abundancia de iones secundarios debe aumentarse por adición de energía a los iones primarios. Esto se logra por activación por colisión (CAD). De esta forma, la energía traslacional de los iones se convierte en energía interna por excitación colisional lo cual conduce a la posterior descomposición unimolecular de los iones excitados.

El nuevo enfoque descrito por Yost y Enke en 1978 resolvió el problema del redireccionado de aquellos iones de baja energía de traslación que, seleccionados en el MS1 y sometidos a choques de "bolas de billar" en la cámara de colisión, se enviaban al quadrupolo del MS3 utilizando un cuadrupolo de sólo radiofrecuencia (RF) (no separación de masas) como cámara de MS2 para los procesos de disociación por activación colisional (CAD). Los iones ampliamente dispersados por las colisiones en esa cámara quedaban focalizados de manera eficiente por el campo de RF para su transmisión al segundo analizador de masas MS3 (véase el articulo de revisión de EM en tándem por Yost y Boyd (53). En comparación con las células electrostáticas de colisión utilizadas en espectrómetros de masa magnéticos, la celda de colisión en el sistema de Yost y Enke era de 20 a 100 veces mas eficiente y por tanto mas sensible. Por otra parte, la transmisión de iones producto variaba de 20 a 95% en relación con el 1% en las células de colisión electrostáticas.

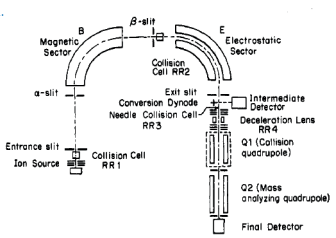
16. Los primeros Espectrómetros de Masas Híbridos.

Como se ha indicado, antes del triple cuadrupolo, el análisis de masas en tandem se efectuaba con espectrómetros de geometría reversa de doble enfoque en el modo de barrido acoplado (linked scan) y, aunque se obtenían así espectros de CID a resolución unitaria, la baja resolución de los iones precursores originados en el MS1 causaba interferencias con la fragmentación de otros precursores y obstaculizaba la posibilidad de seleccionar un ion monoisotópico en la región del peso molecular del espectro así como la posibilidad de distinguir compuestos isoméricos e isobáricos. Por ello, en búsqueda de mayor resolución de iones precursores así como de mayor sensibilidad mediante regimenes de activación colisional de alta energía, se desarrollaron los instrumentos de tres sectores y las configuraciones híbridas. (Tabla 1 y Figuras 4a y 4b), La figura 4a ilustra el esquema óptico de un instrumento híbrido diseñado en torno a la configuración BEqQ (54). El instrumento real se ilustra en la figura 4b y representa el espectrómetro de masas MS/MS HSQ 30 de Finnigan MAT (55).

Tabla 1. Resumen de configuraciones de Espectrómetros de Masas en Tándem (en el espacio).

	MS1	Célula q para CID	MS2	
Configuración	Q1	CID*	Q2	TAGA 6000, TSQ 45, TSQ 700 &7000
Resolución	1		1	
Configuración	B	CID	E	Finnigan MAT 312, VG ZAB 2F
Resolución	«1		≤ 1	
Configuración	EB	CID	E	Kratos MS 50 TA, VG Autospec
	BE	CID	B	ZAB 3HF
	BE	CID	Q	ZAB-EQ, VG ZAB-HSQ, Finnigan MAT HSQ 30
	EB	CID	Q	VG 70-SEQ, VG 7070EQ, Kratos MS890
Resolucion	» 1		≤ 1	
Configuración	B	CID	EB	ZAB 3F
Resolucion	1		» 1	
Configuración	EB	CID	EB	
	BE	CID	EB	VG ZAB-SE 4F, JEOL HX110/HX110
	BE	CID	BE	VG ZAB-T
Resolución	» 1		» 1	
CID: "Collisionally Induced Dissociation" en cámaras de colisión				
CID* : collision-induced dissociation a bajas energías.				

a.



b.

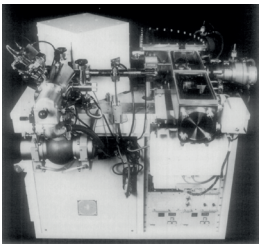


Figura 4a. Diagrama esquemático del instrumento hibrido en configuración BEqQ el que se aprecian las cuatro regiones de reacción (RR1 a RR4). [Reproducido con el permiso de la American Chemical Society: ver Ref. [54], Copyright (1985)].

Figura 4b. Sistema MS/MS (configuración BEqQ) del espectrómetro Finnigan MAT HSQ 30. [Reproducido con el permiso de Elsevier Science B.V.: ver Ref. [55], Copyright (1987)]. En esta imagen puede discernirse la distribución de los diferentes sectores mostrados en el diagrama en (a). En la parte trasera pueden apreciarse el imán (caja blanca) y sector electrostático a la derecha. La situación de los quadrupolos puede verse en la parte frontal derecha de la imagen.

A energías de colisión bajas (20-100 eV) los espectros resultantes de CID no eran lo suficientemente reproducibles para el uso de librerías de CID y, en consecuencia, los híbridos que utilizaban un cuadrupolo para la segunda etapa adolecían de la ineficiencia característica de las colisiones de baja energía y de falta de sensibilidad ya que para registrar el espectro del ion producto tenia que barrerse el cuadrupolo. En conjunto, esto impulsó la introducción de los instrumentos de cuatro sectores (56) (Figura 5), que proporcionaban alta sensibilidad y resolución en todo el rango de masas, aunque no se comercializaron hasta el final de la década de 1980. En efecto, estos instrumentos se convirtieron en prácticos, aunque incómodos, "analizadores de péptidos" proporcionando una sensibilidad muy alta con detectores matriciales y caracterizaciones estructurales inequívocas de los iones de masa alta y de péptidos isoméricos (Vg. la asignación de péptidos con isoleucina o leucina). Los costes del instrumento y del personal, más el mantenimiento de estos espectrómetros de masas limitaba sus aplicaciones analíticas de rutina y es por eso por lo que al final los triples cuadrupolos, a pesar de sus limitaciones de resolución unitaria y rango de masas, alcanzaron un éxito inmediato, ya que proporcionaban alta sensibilidad de detección y la posibilidad de automatizar la configuración del instrumento.

Los primeros cuadrupolos triples comerciales para MS/MS los desarrollaron en 1980 Sciex Inc. (Ontario, Canadá) y Finnigan Corporation (Sunnyvale, California) que los comercializaron

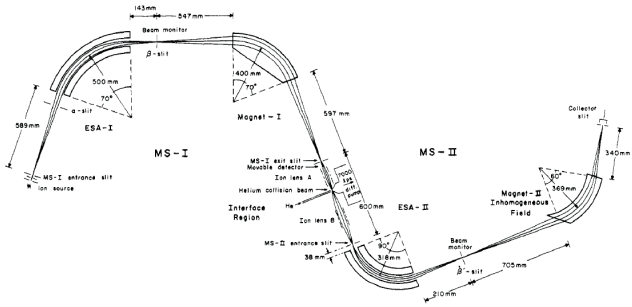


Figura 5. Esquema de uno de los primeros instrumentos MS/MS de cuatro sectores de alta resolución. [Reproducido con el permiso de la American Chemical Society: ver Ref. [56], Copyright (1980)].

en la Conferencia de Pittsburg en Marzo de 1981. Allí presentaron la serie TAGA 6000 (Figura 6) y el triple cuadrupolo para CG/EM, el TSQ45. El primer prototipo del modelo TSQ MS/MS se instaló en la Shell Research and Development en Houston a principios de noviembre de 1980 (Figura 7). Por lo que respecta a los instrumentos de tres sectores, Kratos Scientific Instruments (Westwood, Nueva Jersey) fue la primera compañía en comercializar un triple analizador de configuración EBE, el MS30 MS50TA de doble haz mientras que VG Micromass Ltd. (Cheshire, Inglaterra) presentó la línea del ZAB 3F basada en una configuración BEB. El coste de estos instrumentos osciló entre los 500.000 y los 800.000 dólares, pero los instrumentos de cuatro sectores no estuvieron disponibles hasta finales de la década.

RM Hewick y cols. (57) en 1981 revolucionaron la secuenciación de proteínas por métodos no espectrométricos con un nuevo tipo de microsecuenciador, comercializado por la nueva empresa Applied Biosystems, Foster City, CA. que utilizaba como cámara de reacción un cartucho de nuevo diseño y productos químicos en fase gas, permitiendo la secuenciación de unos pocos picomoles.

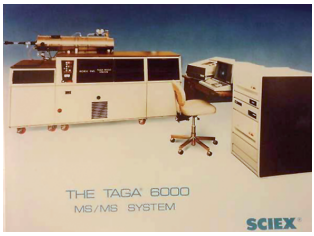


Figura 6. Instrumento TAGA 6000 de triple cuadrupolo fabricado por SCIEX (cortesía de T. Covey) [Reproducido con el permiso de John Wiley & Sons, Ltd.: ver Ref. [1], Copyright (2008)].

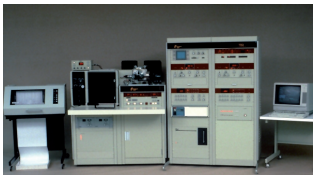


Figura 7. Primera unidad del triple cuadrupolo TSQ de Finnigan (cortesía de D. Cho). [Reproducido con el permiso de John Wiley & Sons, Ltd.: ver Ref. [1], Copyright (2008)].

17. Bombardeo con átomos pesados (FAB) y SIMS líquido.

En 1981 el profesor de química Michael Barber y colaboradores en el Instituto de Ciencia y Tecnología de la Universidad de Manchester en Inglaterra, desarrollaron la EM por bombardeo con átomo pesados (FAB MS), o “SIMS líquido” (58). Este fue un avance verdaderamente revolucionario, que anunció el comienzo de una nueva era en la EM biológica. En FAB MS, la muestra se deposita sobre un líquido viscoso (Vg., glicerol) y se utilizan haces de átomos neutros de argón de alta energía (keV) para vaporizar e ionizar compuestos polares en la superficie del líquido con una mínima descomposición. Los iones son expulsados de la fase liquida a la fase gaseosa en una especie de proceso de desorción/ionización. Esto produce espectros de las grandes moléculas orgánicas no volátiles con una relación señal-ruido del mismo orden que en impacto electrónico convencional, pero en general con poca fragmentación y por tanto poca información estructural. La SIMS (espectrometría de masas de iones secundarios) fue introducida por primera vez en 1950 por Richard E. Honig en los Laboratories RCA, de Prince-

ton, NJ. En SIMS, se utiliza un haz de iones para bombardear una muestra sólida produciendo iones secundarios de baja intensidad.

En su primera publicación en 1981 sobre la utilización de FAB (58), los autores presentaron un espectro de masas de un undecapéptido, obtenido a partir de 1 µg de material, con alta sensibilidad en la región del ión molecular a 1319 Da y con la fragmentación suficiente para permitir la secuenciación del péptido. También informaron de que habían tenido éxito en ionizar Gastrinas (Mr 2096) y Melitina (Mr 2845 con 26 unidades AA). Hasta entonces la aplicación de la EM a la química de péptidos y proteínas estaba muy limitada a un rango de masa muy bajo, por la necesidad de volatilizar e ionizar la muestra antes del análisis. Esta situación, a pesar del anterior desarrollo del PDMS y las técnicas de desorción por campo, cambió de facto con la introducción del FAB, una técnica que podía incorporarse fácilmente a los instrumentos existentes abriéndolos al rango de los pocos kilodaltons y por tanto cambiando el enfoque de la moderna EM biológica.

Casi al mismo tiempo, el grupo de Donald Hunt, presentó un primer informe sobre el uso de un triple cuadrupolo de CID de iones secundarios para el análisis directo de la secuencia de mezclas de oligopéptidos, sin separación previa de los componentes (59). Por lo tanto, Hunt y cols. demostraron que la disociación a baja energía de colisión inducida en analizadores de masas de triple quadrupolo, comparada con los más sofisticados espectrómetros de masas multisectoriales, también podría ser útil para el análisis de la secuencia de un péptido. Esto más tarde tendría importantes consecuencias prácticas, como este mismo grupo demostró a principios de los 90 con la caracterización de péptidos unidos a moléculas de MHC de clase I o clase II por EM de triple cuadrupolo (60).

Un año más tarde Anne Dell y Howard Morris, con un espectrómetro de masas equipado con imán de alto campo, el ZAB 1F de VG Analytical, demostraron la capacidad de observar con precisión los pesos moleculares de polipéptidos con un rango de masas de hasta 6000 Da, cerca de los límites de tecnología de la época (61). En ese artículo se presentó el espectro de masas FAB de la insulina intacta no reducida, el mayor péptido natural determinado por espectrometría de masas. Al mismo tiempo, el grupo de B. Sundqvist en Suecia, junto con McFarlane anunciaron que habían realizado la primera observación de la masa molecular de la insulina en un espectro de masas, pero obtenido en este caso por PDMS (62).

También en 1982, William Aberth, Kenneth Straub y Alma Burlingame desarrollaron la fuente de iones de cesio líquido SIMS (FAB) para la secuenciación de péptidos (63). Este trabajo demostró que el “SIMS líquido” era posible usando iones como haz primario de bombardeo y que ello no afectaba a los espectros de iones secundarios así obtenidos.

Poco después de la introducción de la ionización por FAB, esta técnica en instrumentos híbridos y de cuatro sectores se convirtió en un método bien establecido para el análisis estructural de péptidos derivados de proteínas de origen natural (para referencias, véase Bean et al. 64). Su rápida introducción y aceptación se debió principalmente a dos hechos. Por un lado, se eliminó la necesidad de tediosas y largas derivatizaciones químicas, necesarias para lograr que estas moléculas complejas resultasen susceptibles a la ionización química o por electrones y, por otro lado, podía adaptarse fácilmente a los espectrómetros de masas de cuadrupolo o sector, disponibles en aquella época. Esto último contrasta con la PDMS, que desde su desarrollo en 1974 se había limitado a unos pocos laboratorios debido a la falta de instrumentos comerciales. Del mismo modo, las técnicas de CID, se convirtieron de repente en un complemento necesario a los protocolos de EM que se utilizaban para obtener información estructural detallada de biomoléculas ya que el FAB, siendo una técnica de ionización suave, genera sobre todo moléculas protonadas de baja energía interna y por lo tanto no favorece la fragmentación de modo que en general el FAB no proporciona este tipo de información.

Por otra parte, desde la introducción de la 2D PAGE para estudios de proteínas en 1975, su uso se había sido retrasado por problemas prácticos relacionados con la inestabilidad del gradiente de pH y su limitada reproducibilidad, Principalmente por la variabilidad lote a lote de las preparaciones del anfólito portador. Estos problemas los resolvieron Bjellqvist y cols. (65) en 1982 con la introducción de gradientes de pH inmovilizados (IPGs) para la aumentar reproducibilidad de los geles 2-D. Finalmente, en 1989, Pharmacia Amersham comercializo tiras de IPGs que permiten la generación de gradientes de pH en cualquier rango entre pH 3 y 12 y por estas razones la 2D-PAGE con IPGs se convirtió en el método de elección para la separación e identificación micropreparativa.

En la cuarta Conferencia Internacional sobre Métodos de Análisis de Secuencias de Proteínas (MPSA), celebrada en 1982, Ken Walsh discutió nuevas posibilidades de secuenciación, y mencionó la EM como una "técnica de futuro". El año anterior, Walsh había publicado una visión general sobre los avances en la secuenciación de proteínas que sobre todo trataba de técnicas de separación y la química de la degradación de Edman (66), y en esta reunión Howard Morris, Steven Carr y VN Reinhold, Shimonishi Yasutsugu y Klaus Biemann presentaron varios trabajos sobre la secuenciación de proteínas por EM. El informe de esta reunión, que se encuentra en Http://www.bnl.gov/iapsap/linkable_files/pdf/MPSA1975-2004Wittmann-Liebold-Reflections.pdf) describe la impresionante irrupción en la reunión de las nuevas técnicas de HPLC en columnas de fase reversa, para la purificación de péptidos, proteínas y aminoácidos en el rango picomolar. En aquella conferencia, Ken Walsh, vaticino: "El análisis de secuencia se llevará a cabo en el rango del femtomol". Aunque ello era todavía un sueño en aquel momento, en efecto, tenía razón.

18. Nuevas posibilidades para el acoplamiento CL-EM.

En los años 80, con el acoplamiento de la CG a la EM ya implantado como técnica analítica bien establecida y aceptada el interés se centraba en desarrollar una solución viable para el acoplamiento, al parecer antagónico, entre la CL de alta eficacia y la EM. En esta línea, y después de la introducción en 1981 del triple cuadrupolo TAGA 6000 de Sciex que utilizaba una fuente de APCI, una colaboración entre Jack Henion en la Universidad de Cornell, Bruce Thomson en Sciex y Peter Dawson en el NRC de Canadá condujo a un trabajo preliminar sobre el acoplamiento en-línea LC-MS y LC-MS/MS. Esto se logró mediante la interconexión de una sonda de introducción directa de líquido (DLI) y la API de Sciex de diseño abierto (67). La sensibilidad, no obstante, resultó baja por la división del 10% del flujo del efluente y la desolvatación incompleta de las microgotas en la sonda DLI. Sin embargo, este trabajo estimuló la investigación sobre la vaporización de todo el flujo de líquido en la fuente APCI y condujo al desarrollo del nebulizador neumático que se presentó en la reunión de la ASMS en Washington en 1983, pero que no estuvo disponible comercialmente hasta finales de la década.

19. Ionización por Thermospray. Consolidación de la Cromatografía Liquida-Espectrometría de Masas.

En 1983 la introducción tanto de la interfaz de Termospray para LC/MS por Marvin Vestal y de los primeros analizadores comerciales de trampa de iones (Finnigan ITD 700 GC/MS), fueron dos eventos muy importantes en el avance de la EM. También, en ese año Dominic Desiderio

confirmó la identificación de la sustancia P por FAB, un undecapéptido amidado de Mr 1346 que se encuentra en el tejido cerebral. Para ello utilizó un Finnigan MAT de doble enfoque en el modo de barrido acoplado (68). Para entonces la EM en tándem se había convertido en una técnica bien aceptada para el estudio de mezclas complejas como apuntan Yost y Fetterolf (69) y empresas como Extranuclear Laboratories Inc. (EE.UU.), Finnigan MAT (EE.UU.), Kratos Ltd . (Reino Unido), Nicolet Instrument Corp. (EE.UU.), Sciex Inc. (Canadá), y VG-Micromass (Reino Unido) ya comercializaban varios instrumentos de este tipo.

La ionización por termospray (TSP), presentada por primera vez por Marvin Vestal y cols. en la reunión de la ASMS en Minneapolis en 1981, era capaz de producir iones a partir de una disolución acuosa que se había pulverizado directamente en el interior del espectrómetro de masas (70). Si bien se requiere calor para generar el aerosol en la sonda de TSP, la ionización por termospray se considera una técnica de ionización suave que genera analitos cargados a través del mecanismo de desorción por campo de los iones presentes en gotas en evaporación con carga iónica. La interfaz TSP fue un éxito comercial inmediato, ya que permitía introducir directamente flujos de hasta 1 ml/min en el espectrómetro de masas, manteniendo bajo control la descomposición térmica de los analitos. Por su rendimiento y facilidad de operación, el TSP experimento un aumento explosivo desde 1983 y se convirtió en el sistema más popular como interfaz LC/MS de fácil uso hasta su punto máximo en 1991/1992 y posterior desaparición debido a la introducción de la ionización por electrospray y las interfaces de nebulizador neumático calentado.

20. Las trampas de iones y el mapeado por FAB.

Figura 8. Trampa de Iones modelo ITD700 (Cortesía de D. Cho). [Reproducido con el permiso de John Wiley & Sons, Ltd.: ver Ref. [1], Copyright (2008)].

Con la introducción comercial por Finnigan MAT en 1983 de un analizador CG/EM de Trampa de Iones (Figura 8), se dispuso de un nuevo instrumento, que por su alta velocidad de barrido, su alta sensibilidad, rango de masas de alta resolución y su elevada versatilidad de uso (71), se convirtió rápidamente en una tecnología de base para el análisis estructural de proteínas y péptidos. De hecho, se ha dicho que el tamaño de una trampa de iones, aproximadamente el de una pelota de tenis, es inversamente proporcional a su versatilidad y aunque entonces el término “Proteómica” todavía no se había acuñado, con el tiempo este instrumento se convertiría en el caballo de batalla de la proteómica. En sólo dos décadas se pasó de los instrumentos grandes y aparentemente complejos, a una pequeña “caja blanca” de sobremesa que contenía un espectrómetro de masas digital de tándem (en el tiempo). En el 1950, Wolfgang Paul inventó el filtro de masas cuadrupolar, usado en 1968 por Robert Finnigan para introducir en el mercado los espectrómetros de masas de cuadrupolo y también la trampa de iones. Sin embargo, el uso de la trampa de iones se limitaba principalmente a la comunidad de la física hasta que en 1983, George Stafford y cols. también en Finnigan MAT, desarrollaron un nuevo modo de funcionamiento de la trampa. Dicho método se basaba 1) en la llamada inestabilidad selectiva de masas, lo que permitió el almacenamiento de todos los iones y su eyección secuencial selectiva, y 2) la adición helio como gas amortiguador dentro del volumen de captura para incrementar la resolución de masas. Ambos descubrimientos revitalizaron el interés por el uso práctico de la trampa y llevaron a su desarrollo comercial (72).

Cabe destacar también en 1983, la publicación por H. Morris y cols. describiendo un nuevo método para la detección y el establecimiento de la estructura primaria de las proteínas que

llamaron Mapeado por FAB (73). Este trabajo puede ser considerado como el precursor de las estrategias de la proteómica basadas en el mapeo o huellas digitales de la masa de proteínas, publicado en 1993 por Stults y cols. y que con tanto éxito se han venido usando desde entonces, por ejemplo para el estudio de modificaciones post-translacionales de proteínas por espectrometría de masas.

21. Ionización por Electrospray.

Mientras se desarrollaba la Trampa de Iones (tándem en el tiempo), el tándem MS en cuadrupolos triples y espectrómetros de masas híbridos (tándem en el espacio) se estaba afianzando en muchos laboratorios como técnica de referencia (74) y además el futuro de las aplicaciones analíticas de estas técnicas estaba a punto de recibir un tremendo empuje. Así, 1984 fue un verdadero año clave para le determinación de los pesos moleculares y secuencias de péptidos y proteínas. En primer lugar, por la introducción por la compañía sueca Nordic AB Biolon (Uppsala), del Biolon 10. Este era un Espectrómetro de Masas de desorción de plasma por bombardeo con 252Cf, totalmente automatizado que, tanto en sus versiones iniciales como mas tarde hasta el inicio de este siglo seguía operativo para la determinación directa de los espectros de masas de proteínas intactas (75). En segundo lugar, porque fue entonces cuando John Fenn y colaboradores describieron el uso del electrospray para ionizar moléculas de gran tamaño (76, 77). Según lo indicado por estos autores, los primeros resultados que obtuvieron con el electrospray (ESI), a la luz de los trabajos experimentales anteriores de Dole en 1968-1973 y de Iribarne y Thomson de 1976 a 1982, representaron un enfoque muy prometedor para el análisis de muestras que contenían moléculas demasiado grandes, demasiado complejas, no volátiles o frágiles para ionizarlas por métodos convencionales.

El mecanismo de ionización por electrospray aún se está debatiendo, pero en esencia se puede describir por dos modelos: el modelo de Carga Residual de Dole basado en la división continua de pequeñas gotas supercargadas hasta que queda un único ion en la microgota y el Modelo de Evaporación de Iones de Iribarne y Thomson basado en eyección directa de iones a partir de una gota supercargada (78). John Fenn publicó un interesante artículo de revisión en el Journal of Biomolecular Techniques en 2002, que es de lectura recomendada (22) sobre los recuerdos históricos y personales de los pros y los contras de “cómo empezó todo” para el ESI.

Los trabajos de Yamashita y Fenn en 1984 sentaron las bases para el desarrollo de una interfaz efectiva y “fácil de usar” entre CL y EM (79), donde mediante el uso del cuadrupolo de VG Micromass 1212 de un rango de masas más alto, los autores pudieron demostrar la formación de iones MH^+ a una relación m/z de hasta 1203 para la ciclosporina A y m/z 1141 para la gramicidina S. Sin embargo, la realización de todo el potencial de la CL-ESI/EM aún requeriría algún tiempo hasta que en 1988 se demostró la carga múltiple de proteínas de gran tamaño (80). Cabe también señalar que la existencia de iones múlticargados (1^+, 2^+, 3^+) ya había sido demostrada anteriormente por el grupo de Iribarne y Thomson para la ATP en la reunión de la ASMS de 1980 en Boston (T. Covey, comunicación personal). En consecuencia, aunque en 1984 estábamos ya en el camino hacia los métodos de ionización más universales y suaves y hacia interfases para LC-MS fáciles de usar, la técnica del Termospray, como se indicó anteriormente, dominó la década de los 80 como el enfoque más práctico para el acoplamiento directo de la CL y la EM.

22. Resumen de los primeros 20 años.

Como hemos visto, el período de 20 años antes de 1985 estuvo claramente marcado por importantes avances técnicos y fundamentales en la EM, así como por el uso de esta técnica en el estudio de proteínas y péptidos, entre otras moléculas biológicamente relevantes. En esos primeros días, pocos de los que trabajábamos en EM biológica por impacto electrónico (EI) o por ionización química (IQ) hubiésemos anticipado la relativa facilidad con que, en un futuro no muy lejano, podríamos ionizar macromoléculas. Después del debut del PDMS y el FAB, el progreso fue rápido y pronto las nuevas técnicas de ionización habían hecho factible la ionización de moléculas pesadas no volátiles y térmicamente frágiles de hasta 10.000 Da de peso molecular. La base fundamental de este cambio fue que, si bien el IE y la IQ requieren la previa vaporización de la muestra, ya que ambos son procesos en fase gaseosa, las técnicas de ionización más recientes se basaban en un proceso que permite la desorción de iones desde una fase condensada. Un buen resumen y el cronograma de las técnicas de desorción vigente hasta la introducción del FAB en 1981 se encuentra en una publicación de K.L. Busch y R.G. Cooks (81). También digno de mención es el resumen de Richard Kerr sobre el incremento del límite de peso en EM, publicado en Science en abril de 1982 (82).

Puesto que el alcance del FAB se vio limitado por sus características de muestreo estático, lo que impide el muestreo dinámico de flujos líquidos, en diciembre de 1986 Richard Caprioli en colaboración con John Cottrell en Kratos Analytical Instruments publicó la primera descripción de una sonda de muestreo de flujo continuo para FAB que podía montarse fácilmente en la mayoría de los instrumentos comerciales (83). Véase más adelante.

23. Nuevos instrumentos de FTICR.

En búsqueda de la máxima capacidad de análisis de masas, combinada con la mayor sensibilidad posible, D. Hunt y cols, en las universidades de Virginia y California y en colaboración con Finnigan MAT-Corp. construyeron en 1985 un espectrómetro de masas de tándem cuadrupolar y de transformada de Fourier equipado con un imán 7 Teslas y una fuente de iones de cesio para el bombardeo de las muestras no volátiles disueltas en una matriz líquida volátil (84). En este instrumento, la introducción de la muestra y su ionización se realizó en una fuente de iones evacuada diferencialmente y el análisis fue posible por transferencia de los iones seleccionados a la celda del ICR. La ventaja de este novedoso sistema híbrido QqFTICR es que fue capaz de proporcionar un rango de masas de más de 10,000 Da y también que podía acumular iones de baja abundancia a través de sus funciones como dispositivo de almacenamiento. Con este enfoque, los autores presentaron un espectro de masas de melitina, un péptido de 26 residuos con un peso molecular de 2844. Aunque, en principio, la célula ICR podría haber proporcionado un rango de masas de hasta 10,000 Da o más, el alcance práctico del prototipo montado se limitó a un m/z de 3,000 con baja resolución y sensibilidad.

Al mismo tiempo y con el tándem comercial de Nicolet, el FTMS-1000, equipado con un imán 3 Teslas y un láser de infrarrojos para la ionización de la muestra por desorción, McLafferty y cols. en la Universidad de Cornell presentaron espectros de FTMS de los péptidos lineales y cíclicos de la gramicidina D y gramicidina S, respectivamente (85). A través de los iones $[M + K]^+$ de masa 1881-1934 demostraron que la Gramicidina D era una mezcla de cuatro componentes alcalinos. Los autores separaron los cuatro iones moleculares en MS1 y obtuvieron datos de secuencias en MS2. El CAD selectivo de los iones 1920-1921 dio un espectro con picos de secuencia. Esto puede citarse como un ejemplo de la utilización del concepto aún no definido de caracterización de proteínas por espectrometría de masas 'top-down' (ver sección sobre "Las nuevas tecnologías y mecanismos de ionización"). Como veremos más adelante, la proteómica

basada en la MS se ha llevado a cabo hasta la fecha mediante el uso de cualquiera de los llamados enfoques «de abajo-arriba" o "de arriba-abajo" ('bottom-up' o 'top-down'). En 'top-down, la proteína o péptido intactos se fragmentan en su totalidad en el espectrómetro de masas, sin digestión proteolítica previa, en contraste con el enfoque de bottom-up centrado en una primera digestión de la proteína o péptido a una mezcla en péptidos cortos que luego son analizados por EM. El desarrollo de enfoques "bottom-up" se vio obligado en su momento porque las proteínas y péptidos no eran fácilmente vaporizables, pero si podían recuperarse fácilmente de un gel de SDS después de la proteólisis. Tampoco en esta ocasión estos autores presentaron una prueba de la viabilidad de un trabajo a gran rango de masas. Ambos grupos, de Hunt y McLafferty, colaborarían más tarde en el acoplamiento del ESI al FTMS (ver más adelante).

Modificaciones posteriores del instrumento QqFTICR diseñado por D. Hunt y cols. (86) produjeron un aumento de la velocidad de bombeo y de la eficiencia de captura de iones. Ello amplió el rango de masas útil más allá de la masa 10,000, lo que permitía la caracterización de la masa molecular de oligopéptidos y proteínas pequeñas, como la insulina o el citocromo c (12.384 Da) a nivel de 100 a 200 pmol (Figura 9).

El diseño de este instrumento híbrido debe considerarse un importante avance para la época, ya que el electroimán de alto campo y los instrumentos en tándem de cuatro sectores no habían sido capaces de superar la masa 6000 (87) por los problemas de baja transmisión de iones y de baja resolución a estas masas altas. Sin embargo, el uso de FTICR para el trabajo biológico real con biopolímeros grandes a alta resolución y sensibilidad no se consiguió hasta principios de los años 90, como se verá más adelante, y los instrumentos comerciales de Bruker, Ionspec y Thermo Electrón no aparecerían en el mercado hasta pocos años después (ver Tabla 2, Capitulo 7).

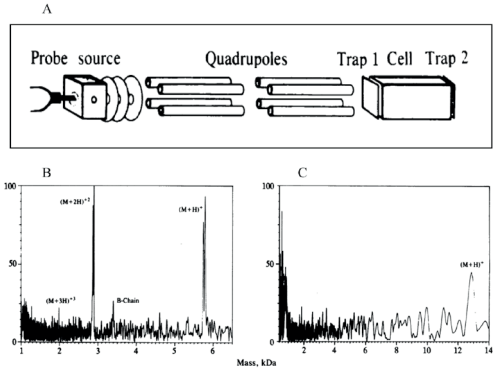


Figura 9. (A). Diagrama del espectrómetro de masas de quadrupolo y transformada de Fourier en tándem. Espectros de masas de iones secundarios. (B). Insulinas bobina (PM = 5733.5) y porcina (PM = 5777.6) (200 picomoles de cada una). (C). Citocromo *c* equino (PM = 12384). [Reproducido con el permiso de la National Academy of Sciences, USA.: ver Ref. [86], Copyright (1987)].

24. Electrospray LC-MS.

Como ya se indicó, las publicaciones de Yamashita y Fenn en 1984 sobre la fuente de iones de electrospray sentaron las bases de una interfaz de LC-MS de fácil utilización. Un año después esto se materializó en la publicación de Whitehouse y cols (88) que describía el uso de esta nueva interfaz para el acoplamiento CL y EM. En esta publicación los autores señalan que la ionización por electrospray comparte un mecanismo común de "desorción de iones por campo" a partir de líquidos con otras fuentes de iones existentes, como las de ionización por campo (FI), desorción por Campo (FD) ambas de Beckey y cols, la evaporación de iones a presión atmosférica (APIE) de Iribarne y Thomson (o Aerospray, como se conocería más adelante) y el Termospray (TSP) de Vestal y cols. También señalan que, si bien el ESI es operacionalmente

diferente del TSP y la APIE, ya que provoca la atomización de la gota de líquido por inducción de cargas eléctricas, el TSP y la APIE producen cargas por atomización, pero las tres utilizan un jet supersónico para el transporte de los iones desde la región de desorción a la región de alto vacío del espectrómetro de masas. Por tanto, estos autores apropiadamente afirman que, "el mismo jet supersónico que ayudo a la unión de la EM con la CG en 1965, 20 años más tarde facilitaba la unión con la CL".

No obstante, esta nueva interfaz no era más que una promesa cuando se introdujo por primera vez en 1985 y proporcionó a los autores una limitada experiencia en el acoplamiento CL-EM y acceso relativamente fácil a iones moleculares protonados como los de la adenosina, AMP, ADP, gramicidina S y la ciclosporina A. Esto no era excepcional en 1985, excepto por que el ESI parecía capaz de generar iones con alta eficiencia a partir de moléculas de gran tamaño, en relación a otras técnicas de ionización, como la EM de desorción por plasma (PDMS) que utilizaba iones de 252Cf a 100 MeV, el FAB con átomos de xenón o la EM de iones secundarios en fase líquida (LSIMS) que usaba iones cesio a 8 keV para bombardear la muestra, todas las cuales fueron capaces de proporcionar acceso a masas de hasta 10-14 kDa. En 1987, el rango útil de masas se amplió a pequeñas proteínas con pesos moleculares de hasta el 24 kDa. Este logro fue posible gracias a la colaboración entre Michael Barber en UMIST y Green Brian en VG Analytical quienes, mediante el uso de partículas primarias de alta energía en una fuente de iones cesio de 35 keV instalada en un VG-SE ZAB, fueron capaces de alcanzar estas masas con la suficiente precisión y sensibilidad. Este interesante trabajo, que extendió los límites de la EM de sector se publicó en septiembre de 1987 en RCM (89) y también ese mismo mes Brian Green lo presentó en el Applied MS Meeting in the Health Sciences en Barcelona.

25. Inicios de la Desorción por Láser Asistida por Matriz (Matrix Assisted Laser Desorption).

En 1985 ocurrió otro desarrollo interesante y con potencial para cambiar el futuro, aunque en ese momento no recibió mucha atención. Esta fue la primera manifestación del fenómeno de la "desorción por láser asistida por matriz" y la primera introducción de esta terminología en la literatura (90). Los autores, en un estudio sobre la influencia de la longitud de onda de láser sobre la desorción de iones a partir de AAs y dipéptidos, observaron que una mezcla de triptófano y alanina producía un espectro de ambos a pesar de que, al flujo de energía láser que estaban utilizando, sólo se esperaba visualizar el triptófano. Así, el triptófano absorbía la energía láser y ayudaba a ionizar la alanina no absorbente. Esta observación de la co-desorción de alanina (Figura 10) fue la primera observación de un efecto de la matriz a la longitud de onda de 266 nm y por lo tanto el MALDI nació para indicar una ionización suave por desorción con láser asistida por un compuesto matriz diferente.

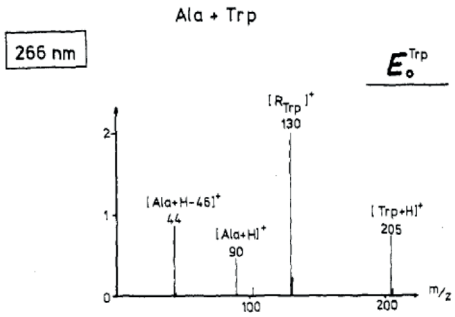


Figura 10. Espectro de Masas obtenido por LD de una mezcla 1:1 de alanina y triptófano a 266 nm y irradiancia límite para el triptófano. [Reproducido con el permiso de la American Chemical Society: ver Ref. [90], Copyright (1985)].

26. Primeras bases de datos.

En aquel momento, todos las esfuerzos de secuenciación de péptidos y proteínas señalaban la necesidad de reunir y anotar los datos de secuencias en bases de datos de referencia, como ya inició M. Dayhoff en 1965 (33). Los investigadores tenían dificultades para lograr la plena identificación de secuencias de fragmentos de las proteínas digeridas si la secuencia de consultas no podía compararse con todas las secuencias en una base de datos especificada y de libre acceso a la comunidad académica.

El 21 de julio de 1986, apareció la Swiss-Prot. Contenía cerca de 4,000 entradas de secuencia de proteínas y su autor fue Amos Bairoch, un estudiante de postgrado del Depto. de Bioquímica Médica de la Universidad de Ginebra y desde 1987 se mantuvo como colaboración entre el Departamento y el European Molecular Biology Laboratory (EMBL). Ahora es una colaboración de Instituto Suizo de Bioinformática (SIB) y el Instituto Europeo de Bioinformática (EBI) (91).

27. Primeras secuencias completas de dos proteínas.

De la literatura de aquella época cabe citar dos artículos publicados en septiembre de 1986, relativos a la determinación de secuencias de proteínas por métodos de EM. Uno de ellos del laboratorio de D. Hunt, de la Universidad de Virginia, describió la secuencia de los péptidos de la apolipoproteína B, la separación por HPLC de los péptidos digeridos y la EM de iones líquidos secundarios en un triple cuadrupolo de las fracciones recogidas preparativamente. La disociación activada por colisión de los iones protonados resultantes generaba iones característicos de la secuencia de aminoácidos en el péptido. Los autores reconocieron que este enfoque era limitado por el rango de masas de m/z 1800 del espectrómetro de masas. El estudio del digerido de una sola proteína requería una media de 4-5 días de trabajo, proporcionando información de la secuencia a nivel 10nmol para el 50-60% de una proteína de 50-kDa (92). Esto puede citarse como un ejemplo importante de la EM “de abajo-arriba» puesto que la proteína se digiere a péptidos más pequeños que luego son separados por HPLC y secuenciados de forma individual por EM. La otra colaboración fue la de John Crabb, Steve Carr y Bob Bordoli y cols en Alton Jones Cell Centre, y los laboratorios SmithKline and French y V.G. Analytical, respectivamente. Estos autores, utilizando espectros de masas convencionales, obtenidos por FAB en un espectrómetro de masas de doble enfoque VG ZAB 1F-1F y un espectrómetro de masas tándem de cuatro sectores VG ZAB SE-4F describieron la estructura completa de la forma principal de la prostatropina de cerebro bovino con terminal acetilado y señalaron que la información hubiera sido difícil de obtener por otras técnicas (93). Asimismo, el grupo de Klaus Biemann en la edición marzo de 1987 describió en Biochemistry su planteamiento tándem de alto rendimiento que, en un JEOL de cuatro sectores HX110/HX110, consiguió la secuenciación completa de la tiorredoxina por FAB MS/MS (94). Estas dos proteínas fueron las dos primeras secuenciaciones completas por EM.

28. Tiempo de vuelo en tándem. Tandem Time-of-Flight.

Ese mismo año Schey y Cooks en Purdue, en colaboración con Grix y Wöllnik de la Universidad Justus Liebig en Alemania, publicaron la primera descripción de una nueva clase de espectrómetro de masas. Se trataba de un instrumento tándem de Tiempo de vuelo (TOF-TOF) que incorporaba la disociación inducida por superficie (95). Digno de mención también como un hito del 1987 es el primer uso de una trampa de iones para MS/MS. Esto fue el resultado de un trabajo de colaboración entre Louris y Cooks en Purdue, Syka, Kelley y Stafford de Finnigam Mat en San José, California y John Todd en la Universidad de Kent (96). Estos autores describieron la generación de espectros de iones producto a partir de iones preseleccionados, tras expulsar los iones no deseados de la trampa. A continuación, los iones preseleccionados se sometieron a procesos muy eficientes de CAD para generar iones fragmento diagnósticos. Aunque el trabajo se hizo con moléculas pequeñas, como el nitrobenceno, en ese momento representó un avance fundamental con grandes consecuencias prácticas para el trabajo posterior sobre la química de proteínas.

29. La llegada definitiva del electrospray y del MALDI.

También en 1987, Michael Karas y colaboradores publicaron un interesante manuscrito sobre desorción por láser UV (LD) asistida por matriz de compuestos no-volátiles como la estaquiosa, un tetrasacárido de Mr 666, la eritromicina, un antibiótico de Mr 733, la vitamina B12 de Mr 1355 y la melitina, un polipéptido de 26 AAs de Mr 2843 (97). Al igual que en FAB y SIMS, cuyos éxitos en el momento se debieron a la utilización de glicerol líquido como matriz, estos autores afirman que teniendo en cuenta sus propias experiencias y las de otros autores sobre la aplicación de la EM de desorción por láser a muestras no volátiles sería razonable comprobar si una matriz resonantemente absorbente podría comportarse mejor que las matrices utilizadas anteriormente y producir desorción “muy suave”, incluso para moléculas que no absorbían a la longitud de onda utilizada. Ellos demostraron en aquel trabajo la importancia de la matriz, además de la fluencia del láser y señalaron que este trabajo abría nuevas perspectivas para la LD.

Sin embargo, el verdadero progreso en el análisis de macromoléculas no volátiles, así como el nacimiento de una era totalmente nueva en EM había de llegar al año siguiente con la presentación pública de dos de los avances más importantes de la historia de la EM. Por un lado, John Fenn y colaboradores describieron en 1988 el fenómeno de la carga múltiple en polietilenglicoles (PEG) (98) demostrando que el ESI podía depositar hasta 23 iones Na^+ sobre moléculas de PEG con masas moleculares del orden de 17,500. Este trabajo, publicado el 28 de enero en el Journal of Physical Chemistry, demostró la posibilidad de analizar por primera vez iones con masas de 20 kDa, aunque el límite nominal superior del filtro de masas del cuadrupolo utilizado por los autores era sólo de 1500 Da, ya que el analizador de masas aumenta el rango en un factor igual a la cantidad de cargas por ion. Fue en efecto una observación experimental crucial, aunque no obstante, se vio afectada por las limitaciones prácticas debido a la amplia distribución de los pesos moleculares de los oligómeros de PEG que derivó en la imposibilidad de resolver los diferentes picos de masa para lograr su asignación inequívoca.

Por un breve lapso de sólo cuatro meses este trabajo quedo relativamente ignorado hasta que Fenn en junio de 1988 presentó los primeros resultados obtenidos sobre proteínas intactas en la 36 Annual Conference on Mass Spectrometry and Allied Topics, celebrada en San Francisco, CA. Poco después de la reunión, estos nuevos y revolucionarios resultados aparecieron

en junio de 1988 en el Zeitschrift für Physik D - Atoms, Molecules and Clusters (99). En esta publicación el grupo extendió observaciones anteriores a moléculas más pesadas, pero puras, como la deshidrogenasa del alcohol del hígado equino con un peso molecular de 39.830 Da y con capacidad para acomodar hasta 46 cargas por molécula. Además, la contribución específica a esta obra de Matthias Mann, entonces un joven estudiante en el grupo de J. Fenn en Yale, se reflejó en el desarrollo de un algoritmo informático para promediar las señales y lograr la deconvolución computacional del espectro de ESI, que se publicó en agosto de 1989 (100). De esta manera el espectro de ESI se transforma de una curva en forma de campana de Gauss que aloja los múltiples picos de distinta carga en un solo pico cargado que aparece a la masa molecular de las especies ionizadas. En esta publicación se mencionó que el algoritmo era capaz de extraer datos de cada especie cuando la disolución que se inyectaba era una mezcla compuesta de dos solutos. Una de las primeras aplicaciones publicadas de este algoritmo se muestra en la Figura 11 (101). En esta publicación, John Fenn y cols. comunicaron espectros de masas de proteínas representativas con PMs que iban desde 5,000 hasta 76,000.

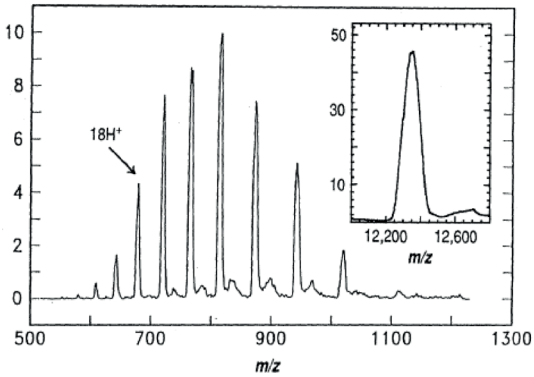


Figura 11. Aplicación del algoritmo de deconvolución a un espectro de citocromo *c*. Figura insertada: pico deconvolucionado resultante de la aplicación del algoritmo. [Reproducido con el permiso de la AAAS: ver Ref. [101], Copyright (1989)].

Por otra parte, poco después de la 36 Annual Conference on Mass Spectrometry and Allied Topics, Karas y Hillenkamp presentaron en la 11 ª International Conference of Mass Spectrometry celebrada en agosto-septiembre en Burdeos, Francia, su primer espectro de masas de una enzima de masa a 117,000 Da, obtenido por desorción con láser UV (Figura 12). Además, Karas y Hillenkamp, habían presentado el 16 de mayo un trabajo a Analytical Chemistry, que fue aceptado el 5 de julio y salió en el numero de octubre (102). En esta publicación, los autores demostraron la desorción por láser ultravioleta (UVLD) de cuatro proteínas (la lisozima, la β-lactoglobulina, la tripsina y la albúmina, con PMs desde 14,306 hasta 67,000) con ácido nicotínico como matriz absorbente. Al igual que en ESI y como señalaron los autores del trabajo, la ionización de desorción por láser también produce picos doble y triplemente cargados en los espectros. El progreso en la aplicación y mejora de la técnica fue rápido y en junio de 1989 los mismos autores publicaron un informe que mostró un espectro mucho más mejorado de β-D-galactosidasa (Figura 12), junto con los espectros, obtenidos a nivel femtomolar, de la isomerasa de glucosa con un Mr medido de 172,420 y la catalasa con Mr de 236.230 (103). Para una valoración global del nacimiento del MALDI ver el artículo de revisión (104) y posterior recapitulación histórica publicado por F. Hillenkamp y M. Karas (105).

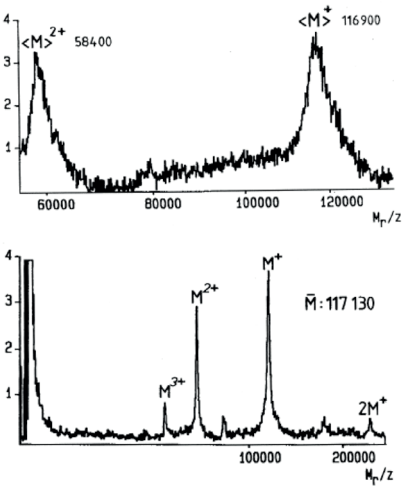


Figura 12. Espectro MALDI de la β-D-Galactosidasa obtenido con acido nicotínico como matriz y utilizando una longitud de onda del láser de 266 nm. (Arriba) Primer espectro presentado en la IMSC de Burdeos en 1988. Suma de 200 disparos de laser. (Abajo) El mismo espectro obtenido unas pocas semanas mas tarde. Suma de 30 disparos. [Reproducido con el permiso de Elsevier: ver Ref. [105], Copyright (2000)].

Además, entre la presentación de ESI por Fenn en San Francisco y los espectros de masas de MALDI por Hillenkamp y Karas en Burdeos, Koichi Tanaka y sus colaboradores en Shimadzu Corporation en Japón publicaron en Agosto de 1988 su descripción de un nuevo espectrómetro de masas TOF con ionización por láser (Shimazu LAMS-50K), que fue aceptada por RCM el mismo día en que se recibió (6 de junio) (106). Este instrumento, junto con su nuevo método de preparación de muestras para ionización por láser, conocido como el "método de la matriz plus ultra fina de metal líquido (UFP) permitió el análisis de proteínas y polímeros de m/z hasta 100,000. El papel fundamental de la matriz en estos experimentos, como se verá más adelante, fue reconocido adecuadamente por estos autores, quienes señalaron que era imposible detectar iones a masa alta sin la UFP.

Aunque Hillenkamp y Karas (104), en su publicación histórica citan el desarrollo del primer instrumento comercial de MALDI-TOF por la corporación Vestec de Marvin Vestal, basado en un trabajo previo de Ron Beavis y Brian Chait, parece que, como afirma Shimadzu, el primer instrumento comercial de MALDI-TOF fue el LAMS-50K de Shimadzu. (Véase http://www.shimadzu.com/about/oh80jt0000002x31.html). Por lo que he podido averiguar, Beavis y Chait publicaron un trabajo (107) en julio de 1989, reproduciendo los primeros espectros de MALDI presentados un año antes por Hillenkamp y Karas (102) que obtuvieron en un instrumento TOF lineal simple y por lo tanto el reclamo de Shimadzu es probablemente correcto ya que la introducción comercial de un espectrómetro de masas MALDI-TOF por Vestec se produjo en agosto de 1991 (véase más adelante).

La diferencia fundamental entre el trabajo de Tanaka y el de Karas/Hillenkamp, aparte de las diferencias en el momento de la presentación y aceptación de sus artículos, se encuentra en el uso de una muestra depositada en un líquido (mezcla de glicerol más nanopartículas de cobalto) frente a una matriz cristalina (ácido nicotínico desecado). Por ello, la concesión posterior del Premio Nobel de Química a K. Tanaka en 2002 por la ionización de biomoléculas de grandes dimensiones, utilizando la ionización por desorción con láser asistida por matriz condujo a una controversia que no ha sido claramente resuelta hasta la fecha.

El trabajo de Fenn con ESI se llevó a cabo en un filtro de masas de cuadrupolo de VG Micromass 1212 con un límite de m/z nominal superior a 1500, mientras que el trabajo Karas y Hillenkamp sobre UVLDMS se llevó a cabo en un espectrómetro de masa de tiempo de vuelo (TOF) con reflector, diseñado inicialmente para el análisis de compuestos inorgánicos y orgánicos por microsonda en el rango de masa baja (LAMMA 1000). Por lo tanto, a los TOF de rango de masas elevado se les unieron, por las características de carga múltiple del ESI, los cuadrupolos simples que, en su capacidad para determinar la masa de las moléculas biológicas, acapararon el centro de atención. Sin embargo, los cuadrupolos coexistieron por un tiempo con los instrumentos magnéticos, como se vio en la exposición en Burdeos con el anuncio del espectrómetro de masas sector de alta resolución (200.000) "más grande del mundo", el JMS HM 1000 (geometría EBEB). Este era un espectrómetro de masas en tándem de cuatro sectores equipado con un detector matricial y con un amplio rango de masa de casi 48,800 a voltaje de aceleración total. También en la reunión de Burdeos VG expuso por primera vez el popular Autospec, un nuevo sector triple de doble enfoque automático (EBE) de alto rendimiento con tecnología de imán superlaminado y fuentes 'Plug in' para un cambio rápido de un modo de ionización a otro.

30. Diferencias entre métodos de ionización.

En este punto de la evolución histórica de las nuevas técnicas de ionización para macromoléculas sería conveniente considerar las diferencias fundamentales entre el uso de campos

electrostáticos frente al bombardeo con partículas para la desorción de iones. Como se ha comentado los primeros métodos “blandos” de ionización fueron la desorción por campo y la ionización electrohidrodinámica (EHD), descritos por primera vez en 1974 para compuestos orgánicos inestables al calor. El método de EHD produce gotas cargadas y/o iones directamente en vacío a partir de la interacción de un campo electrostático fuerte con un menisco líquido (glicerol) en el extremo de un tubo capilar. Esto, aunque no atrajo mucha atención en su momento sin duda anunciaba nuevos e interesante logros por llegar en este ámbito, tanto en el uso del glicerol como una matriz líquida como en el uso de campos electrostáticos en tubos capilares. Así, a principios de los años 80, estos métodos dieron paso a una nueva familia de técnicas de ionización también basadas en la utilización de campos eléctricos para la desorción de iones desde pequeñas gotas de líquido cargadas a la atmósfera en lugar de al vacío. Me refiero al Termospray, el Aerospray y el electrospray.

Los otros métodos de ionización suave en fase de desarrollo en los años 70 y principios de los 80 derivan todos de la idea propuesta en 1974 por Beuhler y cols. de lograr una súbita vaporización por aporte rápido de energía antes de que tenga lugar la descomposición (108). Por lo tanto, la deposición rápida de energía fue posible gracias a la PD, el SIMS, el FAB o la LD. Estas técnicas “violentas” o de “ energía repentina “ para la ionización de moléculas de la muestra dispersas sobre una capa de un sustrato inerte proporcionaron, en octubre de 1989, mediciones de moléculas de Mr hasta 210.000 por LD, de 24.000 por FAB y 45.000 por PD.

Justo en el momento en que estas técnicas de ionización suave hacían su debut, se estaban analizando en línea digeridos proteolíticos mediante la ingeniosa combinación de las técnicas de HPLC con FAB-MS, según había descrito un par de años antes el grupo de Richard Caprioli. Caprioli y sus colaboradores, entonces en el Texas Medical School en Houston publicaron su trabajo sobre la interfaz de flujo continuo FAB para el acoplamiento directo columnas microbore y capilares de sílice fundida empacadas con C18 (109).

31. Ion Spray y el principio de la moderna LC-MS.

Otro tema que debe situarse correctamente en el contexto de los estudios sobre el electrospray y las nuevas interfaces de CL-EM es el trabajo paralelo que llevó al desarrollo del “Ion Spray”, como describieron en 1987 Jack Henion, Tom Covey y Andries Bruins en la Universidad de Cornell. Estos autores propiciaron el desarrollo de la Ionización a Presión Atmosférica (API) basada en el trabajo anterior sobre APCI/EM por el grupo de Evan Horning en Houston y descrito en diciembre 1975 en Analytical Chemistry (110). En este enfoque los analitos más el disolvente se evaporaban en un bloque caliente, no al vacío, sino a presión atmosférica, y el disolvente se utilizaba como reactivo de ionización química. Sin embargo, debido a la falta de la fragmentación no despertó interés, hasta que en 1981 pudo aplicarse CAD MS/MS gracias a los cuadrupolos triples. Basándose en estas ideas y utilizando un Sciex TAGA 6000 (Sciex, Thornhill, Ontario), Bruce Thomson describió la interfaz de nebulizador térmico para CL-EM en la 31 ª Annual Conference on Mass Spectrometry and Allied Topics, celebrada en Washington en 1983.

Este dispositivo surgió del trabajo realizado el año anterior por Henion, Thomson y Dawson quienes habían acoplado una fuente de DLI y la fuente de APCI de Sciex para un trabajo preliminar pero innovador de CL-EM (111). La rápida vaporización del eluyente y de los analitos reducía o evitaba la descomposición térmica y esta característica fue bien explotada ya en 1986 por Covey, Lee y Henion que demostraron la viabilidad del análisis rápido de compuestos polares, lábiles y de bajo Mr para un gran numero de muestras y señalaron la necesidad del desarrollo de sistemas automáticos de LC-MS/MS “de sobremesa”, anunciando así la llegada de las

técnicas modernas de CL-EM (112). Toda esta actividad finalmente dio lugar a la introducción del "Ion Spray" en 1987, que se basaba en el principio de adición de un gas de nebulización al electrospray y abrió la posibilidad de trabajar a flujos compatibles con los de la CL (113). Al igual que en electrospray, el ion spray producía iones en fase gas a través del mecanismo de evaporación de Iribarne y Thomson, ocurriendo la nebulización y la ionización de la muestra a presión atmosférica en un caudal de líquido. El uso de la nebulización neumática, además del campo eléctrico, mejora la estabilidad de la corriente de iones en una amplia gama de caudales líquidos y permite el uso de fases móviles de mayor tensión superficial en CL-EM.

Aunque el trabajo del grupo de Jack Henion no involucraba directamente a péptidos o proteínas en su desarrollo, he considerado necesaria su inclusión en el marco de los principales hitos de la EM de los años 80. Hay que tener en cuenta que este trabajo proporcionó las bases para las técnicas que más tarde serían ampliamente aceptadas y utilizadas en proteómica y me refiero a LC-MS/MS y Ion Spray y que comercialmente se materializó en instrumentación especifica como el primer APCI API III/MS/MS fabricado por Sciex en 1989 o los más modernos Sciex API 3000, Sciex Q Star Pulsar, Micromass Q- TOF, PE/Sciex API 365, etc. (ver Tabla 2, Capitulo 7).

Teniendo en cuenta todos los avances en materia de ionización a presión atmosférica predominantes en 1989 se puede observar que la fuente de iones Sciex API III lideró el camino hacia la aceptación de las técnicas API de CL-EM por la comunidad analítica. Sin embargo, en lo que respecta a la caracterización por EM de péptidos y proteínas, la técnica más generalizada en aquel tiempo aun era el FAB, aunque esto estaba a punto de cambiar en unos 4 años.

32. Nueva instrumentación y aplicaciones para MALDI-TOF y ESI.

En ese momento, Jim Dawson y Michael Guilhaus propusieron un espectrómetro de masas de tiempo de vuelo con aceleración ortogonal como una nueva forma de acoplamiento del TOF a fuentes de producción continua de iones (114). Aunque el desarrollo del reflectrón en 1973 había dado un impulso muy necesario para el aumento de la resolución de masa en los instrumentos TOF, no era posible utilizar en rutina todas las fuentes convencionales de ionización continua. En la práctica esto se resolvió con la introducción del principio de aceleración ortogonal (TOF-oa). Así, el nuevo prototipo TOF fue capaz de obtener espectros con una fuente convencional de IE. Se ha afirmado que el renacimiento del TOF se debió primero a la llegada del MALDI y segundo al "redescubrimiento" de la aceleración ortogonal por estos autores, que en su revisión publicada en 2000, reconocen que "en mayor medida la reducción del desarrollo de la EM de 4 sectores para MS/MS se debió, posiblemente, a que los híbridos de sector o cuadrupolo estaban equipados con un sistema de TOF-oa como MS2. La excelente sensibilidad, velocidad, resolución, exactitud de masa y estabilidad de calibración masas de los instrumentos TOF-oa que se introdujeron comercialmente en la década de los años 90 proporcionó una mejora significativa sobre los instrumentos tándem de 4-sectores para el análisis de iones producto generados por "disociaciones inducidas por colisión" (115). Una declaración como esta también podría ampliarse, según la opinión de Graham Cooks (comunicación personal), para incluir a los triples cuadrupolos que probablemente tuvieron una influencia comparable sobre el uso cada vez menor de la EM de sector magnético.

En 1989 la mitad del Premio Nobel de Física fue compartido por Hans G Dehmelt y Wolfgang Paul "por el desarrollo de la técnica de la trampa de iones", que en ese momento era ya un elemento de poyata de laboratorio comercializado por Finnigan MAT desde 1983. Otro aconte-

cimiento digno de mención fue la primera publicación del acoplamiento de ESI a FTMS para el estudio de una variedad de péptidos y proteínas con pesos moleculares de hasta 29 kDa (116).

En abril de 1990, Finnigan Corporation anunciaba su nuevo TSQ 700 como un instrumento de alto rendimiento, basado en el exitoso TSQ 70, y en la Conferencia de 1990 en Pittsburgh esta misma empresa comunico la introducción del MAT 95, un instrumento de alta resolución de sector magnético, con un rango de masas de 3500 y resolución de más de 60.000 (a 10% de valle).

Además, en abril de ese año, la fragmentación directa de una proteína intacta de peso molecular 13 kDa por ESI tándem MS de cuadrupolo, fue una contribución top-down histórica como demostró el grupo de Richard Smith (117). La EM en tándem con la asistencia de los procesos de CID se estaba utilizando para la determinación de la estructura primaria de oligopéptidos e, indirectamente, proteínas, pero nunca para la secuenciación directa de una proteína intacta. Los autores describieron el CID tándem MS en un Sciex TAGA 6000 de una ribonucleasa A intacta multiplemente protonada con un Mr de 13,682 y 124 residuos de AAs. El ESI indujo hasta 15 cargas sobre la molécula protonada, un número que se elevó a 23 cargas tras la reducción de los cuatro enlaces disulfuro de la proteína. Los espectros de masas CID proporcionaron fragmentaciones específicas asignables a la proteína intacta, lo que ahorraba tiempo de secuenciación en un factor igual al número de fragmentos peptídicos generados por digestión proteolítica para su posterior análisis por MS/MS.

En 1990 también vimos el informe de Nohmi y Fenn sobre la generación y medida por ESI de varios iones de oligómeros intactos de óxido de polietileno con hasta 4000 cargas y con pesos moleculares de 5.000.000 (118), así como la primera demostración de la utilización de ESI-MS para estudiar la conformación de una proteína en disolución. En este trabajo pionero, los autores utilizaron la dependencia del mecanismo de captación de protones en electrospray sobre la accesibilidad a los centros básicos ionizables del citocromo *c* de bovinos y caballos, así como de la ubiquitina bovina y de la levadura. La idea era que una proteína en una conformación muy plegada tendría menos sitios disponibles para la protonación y por lo tanto los espectros ESI reflejarían los cambios en la conformación de proteínas a diferentes valores de pH (119). Adicionalmente, en 1991, Bruce Ganem y Jack Henion en Cornell demostraron la aplicabilidad del ESI-MS para el estudio de complejos no covalentes (120, 121) y Viswanatham Katta y Brian Chait informaron sobre la unión no covalente del heme a la mioglobina (122).

En agosto de 1991 Vestec Corporation anunció la introducción comercial de un espectrómetro de masas MALDI-TOF (Vestec LASERTEC 2000), un prototipo del que había sido entregado un año antes a Klaus Biemann. También en dos publicaciones de los laboratorios de Brian Chait, y de Tom Covey y Jack Henion se señaló la importancia del CID de baja energía para la secuenciación de péptidos, tanto en cuadrupolos simples como triples. Así, en enero de ese año, Brian Chait y cols. publicaron una descripción de la utilización de un cuadrupolo simple para los estudios por CID de iones péptidicos con cargas múltiples, como las bradicininas, la angiotensina II, el fibrinipeptido(Glu-1) y un inhibidor sintético de la renina (123). Estos autores ampliaron trabajos anteriores que habían demostrado la posibilidad de secuenciación de péptidos por CID de iones con doble carga obtenidos por FAB o SIMS. Para ello, utilizaron un modelo modificado de espectrómetro de masas Vestec 201 de ionización por Termospray, adaptado para electrospray según el diseño de una nueva fuente de iones de electrospray que habían publicado un año antes. La activación colisional se llevó a cabo en la región de alta presión entre la salida capilar del electrospray y la entrada del skimmer en el analizador de masas del cuadrupolo simple. Esta idea resultó popular en laboratorios con acceso a la EM limitado a instrumentos simples y de bajo costo. En esta línea, y sobre la base de un trabajo de Tom Covey y cols. presentado en 1989 en la 37 ª ASMS, en el que habían demostrado que como los péptidos trípticos contienen, por lo general, un grupo amino terminal y una base de arginina o lisina en el extremo carboxilo, con frecuencia producen por ion spray iones moleculares doblemente cargados, el mismo Tom Covey en Sciex en colaboración con Huang y Henion de la Universidad de Cornell describieron,

en julio de 1991, la caracterización estructural de proteínas por CID de iones con doble carga utilizando una fuente Sciex API III y el triple cuadrupolo TAGA 6000 equipado con interfaz de ion spray para el análisis en línea por CL-MS/MS (124).

En la misma línea que Dawson y Guilhaus habían descrito un par de años antes, en agosto de 1991, Dodonov y sus colegas presentaron en la 12 ª Conferencia Internacional de Espectrometría de Masas en Amsterdam, el acoplamiento de una fuente de iones de electrospray a un TOF con un reflector de iones y "aceleración ortogonal" (125). Este trabajo se baso en una patente Rusa obtenida en 1987, pero no accesible a la literatura occidental y que no se publicó hasta 1994, (126) (Figura 13). Casi simultáneamente, en septiembre de 1991, Boyle, Whitehouse y Fenn (127) publicaron una nueva aproximación para el acoplamiento directo del electrospray y la EM de tiempo de vuelo basada en un enfoque de almacenamiento de iones. Estos autores no utilizaron un espejo de iones, sino una distribución especial de la óptica de almacenamiento de iones y circuitos pulsantes para generar paquetes discretos de iones. Hay que señalar, sin embargo, que el concepto de extracción de iones por impulsos ya lo habían demostrado en 1955 WC Wiley y JH McLaren (128) describiendo una técnica que llamaron "time lag energy focusing" para la corrección de primer orden de la de distribución inicial de los iones en el espacio.

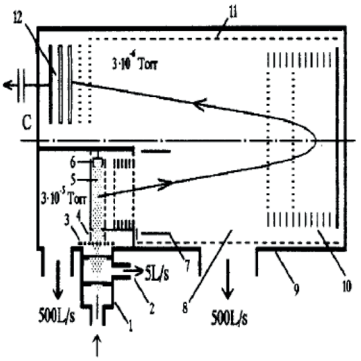


Figura 13. Diagrama esquemático de un espectrómetro ESITOF: (1) fuente de electrospray, (2) primera etapa del bombeo diferencial, (3) grM de enfoque, (4) lente de cuadrupolo, (5) modulador de acumulación de iones, (6) colector de iones, (7) placas deflectoras, (8) región de deriva de iones, (9) cámara de vacío, (10) espejo de iones, (11) malla protectora a alto potencial, y (12) detector de iones. [Reproducido con el permiso de la American Chemical Society: ver Ref. [126], Copyright (1994)].

Bernhard Spengler y cols en la Universidad de Düsseldorf y el Instituto Max-Plank de Bioquímica describieron en 1992 un nuevo método de secuenciación de péptidos por EM. El método se basó en el análisis de masas de los fragmentos iónicos de péptidos y proteínas generados por procesos de desintegración metastable uni- y bi-molecular de iones desorbidos por láser en la región libre de campo de un instrumento TOF. Aunque no se le llamó así en el artículo original, la técnica sería conocida popularmente más tarde como EM de tiempo de vuelo con desintegración-post-fuente MALDI (PSD) o MALDI-PSD y fue durante un tiempo una herramienta útil para la secuenciación de proteínas y para caracterizar o elucidar por MS/MS (129) modificaciones postraslacionales o inducidas químicamente, aunque las mediciones de la sensibilidad y precisión de masas eran relativamente bajas y el grado de fragmentación de difícil control.

33. Técnicas de la Huella de masas peptídicas.

El año 1993 podría ser conocido por "el año de la huella peptídica", ya que este fue un año histórico en relación con la rápida búsqueda de métodos de identificación "bottom-up de un gran número de proteínas por EM. En ese momento, la mayor parte de la secuenciación directa de los péptidos y las proteínas aun se llevaba a cabo por degradación automatizada de Edman en secuenciadores en fase gas. Estos instrumentos, aunque mucho más lentos en tiempo (alrededor de una hora por residuo de AA) cubrían un máximo de 25-50 residuos al día a un nivel de sensibilidad (10 pmol), aún de 100 a 1000 veces mejor que con el obtenido utilizando métodos de EM. A pesar de su alto grado de heterogeneidad, las proteínas constituyen secuencias únicas que podrían ser una herramienta eficaz para su identificación. Esta idea ya fue probada en 1989 por Henzel, Stults y Watanabe, como describen en su artículo de revisión sobre el

origen del "peptide mass fingerprinting" (130). Así, rompiendo una proteína de una manera predecible, se podrían obtener pequeños péptidos que forman una huella digital de esa proteína y, además, si dicha huella digital podía aparejarse en una base de datos con la resultante de un corte *in silico* de la proteína, esto serviría para identificarla. De hecho, estas huellas podían ser tan únicas que, como señalaron Henzel y cols. es admirable que en sus primeros experimentos con sólo dos péptidos se identificó una proteína. El trabajo base para este desarrollo había sido establecido por la publicación sobre "FAB-mapping" de H. Morris y cols (73), el electroblotting y los métodos de digestión *in situ* publicados por Aebersold y cols y por la presentación de Henzel, Stults y Watanabe en el Third Symposium of the Protein Society en 1989. Además, es notable que el concepto de mapeo de péptidos ya había sido demostrado en 1977 (131) por SDS-PAGE. La idea era buena, pero aunque el potencial de la huella de masas peptídicas había quedado demostrado, su aplicación práctica tuvo que esperar a la evolución de la instrumentación desde los métodos de FAB y PD/MS utilizados en 1989 a los recién desarrollados ESI y MALDI que permitían sensibilidades picomolares a un alto rango de masas.

En busca de "el" método de identificación rápida de nuevas proteínas, entre junio y noviembre de 1993, varios grupos publicaron simultáneamente informes sobre nuevos enfoques para desarrollar algoritmos de búsquedas en bases de datos capaces identificar unívocamente una proteína a través de las masas de los péptidos resultantes de su digestión con una enzima de rotura específica conocida. En conjunto, este enfoque se conoce como "mass mapping", "mass fingerprinting" or "ladder sequencing" de proteínas.

Todos los métodos descritos en 1993 se basaban en el aislamiento de la proteína en cuestión por electroforesis mono- o bi-dimensional, su digestión enzimática o química, y la determinación directa por EM de las masas exactas de los péptidos generados, o bien separación de los péptidos digeridos en HPLC microcapilar y determinación de fragmentos específicos de secuencia (Proteómica bottom-up). La huella peptídica resultante, el reactivo utilizado para romper la proteína y el peso molecular de la proteína intacta fueron los parámetros utilizados en el rastreo de bases de datos de proteínas, por lo que este tipo de enfoque tuvo que ir de la mano con el desarrollo de algoritmos computacionales y poder de computación adecuado junto con la disponibilidad de bases de datos de proteínas, como John Yates y cols. (132) demostraron un año más tarde. Evidentemente, las limitaciones a la identificación mediante mapeo de masas, estriban en que las muestras han de ser relativamente homogéneas para evitar ambigüedades en la asignación de la secuencia de aminoácidos, que las identificaciones de secuencia deben ser confirmadas por EM en tándem y que la secuencia de la proteína ha de estar presente en bases de datos rastreables.

Las tres primeras de esta serie de publicaciones aparecieron todas en junio de 1993. Dos de ellas eran de Darryl Pappin y cols, (133) y Peter Roepstorff y cols. (134) quienes describieron programas de búsqueda para rastrear bases de datos por el peso molecular de la proteína y por los mapas de péptidos. También en junio, WJ Henzel y cols. (135) describieron una digestión *in situ* sobre una membrana de PVDF con análisis por MALDI. A estos trabajos les siguió en Agosto la publicación de Peter James y colaboradores en Zurich (136) que después de dos separaciones por electroforesis en gel bidimensional recortaron la proteína del gel, realizaron la digestión *in situ* con una proteasa y determinaron las masas de los péptidos resultantes por ESI en un TSQ 700 (tabla 2, Cap. 7). Esto permite la búsqueda en una base de datos de los perfiles de masa generados por la fragmentación teórica de todas las secuencias conocidas de la proteína con el perfil obtenido experimentalmente. En octubre, Brian Chait y cols. publicaron su novedoso método de Secuenciación en Escalera (Ladder Sequencing) (137), que describieron como un nuevo principio en la secuenciación de proteínas, combinando múltiples pasos de química de degradación húmeda con un solo paso de lectura por MALDI/MS de la secuencia de AAs a una alta sensibilidad (2-5 fmol) y gran cantidad de muestras a velocidades de más de 100 residuos por hora. Por último, en noviembre John Yates y colaboradores describieron su método de asignación de masas basado en la generación de

digestados proteolíticos de proteínas y su análisis por HPLC microcapilar ESI-MS en un triple cuadrupolo TSQ 700 (138).

El éxito de estas técnicas "bottom-up' depende de la especificidad de la enzima proteolítica, de la exactitud de la masa y de la resolución del espectrómetro de masas para identificar cuantos mas péptidos posibles. Esto es aún más crítico en la EM top-down, cuando dicha identificación se lleva a cabo sobre la proteína intacta, como demostraron en julio de 1993 McLafferty y cols que describieron lo que puede considerarse como el primero de los instrumentos reales de top-down. Se trataba de un espectrómetro de masas de transformada de Fourier (FTMS) con fuente de ESI y equipado con un imán 6,2 Teslas y una resolución de más de $l0^6$ y mas de 1 ppm en exactitud de masa (139) para una proteína de 29 kDa. Con tal poder de resolución y las características de precisión de masa, la disociación del ion molecular de la proteína producía una serie de iones fragmento cuya masa exacta podía identificarse a partir de sus valores de relación masa-carga y el espaciamiento isotópico.

La búsqueda de exactitud de masa y poder de resolución cada vez mayores ha sido uno de los motores para los desarrollos instrumentales. La Tabla 2 del Capitulo 7 muestra la evolución comercial de estos parámetros a lo largo de los años.

34. Extracción Retardada de Iones.

Además de las técnicas de la huella peptídica, el año 1993 también fue importante para el desarrollo de técnicas para el aumento de la resolución de la masa de los iones generados por MALDI en instrumentos TOF lineales o de reflectrón. En este sentido, Robert S. Brown y sus colaboradores de la Universidad de Utah (P. Chaurand, comunicación personal) presentaron el primer informe sobre lo que hoy podríamos llamar el redescubrimiento de la extracción retardada de iones o extracción por pulsos (DE) para aumentar el poder de resolución del MALDI-TOF, según lo descrito casi 40 años antes por Wiley y McLaren (128). Este trabajo lo presentaron en la Rocky Mountain Conference on Analytical Chemistry, celebrada en agosto de 1993 en Denver, Colorado, y fue seguido por dos presentaciones del mismo grupo al año siguiente. Una de ellas tuvo lugar en marzo en la reunión 'Desorción 94' en Sunriver, OR, titulada Mass Spectrometry of Large Organic Ions by Particle and Photon Induced Desorption y la otra se presento en la 42 ASMS en Chicago. Poco después de su reactivación por Brown y sus colaboradores, esta técnica demostró su aplicabilidad práctica y sus ventajas para el trabajo con proteínas y péptidos en una serie de artículos, que aparecieron en 1995, y que describían varios enfoques instrumentales para incorporar la DE en los instrumentos TOF con una mejora en la resolución de masa de 3 a 6 veces mayor para proteínas como el citocromo c, la insulina bovina, la anhidrasa carbónica o la RNAsa B. En el intervalo de unos meses los tres principales fabricantes de MALDI-TOF MS en aquel momento (Applied Biosystems, Bruker y VG/Fisons) propusieron sistemas equipados con DE (p. Chaurand, comunicación personal).

35. SELDI.

En la 41 ASMS en San Francisco en 1993 Hutchens y Yip describieron otro enfoque novedoso para la obtención de perfiles de proteínas que publicaron en RCM (140). El enfoque se basaba en el desarrollo de dos estrategias de desorción para el análisis de macromoléculas

por EM, ambas basadas en el diseño molecular y la construcción de dos tipos de superficies como sonda de muestreo. Una de ellas, diseñada para mejorar la desorción de macromoléculas intactas (puras), se denominó surface-enhanced neat desorption (SEND). La otra diseñada para mejorar la desorción de macromoléculas específicas capturadas directamente de extractos y fluidos biológicos no fraccionados se denominó captura de afinidad asistida por superficie (surface-enhanced affinity capture, SEAC). Se esperaba que estas nuevas superficies de captura por sonda facilitaran el descubrimiento de proteínas a través del reconocimiento molecular *in situ*, y también el análisis de la estructura de macromoléculas a través de la química secuencial y/o modificación enzimática del analito adsorbido *in situ*. La técnica se convertiría más tarde en el conocido método SELDI (espectrometría de masas de ionización/ desorción por láser mejorada por superficie (Surface Enhanced Laser Desorption/ Ionization Mass Spectrometry).

36. Fragmentos de secuencia.

Después de la explosión inicial de publicaciones sobre métodos de "protein mass fingerprinting", Mann y Wilm introdujeron en 1994 el concepto de "fragmentos de secuencias (sequence tags)" como un nuevo enfoque para la identificación de péptidos fragmentados por EM (141). Esto fue un primer ejemplo de la identificación de una proteína a partir de un único espectro ESI MS/MS de iones fragmento de péptidos trípticos con doble carga, producidos en espectrómetros de masas de triple cuadrupolo. Como que el espectro MS/MS de un péptido proporciona información sobre su secuencia de AAs, este tipo de mediciones ofrece más especificidad que el mapeado peptídico. Al mismo tiempo, John Yates, y cols. presentaron un nuevo método de correlación cruzada de los datos de EM en tándem para péptidos fragmentados, con las secuencias de AAs en una base de datos de proteínas. Este método facilita la identificación automatizada de proteínas a partir de un conjunto de espectros de EM en tándem por la capacidad de predecir los iones fragmento de una secuencia de AAs y luego correlacionarlos directamente con los datos experimentales obtenidos de espectros de EM en tándem sin interpretar (132). Este manuscrito, presentó una aplicación específica para la identificación de péptidos MHC de clase II, confirmando el papel cada vez mayor de la EM en inmunología, como también demostró aquel mismo año una importante publicación de Donald Hunt y sus colaboradores sobre la identificación y secuenciación de uno de los epítopos de un péptido a partir de un total de nueve péptidos reconocidos por cinco líneas citotóxicas de las células T específicas de melanoma (142).

Todos los métodos basados en la correlación cruzada de datos de EM con las secuencias almacenadas en las bases de datos apropiadas dependían del desarrollo de algoritmos informáticos, así como del acceso a la World Wide Web. Uno de los primeros de estos algoritmos fue el Protein Prospector, inicialmente desarrollado por AL Burlingame y cols. (143) en abril de 1995 y disponible en la Web como el Protein Prospector v1.06/1996.

37. Nuevos instrumentos de Tiempo de Vuelo.

Como ya se ha citado, en el numero de Enero de 1994 de Analytical Chemistry Alexander Dodonov (126) publico el diseño de un espectrómetro de masas ESI-TOF con aceleración ortogonal y en el mismo número el grupo de Kenneth Standing en Manitoba (144) también publicó su des-

cripción del acoplamiento directo de una fuente de electrospray a un TOF en el que los iones se introducían mediante un haz continuo, perpendicularmente al eje del TOF y se extraían en ráfagas cortas paralelas al eje mediante impulsos eléctricos aplicados a los electrodos de inyección.

En cuanto a nuevos instrumentos TOF, en el año 1994 también apareció la descripción de dos interesantes técnicas para aumentar la capacidad de identificación de los iones metastables formados después de la extracción inicial en la región de la fuente. Estos iones pueden ser extremadamente útiles en la determinación de la secuencia de AAs péptidicos. Una de estas técnicas fue la del enfoque post-fuente (post-source focusing, PSF) de Kaufmann y cols. destinada a la adquisición de la información de la masa de iones metastables, que normalmente se pierde en instrumentos TOF convencionales (145), y la otra fue la introducción del reflectrón de campo curvado por Cornish y Cotter, también para el mismo fin (146).

38. Micro y nano-electrospray.

Al mismo tiempo, en la 42nd ASMS Conference, celebrada en Chicago en mayo-junio de 1994 asistimos a un hito muy importante en el camino hacia la plena identificación de péptidos y proteínas por EM. Allí, Emmett y Caprioli y por separado Wilm y Mann presentaron cada uno sus fuentes de electrospray miniaturizadas, ambas llamadas de ionización por "microelectrospray". En el primer caso, Emmett y Caprioli integraron un relleno de cromatografía líquida de C18 en la punta de la aguja del electrospray y optimizaron la fuente para acomodar caudales de nanolitros. Este trabajo se publicó en la edición de julio de la Revista de la Sociedad Americana de Espectrometría de Masas (JASMS), donde los autores demostraron sensibilidades del orden del attomol para péptidos y proteínas (147). Por otra parte, el trabajo de Wilms y Mann no se publicó en detalle hasta Enero de 1996 en Anaytical Chemistry (148). Previamente, en la edición de Septiembre del International Journal for Mass Spectrometry and Ion Processes, estos autores habían publicado un estudio de su fuente microelectrospray montada en un espectrómetro de masas de electrospray API III (149). Sus principales características diferenciales eran un orificio de pulverización de solo 1-2 micras, la velocidad de flujo muy baja de ~ 20 nL/min y la ausencia de bombas y válvulas de entrada de disolvente. Así, este dispositivo, que proporcionó un tiempo ilimitado para los estudios de EM, operaba con el flujo dictado por el propio proceso de electrospray y los autores, lo redenominaron nanoelectrospray (nanoES) que claramente lo diferenciaba de otras fuentes de microelectrospray. En ese momento, el grupo de McLafferty en la Universidad de Cornell también estaba desarrollando una fuente de nanoES o más bien "picospray" para ampliar las capacidades del ESI-FTMS, que fue descrita en octubre de 1995 en Analytical Chemistry (150).

39. Primer uso del termino Proteoma.

Otro hito que cabe destacar de 1994 fue la acuñación del término "proteoma" (el complemento proteico del genoma) por Marc Wilkins en el primer Simposio sobre "2D electrophoresis: from protein maps to genomes" ("electroforesis 2D: de los mapas de proteínas a los genomas"), que se celebró en Siena, Italia, y que se ha convertido en un evento regular cada dos años. La primera publicación en mencionar la palabra Proteoma salió en 1995 (151). Así, 1995 puede citarse como el año del comienzo de la era del Proteoma.

40. La EM se impone a las técnicas clásicas.

Al final del año, un interesante artículo de los grupos de Fred McLafferty en la Universidad de Cornell y George Kenyon en la Universidad de California en San Francisco, describió como la EM de alta resolución de transformada de Fourier con ionización por electrospray permitió el análisis de proteínas intactas, sin previa degradación enzimática o química. Esto se demostró para la creatina cinasa, una proteína 43 kDa (152) y también para la tiaminasa I (42 kDa) y la albúmina (67 kDa).

A principios de 1996, Wilm y Mann publicaron formalmente la descripción de su fuente de nanoESI y un mes después, en Febrero, Wilms y cols. describieron el uso de la nanoESI-EM en tándem para tiras de secuencia de hasta 16 AAs de proteínas aisladas por electroforesis en gel que luego se podían identificar inequívocamente a través de las bases de datos. El procedimiento resultaba de 10-100 veces más sensible que las técnicas de degradación de Edman cuyo límite práctico era de 10-50 pmol (153). Esto dejó en claro que la EM finalmente había superado a las técnicas clásicas en términos de sensibilidad. Más tarde, en Agosto de ese año McLafferty y cols. (154) demostraron el acoplamiento de la electroforesis capilar con la EM en tándem de transformada de Fourier para el análisis de la anhidrasa carbónica (CA) con un poder de resolución de 60.000 pora cantidades subatomolares de la proteína. En este trabajo los autores identificaron la CA en un solo glóbulo rojo (RBC) que tan sólo contenía 7 amoles de la proteína. El análisis de las proteínas en una sola RBC también había sido descrito simultáneamente por Smith y cols (155) quienes describieron el primer análisis de una célula individual llevado a cabo mediante la inyección de un eritrocito humano en un tubo capilar de CZE y, después de una lisis en columna, la detección de la hemoglobina mediante ESI-MS-FTICR. Este tipo de enfoques demuestran la posibilidad no sólo del análisis de células individuales, sino también de análisis subcelulares, con el uso de electrofisiología de una sola célula y técnicas de “patch camping”.

41. La llegada de las nuevas Trampas de Iones y las técnicas de bioimagen por EM.

Finnigan MAT mostró en Pittcon ‘96’ su LCQ™, una nueva trampa de iones para LC (Tabla 2 en Capitulo 7 y Figura 14). Este instrumento, que el año anterior se había presentado en la reunión de ASMS en Atlanta era un detector de sobremesa para HPLC basado en la tecnología de la trampa de iones que representó una significativa mejora respecto a la anterior versión ITD 700 para EM en tándem y que adquirió mucha popularidad. El instrumento era capaz de realizar experimentos en etapas múltiples, hasta MS^{10}. En general, las principales ventajas de este sistema de trampa de iones dedicada residen en su costo significativamente reducido y en una alta sensibilidad, del orden de 5-10 veces mejor que con los cuadrupolos triples. También en Pittcon, Bruker y Hewlett-Packard (HP) anunciaron conjuntamente la introducción del espectrómetro de masas Bruker Esquire Ion Trap, equipado con el sistema HP 1100LC, que formalmente se presentó el año siguiente en PITTCON 97 (Tabla 2, Capitulo 7). También en esta reunión, el nuevo instrumento de sobremesa Kompact MALDI IV de Shimadzu Scientific Instruments incorporo la tecnología patentada del

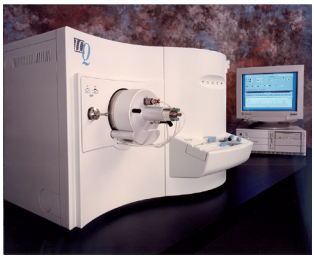


Figura 14. Trampa de Iones LCQ (Cortesía de James W. Hurst de ThermoFisher Scientific). [Reproducido con el permiso de John Wiley & Sons, Ltd.: ver Ref. [2], Copyright (2009)].

reflectrón de campo curvado (146), que permitía focalizar toda la corriente de iones para mayor sensibilidad y velocidad de operación.

42. Mejoras en la electroforesis en gel de poliacrilamida. Electroforesis diferencial.

En 1997 Unlu y cols. en Carniege Mellon University presentaron una modificación del gel de electroforesis de poliacrilamida de dos dimensiones que sólo requería un único gel para detectar diferencias entre dos muestras de proteínas (156). La técnica denominada electroforesis diferencial en gel (DIGE) evita la necesidad de comparar varios geles 2-D. Esto se logra marcando por fluorescencia dos muestras con dos tintes diferentes, desarrollando ambas muestras sobre el mismo gel, y obteniendo imágenes de fluorescencia del gel con la superposición de las imágenes de los dos tintes para revelar las diferencias.

Sin embargo, a pesar de todos los avances en la separación y detección en geles 2-D, los problemas tecnológicos de la electroforesis en gel, como la falta de suficiente resolución para el estudio de proteomas completos provocó el desarrollo de métodos de análisis alternativos para la expresión de proteínas a gran escala. La búsqueda, como ya habían demostrado Hunt y cols, se centro en el poder de la CL-MS/MS combinada (157). Esto, en combinación con la búsqueda en bases de datos, condujo a lo que entonces se denominaba la “proteómica por perdigonada” “shotgun proteomics” a pesar de que estos métodos aún no fueron suficientes para trazar el perfil cuantitativo de proteínas.

43. Hibridos de cuadrupolo y TOF.

También en 1997 y junto con la introducción comercial por Bruker en Pittcon de la trampa de iones como alternativa a la EM de triple cuadrupolo (Esquire-LC, véase el Tabla 2, Capitulo 7), Micromass presentó un espectrómetro híbrido de nueva geometría, un cuadrupolo TOF (Q-TOF) con aceleración ortogonal, previamente desarrollado por Howard Morris y cols. (158). El objetivo alcanzado por los autores fue la secuencia inequívoca de proteínas desconocidas, como los antígenos peptídicos MHC a niveles de femtomol y subfemtomol. Al mismo tiempo, en un esfuerzo de colaboración entre el grupo de proteínas y péptidos del EMBL, la Universidad de Manitoba y PE Sciex también se construyó un sistema ortogonal Q-TOF mediante la combinación de parte de un Sciex API 300 (desde la fuente de iones, e incluyendo, Q2) con el espectrómetro de masas ESI-TOF descrito previamente (159). Este trabajo condujo a principios de 1998 al inicio del proyecto Q-Star en MDS Sciex cuyo objetivo era capturar una cuota de mercado mediante el desarrollo de nuevos espectrómetros de masas híbridos que combinasen la tecnología de tiempo de vuelo con los diseños convencionales de cuadrupolo. El nuevo sistema híbrido QSTAR CL-MS/MS, con un analizador de transmisión cuadrupolar, un quadrupolo como cámara de colisión y un analizador de tiempo de vuelo (Tabla 2, Capitulo 7), fue presentado por Applied Biosystems/MDS Sciex en Pittcon 99 en Orlando, Florida.

44. Imagenes por MALDI.

En diciembre de 1997 Richard Caprioli, entonces, en la Universidad de Texas Medical School en Houston, informó por primera vez sobre el uso del MALDI para generar imágenes de iones de secciones de tejido biológico, lo cual proporcionaba la capacidad de crear un mapa de moléculas específicas sobre las coordenadas bidimensionales de la muestra original (160). Aunque los trabajos anteriores de obtención de imágenes por EM se debían al grupo de Bernhard Spengler quien utilizo un equipo LAMMA comercial para análisis elemental y a los grupos de George Morrison y otros que utilizaron SIMS, así como a los trabajos de David Hércules y cols. sobre imágenes de cromatógramas en capa fina mediante MALDI MS (RG Cooks y P.Chaurand, comunicación personal), fue el trabajo de Houston lo que inició la era de la bioimagen a través de la EM de MALDI y todos los acontecimientos venideros en este nuevo campo, especialmente los que por primera vez realizo el grupo Caprioli en aplicaciones clínicas. En esta primera publicación, los autores utilizaron un espectrómetro de masas PerSeptive Voyager Elite de tiempo de vuelo, equipado con sistema de extracción retardada (DE) y un láser de nitrógeno (337 nm), para determinar los pesos moleculares de las moléculas en las capas superficiales del tejido después de haber aplicado una fina capa de la matriz sobre la superficie para el análisis directo por MALDI MS. Esto se demostró con el mapeado de moléculas en zonas de menos de 1 mm^2. Por ejemplo, la insulina en un trozo de tejido pancreático de ratas, péptidos hormonales en una sección de hipófisis de rata, y una pequeña proteína unida a la superficie de células endoteliales humanas.

45. Nueva tecnología y mecanismos de ionización.

El mecanismo de la disociación por captura de electrones (ECD) se conocía desde principios de 1980. Sin embargo, no fue hasta 1998 cuando Zubarev, Kelleher y McLafferty en la Universidad de Cornell lo aplicaron a proteínas (161). Este renacimiento del ECD se debió al hallazgo fortuito de que una o más de las cargas de los cationes de proteínas con cargas múltiples generados por ESi se podían neutralizar con electrones de baja energía, lo que provocaba la rotura especifica del enlace amida a productos *c*, *z* en contraste con los *b*, *y*, formados por CAD u otros mecanismos de fotodisociación y métodos LID de fragmentación explícitamente lenta de electrones pares. Debido a la escisión nonergodica (picosegundo) del ECD a menudo se conservan las modificaciones co- y post-traslacionales lo que permite el análisis específico del sitio de fragmentación (site-specific). En consecuencia, la técnica ha evolucionado rápidamente a un método alternativo para la secuenciación de proteínas y péptidos, pero su uso quedó limitado al FTICR-MS hasta su extensión a la Disociación por Transferencia Electrónica (ETD) de Hunt y cols. en 2004 (162). Las características del ECD son especialmente útiles para fragmentar proteínas intactas en el EM ("proteómica Top-down"). De hecho, Fred McLaffertty recientemente en una presentacion en una reunion de la ASMS afirmó que la espectrometría de masas ha sido siempre Top-down desde 1950!

Siguiendo la tendencia encaminada a encontrar métodos alternativos a la electroforesis en gel y espectrometría de masas, en 1998 en nuestro laboratorio en el CSIC en Barcelona aplicamos con éxito un método de micro LC-ESI-MS/MS para el análisis directo de mezclas complejas de péptidos. Se basaba en el trabajo que habíamos iniciado en 1996 sobre el desarrollo de las columnas de nano-LC y una interfaz de micro-electrospray (163) para el análisis de péptidos de endotelina endógena. Esta fue una de las primeras aplicaciones cuantitativas prácticas de nano-LC y micro-ESI al estudio de muestras biológicas. Para hacer frente a las bajas concentraciones de los péptidos, el equipo de nanoLC-ESI MS/MS se acopló a un sistema de

preconcentración en línea basado en un material de intercambio aniónico fuerte. Así, después de una etapa de preconcentración en línea, los péptidos de interés se separaron en la columna de nano LC y se cuantificaron por electrospray tándem EM mediante un péptido sintético análogo como estándar interno (Figuras 15, 16). Más tarde, en 1999 (164) describimos una evaluación detallada de este sistema, así como un resumen y comparación de diferentes tipos de columnas (convencional, diámetro estrecho y capilar) y los parámetros de la miniaturización en la práctica de la CL-EM.

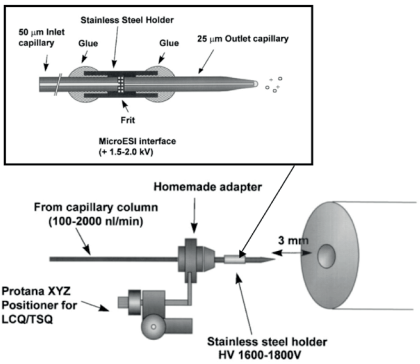


Figura 15. Configuración micro-ESI. (Reproducido de la referencia 165 con permiso). [Reproducido con el permiso de John Wiley & Sons, Ltd.: ver Ref. [165], Copyright (1999)].

Aparte de la introducción del QStar, mencionada anteriormente, 1999 fue un año interesante desde el punto de vista del método y el desarrollo de instrumentos, tanto para la proteómica como para EM. Fue entonces cuando Alexander Makarov obtuvo una patente de los EE.UU. para un nuevo tipo de analizador de masas que se conocería unos años más tarde como el Orbitrap. Este nuevo concepto de instrumento, que fue lanzado comercialmente en 2005 (Tabla 2, Capitulo 7), fue el resultado de su liderazgo en un equipo de desarrollo cuando él empezó a trabajar para HD Technologies en 1998 antes de que la compañía fuese adquirida por Thermo Fisher Scientific. Además, en ese momento comenzaron a aparecer en el mercado las plataformas de soluciones integradas para proteómica. En esta linea, Micromass y PE Biosystems presentaron en PITTCON '99 el ProteomeWorks™ y la Proteomics Solution 1™, respectivamente. Estos sistemas trataban de integrar la preparación de muestras y la espectrometría de masas.

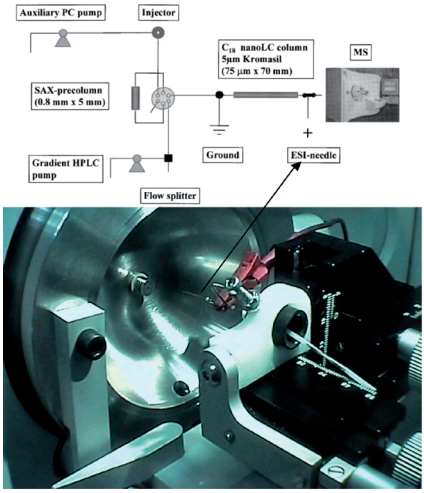


Figura 16. Sistema de microconcentración para un equipo de nanoLC acoplado a una fuente de micro-ESI y detalle grafico de la interfaz. [Reproducido con el permiso de John Wiley & Sons, Ltd.: ver Ref. [165], Copyright (1999)].

Aquel año, en mayo, Nature presento la publicación de Gary Siuzdak del Scripps Research Institute quien, en colaboración con Jillian Buriak de la Universidad de Purdue, describió una estrategia libre matriz para EM biomolecular basada en la ionización por desorción con láser pulsado desde una superficie de silicio poroso denominada ionización por desorción sobre silicio (DIOS) (165). Sin embargo, aunque el método ofrece buena sensibilidad y alta tolerancia a los contaminantes, está libre de interferencias de la matriz y también puede aplicarse a los análisis de moléculas pequeñas, todavía no ha alcanzado gran popularidad en términos de un número importante de aplicaciones en la química de péptidos y proteínas.

Con el cambio de siglo era evidente que cuando se separaban lisados celulares sobre placas de 2-D, sólo podían observarse pequeños subgrupos del proteoma celular y que la incidencia de la co-migración de proteínas era más común de lo esperado, lo que limitaba los enfoques cuantitativos (166) y por lo tanto era natural que los laboratorios especializados se vieran obligados a explorar las ventajas de combinar los bien conocidos conceptos de la dilución isotópica con el fraccionamiento de mezclas de péptidos complejos.

46. Marcado con isótopos estables y de fragmentos de afinidad codificados por isótopos.

En 1999 ya se había logrado la secuenciación automatizada de los genomas completos de diversos organismos unicelulares. Sin embargo, el estudio de los resultados de la expresión génica, y su traducción en proteínas exigía tener en cuenta las modificaciones covalentes co- o post-traslacionales en la secuencia de AAs y esto no resultaba un proceso de fácil cuantificación. En consecuencia, la aparición por primera vez en junio y después, en octubre en ese año, de las ponencias de los grupos de Brian Chait (167) en la Universidad Rockefeller, Ruedi Aebersold (168) en la Universidad de Washington y Bertrand Séraphin (169) en el European Molecular Biology Laboratory sobre el desarrollo de métodos de alta afinidad para el marcado de proteínas diana tuvo un alcance de grandes consecuencias prácticas. De estos tres, el grupo de Brian Chait parece haber sido el primero en describir el uso de isótopos estables para el marcado metabólico de las células en proteómica cuantitativa. Según lo descrito por estos autores, un grupo de células (por ejemplo, el tipo salvaje) se cultiva sobre un medio con abundancia natural de los isótopos de nitrógeno, mientras que un segundo grupo (por ejemplo, un mutante) se cultiva sobre el mismo medio pero enriquecido en ^{15}N. Después de un período de crecimiento, los grupos de células se combinan y las proteínas de interés se extraían y separaban parcialmente (por ejemplo, por PAGE). Las manchas de interés en los geles se sometían a un proceso de digestión por una proteasa. Los fragmentos peptídicos resultantes se extraían del gel y se analizaban por EM.

La situación hasta ese momento en términos del análisis cuantitativo de proteínas se había centrado en la combinación de las técnicas de trabajo intensivo por separación de proteínas en 2D-PAGE con la identificación de secuencia de las proteínas seleccionadas por tándem MS / MS. Los resultados en general no fueron cuantitativos y por lo general se ignoraban las proteínas de poca abundancia. Por lo tanto, la introducción del marcado con isótopos estables por Chait y cols. y de fragmentos de afinidad codificados por isótopos (ICAT) por Aebersold y cols. para la cuantificación de la expresión diferencial de proteínas sin usar geles 2-D proporcionó

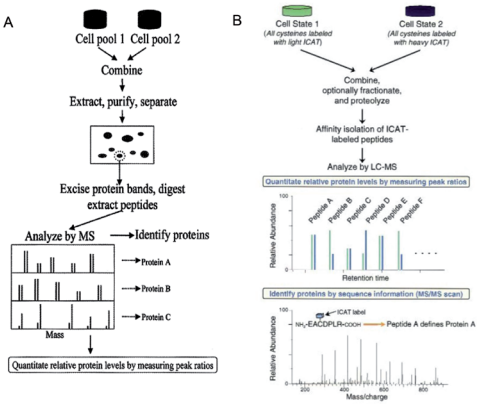


Figura 17 A. Método para la cuantificación de la expresión diferencial de proteínas. En el ejemplo, se asume que las proteínas que permanecen inalteradas en los dos pools celulares están presentes en iguales cantidades. (En la práctica, las proteínas que no se modifican en los dos pools celulares pueden no estar presentes en cantidades equivalentes, y la relación entre los picos de péptidos será un valor constante no igual a la unidad.) [Reproducido de Ref. [168] Copyright 1999 National Academy of Sciences, USA].

Figura 17 B. Estrategia ICAT para cuantificar la expresión diferencial de proteínas.Dos mezclas de proteínas representando dos estados celulares se han tratado con los reactivos ICAT isotópicamente ligeros y pesados, respectivamente; un reactivo ICAT se halla covalentemente ligado a cada residuo cisteinilo en cada proteína. Las proteínas procedentes del estado celular 1 se ilustran en color verde y las proteínas del estado 2 en color azul. Las mezclas proteicas se combinan y someten a proteólisis para obtener los péptidos marcados con el reactivo ICAT que se aíslan utilizando el marcador de biotina. Estos péptidos se separan por HPLC microcapillar. Un par de péptidos marcados con ICAT son químicamente idénticos y se visualizan fácilmente porque esencialmente co eluyen y presentan una diferencia de masa de 8 daltons medida en un espectrómetro de masas de barrido (cuatro unidades de m/z para un ion doblemente cargado). Las relaciones entre las cantidades originales de las proteínas procedentes de los dos estados celulares se mantienen estrictamente en los fragmentos peptídicos. La cuantificación relativa se determina por el cociente entre los pares de péptidos. Cada barrido alternativo se dedica a la fragmentación y luego al registro de la información de secuencia sobre cada péptido eluido (espectro de masas en tándem). La proteína se identifica por búsqueda computacional de la información registrada de secuencia frente a grandes bases de datos de proteínas. [Reproducido con el permiso de Macmillan Publishers Ltd: ver Ref. [169], Copyright (1999)].

la capacidad de medir con precisión pequeños cambios en los niveles relativos de la proteína. Esta estrategia, que en ambos casos se aplicó en un espectrómetro de masas LCQ de trampa de iones, se resume en la Figura 17. Del mismo modo, el enfoque adoptado por el Grupo de Séraphin, basado en el uso de fragmentos de purificación por afinidad en tándem (TAP), representó un procedimiento genérico para purificar proteínas que se expresan a su nivel natural. La estrategia de TAP, combinada con la EM en un espectrómetro de masas API III equipado con una fuente nanoelectrospray, permitió la identificación de las proteínas que interactúan con una proteína objetivo determinada.

47. Contribución de la EM a la secuenciación del Genoma Humano.

En el año 2000 los trabajos realizados al amparo del Proyecto Genoma Humano produjeron la presentación por Celera Genomics de un primer borrador de trabajo del genoma que, eventualmente, se completó en 2003. El envío del primer instrumento MALDI TOF/TOF de PE Biosystems a Celera Genomics tuvo lugar en 2000. El prototipo de este instrumento se había presentado en la 47 Conferencia ASMS, celebrada en Dallas, Texas, en 1999 como un desarrollo conjunto del grupo de Al Burligame de la Universidad de California en San Francisco y de Marvin Vestal en PE, Framingham (170). Este instrumento se destacó luego como la tecnología tándem TOF de siguiente generación y un año más tarde se comercializó como el 4700 MALDI TOF/TOF Analyzer. Simultáneamente a este desarrollo, el grupo de Kenneth Standing de la Universidad de Manitoba en Canadá, bajo el patrocinio de MDS Sciex, describió un MALDI QqTOF para mediciones de MS/MS de alta precisión que también había sido descrito en la reunión de la ASMS el año antes en Dallas y que había estado en fase de desarrollo desde 1997 (171). Los autores utilizaron este instrumento para el análisis de la citrato sintasa (Mr 47.885 Da) en un digerido de *E. coli* demostrando también con la albúmina de suero bovino que era capaz de analizar proteínas de hasta unos 80 kDa.

Otra novedad en el año 2000 fue la descripción del MALDI a presión atmosférica (PA) por varios grupos. En octubre de 1999 Laiko y Burlingame habían obtenido una patente de los EE.UU. sobre el sistema que habían presentado también en la reunión de Dallas ese mismo año y que describieron al año siguiente en la edición de febrero de Analytical Chemistry (172). En este sistema, la transferencia de los iones de la región de ionización a presión atmosférica a la de alto vacío fue asistida neumáticamente (PA). El PA-AP MALDI se demostró capaz de proporcionar iones diagnóstico de los péptidos y hasta iones moleculares protonados de proteínas pequeñas como la insulina y los autores señalaron que podría intercambiarse fácilmente con la ionización por electrospray en un instrumento oaTOF (PE BioSystems Mariner oaTOF). Sin embargo, la producción de una corriente de iones a PA resulto alrededor de 1 orden de magnitud inferior a la de una típica fuente de ESI, debido a la menor eficiencia en la transferencia de iones desde la fuente externa a la cámara de vacío del espectrómetro de masas. De alguna manera esto podía compensarse por el consumo casi total de la sustancia analizada. El equipo demostró sensibilidad femtomolar para mezclas de péptidos. Poco después de esta publicación, en la edición de noviembre de Analytical Chemistry, Robert Cotter y cols., incluyendo a Víctor Laiko que en el ínterin se había trasladado de la UCSF a la Universidad John Hopkins, describió un sistema similar para la EM de trampa iónica (AP MALDI / trampa de iones) que demostró la posibilidad de análisis estructurales y experimentos MS^3 (173). Véase también la revisión de Moyer y Cotter (174).

Una imagen claramente emergente en el cambio de siglo en el nuevo campo de la Proteómica era la necesidad de alternativas, complementarias y estrategias de análisis más globales relativas a la electroforesis en gel 2-D. Estas alternativas debían ser capaces de proporcionar más

potencia de separación y enriquecimiento de las proteínas de baja abundancia para superar las limitaciones que planteaban la amplia gama de niveles de proteínas y la gran diversidad de expresión de las mismas. En este sentido y como se indicó anteriormente, la combinación de técnicas de CL-EM/MS ya había sido seleccionada de forma natural por varios grupos como la mejor opción fundamental para este propósito. Esto fue muy bien discutido por Steven Gygi en un artículo de tutoría para el Journal of Mass Spectrometry (175).

A principios de 2002 Applied Biosystems/MDS Sciex introdujo comercialmente en Pittcon 2002 el QTrap 2000, un triple quadrupolo al que había incorporado un analizador de trampa lineal de iones como MS2 (ver capitulo 7) de modo que este segundo quadrupolo podía funcionar como filtro m/z cuadrupolar de transmisión o como trampa de iones.

Finalmente este mismo año, y después de que Wolfgang Paul compartiese en 1989 el Premio Nóbel de Física por su trabajo en el desarrollo de la trampa de iones, el Premio Nóbel se concedió en 2002 a dos investigadores que en la década de los años 80 habían desarrollado los dos métodos de ionización que mas han aportado para hacer posibles los espectaculares avances de la espectrometría de masas biológica. El premio Nóbel de Química 2002 se concedió *"por el desarrollo de métodos para la identificación y análisis estructural de macromoléculas biológicas "* compartiendo John B. Fenn y Koichi Tanaka la mitad del premio *"por su desarrollo de métodos de ionización por Desorción suave para el análisis de macromoléculas biológicas por espectrometría de masas "* y la otra mitad a Kurt Wüthrich *"por su desarrollo de la espectroscopia de resonancia nuclear magnética para la determinación de la estructura tridimensional de macromoléculas biológicas in disolución"*. Sin embargo en su mitad compartida entre J. Fenn de la Virginia Commonwealth University, Richmond, USA y K. Tanaka de Shimadzu Corp. en Kyoto, Japón el premio no estuvo exento de polémica como recogió la revista The Scientist en un artículo del 11 de diciembre de 2002 el día después de la ceremonia de entrega de los premios en Estocolmo (http://cmbi.bjmu.edu.cn/news/0212/55.htm). El articulo dejaba patente el malestar de muchos científicos por la "omisión de los dos mayores contribuidores en el campo de la espectrometría de masas en química de proteínas". En opinión de estos investigadores críticos el premio debería haber recaído en lugar de en Tanaka en Franz Hillenkamp del Institute for Medical Physics and Biophysics en la Universidad de Munster y en su colaborador Michael Karas del Institute for Physical and Theoretical Chemistry en la Universidad de Frankfurt.

También en 2002 Shimadzu-Biotech introdujo un nuevo híbrido basado en la combinación de una trampa de iones y un TOF en el que se reunían las ventajas de la trampa de iones para MS^n y la resolución y exactitud de masas del TOF. Este sistema MALDI-Trampa de Iones Cuadrupolar-analizador de masas de tiempo de vuelo se comercializo con el nombre de MALDI-AXIMA-QIT.

En la reunión anual de la ASMS en Montreal en Mayo de 2003 Applied Biosystems/MDS Sciex presentó el QTrap 4000 como modificación del API 4000. En esta configuración la trampa lineal actuaba como MS3 después de la célula de colisión LINAC de tecnología patentada para la aceleración de iones en un quadrupolo de radiofrecuencia.

En Junio de 2004 John E.P. Syka de Thermo Electrón en San José, California y el grupo de Donald Hunt en la Universidad de Virginia, Charlottesville describieron en PNAS un método para el análisis de secuencia de péptidos y proteínas por espectrometría de masas de ETD (Electrón Transfer Dissociation) que simplifica la secuenciación de péptidos mediante una combinación de interacciones ion/ion y MS/MS (176). Los autores demostraron en este trabajo que los iones de antraceno podían transferir un electrón a los péptidos múltiplemente protonados en un instrumento del tipo QLT (trampa cuadrupolar lineal), adecuadamente modificado para este tipo de interacciones químicas entre iones e inducir la fragmentación del péptido de forma análoga a la observada por ECD (Electrón Capture Dissociation). Como ya se ha indicado anteriormente

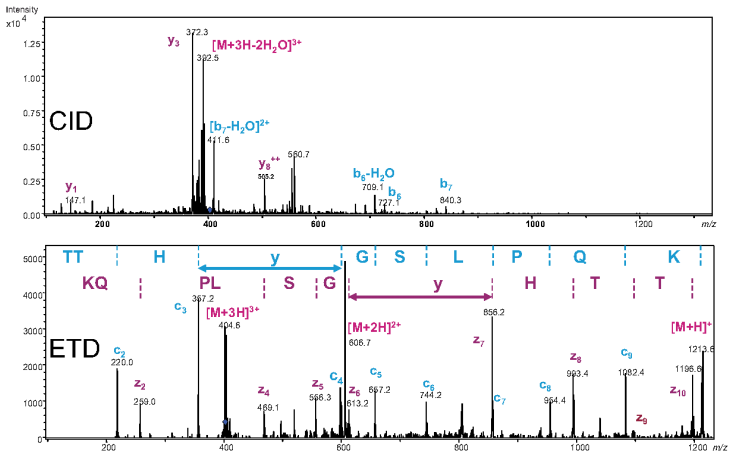


Figura 18. Comparación de los espectros obtenidos por CID y ETD del fosfopéptido TTHyGSLPQK (*m/z* = 404.7) usando HPLC-Chip nanospray. El espectro de CID presenta muy pocos iones fragmento, mientras que el de ETD proporciona una cobertura de secuencia casi completa con la indicación de la posición de la fosfotirosina. Reproducido de Figura 2 en Application Note Agilent Technologies, Inc, 2006. Characterization of Protein Phosphorylation Using HPLC-Chip Electron Transfer Dissociation Ion Trap Mass Spectrometry by N. Tang y P. Goodley

FW McLafferty introdujo en 1998 la técnica de ECD (161) la cual proporcionaba espectros de MS/MS complementarios a los obtenidos por CID en cuanto al tipo de fragmentación de la cadena peptídica. Los procesos de disociación por transferencia de electrones (ETD) constituyen una útil alternativa a la inducción de fragmentación peptídica por CID ya que al ser el proceso de colisión ion/molécula mucho más suave producen menos fragmentación de las cadenas laterales de AAs y en especial de aquellos que están modificados. Por ello el grupo fosfato permanece inalterado sobre el aminoácido portador lo que facilita su localización especifica (ver figura 18). Además facilita la fragmentación e identificación de iones precursores de mayor tamaño y carga así como de proteínas intactas. Por todo ello, el método de ETD resulta particularmente apropiado para la caracterización de péptidos con modificaciones postraslacionales. En la literatura se pueden encontrar trabajos de revisión sobre la utilidad de este método (177, 178).

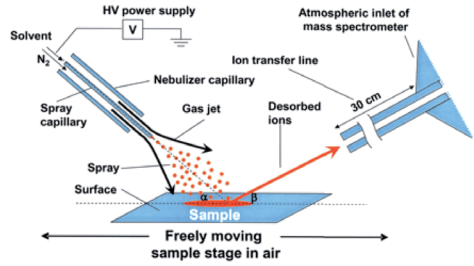


Figura 19. Representación esquemática de un experimento DESI. La muestra se deposita y se seca sobre una superficie de PTFE sobre la cual se pulveriza, mediante un gas nebulizador a velocidad de 350 m/s, una disolución de metanol-agua con un 1% de acido acético a un flujo de 3-15 µL/s. El voltaje aplicado es de 4 kV. [Reproducido con el permiso de AAAS.: ver Ref. [180], Copyright (2004)].

También en 2004, el 15 de Octubre apareció en Science (179) la descripción de un nuevo método para el análisis directo de muestras en condiciones ambientales denominado DESI (Desorption Electrospray Ionization), Figura 19. En el método DESI se dirige el chorro electropulverizado de gotas cargadas e iones de disolvente sobre la superficie de la muestra a analizar. El impacto de las gotas partículas produce iones gaseosos del material de la superficie. Por primera vez el procedimiento de muestreo es extremadamente simple y versátil a la vez. En esta publicación Graham Cooks y colaboradores describen varias interesantes aplicaciones demostrando el potencial del DESI para el análisis espacial directo de superficies de plantas o tejidos animales. Por ejemplo una de las figuras ilustra el espectro del antihistaminico Loratadine registrado mediante la dirección del jet de electrospray sobre el dedo de una persona 50 minutos después de haber ingerido el fármaco. En este experimento incluso pueden registrarse picos procedentes del detergente

utilizado por esta persona para lavarse las manos. En resumen la característica estrella de este método es que no hay que aplicar ningún pretratamiento a la muestra la cual se rocía con un jet acuoso de partículas con carga eléctrica y los iones liberados se transportan al espectrómetro de masas.

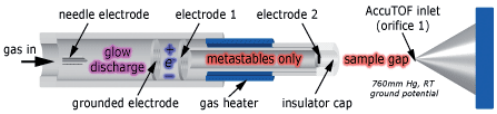


Figura 20. La fuente DART consiste en un tubo divido en varias cámaras por el que fluye nitrógeno o helio. El gas se introduce en una cámara de descarga bajo un potencial eléctrico de varios kilovoltios. La descarga eléctrica que se produce genera un plasma de gas ionizado, electrones, y especies excitadas (metastables). El electrodo 1 filtra los iones y electrones dejando paso solo a las especies excitadas que fluyen al interior de una tercera cámara que puede calentarse hasta 250 °C. Las especies metas tablees pueden interaccionar directamente con la muestra en estado sólido, liquido o gas ionizando los analitos de interés. Copyright de JEOL USA, Inc.

Al año siguiente apareció otro método que puede enmarcarse en la nueva generación de métodos de ionización en condiciones ambientales (180). Se trata del DART (Direct Analisis in Real Time) que fue introducido por Jeol USA en la Conferencia de Pittcon 2005 en una versión comercial denominada AccuTOF DART. Esta fuente de iones fue desarrollada en la sede americana de la compañía Jeol por Robert Cody y James Laramée, probándose el primer prototipo en 2003 en un espectrómetro de masas de tiempo de vuelo AccuTOF de Jeol. En la figura 20 se ilustra un esquema de la fuente DART. En esta configuración se introduce un gas seco como el helio a través de un electrodo produciéndose una descarga eléctrica que genera partículas cargadas y especies excitadas metastables. En el electrodo 1 de la figura se filtran las partículas cargadas dejando pasar solo las metastables. Este haz se calienta y finalmente pasa a través del electrodo 2 para evitar la recombinación de iones negativos y positivos. El proceso de ionización de la muestra ocurre en espacio indicado como “sample gap” que se encuentra a potencial neutro, temperatura ambiente y presión atmosférica (condiciones ambientales). Los iones formados se dirigen hacia el espectrómetro de masas por efecto del flujo de gas y del vacío del equipo. Este flujo se puede dirigir directamente al orificio de entrada del espectrómetro de masas o hacerlo incidir sobre la muestra y reflectarlo hacia la entrada. El Dart permite el análisis directo de cualquier tipo de molécula sobre una superficie así como en un gas, un líquido o un sólido sin ninguna preparación o manipulación de la muestra. Sin embargo, a pesar de la gran promesa de estas técnicas ambientales y de su reciente utilización como técnica de diagnostico del cáncer de ovario a través del análisis directo de una muestra de una sola gota de suero humano (181) todavía esta por desarrollar su potencial en el campo de la Proteómica. Así, mientras que el método DESI es capaz de formar iones multicargados a partir de péptidos y proteínas el DART no ha demostrado ser capaz de generar ni siquiera iones peptídicos monocargados.

No es este el caso sin embargo con la aparición en el mercado del Orbitrap, que Thermo Electron (San José, CA) presentó en la 53 reunión de la ASMS en Junio de 2005 en San Antonio, Tejas (ver capitulo 7). Desde la comercialización de la trampa de iones, también por Thermo Electron pronto hará 30 años este ha sido la primera novedad tecnológica absoluta en analizadores de masas y desde la publicación por el grupo de G. Cooks sobre este equipo (182) en 2005 se pueden encontrar en la bibliografía 170 trabajos publicados que describen aplicaciones del Orbitrap en el análisis de proteínas y en Proteómica. El Orbitrap utiliza métodos de transformada de Fourier análogos a los utilizados en los equipos de FT-ICR pero con la ventaja de que no requiere un imán superconductor lo que simplifica mucho su accesibilidad y mantenimiento.

Finalmente, otro instrumento de nueva tecnología todavía de más reciente introducción es el Synapt de Waters introducido comercialmente como Synapt HDMS en 2006 y premiado en la Conferencia de Pittcon 2007 como “mejor producto nuevo”. Sin embargo, como se describe en el capitulo 7 se trata de un equipo que combina la espectrometría de movilidad iónica con la espectrometría de masas y sus aplicaciones en Proteómica todavía son limitadas. Para mayor información puede consultarse un reciente artículo de revisión (183).

REFERENCIAS.

1. Gelpí E. From large analogical instruments to small digital black boxes: 40 years of progress in mass spectrometry and its role in proteomics. Part I 1965–1984. J Mass Spectrom. 2008; 43: 419–435.

2. Gelpí E. From large analogical instruments to small digital black boxes: 40 years of progress in mass spectrometry and its role in proteomics. Part II 1985–2000. J Mass Spectrom. 2009; 44: 1137–1161.

3. 16. Biemann K, The coming of age of mass spectrometry in peptide and protein chemistry. Protein Science 1995; 4: 19920-1927.

4. Abian J. The coupling of gas and liquid chromatography with mass spectrometry. Journal of Mass Spectrometry 1999; 34: 157.

5. Barber M, Jolles P, Vilkas E, Lederer E. Determination of amino acid sequences in oligopeptides by mass spectrometry. I. The structure of fortuitine, an acylnonapeptide methyl ester. Biochemical and Biophysical Research Communications 1965; 18: 469.

6. Barber M, Wolstenholme WA, Guinard M, Michel G, Das BC, Lederer E. Determination of amino acid sequences in oligopeptides by mass spectrometry. II. The structure of peptidolipin NA. Tetrahedron Letters 1965; 18: 1331.

7. Bailey K, Sanger F. The chemistry of amino acids and proteins. Annual Review of Biochemistry 1951; 20: 103.

8. Edman P. Method for the determination of amino acid sequence in peptides. Acta Chemica Scandinavica 1950; 4: 283.

9 Biemann K, Gapp F, Seibl J. Application of mass spectrometry to structure problems. I. Amino acid sequence in peptides. Journal of the American Chemical Society 1959; 81: 2274.

10. Biemann K, Vetter W. Separation of peptide derivatives by gas chromatography combined with the mass spectrometric determination of the amino acid sequence. Biochemical and Biophysical Research Communications 1960; 3: 578.

11. Bricas E, van Heijenoort J, Barber M, Wolstenholme WA, Das BC, Lederer E. Determination of amino acid sequences in oligopeptides by mass spectrometry. IV. Synthetic N-Acyl Oligopeptide methyl esters. Biochemistry 1965; 4: 2254.

12. Thomas DW, Das BC, Gero SD, Lederer E. Advantages and limitations of the mass spectrometric sequence determination of permethylated oligopeptide derivatives. Biochemical and Biophysical Research Communications 1968; 32: 199.

13. Biemann K, Cone C,Webster BR, Arsenault GP. Determination of the amino acid sequence in oligopepties by computer interpretation of their high-resolution mass spectra. Journal of the American Chemical Society 1966; 88: 5598.

14. Barber M, Powers P, Wallington MJ, Wolstenholme PA. Computer interpretation of high-resolution mass spectra. Nature 1966; 212: 784.

15. Senn M, McLafferty FW. Automatic amino acid sequence determination in peptides. Biochemical and Biophysical Research Communications 1966; 23: 381.

16. VanLear GE, McLafferty FW. Biochemical aspects of high resolution mass spectrometry. Annual Review of Biochemistry 1969; 39: 289.

17. Munson MSB, Field FH. Chemical ionization mass spectrometry. I. General introduction. Journal of the American Chemical Society 1966; 88: 2621.

18. Beckey HD. Massenspektrometrische untersuchung von chemisorptions- und assoziationsreaktionen mit hilfe einer neuen feldemissionsionen quelle. Naturwissenschaften 1958; 45: 259.

19. Edman P, Begg G. A protein sequenator. European Journal of Biochemistry 1967; 1: 80.

20. Zimmermann CL, Pisano JJ, Appella E. Analysis of amino acid phenylthiohadantoins by high speed liquid chromatography. Biochemical and Biophysical Research Communications 1973; 272: 1220.

21. Dole M, Mach LL, Hines RL, Mobley RC, Ferguson LP, Alice MB. Molecular beams of macroions. Journal of Chemical Physics 1968; 49: 2240.

22. Fenn JB. Electrospray ionization mass spectrometry: how it all began. Journal of Biomolecular Techniques 2002; 13: 101.

23. Jennings KR. Collision-induced decomposition of aromatic molecular ions. International Journal of Mass Spectrometry and Ion Physics 1968; 1: 227.

24. Haddon WF, McLafferty FW. Metastable ion characteristics. VII. Collision-induced metastables. Journal of the American Chemical Society 1968; 90: 4745.

25. Cooks GR. Collision-induced dissociation: readings and commentary. Journal of Mass Spectrometry 1995; 30: 1215.

26. Beckey HD. Field desorption mass spectrometry: a technique for the study of thermally unstable substances of low volatility. Journal of Mass Spectrometry and Ion Physics 1969; 2: 500.

27. Winkler HU, Beckey HD. Field desorption mass spectrometry of peptides. Biochemical and Biophysical Research Communications 1970; 46: 391.

28. Lucas F, Barber M, Wolstenholme WA, Geddes AJ, Graham GN, Morris HR. Mass-spectrometric determination of the amino acid sequences in peptides isolated from the protein silk fibroin of Bombyx mori. The Biochemical Journal 1969; 114: 695.

29. Burgus R, Dunn T, Desiderio DM, Ward DN, Vale W, Guillemin R. Characterization of ovine hypothalamic hypophysiotropic TSH-releasing factor. Nature 1970; 226: 321.

30. Desiderio DM, Burgus R, Dunn T, Vale W, Guillemin R, Ward DN. The elucidation of the primary structure of the hypothalamic thyroid stimulating hormone releasing factor of

ovine origin by means of mass spectrometry. Organic Mass Spectrometry 1971; 5: 221.

31. McLaferty FW, Venkataraghavan R, Irving P. Determination of amino acid sequences in peptide mixtures by mass spectrometry. Biochemical and Biophysical Research Communications 1970; 39: 274.

32. Laursen RA. Solid-phase Edman degradation. An automatic peptide sequencer. European Journal of Biochemistry 1971; 20: 89.

33. Dayhoff MO, Eck R, Chang MA, Sochard MR. Atlas of Protein Sequence and Structure. National Biomedical Research Foundation: Silver Spring, 1965.

34. Förster HJ, Kelley JA, Nau H, Biemann K. In Chemistry and Biology of Peptides, Meienhofer J (ed). Ann Arbor Science Publishers: Ann Arbor, Michigan, 1972; 679.

35. Nau H, Kelley JA, Biemann K. Determination of the amino acid sequence of the C-terminal cyanogen bromide fragment of actin by computer-assisted gas chromatography-mass spectrometry. Journal of the American Chemical Society 1973; 95: 7162.

36. Khorana HG, Gerber GE, Herlihy WC, Gray CP, Anderegg RJ, Nihei K, Biemann K. Amino acid sequence of bacteriorhodopsin. Proceedings of the National Academy of Sciences of the United States of America 1979; 76: 5046.

37. Caprioli RM, Seifert WE Jr, Sutherland DE. Polypeptidesequencing: use of dipeptidylaminopeptidase I and gas chromatography/mass spectrometry. Biochemical and Biophysical Research Communications 1973; 55: 67.

38. Biemann K. In Biochemical Applications of Mass Spectrometry, First Supplementary Volume, Waller R, Dermer C (eds). Wiley: New York, 1980; 469.

39. Beynon JH, Cooks RG, Amy JW, Baitinger WE, Ridley TY. Design and performance of a Mass Analyzed Ion Kinetic Energy (MIKE) spectrometer. Analytical Chemistry 1973; 45: 1023A.

40. Kruger TL, Litton JF, Kondrat RW, Cooks RG. Mixture analysis by mass-analyzed ion kinetic energy spectrometry. Analytical Chemistry 1976; 48: 2113.

41. Kondrat RW, McClusky GA, Cooks RG. Direct mass spectrometric mixture analysis by negative chemical ionization/mass-analyzed ion kinetic energy spectrometry. Analytical Chemistry 1978; 50: 1222.

42. Horning EC, Horning MG, Carroll I, Dzidic I, Stillwell RN. New picogram detection system based on a mass spectrometer with external ionization source at atmospheric pressure. Analytical Chemistry 1973; 6: 936.

43. Horning EC, Carroll DI, Dzidic I, Haegele K, Lin S-N, Oertli CU, Stillwell RN. Development and use of analytical system based on mass spectrometry. Clinical Chemistry 1977; 23: 13.

44. Torgeson DF, Skowronski RP, Macfarlane RD. New approach to the mass spectroscopy of non-volatile compounds. Biochemical and Biophysical Research Communications 1974; 60: 616.

45. Hughes J, Smith TW, Kosterlitz HW, Fotherhill LA, Morgan BA, Morris HR. Identification of two related pentapeptides from the brain with potent opiate agonist activity. Nature 1975; 258: 577.

46. O'Farrell PH.High resolution two-dimensional electrophoresis of proteins. The Journal of Biological Chemistry 1975; 250: 4007.

47. Anderson L, Anderson NG. High resolution two-dimensional electrophoresis of human plasma proteins. Proceedings of the National Academy of Sciences of the United States of America 1977; 74: 5421.

48. Hunkapiller MW, Hood LE. Direct mirosequencing analysis of polypeptides using an improved sequenator, a nonprotein carrier (Polybrene), and high pressure liquid chromatography. Biochemistry 1978; 17: 2124.

49. Brunnée C. 50 years of MAT in Bremen. Rapid Communications in Mass Spectrometry 1997; 11: 694.

50. Wipf HK, Irving P, McCamish M, Venkataraghavan R, McLafferty FW. Mass spectrometric studies of peptides. V. Determination of amino acid sequences in peptide mixtures by mass spectrometry. Journal of the American Chemical Society 1973; 95: 3369.

51. Kondrat RW, Cooks RG. Direct analysis of mixtures by mass spectrometry. Analytical Chemistry 1978; 50: 81A.

52. Yost RA, Enke CG. Selected ion fragmentation with a tandem quadrupole mass spectrometer. Journal of the American Chemical Society 1978; 100: 2274.

53. Yost RA, Boyd RK. Tandem mass spectrometry: quadrupole and hybrid instruments. Methods in Enzymology 1990; 193: 154.

54. Louris JN, Wright LG, Cooks RG, Schoen AE. New scan modes accessed with a hybrid mass spectrometer. Analytical Chemistry 1985; 57: 2918.

55. Brunnée C. The ideal mass analyzer. International Journal of Mass Spectrometry and Ion Processes 1987; 76: 125.

56. McLafferty FW, Todd PJ, McGilvery DC, Baldwin MA. High resolution tandem Mass Spectrometer (MS/MS) of increased sensitivity and mass range. Journal of the American Chemical Society 1980; 102: 3360.

57. Hewick RM, Hunkapiller MW, Hood LE, Dreyer WJ. A gas liquid solid phase peptide and protein sequenator. The Journal of Biological Chemistry 1981; 256: 7990.

58. Barber M, Bordoli RS, Sedgwick RD, Tyler AN. Fast atom bombardment of solids (F.A.B.): a new ion source for mass spectrometry. Journal of the Chemical Society, Chemical Communications 1981; 325, DOI: 10/1039/C39810000375.

59. Hunt DF, Bone WM, Shabanowitz J, Rhodes J, Ballard JM. Sequence analysis of oligopeptides by secondary ion/collision activated dissociation mass spectrometry. Analytical Chemistry 1981; 53: 1704.

60. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH. Characterization of peptides bound to the class-I Mhc molecule Hla-A2.1 by mass spectrometry. Science 1992; 255: 1261.

61. Dell A, Morris HR. Fast atom bombardment – high field magnet mass spectrometry of 6000 dalton polypeptides. Biochemical and Biophysical Research Communications 1982; 106: 1456.

62. Hakanson P, Kamensky I, Sundqvist B, Fohlman J, Peterson P, McNeal CJ, Mcfarlane RD. 127I-Plasma desorption mass spectrometry of insulin. Journal of the American Chemical Society 1982; 104: 2948.

63. Aberth W, Straub K, Burlingame A. Secondary ion mass spectrometry with cesium ion primary beam and liquid target matrix for analysis of bioorganic compounds. Analytical Chemistry 1982; 54: 2029.

64. Bean MF, Carr SA, Thorne GC, Reilly MH, Gaskell SJ. Tandem mass spectrometry of peptides using hybrid and four-sector instruments: a comparative study. Analytical Chemistry 1991; 63: 1473.

65. Bjelqvist B, Ek K, Righetti PG, Gianazza E, Görg A, Westermeier R, Postel W. Isoelectric focusing in immobilised pH gradients: principle, methodology and some applications. Journal of Biochemical and Biophysical Methods 1982; 6: 317.

66. Walsh KA, Ericsson LH, Parmelee DC, Titani K. Advances in protein sequencing. Annual Review of Biochemistry 1981; 50: 261.

67. Henion J, Thomson BA, Dawson P. Determination of sulfa-drugs in biological fluids by liquid chromatography/mass spectrometry/mass spectrometry. Analytical Chemistry 1982; 54: 451.

68. Desiderio D, Katakuse I. Fast atom bombardment-collision activated dissociation-linked field scanning mass spectrometry of the neuropeptide substance P. Analytical Biochemistry 1983; 129: 425.

69. Yost RA, Fetterolf DD. Tandem Mass Spectrometry (MS/MS) instrumentation. Mass Spectrometry Reviews 1983; 2: 1.

70. Blakely CR, Vestal ML. Thermospray interface for liquid chromatography/mass spectrometry. Analytical Chemistry 1983; 55: 750.

71. Cooks RG, Glish GL, McLuckey SA, Kaiser RE. Ion trap mass spectrometry. Chemical and Engineering News 1991; 69: 26.

72. Jonscher KR, Yates JR III The quadrupole ion trap mass spectrometer – a small solution to a big challenge. Analytical Biochemistry 1997; 244: 1.

73. Morris HR, Panico M, Taylor GW. FAB-mapping of recombinant-DNA protein products. Biochemical and Biophysical Research Communications 1983; 117: 299.

74. Glish GL, Goeringer DE. Tandem quadrupole/time-of-flight instrument for mass spectrometry /mass spectrometry. Analytical Chemistry 1984; 56: 2291.

75. Sundqvist B, Roepstorff P, Fohlman J, Hedin A, Hakansson P, Kamensky I, Lindberg M, Salehpour M, Saweand G. Molecular weight determinations of proteins by californium plasma desorption mass spectrometry. Science 1984; 226: 696.

76. Yamashita M, Fenn JB. Electrospray ion source. Another variation on the free-jet theme. Journal of Physical Chemistry 1984; 88: 4451.

77. Yamashita M, Fenn JB. Negative ion production with the electrospray ion source. Journal of Physical Chemistry 1984; 88: 4671.

78. Thomson BA, Iribarne JV, Dledzic PJ. Liquid ion evaporation/mass spectrometry/mass spectrometry for the detection of polar and labile molecules. Analytical Chemistry 1982; 54: 2219.

79. Whitehouse CM, Dreyer RN, Yamashita M, Fenn JB. Electrospray interface for liquid chromatographs and mass spectrometers. Analytical Chemistry 1985; 57: 675.

80. Meng CK, Mann M, Fenn JB. Of protons or proteins. Zeitschrift für Physik D 1988; 10: 361.

81. Busch KL, Cooks RG. Mass spectrometry of large, fragile, and involatile molecules. Science 1982, 218, 247.

82. Kerr RA. Weight limit for mass spectroscopy raised. Science 1982, 216, 163.

83. Caprioli RM, Fan T, Cottrell JS. Continuous-flow sample probe for fast atom bombardment mass spectrometry. Analytical Chemistry 1986, 58, 2949.

84. Hunt DF, Shabanowitz J, Yates JR, McIver RT, Hunter RL, Syka JEP, Amy J. Tandem quadrupole-Fourier transform mass spectrometry of oligopeptides. Analytical Chemistry 1985, 57, 2728.

85. Cody RB, Amster IJ, McLafferty FW. Peptide mixture sequencing by Tandem Fourier-transform mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America 1985, 82, 6367.

86. Hunt DF, Shabanowitz J, Yates JR, Zhu NZ, Russell DH, Castro E. Tandem quadrupole Fourier-transform mass spectrometry of oligopeptides and small proteins. Proceedings of the National Academy of Sciences of the United States of America 1987, 84, 620.

87. Dell A, Morris HR. Fast atom bombardment – high field magnet mass spectrometry of 6000 dalton polypeptides. Biochemical and Biophysical Research Communications 1982, 106, 1456.

88. Whitehouse CM, Dreyer RN,Yamashita M, Fenn JB. Electrospray interface for liquid chromatographs and mass spectrometers. Analytical Chemistry 1985, 57, 675.

89. Barber M, Green BN. The analysis of small proteins in the molecular weight range 10–24 kDa by magnetic sector mass spectrometry. Rapid Communications in Mass Spectrometry 1987, 1, 80.

90. Karas M, Bachmann D, Hillenkamp F. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. Analytical Chemistry 1985, 57, 2935.

91. Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. Nucleic Acids Res. 1999, 27, 49.

92. Hunt DF, Yates JR III, Shabanowitz J, Winston S, Hauer CR. Protein sequencing by tandemmass spectrometry. Proceedings of the National Academy of Sciences of the United States of America 1986, 83, 6233.

93. Crabb JW, Armes LG, Carr SA, Johnson CM, Roberts GD, Bordoli RS, McKeenan WL. Biochemistry 1986, 25, 4988.

94. Johnson RS, Biemann K. The primary structure of Thioredoxin from Chroamtium vinosum determined by high-performance tandem mass spectrometry. Biochemistry 1987, 26, 1209.

95. Schey K, Cooks RG, Grix R, Wöllnik H. Tandem time-of-flight mass spectrometer for surface induced dissociation. International Journal ofMass Spectrometry and Ion Processes 1987, 77, 49.

96. Louris JN, Cooks RG, Syka JEP, Kelley PE, Stafford GC,. Todd JFJ. Instrumentation, applications and energy deposition in quadrupole ion trap MS/MS spectrometry. Analytical Chemistry 1987, 59, 1677.

97. Karas M, Bachman D, Bahr U, Hillenkamp F. Matrix assisted Ultraviolet laser desorption of non-volatile compounds. International Journal of Mass Spectrometry and Ion Processes 1987, 78, 53.

98. Wong SF, Meng CK, Fenn JB. Multiple charging in electrospray ionization of poly(ethylene glycols). Journal of Physical Chemistry 1988, 92, 546.

99. Meng CK, Mann M, Fenn JB. Of protons or proteins. Zeitschrift Fur Physik D-Atoms Molecules and Clusters 1988, 10, 361.

100. Mann M, Meng CK, Fenn J. Interpreting mass spectra of multiple charged ions. Analytical Chemistry 1989, 61, 1702.

101. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science 1989, 246, 64.

102. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10.000 Daltons. Analytical Chemistry 1988, 60, 2299.

103. Karas M, Bahr U, Ingendoh A, Hillenkamp F. Laser desorption/ionization mass spectrometry of proteins of mass 100.000 to 250.000. Angewandte Chemie-International Edition in English 1989, 28, 760.

104. Karas M, Bahr U, Giessmann U. Matrix-assisted laser desorption ionization mass spectrometry. Mass Spectrometry Reviews 1991, 10, 335.

105. Hillenkamp F, Karas M. Matrix assisted laser desorption/ionization, an experience. International Journal of Mass Spectrometry 2000, 200, 71.

106. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T. Protein and polymer analyses up to m/z 100.000 by laser ionization time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 1988, 2, 151.

107. Beavis RC, Chait BT. Factors affecting the ultraviolet laser desorption of proteins. Rapid Communications in Mass Spectrometry 1989, 3, 233.

108. Beuhler RJ, Flanigan E, Green LJ, Friedman L. Proton transfer mass spectrometry of peptides. Rapid heating technique for underivatized peptides containing arginine. Journal of the American Chemical Society 1974, 96, 3990.

109. Caprioli RM, Moore WT, DaGue B, Martin M. Microbore highperformance liquid chromatography mass spectrometry for the analysis of proteolytic digests by continuous-flow fast-atom bombardment mass spectrometry. Journal of Chromatography 1988, 443, 355.

110. Carroll DI, Dzidic I, Stillwell RN, Haegele KD, Horning EC. Atmospheric pressure ionization mass spectrometry. Corona discharge ion source for use in a liquid chromatograph-mass spectrometer-computer analytical system. Analytical Chemistry 1975, 47, 2369.

111. Thomson BA. Atmospheric pressure ionization and liquid chromatography/mass spectrometry-together at last. Journal of the American Society for Mass Spectrometry 1998, 9, 187.

112. Covey TR, Lee ED, Henion JD. High-speed liquid chromatography/tandem mass spectrometry for the determination of drugs in biological samples. Analytical Chemistry 1986, 58, 2453.

113. Bruins AP, Covey TR, Henion JD. Ion spray interface for combined liquid chromatography/ atmospheric pressure ionization mass spectrometry. AnalyticalChemistry 1987, 59, 2642.

114. Dawson JHJ, Guilhaus M. Orthogonal-acceleration time of flight mass spectrometer. Rapid Communications in Mass Spectrometry 1989, 3, 155.

115. Guilhaus M, Selby D, Mlynski V. Orthogonal acceleration time of-flight mass spectrometry. Mass Spectrometry Reviews 2000, 19, 65.

116. Henry KD, Williams ER, Wang BH, McLafferty FW, Shabanowitz J, Hunt DF. Fourier-transform mass spectrometry of large molecules by electrospray ionization. Proceedings of the National Academy of Sciences of the United States of America 1989, 86, 9075.

117. Loo JA, Edmonds CG, Smith RD. Primary sequence information from intact proteins by electrospray ionization Tandem mass spectrometry. Science 1990, 248, 201.

118. Nohmi T, Fenn JB. High mass ions. Howhigh is high? Proceedings of the 38th Conference on Mass Spectrometry and Allied Topics, Tucson, 1990.

119. Chowdhury SK, Katta V, Chait BT. Probing conformational changes in protein by mass spectrometry. Journal of the American Chemical Society 1990, 112, 9012.

120. Ganem B, Li JD, Henion JD.Detection of noncovalent receptorligand complexes by mass spectrometry. Journal of the American Chemical Society 1991, 113, 6294.

121. Ganem B, JD Li, Henion JD. Observation of noncovalent enzymesubstrate and enzyme-product complexes by ion-spray mass spectrometry. Journal of the American Chemical Society 1991, 113, 7818.

122. Katta V, Chait BT. Observation of the hemoglobin complex in native myoglobin by electrospray-ionization mass spectrometry. Journal of the American Chemical Society 1991, 113, 7818.

123. Katta V, Chowdhury SK, Chait BT. Use of single quadrupole mass spectrometer for collision-induced dissociation studies of multiply charged peptide ions produced by electrospray ionization. Analytical Chemistry 1991, 63, 174.

124. Covey TR, Huang EC, Henion JD. Structural characterization of protein tryptic peptides via liquid chromatography/mass spectrometry and collision-induced dissociation of their doubly charged molecular ions. Analytical Chemistry 1991, 63, 1193.

125. Dodonov AF, Chernushevich IV, Laiko VV. Atmospheric pressure ionization time-of-flight mass spectrometer. Abstract 153, International Mass Spectrometry Conference, Amsterdam, 1991.

126. Mirgorodskaya OA, Shevchenko AA, Chernushevic IV, Dodonov AF, Miroshnikov AI. Electrospray ionization time of-flight mass spectrometry in protein chemistry. Analytical Chemistry 1994, 66, 99.

127.Boyle JG, Whitehouse CM, Fenn JB. An ion-storage time-of flight mass spectrometer for analysis of electrospray ions. Rapid Communications in Mass Spectrometry 1991, 9, 400.

128. Wiley WC, McLaren IH. Time-of-flight mass spectrometer with improved resolution. The Review of Scientific Instruments 1955, 26, 1150

129. Chaurand P, Luetzenkirchen F, Spengler B. Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and MALDI-post-source decay time-of-flight mass spectrometry. Journal of the American Society for Mass Spectrometry 1999, 10, 91.

130. Henzel WJ, Watanabe C, Stults JT. Protein identification, the origins of peptide mass fingerprinting. Journal of the American Society for Mass Spectrometry 2003, 14, 931.

131. Cleveland DW, Fischer SG, Kirschner MW, Laemmli UK. Peptide mapping by limited proteolysis in sodium dodecyl sulphate and analysis by gel electrophoresis. Journal of Biological Chemistry 1977, 252, 1102.

132. Eng JK, McCormack AL, Yates JR III An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry 1994, 5, 976.

133. Pappin DJC, Hojrup P, Bleasy AJ. Rapid identification of proteins by peptide-mass fingerprinting. Current Biology 1993, 3, 327.

134. Mann M, Hojrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biological Mass Spectrometry 1993, 22, 338.

135. Henzel WJ,. Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proceedings of the National Academy of Sciences 1993, 90, 5011.

136. James P, Quadroni M, Carafoli E, Gonnet G. Protein profiling by mass profile fingerprinting. Biochemical and Biophysical Research Communications 1993, 195, 58.

137. Chait B, Wang R, Beavis RC, Kent SBH. Protein ladder sequencing. Science 1993, 262, 89.

138. Yates JR III, Speicher S, Griffin PR, Hunkapiller T. Peptide mass maps, a highly informative approach to protein identification. Analytical Biochemistry 1993, 214, 397.

139. Beu SC, Senko MW, Quinn JP, Wampler FM III, McLafferty FW. Fourier-transform electrospray instrumentation for tandem high-resolution mass spectrometry of large molecules. Journal of the American Society for Mass Spectrometry 1993, 4, 557.

140. Hutchens TW, Yip TT. New desorption strategies for the mass spectrometric analysis of macromolecules. Rapid Communications in Mass Spectrometry 1993, 7, 576.

141. Mann M, Wilm M. Error tolerant identification of peptides in sequence databases by peptide sequence tags. Analytical Chemistry 1994, 66, 4390.

142. Cox AL, Skipper J, Chen Y, Henderson RA, Darrow TL, Shabanowitz J, Engelhard VH, Hunt DF, Slingluff CL Jr. Identification of a peptide recognized by five melanoma-specific human cytotoxic T cell lines. Science 1994, 264, 716.

143. Clauser KR, Baker PR, Burlingame AL. Peptide fragment-ion tags from MALDI/PSD for error tolerant searching of genomic databases. 44th ASMS Conference on Mass Spectrometry and Allied Topics, Seattle, Washington, D.C. May 28th–June 1st 1996.

144. Verentchikov A, Ens W, Standing KG. Reflecting time-of flight mass spectrometer with an electrospray ion source and orthogonal extraction. Analytical Chemistry 1994, 66, 126.

145. Kaufmann R, Kirsch D, Spengler B. Sequencing of peptides in a time-of-flight mass spectrometer, evaluation of postsource decay following matrix-assisted laser desorption ionisation (MALDI). International Journal of Mass Spectrometry and Ion Processes 1994, 131, 355.

146. Cornish TJ, Cotter RJ. Curved field reflectron time-of-flight mass spectrometer for the simultaneous focusing of metastable product ions. Rapid Communications in Mass Spectrometry 1994, 8, 781.

147. Emmett MR, Caprioli RM. Micro-electrospray mass spectrometry, ultra-high sensitivity analysis of peptides and proteins. Journal of the American Society for Mass Spectrometry 1994, 5, 605.

148. Wilm M, Mann M. Analytical properties of the nanoelectrospray ion source. Analytical Chemistry 1996, 68, 1.

149. Wilm MS, Mann M. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? International Journal of Mass Spectrometry and Ion Processes 1994, 136, 167.

150. Valskovic GA, Kelleher NL, Little DP, Aaserud DJ, McLafferty FW. Attomole-sensitivity electrospray source for large-molecule mass spectrometry. Analytical Chemistry 1995, 67, 3802.

151. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I. Progress with gene-product mapping of the Mollicutes, Mycoplasma genitalium. Electrophoresis 1995, 16, 1090.

152. Wood TD, Chen LH, Keheller NL, Little DP, Kenyon GL, McLafferty FW. Direct sequence data from heterogeneous Creatine Kinase (43 kDa) by high-resolution tandem mass spectrometry. Biochemistry 1995, 34, 16251.

153. Wilm M, Schevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T, Mann M. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature 1996, 379, 466.

154. Valaskovic GA, Kelleher NL, McLafferty FW. Attomole protein characterization by capillary electrophoresis-mass spectrometry. Science 1996, 273, 1199.

155. Hofstadler SA, Severs JC, Smith RD,.Swanek FD, Ewing AG. Analysis of single cells with capillary electrophoresis electrospray ionization fourier transform ion cycloton resonance mass spectrometry. Rapid Communications in Mass Spectrometry 1996, 10, 919.

156. Ünlu M, Morgan ME, Minden JS. Difference gel electrophoresis, a single gel method for detecting changes in protein extracts. Electrophoresis 1997, 18, 2071.

157. Yates JR, McCormack AL, Schieltz D, Carmack E, Link A Direct analysis of protein mixtures by Tandem mass spectrometry. Journal of Protein Chemistry 1997, 16, 495.

158. Morris HR, Paxton T, Panico M, McDowell R, Dell A. A novel geometry mass spectrometer, the Q-TOF, for low femtomole/attomole-range biopolymer sequencing. Journal of Protein Chemistry 1997, 16, 469.

159. Shevchenko A, Chernushevich I, Ens W, Standing KG, Thomson B, Wilm M, Mann M. Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labelling and a quadrupole/time of flight mass spectrometry. Rapid Communications in Mass Spectrometry 1997, 11, 1015.

160. Caprioli RM, Farmer TB, Gile J. Molecular imaging of biological samples, localization of peptides and proteins using MALDI-TOF MS. Analytical Chemistry 1997, 69, 4751.

161. Zubarev RA, Kelleher NL, McLafferty FW. Electron capture dissociation of multiply charged protein cations. A nonergodic process. Journal of the American Chemical Society 1998, 120, 3265.

162. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America 2004, 101, 9528.

163. Oosterkamp AJ, Gelpí E, Abian J. Quantitative peptide bioanalysis using column-switching nano liquid chromatography/mass spectrometry. Journal of Mass Spectrometry 1998, 33, 976.

164. Abian J, Oosterkamp AJ, Gelpí E. Comparison of conventional, narrow-bore and capillary liquid chromatography/mass spectrometry for electrospray ionization mass spectrometry, Practical considerations. Journal of Mass Spectrometry 1999, 34, 244.

165. Wei J, Buriak JM, Siuzdak G. Desorption-ionization mass spectrometry on porous silicon. Nature 1999, 399, 243.

166. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. Evaluation of two dimensional gel electrophoresis-based proteome analysis technology. Proceedings of the National Academy of Sciences 2000, 97, 9390.

167. Oda Y, Huang K, Cross FR, Couburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. Proceedings of the National Academy of Sciences of the United States of America 1999, 96, 6591.

168. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope coded affinity tags. Nature Biotechnology 1999, 17, 994.

169. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nature Biotechnology 1999, 17, 1030.

170. Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML, Burlingame AL. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF Tandem mass spectrometer. Analytical Chemistry 2000, 72, 552.

171. Loboda AV, Krutchinsky AN, Bromirski M, Ens W, Standing KG. A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source, design and performance. Rapid Communications in Mass Spectrometry 2000, 14, 1047.

172. Laiko VV, Baldwin MA, Burlingame AL. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. Analytical Chemistry 2000, 72, 652.

173. Laiko VV, Moyer SC, Cotter RJ. Atmospheric pressure MALDI/ion trap mass spectrometry. Analytical Chemistry 2000, 72, 5239.

174. Moyer AC, Cotter RJ. Atmospheric pressure MALDI. Analytical Chemistry 2002, 469A, 469A–476A.

175. Peng J, Gygi SP. Proteomics: the move to mixtures. Journal of Mass Spectrometry 2001, 36, 1083.

176. Syka JP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, Proc. Natl. Acad. Sciences 2004; 101: 9528-9533

177. Mikesh LM, Ueberheide B, Chi A, Coon JJ, Syka JE, Shabanowitz J, Hunt DF. The utility of ETD mass spectrometry in proteomic analysis. Biochim Biophys Acta. 2006;1764:1811-22.

178. Wiesner J, Premsler T, Sickmann A. Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. Proteomics. 2008 Nov;8(21):4466-83.

179. Takáts Z, Wiseman JM, Gologan B, Cooks RG, Mass Spectrometry Sampling

Under Ambient Conditions with Desorption Electrospray Ionization, Science 2004; 306: 471- 473.

180. Cody RB, Laramée JA, Dupont Durst H. Versatile New Ion Source for the Analysis of

Materials in Open Air under Ambient Conditions, Anal. Chem. 2005; 77: 2297-2302

181. Zhou M, Guan W, DeEtte Walker L, Mezencev R, Benigno BB, Gray A, Fernández FM, McDonald JF, Rapid Mass Spectrometric Metabolic Profiling of Blood Sera Detects Ovarian Cancer with High Accuracy, Cancer Epidemiol Biomarkers Prev. 2010; 19: 2262–71.

182. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks G R. The Orbitrap: a new mass spectrometer. J Mass Spectrom. 2005;40:430-43.

183. Michaelevski I, Kirshenbaum N, Sharon M. T-wave ion mobility-mass spectrometry: basic experimental procedures for protein complex analysis. J Vis Exp. 2010; 31: 41.

Consideraciones sobre la interpretación de espectros de masas

08

Juan J. Calvete

Juan J. Calvete, Laboratorio de Proteinómica Estructural, Instituto de Biomedicina de Valencia.

La identidad de todo compuesto está definida en su estructura química. La espectrometría de masas es una potente herramienta para investigar la estructura de los compuestos químicos, incluídas la masa y la ordenación secuencial de sus elementos constituyentes. Sin embargo, los analizadores de masas solo pueden detectar iones. Las partículas neutras, sean moléculas o radicales, representan *materia oscura* para el analizador de masas y son eliminadas por el sistema de vacío del espectrómetro de masas. Los iones son partículas cargadas y, por tanto, su posición en el espacio puede ser manipulada mediante campos eléctricos y magnéticos. Conceptualmente, pues, un espectrómetro de masas puede considerarse un manipulador de iones. La interpretación del espectro de masas es la clave para la asignación estructural del correspondiente ión. Y, por complicado que pudiera parecer, interpretar espectros de masas es una cuestión sencilla: sólo se requiere aplicar rigurosamente unas pocas reglas generales derivadas de los procesos de ionización y fragmentación de los analitos y un nivel de cálculo de matemáticas elementales. De eso trata esta revisión.

1. Brevísima historia del nacimiento de la espectrometría de masas.

Durante la última década del siglo XIX, el científico británico Joseph John Thomson (1856-1940) realizó una serie de experimentos con rayos catódicos que condujeron al descubrimiento del electrón [1]. Utilizando un tubo de rayos catódicos en alto vacío, Thomson demostró que los rayos podían deflectarse bajo la influencia de un campo magnético o eléctrico y que la carga negativa era una propiedad intrínseca de dichos rayos. En 1897, Thomson determinó que la relación m/z de los rayos catódicos era más de mil veces menor que la del ión H^+ (hoy sabemos que un protón es unas 1836 veces más masivo que un electrón), sugiriendo que se trataba de partículas muy pequeñas o altamente cargadas [1]. Esta revolucionaria conclusión implicaba que los rayos catódicos estaban constituídos por partículas subatómicas emitidas por los átomos del electrodo, a las que Thomson denominó “corpúsculos”, y que - por tanto- los átomos (del griego *a*τομον, *indivisible*) no eran las entidades elementales de la materia postuladas por la escuela atomista en la Antigua Grecia. Los *corpúsculos* descubiertos por Thomson resultaron ser *electrones*, término acuñado en 1891 por el físico irlandés George Johnstone Stoney (1826-1911) para “la unidad fundamental de carga electrica”.

En 1913, en el marco de sus investigaciones sobre rayos anódicos, Thomson midió la influencia de campos eléctrico y magnético sobre un haz de neón ionizado, observando su descomposición en dos parábolas de deflección. Thomson concluyó que el neón está compuesto por una mezcla de dos isótopos de masas atómicas 20 y 22. Esta observación representó la primera evidencia de la existencia de isótopos de un elemento estable. La separación de los isótopos del neón se considera el experimento que supuso el nacimiento de la espectrometría de masas, aunque la técnica, como método analítico capaz de determinar la composición isotópica de los elementos, fué posteriormente desarrollada por Arthur Jeffrey Dempster y Francis William Aston [2] entre 1918 y 1919. Thomson (Premio Nobel de Física de 1906), Demster, y Aston (Premio Nobel de Química de 1922) son considerados los fundadores de la espectrometría de masas moderna [3].

2. Unidades de m/z y el defecto de masa atómica.

Aunque resulte paradójico, los analizadores (espectrómetros) de masas no miden la masa sino la relación masa/carga (m/z) de compuestos ionizados, siendo "m" la masa molecular del compuesto y "z" un múltiplo entero de la carga elemental (e) del electrón. La relación m/z es adimensional al estar basada en la escala de Unidades Atómicas (UA) de Hartree (http://en.wikipedia.org/wiki/Atomic_unit) desde que en 1961 la IUPAC adoptó como unidad de masa atómica (uma) la doceava parte de la masa del ^{12}C, el isotopo natural más abundante del carbono, como base para asignar la masa atómica relativa de todos los demás elementos químicos. De acuerdo a esta convención, la masa de 1 átomo de ^{12}C es 12 uma. También por definición, la unidad de carga eléctrica, e, en la escala UA es exactamente igual a 1.

La denominación más frecuentemente utilizada para la unidad de masa atómica (uma) es el "Dalton" (Da), término acuñado en honor de John Dalton (1766-1844), químico y físico británico, quién en 1803 formuló la *Ley de las Proporciones Múltiples* que postula que los pesos de dos elementos siempre se combinan entre sí en proporciones de números enteros pequeños. En ese mismo año Dalton publicó su primera lista de pesos atómicos y símbolos, y en 1808 expuso la teoría atómica en la que se basa la física moderna. 1 Da se corresponde aproximadamente con la masa de un protón (o un átomo de Hidrógeno, ^{1}H). En honor del descubridor del electrón y fundador de la espectrometría de masas, Cooks y Rockwook propusieron en 1991 el Thomson (Th) como unidad de m/z [4]. 1 Th equivale a 1 Da/e y, por tanto, para iones monocargados, y sólo en este caso, la masa en Da de un ión coincide con su relación m/z.

1 gramo equivale a la masa 1 mol (N_A) de unidades de masa atómica. Así pues, un mol de átomos de ^{12}C tiene una masa de 12 g y, por tanto, el valor de 1 uma en unidades del SI se obtiene mediante la fórmula $12\ g/(12N_A) = 1/(6{,}02214199 \times 10^{23}) = 1{,}66053873 \times 10^{-27}$ kg. Un átomo de ^{12}C está formado por 6 protones, 6 neutrones y 6 electrones. Las masas experimentales en reposo del protón (m_p), neutrón (m_n) y electrón (m_e) son $1{,}67265 \times 10^{-27}$ kg (1,0072935 u.m.a), $1{,}67495 \times 10^{-27}$ kg (1,0086786 uma) y $9{,}10938215 \times 10^{-31}$ kg (0,0005485799 uma), respectivamente. La masa atómica asignada al ^{12}C (m_{asig}, 12,0000) es menor que su masa atómica calculada (m_{calc}) a partir de su composición subatómica, incluso tomando en consideración el efecto relativista de pérdida de masa debido a la energía ($E = mc^2$) liberada en la formación del núcleo atómico.

El defecto de masa (m_{def}) se calcula mediante la ecuación: $m_{def} = m_{asig} - m_{calc} = 12{,}0000 - (1{,}0072935\ n_p + 1{,}0086786\ n_n + 0{,}0005485799\ n_e) = -0{,}0958$ uma, siendo n_p, n_n y n_e el número de protones, neutrones y electrones constituyentes del átomo. Por convención, en la escala de UA el defecto de masa del ^{12}C se cancela aproximando las masas del protón y del

neutrón a 1 uma [5]. De esta manera la masa atómica del ^{12}C (12 uma) coincide con su número másico (6p + 6n). Las masas de todos los demás átomos calculadas en relación a la masa del ^{12}C se desvían, por exceso o defecto, de su número de nucleones. En el caso de los átomos más abundantes en las biomoléculas la desviación entre masa atómica y número másico es: ^{14}N, +0.00307 uma; ^{16}O, -0.00509 uma; ^{1}H, +0.007825 uma; ^{31}P, 0,02624 uma; ^{32}S, 0,02793 uma (Tabla). Aunque pequeñas, estas diferencias deben tenerse en cuenta a la hora de comparar masas experimentales determinadas por espectrometría de masas con valores teóricos calculados sumando las masas atómicas de los átomos constituyentes del ión. Además, la comparación entre masa experimental y teórica sólo tiene sentido si ambas magnitudes están expresadas en las mismas unidades, que en el caso que nos ocupa son unidades de masa atómica relativas a la del ^{12}C.

3. Iones quasimoleculares.

El cálculo de la masa molecular de un compuesto a partir de su espectro de masas requiere también un entendimiento del proceso de ionización (¿resultó alterada la composición y/o la ordenación atómica del compuesto al ionizarse?), así como la determinación de la carga (z) del correspondiente ión. En el caso de moléculas biológicas, las técnicas de ionización más utilizadas son la electronebulización (*electrospray*, ESI) [6] y la desorción por radiación láser asistida por una matríz (MALDI, *matrix-assisted laser-desorption ionization*) [7,8]. ESI y MALDI se engloban en el conjunto de técnicas de ionización "suave" (*soft ionization*) caracterizados por producir iones de analitos intactos que no muestran una tendencia significativa a fragmentarse espontáneamente. Una descripción pormenorizada de los principios operacionales de estas técnicas excede los objetivos de este artículo. Baste mencionar aquí que tanto ESI como MALDI producen iones quasimoleculares $(M+nH)^{n+}$ y $(M-nH)^{n-}$ por captación o pérdida de protones [9]. La detección de iones quasimoleculares positivos o negativos está en función de la polaridad de medida del analizador de masa seleccionada por el operario, y el número de protones (nH^+) captados o cedidos por el analito depende tanto de la estructura del compuesto como de las condiciones de ionización empleado. La tendencia de un compuesto (Y) por generar iones protonados es proporcional a su basicidad relativa (ΔG_B) en la fase gaseosa, considerando la basicidad en el sentido empleado por Lewis-Brønsted, como una medida de la tendencia del ión a captar un protón. En el equilibrio $XH^+ + Y \rightleftarrows X + YH^+$ (donde XH^+ representa un donante de protones, como es el solvente en ESI y la matríz en MALDI), $\Delta G_B = -RT \ln ([X]\,[YH^+])/([Y]\,[XH^+])$. Por tanto, las condiciones experimentales para producir iones protonados mediante ESI o MALDI deben ajustarse para favorecer el desplazamiento del equilibrio hacia la derecha. MALDI tiende a producir iones monocargados mientras que ESI genera series de iones multicargados. Más adelante veremos como esta última circunstancia posibilita la determinación de la masa de macromoléculas en instrumentos cuyo rango de medida de m/z (< 4000) es muy inferior a la masa molecular del analito..

En experimentos con mezclas se observa frecuentemente el fenómeno de supresión de la ionización de compuestos de baja basicidad relativa por aquellos de más alta ΔG_B. Conviene recordar este hecho pues un espectro de masas es un refleja la intensidad relativa de los iones detectados, qué no tiene por qué equivaler a "la abundancia relativa de los compuestos en la muestra analizada"... En el caso polipéptidos, los protones se adicionan a sitios de alta basicidad (grupos amino terminal, ε-amino de lisinas, guanidino de argininas, e imidazol de histidinas), o se liberan de grupos ácidos (aspártico, glutámico, grupo fenol de la tirosina, fosfórico, etc.), generándose, en ambos casos los iones quasimoleculares $M+H^+$ y $M-H^-$, respectivamente. La masa molecular (M) de estos iones puede calcularse a partir de las ecuaciones $M = [(m/z) \times z] - nm_{H^+}$ ó $M = [(m/z) \times z] + nm_{H^+}$, dependiendo de que la medida de m/z se haya realizado

en modo positivo o negativo, respectivamente, y donde nm_H^+ es la masa de los protones captados o cedidos en la ionización. Pero no nos adelantemos, porque antes de pasar a describir cómo se deduce una masa molecular a partir de un espectro de masas debemos hacernos la siguiente pregunta: ¿cada compuesto químico tiene una masa molecular única? En otras palabras, ¿cómo se calcula la masa molecular de un compuesto a partir de su fórmula química? La respuesta inmediata es "sumando las masas de todos los átomos de la estructura", y ahí es donde reside el *quid* de la cuestión: ¿qué valores de masa atómica debemos considerar? Para responder a esta pregunta retrocedamos a los orígenes de la materia...

4. Isótopos y distribución de masa molecular.

La nucleosíntesis ocurrió en los primeros tres minutos despues del Big Bang cuando el Universo en expansión se hubo enfriado lo suficiente como para formar protones y neutrones estables [10,11]. La nucleosíntesis del Big Bang es responsable de las relaciones de abundancia de los elementos más ligeros que el Berilio, ^{1}H, ^{2}H, ^{3}He, ^{4}He, y ^{7}Li. El resto de los elementos de la tabla periódica se sintetizaron posteriormente mediante procesos de nucleosíntesis estelar, mediante el ciclo del carbono-nitrógeno-oxígeno (CNO) propuesto en 1938 por Hans Bethe [12] y cuyas ecuaciones más básicas son:

$$
\begin{array}{lcl}
{}^{4}He + {}^{4}He & \leftrightarrow & {}^{8}Be \\
{}^{8}Be + {}^{4}He & \leftrightarrow & {}^{12}C + \gamma \\
{}^{12}C + {}^{4}He & \rightarrow & {}^{16}O + \gamma \\
{}^{12}C + {}^{1}H & \rightarrow & {}^{13}N + \gamma \\
{}^{13}N & \rightarrow & {}^{13}C + e^{+} + \nu_e \\
{}^{13}C + {}^{1}H & \rightarrow & {}^{14}N + \gamma \\
{}^{14}N + {}^{1}H & \rightarrow & {}^{15}O + \gamma \\
{}^{15}O & \rightarrow & {}^{15}N + e^{+} + \nu_e \\
{}^{15}N + {}^{1}H & \rightarrow & {}^{12}C + {}^{4}He
\end{array}
$$

Las reacciones de nucleosíntesis por fusión nuclear en estrellas de masa menor a 1.4 masas solares (límite de Chandrasekhar) sólo producen elementos hasta el ^{56}Fe, el núcleo atómico más estable; los elementos más pesados se producen por procesos de captura neutrónica en estrellas más masivas. Una consecuencia de la nucleosíntesis estelar es la existencia de isótopos. Nombrados a partir del griego *isos*, mismo, y *topos*, lugar, debido a que ocupan el mismo lugar en la Tabla Periódica de los Elementos, los isótopos son átomos que tienen el mismo número atómico (número de protones en el núcleo), pero diferente número másico (suma del número de neutrones y el de protones en el núcleo). Difieren pues en el número de neutrones, por lo que los isótopos tienen el mismo número de electrones (y carga eléctrica) y, por tanto, difieren en masa pero poseen las mismas propiedades químicas.

Exceptuando los elementos radiactivos, la composición isotópica de los átomos que se encuentran en la naturaleza es esencialmente constante (al menos desde que sabemos como determinarla). La siguiente Tabla muestra los isótopos estables de los principales átomos que forman las estructuras covalentes de toda la materia biológica conocida [13]:

Elemento	Isótopo	Abundancia (%)	Masa atómica (uma)
Hidrógeno	^{1}H	99,9885	1,007825
	^{2}H	0,0115	2,014101
Carbono	^{12}C	98,93	12,00000
	^{13}C	1,07	13,00335
Nitrógeno	^{14}N	99,632	14,00307
	^{15}N	0,368	15,00010
Oxígeno	^{16}O	99,757	15,99491
	^{17}O	0,038	16,99491
	^{18}O	0,205	17,99716
Fósforo	^{31}P	100	30,97376
Azúfre	^{32}S	95,02	31,97207
	^{33}S	0,75	32,97145
	^{34}S	4,21	33,96786
	^{36}S	0,02	35,96708

Las masas atómicas de los átomos de elementos di- o poliisotópicos de la Tabla Periódica (http://tablaplus.awardspace.com/Tabla.html) se calculan como la media ponderada de las masas de sus distintos isótopos teniendo en cuenta la abundancia relativa de cada uno de ellos. Consideremos el carbono como ejemplo. Este elemento consta de dos isótopos estables (Tabla): ^{12}C (abundancia relativa, 98,93%) y ^{13}C (1,07%). La suma ponderada de las masas de cada isótopo da: (0,9893 x 12,000) + (0,0107 x 13,00335) = 12,0107 uma.

La composición química de toda la materia de nuestro planeta refleja las relaciones de abundancia de los átomos de la nube primigenia de polvo y gas expulsada al espacio en explosiones de Supernova y que constituyó la materia prima a partir de la cual se formó el Sistema Solar. Ello implica que, por ejemplo, por término medio, de cada 100.000 átomos de carbono 1070 son ^{13}C. Igualmente, existen, respectivamente, 11,5 átomos de ^{2}H y 368 de ^{15}N en cada 100.000 átomos de hidrógeno y nitrógeno. Esta circunstancia equivale a decir que las moléculas no poseen una masa molecular única sino una distribución de masas. En el caso de moléculas que, como las proteínas y otras macromoléculas biológicas, están constituídas por elementos cuyos isótopos más abundantes son los de menor masa atómica (Tabla), el primer pico (M) de la distribución isotópica corresponde al ión *monoisotópico*. Este ión es único en la distribución isotópica al ser una especie homogénea constituida por los isótopos, ^{1}H, ^{12}C, ^{14}N y ^{16}O, ^{32}S, etc. Todos los demás picos corresponden a poblaciones heterogéneas de iones que contienen combinaciones de isótopos sucesivamente más pesados. Por ejemplo, el pico M+1 (segundo pico o pico del primer isotopo) corresponde a iones que contienen un ^{13}C o un ^{15}N o un ^{33}S etc. y, por tanto, aproximadamente una unidad de masa más que el monoisotópico; los iones M+2 (pico del segundo isótopo) poseen 2 ^{13}C ó 2 ^{15}N ó 1 ^{18}O, ó (1 ^{13}C + 1 ^{15}N) ó (1 ^{13}C + 1 ^{33}S, etc..) y son aproximadamente 2 unidades de masa atómica (uma) más pesados que M; y los iones M+3, ... M+n poseen, sucesivamente, masas que exceden desde ~ 3 hasta "n" uma a la masa del ión monoisotópico. Es importante insistir en el carácter aproximado de los cálculos anteriores, dado que la diferencia de masa entre isótopos consecutivos de diferentes átomos no es exactamente igual a 1 uma, ni siquiera es la misma para todos los átomos poliisotópicos: ^{13}C - ^{12}C = 1,00335 uma; ^{15}N - ^{14}N = 0,99703 uma; ^{2}H - ^{1}H = 1,006276 uma, etc. (Tabla). De hecho, en analizadores de masa de muy alta resolución los picos M+1, M+2, etc. llegan a desdoblarse en los iones correspondientes a los isótopos de los diferentes átomos constituyentes[1].

[1]. Por esta razón se utilizan las masas de iones monoisotópicos de péptidos trípticos para identificar proteínas mediante experimentos de huella peptídica (*peptide mass fingerprinting*, PMF). La especificidad de este método, desarrollado en 1993 independientemente por varios grupos de investigación [16-20], depende criticamente de la exactitud de las masas utilizados para interrogar al banco de datos de referencia. El ión monoisotópico es la única especie cuya masa está perfectamente definida por su composición atómica y, por tanto, su medida conlleva la mínima dispersión respecto a los demás picos iónicos de la distribución isotópica.

Una consecuencia de la disparidad de incrementos de masa entre los isótopos consecutivos de diferentes átomos es que compuestos de fórmulas empiricas diferentes también difieren en su masa exacta. En otras palabras: la masa, medida con suficiente exactitud, es una característica única y discriminatoria de compuestos no isoméricos.

La abundancia relativa de los picos de la distribución isotópica de una molécula puede calcularse a partir de su fórmula química mediante el desarrollo en sumandos de un polinomio del tipo $(a_1 + a_2 + a_3 + ...)^X (b_1 + b_2 + b_3 + ...)^Y (c_1 + c_2 + c_3 + ...)^Z$ ..., representando las series "$n_1 + n_2 + n_3 + ...$" la abundancia relativa de los isótopos del elemento "n" y los exponentes X, Y, Z, etc. el número de átomos del correspondiente elemento en la fórmula [14,15]. La distribución isotópica de un péptido sencillo como es el glucagón, de 29 resíduos, cuya fórmula es $C_{153}H_{224}N_{42}O_{50}S$, contiene un enorme número de términos: $(0,9889 + 0,0107)^{153} (0,999885 + 0,000115)^{224} (0,99632 + 0,00368)^{42} (0,99757 + 0,00038 + 0,00204)^{50} (0,9502 + 0,0075 + 0,0421 + 0,0002)$, que hace su cálculo inviable manualmente. Existen, no obstante, sencillos programas de ordenador para computar por deconvolución matemática la distribución isotópica teórica a partir de la fórmula empírica (ej. http://www.chemcalc.org/).

Es mucho más sencillo calcular la altura relativa de los picos isotópicos. Así, el cociente $(M+1)/M = \Sigma(\%I \times nE)/(100 - \%I)$, siendo %I la abundancia relativa del elemento E y nE el número de átomos de ese elemento en la molécula. Para el glucagón, la abundancia relativa del pico isotópico (M+1) respecto al ión monoisotópico es: (1,07 x 153)/0,9893 + (0,0115 x 224)/0,999885 + (0,368 x 42)/0,99632 + (0,038 x 50)/0,99962 + (0,205 x 50)/0,99795 + 0,75 = 196,49%. Puede demostrarse facilmente que la abundancia relativa de los picos isotópicos se desplaza hacia la derecha a medida que aumenta la masa molecular del compuesto. Este punto queda ilustrado en la Fig. 1.

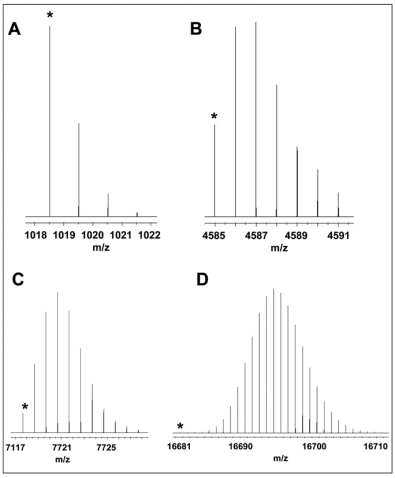


Figura 1. Simulación por deconvolución con ChemCalc (http://www.chemcalc.org) de las distribuciones isotópicas de (A) un nonapéptido (M_{mono}, 1018.5 Da) de fórmula empírica $C_{44}H_{70}N_{14}O_{12}S$, (B) un polipéptido de ~ 4.5 kDa de masa molecular y fórmula $C_{186}H_{293}N_{59}O_{61}S_8$, (C) una proteína de ~7 kDa ($C_{318}H_{502}N_{102}O_{119}S_{14}$) y (D) una proteína de masa molecular monoisotópica 16.681,6 Da y fórmula $C_{1253}H_{224}N_{42}O_{50}S$. Se aprecia claramente el corrimiento hacia la derecha de la intensidad relativa de los picos isotópicos a medida que aumenta la masa molecular. En A-D, el ión monoisotópico está señalado con un asterisco.

La distribución isotópica es una característica diferencial de cada molécula poliatómica compuesta por elementos poliisotópicos. Así, dos compuestos de igual masa molecular pero diferente fórmula química difieren en sus diagramas de distribución isotópica. Este hecho es ampliamente utilizado en Química Analítica como prueba de la correcta asignación estructural de un compuesto [9]. La determinación de la distribución isotópica es la única técnica apta para la detección de pequeñas alteraciones isotópicas en muestras geológicas y medioambientales, que pongan de manifiesto procesos naturales o antropogénicos sobre el entorno, con incertidumbres inferiores al 1‰. Asimismo, la manipulación de la distribución isotópica de una molécula mediante técnicas de marcaje isotópico ha permitido el desarrollo de métodos de comparación de la abundancia relativa de proteínas (ej. ICAT, iTRAQ, SILAC, etc.) con los que

explorar la dinámica molecular del proteoma celular [21,22]. Estas metodologías son de gran relevancia para detectar cambios en la expresión de proteínas asociados a procesos fisiológicos o patológicos celulares. Además, el uso de trazadores isotópicos, enriquecidos en algún isótopo minoritario en la naturaleza, permite medidas absolutas de cantidad de sustancia mediante el método de dilución isotópica [23]. Esta técnica en conjunción con la espectrometría de masas con fuente de plasma (ICP-MS) constituye una herramienta imprescindible para la detección y cuantificación de heteroátomos [24] que, cómo el fósforo, desempeñan un papel fundamental en la regulación de procesos biológicos. Por su relevancia, una descripción detallada de los protocolos de proteómica cuantitativa merece un capítulo aparte.

5. Resolución, precisión, exactitud y reproducibilidad.

La capacidad de un espectrómetro de masas de distinguir entre las masas de dos picos adyacentes se expresa normalmente en términos de su resolución. Con una resolución adecuada, un espectrómetro de masas puede separar los picos (MH^+, MH^++1, MH^++2...) de una distribución isotópica (Fig.2). La resolución (R) se define como M/ΔM. ΔM es la diferencia entre las masas de dos picos adyacentes que están resueltos y M corresponde a la masa sin decimales del primer pico. Dos picos próximos se consideran resueltos cuando la altura del valle entre ellos no sobrepasa un determinado porcentaje, que debe ser especificado, de la altura media de ambos. Así, utilizando esta definición de la resolución, un espectrómetro capaz de resolver picos que tengan valores de m/z de 400,0 y 400,1 debería tener una R ≥ 4000. El valor de la resolución calculada de esta manera varía para diferentes rangos de masa del mismo espectro: en el espectro de la Fig.2, R = 1846 para m/z 1846 pero 3255 para m/z 3255. Por ello, prefiere definirse ΔM como el incremento de m/z a una determinada altura (comunmente 50%) del pico M. En la literatura en lengua inglesa, el término estándar para este valor es "FWHM" (*full width at half mass*). Esta manera de calcular R proporciona valores esencialmente constantes en todo el rango de m/z. El espectro de la Fig.2 fue adquirido con una resolución de ~9000, suficiente para resolver la distribución isotópica de los iones resaltados (R = 1846/0,2 = 9230; R = 3255/0,36 = 9041). En el ejemplo de la Fig.3, la resolución obtenida en el espectro del panel A fue 665,3/0,24= 2.772, casi en el límite del valor requerido para separar los picos isotópicos de este ión doblemente cargado. Esta misma resolución fue insuficiente para resolver adecuadamente (a 50% de la altura) los picos isótopicos del ión triplemente cargado m/z 898 (panel B).

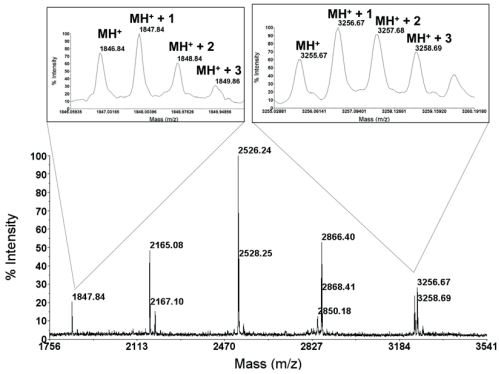


Figura 2. Espectro de masas de una mezcla de péptidos obtenido con un espectrómetro de masas MALDI-TOF Voyager DE-Pro™ (Applied Biosystems) en modo reflectrón. Se muestran las distribuciones isotópicas de los iones monocargados de m/z 1846,48 y 3255,67. MH^+, MH^++1, MH^++2, y MH^++3 corresponden, respectivamente, a los picos monoisotópicos, del primer isótopo (también denominado del ^{13}C), del segundo isótopo, y del tercer isótopo.

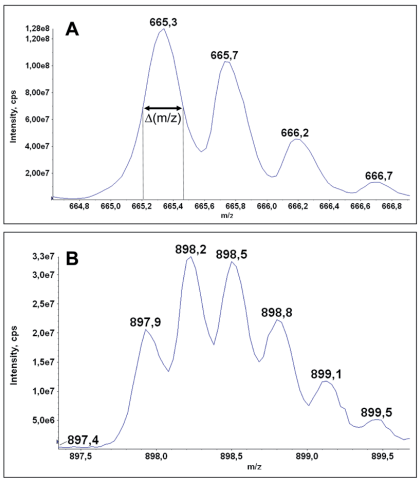


Figura 3. Espectro de resolución de los iones m/z 665,3 y 897,9 obtenidos mediante ionización por electrospray en un espectrómetro QTrap2000 (Applied Biosystems). La diferencia de 0,5 Th (A) y 0,3 Th (B) entre los picos de la distribución isotópica indica que los iones corresponden a $(M+2H)^{2+}$ (dicargado) y $(M+3H)^{3+}$ (tricargado), respectivamente, y sus masas moleculares son pues: (665,3 x 2) - (2 x 1,007825) = 1328,58 Da y (887,9 x 3) - (3 x 1.007825) = 2260,67 Da.

Para lograr resolución isotópica los centroides de los iones de la distribución deben estar separados, a mitad altura, por un $\Delta(m/z) \geq 1/2z$. Los analizadores de masas más comúnmente utilizados en Proteómica (de tiempo de vuelo -TOF-, de transformada de Fourier -FT-, triple cuadrupolos, trampas iónicas, Orbitraps, etc.) alcanzan resoluciones > 10.000, pudiendo llegar en el caso del Orbitrap a superar 100.000 y alcanzar >2.000.000 en el de los analizadores de transformada de Fourier (FTMS) [25]. No hay que confundir resolución, que es una característica propia de cada tipo de espectrómetro de masas, con precisión o con exactitud. La precisión se refiere a lo cerca que están unas de otras un conjunto de mediciones diferentes realizadas en las mismas condiciones. Esta cualidad debe evaluarse a corto plazo y tampoco debe confundirse con reproducibilidad (capacidad de dar el mismo resultado en mediciones diferentes realizadas en las mismas condiciones experimentales a lo largo de periodos dilatados de tiempo). La precisión se expresa como una indicación del error máximo estimado para la magnitud. Es, pues, es una medida del grado de dispersión de los datos con respecto al valor promedio. Así, la media de las medidas de m/z {479,220; 479,229; 479,218; 479,227} de un ión de fórmula $C_{18}H_{33}N_5O_{10}$ (masa exacta: 479,2230 Da) es un valor muy exacto (479,2235) pero poco preciso (± 0,06); en cambio el conjunto de medidas {482,220; 482,221; 482,220; 482,219} corresponde a un valor medio muy preciso (± 0,01) pero poco exacto. Quizás los términos más utilizados para cuantificar el error experimental son la *desviación estándar* o *desviación típica* (σ) [26], incorporado a la estadística por Karl Pearson en 1894 y definido actualmente como la raíz cuadrada de la varianza (la media de las diferencias con la media elevadas al cuadrado [27]), y la *desviación media* (media de las diferencias en valor absoluto de los valores a la media). Así, la desviación estándar de los valores de m/z 1235,456, 1235,458, 1235,452 y 1235,454 obtenidos de mediciones repetidas para el mismo ión es ± 0.0022 Th, y la desviación media es ± 0,002 Th, sobre un valor promedio de 1235,455 Th.

La exactitud es una estimación de la cercanía entre el valor real y el valor medido. En espectrometría de masas representa la proximidad del valor de m/z registrado experimentalmente (M_{medida}) a su valor calculado ($M_{calculada}$). La exactitud de una medición de masa se expresa mediante el error relativo, frecuentemente en partes por millón (ppm), que representa la diferencia entre el valor experimental y el valor verdadero en relación al valor experimental:

$$ppm = [(M_{calculada} - M_{medida})/M_{calculada}] \times 10^6.$$

Así, la masa exacta calculada para el ión monoisotópico a m/z 1846,84 de la Figura 1 es 1846,836. La exactitud de la medida fue, por tanto, (1846,84 - 1846,838)/1846,836 = 2,16 x 10^{-6}, o 2 ppm.

El poder resolutivo es una propiedad intrínseca de cada analizador de masas; la exactitud de una medida de masa depende, además del sistema de medida, de parámetros externos como la calibración del instrumento y de la intensidad (concentración) de los iones. Las medidas realizadas en un instrumento de alta resolución pero mal calibrado serán poco exactas pero muy precisas, y en general reproducibles al tratarse de un error sistemático, un sesgo que hace que todas las medidas estén desviadas en una cierta cantidad. Por otra parte, en el rango de masas altas (ej. proteínas medidas en un analizador de tiempo de vuelo como el ejemplo mostrado en las Figs.4 y 5), cuando la resolución es insuficiente para separar los iones de la distribución isotópica (como es el caso de los espectros de ionización por electrospray de proteínas mostrados en las Fig.6), la geometría de los picos espectrales de los iones depende en gran medida del número de espectros acumulados. Además, la geometría de un pico espectral está en función de la resolución del espectrómetro de masas, pudiendo modelarse como una mezcla entre doble exponencial y gaussiana a resoluciones > 4000 y como una sigmoide (o gaussiana) doble asimétrica a resoluciones < 1000. Dado que se considera como medida promedio de la masa el valor de m/z en el centroide del pico, cualquier fluctuación que distorsione la geometría (Fig.5B) de un pico iónico afectará pues al cálculo ulterior de su masa.

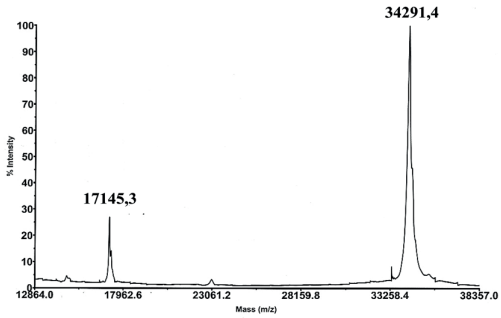


Figura 4. Panel superior, espectro de masas de una proteína obtenido con un espectrómetro de masas MALDI-TOF Voyager DE-Pro™ (Applied Biosystems) en modo lineal. El ión a m/z 34.291,4 puede teoricamente corresponder la especie quasimolecular $M+H^+$ o a un dímero $(2M+H)^+$ del ión a m/z 17.145,3.

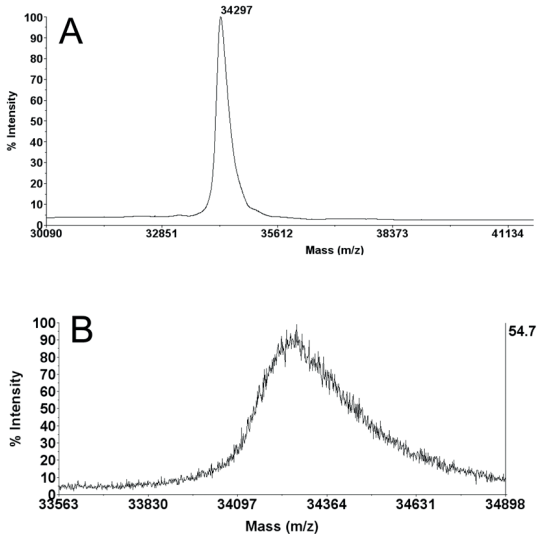


Figura 5. Panel A, espectro MALDI-TOF de una proteína de masa molecular 34,3 kDa obtenido en modo linear. El panel B muestra un detalle del ápice del ión a m/z 34.297. La ampliación muestra el perfil en sierra de la envolvente del pico. El valor 34.297 es el calculado en el centroide del pico, pero la masa promedio exacta podría corresponder a cualquier valor incluído en 34.297 ± 35 Da.

5.1. Cálculo de la carga (z) y de la masa molecular monoisotópica.

La distribución isotópica contiene las pistas necesarias para determinar la carga y la masa del correspondiente ión. Basta con aplicar consecuentemente la siguiente regla general para cualquier espectro de masas:

Regla 1: Los picos de la distribución isotópica de un ión monocargado son siempre consecutivamente mayores en ~una unidad de masa (un neutrón), comenzando por el ión monoisotópico.

Dado que la masa molecular de un ión quasimolecular positivo se calcula mediante la fórmula M = [(m/z) x z] - zH^+, donde zH+ es el numero de protones que el ión adquirió durante el proceso de ionización ($M \rightarrow M+zH^+$), la Regla 1 implica que el Δ(m/z) entre picos isotópicos vecinos es igual a 1/z. En otras palabras: z = 1/Δ(m/z). En los ejemplos mostrados en la Fig.1 los iones a m/z 1847 y 3256 son ambos monocargados y sus masas quasimoleculares (MH^+) monoisotópicas son 1845,84 y 3254,67 Th. Para iones con z = 1, 1 Th equivale a 1 Da, y por tanto las masas moleculares monoisotópicas de los iones de la Fig.1 son 1845, 83 y 3256,66 Da.

5.2. Determinación de z en espectros de masas moleculares promediadas.

Cuando la resolución del analizador de masas no alcanza a resolver la distribución isotópica del ión la aproximación basada en la Regla 1 no es aplicable. En estas circunstancias el espectro de masas sólo muestra la envoltura de la distribución isotópica cuyo ápice corresponde al valor promediado por las abundancias relativas de los picos isotópicos (Fig.4). Esta masa se denomina masa molecular promedio (*isotope-averaged mass*, M_{av}). La M_{av} de un compuesto se calcula a partir de su fórmula química utilizando los valores de masas atómicas de la Tabla Peródica de los elementos. Retomando el ejemplo del glucagón ($C_{153}H_{224}N_{42}O_{50}S$), de masa molecular monoisotópica (M_{mono}) = (153 x 12,000) + (224 x 1,007825) + (42 x 14,00307) + (50 x 15,99491) + 31,97207 = 3481,59931 Da, tiene una masa molecular promedio igual a (153 x 12,0111) + (224 x 1,00797) + (42 x 14,0067) + (50 x 15,9994) + 32,064 = 3483,7989 Da. La diferencia entre M_{mono} y M_{av} de moléculas como las biológicas formadas por los elementos C, H, N, O, S y P aumenta linearmente en función de su fórmula empírica de acuerdo a la ecuación:

$$M_{av} = M_{mono} + (0{,}0107C + 0{,}000075H + 0{,}00335N + 0{,}005030\,O + 0{,}08793S)$$

Los espectros obtenidos mediante MALDI (*matrix-assisted laser-desorption ionization*) corresponden generalmente a iones monocargados (MH^+), pero la formación de iones doblemente cargados $(M+2H)^{2+}$ y de aductos multiméricos del ión monocargado ($MH^+ + M \rightarrow (2M+H)^+$) no es infrecuente, particularmente en el caso de macromoléculas biológicas. El espectro de la Fig.4 probablemente corresponda a una proteína de masa molecular promedio 34.290 Da, aunque la posibilidad de que la especie m/z 17.145,3 sea el ión quasimolecular MH^+ y el ión a m/z 34.291,4 corresponda a un dímero no puede formalmente descartarse. Para distinguir entre ambas posibilidades se necesita información adicional (ej. SDS-PAGE de la proteína) o aplicar un método para determinar la carga de los iones. Afortunadamente, para ello sólo tenemos que aplicar otra regla de cumplimiento obligatorio en espectros de moléculas multicargadas:

Regla 2: Los iones multicargados de una serie poseen cargas enteras consecutivas.

La Regla 2 implica que en una serie de iones multicargados no puede faltar ningún valor intermedio de z. Antes de volver al ejemplo de la Fig. 4, apliquemos esta regla al espectro de la Fig. 6A. Para ello asignamos carga "n" a cualquier ión de la serie, quedando entonces definada la carga de los iones consecutivos de menor masa como n+1, n+2, etc. y los de mayor m/z como n-1, n-2, etc. Si todos los iones del espectro pertenecen a una serie de iones multicargados de la misma molécula (en este caso una proteína) podemos plantear una ecuación para cada valor de m/z:

M = 1304,4 n -n
M = 1239,2 (n+1) - (n+1)
M = 1376,8 (n-1) - (n-1)
etc.

Igualando cualquier par de ecuaciones y despejando "n" obtenemos:

1304,4n - n = 1239,2n +1239,2 -n -1
(1304,4 - 1239,2)n = 1238,2
n = 1238,2/65,2 = 18,99 (= 19, pues la carga tiene que ser un número entero).

De forma general, $n = (M^{n+1}-1)/(M^{n}-M^{n+1})$, donde M^{n+1} y M^{n} corresponden a los valores de m/z de los iones a los que se asigna carga "n+1" y "n", respectivamente. Habiendo determinado la carga de cada ión de la serie, el cálculo de la masa molecular promedio de la proteína es inmediato:

M = (1304,4 x 19) - (19 x 1,007825) = (1239,2 x 20) - (20 x 1,007825) = (1376,8 x18) - (18 x 1,007825), etc., obteniendose un valor de 24764,5 ± 1 Da. Existen, no obstante, programas sencillos que convolucionan los datos del espectro ESI en un espectro que muestra en un único pico la distribución de masa molecular promedio modelada matematicamente para el ión monocargado.

La Regla 2 puede aplicarse a cualquier espectro con tres o más iones de una misma serie para determinar inequivocamente sus cargas. Por ejemplo, en el caso de MALDI, la presencia de un tercer pico en el espectro de la Fig. 6B garantiza la deducción de la masa molecular de la correspondiente proteína sin ambigüedades. En efecto, la única solución posible es que el ión m/z 9357,4 corresponda a la especie quasimolecular $M+H^+$ y, por tanto, m/z 4678,6 sea $(M + 2H^+)^{2+}$. Nótese que el ión m/z 18.714,1 no pertenece a la misma serie iónica que m/z 4678,6 y 9357,4, ya que ello implicaría una carga 0 para esta especie y recordemos que las partículas no cargadas son invisibles (materia oscura) para un espectrómetro de masas. m/z 18.714,1 sólo puede asignarse a un dímero monocargado $2M+H^+$. Cualquier otra solución es incongruente con la regla 2: si m/z 4678,6 fuera $M+H^+$, entonces m/z 9357,4 debería asignarse al dímero $2M+H^+$, lo cual no sería descabellado si no fuera porque en esta serie m/z 18.714,1 debiera corresponder al trímero $3M+H^+$, y claramente ese no es el caso. Igualmente, si m/z 18.714,1 fuera el ión quasimolecular, $M+H^+$, los iones m/z 9357,4 y m/z 4678,6 deberían asignarse a $(M+2H^+)^{2+}$ y $(M+3H^+)^{3+}$, siendo esta posibilidad nuevamente inconsistente con la Regla 2.

5.3. Polímeros *versus* series de iones multicargados.

En una serie de iones multicargados el Δ(m/z) entre iones consecutivos varía gradualmente, decreciendo a medida que aumenta la carga. Esta característica puede utilizarse para distinguir el espectro de masas de una serie de iones multicargados y del de iones que no están relacionados entre sí por su carga. Así, los iones del espectro de la figura 7 están espaciados entre sí por un Δ(m/z) constante de aproximadamente 44 y están relacionados entre sí mediante la fórmula $m/z = [22.5 + 44(x+1)] + H^+$, siendo x = 11 para el ión de menor m/z. Este patrón de variación de m/z claramente indica que no puede tratarse de una serie de iones multicargados sino de una mezcla de polímeros monocargados compuestos por un número creciente de unidades de 44 Da (¿quizás bloques de CH2-CHOH, una estructura común a muchos detergentes utilizados para solubilizar proteínas hidrofóbicas?).

5.4. Espectros con más de una serie de iones. Regla de conservación de la topología.

Un espectro de masas puede contener más de una serie de iones, por ejemplo porque se trate de una mezcla de proteínas, como en el ejemplo de la Fig. 8A. La Regla 2 nos permite identificar sin ambigüedad cada una de las series coexistentes en un mismo espectro de masas con tal

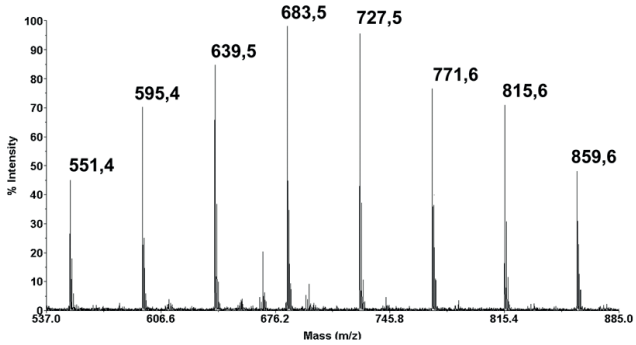


Figura 7. Espectro de masas de un polímero de fórmula m/z = [22.5 + 44(x+1)] + H^+ resuelto en un espectrómetro MALDI-TOF Voyager DE-Pro™ (Applied Biosystems) en modo lineal.

de plantear las ecuaciones adecuadas. El caso más simple es cuando las series iónicas están separadas (Fig.8A). En este supuesto el inicio de la segunda serie comienza con el primer ión para el que la masa calculada asumiendo que pertenece a la primera serie se desvía significativamente del valor promedio de la serie. Así, en el ejemplo mostrado en el espectro de la Fig.8A, es fácil demostrar que el ión m/z 1102,8 no puede pertenecer a la misma serie que los iones de mayor m/z, pues para que así fuera debería tener carga 27 y su masa molecular calculada (29.748,6 Da) es significativamente diferente del valor 29.337,5 ± 4 Da calculado para la serie que incluye a los iones de m/z 1129,1 a 1629,7. Además, si asignamos carga "n" a m/z 1129,1, la solución de la igualdad M= 1129,1n -n = 1102,8(n+1)-(n+1) sería n= 41.9, que es un resultado absurdo. Por el contrario, los iones m/z 1102,8 (12+)-778,9 (17+) comprenden una serie que definen a una proteína de 13.222,3 ± 2 Da. Nuevamente, tanto la asignación de z = 11 al ión m/z 1129,1 (M_{calc} = 12.409,1 Da), como el valor de n= 42,8 calculado a partir de la ecuación M= 1102,8n-n = 1129,1(n-1)-(n-1), son incongruentes con la hipótesis de que m/z 1102,8 pertenezca a la misma serie que los iones con carga +17 a +12.

Regla 3: La topología de los grupos de iones formados por la superposición de series de iones multicargados se repite a intervalos regulares de m/z.

Varias series de iones multicargados pueden también estar superpuestas en un mismo espectro. Cuando esto ocurre, y la masa molecular de las especies ionizadas es similar, la topología de cada grupo de iones se repite en cada rango de m/z. Esta regla de "conservación de la topología" se ilustra en el espectro del panel B de la Fig.8, que corresponde a una mezcla de isoformas de una proteína de ~7.3 kDa, y simplifica enormemente la identificación de los iones pertenecientes a cada serie. La Regla 3 se cumple aunque los iones de las diferentes series, agrupados por la superposición, difieran en su estado de carga (Fig.8C). Sin embargo, cuando las masas moleculares de las proteínas de la mezcla son muy diferentes la distribución de iones en el espectro no sigue la regla de conservación de la topología. El espectro de la figura 8D ilustra este punto. Ello es debido a que, de acuerdo a la Regla 2, en el rango de m/z 1000-1600, la serie teoricamente observable de la proteína de 13.783 Da abarca 5 iones, con z= M/(m/z), mientras que en el mismo intervalo de medida las proteínas de masas moleculares 9001 Da y 4793 Da sólo pueden producir un máximo de 4 y 2 iones, respectivamente. No obstante, el número de iones producidos en la ionización por electrospray depende del número de protones captados (en modo positivo) o cedidos (en modo negativo) por el analito. En el caso de péptidos y proteínas, el extremo N-terminal y los residuos de histidina, arginina y lisina son los sitios que pueden protonarse a pH ácido.

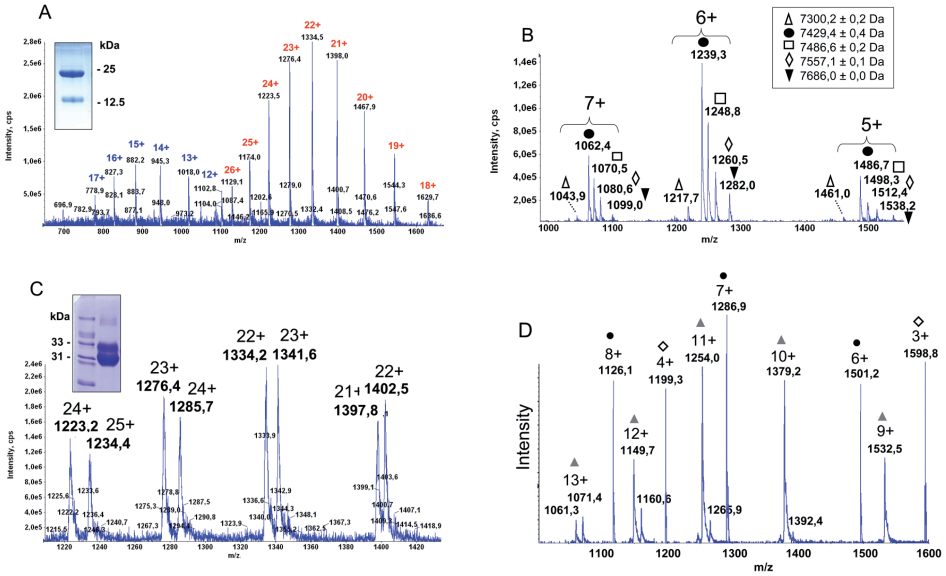


Figura 8. (A) Espectro de masas de una mezcla de dos proteínas. Las series de iones cargados con z = 26+ a 18+ y con z= 17+ a 12+ corresponden a toxinas aisladas de venenos de serpientes, CRISP (cysteine-rich secretory protein) cuya masa molecular promedio es 29.337,5 ± 4 Da y una fosfolipasa A_2 de masa molecular 13.222,3 ± 2 Da. (B) Espectro de masas de una mezcla de isoformas de la disintegrina barbourina aislada del veneno de Sistrurus miliarius barbouri. La similitud topológica de los grupos de iones de cada rango de m/z facilita la asignación de las siguientes series de iones, que en el espectro están identificados con el mismo símbolo. (C) Detalle de un espectro de masas de una mezcla de las dos serinoproteasas de alrededor de 30 kDa analizadas en el gel de la izquierda. Las masas moleculares promedio calculadas son 29.332,8 ± 1,1 Da y 30.833,6 ± 0,8 Da. (D) Espectro de masas de una mezcla de toxinas de masas moleculares 13.783,2 ± 1 Da (PLA_2, serie 9+ a 13+), 9001,3 ± 0,1 Da (disintegrina bitistatina, serie 6+ a 8+) y 4793,3 ± 0,1 Da (miotoxina crotamina, iones de z = 3+ y 4+). Los espectros de masas de los paneles A-D corresponden a proteínas ionizadas por electrospray y fueron obtenidos en un espectrómetro QTrap2000™ (Applied Biosystems).

6. Algunos enlaces útiles.

Muchos son los portales de Internet dedicados a aspectos relacionados con la respectrometría de masas y sus aplicaciones. Sería interminable citarlos todos, por lo que además de las páginas Web citadas anteriormente en el texto, he seleccionado los portales del Swiss Institute of Bioinformatics, (http://us.expasy.org/tools), del Institute for Systems Biology (http://db.systemsbiology.net:8080/proteomicsToolkit/index.html) y de Thermo-Fisher Scientific (http://www.ionsource.com) por contener informaciones complementarias muy útiles sobre conceptos generales de espectrometría de masas, tablas de unidades y de masas atómicas y moleculares, calculadoras diversas y programas de conversión de datos, manuales para iniciarse en la interpretación de espectros, enlaces a otros portales de interés, etc. Espero que el lector interesado en la espectrometría de masas, especialmente en su vertiente biológica, disfrute navegando por estos hiperespacios.

7. Corolario.

Arthur C. Clarke, el genial científico británico (1917-2008) que en 1945 sentó las bases de los satélites artificiales en órbita geoestacionaria (llamada en su honor órbita Clarke), es quizás más conocido por sus famosas Leyes de Clarke. Publicadas en su libro de divulgación científica *Perfiles del Futuro* (1962), la más popular (y citada) de ellas es la llamada «Tercera Ley de

Clarke»: *Toda tecnología lo suficientemente avanzada es indistinguible de la magia*. Y ¿acaso no es magia hacer volar elefantes moleculares? La diferencia entre ilusionismo y espectrometría de masas es que los magos crean la ilusión de que hacen volar elefantes desafiando las leyes de la física y la lógica, en tanto que el espectrometrista aplica la lógica de los principios físicos para hacer volar elefantes y deducir sus propiedades específicas, vetadas a nuestros sentidos, unicamente observando su vuelo. La magia de la espectrometría de masas reside en que ello es posible con solo aplicar un par de reglas generales y matemáticas elementales...

REFERENCIAS.

1. **Thomson, J.J. (1906) Carriers of negative electricity.** Nobel Lecture. http://nobelprize.org/nobel_prizes/physics/laureates/1906/thomson-lecture.html

2. http://nobelprize.org/nobel_prizes/chemistry/laureates/1922/aston-lecture.pdf

3. http://masspec.scripps.edu/mshistory/mshistory.php

4. **Cooks, R. G., Rockwood, A.L. (1991) The 'Thomson'.** A suggested unit for mass spectroscopists. Rapid Communications in Mass Spectrometry 5, 93.

5. **Weast, R.C.** (1974) Handbook of Chemistry and Physics, 60th Ed., CRC Press, Boca Raton, FL.

6. **Fenn, J.B.** (2003) Electrospray Wings for Molecular Elephants (Nobel Lecture). Angew. Chem. Int. Ed. 42, 3871-3894.

7. **Tanaka, K.** (2003) The origin of macromolecule ionization by laser irradiation (Nobel Lecture). Angew. Chem. Int. Ed. 42, 3861-3870.

8. **Hillenkamp, F., Peter-Katalinic, J. (2007) MALDI MS: a practical guide to instrumentation, methods and applications.** Wiley -VCH, Weinheim, ISBN 978-3-527-31440-9.

9. **Watson, J.T., Sparkman, O.D. (2007) Introduction to Mass Spectrometry.** Instrumentation, Applications and Strategies for Data Interpretation (4th Edition), John Wiley & Sons Ltd., Chichester, ISBN 978-0-470-51634-8.

10. **Burbidge, E.M., Burbidge, G.R., Fowler, W.A. & Hoyle, F. (1957) Synthesis of the Elements in Stars.** Rev. Mod. Phys. 29, 547-650.

11. **Weinberg, S. Los tres primeros minutos del universo.** Alianza Editorial, 1ª Ed. 1978. ISBN 84-206-2216-8.

12. **Bethe, H.A. (1938) Energy production in stars.** Phys. Rev. 55, 434 - 456.

13. **Coplen, T.B. (2001) Atomic Weights of the Elements 1999, Pure Appl. Chem. 73, 667-683.** (http://physics.nist.gov/PhysRefData/Compositions/index.html).

14. **Gross, J.H. (2004) Mass Spectrometry.** Springer-Verlag, Berlin, ISBN 3-540-40739-1

15. **Kubinyi, H. (1991) Calculation of isotope distributions in mass spectrometry.** A trivial solution for a non-trivial problem. Anal. Chim. Acta 247, 107-119.

16. **Pappin, D.J., Højrup, P., Bleasby, A.J. (1993) Rapid identification of proteins by peptide-mass fingerprinting.** Curr. Biol. 3, 327-332.

17. **Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C., Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases.** Proc. Natl. Acad. Sci. U.S.A. 90, 5011-5015.

18. **Mann, M., Højrup, P., Roepstorff, P. (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases.** Biol. Mass Spectrom. 22, 338-345.

19. **James, P., Quadroni, M., Carafoli, E., Gonnet, G. (1993) Protein identification by mass profile fingerprinting.** Biochem. Biophys. Res. Commun. 195, 58-64.

20. **Yates, J.R., Speicher, S., Griffin, P.R., Hunkapiller, T. (1993) Peptide mass maps: a highly informative approach to protein identification.** Anal. Biochem. 214, 397-408.

21. **Ong, S.E., Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative.** Nature Chem. Biol. 1, 252–262.

22. **Pan, S., Aebersold, R., Chen, R., Rush, J., Goodlett, D.R., McIntosh, M.W., Zhang, J., Brentnall, T.A. (2009) Mass spectrometry based targeted protein quantification: methods and applications**. J. Proteome Res. 8, 787–797.

23. **Brun, V., Masselson, C., Garin, J., Dupuis, A. (2009) Isotope dilution strategies for absolute quantitative proteomics.** J. Proteomics 72, 740-749.

24. **Bettmer, J., Montes-Bayón, M., Encinar, J.R., Fernández-Sánchez, M.L., Fernández de la Campa, M. R., Sanz-Medel, A. (2009) The emerging role of ICP-MS in proteomic analysis.** J. Proteomics 72, 989-1005.

25. **Gelpí, E. (2009) From large analogical instruments to small digital black boxes: 40 years of progress in mass spectrometry and its role in proteomics.** Part II 1985–2000. J. Mass Spectrom. 44, 1137-1161.

26. http://www.disfrutalasmatematicas.com/datos/desviacion-estandar-calculadora.html

27. http://www.disfrutalasmatematicas.com/datos/desviacion-estandar.html

Analizadores de masa en proteómica.
Principios operacionales y parámetros de rendimiento. Triple cuadrupolo. QTrap

Marina Gay, Montserrat Carrascal, Joaquin Abian
Laboratorio de Proteómica CSIC/UAB, ProteoRed, Instituto de Investigaciones Biomédicas de Barcelona. Consejo Superior de Investigaciones Científicas. Barcelona. Spain.

1. Introducción.

La aplicación de la espectrometría de masas a la secuenciación de péptidos fue descrita por primera vez por Biemann en 1959 [1], pero no fue hasta el 1965 que el grupo de Lederer determinó la estructura de un lipopéptido natural, fortuitine, mediante espectrometría de masas [2]. Estos mismos autores en 1967 aplicaron diversos métodos de acetilación y permetilación para aumentar la volatilidad y estabilidad térmica de los péptidos de forma que pudieran ser analizados por espectrometría de masas. Estos primeros estudios se llevaron a cabo en instrumentos de sectores (MS9) mediante introducción directa en una fuente de impacto electrónico utilizando una sonda de sólidos [3, 4]. Las cantidades de péptido utilizadas en estos análisis fueron de 0.1 mg.

En esa época los instrumentos de sectores constituían los instrumentos de referencia en espectrometría de masas y proveían la máxima resolución y sensibilidad entonces disponible. Sin embargo, el uso de estos aparatos era muy complejo y la dificultad para mantener el campo magnético del espectrómetro estable y uniforme requería un continuo ajuste de las condiciones de trabajo para obtener una buena resolución y precisión en la medida de masa. A principios de los años 50, W. Paul desarrolló el primer analizador de cuadrupolo que en lugar de usar campos magnéticos usaba campos eléctricos cuadrupolares para definir la trayectoria de los iones [5]. Posteriormente, a finales de los setenta Yost y Enkel desarrollaron el primer espectrómetro de masas de triple cuadrupolo [6]. A partir de este momento los espectrómetros de masas con analizadores de cuadrupolo fueron desplazando los espectrómetros de sectores, debido, en gran parte, a su menor tamaño, menor coste y mayor resolución en el análisis MS/MS.

En 1981, Hunt y col. reportaron la secuenciación de diversos neuropéptidos con longitudes de secuencia entre 7-9 aminoácidos derivatizados a niveles de picomoles utilizando un espectrómetro de masas de triple cuadrupolo con una fuente de ionización química [7, 8]. Ese mismo año, Barber y Bordoli acoplaron el analizador de cuadrupolo a una fuente de ionización suave, el FAB [9], que se convirtió en la técnica más utilizada durante los años 80 para el análisis de péptidos [10]. Mediante FAB-MS/MS era posible obtener, en un único experimento, información de secuencia con una cobertura del 25-60% a partir de una proteína de 50 kDa en cantidades de 10 nmoles. El peso molecular máximo de los péptidos analizados se encontraba entre 0.5

kDa (IE) y 1-5 kDa (FAB) debido tanto a las dificultades para ionizar moléculas de gran tamaño y termolábiles como para obtener información de secuencia útil mediante espectrometría de masas en tándem.

Los sistemas de triple cuadrupolo pueden considerarse, por tanto, los primeros analizadores usados de forma general y eficiente para la identificación y secuenciación de péptidos. A partir de la segunda mitad de los años 90, la aparición de las trampas iónicas más sensibles y rápidas y de mayor resolución desplazó a los cuadrupolos en los trabajos de proteómica. Sin embargo, actualmente hay un reflorecimiento en el uso de los triples cuadrupolos para el análisis cuantitativo de péptidos y proteínas. Estos análisis aprovechan los modos específicos de barrido de los triples cuadrupolos para lograr una mayor sensibilidad y selectividad.

El analizador de masas de cuadrupolo es un instrumento que juega con la diferente estabilidad de las trayectorias de los iones cuando estos atraviesan campos eléctricos oscilantes para separarlos en función de su relación masa/carga. Otros analizadores como la trampa iónica o el ICR también están basados en este principio. El funcionamiento de este analizador fue descrito por Paul y Steinweden en 1953 [5]. Fruto de estos trabajos, Paul recibió el premio Nobel de física en 1989. Los cuadrupolos son un caso particular de sistemas multipolares que, como los hexapolos u octapolos, están presentes en casi todos los tipos de espectrómetros de masa modernos como filtros y guías de banda ancha de iones. Los espectrómetros de masas con analizador de cuadrupolo han sido durante muchos años instrumentos de referencia en los análisis realizados por industrias, principalmente farmacéuticas y petroleras, tanto acoplados a sistemas de cromatografía de gases, como de cromatografía líquida.

2. Principios de funcionamiento y características.

Los cuadrupolos están formados por cuatro barras o polos alineadas paralelamente y equidistantes a una distancia r_0 de un eje central situado sobre el eje z [11] y forman un túnel por el que pasan los iones (Figura 1). Las barras son de sección cilíndrica o hiperbólica en la cara interna, generalmente de unos 15-20 cm de largo y 0.5 cm de radio, separadas entre sí unos 2 cm.

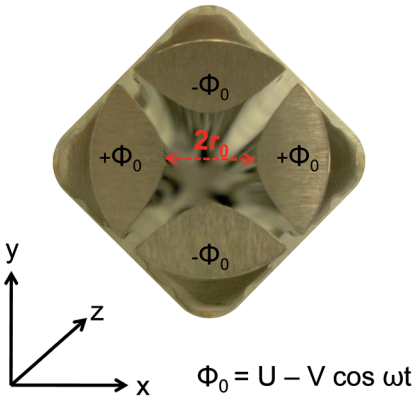


Figura 1. Cuadrupolo de barras hiperbólicas, los potenciales aplicados a cada una de las barras y el diámetro que las separa. El analizador mostrado corresponde al de un sistema TSQ700.

A cada barra se aplica un potencial eléctrico que es la combinación de un voltaje de corriente continua (DC) y otro de corriente alterna en el rango de las radiofrecuencias (RF), ambos de amplitud variable, y que crean en su interior un campo eléctrico denominado cuadrupolar. Los iones generados en la fuente deben atravesar longitudinalmente el recinto limitado por estas barras para incidir en el detector. Estos iones, que entran en el analizador a una energía de unos pocos electronvoltios, son sometidos al efecto del campo cuadrupolar que los hace oscilar y los desvía en función de su valor m/z (masa por unidad de carga) de forma que para una combinación de potenciales DC y RF sólo los iones en un estrecho rango de valores m/z llegan a incidir en el detector. El movimiento de los iones en el campo cuadrupolar viene definido por las ecuaciones de movimiento que se describen en el subapartado 2.1. Los analizadores de cuadrupolo actúan por tanto como filtros de iones y los

espectros de masas en estos sistemas se obtienen mediante un barrido del potencial aplicado a las barras. En consecuencia, en cada instante sólo una pequeña fracción del total de iones es monitorizada mientras que el resto se desecha. Este modo de funcionamiento repercute negativamente en el límite de detección de estos instrumentos, especialmente cuando se requiere la obtención de espectros completos en rangos de masa amplios.

El campo eléctrico del analizador de cuadrupolo en el plano xy, Φ *(x,y),* viene dado por:

$$\Phi(x, y) = \left(\frac{x^2 - y^2}{r_0^2}\right)\Phi_0 \qquad \{1\}$$

donde *x* e *y* son las coordenadas cartesianas, r_0 es la distancia del centro del cuadrupolo a uno de los electrodos y Φ_0 es el potencial aplicado a los electrodos. El potencial Φ_0, que se aplica con valores de signo contrario a los pares de electrodos en el eje *x* e *y* ($+\Phi_0$, $-\Phi_0$ en Figura 1), es la combinación de un voltaje de corriente continua *U* y de un potencial de radiofrecuencia de amplitud *V* y frecuencia ω:

$$\Phi_0 = U - V \cos \omega\tau \qquad \{2\}$$

Los iones que viajan a lo largo del eje z están sujetos a la influencia de este campo cuadrupolar. La relación entre los potenciales DC y RF determina la capacidad del espectrómetro de separar iones de diferente relación m/z [11, 12].

2.1. Ecuaciones de movimiento

Las ecuaciones que definen el movimiento de las partículas cargadas en el espacio cuadrupolar son conocidas como las ecuaciones de Mathieu [11, 12]. Estas ecuaciones se deducen a partir de las ecuaciones del campo eléctrico cuadrupolar {1}, del potencial aplicado a las barras del cuadrupolo {2} y de las ecuaciones de fuerza de los iones (segunda ley de Newton y teorema de Floquet) {3}.

$$F_x = m\frac{d^2x}{dt^2} = -ze\frac{\delta\Phi}{\delta x}$$

$$F_y = m\frac{d^2y}{dt^2} = -ze\frac{\delta\Phi}{\delta y} \qquad \{3\}$$

Reordenando los términos de las ecuaciones {1}, {2} y {3} se obtienen las ecuaciones de Mathieu {4}.

$$\frac{d^2u}{dx^2} + (a_u - 2q_u \cos 2\xi)\, u = 0 \qquad \{4\}$$

donde $\xi = \frac{\omega\tau}{2}y$ y

$$a_u = a_x = -a_y = \frac{8zeU}{mr_0^2\omega^2}$$

$$q_u = q_x = -q_y = \frac{4zeV}{mr_0^2\omega^2} \qquad \{5\}$$

ξ es un parámetro adimensional, *a* y *q* son los denominados parámetros de atrapamiento cuyo valor depende de la masa del ión, de la frecuencia, del radio del cuadrupolo y del voltaje aplicado, el DC en el caso de a y el RF en el caso de *q*.

Es importante tener en cuenta que estas ecuaciones establecen una relación entre la posición de un ión, su masa y el tiempo. Las soluciones a las ecuaciones de Mathieu se clasifican como estables o inestables. Si *x* e *y*, que determinan la posición de un ión respecto al centro de las barras del cuadrupolo, se mantienen inferiores a r_0, el ión pasará a través del cuadrupolo sin tocar las barras y su trayectoria será estable. Por el contrario si *x* e *y* son superiores a r_0 el ión chocará contra la barra y no llegará al detector, su trayectoria será inestable (Figura 2).

Las áreas de estabilidad de los iones se pueden representar en un diagrama de a_u frente a q_u (Figura 3). Las regiones donde las trayectorias de los iones en ambas direcciones (*x*, *y)* son estables se denominan regiones estables [13]. En el gráfico se muestran cuatro regiones estables, pero en realidad existe un número infinito de regiones estables. Sin embargo, casi todos los cuadrupolos comerciales usan la región estable I.

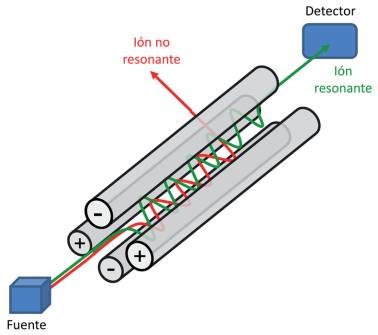


Figura 2. Esquema de un instrumento de cuadrupolo y paso de los iones a través de éste con una trayectoria estable que les permite llegar al detector (ión resonante) o bien con una trayectoria inestable que los hace chocar contra las barras (ión no resonante). Las trayectorias de los iones son estables o inestables en función de la solución de la ecuación de Mathieu.

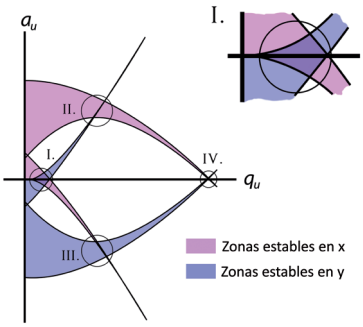


Figura 3. Diagrama de estabilidad de los iones en el campo cuadrupolar. Se muestran las zonas estables en x y las estables en y. Cuando las trayectorias de los iones son estables tanto en el eje x como en el eje y se denominan regiones estables.

A partir de las ecuaciones de a_u y q_u {5}, se puede deducir,

$$U = a_u \frac{m}{z} \frac{\omega^2 r_0^2}{8e}$$

$$V = q_u \frac{m}{z} \frac{\omega^2 r_0^2}{4e} \qquad \{6\}$$

donde es relevante que los potenciales *U* y *V* son función de la relación masa carga, m/z.

En la Figura 4 se muestra la posición de tres iones con una relación m/z $m_1 < m_2 < m_3$ en el diagrama de estabilidad derivado de las ecuaciones {5} y {6} para valores *U* y *V* constantes. En este ejemplo, m_2 está dentro de la región de estabilidad. Como se observa en la ecuación {5} los valores de a_u y q_u son inversamente proporcionales a la masa del ión. Así, para los mismos voltajes de *U* y *V* los iones más pesados m_3 tienen valores menores de *a* y *q;* por el contrario, los iones más ligeros m_1 tienen valores más altos. Los iones con distinta relación masa/carga dibujan en el diagrama de estabilidad una recta, para valores fijos de voltaje, que se denomina

línea de operación o línea de barrido. Los iones m_1 y m_3 están fuera de la región de estabilidad y sus trayectorias serán inestables, de manera que chocarán contra las barras del cuadrupolo sin llegar al detector. En este ejemplo el cuadrupolo actúa como filtro de masas para los iones m_2. La pareja de polos situados en el plano XZ (izquierdo y derecho en Figura 1) actúa como un filtro de masas bajas, ya que, para determinados valores de voltaje y radiofrecuencia, solo consiguen pasar con trayectorias estables las masas mayores de un determinado valor. Por el contrario, los polos situados en el plano YZ actúan como un filtro que elimina las masas altas, rechazando aquellas que sean mayores de un cierto valor [14].

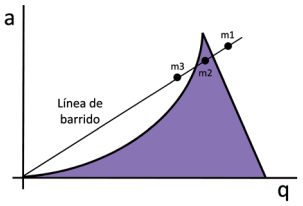


Figura 4. Línea de barrido y posición en la región de estabilidad de tres iones con m/z m1, m2 y m3 para los potenciales constantes de corriente continua, U, y de radiofrecuencia, V. Solo los iones m2 seguirán trayectorias estables y llegarán al detector.

Tal como se muestra en el ejemplo de la Figura 4 comentado arriba, un cuadrupolo es capaz de filtrar un determinado ión manteniendo fijos los potenciales *U* y *V*. Para transmitir iones de otro valor m/z hay que ajustar los valores de *U* y *V* de manera que el área de la región de estabilidad se desplace y la masa de interés se sitúe dentro de ella. Si *U* y *V* se barren de forma conjunta a relación *U/V* constante, los iones con masa creciente se irán situando en la región de estabilidad y alcanzarán secuencialmente el detector (Figura 5). Esto es lo que se denomina hacer un barrido y dará lugar a un espectro de masas.

La resolución de un espectrómetro de masas es un parámetro que define la capacidad de separar y medir masas de iones de similar masa molecular. La resolución de un cuadrupolo puede modificarse cambiando la relación *U/V* (potencial de corriente continua y radiofrecuencia), de manera que se modifica la pendiente de la línea de barrido. Cuanto mayor sea la pendiente mayor será la resolución, pero peor será la sensibilidad (Figura 5). Si un cuadrupolo opera a una relación constante de *U/V,* la resolución será constante a lo largo de todo el barrido de masas. Cuanto más cercano sea el campo electrostático generado por los cuadrupolos a la geometría hiperbólica, mejor serán las características del cuadrupolo. La sensibilidad y la resolución del cuadrupolo aumentan con su radio, con su longitud y con el valor de la frecuencia empleada [15].

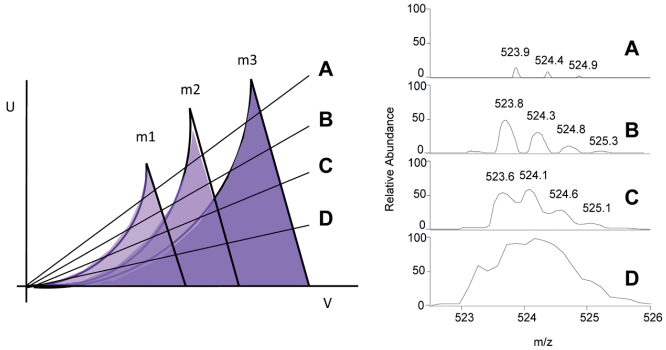


Figura 5. Desplazamiento de la región de estabilidad en función de los valores de U y V. Si la relación U/V es constante los iones con relación m/z se irán situando en la región de estabilidad y llegarán de forma secuencial al detector. La pendiente de la línea de barrido determina la resolución obtenida, a mayor pendiente mayor resolución.

Los cuadrupolos son analizadores de baja resolución. Convencionalmente la resolución de estos analizadores se había indicado en sus especificaciones con el término resolución unidad, indicando su capacidad para separar dos picos que difieren en una unidad de masa en todo el rango de trabajo. Esto corresponde aproximadamente a resolución 1500 FWHM a m/z 1000. Sin embargo, los nuevos analizadores de cuadrupolo pueden trabajar de forma rutinaria y sin perder demasiada sensibilidad a resoluciones superiores a 5000 FWHM a ese valor de masa carga.

Los sistemas electrónicos de los espectrómetros de masas permiten variar los potenciales que se aplican a las barras de forma muy rápida y precisa. Por ejemplo, en el sistema de triple cua-

drupolo TSQ Quantum Discovery el instrumento puede barrer todo el rango de masas (m/z 30 a 1500) en tan solo 0.1s. En los analizadores de cuadrupolo convencionales el valor m/z más alto alcanzable se sitúa entre los 1000 y 4000 Da dependiendo del instrumento. Modificando la frecuencia RF utilizada es posible obtener límites mayores habiéndose descrito sistemas capaces de alcanzar 10000 Da [16].

2.2. El cuadrupolo como guía de iones y como cámara de colisión

Cuando se elimina el voltaje de corriente continua aplicado a las barras de un cuadrupolo ($U = 0$) dejando únicamente el potencial de radiofrecuencia, este sistema actúa como un filtro de paso de banda ancha. En estas condiciones, la resolución del instrumento es 0 y el valor de V determina el valor mínimo de las masas que son estables.

Este modo de trabajo, conocido como *RF-only*, permite trabajar al cuadrupolo como una guía de iones. Las guías de iones permiten transferir los iones de un compartimiento del espectrómetro de masas al otro sin afectar significativamente su energía cinética axial, y por tanto manteniéndolos en un haz compacto en el eje central del cuadrupolo. En instrumentos de baja energía, las guías de iones formadas por cuadrupolos, resultan mucho más eficientes que las formadas por lentes electrostáticas.

Las guías de iones pueden construirse con sistemas multipolares formados por más de cuatro barras como son los hexapolos u octapolos. Estos sistemas multipolares presentan características de transferencia superiores a las del cuadrupolo por lo que en la práctica las guías de iones se construyen con multipolos mientras que los cuadrupolos son utilizados de forma casi exclusiva como analizadores de masa (Tabla 1).

Tabla 1. Principales características y diferencias entre distintos multipolos.

Multipolo	Poder de focalización	Rango de masas para la transmisión simultánea de iones
Cuadrupolo	alto	reducido
Hexapolo	medio	medio
Octapolo	bajo	amplio

Este tipo de guías de iones puede trabajar a presiones relativamente altas por lo que se utilizan como cámaras de colisión en los instrumentos de triple cuadrupolo y en otros espectrómetros en tándem de baja energía como es el caso de la celda de HCD implementada en los sistemas LTQ-Orbitrap. Se utilizan también para conectar regiones del espectrómetro de masas que se encuentran a diferente presión como es la entrada de las fuentes de ionización atmosférica de electrospray o APCI [11]. Así pues, podemos encontrar multipolos en prácticamente todos los espectrómetros de masas usados actualmente en proteómica (LIT, Orbitrap, FT, QqTOF, qTRAP).

3. Triple cuadrupolo. Espectrometría de masas en tándem.

Los espectrómetros de triple cuadrupolo disponen de un sistema de análisis de iones formado por tres cuadrupolos colocados consecutivamente (Figura 6). El primer y tercer cuadrupolo (denominados Q1 y Q3) pueden utilizarse como analizadores o como guía de iones mientras que el segundo cuadrupolo se utiliza como cámara de colisión y opera siempre en modo *RF-only*. Este cuadrupolo intermedio, como ya se indicó, en realidad suele ser un hexapolo u octapolo.

Cuando un cuadrupolo puede operar en modo DC y RF se usa el acrónimo Q. Si se trata de un cuadrupolo que solo puede operar en modo RF se usa el acrónimo q. Por lo tanto la configuración de un instrumento de triple cuadrupolo suele indicarse como QqQ [17].

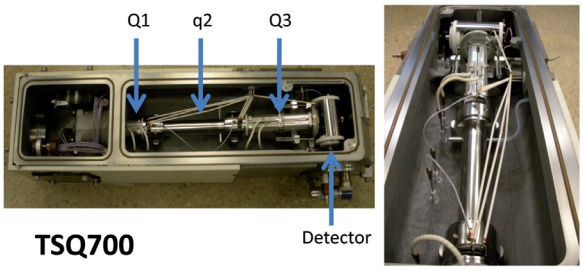


Figura 6. Fotografía de un triple cuadrupolo TSQ700 donde se muestran los tres cuadrupolos y el detector. Este instrumento estuvo operativo en el LP-CSIC/UAB desde 1992 hasta 2006.

Los espectrómetros de triple cuadrupolo permiten llevar a cabo análisis de espectrometría de masas en tándem. En estos procedimientos, los iones seleccionados en Q1 se fragmentan en la cámara de colisión q2 y los fragmentos producidos son analizados en Q3. Este tipo de análisis es normalmente utilizado para caracterizar compuestos individuales en una mezcla compleja o para identificar estructuralmente un compuesto mediante el estudio de los fragmentos producidos.

Para fragmentar los iones seleccionados en Q1, se introduce en la cámara de colisión un gas neutro, por ejemplo argón, helio o nitrógeno. La cámara de colisión se encuentra encerrada dentro de una camisa hermética por lo que la presión del gas no afecta a la transmisión de iones en Q1 y Q3. Los iones seleccionados en Q1 entran en esta cámara colisionando con los átomos o moléculas neutras del gas y debido a esta colisiones, se disocian en fragmentos más pequeños. Algunos de estos fragmentos mantendrán la carga del ión mientras que otros corresponderán a fragmentos sin carga a veces denominados también pérdidas neutras. Este proceso de fragmentación se denomina CID (*collision-induced dissociation*) o CAD (*collisionally activated dissociation*) [18]. El mecanismo de disociación implica la conversión de parte de la energía cinética traslacional del ión en energía interna de manera que el ión pasa a un estado excitado. Si la energía interna es suficiente el ión se fragmenta. Las colisiones en este tipo de instrumentos se llevan a cabo a energías de 20-90 eV por lo que se consideran colisiones de baja energía (Figura 7).

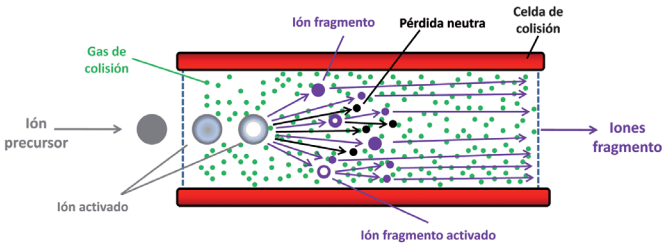


Figura 7. Esquema de fragmentación CID de los iones en la cámara de colisión q2. El ión seleccionado en Q1 colisiona con una molécula neutra y se disocia en fragmentos más pequeños, que pueden ser iones o pérdidas neutras. Los iones producto pueden analizarse en el tercer cuadrupolo Q3.

La eficiencia del proceso de CID viene determinada por la eficiencia de dos fenómenos diferentes:

- Eficiencia de colección de iones: es la relación entre el flujo de iones a la salida de la celda de colisión y a la entrada. Cuando no hay gas de colisión, esta eficiencia es teóricamente del 100%. Es un parámetro dependiente de la masa. Para unos determinados parámetros de colisión, esta eficiencia puede variar de un 50% para iones menos pesados hasta un 75% para los más pesados. Esta diferencia se debe a la mayor difusión de los iones de menor masa, como resultado de las colisiones que tienen lugar en la cámara.

- Eficiencia de fragmentación: es la fracción del flujo de iones a la salida de la cámara de colisión que son iones fragmentos. Esta eficiencia depende directamente de la estabilidad del ión e indirectamente de su masa. Cuanto más estable es un ión, más difícil es producir iones fragmento a una determinada energía y presión de colisión. Cuanto más pesado es un ión, mayor es su habilidad para distribuir la energía vibracional impartida por la colisión, y más difícil es que se produzcan iones fragmento.

De este modo, la eficiencia total es el producto de la eficiencia de colección y de la de fragmentación. Esta eficiencia tiene un máximo con valores de presión de gas del orden de 0.3 - 1 mTorr. A mayor presión se producen más colisiones y aumentan los procesos de difusión de los iones, disminuyendo la eficiencia de colección. A una presión menor de la óptima hay pocas colisiones y la eficiencia de fragmentación disminuye. También existe una energía de colisión óptima para cada ión precursor y cada uno de los iones producto. En el caso de péptidos, esta energía es de alrededor de -30 V a 0.5 mTorr. Para iones de bajo peso molecular, como iones imonio o iones reportero, la energía de colisión es más elevada, del orden de -50 V (Figura 8).

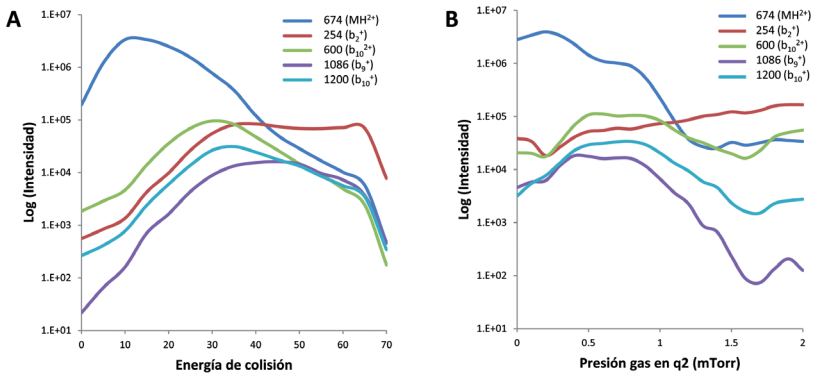


Figura 8. Efecto de la energía y presión de colisión sobre la eficacia de fragmentación de un ión dicargado derivado de la substancia P, utilizando un sistema de triple cuadrupolo TSQ Quantum (ThermoFisher). A) La energía de colisión alcanza un máximo a valores de aproximadamente -30 V para iones derivados de la fragmentación de la cadena lineal del péptido y valores superiores (-40 V) para iones de bajo peso molecular como iones imonio o iones reportero. B) La presión de colisión optimizó a 0.4-0.5 mTorr para todos los iones seleccionados.

3.1. Tipos de barrido en un triple cuadrupolo

Los principales modos de barrido en espectrometría de masas en tándem con un analizador de triple cuadrupolo son los que se indican a continuación [19] (Figura 9).

Modo de barrido de iones producto (antes también denominado de iones fragmento o iones hijo, en inglés *fragment* y *daughter ions*). En este caso se selecciona un ión de una masa determinada en el primer cuadrupolo. Este ión fragmenta en la cámara de colisión y los iones fragmento obtenidos se analizan en el tercer cuadrupolo. Este tipo de barrido es muy útil en estudios de elucidación estructural y es el que se utiliza para secuenciar péptidos.

Modo de barrido de iones precursores (antes también denominado de iones padre, *parent ions*). En este caso se barren todas las masas en Q1 y se fragmentan los iones en q2. En Q3 se filtra un determinado ión denominado fragmento diagnóstico. Todos los iones que cuando se fragmentan en la cámara de colisión producen el ión seleccionado en Q3 son detectados. Este tipo de barrido es muy útil para identificar miembros de una misma familia definidos por la presencia de uno o varios iones producto comunes.

Modo de barrido de pérdidas neutras. En este caso ambos analizadores están barriendo a la vez pero con una diferencia de masas entre los dos. Esta diferencia de masa corresponde a

una pérdida de un fragmento o molécula neutra. Si el ión que pasa por Q1, al colisionar en q2, sufre esta pérdida, será detectado en Q3. Este tipo de barrido es útil para identificar la presencia de grupos de compuestos que contienen un determinado grupo funcional.

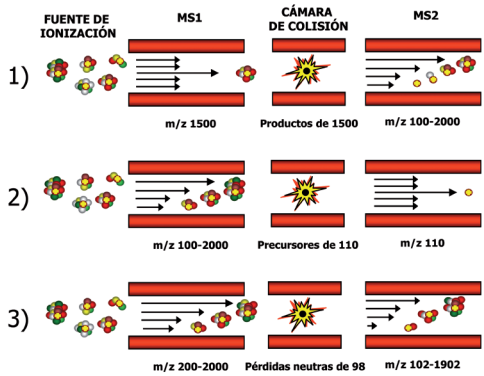


Figura 9. Modos de barrido en un triple cuadrupolo.
1) Modo de barrido de iones fragmento.
2) Modo de barrido de iones precursores.
3) Modo de barrido de pérdidas neutras.

Dentro del tipo de barrido de iones producto existe la posibilidad de monitorizar de forma selectiva en Q3 uno o varios iones fragmento previamente seleccionados. Esto es lo que se denomina *Selected Reaction Monitoring (SRM)* o *Multiple Reaction Monitoring* (MRM). En este modo de barrido tanto el primer (Q1) como el último (Q3) analizador de masas se usan como filtro de iones. En Q1 se filtra el ión de interés que se fragmenta en q2, y Q3 selecciona el ión producto. Este tipo de barrido se aplica a la cuantificación de moléculas pequeñas desde principios de los setenta pero no ha sido hasta recientemente que su uso se ha introducido en el análisis de proteómica cuantitativa gracias a la aparición de nuevos espectrómetros de triple cuadrupolo capaces de realizar más de 300 transiciones de SRM por segundo y más de 40000 en un solo método. Además estos nuevos espectrómetros alcanzan límites de cuantificación de péptidos en muestras complejas inferiores a los femtomoles, con un rango dinámico de hasta 6 órdenes de magnitud.

4. Espectrómetros híbridos.

El cuadrupolo se ha utilizado en instrumentos para espectrometría de masas en tándem acoplado a otros tipos de analizadores como los de sectores, TOF, trampas iónicas (IT) y Orbitrap. Estos espectrómetros híbridos permiten aprovechar al máximo las características de cada tipo de analizador. En proteómica se utilizan principalmente dos tipos de sistemas híbridos, los espectrómetros tipo QqTOF, disponibles en el mercado desde mediados de los años noventa, y los QqIT, de aparición posterior.

Durante mucho tiempo, los analizadores TOF se utilizaron casi exclusivamente acoplados a fuentes MALDI ya que la forma de trabajo pulsátil de estos sistemas los hacía complementarios. Esta misma característica dificultó el uso de estos analizadores con sistemas de electrospray o como analizadores en sistemas de espectrometría de masas en tándem. El desarrollo de la inyección de iones pulsada ortogonal ha permitido el acoplamiento ESI-TOF y la utilización de sistemas de analizadores híbridos QqTOF capaces de realizar experimentos de MS y MS/MS de alta resolución. La operación en este tipo de espectrómetro de masas en tándem es esencialmente la misma que para el triple cuadrupolo excepto que los iones productos que salen de q2 después de la colisión son ortogonalmente introducidos en el analizador TOF para adquirir el espectro [20]. Los instrumentos QqTOF permiten aprovechar las capacidades de alta resolución del TOF para la asignación de los iones fragmento, junto con la habilidad de obtener una selección de iones eficiente y una buena disociación utilizando el filtro cuadrupolar [21].

La trampa iónica es el análogo tridimensional del analizador de cuadrupolo [22]. Consiste en un recinto definido por tres electrodos: un electrodo anular (*ring electrode*) y dos electrodos de

superficie hiperbólica denominados electrodos de cierre (*endcap electrodes*) que se sitúan a modo de tapaderas a cada lado del electrodo anular. Estos electrodos de cierre tienen en el centro un pequeño agujero que permite a los iones entrar o salir de la cavidad. El electrodo anular está situado entre los electrodos de cierre. En el modo de trabajo normal los electrodos de cierre están conectados a tierra y es en el anular donde se aplican los voltajes. Los iones producidos en la fuente son introducidos en la trampa donde pueden acumularse durante un cierto tiempo. Posteriormente, mediante un barrido de potencial se determina el valor de m/z de todos los iones atrapados en la cavidad. A diferencia de los cuadrupolos donde en cada momento se monitoriza un tipo de ión mientras que se filtran todos aquellos de valor m/z diferente, en las trampas se analiza todo el paquete de iones acumulado. Este modo de trabajo confiere a la trampa una mayor sensibilidad en el modo de barrido completo, mientras que el cuadrupolo la supera en el caso de barrer en modo *selected ion monitoring* (SIM). Las desventajas de cuadrupolo frente a la trampa es que este no tiene la posibilidad de realizar, de forma eficiente, barridos de alta resolución, mientras que la trampa sí.

Las trampas iónicas son capaces de llevar a cabo análisis de espectrometría de masas en tándem mediante la eliminación selectiva de todos los iones acumulados excepto los del precursor seleccionado, su fragmentación mediante colisiones energéticas con moléculas de gas presentes en el propio recinto y el análisis de los fragmentos producidos de la misma forma que en el caso del análisis MS simple. Así, mientras que en los sistemas de triple cuadrupolo los análisis de espectrometría de masas en tándem son llevados a cabo en un proceso de separación en el espacio, un único analizador de trampa iónica puede llevar a cabo este proceso en el tiempo. En contrapartida un único analizador de trampa iónica no puede llevar a cabo dos de los modos de barrido característicos del triple cuadrupolo: el de iones precursores y el de pérdidas neutras (Tabla 2) y aunque puede mimetizarlas a nivel de análisis de datos, no alcanza la eficiencia y sensibilidad de aquellos.

El espectrómetro híbrido QqIT es una variante de un triple cuadrupolo en el que el tercer cuadrupolo Q3 se sustituye por una trampa iónica lineal. Este instrumento supera las limitaciones de la baja resolución de los cuadrupolos y a la vez permite la espectrometría de masa en tándem múltiple exclusiva de las trampas iónicas. Además se pueden realizar todos los tipos de barrido propios de un triple cuadrupolo (ver apartado 3.1) y no tiene los problemas de la limitación en masas bajas de las trampas iónicas en los espectros de MS/MS ya que la fragmentación se produce en q2 y no en la trampa [23].

Tabla 2. Principales diferencias entre trampas iónicas y cuadrupolos.

Características	Trampa iónica	Cuadrupolo
Separación de masas	En el tiempo	En el espacio
Sensibilidad en modo de barrido completo	Alta	Baja
Sensibilidad en modo SRM	Baja	Alta
Espectrometría de masas en tándem múltiple	Si	No
Modo de barrido de iones precursores y de pérdidad neutras	No	Si

5. Aplicaciones en proteómica.

Un rasgo que caracteriza a los triples cuadrupolos es la sensibilidad y especificidad de sus modos característicos de barrido en espectrometría de masas en tándem. En este capítulo vamos a centrarnos en dos de las aplicaciones principales: la fosfoproteómica y la proteómica cuantitativa.

5.1. Fosfoproteómica

Modificaciones postraduccionales como la fosforilación, acetilación o nitrosilación implican la adición de un grupo químico a ciertos aminoácidos de la secuencia de la proteína. El análisis por espectrometría de masas en tándem de los correspondientes péptidos modificados puede dar lugar a fragmentos neutros o iones característicos de esa modificación. En los sistemas de triple cuadrupolo es posible utilizar el modo de barrido de iones precursores para detectar en una mezcla de péptidos únicamente aquellos precursores que producen eficientemente un ión producto específico. De forma similar, usando el modo de barrido de pérdidas neutras estos sistemas permiten la detección de aquellos iones precursores que al fragmentarse dan lugar a la eliminación de una misma molécula neutra [24].

Utilizando el modo de barrido de iones precursores y fijando en Q3 el ión correspondiente al ión fosfato (PO_3^-, m/z -79) se puede detectar la presencia de fosforilación y determinar aquellos precursores que al fragmentar producen este ión. Esto ocurre para todos los aminoácidos fosforilados. Con el modo de barrido de iones precursores también es posible detectar péptidos fosforilados en tirosina monitorizando los iones precursores que como fragmento dan el ión imonio de fosfotirosina m/z 216.043.

Asimismo, en el caso de péptidos fosforilados en serina y treonina podemos usar el modo de barrido de pérdidas neutras para la identificación de péptidos fosforilados. En este caso se pueden monitorizan los iones precursores que al fragmentarse producen una pérdida neutra de 98, correspondiente a la pérdida de ácido fosfórico (H_3PO_4) (Ver Capítulo 22, Figura 7).

En ambos casos el espectro resultante que obtendremos contendrá únicamente los iones moleculares de los fosfopéptidos. Así pues, si tenemos información de secuencia de la proteína, en una mezcla sencilla, esta información es suficiente para identificar los péptidos fosforilados. Sin embargo, para determinar el sitio exacto de fosforilación hay que obtener información de secuencia a partir del espectro de MS/MS de iones fragmento. En general, se combina el barrido de iones precursores o el barrido de pérdidas neutras con el de iones fragmento para determinar qué péptidos están fosforilados, la secuencia y el punto de fosforilación de los mismos.

Una de las ventajas de los instrumentos híbridos que combinan el analizador de cuadrupolo con la trampa iónica es la posibilidad de cambiar del modo de barrido de iones precursores al modo positivo de MS/MS en un ciclo que puede durar menos de 5 segundos. De este modo se pueden detectar específicamente los iones fosforilados aprovechando la selectividad y sensibilidad del modo de barrido de iones precursores y a la vez aprovechar la alta sensibilidad de la trampa iónica en modo MS/MS para la secuenciación de los fosfopéptidos [25].

5.2. Proteómica cuantitativa mediante SRM

El análisis mediante SRM se introdujo a finales de los años 70 [26, 27] junto con el desarrollo de los triples cuadrupolos y desde entonces es la técnica de referencia para el análisis cuantitativo de pequeñas moléculas, sobre todo en aplicaciones clínicas y en la industria farmacéutica. Sin embargo, solo recientemente se ha extendido el análisis mediante SRM de péptidos diana en el campo de la proteómica. De hecho, se trata de una de las aplicaciones de la espectrometría de masas que ha experimentado un mayor auge en proteómica en los últimos años [28-30].

La aproximación proteómica denominada de perdigonada o *shotgun proteomics* es uno de los métodos más extendidos para el estudio de proteomas complejos [31, 32]. En esta aproximación las proteínas del proteoma de estudio se digieren y los péptidos resultantes se analizan mediante cromatografía líquida acoplada a la espectrometría de masas. Se trata de una herramienta muy potente para la identificación de péptidos y proteínas, pero que no es óptima para el análisis cuantitativo. La técnica de SRM, capaz de cuantificar de forma precisa, se presenta como el complemento ideal a la proteómica de *shotgun* en los estudios proteómicos de mues-

tras biológicas complejas. Por ejemplo, se han usado los triples cuadrupolos y la técnica de SRM para la cuantificación de biomarcadores en distintos fluidos biológicos como el suero, el plasma o la orina [33-35].

En el modo de barrido de SRM el primer y el tercer cuadrupolo actúan como filtros de masas, de manera que seleccionan de forma específica los iones predefinidos, correspondientes al ión precursor y a uno o más iones fragmento. Las transiciones se monitorizan a lo largo del tiempo dando lugar a una traza cromatográfica, de la que obtendremos información sobre el tiempo de retención y la intensidad de la señal para un determinado péptido (Figura 10), y de donde obtendremos la información cuantitativa.

Inicialmente se diferenciaba entre los términos *selected reaction monitoring* (SRM) y *multiple reaction monitoring* (MRM) en función de si se monitorizaba una o más transiciones de un ión precursor. Actualmente, se han unificado estos dos términos en SRM para evitar ambigüedades sobre el número de transiciones o el número de estados usados en el análisis de espectrometría de masas [36].

El modo de barrido de SRM implica un filtrado de los iones y una selección de masas a dos niveles, de manera que se obtiene una elevada selectividad en el análisis LC-MS, eliminándose se forma efectiva iones derivados de sustancias interferentes de estructura diferente a la monitorizada. Además, mediante SRM no se realiza un barrido de masas completo, sino que el analizador monitoriza continuamente los iones seleccionados. Esto aumenta la sensibilidad del método en dos órdenes de magnitud en comparación con las técnicas que obtienen barridos completos de MS/MS. Estas dos características proporcionan a la técnica una respuesta lineal en un rango dinámico de hasta cinco órdenes de magnitud, de manera que pueden detectarse y cuantificarse péptidos poco abundantes en muestras complejas. Todo esto hace del SRM la técnica idónea para el análisis cuantitativo. Sin embargo, a diferencia de los análisis mediante proteómica *shotgun,* el análisis mediante SRM requiere un conocimiento previo de la muestra que nos permita elegir la proteínas, los péptidos y las transiciones que se van a monitorizar. Existen ciertas especificaciones que hay que tener en cuenta cuando se quiere aplicar el SRM al estudio de proteomas, tanto en la selección de las proteínas y péptidos, como en la selección de las transiciones.

Generalmente, el análisis cuantitativo en proteómica se dirige a la determinación de una o unas pocas proteínas de interés seleccionadas previamente mediante, por ejemplo, un análisis *shotgun* o en base a hipótesis funcionales derivadas de un conocimiento previo. Cada unas de estas proteínas da lugar a decenas o centenares de peptidos trípticos, que son los que finalmente son cuantificados en técnicas de SRM. En general, para determinar la presencia o la cantidad de una proteína se analizan sólo unos pocos péptidos representativos de ésta. En principio cualquier péptido es susceptible de ser analizado mediante SRM, pero en la práctica es muy importante seleccionarlos cuidadosamente para asegurar el éxito del experimento. Existen una serie de reglas para la selección previa de péptidos en un análisis mediante SRM que vamos a discutir a continuación [37, 38].

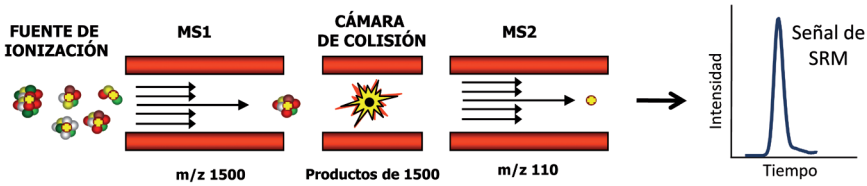


Figura 10. Modo de barrido de SRM y traza cromatográfica resultante. En el primer cuadrupolo se filtra el ión de interés y en el tercer cuadrupolo se filtra uno o más iones producto. Este tipo de barrido ofrece una gran sensibilidad y especificidad y se usa sobretodo en experimentos de cuantificación.

En primer lugar, deben elegirse los que se denominan péptidos proteotípicos, es decir, péptidos cuya secuencia se encuentra únicamente en la proteína a cuantificar y que por tanto la identifican. Si un péptido puede pertenecer a más de una proteína el resultado de cuantificación será ambiguo. Los péptidos proteotípicos deben tener características fisicoquímicas que permitan un análisis mediante LC-MS sensible y reproducible. Deben descartarse aquellos péptidos muy hidrofílicos, que no se retienen bien en la columna, o muy hidrofóbicos, que se retienen demasiado. Además su eficiencia de ionización tiene que ser buena, y su relación m/z tiene que estar comprendida dentro del rango de análisis del cuadrupolo.

Dentro de la preparación de la muestra hay otros parámetros que hay que considerar, como es su solubilidad después de la digestión. Los péptidos que no presenten una digestión completa o que sean degradables no son una buena elección ya que pueden dar lugar a resultados inconsistentes. Finalmente, hay que tener en cuenta que es mejor evitar los péptidos con aminoácidos susceptibles de ser modificados químicamente ya que pueden dar lugar a un sesgo en la cuantificación. Dentro de estas modificaciones encontramos la alquilación de la cisteína, la oxidación de la metionina o la deamidación de la asparagina. Los péptidos con modificaciones postraduccionales, a no ser que sean el objeto del estudio, tampoco son buenos candidatos para su cuantificación mediante SRM ya que pueden encontrarse en diferentes formas.

Para elegir los péptidos proteotípicos podemos recurrir a información basada en evidencias experimentales o bien basada en predicciones. Los datos de evidencias experimentales se pueden obtener o bien de análisis previos realizados en el laboratorio mediante otras aproximaciones como la proteómica de *shotgun*, o bien a través de bases de datos que contienen información espectrométrica de péptidos de distintas especies. Las bases de datos más utilizadas actualmente y que contienen información de los péptidos detectados en determinadas muestras y condiciones, así como el espectro de masas obtenido, son el PeptideAtlas [39], PRIDE [40] o GPMDB [41]. Para péptidos fosforilados se puede consultar también Phosida [42] o LymPhos [43].

Cuando no tenemos información experimental de nuestra muestra podemos recurrir a programas que permiten predecir los péptidos que mejor responden a este tipo de análisis. Algunos de ellos son ESP predictor [44] o PeptideSieve [45]. Se trata de programas que tienen en cuenta las propiedades físico-químicas de los péptidos para predecir su comportamiento en un análisis mediante LC-MS.

En la práctica, la selección de péptidos proteotípicos no asegura la especificidad del análisis SRM. Péptidos de diferente secuencia pero de la misma masa pueden coeluir con nuestro péptido y sus iones producir fragmentos cuya masa coincidan con los de aquél. Para asegurar la especificidad del análisis es por tanto necesario seleccionar cuidadosamente las transiciones a monitorizar, evitándose aquellas derivadas de transiciones no específicas de secuencia como son las perdidas neutras. De esta elección dependerá además la sensibilidad, dependiente de la intensidad de la transición o transiciones seleccionadas.

Generalmente se seleccionan dos o tres transiciones para cada uno de los péptidos. Del ión precursor se elige el valor de m/z de mayor intensidad. Es difícil predecir la distribución de cargas de un determinado péptido, ya que es dependiente de las condiciones experimentales, de manera que cualquier información previa que nos indique que nuestro péptido se detecta mayoritariamente con carga dos o carga tres en nuestras condiciones experimentales de LC-MS es de gran utilidad para maximizar la sensibilidad del experimento. Sin esta información podemos elegir transiciones correspondientes al péptido doblemente cargado y al péptido triplemente cargado, pero esto reducirá a la mitad el número de péptidos a analizar en el mismo experimento.

Los iones fragmento a monitorizar dependerán del tipo de instrumento y de la energía de colisión. Como en el caso de la elección de los iones precursores, para elegir los iones producto

se puede recurrir a datos experimentales previos disponibles en bases de datos (SRMAtlas [46], PeptideAtlas [39], PRIDE [40], GPDB [41]).

Aunque a una determinada energía de colisión el patrón de fragmentación de los péptidos en distintos espectrómetros es muy similar, la intensidad relativa de éstos puede cambiar de forma importante. Aunque existe una buena correlación entre la fragmentación en instrumentos de Orbitrap con HCD y la de los triples cuadrupolos [47] las diferencias aumentan cuando se comparan trampas y cuadrupolos. Dado que la mayor parte de la información almacenada actualmente en estas bases de datos proviene de datos obtenidos en trampas iónicas, es siempre importante contrastar estos datos con datos experimentales propios obtenidos sobre la muestra o mediante el uso de estándares. En estos casos es posible optimizar la respuesta para cada una de las transiciones SRM seleccionadas determinando la energía óptima de colisión de cada uno los péptidos monitorizados.

Finalmente, si no se dispone de datos experimentales se puede recurrir a herramientas informáticas que nos dan una información de partida para evaluar las condiciones de trabajo [48].

Aunque la especificidad del método de SRM es muy elevada, la complejidad de las muestras analizadas no permite descartar la presencia de interferencias responsables de la producción de datos erróneos o ambiguos. La manera más efectiva de validar los resultados obtenidos es mediante el uso de un péptido marcado isotópicamente como estándar interno de referencia, método que por otra parte es considerado el más preciso desde el punto de vista cuantitativo. Este estándar interno, que tendrá el mismo tiempo de retención que nuestro péptido diana pero con masa distinta, debe además producir intensidades relativas entre los iones fragmentos iguales a las de los del péptido analizado. Una desviación en este sentido puede indicar que los resultados derivados de una determinada transición pueden estar afectados por la presencia de una interferencia.

Otro factor importante a considerar en el análisis SRM es la duración de cada uno de los ciclos de barrido. El tiempo total dependerá del número de péptidos a analizar, el número de transiciones para cada péptido y el tiempo que el espectrómetro de masas tarda en realizar el análisis de SRM. Es importante considerar todos estos parámetros para realizar una buena cuantificación, ya que como mínimo tenemos que tener unos nueve puntos que definan la traza de nuestro ión para obtener un buen pico cromatográfico que se pueda cuantificar. Como consecuencia, el número de péptidos que se puede cuantificar en un determinado tiempo es limitado. Así, con el objetivo de analizar el máximo número de transiciones en un único experimento se puede trabajar con ventanas de tiempo, se pueden definir las transiciones en función del tiempo de retención de cada uno de los péptidos. Actualmente los triples cuadrupolos son muy rápidos y trabajando con ventanas de tiempo pueden llegar a analizar hasta 3000 péptidos en un único cromatograma de LC-MS [29].

Como ya se ha indicado arriba, el análisis mediante SRM combinado con la utilización de estándares internos de los correspondientes péptidos diana marcados con isótopos estables, constituye el método más sensible, específico y preciso para la cuantificación absoluta de estos péptidos en la muestra problema. Los péptidos marcados se añaden a la muestra a una concentración conocida y la concentración del correspondiente péptido endógeno se determina en base a la relación entre sus señales y en base a una curva de calibrado determinada experimentalmente. Mediante esta estrategia se puede obtener una respuesta lineal de hasta seis órdenes de magnitud en un triple cuadrupolo. Sin embargo la complejidad de la muestra, el rango dinámico de las proteínas y la matriz puede disminuir significativamente los límites de detección y de cuantificación de los péptidos. Unos de los estándares internos más usados en proteómica son los péptidos AQUA, desarrollados por Gygi en el año 2003 [49]. Una excelente revisión sobre el estado actual la cuantificación absoluta de péptidos y proteínas mediante SRM se encuentra en la referencia [29].

6. Perspectiva.

En este capítulo hemos visto que los cuadrupolos fueron los primeros instrumentos usados de forma general para la secuenciación de péptidos. Hemos analizado sus principios operacionales y los parámetros de rendimiento. Los triples cuadrupolos poseen modos de barrido específicos que les otorgan una gran especificidad y sensibilidad. Esto ha hecho, juntamente con la mejora de su velocidad, que actualmente reaparezcan en el campo de la proteómica con gran fuerza. Se trata de instrumentos especialmente útiles en la proteómica cuantitativa cuando operan en modo SRM.

REFERENCIAS.

1. Biemann K, Gapp F, Seibl J. Applications of mass spectrometry to structure problems. I. Amino acid sequence in peptides. JAm Chem Soc. 1959;81:2274-5.

2. Barber M, Jollès P, Vilkas E, Lederer E. Determination of amino acid sequences in oligopeptides by mass spectrometry. I. The structure of fortuitine, an acylnonapeptide methyl ester. Biochem Biophys Res Comm. 1965;18:469-73.

3. Das BC, Géro SD, Lederer E. N-methylation of N-acyl oligopeptides. Biochem BiophysResComm. 1967;29:211-5.

4. Das BC, Géro SD, Lederer E. Detection and location of N-Methyl-amino-acid residues in N-acyloligopeptide methyl esters by mass spectrometry. Nature. 1968;217:547-8.

5. Paul W, Steinwedel H. A new mass spectrometer without a magnetic field. Zeitschrift fuer Naturforschung. 1953;a8:448-50.

6. Yost RA, Enke CG. Selected ion fragmentation with a tandem quadrupole mass spectrometer. JAm Chem Soc. 1978;100:2274-5.

7. Hunt DF, Bone WM, Shabanowitz J, Rhodes J, Ballard JM. Sequence analysis of oligopeptides by secondary ion/collision activated dissociation mass spectrometry. Anal Chem. 1981;53:1706-8.

8. Hunt DF, Buko AM, Ballard JM, Shabanowitz J, Giordani AB. Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. Biomed Mass Spectrom. 1981;8:397-408.

9. Barber M, Bordoli RS, Sedgwick RD, Tyler AN. Fast atom bombardment of solids (F.A.B.): a new ion source for mass spectrometry. J Chem Soc, Chem Commun. 1981:325-7.

10. Abian J. Espectrometría de masas de péptidos. In: Andreu D, Rivas L, editors. Péptidos en biología y biomedicina. Madrid: CSIC; 1997.

11. Hoffmann Ed, Stroobant V. Mass spectrometry: principles and applications. 3rd ed ed. Chichester, West Sussex (England): John Wiley & Sons; 2007.

12. Douglas DJ. Linear quadrupoles in mass spectrometry. Mass Spectrom Rev. 2009;28:937-60.

13. Du Z, Douglas DJ, Konenkov N. Elemental analysis with quadrupole mass filters operated in higher stability regions. J Anal At Spectrom. 1999;14:1111-9.

14. Olimpieri L, Traldi P. Quadrupole mass filters, quadrupole ion traps and Fourier transform ion cycotron resonance spectrometers. In: (eds.) RMCea, editor. Mass spectrometry in biomolecular sciences. Netherlands: Kluwer Academic Publishers; 1996. p. 177-200.

15. Esteban L. La espectrometría de masas en imágenes. Madrid1993.

16. Pedder RE, Lynch D, J. W. Optimizing quadrupole transmission for wide mass range to 10,000 amu. 47th ASMS Conference on Mass Spectrometry and Allied Topics. Dallas, Texas1999.

17. Yost RA, Enke CG. Triplequadrupolemass spectrometry for direct mixture analysis and structure elucidation. Anal Chem. 1979;51:1251-64.

18. Sleno L, Volmer DA. Ion activation methods for tandem mass spectrometry. J Mass Spectrom. 2004;39:1091-112.

19. Hoffmann Ed. Tandem mass spectrometry: a primer. J Mass Spectrom. 1996;31:129-37.

20. Morris HR, Paxton T, Panico M, McDowell R, Dell A. A novel geometry mass spectrometer, the Q-TOF, for low-femtomole/attomole-range biopolymer sequencing. Journal of protein chemistry. 1996;5:469-79.

21. Chernushevich IV, Loboda AV, Thomson BA. An introduction to quadrupole-time-of-flight mass spectrometry. J Mass Spectrom. 2001;36:849-65.

22. March RE. Quadrupole ion traps. Mass Spectrom Rev. 2009;28:961-89.

23. Hager JW, Yves Le Blanc JC. Product ion scanning using a Q-q-Q linear ion trap (Q TRAP) mass spectrometer. Rapid Commun Mass Spectrom. 2003;17:1056-64.

24. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ. A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. Anal Chem. 2004; 76:3590-8.

25. Williamson BL, Marchese J, Morrice NA. Automated identification and quantification of protein phosphorylation sites by LC/MS on a hybrid triple quadrupole linear ion trap mass spectrometer. Mol Cell Proteomics. 2006;5:337-46.

26. Baty JD, Robinson PR. Single and multiple ion recording techniques for the analysis of diphenylhydantoin and its major metabolite in plasma. Biomedical mass spectrometry. 1977;4:36-41.

27. Zakett D, Flynn RGA, Cooks RG. Chlorine isotope effects in mass spectrometry by multiple reaction monitoring. J Phys Chem. 1978;82:2359.

28. Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol. 2008;4:222.

29. Gallien S, Duriez E, Domon B. Selected reaction monitoring applied to proteomics. J Mass Spectrom. 2011;46:298-312.

30. Elschenbroich S, Kislinger T. Targeted proteomics by selected reaction monitoring mass spectrometry: applications to systems biology and biomarker discovery. Mol Biosyst. 2010;7:292-303.

31. Washburn MP, Wolters D, Yates JR, 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19:242-7.

32. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422:198-207.

33. Selevsek N, Matondo M, Sanchez Carbayo M, Aebersold R, Domon B. Systematic quantification of peptides/proteins in urine using selected reaction monitoring. Proteomics. 2011;11:1135-47.

34. Barnidge DR, Goodmanson MK, Klee GG, Muddiman DC. Absolute quantification of the model biomarker prostate-specific antigen in serum by LC-Ms/MS using protein cleavage and isotope dilution mass spectrometry. J Proteome Res. 2004;3:644-52.

35. Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol Cell Proteomics. 2006;5:573-88.

36. Murray KK, Boyd RK, Eberlin MN, Langley GJ, Li L, Naito Y. Standard definitions of terms relating to mass spectrometry. IUPAC Recomendations, Public Draft. 2006.

37. Bronstrup M. Absolute quantification strategies in proteomics based on mass spectrometry. Expert Rev Proteomics. 2004;1:503-12.

38. Kirkpatrick DS, Gerber SA, Gygi SP. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. Methods. 2005;35:265-73.

39. Deutsch EW, Lam H, Aebersold R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. EMBO Rep. 2008;9:429-34.

40. Jones P, Cote RG, Cho SY, Klie S, Martens L, Quinn AF, et al. PRIDE: new developments and new datasets. Nucleic Acids Res. 2008;36:D878-83.

41. Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. J Proteome Res. 2004;3:1234-42.

42. Gnad F, Ren S, Cox J, Olsen JV, Macek B, Oroshi M, et al. PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. Genome Biol. 2007;8:R250.

43. Ovelleiro D, Carrascal M, Casas V, Abian J. LymPHOS: design of a phosphosite database of primary human T cells. Proteomics. 2009;9:3741-51.

44. Fusaro VA, Mani DR, Mesirov JP, Carr SA. Prediction of high-responding peptides for targeted protein assays by mass spectrometry. Nat Biotechnol. 2009;27:190-8.

45. Mallik B, Zhang L, Koide S, Morikis D. pH dependence of stability of the 10th human fibronectin type III domain: a computational study. Biotechnol Prog. 2008;24:48-55.

46. Picotti P, Lam H, Campbell D, Deutsch EW, Mirzaei H, Ranish J, et al. A database of mass spectrometric assays for the yeast proteome. Nat Methods. 2008;5:913-4.

47. de Graaf EL, Altelaar AF, van Breukelen B, Mohammed S, Heck AJ. Improving SRM Assay Development: A Global Comparison between Triple Quadrupole, Ion Trap, and Higher Energy CID Peptide Fragmentation Spectra. J Proteome Res. 2011;10:4334-41.

48. Cham Mead JA, Bianco L, Bessant C. Free computational resources for designing selected reaction monitoring transitions. Proteomics.10:1106-26.

49. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A. 2003;100:6940-5.

Espectrómetros de masa basados en tiempo de vuelo: La importancia de la masa exacta en Proteómica

10

Isidre Masana González

Especialista Productos LC/MS. Agilent Technologies – Barcelona.

1. Introducción.

Los inicios de la cromatografía y espectrometría de masas coinciden a principios del siglo XX: en 1906 se inició el desarrollo de ambas, y cuya posterior evolución tanto ha revolucionado el análisis proteómico:

- En 1906 el botánico ruso M.S. Tswett, publicó los primeros trabajos (1)(2) de una nueva técnica para separar pigmentos vegetales, a la que llamó cromatografía. Como técnica analítica instrumental, nació en la década de los 50's con la aparición de la "cromatografía de gases" (GC); para aplicarla al análisis de compuestos volátiles. La "cromatografía líquida de alta eficacia" (HPLC), utilizada hoy en día en proteómica, inició su desarrolló a finales de los años 60.

- En 1906 J. J. Thomson, de la Universidad de Cambridge recibió el premio Nobel de Física. Thomson fue el descubridor del electrón y los isótopos, y se le considera el inventor de la espectrometría de masas, por los estudios realizados con un instrumento entonces denominado "espectrógrafo de parábola". Éste se considera el primer espectrómetro de masas de la historia, y con él relacionaba carga y masa por la desviación de su movimiento (generado por un campo magnético).

- En 1946 William E. Stephens de la Universidad de Pennsylvania propuso (3) el concepto de analizador de masas basado en "Tiempo de Vuelo" (TOF).

- En la década de los 90, los analizadores de masas basados en "Tiempo de Vuelo" (TOF) empezaron a adquirir una cierta popularidad; gracias a los espectaculares avances de la electrónica e informática. Hoy en día se pueden determinar tiempos de vuelo de los iones, con precisiones por debajo del nanosegundo, y diferencias de masa inferiores al peso de un electrón .

Aplicando el principio de "divide y vencerás", la simbiosis entre el excelente "poder separador" de la cromatografía y la "gran selectividad y enorme poder identificador" de la espectrometría de masas, han convertido al acoplamiento HPLC/MS (simplificado: LC/MS) en la herramienta analítica más potente de la que se dispone hoy en día para el estudio del proteoma humano.

1.1 Evolución histórica de los tiempos de vuelo.

El concepto de analizador de masas basado en "Tiempo de Vuelo" (TOF), nace en 1946 cuando William E. Stephens de la Universidad de Pennsylvania, lo presentó a la Sociedad de Físicos Americanos (3). Cada milisegundo, se generaba un pulso de una duración de microsegundos, que empujaba a los iones a volar en el interior de un tubo de vacío, donde los iones con distinta *m/z* se movían a distinta velocidad. Los iones se recogían en una jaula de Faraday, y los pulsos de la señal amplificada se recogían en un oscilógrafo cuyo barrido se sincronizaba con los pulsos, para poder mostrar en pantalla el espectro de MS. En la tabla 1 se muestra un resumen de su evolución histórica, y que a continuación se describe con más detalle:

- En 1955 Wiley y McLaren (4) mejoran la focalización y detección, desarrollando un espectrómetro capaz de resolver masas nominales por encima de 100 Daltons, y espectros con una resolución de hasta 300. Los espectros se registraban mediante osciloscopios electrónicos. El desarrollo de éste (5), da lugar en 1958 al primer instrumento comercial: el Bendix TOFMS fabricado por Bendix Corp. (Detroit). Aunque su resolución espectral era muy pobre comparada con los sectores magnéticos ya comerciales en aquella época.

- En 1966 Mamyrin y colaboradores (6) describen el reflectrón de masa, un dispositivo para refocalizar la dispersión de energía de iones con la misma relación *m/z*, y aumentar enormemente la resolución espectral; elemento que ha resultado clave para convertir a los TOF en espectrómetros de masa de alta resolución.

- En 1984 Glish y Goeringer (7) desarrollan un nuevo espectrómetro de masas híbrido para estudios MS/MS: el Cuadrupolo-Tiempo de Vuelo (QTOF). El espectrómetro basado en cuadrupolos había sido desarrollado previamente por Paul en 1953.

- En 1985 M. Karas y F. Hillenkamp (8) desarrollan la tecnología de ionización MALDI ("Matrix-Assisted Laser Desorption Ionization"), aplicada al análisis de aminoácidos y dipéptidos. En 1987 K. Tanaka desarrolló la aplicación de MALDI-TOF, para analizar proteínas intactas (Lysozima y Carboxipeptidasa A) y moléculas orgánicas de hasta *m/z* = 71.000. Estos desarrollos han sido claves para popularizar el uso de los TOF en proteómica, y permitieron a Tanaka obtener el premio Nobel de Química en el 2002.

- En 1989 J. Fehn y colaboradores (9) aplican la tecnología electrospray (desarrollada en 1968 por Dole), a la ionización de biopolímeros de hasta 130.000 Daltons, incluyendo proteínas y oligonucleótidos. En el 2002 le conceden el premio Nobel de Química (compartido con Tanaka) por sus trabajos para el análisis de macromoléculas biológicas.

- En 1989 Dawson y Gilhaus (10) presentan por primera vez un tiempo de vuelo con aceleración ortogonal (oa-TOF), posteriormente ampliamente utilizado en combinación con fuentes de ionización a presión atmosférica, en especial con electrospray.

- En 1996 la firma Hewlett-Packard/Agilent Technologies patentó (11) e introdujo por primera vez en el mercado, un LC/MS con fuentes de ionización Electrospray y APCI ortogonales (ángulo de 90º entre las direcciones de nebulización y entrada en el MS), y desolvatación en contracorriente. Este nuevo tipo de fuentes ("off axis") significó, con respecto a la tecnología "on-axis" existente hasta entonces, un avance muy importante en:

  - Versatilidad: poder trabajar hasta más de 1mL/min (en lugar de los 50-200µL/min del resto de ESI).
  - Robustez: el interior del MS apenas se ensucia.
  - Facilidad de uso de la tecnología: limpieza sencilla y ya no se requiere ajustar la posición del nebulizador.
  - Sensibilidad.

Este desarrollo, aún vigente en la actualidad (ver zona izquierda figura 3), inició e inspiró una nueva generación de fuentes de ionización a presión atmosférica, mucho más robustas y fáciles de usar, gracias a las cuales la tecnología LC/MS se ha popularizado en muchos laboratorios, y ha experimentado un muy importante crecimiento.

Todos estos avances, en combinación con el desarrollo de la informática y la electrónica, han permitido que hoy en día se disponga de una muy avanzada tecnología de tiempo de vuelo, que compite en prestaciones (exactitud de masa, resolución espectral,….) con otras de mayor coste:

- En la conferencia ASMS (American Society Mass Spectrometry) del 2007, Muddiman y colaboradores de la Universidad de Carolina del Norte (12) presentaron un poster en el que compararon la exactitud de masa de un sector magnético de doble foco frente a 3 LC/MS-TOF de 3 firmas distintas; a pesar de la menor resolución espectral de los TOF, estos presentaron resultados semejantes en exactitud y precisión de masa, e incluso considerablemente mejores en uno de los modelos de TOF (con especificación <3ppm y 10K resolución).

- Smith, Lubda y colaboradores (13) comparan un LC/MS-TOF frente a un ESI-FTICR en condiciones de cromatografía rápida (1 a 5 min.) obteniendo resultados similares.

La limitada aplicabilidad cuantitativa que proporcionaban los TOF's y QTOF's hasta mediados de la década de los 2000, han dado lugar hoy en día a equipos con unas muy buenas prestaciones cuantitativas (buena sensibilidad, linealidad y repetibilidad), y que además proporcionan unas excelentes prestaciones cualitativas (masa exacta, resolución espectral y fidelidad isotópica).

Tabla 1.- Resumen de la evolución histórica de espectrómetros de tiempo de vuelo.

Año	Evento	Autor
1946	Concepto de Tiempo de Vuelo	Stephens WE
1958	Primer TOFMS comercial	Bendix Corp. (Detroit).
1966	Reflectrón	Mamyrin
1984	Cuadrupolo-Tiempo de Vuelo (QTOF)	Glish y Goeringer
1985	MALDI	Karas y Hillenkamp
1987	Aplicación MALDI-TOF a proteínas	K. Tanaka
1989	Aplicación Electrospray a proteínas y oligonucleótidos	J. Fehn
1989	TOF con aceleración ortogonal (oa-TOF)	Dawson y Gilhaus
1996	Electrospray con nebulización ortogonal	Hewlett Packard/ Agilent Technologies

1.2. Prestaciones de los tiempos de vuelo actuales.

Exactitud de masa, resolución espectral, sensibilidad, rango dinámico, fidelidad isotópica y herramientas informáticas para automatizar la extracción de la ingente cantidad de información, que puede almacenar un fichero de datos de LC/MS con masa exacta, son sin duda las prestaciones claves que deben reunir los sistemas LC/MS/MS-QTOF o LC/MS-TOF. Buena parte de estos conceptos fueron descritos en el capítulo de "Espectrometría de masas biológica. Conceptos generales".

Las elevadas prestaciones que proporcionan los espectrómetros actuales (ver Figura 1), con acelerador ortogonal y reflectrón, se pueden resumir en:

- **Exactitud de masa: 1 - 5ppm.** Los mejores equipos dan especificaciones de <1-2ppm en modo MS, y del orden del doble trabajando en modo MS/MS. Un error de 1 ppm significa un error de 1 millonésima parte:

  - 0.001Daltons en un péptido de 11.000Da.
  - El peso de 1 electrón en una molécula de 600Da.
  - Estos niveles de exactitud compiten con los proporcionados por otras tecnologías (sectores magnéticos, Orbitraps, FTICR,…) de coste considerablemente superior.

Figura 1.- Resumen aportaciones analíticas de los espectrómetros de masas basados en tiempo de vuelo: exactitud de masa, resolución espectral y capacidad de analizar retrospectivamente (incluso a nivel cuantitativo), con excelente sensibilidad espectral en equipos de última generación, son las características más relevantes de los tiempos de vuelo.

- **Resolución espectral: 10.000-50.000** (FWHM: Full Width at Half Maximum), 35-50K en los instrumentos de gama alta. Ello permitirá resolver espectralmente compuestos que difieran en centésimas de Dalton. Aunque no en todos, en buena parte de equipos el trabajar a máxima resolución significa reducir considerablemente la sensibilidad. En los TOF, el poder resolutivo (resolución) se calcula como el cociente entre la masa y el ancho del pico de masa a la mitad de su altura (FWHM). Por tanto a mayor resolución más fino será el pico de masa; así para una masa de 1.000 Da el ancho del pico de masa al 50% de su altura (FWHM) sería:

  - 1.000 Da con Rs 10.000 = 0.1Da.
  - 1.000 Da con Rs 50.000 = 0.02Da.

- **Sensibilidad en modo MS (espectro completo): 0.2-100 pg** de reserpina inyectados en columna como límites de cuantificación en los instrumentos actuales, con un rango de 0.2-5pg en los mejores equipos. Las especificaciones de los equipos se dan en condiciones de máxima sensibilidad, trabajando en modo de baja resolución en instrumentos que no pueden trabajar simultáneamente a máxima resolución y sensibilidad.

  Hoy en día, la sensibilidad de los QTOF más sensibles está al mismo nivel que los triples cuadrupolos de gama media en modalidad MRM; con la ventaja del QTOF de siempre proporcionar el espectro completo.

- **Rango dinámico: 2–5 órdenes de magnitud** (100 a 100.000)**.** Éste se desdobla en 2 tipos:

  - **Rango dinámico espectral:** independencia de la exactitud de masa con respecto a la concentración del analito. Cambia mucho dependiendo del sistema de medida del tiempo de vuelo (ver final apartado 2.1):
    - 2 órdenes en sistemas con conversor digital del tiempo de vuelo (TDC).
    - 3-4 órdenes en sistemas con conversor analógico digital (ADC).
    - + 5 órdenes en sistemas ADC con doble amplificador.

  - **Rango dinámico cuantitativo:** los sistemas actuales proporcionan 4-5 órdenes, incluyendo los tipo TDC, que para solventar su facilidad de saturación recurren a desechar parte del haz de iones cuando la señal es excesivamente intensa. Un rango dinámico de 4 órdenes, indica que el equipo permite crear una curva de calibrado con un ratio, entre el patrón más concentrado y el más diluido, de hasta 10.000 veces (10^4).

- **Buena Fidelidad Isotópica:** buena parte de los TOF y QTOF, especialmente los de mayor rango dinámico espectral (se saturan con mayor dificultad), proporcionan ratios de intensidades del perfil de distribución isotópica muy próximos a los teóricos; aspecto clave para la determinación de fórmulas empíricas e identificación de compuestos desconocidos.

- **Buena capacidad cuantitativa** tanto en MS como en MS/MS. La mayoría de los TOF y QTOF que actualmente se ofrecen en el mercado, ofrecen unas buenas o excelentes prestaciones cuantitativas con buena sensibilidad, selectividad, repetibilidad y rango dinámico. Trabajando en modo MS, permiten cuantificar cualquier compuesto ionizado (incluso retrospectivamente); la robustez de su masa exacta (independencia con respecto a la concentración del analito), será clave para disponer de una excelente selectividad (tipo MS/MS); en este punto (se abordará en secciones 3.3. y 3.3.1) hay diferencias significativas entre los equipos actuales.

- **Capacidad Cualitativa y Cuantitativa simultánea (Qual/Quant):** la cada vez mayor sensibilidad y rapidez de los LC/MS-QTOF, permite en algunos casos llegar a combinar en1 sola inyección, tareas identificativas (Qual) y cuantitativas (Quant). Por ejemplo (p.e.) en modo Auto MS/MS, combinando MS para poder cuantificar cualquier compuesto ionizado y MS/MS para identificarlo. Se deberán balancear adecuadamente los parámetros de adquisición, para tomar un mínimo de 10-12 puntos de medida/pico en la dimensión de MS para cuantificar; al tener menos tiempo para cada ciclo de medida, se perderá algo de sensibilidad con respecto a adquisiciones específicas para cuantificar o identificar por separado.

2. Fundamentos teóricos de la espectrometría de masas basada en Tiempo de Vuelo.

Los espectrómetros de masas basados en tiempo de vuelo, se fundamentan en la determinación de la masa de los iones, a partir del tiempo que éstos "vuelan" para recorrer una determinada distancia. El enorme desarrollo de la electrónica e informática, permite disponer hoy en día de equipos capaces de medir tiempos de vuelo con precisiones por debajo de los nanosegundos, proporcionando rutinariamente exactitudes de masa por debajo de 1ppm, resoluciones espectrales de 50.000 y límites de detección por debajo del picogramo; todo ello unido a unas excelentes prestaciones cuantitativas.

En la figura 2 se representa un esquema simplificado del sistema de medida del tiempo de vuelo de los iones (ejemplo con iones monocargados). Para simplificar el esquema, no se ha incluido el espejo de iones (reflectrón) que incorporan los instrumentos actuales. Los iones, después de pasar por unas lentes de focalización vertical, entran alineados en la zona inferior del tubo de vuelo, donde la aplicación de un pulso de elevado voltaje, envía a los iones a volar ortogonalmente hacia la zona superior donde se sitúa el detector. La velocidad (v) a la que se mueven los iones dentro del tubo de vuelo, se puede calcular a partir de la energía cinética (E) que reciben del pulso de voltaje, mediante la fórmula:

$$E = (1/2)\, m\, v^2 \quad \rightarrow \quad v = (2E/m)^{1/2}$$

Es decir, la velocidad de los iones es inversamente proporcional a la raíz cuadrada de su masa (m); en realidad se debería sustituir m por *m/z* (masa/carga). En definitiva, si todos tienen la misma carga, los de menor masa serán los primeros en llegar al detector. El tiempo de vuelo

(t) de los iones, o su masa, se podrán calcular a partir de la ecuación que relaciona la velocidad (v) con la distancia a recorrer (d):

$$t = d / v \rightarrow t = d\,(m/2E)^{1/2} \rightarrow m = 2E\,t^2/d^2$$

Dada la enorme velocidad a la que mueven los iones en el interior del tubo de vuelo (típica entre 20 y 100Km/segundo), y los reducidos tiempos de vuelo (típico <100µs), el pulso se va repitiendo varias miles de veces por segundo (típico 10.000 pulsos/s. -dependerá del rango de masa a medir).

En la práctica existe un retraso entre el tiempo en que la electrónica envía la orden de iniciar un pulso de voltaje, y cuando realmente ese voltaje empieza a generarse en el pulsador. Por otro lado, existe un retraso adicional entre el tiempo en el que un ión alcanza la superficie del detector, y la electrónica genera la señal digitalizada. Por ello, en las anteriores ecuaciones, el tiempo (t) debería ser sustituido por la diferencia entre el tiempo medido (t_m) y la suma de tiempos de retraso (t_0): $t = t_m - t_0$. Adicionalmente el campo eléctrico (E) y la distancia (d) son constantes que pueden ser sustituidas por otra constante $A = 2E/d^2$. Consecuentemente la ecuación que relaciona la masa (m) -en realidad masa/carga (*m/z*)- con el tiempo quedará simplificada como:

$$m/z = A\,(t_m - t_0)^2$$

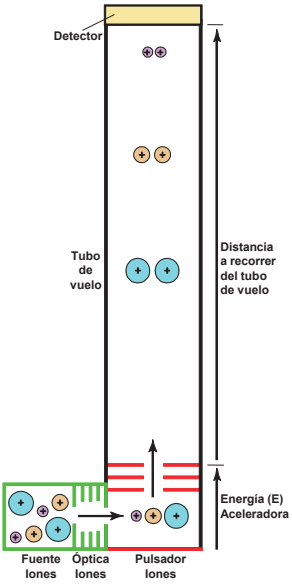


Figura 2.- Esquema simplificado (sin "reflectron") de sistema de medida del tiempo de vuelo de iones. En la base del tubo de vuelo los iones se aceleran ortogonalmente (mediante un pulso de alto voltaje) a su dirección de entrada, y se separan de acuerdo con su relación masa/carga; si todos tienen la misma carga, los de menor masa serán los primeros en llegar al detector. El pulso se va repitiendo varias miles de veces por segundo.

Los valores de A y t_0 se calculan mediante una regresión no lineal, cuando se efectúa un calibrado del eje de masa, utilizando un conjunto de compuestos de masa exacta conocida (típico 6-7 masas). Para mejorar la exactitud de masa, una vez calculados las constantes A y t_0, las pequeñas desviaciones observadas al aplicar estas constantes a la ecuación anterior, se corrigen mediante un segundo ajuste que utiliza una función polinomial de alto grado. Estas desviaciones suelen ser pequeñas y relativamente constantes a lo largo del tiempo.

Si se desea mayor exactitud se podrá trabajar introduciendo en continuo 2 (o más) compuestos de masa exacta conocida, para espectro a espectro, recalcular la A y t_0 específicas de cada espectro.

Para una descripción más detallada se puede consultar la publicación de J.Fjeldsted, "Time-of-Flight Mass Spectrometry Technical Overview" (14).

2.1. Zonas y Funciones de los Componentes de espectrómetros de Tiempo de Vuelo.

Este apartado explicará cómo funciona un espectrómetro de masas Cuadrupolo-Tiempo de Vuelo (QTOF/TOF). Se describirán sus distintas zonas y componentes, sus funciones y algunos comentarios sobre distintos diseños que actualmente (2011) se pueden encontrar en el mercado.

En la figura 3 se muestra el esquema de un espectrómetro de masas Cuadrupolo-Tiempo de Vuelo (QTOF); en él se pueden diferenciar 5 zonas. En el caso de los TOF's se reduciran a 3, al eliminarse la zona central de aislamiento y fragmentación de iones precursores, y unificarse las

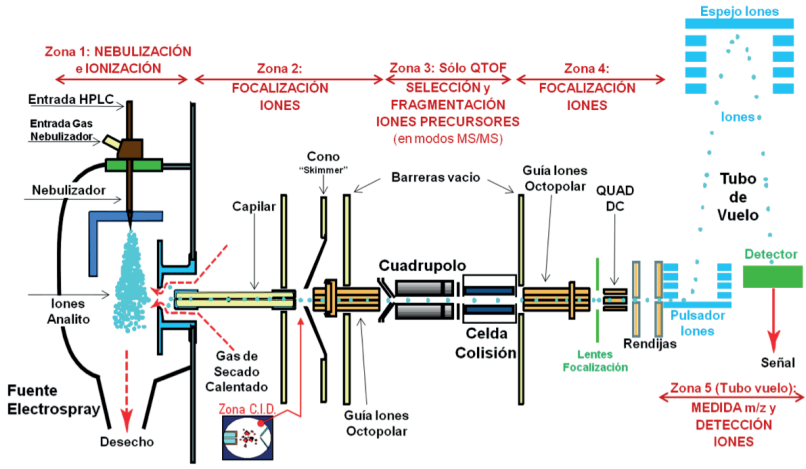


Figura 3.- Esquema de un espectrómetro de masas Cuadrupolo-Tiempo de Vuelo (QTOF): los iones formados en zona 1 se focalizan en la zona 2, se aíslan y fragmentan en zona 3 cuando se trabaja en modo MS/MS (en modo MS se transmitirían todos los iones sin fragmentar). En la zona 4 los iones se focalizan para entrar en la zona 5 (tubo de vuelo) donde el pulsador los envía a volar ortogonalmente para medir su masa exacta. En un TOF se eliminaría la zona 3 y se juntarían 2 y 4.

2 zonas de focalización. Los componentes que se utilizan en cada una de las zonas variarán en función del diseño de cada modelo y fabricante, pero en esencia sus funciones se podrían describir como:

Zona 1.-VAPORIZACIÓN e IONIZACIÓN: en esta zona denominada **Fuente de Iones** tiene lugar la nebulización, ionización y desolvatación de los analitos procedentes del HPLC. El objetivo es conseguir que los analitos de interés adquieran una carga, y separar los analitos del disolvente que les rodea. Existen diversos tipos de fuentes fácilmente intercambiables entre sí:

- **Fuentes online** para acoplamientos en línea con HPLC:
  Electrospray (ESI) y nano-Electrospray (o HPLC/Chip).
  Iónización Química (APCI).
  Fotoionización (APPI).
  Mixtas ESI+APCI

- **Fuentes offline:**
  MALDI (AP-MALDI).
  Desorption ElectroSpray Ionization (DESI).
  Direct Analysis in Real Time (DART).

Estas fuentes operan a presión atmosférica, la selección de unas u otras, dependerá del tipo de analito. En proteómica, cuando se trabaja con acoplamiento a cromatografía líquida, se utilizan fuentes basadas esencialmente en la ionización del analito por el pH de la disolución en la que se encuentran:

- **Electrospray (ESI):** para flujos desde unos pocos µL/min hasta 1-2mL/min, con columnas de diámetro interno desde 180µm hasta 4.6mm. La nebulización del analito se realiza mediante la utilización de un flujo de gas (típico nitrógeno). Para flujos por debajo de 50-100µL/min se suele recomendar utilizar un micronebulizador. También se podrá acoplar una electroforesis capilar (mediante el uso de un micronebulizador triconcéntrico), cuando interese una elevada capacidad de separación de péptidos mayoritarios; su mecanismo de separación ortogonal con respecto a HPLC con fase reversa, proporcionará información complementaria.

- En la figura 3 puede observarse una fuente electrospray con diseño ortogonal: nebulización asistida por nitrógeno en ángulo de 90º con respecto a la dirección de entrada de los iones al interior del espectrómetro, a través de un capilar. El gas de secado se suministra a través del propio orificio por el que han de entrar los iones para acceder al capilar (impulsados por el gradiente de voltaje existente en la zona); este diseño de secado en contracorriente (iones y gas en direcciones opuestas), proporciona una muy elevada capacidad de desolvatación y robustez en la operación, al evitar la entrada de neutros e iones de carga contraria al interior del espectrómetro.

- En la figura 4 puede observarse una fuente electrospray mejorada; en ella la pluma de nebulización se rodea de un cilindro de gas de secado, que focaliza el haz de iones y típicamente permite mejorar la sensibilidad por un factor de 5 veces, y aumentar el flujo máximo de trabajo (útil en UHPLC).

- Nano-Electrospray: en este caso el proceso de nebulización está asistido por un intenso campo eléctrico (no por nitrógeno); éste generará la pluma de nebulización a flujos por debajo de típicamente 1 µL/min, utilizando una punta de nebulización conductora que termina en un orificio de unas pocos µm (típico 8-12 µm). Estas fuentes ofrecen una sensibilidad de 10 a más de 100 veces superior a la que ofrecen las fuentes ESI (dependiendo del diámetro de la columna y tipo de fuente utilizada); no obstante son mucho más delicadas y menos robustas que las ESI; además el trabajar con nanocromatografía requiere de un cierto nivel de experiencia. Una alternativa es utilizar la versátil tecnología HPLC-Chip/MS; ésta solventa buena parte de estos inconvenientes (está descrita en el capítulo 14 de este libro).

Los fundamentos de estas técnicas de ionización se describen en profundidad en un capítulo previo dedicado a métodos y mecanismos de ionización.

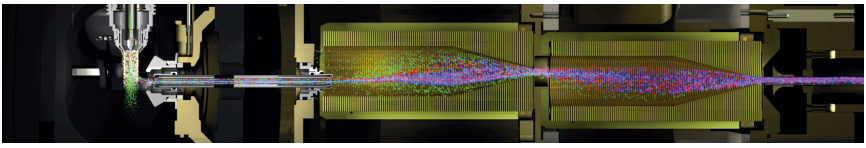


Figura 4.- Componentes de la tecnología iFunnel de un espectrómetro de masas Cuadrupolo-Tiempo de Vuelo (QTOF): Agilent Jet Stream, capilar de muestreo con 6 orificios y doble embudo de iones (desalineados).

Zona 2.-FOCALIZACIÓN de IONES: en esta zona se pueden diferenciar 2 fases.

Descripción Primera Fase: los analitos ya ionizados y desolvatados, se transmiten y focalizan hacia el interior del espectrómetro. Para separarlos del gas de secado, los iones empiezan a circular en zonas que cada vez están a mayor vacío (excepto en la celda colisión). En buena parte de los equipos, como en el representado en la figura 3, la entrada de los iones al interior del espectrómetro, se realiza mediante un capilar dieléctrico con un fino conducto en su interior; por él circulan los iones y el gas de secado aspirado por el gradiente de vacío. A la salida del capilar, los iones y el gas de secado se encuentran una zona de expansión; en ésta:

- El gas (sin carga) difunde y empieza a ser eliminado por el sistema de vacío y con ayuda del cono ("skimmer" muy próximo a la salida del capilar), y de las diferentes barreras de vacío que le siguen (2ª fase).

- Los iones (con carga) continúan con una mínima dispersión, debido al gradiente axial de voltaje que existe entre la salida del capilar y el cono (se le denomina voltaje de fragmen-

tador o de cono); ésta mínima dispersión permite que la mayoría de los iones atraviese el orificio del cono y pasen a los siguientes dispositivos de focalización.

- **Disociación Inducida por Colisión** (CID): si el voltaje de fragmentador/cono se aumenta suficientemente, los iones fragmentan al acelerarse en exceso e impactar violentamente con las moléculas de gas de secado; ello permite obtener un "pseudo MS/MS" con un TOF o un "pseudo MS/MS/MS" con un QTOF.

Últimos desarrollos: para poder captar una mayor cantidad de iones y mejorar la sensibilidad, en algunos de los últimos diseños se ha optado por incrementar el diámetro del orificio del capilar, o utilizar un capilar con múltiples conductos de entrada (tecnología "i-Funnel"). Éste tipo de diseños, incrementa muy considerablemente la cantidad de gas de secado captado, y requiere que el tradicional esquema de conos y barreras de vacío, sea sustituido por un sistema de doble embudo de iones (ver figura 4). En este sistema, el primer embudo se encarga de eliminar la gran cantidad de gas de secado y moléculas neutras que le llegan, mientras el segundo embudo focaliza los iones hacia la primera guía de iones.

Descripción Embudo de Iones: es un dispositivo que va reduciendo su diámetro, y al que se aplican 2 tipos de campos eléctricos; uno uno aplicado radialmente de corriente alterna que al ir cambiando de polaridad va focalizando los iones hacia el centro del embudo, y otro de corriente continua que empuja a los iones hacia la salida (se aplica longitudinalmente).

El uso de un capilar de entrada, y consecuentemente la gradual pérdida de presión, tiene la ventaja de reducir la formación de aductos del analito con el disolvente, pero no todos los diseños lo utilizan.

Diseños alternativos sin capilar de entrada: en otros equipos los iones en lugar de entrar a través de capilar lo hacen a través de un:

- Opción A: **Orificio frontal en forma cónica**, por el que los iones entran a una cámara contigua con una cortina de gas, que al atravesarla puede dar acceso a dos tipos de diseños distintos:

  - Diseño A1: un cono ("skimmer") después del cual los iones van a parar a una guía de iones cuadrupolar, antes de entrar en la siguiente zona donde ya se encontrarán la celda de colisión.

  - Diseño A2 (ofrece mayor sensibilidad): Un segundo orificio de entrada, a cuya salida el gas y los iones se expanden; los iones se capturan mediante el uso de una primera guía de iones cuadrupolar, que los capta para separarlos del gas. Una segunda guía cuadrupolar, focaliza y acondiciona a los iones para entrar en la siguiente zona (donde está la celda de colisión).

- Opción B: **Cono de muestra situado en la propia fuente**; éste da paso a un cono de extracción situado en una **cámara contigua**, después de la cual ya se accede a una guía de iones hexapolar.

Descripción Segunda Fase: una vez los iones ya se han separado mayoritariamente del gas de secado (primera fase), los iones han de atravesar diversos compartimentos con cada vez un vacío más elevado; éstos se utilizan para alcanzar el elevado vacío que se requiere en un espectrómetro de masas. Para ello, se utilizan unas guías de iones y lentes de focalización, que tienen por objeto mantenerlos focalizados para proporcionar un elevado grado de transmisión, y que los iones entren en condiciones adecuadas de energía al interior del cuadrupolo (en el caso en los QTOF). Estas guías de iones suelen ser octopolos, hexapolos o cuadrupolos a los

que se aplica una radiofrecuencia pura (RF: corriente alterna sin componente de corriente continua) que opera a varios MHz de frecuencia (la polaridad cambia varios millones de veces por segundo). Esta RF focaliza a los iones en su interior, y proporciona un elevado grado de transmisión en un amplio rango de masas (especialmente con octopolos); en la figura 5 se puede observar un octopolo (aunque el de la figura pertenece a una zona de focalización posterior).

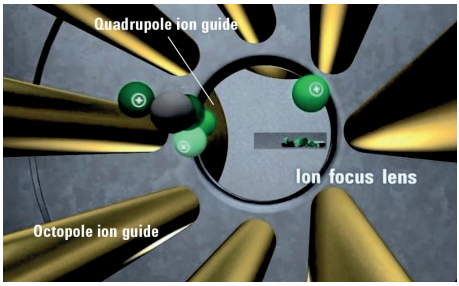


Figura 5.- Visualización desde la parte posterior del octopolo situado después de la celda de colisión en dirección hacia el tubo de vuelo. En los Q-TOF, la óptica de iones situada después de la celda de colisión ayuda a eliminar el gas de colisión, a dar una forma adecuada al haz de iones, y a eliminar los iones con una velocidad vertical no uniforme; todo ello mejora la precisión y resolución de masas del espectrómetro.

Zona 3.- SELECCIÓN y FRAGMENTACIÓN de IONES PRECURSORES (sólo QTOF): esta zona está formada por un cuadrupolo y por una celda de colisión, que pueden operar en diversos modos según el tipo de información que se quiera obtener.

Modalidades Operación Cuadrupolo:

- Modalidad QTOF-MS (QTOF utilizado como TOF): el cuadrupolo opera como una guía de iones, transmitiendo todos los iones que le entran sin filtrarlos por su masa. Es la modalidad que se utiliza cuando el QTOF se quiere utilizar en modo MS, sin hacer MS/MS; es decir operando como un TOF. En esta modalidad al cuadrupolo sólo se le aplica una radiofrecuencia pura (RF); es decir corriente alterna sin componente de corriente continua.

- Modalidad QTOF-MS/MS: el cuadrupolo sólo deja pasar iones con una determinada relación masa/carga (*m/z*); la del ión precursor que interese. El resto de iones se descarga al tocar las barras ("rods") del cuadrupolo, siendo eliminados por el sistema de vacío. Es la modalidad utilizada cuando el QTOF opera en modo MS/MS. En este modo al cuadrupolo se le aplica una RF que incorpora una componente de corriente continua; consecuentemente, una polaridad predominará sobre la otra en función de esta componente de corriente continua. Así se consigue que sólo iones de una determinada *m/z*, tengan una trayectoria oscilante pero estable (sin tocar las paredes del cuadrupolo). La *m/z* que es capaz de atravesarlo, se define por el ratio entre la amplitud del voltaje de corriente alterna aplicada, y la del voltaje de corriente continua utilizado. Las dos barras opuestas ("rods") a las que se aplica una componente positiva de corriente continua, se encargarán de filtrar las masas bajas (por debajo de la estabilizada); las otras 2 barras, a las que se aplica una corriente continua negativa, serán las encargadas de descargar eléctricamente los iones cuyas masas están por encima de la *m/z* seleccionada como precursora.

Tipos de Celda de Colisión: la mayoría de las celdas de colisión son cuadrupolares o hexapolares; las tradicionales son un cuadrupolo (de ahí el nombre de triple cuadrupolo), la tendencia es a utilizar hexapolos (como el equipo de la figura 3) por su mayor rango de transmisión de masas. También las hay que consisten en un conjunto de más de 100 electrodos en forma de anillo, a los que alternativamente se les aplica una RF con componente de corriente continua positiva y negativa. Otras diferencias entre los distintos diseños, se encuentran en las presiones a las que operan (las hay de "alta" y "baja" presión), en el gas que utilizan (nitrógeno o argón), y en el sistema que utilizan para eliminar el "efecto memoria" ("cross talk").

Modalidades Operación de la Celda de Colisión:

- **Modalidad QTOF-MS** (QTOF utilizado como TOF)**:** la celda opera como una simple guía de iones, transmitiendo todos los iones que le entran sin fragmentarlos. En este modo, los parámetros de la celda se ajustan para que todos los iones de todas las masas salgan con una mínima y muy uniforme energía ("ion cooling").

- **Modalidad QTOF-MS/MS:** la celda opera como una simple guía de iones transmitiendo todos los iones que le entran, pero en unas condiciones en las que éstos se aceleran suficientemente, para que su impacto con el gas de colisión, sea lo suficientemente violento para fragmentar al ión precursor seleccionado en el cuadrupolo previo.

- **Proceso de Fragmentación:** tienen lugar cuando la energía vibracional que recibe el ión (por el impacto con el gas de colisión), supera la energía de sus enlaces más lábiles. El tipo de roturas que se generan son mayoritariamente del tipo heterolítico (rotura de enlaces entre átomos distintos); éstas requieren menos energía que las de tipo homolítico (p.e. C-C) muy típicas en GC/MS con impacto electrónico. En las roturas en la celda de colisión, los 2 electrones del enlace covalente roto se suelen quedar en uno de los 2 fragmentos, generando un ión y un fragmento neutro, ambos con un número par de electrones; al igual que el ión precursor. En contadas ocasiones se pueden obtener también iones producto con un número impar de electrones. Esto es importante tenerlo en cuenta a la hora de calcular fórmulas moleculares; para una molécula de 600Da, no considerar el peso de 1 electrón significa un error de 1ppm.

- **Diseños Eliminación Efecto Memoria ("cross talk") en MS/MS:** el procedimiento utilizado para acelerar los iones, y vaciar rápidamente la celda de colisión de restos de iones de la transición anterior, varía de unos equipos a otros; un par de diseños actuales:

  - Diseño A: **Hexapolos con Aceleración Axial Lineal.** En éste tipo de diseño (el de la figura 3), las barras del hexapolo de la celda de colisión están recubiertas con un material semiconductor. Al aplicarse una diferencia de voltaje continuo entre sus extremos, se generará a lo largo de las barras, un gradiente continuo de voltaje en dirección axial, muy próximo a la zona de circulación de los iones; éste facilita un muy rápido vaciado de los iones (<1ms), eliminando el fenómeno de "cross talk".

  - Diseño B: **Onda de Pulsos de Voltaje Continuo** que se desplazan a través de la celda de colisión, constituida por un conjunto de más de un centenar electrodos en forma de finos anillos (arandelas), separados entre sí por un milímetro; a éstos, se les aplica una radiofrecuencia, alternativamente con componente de corriente continua positiva y negativa. Estos voltajes se aplican formando una onda de voltaje "viajera", que se va desplazando axialmente a lo largo del dispositivo, para arrastrar axialmente a los iones a través de la celda.

Zona 4.-FOCALIZACIÓN de IONES: esta segunda zona de focalización de iones, situada en el caso de los QTOF después de la celda de colisión, prepara los iones para su adecuada introducción en el tubo de vuelo. En el caso de los TOF, el cuadrupolo y celda de colisión de la zona 3 (figura 3) desaparecerían, y los 2 octopolos de las zonas 2 y 4 quedarían separados por una simple barrera de vacío. En el diseño de la figura 3, los diversos componentes de esta zona son:

Guía Octopolar de Iones: situada después de la celda de colisión, tiene por objeto mantener focalizados los iones mientras se elimina el gas de colisión que sale de la celda de colisión, para llegar al elevado vacío que se necesita en el tubo de vuelo (10^{-7} Torr). En la figura 5 se visualiza este octopolo en dirección hacia el tubo de vuelo.

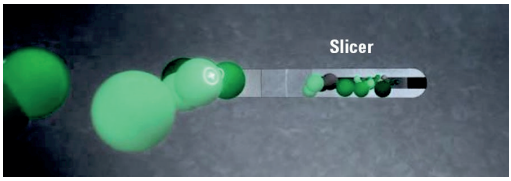


Figura 6.- Visualización de la zona de rendijas horizontales situadas justo antes del tubo de vuelo.

Lente de Focalización de Iones ("Focus lens" en la figura 3)**:** es una lenta plana redonda, a la que se aplica un voltaje de corriente continua, con el objetivo de reacelerar a los iones. Se encuentra situada a la salida del octopolo (ver figura 5)

Guía Cuadrupolar de Iones ("DC quad" en la figura 3)**:** situada a la salida de la lente anterior, reduce la velocidad de los iones en el eje vertical, para disminuir su dispersión vertical y conseguir un haz de iones "plano", que adapte su perfil al de las rendijas situadas después de esta guía (ver figura 5 –justo detrás del "Ion Focus" y antes de la rendija).

Zona de Rendijas Horizontales ("Slits" en la figura 3)**:** está constituida por una serie de rendijas rectangulares en configuración horizontal, cuyo objetivo es eliminar, antes de entrar en el tubo de vuelo, aquellos iones que tienen un exceso de dispersión/velocidad en el eje vertical. Mejora la exactitud y especialmente la resolución de masa del tiempo de vuelo (ver figuras 5 y 6).

Diseños Alternativos:
Compresor del Haz de Iones (IBC): situado a la salida de la celda de colisión, mediante el uso de una guía de iones hexapolar convergente (ver figura 7), consigue enfriar energéticamente y reducir 10 veces la sección del haz de iones, proporcionando una muy importante mejora en la resolución espectral sin sacrificar sensibilidad. La compresión se consigue al ir reduciéndose la sección de la guía a medida que los iones van avanzando en su interior. Después unas lentes y guías optimizan las condiciones de entrada de los iones a la zona de rendijas situadas justo antes del tubo de vuelo. Este dispositivo, conjuntamente con otras mejoras en el tubo de vuelo, permiten doblar la resolución espectral dando lugar a los denominados UHR-QTOF (QTOF de Ultra Alta Resolución). En la figura 3 este dispositivo sustituirá a los 3 primeros componentes descritos de la zona 4; se encuentra entre la celda de colisión y la zona de rendijas.

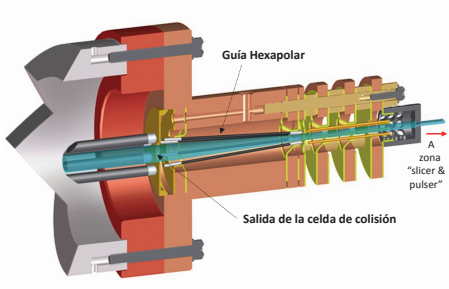


Figura 7.- Tecnología de Compresión del Haz de Iones (Agilent IBC): a la salida de la celda de colisión (zona izquierda figura), una guía de iones hexapolar convergente (zona central figura), enfría y reduce 10 veces la sección del haz de iones, proporcionando una muy importante mejora en resolución espectral sin sacrificar sensibilidad.

Zona 5.-Tubo de Vuelo - MEDIDA *m/z* y DETECCIÓN IONES: esta **última zona m**ide la masa exacta y cantidad de iones que entran en ella; opera a un vacío muy elevado (10^{-7} Torrs), para evitar que el impacto de los iones con moléculas de nitrógeno interfiera en la medida de su masa exacta.

Influencia de la Temperatura del Laboratorio: una mínima dilatación del tubo de vuelo afectar**á** a la distancia que los iones han de recorrer y consecuentemente a su tiempo de vuelo; para disponer de una masa exacta robusta, será muy importante que los equipos utilizados incorporen medidas para reducir el impacto de cambios en la temperatura del laboratorio en la masa exacta medida:

- Aleaciones especiales con mínimos coeficientes de expansión térmica.
- Tubo de vuelo con un muy buen aislamiento térmico del exterior.
- Control de la temperatura del tubo de vuelo y su electrónica.

En estos puntos hay bastantes diferencias entre fabricantes. Dependiendo del instrumento que se disponga, será más o menos importante trabajar en un laboratorio con muy buen control de su temperatura; especialmente cuando se trabaje en modo de patrón externo, es decir, sin adición en continuo de masas exactas conocidas para recalibrar internamente con ellas, en continuo/espectro a espectro. Sin esta adición con algunos modelos se deberán efectuar frecuentes recalibraciones del eje de masas, para corregir los efectos de cambios de la temperatura del laboratorio.

Los dispositivos incorporados en el tubo de vuelo son:

Pulsador de Iones: localizado a la entrada del tubo de vuelo, acelera ortogonalmente a los iones (procedentes de las rendijas de la zona anterior) en dirección al espejo de iones. Esta aceleración se realiza mediante la aplicación de un pulso de elevado voltaje (varios Kv) en la placa inferior del dispositivo (ver figura 8); éste repele a los iones de la misma polaridad y los acelera ortogonalmente, para pasar a través de los orificios de las placas superiores que definen el ancho del haz de iones que se envían a volar (ver parte derecha en figura 3). El pulso en cuestión se va repitiendo, dependiendo del rango de masa de interés, con una frecuencia superior a 10KHz (más de 10.000 veces por segundo). Cada uno de estos pulsos genera el vuelo del "paquete de iones" que en ese momento están pasando por el pulsador, y que al ir llegando al detector a distinto tiempo, dará lugar a un registro denominado "transient" ("mini espectro" super-rápido construido con una muy pequeña cantidad de iones). **La suma de unos miles de "transients" generará un espectro de masas**; el número dependerá de la frecuencia de adquisición y masa máxima del calibrado.

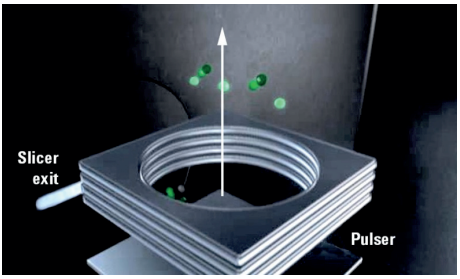


Figura 8.- Pulsador de Iones: la aplicación de un pulso de elevado voltaje en la placa inferior del dispositivo repele a los iones de la misma polaridad y los acelera ortogonalmente (en dirección al tubo de vuelo).

Tubo de vuelo: las zonas donde los iones vuelan libremente, se encuentran libres de campos eléctricos. La longitud del tubo de vuelo varía dependiendo de los modelos (de 1 a 2.7 metros), distancia que se amplificará mediante el uso de espejos de iones. En general un recorrido mayor incrementará la resolución, pero en algunos sacrifica sensibilidad.

Espejo de Iones (Reflectrón): ésta es una pieza clave para que un tiempo de vuelo pueda proporcionar alta resolución y masa exacta. Sin él, los tiempos de vuelo son equipos de baja resolución y masa nominal, al reducirse su resolución en un orden de magnitud, pero pueden ofrece un mayor rango de masa. El objetivo del espejo de iones es doblar el camino óptico y, sobre todo, compensar las diferencias en energía cinética con la que llegan los iones de la misma masa (p.e. por entrar a distinta distancia de la superficie del pulsador de iones que los envía a volar). El espejo iónico se genera en la zona final del tubo de vuelo mediante un campo electrostático uniforme (E), que repele/refleja los iones que entran en el dispositivo (ver parte derecha superior en figura 3). Los iones de la misma masa que llegan con mayor energía cinética/velocidad, recorrerán una mayor trayectoria en el interior del reflectrón; la distancia de penetración en el reflectrón (d) será proporcional a su energía cinética (K) de acuerdo con la fórmula:

$$d = K / q E$$

Ajustando adecuadamente los parámetros del reflectrón, se podrá compensar la diferencia de energía cinética entre iones con la misma masa, aumentando dramáticamente la resolución espectral y exactitud de masa del tiempo de vuelo. Dependiendo de los fabricantes y tipos de equipos, existen diversos esquemas y posibilidades de uso de los reflectores de iones:

- **Modo V:** es el más ampliamente utilizado por su mejor sensibilidad. En este modo los iones son reflejados 1 vez.

- **Modo W:** en este modo los iones son reflejados 2 veces a fin y efecto de incrementar el recorrido de los iones y consecuentemente su resolución. Este diseño W es utilizado para conseguir resoluciones de 20-40K por algún fabricante cuyo diseño en el modo en V proporciona resoluciones espectrales apreciablemente inferiores a las de otras firmas. El inconveniente del modo W es que reduce de 3 a 5 veces la sensibilidad con respecto a la operación en modo V.

- **Modo Lineal:** en el caso de equipos dedicados de MALDI-TOF, para poder trabajar con proteínas intactas grandes (típico >20-50KDa), ofrecen la posibilidad de trabajar en modo lineal (sin utilizar el reflectrón), y poder así detectar proteínas monocargadas de elevado peso molecular. En el caso de LC/MS, al ionizarse las proteínas con múltiples cargas ésta opción no es necesaria y no suele ofrecerse. En electrospray las envolventes de la distribución de iones con múltiples cargas de proteínas, incluso de 150KDa, como los anticuerpos monoclonales, suelen estar por debajo de *m/z* 4000, y no requieren utilizar el modo lineal, como sería necesario en el caso de ionizarlas con una fuente MALDI.

Detector: en los sistemas de tiempo de vuelo el sistema de detección más ampliamente utilizado, es el basado en un "microchannel plate" (MCP) por su elevado nivel de amplificación de la señal (superior a 1 millón de veces) y bajo ruido. Consiste en una fina placa perforada por muchos precisos tubos microscópicos (microcanales); cuando un ión con suficiente energía lo golpea se liberan varios electrones (ver figura 9 y parte inferior derecha de figura 3). Cada microcanal actúa como un multiplicador de electrones; a "groso modo" por cada ión que impacta en el MCP se generan del orden de 10 electrones. Estos electrones son acelerados por un campo eléctrico hacia un centelleador, donde al impactar generan fotones que mediante el uso de lentes ópticas son focalizados en un tubo fotomultiplicador, que generará una señal eléctrica proporcional al número de fotones recibidos, y consecuentemente, al número de iones que inicialmente impactaron en el MCP. Este sistema de detección tiene una larga vida, un rango dinámico superior a 4 órdenes de magnitud, y una ganancia (capacidad de amplificación de la señal) de varios millones.

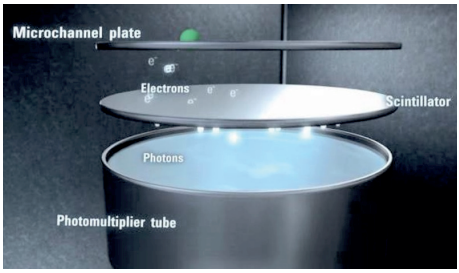


Figura 9.- Detector de Iones: los iones al impactar en la placa de microcanales generan una cantidad proporcional de electrones; éstos al impactar en el centelleador generan proporcionalmente una señal lumínica (fotones) que se recoge en un fotomultiplicador. Este dispositivo genera una señal amplificada varios millones de veces, proporcional al número de iones que entran en el detector.

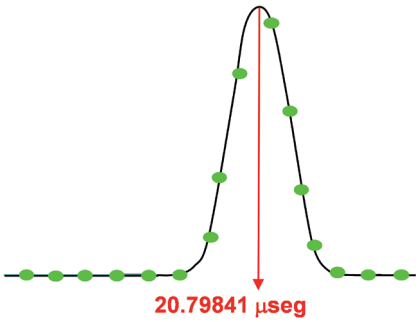


Figura 10.- Medida del tiempo de vuelo de un pico de masa: un avanzado algoritmo ajusta a los puntos de muestreo (puntos verdes) una función continua que permite determinar con gran precisión el máximo del pico de masa.

Digitalización de la señal y obtención del espectro de masas: la siguiente etapa del

proceso de detección, es digitalizar la señal que se obtiene en el detector por los impactos de los iones. Cada uno de los pulsos generados en el pulsador iniciará el vuelo del "paquete de iones" (que en ese momento está pasando por el pulsador), a la vez que inicia el registro en el tiempo de los distintos iones que van llegando al detector. Este registro de la señal de llegada de iones a lo largo del tiempo entre 2 pulsos, recibe el nombre de "transient". La suma de n "transients" generará un espectro de masas; usualmente se suman varios miles de "transients" para generar 1 espectro de MS o MS/MS con masa exacta. Una vez sumados los n "transients", un avanzado algoritmo ajusta, a los puntos de muestreo (ver figura 10), una función continua que permite determinar con gran precisión el máximo de los picos de masa. Existen 2 tipos de tecnologías utilizadas para el registro de un "transient":

- Tecnología ADC ("Analog to Digital Converter"): registra la intensidad de señal a lo largo del tiempo que dura cada "transient" (como si de un registrador se tratará), para ello se monitoriza ésta intensidad de señal cada x picosegundos; habitualmente entre 250 ps y 1ns (4 y 1GHz), es decir entre 1 y 4 billones de veces por segundo. La intensidad de cada medida se digitaliza convirtiéndola en un valor digital representado por 8 bits (0 a 255 cuentas), proporcional al número de iones que impactan en un instante dado en el detector; los equipos más avanzados son hoy en día capaces de proporcionar una muy elevada capacidad de muestreo de la señal, pudiendo llegar a recoger más de 30Gbits/s de información espectral. Para evitar que al ir sumando "transients" se vaya también sumando ruido, es importante que el sintonizado del equipo ajuste adecuadamente el offset del amplificador, para que ya registre la llegada de 1 sólo ión, pero que en cambio, no dé ninguna señal cuando no está llegando ningún ión. Éste es el tipo de digitalizador empleado por el equipo de la figura, y el más utilizado en la actualidad.

- Tecnología TDC ("Time to Digital Converter"): registra directamente el tiempo al que se producen impactos de iones en el detector a lo largo de un "transient", pero no es capaz de diferenciar si el impacto ha sido producido por un único ión, o por varios que han llegado al mismo tiempo; por ello estos sistemas tienen:

  - Un rango dinámico limitado.
  - Una apreciable dependencia de la masa medida con respecto a la concentración del analito.
  - Dificultad en proporcionar ratios de perfiles isotópicos correctos.

  Una descripción detallada del último desarrollo con esta tecnología puede encontrarse en un muy reciente artículo de Andrews, Simons, Muddiman: (16: Performance Characteristics of a New Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometer). El sistema TDC se basa en utilizar un discriminador que cuando la intensidad de señal del detector supera un umbral (por el impacto de 1 ión), dispara el registro del tiempo de ese impacto. Si en 1 "transient" llegan varios iones de la misma masa, la pequeña dispersión en el tiempo de llegada, hace que se registre un tiempo de vuelo distinto al que se obtendría si sólo hubiera llegado 1 único ión de una masa; generando así un cambio de la masa medida con la concentración del analito, y proporcionando una masa exacta mucho menos robusta que el sistema anterior (ADC). Utilizando un MCP ("microchannel plate") con varios ánodos, conectados cada uno a un TDC independiente, se puede mejorar algo su rango dinámico. Al registrar directamente el tiempo de vuelo de los iones, el sistema TDC requiere utilizar una electrónica de muestreo de la señal, mucho más rápida que la necesaria en los sistemas ADC, para obtener un nivel de resolución espectral similar; por ejemplo (p.e.), a m/z ~ 900 un TDC multicanal de 40GHz es capaz de proporcionar >35.000 de resolución, mientras que un ADC a 4GHz es capaz de proporcionar >40.000. Hasta hace unos pocos años, el sistema TDC fue el más utilizado, y los TOF y QTOF eran equipos que tenían importante limitaciones cuantitativas, y una masa exacta muy poco robusta. Incluso en el último desarrollo con esta tecnología, la

exactitud de masa continua siendo poco robusta; cambios de concentración inferiores a 2 órdenes de magnitud (<100 veces) entre el principio y el máximo de un pico, proporcionan desviaciones de varias ppm (3-6ppm) (16). Por todo ello, hoy en día la tendencia es utilizar sistemas de digitalización del tipo ADC.

Si interesa una descripción más detallada de estos puntos, se puede consultar las publicaciones J.Fjeldsted, "Time-of-Flight Mass Spectrometry Technical Overview" (14), e "Ion optics innovations for increased sensitivity in hybrid MS systems" (15).

2.2. Acoplamientos Cromatografía Líquida/ Tiempo de Vuelo: LC/MS-TOF.

Los sistemas de tiempo de vuelo se dividen en 2 grandes categorías, dependiendo de su capacidad de realizar MS/MS:

- **LC/MS-TOF:** son equipos de masa exacta que NO permiten realizar MS/MS, sólo proporcionan espectros de MS. El esquema de uno de estos equipos correspondería al de la figura 3, eliminando la zona central –zona 3- donde están el cuadrupolo y la celda de colisión.

- **LC/MS-QTOF:** son equipos de masa exacta (como el de la figura 3) que sí permiten realizar MS/MS, además de proporcionar espectros de MS (operando como un TOF).

La ventaja de los LC/MS-TOF es su menor coste. Para tareas y aplicaciones que NO requieren MS/MS, es un equipo que da unas excelentes prestaciones y posibilidades a un coste muy competitivo; entre otras se podrían citar:

- **Análisis comparativos** de muestras para descubrimiento de Biomarcadores

- **Cribados masivos** sobre la presencia de listados de centenares/miles de compuestos (basta conocer su fórmula molecular).

- **Cuantificación mediante masa exacta de cualquier compuesto ionizado**.

- **Confirmación de la identificación** de un compuesto mediante el cálculo de su fórmula molecular, a partir de su masa exacta y conformidad de su perfil isotópico. Si interesa confirmar la identificación utilizando información estructural, se necesita disponer de separación cromatográfica del compuesto de interés; así se podrá fragmentar la molécula sola, en la zona del cono (ver zona 2 – CID figura 3) y obtener su "pseudo MS/MS". Ello permitirá obtener las fórmulas moleculares de fragmentos, que ayudaran a corroborar la identificación de la molécula. Si la columna no ha conseguido aislar el compuesto (buena separación cromatográfica), se necesitaría MS/MS para aislarlo.

- **Identificar proteínas** mediante búsquedas PMF "Peptide Mass Fingerprint Data", al sólo considerar la masa exacta del péptido, ésta es una técnica mucho menos potente que las búsquedas utilizando MS/MS, que también tendrían en cuenta la secuencia de aminoácidos de los péptidos.

El desarrollo de estas capacidades aplicativas será descrito en un apartado posterior centrado en aplicaciones.

2.3. Acoplamientos Cromatografía Líquida/ Cuadrupolo-Tiempo de Vuelo: LC/MS-QTOF.

En la figura 3 se muestra el esquema de un LC/MS-QTOF. Su capacidad de proporcionar

espectros de MS/MS con masa exacta, es una valiosísima herramienta en tareas identificativas. En la figura 11 se representa el proceso de MS/MS en un LC/MS-QTOF. Aunque la cromatografía simplifica la muestra que le va llegando al espectrómetro, en muestras complejas, de la columna cromatográfica suelen coeluir diversos compuestos que se ionizan conjuntamente en la fuente (ver espectro de MS en lado izquierdo). Al disponer de un cuadrupolo, éste puede actuar como filtro de masa y sólo dejar pasar, durante un determinado tiempo definido por el usuario (p.e. 300 ms), la masa *m/z* 374.7 (correspondiente al péptido ALELFR con 2 cargas+); éste fragmenta en la celda de colisión formando los iones producto 564.3140 (ELFR con 1 sola carga) y 157.1335, entre otros varios. Estos iones productos darán información de la secuencia de aminoácidos del péptido. Transcurridos esos 300ms, el equipo puede ir aislando sucesivamente otros iones precursores para fragmentar otros péptidos que coeluyan en el mismo pico cromatográfico. Esta programación de iones precursores la puede decidir el propio equipo (Auto MS/MS o "Data Dependent MS/MS"), o ser programada en el tiempo por el propio usuario ("Targeted MS/MS"). Para más detalles ver sección 4.4.

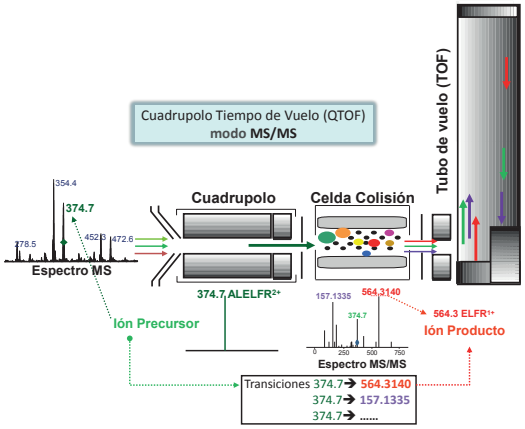


Figura 11.- Proceso MS/MS en un QTOF: de los diversos compuestos que coeluyen de la columna cromatográfica (ver espectro de MS en lado izquierdo), el cuadrupolo sólo dejar pasar el péptido con doble carga de m/z 374.7 (ALELFR); éste fragmenta en la celda de colisión formando los iones producto 564.3140 (ELFR con 1 sola carga) y 157.1335, entre otros. La m/z exacta de todos los iones producto se mide en el tubo de vuelo.

Otra alternativa es fragmentar a todos los iones sin aislarlos o MS/MS Independiente de los Datos ("MS/MS All ions"); en esta modalidad el cuadrupolo transmite todos los iones sin filtrarlos, y la celda de colisión fragmenta todos los iones (para más detalles ver apartado 5.3).

Fragmentación de iones precursores con múltiples cargas: como se puede observar en el espectro de masas de iones producto (espectro MS/MS) de la figura, algunos fragmentos tienen una m/z superior a la del precursor (374.7). Ello es habitual al fragmentar precursores múltiplemente cargados. Éstos cuando dan fragmentos no mucho más pequeños, pero que en cambio si reducen su número de cargas, es fácil que generen fragmentos a mayor m/z que la del precursor.

Además de las posibilidades aplicativas descritas para los TOF, la capacidad MS/MS de los QTOF les permite proporcionar información sobre la estructura molecular de los compuestos analizados, información muy valiosa para poder abordar su identificación:

- **Identificar complejas mezclas de proteínas digeridas:** el espectro de MS/MS con masa exacta proporcionará información de que aminoácidos están presentes (masa exacta precursor) y en qué orden (MS/MS: masas exactas de los fragmentos).

- **Identificar compuestos** a partir de las fórmulas moleculares del ión precursor e iones producto, calculadas a partir de su masa exacta y perfil isotópico (ratio de abundancias y distancia en m/z entre señales isotópicas).

2.4. Aportaciones de la Masa Exacta en espectrometría de masas LC/MS y LC/MS/MS en tándem con masa exacta.

La masa exacta proporciona una especificidad y selectividad que resulta muy útil en el análisis

de muestras muy complejas, como es habitual en proteómica. Sus aportaciones las podríamos enumerar:

- **Selectividad "tipo MS/MS" a partir de un cromatograma adquirido en modo MS** (espectro completo)**.** Al adquirir en modo MS (sin hacer MS/MS), en el disco duro se dispone de la información de todos los compuestos ionizados. Para obtener una señal muy específica de un determinado analito ("selectividad tipo MS/MS"), bastará extraer cromatogramas de iones con una muy fina ventana de extracción (unas decenas de mDa o +/-10-30ppm) alrededor de la *m/z* del ión molecular protonado o desprotonado ($M+H^+$ o $M-H^-$), posibles aductos ($M+Na^+$, $M+NH4^+$, $M+HCOO^-$,…), o como en el caso de péptidos y proteínas, de iones con múltiples cargas ($[M+nH^+]/n$). Esta característica como se comentará en el apartado 3.1.1.2, facilitará enormemente el descubrimiento de biomarcadores a nivel de péptidos minoritarios.

- **Aumento del número de proteínas identificadas y reducción del "False Discovery Rate" (FDR).** El disponer de la masa exacta del ión precursor y de los iones productos cuando se trabaja en MS/MS automático (o "Data Dependent MS/MS") proporcionará una muy importante reducción del FDR y consecuentemente un aumento del número de proteínas identificadas con un determinado porcentaje de confianza. Miller (17) lo demuestra en la tabla 2, al aplicar distintas tolerancias de exactitud de masa, al análisis de un lisado digerido de células HeLa S3. El lisado digerido se fraccionó en el rango de pH 3-10, en 23 fracciones con un Agilent 3100 OFFGEL Fractionator. Para evaluar el impacto de la exactitud de masa de precursor y producto, en el nº de proteínas identificadas, y eliminar el factor de la adquisición, se utilizó el mismo conjunto de datos adquiridos. Se evaluaron 3 condiciones emulando 3 diferentes tecnologías de MS.

  En experimentos con una cantidad masiva de datos, será importante definir adecuadamente las tolerancias de masa a utilizar, especialmente cuando también interesen péptidos muy minoritarios, que empeorarán la máxima desviación de masa MDM En este experimento se utilizaron los MDM calculados de acuerdo con Zubarevy y Mann (18).

- **Especificidad en la secuenciación del péptido mediante MS/MS con masa exacta.** En la secuenciación De Novo, el disponer de un espectro de MS/MS del péptido con masa exacta, proporcionará una especificidad adicional, que permitirá reducir la incertidumbre en la asignación de la secuencia de aminoácidos que lo componen.

Tabla 2.- Comparación del impacto de la exactitud de masa de iones precursor y producto, en el número de identificaciones fiables, para tres diferentes combinaciones de exactitud ("emulan" 3 tecnologías distintas). La tabla demuestra la importancia de la masa exacta de precursores y productos, para aumentar el número de proteínas y péptidos identificados con fiabilidad.

Tipo de Iones y Tolerancia m/z		Tipo Tecnol. Emulada	Número de Péptidos Validados con			Número de Proteínas Validadas (mínimo con dos o más péptidos únicos/proteína)		
Ión Precursor	Iones Producto		% confianza					
			>95%	>75%	>50%	>95%	>75%	>50%
1000 ppm	1000 ppm	Trampas de iones	954	2220	3361	102	213	282
5 ppm	1000 ppm	Híbridos con masa exacta lenta	4181	6590	7645	338	478	524
10 ppm	40ppm	QTOF's e híbridos con masa exacta "rápida"	6590	11552	15509	478	728	829

- **Especificidad en la determinación de la fórmula molecular / composición de aminoácidos del péptido.** El conocer con exactitud la masa exacta de un péptido, permitirá conocer sus posibles fórmulas moleculares o los aminoácidos que lo componen (aunque no su orden ni diferenciar entre isómeros como ile y leu). Cuanto mayor sea el tamaño del péptido, mayor será la incertidumbre en estas determinaciones, y más importante será una buena correspondencia entre el perfil isotópico teórico y experimental (ver siguiente apartado 2.5).

Para poder aprovechar todas las prestaciones que puede llegar a ofrecer la masa exacta, será muy importante disponer de un equipo que ofrezca un amplio rango dinámico espectral, o en otras palabras: que la masa exacta medida no cambie con la concentración del analito, y el detector no se sature para poder proporcionar una buena fidelidad isotópica. Éstos son puntos con diferencias muy significativas entre las diversas tecnologías y modelos; los mejores son los que utilizan un sistema de medida del tiempo de vuelo basados en conversores tipo ADC, con doble amplificación de señal; éstos ofrecen un rango dinámico espectral de 10^5 frente al típico 10^2 de los conversores tipo TDC. En un campo como el de la proteómica, en donde los niveles de expresión de las proteínas abarcan 11 órdenes de magnitud (9 si eliminamos las 5-10 mayoritarias), éste será un punto importante a tener en cuenta.

Una masa exacta más robusta permitirá:

- La utilización de ventanas de masa más finas para:
  - La extracción de cromatogramas de iones más selectivos.
  - La identificación/confirmación de desconocidos con mayor especificidad.

- Que los compuestos de referencia para corrección de masa, añadidos en continuo para el cálculo de la masa exacta, estén a concentraciones mucho más bajas que las de los analitos de interés.

2.5. La importancia del perfil isotópico en la identificación de péptidos.

El perfil isotópico (como se indica en capítulo que describe la "Distribución isotópica y carga iónica"), corresponde a la distribución en el espectro de masas, de la abundancia relativa de las señales proporcionadas por isótopos estables como carbono-13, azufre-34, nitrógeno-15, entre otros. **El perfil de la distribución isotópica de una molécula es característico de su composición** (/formula molecular), y constituye una "huella" muy útil para confirmar la identificación de un compuesto por espectrometría de masas. La masa exacta del analito permite calcular las posibles fórmulas moleculares, mientras que la discordancia entre el perfil isotópico teórico y experimental, permitirá descartar muchas de ellas.

La masa exacta de un péptido no es suficiente para definir inequívocamente su fórmula molecular/ composición de aminoácidos. En general se encontraran muchas posibles fórmulas (tantas más cuanto mayor sea el péptido); para reducir las opciones, será clave descartar todas aquellas fórmulas moleculares cuya distribución isotópica teórica no sea muy semejante a la experimental; ello requerirá disponer de un equipo que proporcione una buena fidelidad isotópica. Algunas de las tecnologías TOF y QTOF más actuales (tipo ADC) son capaces de proporcionar una muy buena fidelidad isotópica, a diferencia de algunos modelos de tecnología de alta resolución, que tiene dificultades en proporcionar rutinariamente un perfil de intensidades isotópicas relativas correcto. En el ejemplo de la figura 12, para el péptido HLVDEPQNLIK de 1304,71Da, se obtienen sólo 4 fórmulas moleculares con una concordancia de la abundancia relativa del perfil isotópico, superior al 99% (incluyendo la C58 H96 N16 O18 correspondiente al péptido en cuestión); en cambio, por masa se obtienen +100 posibles fórmulas moleculares, con una concordancia igual o superior al 99,9% (implica un error de masa ≤ 0,1ppm).

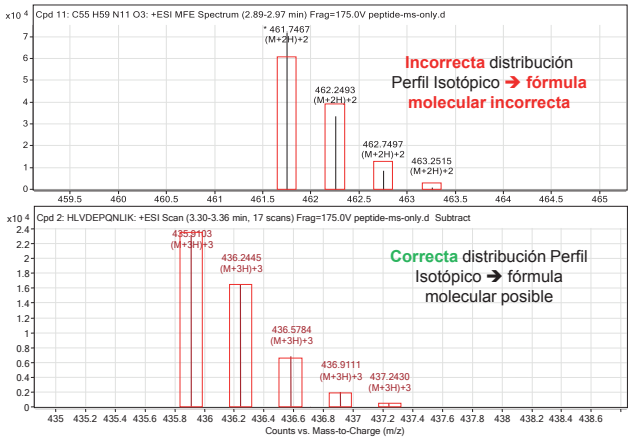


Figura 12.- Fidelidad del perfil isotópico péptido HLVDEPQNLIK de 1304,71Da (con 3 cargas -ver figura inferior): el recuadro en rojo indica el perfil teórico correspondiente a la composición del péptido: C58 H96 N16 O18. La buena correspondencia entre el perfil teórico y experimental ayuda a confirmar la fórmula molecular del péptido, y sobre todo para descartar fórmulas incorrectas (ver ejemplo figura superior).

Las **aportaciones del uso del perfil de abundancias isotópicas relativas** se pueden resumir en:

- **Reducir el número de falsas identificaciones de péptidos**/proteínas propuestos por los motores de búsqueda en bases de datos de proteínas, o secuenciaciones "De novo" de péptidos. Aunque no servirá para diferenciar isómeros (péptidos con la misma fórmula molecular), si se utiliza exclusivamente la información del espectro de MS sin fragmentar.

- **Diferenciar entre isómeros cuando también se dispone del espectro de MS/MS** del péptido, al obtenerse información de la secuencia de aminoácidos (su orden en la cadena). La suma de las fórmulas moleculares de los fragmentos del MS/MS, más la de las correspondientes pérdidas neutras, deberá dar la fórmula del ión precursor; consecuentemente deberán encajar como si de un puzle se tratará. La fórmula molecular del precursor, deberá encajar con TODAS las de sus fragmentos y pérdidas neutras; así se dispondrá de un mayor nº de parámetros para validar la fórmula molecular del péptido propuesto.

Adicionalmente la **distancia en masa/carga (*m/z*) entre las señales isotópicas aportará:**

- **Poder calcular el nº de cargas del ión analizado**; éste es igual al inverso de la distancia en *m/z* entre las señales isotópicas del ión:

$$\text{Nº cargas ión (z)} = 1 / (\Delta \text{ m/z entre isótopos}).$$

- **Diferenciar con respecto a señales de masa procedentes de otros analitos** coeluyentes (requiere disponer de masa exacta)

  Aportar un parámetro adicional para el cálculo de la concordancia del perfil isotópico (abundancias relativas e incremento de masa).

En proteínas intactas de gran tamaño, el elevado nº de cargas de los iones, produce que la distancia en masa entre las diferentes señales isotópicas, pueda llegar a ser muy pequeña. Para resolver su perfil isotópico se requerirán equipos de muy alta resolución espectral; aunque ello

no es necesario para calcular su peso molecular con elevada precisión, mediante la deconvolución de su espectro de MS (ver apartado 3.4.1.-Determinación del peso molecular de péptidos y proteínas intactas).

En resumen, la confirmación de una fórmula molecular se podrá evaluar a partir de:

- Similitud de las intensidades relativas de las señales del perfil isotópico.
- Distancia en masa (*m/z*) entre las señales isotópicas.
- Exactitud de masa del péptido.
- En MS/MS: Compatibilidad de la fórmula molecular del precursor con todas las de sus fragmentos y pérdidas neutras.

Toda esta información isotópica disponible en los espectros de MS y MS/MS de un péptido, permitirá mejorar la fiabilidad y calidad de los resultados reduciendo el nº de falsas identificaciones.

3. Aplicaciones y posicionamiento con respecto a otros analizadores de masa.

El primer objetivo de este extenso apartado, es describir y comentar el uso de TOF y QTOF's en distintos tipos de aplicaciones proteómicas:

- Descubrimiento de Biomarcadores.
- Identificación de Proteínas.
- Proteómica Cuantitativa.
- Caracterización de péptidos y proteínas intactas

El segundo objetivo es posicionar para estas aplicaciones los TOF y QTOF "versus" otras tecnologías. En principio cualquier tipo de espectrómetro podrá ser utilizado para todo tipo de analitos (usando fuentes de ionización intercambiables adecuadas); pero dependiendo del tipo de información que se quiera obtener (tipo de aplicación), sólo algunos de ellos podrán proporcionar la información deseada. Independientemente de su coste, no existe hoy en día ningún espectrómetro de masas que sea el mejor para todo tipo de aplicaciones. Aunque algunos de ellos darán un buen rendimiento en múltiples tipos de aplicaciones; otros serán claramente los mejores en un área de aplicación, pero no serán adecuados para otros fines. Por tanto, será importante definir el tipo de información que se quiere obtener de la muestra, para decidir el tipo de instrumento que mejor resolverá esas necesidades. En la tabla 3 se resume, según el tipo de aplicación, las diversas tecnologías recomendadas; éstas se irán posicionando, más detallada y específicamente, en cada sección aplicativa.

Tabla 3.- Resumen de tecnologías recomendadas según tipo de aplicación.

Aplicación	Mejor Opción	1ª Alternativa	2ª Alternativa
Identificación y Caracterización de Proteínas	MS/MS con masa exacta QTOF, FTICR e híbridos con masa exacta	MSn Trampas de Iones	MS con masa exacta TOF o equivalentes
Cuantificación "Targeted"	QqQ en modalidad MRM	MS o MS/MS (en el espacio) con masa exacta QTOF o TOF	MS/MS en el tiempo Trampas de Iones e híbridos con masa exacta
Cuantificación "UNtargeted"	MS/MS (en el espacio) con masa exacta QTOF	MS o MS/MS (en el tiempo) con masa exacta TOF, FTICR e híbridos con masa exacta	MSn Trampas de Iones

De todos los tipos de espectrómetros de masa, bajo una perspectiva global de aplicación proteómica, probablemente los QTOF de gama alta (resolución 40k) son de los instrumentos, que en promedio, mejor permiten cubrir el amplio rango de tipos de aplicaciones actuales de la proteómica (ver figura 13 y tabla 3). Ello no quiere decir que sean los mejores en cada uno de los campos; por ejemplo en proteómica cuantitativa de péptidos/proteínas concretos, sin duda los triple cuadrupolos proporcionarán mejores prestaciones.

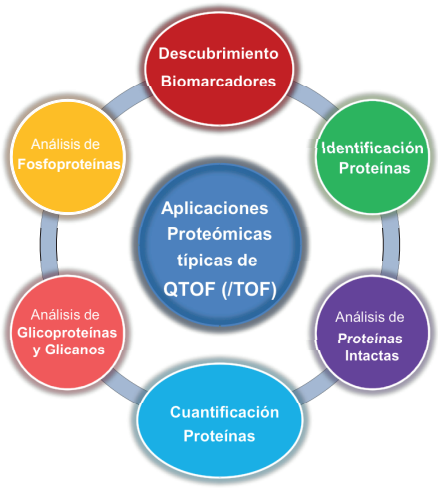


Figura 13.- Típicas aplicaciones proteómicas de LC/MS-QTOF (/TOF): el descubrimiento de biomarcadores e identificación y cuantificación de péptidos y proteínas, son 3 de las más importantes aplicaciones de LC/MS con QTOF; aunque no las únicas.

Otra de las ventajas que proporciona la tecnología QTOF (en modelos de algunas firmas), es la de ser idénticos a un triple cuadrupolo (QqQ) hasta la celda de colisión. Esto facilita considerablemente la transferencia de parámetros de trabajo y datos, al pasar de una proteómica identificativa o de caracterización (QTOF), a proteómica cuantitativa de proteínas/péptidos concretos. Un buen ejemplo de ello es el uso de esta estrategia para la creación del SRM/MRM Atlas Humano, por parte del grupo de R. Moritz (19: "Development of a Complete Human Peptide and SRMAtlas"). Primero utilizan un QTOF para obtener los espectros de MS/MS de alta resolución y masa exacta de los péptidos de interés, mediante el uso de listas de inclusión. Posteriormente, la información del espectro de iones producto (MS/MS) del péptido es utilizada para crear la lista de las mejores transiciones, para cuantificar en un QqQ que hasta la celda de colisión es idéntico al QTOF utilizado (Agilent 6460-QqQ y 6530-QTOF). En el mercado existen motores de búsqueda (ver figura 14) que permiten:

- A partir de las proteínas identificadas (p.e. mediante QTOF), exportar la lista de transiciones MRM programadas en el tiempo, para cuantificar con QqQ los péptidos de las proteínas identificadas.

- A partir de la digestión teórica de las proteínas o secuencias de interés:

  - Seleccionar aquellos péptidos que cumplen unos criterios específicos de filtrado (pe. únicos, fosforilados,..)
  - Proporcionar un listado de iones producto, remarcando aquellos que teóricamente proporcionaran intensas señales de MS/MS.

Además de tener en común buena parte de sus componentes, actualmente los QTOF de la gama más alta, dan un nivel de sensibilidad del mismo orden que triple cuadrupolos de gama media en modalidad MRM. Ello ha sido posible gracias a la aplicación de los últimos avances en tecnología de ionización (como "Agilent Jet Stream"), y captación de iones ("i-funnel" con capilar de múltiples orificios para captar más iones); aplicada a QTOF permite que estos instrumentos de ultimísima generación (presentados en el mercado a mediados del 2011) superen en sensibilidad a QqQ de gama media en MRM (/SRM). Ello ha llevado a AB Sciex a llamar TripleTOF™ a su QTOF de gama alta.

3.1. Descubrimiento de Biomarcadores.

Sin duda, el descubrimiento y validación de nuevos biomarcadores es uno de los objetivos prioritarios de la investigación biomédica actual, en su evolución hacia una medicina predictiva,

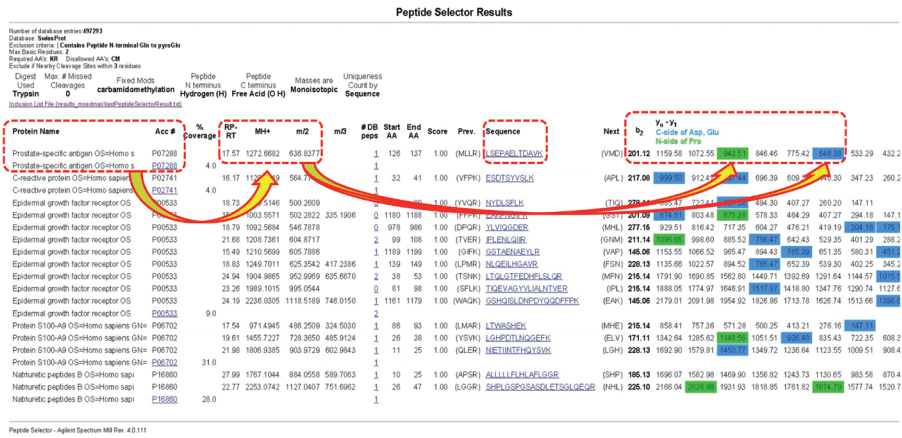


Figura 14.- Software selector de Péptidos y transiciones MRM del motor de búsqueda Spectrum Mill: permite seleccionar los péptidos y sus mejores transiciones a iones producto de las proteínas identificadas o de un listado de interés, para su óptima cuantificación mediante QTOF, o QqQ en modalidad MRM.

preventiva, personalizada y participativa. Esta evolución será posible gracias al descubrimiento de nuevos biomarcadores que permitan:

- El diagnóstico precoz y/o más preciso de las enfermedades.
- Clasificar pacientes: prever si un determinado tratamiento va a ser efectivo para un paciente concreto, o por el contrario le resultará perjudicial.

La medicina actual se basa en el tratamiento de la enfermedad diagnosticada. Los biomarcadores permitirán evolucionar hacia una medicina que promueva la salud, y evite el desarrollo de la enfermedad, mediante el uso de dietas y fármacos correctores adecuados.

La fase de descubrimiento de biomarcadores se basa en la comparación de un conjunto de muestras de diversas poblaciones muy bien caracterizadas; como por ejemplo, entre otras muchísimas posibilidades:

- Población sana “versus” enferma.
- Sana “versus” diversos estados de la enfermedad.
- Muestras, antes de iniciar, y al terminar, un tratamiento de:
    - Una población de individuos que han respondido positivamente al tratamiento.
    - Población que empeora con el mismo.
    - Población a la que no afecta.
- ...

3.1.1. Estrategias en el descubrimiento de Biomarcadores.

Para el descubrimiento de biomarcadores, tradicionalmente en proteómica se ha venido utilizando una estrategia basada en la adquisición en modo “Data Dependent MS/MS”, pero cada vez más, con la irrupción de la “masa exacta” y potentes software de análisis estadístico multivariante, se va evolucionando hacia una estrategia equivalente a la utilizada en metabolómica; capaz de tener en cuenta la información de todos los compuestos ionizados.

3.1.1.1. Estrategia Tradicional: comparación de niveles de expresión de las proteínas identificadas con adquisición en modo “Data Dependent MS/MS”.

La estrategia tradicional para el descubrimiento de biomarcadores en proteómica se podría resumir en 3 etapas:

- La adquisición en modo Auto-MS/MS o “Data Dependent MS/MS”.
- Identificación de las proteínas expresadas.
- La comparación/cuantificación relativa de los niveles de expresión de las proteínas identificadas.

Adquisición en modo Auto-MS/MS: el QTOF opera en modo MS y cuando se detectan iones que superan un determinado umbral de intensidad de señal, en los siguientes ciclos el equipo efectuará MS/MS de iones precursores localizados en el espectro de MS (ver figura 15). Entre cada espectro de MS se intercalarán uno o varios de MS/MS. El usuario definirá entre otros parámetros:

- Nº de espectros de MS/MS que se intercalan entre cada espectro de MS.

- Nº máximo de repeticiones de un mismo ión precursor, durante un determinado intervalo de tiempo de exclusión.

- La preferencia de estado de carga de los iones precursores. En proteómica se suele dar preferencia a iones precursores con dos o más cargas, al generar un espectro de MS/MS con mayor nº de fragmentos ionizados, y por tanto, con más información para poder secuenciar el péptido (más información al final del apartado 3.2.2).

- Listado de *m/z* a excluir/incluir en la selección de iones precursores. Algunos modelos ofrecen adicionalmente:
  - Listado de iones preferidos: si se detectan aunque sean minoritarios, prevalecerán sobre otros iones mayoritarios.
  - Listas programables en el tiempo de iones preferidos, excluidos e incluidos.

A partir de estas directrices definidas por el usuario, y de los iones detectados en el último espectro de MS, el equipo decidirá por si sólo que iones precursores aislará para efectuarles un MS/MS; todo este proceso tiene lugar en tiempo real (más información en apartado 4.4.3).

Identificacón de las proteínas expresadas: mediante motores de búsqueda y bases de datos de proteínas, se identificarán las proteínas detectadas (ver apartado 3.2).

Cuantificación relativa se puede efectuar por técnicas con marcaje isotópico, isobárico o incluso sin etiquetado (“label free”) con la opción que incorporan algunos motores de búsqueda (ver figura 16). Las técnicas de marcaje permiten el análisis simultáneo de múltiples muestras, y facilitan comparar relativamente los niveles de expresión entre las distintas muestras, para correlacionar estos cambios con las propiedades de las distintas poblaciones comparadas. En las figuras 17-18 y tabla 4, se muestran el espectro de MS/MS y los resultados de la cuantificación relativa mediante iTRAQ (4-Plex), de una de las 800 proteínas identificadas en 4 muestras de tejido cerebral. En este estudio Pandey y Waddell “4-Plex iTRAQ Based Quantitative Proteomic Analysis Using an Agilent Accurate-Mass Q-TOF”(20) identifican proteínas expresadas diferencialmente en meningitis tuberculosa (más información en apartado 3.3).

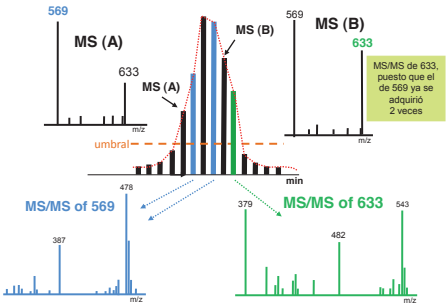


Figura 15.-Modo Auto MS/MS (“Data dependent MS/MS”): cuando la intensidad de señal de un espectro de MS (líneas negras) supera el umbral fijado, en los siguientes n ciclos se aislarán y fragmentarán automáticamente los n posibles iones precursores detectados en el espectro de MS. En este ejemplo n=1; por ello se alternan espectros de MS (en negro) con espectros de MS/MS (en color). Una vez se han adquirido m repeticiones del mismo ión precursor, éste entra temporalmente en una lista de exclusión (Exclusión Activa); en este ejemplo m=2, al haber obtenido 2 MS/MS del ión 569 (azul), en el tercer MS/MS (verde) se aisla el segundo ión en intensidad del espectro de MS (B): el 633 (verde).

Erosiva					No-Erosiva					
E_51_Br38 # spectra intensity	E_59_Br26 # spectra intensity	E_63_Br35 # spectra intensity	E_67_Br42 # spectra intensity	E_70_Br43 # spectra intensity	N_52_Br29 # spectra intensity	N_54_Br28 # spectra intensity	N_58_Br34 # spectra intensity	N_65_Br36 # spectra intensity	N_66_Br39 # spectra intensity	Protein Name
56 4.32e+008	28 2.68e+008	96 3.42e+008	20 1.49e+009	96 4.87e+008	11 1.81e+008	20 3.35e+008	0 0.00e+000	2 3.83e+007	1 3.40e+007	CALGRANULIN A(MIGRATION INHIBIT
0 0.00e+000	42 6.39e+008	23 2.49e+008	136 1.31e+009	25 4.99e+008	21 1.80e+008	0 0.00e+000	0 0.00e+000	4 4.19e+007	12 7.89e+007	C-REACTIVE PROTEIN PRECURSOR
8 6.26e+008	9 3.18e+008	31 7.34e+008	4 1.09e+009	6 6.34e+008	0 0.00e+000	0 0.00e+000	0 0.00e+000	1 4.01e+007	0 0.00e+000	RHO GDP-DISSOCIATION INHIBITOR 2
44 5.22e+008	35 1.17e+009	40 3.76e+008	10 1.06e+009	6 4.18e+008	6 1.96e+008	18 2.43e+008	4 7.43e+008	11 7.22e+007	6 2.03e+008	PROFILIN I
9 2.25e+008	104 6.89e+008	45 3.25e+008	22 5.05e+008	0 0.00e+000	2 3.72e+007	24 1.43e+008	110 1.83e+009	0 0.00e+000	57 9.70e+008	HEMOGLOBIN ALPHA CHAIN
24 7.04e+008	13 5.22e+008	14 5.01e+008	28 1.11e+009	12 1.08e+009	75 9.28e+008	10 7.72e+008	23 3.81e+009	128 3.19e+008	34 1.47e+009	IMMUNOGLOBULIN J CHAIN
3 7.96e+008	0 0.00e+000	3 4.50e+008	0 0.00e+000	4 7.64e+008	21 8.45e+009	13 2.69e+008	8 1.48e+009	56 5.02e+008	4 1.22e+009	APOLIPOPROTEIN D PRECURSOR
5 1.74e+008	4 4.37e+008	6 2.54e+008	3 1.54e+009	3 1.61e+008	12 8.55e+008	5 1.66e+007	6 6.44e+008	5 5.58e+007	0 0.00e+000	inter-alpha-trypsin inhibitor family heavy
3 5.80e+007	4 8.81e+008	14 3.73e+008	3 4.58e+008	0 0.00e+000	0 0.00e+000	1 2.76e+007	0 0.00e+000	1 6.04e+007	1 4.91e+008	14-3-3 PROTEIN BETA/ALPHA (PROTEIN
0 0.00e+000	12 1.16e+008	10 1.99e+008	14 2.50e+008	6 1.09e+009	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	Ig G1 H Nie
0 0.00e+000	10 4.64e+008	21 4.48e+008	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	OSTEOPONTIN PRECURSOR (BONE SI
3 5.42e+008	4 1.03e+009	13 5.30e+008	0 0.00e+000	3 7.20e+008	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	0 0.00e+000	TRIOSEPHOSPHATE ISOMERASE (TIM)

Figura 16.-Ejemplo comparación de nivel de expresión de proteínas sin etiquetado en 10 muestras (motor búsqueda Spectrum Mill) de artritis reumatoide (erosiva y no erosiva). Permite una muy sencilla y rápida comparación de cambios en los niveles de expresión. Cada cambio en la tonalidad del color indica el doble de intensidad de la señal de esa proteína (blanco indica no expresada – rojos y marrones indican mayor intensidad); también se indica su valor numérico, correspondiente al promedio de intensidad de los cromatogramas de iones precursores de los péptidos identificados de la proteína en cuestión.

Tabla 4.- Relación de intensidades de los iones reportero de cuatro péptidos de la proteína glial fibrilar ácida, en 4 muestras distintas etiquetadas isobáricamente con 4-Plex iTRAQ, y analizadas conjuntamente.

Secuencia del Péptido	"Peptide score"	Ratio 115/114	Ratio 116/114	Ratio 117/114
LADVYQAELR	18.4	1.0	2.6	2.8
LEAENNLAAYR	17.4	0.8	2.3	2.2
DNLAQDLATVR	17.3	0.8	2.7	2.6
FADLTDAAAR	16.2	0.9	2.9	2.7

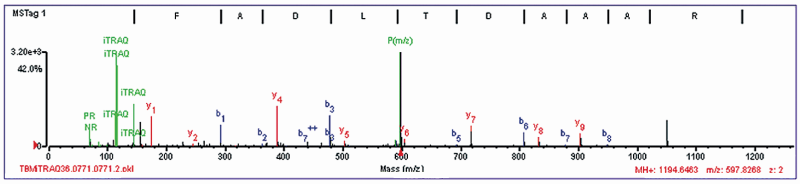


Figura 17.-Espectro MS/MS del péptido FADLTDAAAR 2+ (m/z precursor= 597.827) de la proteína glial fibrilar ácida obtenido mediante un QTOF con etiquetado 4-Plex iTRAQ; a pesar de la baja m/z de los "reporter ions" con respecto al ión precursor (< 1/5: 11X "vs" 598) se muestra un buen balance de intensidades.

Limitaciones estrategias basadas en Auto MS/MS, se podrían resumir en 2 puntos:

- Los cambios de expresión de los péptidos minoritarios no quedarán reflejados en el estudio, a menos que se separen cromatográficamente de los mayoritarios.

- Cuando coeluyan en el mismo tiempo un exceso de péptidos, no se dispondrá de tiempo para adquirir sus espectros de MS/MS; incluso en el caso de precursores no minoritarios.

En definitiva separar previamente las proteínas mayoritarias (p.e. mediante columnas o filtros de inmunoafinidad), fraccionar la muestra (Offgel Electophoresis, LC-2D, Geles 2D,…) y una buena separación cromatográfica, serán de gran ayuda para aumentar el nº de proteínas identificadas. Una alternativa para profundizar más en el proteoma, es realizar sucesivos análisis incluyendo en listas de exclusión programadas en el tiempo, las masas de los iones precursores ya adquiridos en inyecciones previas.

En lugar de adquirir en Auto MS/MS, Domon y Aebersold "Directed Mass Spectrometric Approach for In-depth Characterization of Complex Peptide Mixtures" (21), proponen primero adquirir en modo MS, y después programar en el tiempo un "targeted MS/MS" (aislamiento y fragmentación de los iones precursores detectados en el MS1) para caracterizar en profundidad muestras complejas. La estrategia descrita en el próximo apartado va en esa línea al basarse en la adquisición inicial en modo MS.

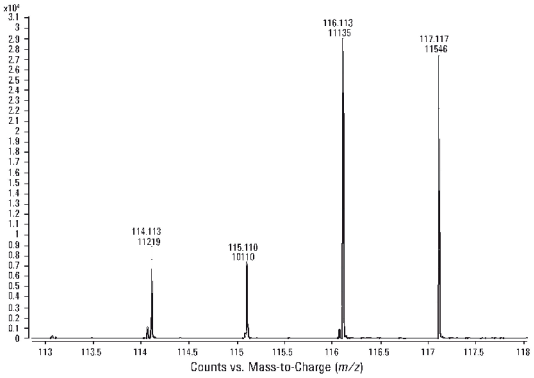


Figura 18. Ampliación de la región de m/z 113 a 117 ("reporter ios") en el espectro MS/MS del péptido FDLTDAAAR de la proteína glial fibrilar ácida. Cada uno de los 4 iones reportero corresponde a una muestra distinta, etiquetadas isobáricamente con 4-Plex iTRAQ y cromatografiadas conjuntamente.

3.1.1.2. Estrategia "Metabolómica": comparación de niveles de expresión de péptidos con adquisición en modo MS con masa exacta. Etapas y Ventajas con respecto a la estrategia tradicional.

La estrategia aquí descrita para el descubrimiento de nuevos biomarcadores, permite tener en cuenta cualquier péptido que se haya ionizado (incluyendo los minoritarios); por tanto ofrece una caracterización en profundidad de muestras complejas. Además, el uso de análisis estadístico multivariante con componentes principales (PCA), permitirá el estudio de complejas correlaciones entre cambios de expresión de péptidos/proteínas, con cambios en las propiedades de las poblaciones sujetas a estudio. Esta estrategia es la misma que se utiliza para el descubrimiento de biomarcadores en metabolómica; los instrumentos y herramientas bioinformáticas utilizadas son los mismos con pequeñas variantes, sobre todo en la identificación de las proteínas relevantes.

Un buen ejemplo de la potencialidad de ésta "estrategia metabolómica" en el estudio de complejos procesos biológicos, es el artículo publicado por Yanes, Clark y Siuzdak, "Metabolic oxidation regulates embryonic stem cell differentiation" (22), éste estudia la pluripotencialidad de las células madre embrionarias; en él se demuestra que el nivel de oxidación metabólica está regulado en los procesos de diferenciación de éste importante tipo de células.

La estrategia "metabolòmica" se podría resumir en las siguientes etapas:

- **1.-Adquisición en modo MS con masa exacta** de las muestras de proteínas digeridas con un LC/MS-QTOF o TOF. En esta fase de descubrimiento, para cada una de las poblaciones sujetas a estudio, se requiere un nº suficientemente elevado de individuos que sea bien representativo de esa población (idealmente 20-30 individuos/población). Para poder comparar resultados con suficiente selectividad y especificidad en muestras complejas, un aspecto clave es disponer de un instrumento con masa exacta robusta, que no cambie con la concentración del analito o la temperatura del laboratorio.

- **2.-Extracción automatizada de la masa exacta e intensidad de señal específica** de cada uno de los péptidos ionizados. Para la obtención de esta información se deberá disponer de algún algoritmo de deconvolución, que agrupe aquellas masas que "suban y bajen" al mismo tiempo en el cromatograma.

- **3.-Análisis estadístico multivariante con componentes principales (PCA),** para descubrir aquellos péptidos cuyos cambios en su respuesta se pueden correlacionar con cambios en las propiedades de las poblaciones sujetas a estudio. Previamente al aná-

lisis estadístico se deberá efectuar un proceso de alineación de tiempos de retención y masas, y de normalización de las respuestas, para una buena comparación de los resultados de muestras que han sido adquiridas durante amplios períodos de tiempo.

Cuando se utiliza componentes principales (PCA), es muy útil ir intercalando entre las muestras, como control de calidad, un "pool" o mezcla a partes iguales de las muestras analizadas. En los gráficos de PCA, estas muestras de control deberían aparecer agrupadas en la zona central entre las diversas poblaciones estudiadas. Se puede ver un ejemplo en la publicación de Coral, Cifuentes y Godzien, "Metabolomic Approach with LC-QTOF to Study the Effect of a Nutraceutical Treatment on Urine of Diabetic Rats" (23)

- **4.-Programar en el LC/QTOF una adquisición en "targeted MS/MS" programada en el tiempo de los péptidos relevantes,** candidatos a biomarcadores, descubiertos en la etapa anterior mediante PCA (Análisis por Componentes Principales).

- **5.-Identificar a que proteínas pertenecen los espectros de MS/MS** con masa exacta de los péptidos relevantes, mediante el uso de un motor de búsqueda y bases de datos de proteínas. Teóricamente los péptidos de la misma proteína deberían mostrar cambios de expresión relativa equivalentes, y ser agrupados conjuntamente por PCA. Ello debería proporcionar un nº elevado de péptidos para cada proteína relevante, y por tanto identificaciones de calidad al proporcionar una elevada cobertura en la secuencia de esas proteínas. Ésta elevada cobertura facilitará la detección de modificaciones post-traduccionales (PMT's) en estos potenciales biomarcadores, y el cálculo del ratio de proteína modificada "versus" nativa.

- **6.-Evaluar en que Rutas Metabólicas están involucradas las proteínas relevantes:** en el mercado existe algún software de análisis estadístico multivariante, que incorpora un módulo de análisis de rutas metabólicas. Éste permite determinar que rutas metabólicas pueden ser relevantes en el experimento biológico realizado; aquellas en las que están involucradas las proteínas relevantes (candidatas a biomarcadores). Otra interesante posibilidad, es la de visualizar el impacto de un determinado tratamiento o enfermedad, sobre los compuestos involucrados en una determinada ruta metabólica; para ello visualiza directamente sobre la ruta los cambios de expresión de las distintas poblaciones estudiadas (ver ejemplo en figura 19).

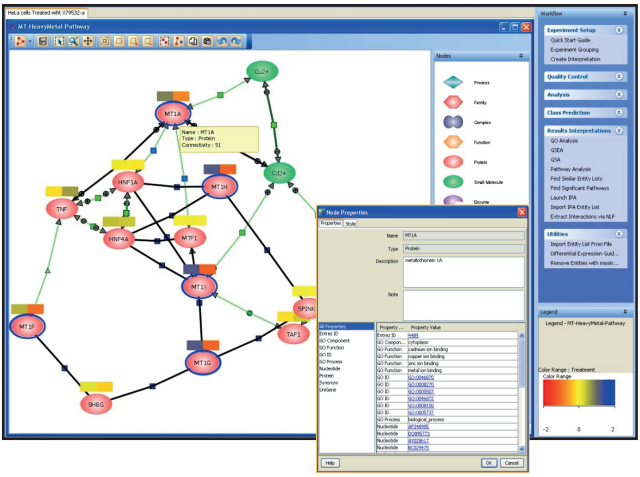


Figura 19.-Visualización en Mass Profiler Profesional de cambios de expresión en una ruta metabólica. Se observa como algunas proteínas no cambian de expresión entre las 2 poblaciones comparadas, otras cambian ligeramente y otras muy intensamente (azul-rojo).

De estas 6 etapas, las cuatro primeras y la última, son idénticas entre proteómica (análisis de péptidos) y metabolómica (análisis de metabolitos), y se podrán utilizar los mismos instrumentos y software. Las diferencias se centran en la quinta etapa; la identificación de las proteínas relevantes requerirá el uso de motores de búsqueda (Spectrum Mill, Mascot, Sequest, Profound,…), en bases de datos de proteínas (Swiss-Prot, NCBInr, dBEST, OWL, DBMS, IPI,…). En cambio la identificación de metabolitos se realiza mediante búsquedas en base de datos de masa exacta (METLIN DB,HMDB, Lipid Maps,…), y en algunos casos mediante bibliotecas específicas de espectros de MS/MS (con QTOF) de metabolitos (METLIN DL). Algunas de estas bases de datos o bibliotecas de espectros, incluyen también los tiempos de retención de varios centenares de metabolitos (para unas condiciones cromatográficas concretas).

La otra diferencia entre proteómica y metabolómica, suele estar en el tipo de cromatografía y columnas habitualmente utilizadas:

- En proteómica se suele usar nano-cromatografía con columnas típicas de 0.075mm de diámetro interno, a flujos de 200-600nL/min y con una fuente de iones nanoelectrospray (nebulización del eluyente mediante campo eléctrico).

- En metabolómica se suele trabajar con columnas convencionales de 2.1mm, a flujos 0.2-0.6mL/min y con fuentes electrospray o APCI (ambas con nebulización asistida por nitrógeno). Además también puede utilizar GC/MS (cromatografía de gases/masas) para metabolitos volatilizables.

De las 6 etapas anteriormente descritas, a continuación se describirán con más de detalle, la segunda (extracción de la información de los péptidos ionizados) y tercera (análisis estadístico):

Ampliación información Etapa 2.-PROCESO DECONVOLUCIÓN MAPA-3D. (Tiempo, *m/z* exacta, Abundancia): permitirá la extracción automatizada del tiempo de retención, la masa exacta, e intensidad de señal específica, de cada uno de los péptidos ionizados; además de su espectro de MS. La adquisición en modo MS (sin fragmentar), tiene la gran ventaja de almacenar en el disco duro, la información de **todos** los compuestos ionizados. Si la adquisición se realiza con masa exacta (como en el caso de un TOF o QTOF), se dispondrá además de una selectividad tipo MS/MS, cuando se extrae un cromatograma de iones con una muy fina ventana de masa (unas pocas centésimas de Dalton); bastará un buen algoritmo de extracción para obtener información de todos los péptidos ionizados. La etapa de deconvolución es una de las más importantes. En el mercado existen diversos algoritmos de deconvolución, que agrupan aquellas masas que "suban y bajan" al mismo tiempo en una "feature" (ver figura 20), alguno como el denominado "Molecular Feature Extractor" (ver figura 21), permite además agrupar esas masas con masa exacta, mediante correlaciones químicas. Estas correlaciones –definibles por el usuario- asocian iones con múltiples cargas, contribuciones de los isótopos estables, pérdidas neutras (17, 18, 44, 49, 98,…), aductos (polaridad +: sodio, amonio,… //polaridad -: formiato, acetato,…). Esta "agrupación

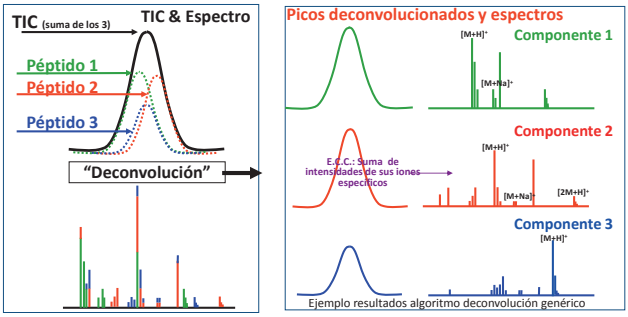


Figura 20.-Ejemplo de resultados de deconvolución genérica; el algoritmo agrupa las masas que "evolucionan conjuntamente" (en el tiempo e intensidad) a lo largo de un pico, proporcionando los espectros y cromatogramas de iones deconvolucionados de los compuestos que coeluyen en el mismo, siempre y cuando presentan pequeñas diferencias en sus tiempos de retención. En el lado izquierdo de la figura se observa un pico en negro (TIC) que corresponde a la coelución de 3 péptidos (color); en el lado derecho se muestran los espectros y cromatogramas específicos de los 3 péptidos extraídos por deconvolución.

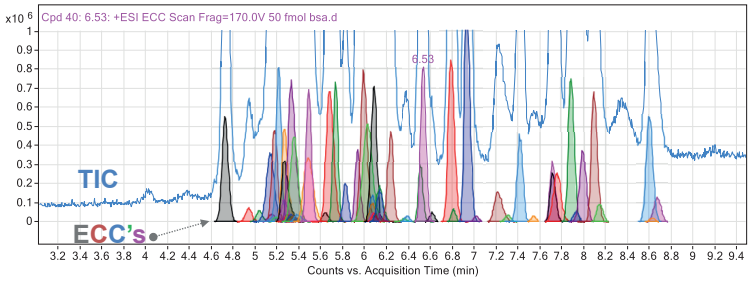


Figura 21.-Deconvolución con algoritmo "Molecular Feature Extractor": en la parte inferior de la figura se muestra un ejemplo de deconvolución de una digestión tríptica. Cada uno de los picos de distinto color que aparecen en la parte inferior corresponden a un péptido distinto (ECC: Extracted Compound Chromatogram); a pesar de las numerosas coeluciones el software proporciona el espectro, tiempo de retención, la masa exacta y abundancia específicos de cada uno de los péptidos ionizados, aunque coeluyan perfectamente (p.e. péptidos a 5.3 y 7.7min). La traza superior en azul corresponde al "Total Ion Chromatogram" (TIC) de la mezcla de péptidos antes de deconvolucionar.

química" de los iones permite resolver espectralmente, solapamientos de péptidos que incluso coeluyan perfectamente y no se diferencien ni por tiempo de retención, ni por distinto perfil cromatográfico; ello resulta de gran utilidad en la deconvolución de muestras complejas. En los algoritmos convencionales, 2 péptidos perfectamente coeluyentes aparecerían como 1 solo péptido. Más información en apartado 4.5.1.

Posteriormente los datos de tiempo de retención, masa exacta y abundancia para cada péptido deconvolucionado, serán exportados al software de análisis estadístico multivariante, con el objetivo de descubrir los péptidos relevantes (aquellos que marcan las diferencias entre las distintas poblaciones comparadas). Otra alternativa es la de no deconvolucionar, y trabajar directamente con los tiempos y las intensidades de las masas individuales, en el software de análisis estadístico; el no agrupar las masas tendrá el inconveniente de no compensar posibles cambios en los ratios de intensidad entre distintos iones del mismo péptido, y el tener que tratar un mayor volumen de datos en el análisis estadístico.

Ampliación información Etapa 3.-FASE ANALISIS ESTADISTICO MULTIVARIANTE. Tienen por objeto descubrir aquellos péptidos cuyos cambios en su respuesta se pueden correlacionar con cambios en las propiedades de las poblaciones sujetas a estudio. Una vez los datos adquiridos en modo MS con masa exacta se han deconvolucionado, los datos de masa exacta, tiempo retención y abundancia, se importarán en el software de análisis estadístico multivariante, donde se someterán a diversas etapas:

Alineamiento y normalización: antes de su análisis estadístico, los datos se alinearán en sus tres dimensiones (tiempo, masa y abundancia), con las herramientas que incorpora el software estadístico, y que permiten el uso de patrones internos para mejorar el alineamiento y corrección/normalización de la respuesta. Estos estudios se suelen prolongar en el tiempo, y pueden producirse cambios en la respuesta del espectrómetro, y en los tiempos de retención cromatográficos (p.e. por envejecimiento o contaminación de la columna). Otro aspecto importante a tener en cuenta, es la posible reducción o exaltación de la respuesta de un péptido (que no cambie su concentración), producidos por cambios, de una muestra a otra, en la concentración de los compuestos que coeluyen con el mismo; una buena cromatografía o fraccionamiento previo de las muestras (para simplificarlas), ayudará a reducir el nivel de interferencias.

Control de calidad/ filtrado de los datos: para descartar aquellos datos cuya variabilidad entre los replicados sea excesiva, siempre es recomendable aplicar filtros de control de calidad a los datos/muestras:

Filtros por frecuencia: descartan los péptidos que superen la máxima proporción de ausencia, que se acepta entre los replicados de cada población.

Filtros por análisis de varianza: descartan de entrada los péptidos cuyas respuestas cambien excesivamente entre los replicados de la misma población; ya sea por cambios reales de concentración, o por cambios de respuesta por los fenómenos anteriormente indicados. Los tests a utilizar para éste cribado, dependerán del nº de factores según los cuales se clasifiquen a las distintas poblaciones a comparar:

- **Monofactoriales:** para un solo factor (pe. estado salud: sano o enfermo) se podrán utilizar un T-test o una ANOVA.

- **Multifactoriales:** en estudios multifactoriales (p.e.: estado salud, sexo, raza,..), se deberá utilizar un test de ANOVA multi-paramétrico de acuerdo con el nº de factores.

Filtro por cambio mínimo en la expresión: exige un mínimo cambio de expresión entre unas poblaciones y otras (p.e. que como mínimo se doble la respuesta).

Clusterización jerárquica: permite visualizar si los replicados de las diversas poblaciones y condiciones, se agrupan correctamente (ver figura 22), indicando que existen cambios en los perfiles de péptidos de las muestras, que permiten su diferenciación y clasificación.

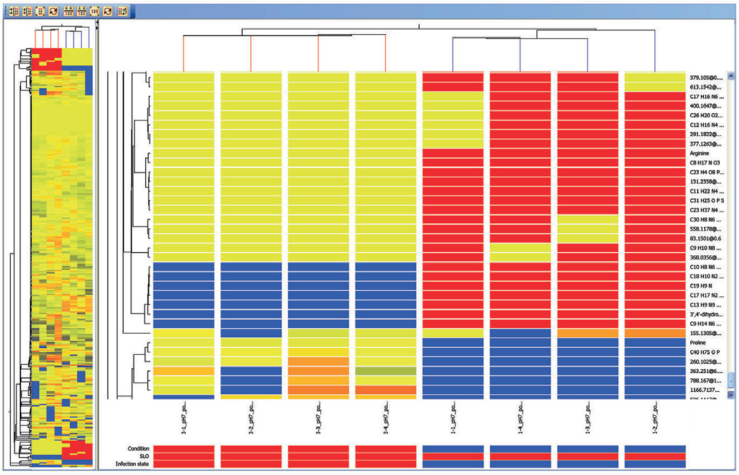


Figura 22.-Ejemplo de clusterización jerárquica de 2 conjuntos de muestras (sanas e infectadas por malaria); además de clasificar las muestras (columnas) según sus similitudes, también agrupa los compuestos (filas) que presentan perfiles semejantes de concentración. Esta agrupación puede revelar interesantes relaciones de compuestos que muestran un comportamiento semejante frente a las distintas condiciones experimentales evaluadas, y que pueden compartir rutas de reacción similares. También agrupará en la misma zona los péptidos procedentes de la misma proteína que presentaran en general perfiles de concentración equivalentes, aunque estos pueden cambiar si la proteína nativa tiene un comportamiento distinto en algunas de sus isoformas. En este caso los péptidos modificados tenderán a separarse (en el gráfico de clusterización) del resto de péptidos de la forma nativa y otras isoformas que compartan comportamiento.

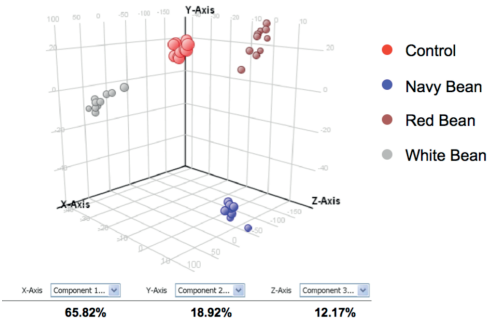


Figura 23.-Ejemplo de gráfico tridimensional de Análisis de Componentes Principales (PCA loadings) de replicados biológicos de 4 poblaciones de ratones sometidas a distinto tratamientos. A partir del mismo se podrán conocer aquellos péptidos que se han utilizado para poder diferenciar entre estas 4 poblaciones.

Componentes Principales (PCA): permite evaluar mediante una sencilla representación ("loading plots") bi o tridimensional (ver figura 23), y sin hipótesis previas (aproximación holística), si se han encontrado cambios en las concentraciones de los péptidos, que permitan diferenciar claramente entre las distintas poblaciones y condiciones experimentales evaluadas. En estos gráficos cada muestra se representa por un punto en un sistema de 2 o 3 componentes principales (dimensiones); los puntos del mismo tipo deberían agruparse, y los grupos de distintas poblaciones separarse. A partir del mismo, se podrán conocer aquellos péptidos que se han utilizado para poder diferenciar entre las distintas poblaciones y condiciones evaluadas, y que corresponden a proteínas potencialmente biomarcadoras.

Herramientas adicionales: los paquetes estadísticos suelen incorporar un amplio conjunto de muy útiles herramientas para análisis y visualización de datos y resultados, clasificación de muestras, predicción de clases, e interpretación de resultados.

3.1.2. Validación de Biomarcadores y Evaluación Significado Biológico.

Un aspecto esencial en todo proceso de búsqueda de nuevos biomarcadores, es su validación. La enorme variabilidad de los organismos vivos, tanto por factores endógenos como exógenos, y especialmente en humanos donde la variabilidad es enorme, **obliga ineludiblemente a validar los presuntos biomarcadores descubiertos** con, idealmente, **miles de individuos; pero usualmente con una selección de unas pocas/decenas de proteínas a cuantificar** (las potenciales biomarcadoras). El uso de extensas poblaciones de individuos, permitirá corroborar los resultados obtenidos inicialmente a partir de una pequeña muestra de individuos, supuestamente representativos de toda una población.

Validación mediante Triple Cuadrupolo:

Dado el elevado nº de individuos de cada población (centenares→miles), la validación de péptidos/proteínas candidatos a biomarcadores, lo ideal es realizarla con un triple cuadrupolo (QqQ) en modalidad MRM por su mayor rapidez y sensibilidad para cuantificar hasta unas pocas centenas de compuestos por inyección. Hoy en día pueden llegar a adquirir hasta varios centenares de transiciones/segundo, permitiendo el seguimiento de centenares (/miles) de péptidos, o de picos cromatográficos muy finos; especialmente con modelos que permitan programar las transiciones, específicamente en una pequeña franja alrededor del tiempo de retención del péptido (MRM dinámico). La sensibilidad proporcionada por algunos triple cuadrupolos puede llegar a muy bajos attomoles de péptidos.

Dado el elevado nº de muestras a analizar, convendrá utilizar cromatografía rápida (si la sensibilidad es suficiente para los péptidos de menor respuesta); por ello será mejor utilizar HPLC convencional o UHPLC con columnas de 2.1mm de diámetro, más rápidas y robustas que los tradicionales sistemas de nano-cromatografía (columnas d.i. 75µm), utilizados en proteómica. Hoy en día existen en el mercado triple cuadrupolos con tecnología "i-funnel", que permiten llegar a bajos niveles de attomoles utilizando columnas de d.i. 2.1mm; con estos sistemas la nanocromatografía sólo mejoraría la sensibilidad por un factor de 10, en lugar de las +100 veces habituales con respecto a los triples tradicionales. Hay que tener en cuenta que en muestras muy complejas, la reducción de los tiempos de cromatograma a muy pocos minutos (p.e. <5min), reducirá la sensibilidad obtenida; al aumentar el ruido con el necesario incremento de la frecuencia de adquisición. Un sistema UHPLC/MS-QqQ con tecnología i-funnel sería ideal para esta fase, especialmente si el modelo de QqQ utilizado es igual hasta la celda de colisión, al QTOF utilizado en la fase de descubrimiento; ello permite obtener los mismos espectros de MS/MS, y poder compartir parámetros entre ambos instrumentos.

Un sistema muy útil para la validación y seguimiento de biomarcadores, es la técnica multiplexada denominada SISCAPA ("Stable Isotope Standard Capture with Anti-Peptide Antibodies"). Ésta permite aislar y preconcentrar selectivamente los múltiples péptidos de interés, mediante

un protocolo automatizable con el uso de robots. En este protocolo, perlas magnéticas recubiertas con múltiples anticuerpos anti-péptido, se unen específicamente a los péptidos de interés procedentes de la digestión tríptica de muestras, tales como suero y plasma; después las perlas se lavan para eliminar los péptidos no enlazados; finalmente los péptidos de interés purificados (los enlazados a los anticuerpos), se eluyen en un pequeño volumen para su inyección en un sistema LC /MS. La técnica aporta las ventajas de:

- Permitir llegar a preconcentrar hasta >100.000 veces los péptidos de interés, y así facilitar enormemente la cuantificación de péptidos muy minoritarios. Permite el uso de tecnologías de MS menos sensibles que los triple cuadrupolos.

- Eliminar el resto de péptidos, para disminuir dramáticamente los fenómenos de supresión de ionización por coelución con otros péptidos mayoritarios.

- Favorecer el uso de cromatografía convencional (columnas 2.1mm d.i.), mucho más robusta, rápida y reproducible que la nanocromatografía.

Para más detalles ver publicación de Anderson L y Miller C, "Automation of a "SISCAPA Magnetic Bead Workflow for Protein Biomarker Quantitation by Mass Spectrometry" (24); el término SISCAPA fue introducido por Leigh Anderson en el 2004 (25).

Validación mediante Cuadruplo-Tiempo Vuelo (QTOF):
En la fase de validación también puede utilizarse un QTOF, pero se obtendrá menor sensibilidad que con un QqQ de gama equivalente (del orden de unas 20 veces). No obstante hoy en día existen ya en el mercado QTOF's de gama muy alta con sensibilidad mejor o equivalente a la de QqQ's de gama media. Las 2 modalidades de adquisición para validar biomarcadores con QTOF son:

- MS (TOF o QTOF): con la extracción de cromatogramas de iones de las masas de interés, con muy finas ventanas de extracción de masa. Esta modalidad proporciona buena selectividad (tipo MS/MS con QqQ), pero menos que la posterior; en cambio, es más rápida (suficiente para picos < 1seg. de ancho) y su sensibilidad no suele ser muy diferente. Además esta modalidad no tiene límite en el número de péptidos coeluyentes, que se pueden cuantificar con 1 sola inyección.

- "Targeted MS/MS" programando en el tiempo la fragmentación de los iones precursores de los péptidos de interés. En este modo los QTOF (u otras tecnologías basadas en trampas de iones) son más lentos que un QqQ en MRM, si se quiere acumular un número suficiente de vuelos ("transients") o iones, para tener una buena sensibilidad. Especialmente si coeluyen muchos péptidos al mismo tiempo, con QTOF (u otras tecnologías en modo MS/MS) se deberá utilizar una cromatografía más lenta que con QqQ en modo MRM.

Una vez adquiridos los datos, se deberá repetir el análisis estadístico multivariante con componentes principales (PCA), con los datos del QqQ en modalidad MRM (o del QTOF), para confirmar o descartar esos péptidos/ proteínas propuestos como biomarcadores. La normalización de las respuestas es un importante aspecto a considerar, para compensar posibles cambios en la señal obtenida debidos a alteraciones en las condiciones de medida. Un calibrado frecuente de la respuesta con patrones o muestras de referencia, y el uso de muestras de control será importante para la buena calidad de los resultados obtenidos. Tampoco se deberá olvidar el posible impacto del efecto de supresión o exaltación en la respuesta de un péptido, por coelución con distintas concentraciones de otros compuestos, cuya concentración puede variar de unos individuos a otros (de ahí la importancia de la cromatografía). El uso de los mismos péptidos

marcados isotópica o isobáricamente, y que coeluyan exactamente con el péptido de interés, compensan estas interferencias; pero se deberá asegurar que la técnica de marcaje utilizada, proporcione una buena reproducibilidad y robustez del proceso de marcaje.(Más información en apartados 3.3 y 4.5.4).

Evaluación Significado Biológico.

En en el existen mercado softwares de análisis estadístico multivariante, que ofrecen un módulo para análisis de rutas metabólicas. Éste permite ver a que rutas metabólicas afectan esos biomarcadores, o proteínas y péptidos que muestran cambios de expresión; incluso permiten crear de nuevas si es el caso. Algunos incluso permiten integrar datos de genómica, transcriptómica, proteómica y metabolómica en una misma plataforma para facilitar el estudio integrado de la biología de sistemas (ver figura 24). Existen organizaciones como Cytoscape (http://www.cytoscape.org/), que ofrecen una plataforma de software en código abierto, para el análisis y visualización de complejas redes de interacciones biomoleculares, para el estudio de la biología de sistemas.

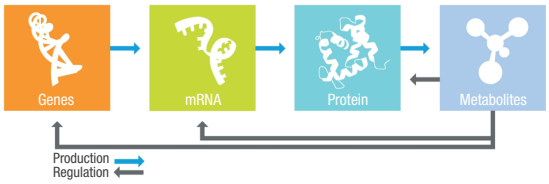


Figura 24.-El genoma de un organismo no es estático. Está influenciado/regulado por el entorno y por el propio metaboloma. Sus cambios modificaran a su vez el proteoma y metaboloma.

Estos softwares de análisis de rutas metabólicas, también pueden utilizarse para **ver exhaustivamente el impacto,** de determinados tratamientos o determinadas poblaciones, **en rutas metabólicas concretas** (ver figura 19). En este caso se puede exportar la lista de compuestos involucrados, para efectuar un análisis dirigido de los péptidos en cuestión. El análisis de este listado de péptidos entre las distintas poblaciones a comparar, se puede efectuar mediante:

- **LC/MS-TOF o QTOF** extrayendo directamente el listado de masas de los péptidos, de los ficheros de MS con masa exacta ya adquiridos.

- **Triple cuadrupolo en modalidad MRM** cuando interese máxima sensibilidad. Para ello se deberán optimizar las condiciones de fragmentación de los péptidos en cuestión; también se puede recurrir a calcular la energía de colisión mediante una fórmula en función de su masa.

El proyecto del SRM/MRM Atlas del Proteoma Humano (19) (http://www.srmatlas.org/), facilita enormemente el obtener las condiciones idóneas de adquisición en MRM de los péptidos de interés cuando se trabaja con QqQ; las transiciones pueden buscarse en la web: https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/GetMRMList.

Adicionalmente se espera la comercialización de bases de datos MRM, que faciliten la implementación de cuantificaciones masivas de péptidos con triple cuadrupolo. También pueden utilizarse softwares como “Skyline” (https://skyline.gs.washington.edu/) y “Peptide Optimizer”; éstos proporcionan herramientas para:

- Localizar péptidos proteotípicos.
- Fragmentar “in silico” secuencias peptídicas, y calcular las transiciones a seguir para su análisis por MRM.
- Ecuaciones para el cálculo de la energía de colisión típica (en función de la m/z del péptido y su número de cargas).
- Herramientas para optimizar las condiciones de fragmentación directamente en modalidad MRM.

3.1.3. Posicionamiento en Descubrimiento de Biomarcadores de los Tiempos de Vuelo con respecto a otros Analizadores de Masa.

Para el descubrimiento de biomarcadores, **es esencial disponer de una masa exacta** robusta. Ello permitirá trabajando en MS con métodos de adquisición genéricos, tener suficiente selectividad (tipo MS/MS) para poder comparar el nivel de expresión de centenares/miles de péptidos en muestras biológicas. Se compararán cromatogramas de iones con muy finas ventanas de extracción de masa. Un aspecto importante en esta comparación son el rango dinámico de la respuesta y el espectral (independencia de la masa exacta con respecto a la concentración).

¿Se Pueden Utilizar Cuadrupolos o Trampas de Iones en la Fase de Descubrimiento de Nuevos Biomarcadores?

En la fase de descubrimiento de nuevos Biomarcadores, **NO** servirá el utilizar equipos de baja exactitud de masa como cuadrupolos, triple cuadrupolos, trampas de iones,…. puesto que **la masa nominal no es suficientemente selectiva para la comparación de analitos desconocidos en muestras complejas**. La utilización de equipos de baja exactitud y resolución, sólo será posible en la fase de validación de Biomarcadores, o en aquellos casos en los que la búsqueda de biomarcadores quede restringida a un conjunto limitado de compuestos conocidos; en definitiva, sólo cuando se puedan realizar los análisis en las modalidades: "targeted MS/MS" para Triple Cuadrupolos y Trampas de Iones, o en modalidad SIM para cuádruplos simples; pero no en la modalidad barrido o scan MS, para poder comparar también analitos desconocidos, en cuyo caso se requerirá masa exacta y una resolución espectral superior a la que proporcionan estos equipos.

Tiempos de Vuelo QTOF o TOF "versus" Otros Analizadores de Masa Exacta.

Exactitud de Masa y Rango Dinámico: La enorme complejidad y rango de concentraciones del proteoma, humano conllevará el tener que comparar péptidos con enormes diferencias en sus niveles de expresión. Se necesitarán espectrómetros con un amplio rango dinámico, cuya masa no cambie con la concentración del péptido, y que proporcionen un nivel suficiente de sensibilidad (bajos ppb's) y resolución espectral (+10.000). **La masa exacta y su independencia de la concentración del péptido, son los 2 aspectos más importantes para la comparación de muestras con analitos "desconocidos".** Hoy en día existen diversas tecnologías capaces de proporcionar masa exacta (<5ppm), e incluso <1-2ppm en los mejores equipos de diversas tecnologías.

Donde existe mayor diferencia, es en los rangos dinámicos espectrales entre las diversas tecnologías y modelos actuales: desde 10^2 a $+10^5$. Al tener que comparar péptidos cuya concentración puede variar enormemente de unas muestras a otras, y coeluir con otros mucho más concentrados, **será muy importante que la masa no cambie con la concentración del péptido**; **el rango máximo del ratio de intensidades entre 2 compuestos coeluyentes puede variar entre unos centenares, 5.000 y 120.000 en los modelos y tecnologías** más actuales. Los mejores instrumentos son modelos de QTOF/TOF con tecnología ADC y doble amplificador, capaces de llegar a rangos dinámicos espectral y cuantitativo de +5 órdenes de magnitud; además su menor propensión a la saturación facilita el proporcionar una excelente fidelidad isotópica. El resto de modelos de QTOF/TOF y otras tecnologías de masa exacta acopladas a trampas de iones, se suelen quedar en alrededor de 3-4 órdenes de rango dinámico (típico 5.000). En los modelos de QTOF/TOF con tecnología de medida del tiempo de vuelo tipo TDC, el rango dinámico espectral suele ser de unos 2 órdenes de magnitud.

Resolución Espectral y Frecuencia de Adquisición.

La resolución espectral de los QTOF/TOF es independiente de la frecuencia de adquisición, pudiendo llegar a proporcionar hasta 50 espectros MS/segundo a resoluciones en el rango de 20.000-50.000 (dependiendo de los modelos y rangos de masa). En la Figura 25 se muestra la resolución a distintas frecuencias de adquisición de datos, proporcionado por un modelo de alta resolución UHD-QTOF para una serie de péptidos de distinta *m/z*; éstos demuestran que los QTOF

no pierden resolución por adquirir a mayor velocidad. En cambio, en otras tecnologías de masa exacta, basadas en la medida de la *m/z* por la frecuencia de movimiento harmónico de iones atrapados orbitalmente, acopladas a trampas de iones o cuadrupolos; su resolución es muy dependiente de la velocidad de adquisición, y si bien son capaces de superar los 100.000 de resolución a masas bajas, necesitan suficiente tiempo para realizar la medida (típico > 1seg). Si la frecuencia de adquisición hay que subirla a 10 espectros/segundo (para cuantificar muy finos picos cromatográficos) su resolución se reduce a 10.000 a masas bajas por debajo de las habituales para péptidos. Por ejemplo al pasar de 1 a 10Hz la resolución a *m/z* 400 se reduciría de 78.000 a tan sólo 8.000; y aún más a *m/z* mayores como suele suceder con los péptidos. Esta fuerte pérdida de resolución con la velocidad es la que lleva a que, en algunos modelos, mientras se realizan varios espectros de MS/MS en la trampa, el módulo de alta resolución determina la masa exacta del ión precursor (A. Makarov 26); consecuentemente sólo se obtiene la masa exacta del ión precursor, pero no la de los iones producto. En la tabla 5, se muestra la comparación de una estimación de la resolución proporcionada por los últimos desarrollos de ambas tecnologías, a distintas velocidades y rangos de masa.

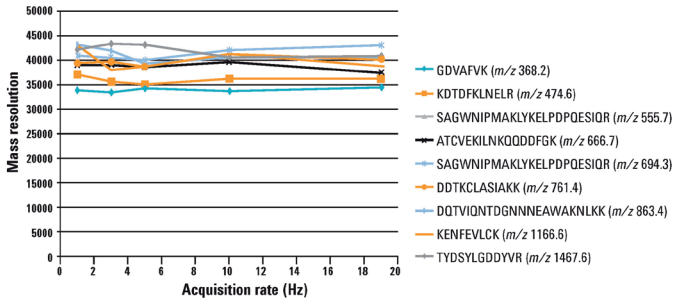


Figura 25.-Ejemplo de la resolución a distintas frecuencias de adquisición de datos proporcionado por un Agilent UHD-QTOF-6540 para una serie de péptidos con m/z de 368 a 1468. Se observa como la resolución es independiente de la frecuencia de adquisición en el rango de 1-20Hz.

Tabla 5.- Comparativa de típicas resoluciones espectrales a distintas frecuencias de adquisición, entre QTOF's de Alta Resolución y Analizadores Orbitales de última generación (2011).

Analizador	Masa m/z	Frecuencia Adquisición (algunos modelos QTOF hasta 50Hz)			
		1 Hz	6 Hz	12 Hz	20 Hz
UHD-QTOF	200	QTOF: Independiente de la frecuencia de adquisición.			
	400	25.000			
	900	35.000			
	2.000	45.000			
		42.000			
Analizador Orbital	200	140.000	35.000	17.500	
	400	100.000	25.000	12.500	
	900	70.000	17.500	8.000	
	2.000	38.000	9.500	4.750	

En LC/MS una mínima de resolución espectral (+10.000), conjuntamente con la masa exacta, es necesaria para disponer la selectividad necesaria para poder comparar perfiles de concentración de péptidos en muestras biológicas complejas. **Con resolución +10.000 (10K) se podrán ya resolver coeluciones de analitos que difieran tan sólo unas pocas centésimas de *m/z*** (masa/carga). Resoluciones mayores pueden ser útiles en LC/MS, para resolver casos muy puntuales de péptidos que se diferencien entre sí unas pocas milésimas de *m/z,* y que además coeluyan cromatográficamente; ello es muy poco probable en LC/MS.

Los diversos modelos actuales de QTOF/TOF del mercado, se podrían clasificar en 2 grandes grupos:

- **HD-QTOF de alta resolución,** capaces de proporcionar resoluciones de 10-20K. Estos suelen tener tubos de vuelo de alrededor de 1m.

- **UHD-QTOF de MUY alta resolución,** capaces de proporcionar resoluciones de 20-40K. Estos suelen tener tubos de vuelo de más de 1.5m.

Dentro de los QTOF de muy alta resolución (UHD-QTOF) se pueden clasificar en diversos grupos según la tecnología utilizada para alcanzar su máxima resolución espectral:

- **Modelos con 2 reflectores de iones** (2º reflectrón opcional): para ampliar su recorrido (W) y conseguir resoluciones de 20-40K; esta doble reflexión de los iones se traduce en una menor sensibilidad (típico 3-5 veces), con respecto a la que obtienen en modo de trabajo con 1 sólo reflectrón (V); éste modo se utiliza cuando se requiere más sensibilidad, pero reduce la resolución a 10K.

- **Modelos con 1 sólo reflectrón y un tubo de vuelo de 2.7 metros:** trabajando en modo "focalización" activada alcanzan resoluciones de hasta 40-60K; éste modo mejora la resolución pero sacrifica sensibilidad (típico 2-4 veces). Sin la focalización activada la resolución se reduce a 20-25K.

- **Modelos con 1 sólo reflectrón, tubo de vuelo de 1,6 metros y un compresor del haz de iones** (guía de iones hexapolar convergente ver figura 7): permiten alcanzan resoluciones de 40-50K sin sacrificar sensibilidad e independientemente de la frecuencia de adquisición.

Resolución Espectral y Rango de Masa.
La resolución espectral cambia con la *m/z* del analito con un comportamiento distinto según el tipo de analizador de masa exacta utilizado (ver tabla 5):

- **QTOF/TOF** la resolución aumenta con el incremento de masa debido al mayor tiempo de vuelo de los iones más grandes. Por ejemplo valores típicos de resolución para UHD-QTOF midiendo a 20Hz: 25.000 a *m/z*=200, 35.000 a *m/z*=400, 45.000 a *m/z*=900.

- **Analizadores Orbitales:** en sistemas basados en la medida de la *m/z* por la frecuencia de movimiento harmónico de los iones atrapados orbitalmente, la resolución disminuye con el incremento de la masa. Esta disminución de resolución con el incremento de la *m/z* del péptido, se debe a que iones con menor densidad de carga se mueven más lentamente, y pueden realizar menos ciclos de oscilación en el tiempo de medida del espectro; por ejemplo al pasar de m/z 200 a m/z 900, la resolución se reduce a ½, y a menos de 1/3 a m/z 2.000.

3.2. Identificación de Proteínas.

La identificación de proteínas se realiza a partir del análisis de una muestra digerida enzimáticamente; se basa en el uso de motores de búsqueda (Spectrum Mill, Mascot, Sequest, Profound, …) y bases de datos de proteínas (Swiss-Prot, NCBInr, dBEST, OWL, DBMS, IPI,…). Esta identificación se fundamenta en la elevada especificidad de la combinación de los 20 aminoácidos esenciales, en secuencias de péptidos con 6 o más aminoácidos.

Bases de datos de proteínas: contienen las secuencias de aminoácidos de las proteínas que un determinado organismo (/taxonomía) puede llegar a expresar. Las secuencias de aminoácidos de las posibles proteínas presentes en un organismo se deducen a partir de la secuencia del genoma del organismo en cuestión.

Motor de búsqueda: se encargará de contrastar la información generada por el espectrómetro de masas, con la información presente en las bases de datos de proteínas. Proporcionará una lista e información de las proteínas identificadas (y sus péptidos). Para ello realiza una digestión y tratamiento "in silico" (teórica) de las proteínas presentes en las bases de datos, de acuerdo con el tratamiento utilizado para preparar la muestra.

Típicos procedimientos de preparación de muestra para la identificación de proteínas:

- **Inmuno-depleción:** consiste en la separación (/eliminación) de las proteínas mayoritarias que no sean de interés. En la identificación de proteínas minoritarias, es clave que éstas no queden enmascaradas por otras muy mayoritarias; ello permitirá profundizar más en el proteoma, cuando se efectúa Auto MS/MS con el QTOF (u otro equipo). En el mercado existen distintos tipos de kits basados en inmunoafinidad, capaces de eliminar hasta más allá del 95% del contenido proteico total. Una interesante comparación entre varios sistemas fue publicada por Björhall y Davidsson, "Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples" (27). Vivanco y Barderas publican un caso práctico en "La depleción de las proteínas mayoritarias del plasma facilita su análisis proteómico en pacientes con síndrome coronario agudo" (28).

- **Fraccionamiento por pI y tamaño:** se suelen realizar mediante geles:

  - Monodimensionales (SDS-PAGE) para separar por tamaño.
  - Bidimensionales: separan por pI y tamaño.

  Si la muestra proviene de un gel con tinción, ésta se deberá desteñir; si proviene de un sistema "Offgel Electrophoresis Fractionator" no será necesario.

  - **"Offgel Electrophoresis Fractionator:** Esta es una herramienta muy útil para fraccionar tanto péptidos como proteínas intactas por su pI (que además se conocerá por el pocillo de donde proceda la muestra analizada); ayuda a aumentar dramáticamente el nº de proteínas identificadas, y da mejor rendimiento que el tradicional fraccionamiento por HPLC con intercambio catiónico (SCX); previo a la separación por fase reversa para realizar HPLC bidimensional. Ejemplos sobre su uso pueden encontrarse en los trabajos de:

    - Waddell y Tang en "Using a Complex Proteomics Standard for Comparison of Protein Identification Workflows" (29) donde lo comparan con diversas estrategias.

    - Mann y Ren en "Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis" (30).

    - Miller y Wenz en "Combining immuno-depletion, protein and peptide OFFGEL electrophoresis for the efficient fractionation of plasma prior to LC-MS analysis" (31) explican el potencial de esta técnica combinada con la de inmuno-depleción.

- **Desnaturalización:** la muestra se desnaturaliza mediante una reducción de los puentes disulfuro (típica con DTT: ditiotreitol), seguida de una alquilación (típica con IAA: iodoacetamida) para prevenir que estos se vuelvan a formar.

- **Digestión enzimática:** es la "última" etapa típica; la enzima más utilizada es la tripsina (calidad para secuenciación /modificada para evitar su auto digestión); suele proporcionar péptidos entre 6 y 20 AAS, ideales para su secuenciación por MS. La tripsina corta a la proteína en los residuos de Lisina (K) y Arginina (R), excepto si están seguidos

de Prolina (P) en la dirección del C terminal (/K-,/R-,\P). Una proteína 50KDa suele dar unos 30 péptidos trípticos. En ocasiones se utilizan otras enzimas para complementar la información proporcionada por la tripsina. Por ejemplo la GLU-C (V8-Proteasa) (/E-,/D*1-,\P) es especialmente útil para el estudio de proteínas con un alto contenido en Lys (K) y Arg (R), que con tripsina darían péptidos excesivamente pequeños (<6AAS), para ser suficientemente específicos para la identificación de la proteína.

- **Desalinizar:** dependiendo del tipo técnica de ionización o del tipo de cromatografía utilizadas, la muestra se debería desalinizar antes de su análisis por MS (usualmente con unas "zip-tip": puntas de micropipeta rellenas con C18).

En el mercado existen robots que permiten automatizar estos procesos de preparación de la muestra, incluyendo la digestión con tripsina y desalinización de la misma.

Las proteínas con su estructura primaria sin modificar ("nativas"), suelen coexistir con otras isoformas de la misma, fruto de modificaciones postraduccionales de la forma nativa. Para reducir el tiempo de procesado para la identificación de las diferentes isoformas de cada proteína presentes en la muestra, algunos motores permiten primero realizar una búsqueda en modo identidad; para identificar las proteínas nativas presentes en la muestra. Después con éstas, se crea una mini-base de datos; ésta se utilizará para evaluar rápidamente, mediante búsquedas en modo homología, todas la posibles mutaciones y modificaciones post-traduccionales que se consideren factibles, de todas las proteínas nativas identificadas (ver figura 26).

Para reducir el tiempo del proceso de búsqueda, algunos motores de búsqueda (p.e. Spectrum Mill) permiten utilizar, sólo aquellos espectros de MS/MS que proporcionen un mínimo de calidad; así se evita perder el tiempo con espectros que tienen demasiado ruido, o que no corresponden a péptidos; los criterios de filtrado son:

Agilent Spectrum Mill - Protein/Peptide Summary

Spectrum Mill | Summary Settings | Autovalidation | Build TIC | MS/MS Search | Spectrum Summary | Tool Belt | Help

Results Shown Filtered by Validation Category: valid
Data Directory: msdataSM\ExampleData\Agilent_LCMSD_Trap\ecoli\heatshock
hit table read - SpecFeatures read
valid hits read from tagSummary file - **Files:** 200 **Hits:** 667
beginning to assemble proteins ... proteins assembled
proteins filtered by unique peptides
proteins filtered by score
proteins rolled up into groups
protein groups ready for display 47 Proteins listed

Group (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Mean Peptide Spectral Intensity	Protein MW (Da)	Database Accession #	Protein Name
1	65	6	107.44	18	3.04e+007	57269.1	41617	GroEL
2	7	5	74.10	13	2.97e+007	69115.3	15799694	Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70)
3	9	3	49.91	7	4.19e+007	77581.7	15803853	Elongation factor G (EF-G)
4	4	3	42.17	8	3.50e+007	56231.1	15834105	Glycerol kinase (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK)
5	6	3	40.74	9	1.34e+007	48414.3	15830232	SeryltRNA synthetase (Serine–tRNA ligase) (SerRS)
6	12	2	40.02	15	6.44e+007	30093.9	10185998	malate dehydrogenase
7	5	3	39.84	15	6.46e+007	31076.5	16130747	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase (5-keto-4-deoxyuronate isomerase) (DKI isomerase)
8	14	2	32.11	7	6.86e+007	43313.8	15834157	Elongation factor Tu (EF-Tu) (P-43)
9	3	2	31.33	27	2.63e+007	10317.7	6176579	Hypothetical protein yfcZ
10	6	2	31.21	18	9.82e+007	20761.5	15800320	Alkyl hydroperoxide reductase C22 protein (SCRP-23) (Sulfate starvation-induced protein 8) (SSI8)
11	2	2	28.75	11	3.50e+007	39147.5	15833050	Fructose-bisphosphate aldolase class II (FBP aldolase)

Figura 26.-Resumen de Proteínas Identificadas con motor de búsqueda Spectrum Mill. La 9ª columna al abrir el desplegable mostrará para cada proteína identificada sus diversas isoformas detectadas. Las columnas 3ª y 5ª muestran el nº de diferentes péptidos detectados para cada proteína identificada y el porcentaje de cobertura de la secuencia de aminoácidos de la proteína. La columna 4ª indica la calidad de la identificación ("MS/MS search score"); este motor utiliza un algoritmo que asigna/descuenta puntos (en mayor o menor grado según el tipo de iones y, b, a, y-NH3,…), según los iones teóricos localizados/no localizados en el espectro de MS/MS de cada uno de los péptidos identificados de la proteína en cuestión (columna nº 9). "Scores" por encima de 10 para péptidos y de 25 para proteínas se pueden considerar como identificaciones fiables. Para una muy fácil comparación semi-cuantitativa en modo "label free" de niveles de expresión de las proteínas, en la columna 6ª se indica el promedio de intensidad de señal de los cromatogramas de iones del ión precursor, de cada uno de los péptidos identificados de la proteína en cuestión. Estas intensidades definen el código de color que indica el nivel de expresión ("label free") de la proteína; cada cambio de tonalidad en el color indica doblar la expresión.

- **Umbral de intensidad mínima** del espectro: para descartar espectros del ruido o de péptidos con respuestas demasiado bajas.

- **Número mínimo de residuos** detectados en el espectro de MS/MS: por ejemplo que al menos 2 diferencias de masa entre iones del espectro, coincidan con residuos de aminoácidos (aminoácido-H2O), para asegurarse que el espectro de MS/MS corresponde realmente a un péptido, y no a un contaminante.

3.2.1. Identificación de Proteínas mediante LC/MS-TOF.

La masa exacta del péptido protonado en polaridad positiva, o desprotonado en polaridad negativa, permitirá determinar la composición en aminoácidos de cada péptido, aunque no el orden de los aminoácidos en la secuencia; para ello se requiere disponer de un MS/MS (ver siguiente apartado 4.2.2). A este tipo de búsqueda por masas de la huella péptidica se le denomina PMF: "Peptide Mass FingerPrint".

Para determinar la masa del ión molecular, se deberá:

- Conocer el nº de cargas de los iones medidos en el masas. Se calcula por el inverso de la diferencia en masa, entre los distintos iones producidos por el ^{13}C en el "cluster" isotópico.

- Tener en cuenta el isótopo estable más abundante (^{12}C, ^{32}S) para ello se debe resolver la envolvente isotópica (debida a la contribución del ^{13}C, ^{34}S,…) para determinar la masa del péptido. En péptidos con más de 80-90 átomos de carbono el ión correspondiente a ^{12}C deja de ser el más intenso del "cluster isotópico", aunque mantiene el ser el de menor masa.

- También se podrán utilizar algoritmos de deconvolución que para su cálculo agrupen:

  - Iones del "cluster" isotópico.
  - Iones con múltiples cargas típicos con péptidos.
  - Iones formados por aductos con Na^+, $NH4^+$,… en positivo o acetato, formiato,… en negativo.
  - Pérdidas neutras muy lábiles (que se puedan formar "en la fuente"); por ejemplo de agua, amoníaco,….

  Estos cálculos con múltiples iones permitirán mejorar la exactitud.

Una búsqueda con espectros de MS con masa exacta (PMF: Peptide Mass FingerPrint), requiere un mínimo de 6-7 péptidos identificados, para una mínima garantía de una correcta identificación de la proteína. En cambio, si se dispone de espectros de MS/MS será suficiente con un número mucho menor (2-3).

3.2.2. Identificación de Proteínas mediante LC/MS-QTOF. Aportaciones de LC/MS/MS con Masa Exacta en la Identificación de Proteínas.

La importancia del espectro de MS/MS en la identificación de proteínas: al disponer de información del orden de aminoácidos en la cadena peptídica, con 2-3 péptidos identificados por MS/MS suele ser suficiente para identificar a una proteína con buena seguridad; incluso en muchas ocasiones, es suficiente con 1 sólo péptido, si se detecta alguno de los péptidos proteotípicos de la proteína. Todo ello es fruto de la elevada especificidad de la combinación de los 20 aminoácidos esenciales en un péptido de 10-20 aminoácidos. Para identificar proteínas minoritarias en muestras complejas, las numerosas coeluciones de sus péptidos con otros péptidos mayoritarios, sólo permiten obtener los espectros MS/MS de unos pocos péptidos

minoritarios (aquellos que eluyen en las zonas del cromatograma mejor resueltas), de ahí que el poder identificar una proteína detectando sólo 1 o 2 péptidos, sea de gran ayuda en la identificación de proteínas minoritarias.

Como se comentará en apartado 3.2.4, la especificidad de los espectros de MS/MS con masa exacta, mejorará muy considerablemente la fiabilidad con respecto a espectrómetros que sólo proporcionen espectros de MS/MS con masa nominal. Por ello, cada vez es mayor la tendencia a utilizar equipos de LC/MS con capacidad de adquirir a elevada frecuencia de adquisición, espectros de MS/MS con masa exacta (al menos 5-10 espectros MS/MS/segundo).

Modos de Adquisición con LC/MS-QTOF para la Identificación de Proteínas.
Los mejores modos y estrategias de adquisición de datos para identificar proteínas variarán según los objetivos del análisis:

- Identificación de Proteínas Mayoritarias en muestras "Desconocidas": si se desconocen que proteínas pueden estar expresadas, la adquisición de datos para identificarlas se realiza en modo Auto MS/MS (o "Data Dependent" MS/MS); especialmente si interesan las de mayor abundancia. En este modo (ver figura 15 en apartado 3.1.1.1), el equipo va adquiriendo en modo MS y, cuando detecta un ión con una intensidad de señal superior a un umbral (y cuya masa no esté en una lista de exclusión), en el siguiente ciclo de medida el equipo lo aísla (en Q1), fragmenta automáticamente (en la celda de colisión), y mide la masa e intensidad de los iones producto formados (en el TOF), proporcionando su espectro de MS/MS con masa exacta. Por su mayor información sobre la secuencia de aminoácidos del péptido, se recomienda activar la opción de escoger como iones precursores, preferentemente iones con múltiples cargas (más información en apartado 4.4.3).

- Identificación de Proteínas Minoritarias en muestras "Desconocidas": para identificar en muestras desconocidas, proteínas minoritarias cuyos péptidos coeluyan con otros de proteínas mayoritarias, se recurre a listas de exclusión (activas/automáticas limitando las repeticiones), y/o a sucesivas inyecciones excluyendo al tiempo de su elución, los precursores ya adquiridos en inyecciones previas. Otra alternativa, es adquirir en modo "targeted MS/MS", programando ventanas al tiempo de retención de cada uno de los iones precursores detectados en un cromatograma previo en modo MS. Se escogerán preferentemente iones precursores con 2 o más cargas.

  Para facilitar la detección de péptidos minoritarios, en el mercado existen algoritmos de deconvolución del mapa 3D (Tr, m/z, abundancia) que como "Molecular Feature Extractor", proporcionan la m/z de todos los péptidos ionizados y su espectro deconvolucionado ("limpio"), aunque estén coeluyendo con otros de muy mayoritarios (ver apartado 3.1.1.2 y Figura 21).

- Identificación de Proteínas "Conocidas": para confirmar la presencia de un listado de proteínas, se buscarán sus péptidos proteotípicos y dependiendo de si se conoce o no su tiempo de retención se optará por:

  1. Péptidos Proteotípicos CON Tiempo de Retención conocido: se programará un "targeted MS/MS" programando en el tiempo las m/z de los iones precursores (escogiendo iones con 2 o más cargas).

  2. Péptidos Proteotípicos SIN Tiempo de Retención: se programará un Auto MS/MS incluyendo como lista de inclusión de iones precursores, sólo las m/z (con 2 cargas/+3 para los de mayor tamaño) de los péptidos proteotípicos. Otra posibilidad que proporcionan algunos equipos, es incorporarlos en una lista de

iones precursores preferidos (en lugar de incluidos); la ventaja de la lista de preferidos, es la de dar preferencia a éstos aunque sean menos intensos, pero sin excluir al resto (como sucedería con los que no estén en la lista de inclusión).

- **Identificación de Modificaciones Postraduccionales (PMT's):** para la detección de las PMT's será clave el tener el mayor grado de cobertura, para asegurarse que "no se escape" justo el péptido modificado. Por ello la mejor alternativa es:

1. Realizar una adquisición en MS con masa exacta.

2. Deconvolucionar los datos para obtener la m/z (preferentemente) con 2 o más cargas), y tiempo de retención de TODOS los péptidos ionizados a utilizar como iones precursores.

3. Programar un targeted MS/MS de esos iones precursores programados a su tiempo de retención.

4. Identificar las proteínas nativas presentes (búsquedas en base de datos en modo identidad).

5. Identificar todas las modificaciones posibles sobre una base de datos creada con las proteínas nativas identificadas en la etapa anterior. Otra alternativa para buscar algunas PMT's como por ejemplo las fosforilaciones o glicosilaciones, sería buscar desde el software de análisis de datos, las típicas pérdidas neutras de las modificaciones de interés, con los MS/MS adquiridos de todos los péptidos. La pérdida neutra es la diferencia entre la masa exacta del precursor, y la del ión producto; en el caso de los fosfopéptidos suele ser 97,9769 /z (H3PO4/z) o 79,9663/z (H3PO4-H2O); z es el nº de cargas del ión precursor. Es importante disponer de la masa exacta de precursor y producto para evitar confusiones con pérdidas isobáricas (similares) como pérdidas de prolina (97.053/z) y valina (99.068/z), o entre lisina y glutamina (difieren tan sólo 36 mDa). Para más detalles ver publicación de Lehmann y Weckwerth: "Neutral loss-based phosphopeptide recognition: a collection of caveats"(32).

Una potente alternativa para caracterizar y profundizar en el estudio de determinadas modificaciones, como fosforilaciones, o N-glycosilaciones, es el uso de dispositivos como el PhosphoChip o el GlycanChip que permitirán aislar fosfopéptidos en un caso, y liberar y caracterizar los N-glicanos de proteínas intactas (especialmente anticuerpos monoclonales), en el otro. Para más detalles ver capitulo HPLC-Chip/MS y apartado 3.4.2 de este capítulo.

Fragmentación peptídica en celda de colisión.

La vibración producida al impactar con el nitrógeno o argón presentes en la celda de colisión, genera mayoritariamente la rotura de los enlaces péptidicos, entre el carbono del carbonilo y el nitrógeno adyacente, dando lugar a iones "b_x" (lado N terminal) e "y_x" (lado C terminal); el subíndice corresponderá al número de grupos carbonilo del ión (número de CO incluyendo el contenido en el grupo carboxílico –COOH) (ver figura 27).

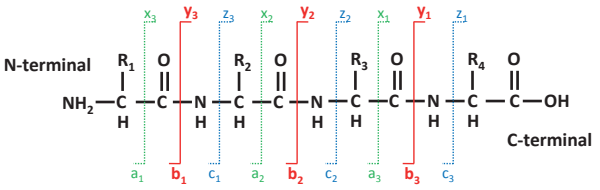


Figura 27.-Fragmentación de péptidos en celda de colisión: suele producir mayoritariamente iones "b_x" (lado N terminal) e "y_x" (lado C terminal).

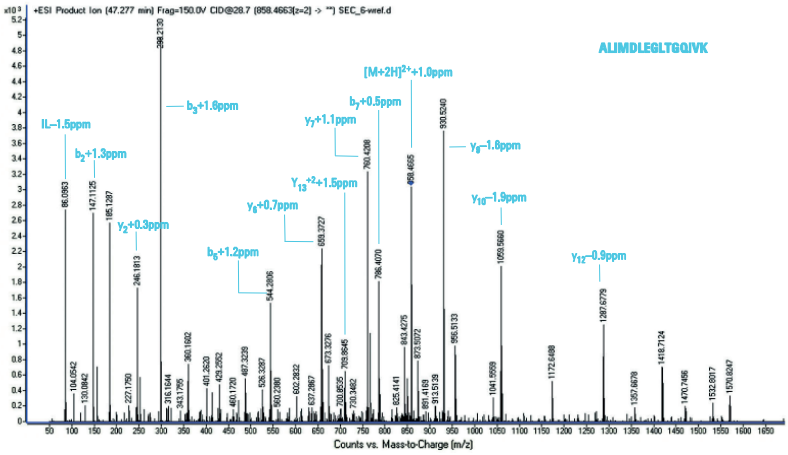


Figura 28.-Espectro MS/MS obtenido con un QTOF del péptido ALIMDLEGLTGQIVK doblemente protonado (2 cargas). Se puede observar una buena asignación de los fragmentos mayoritarios a iones b e y; en muchos casos se observan la detección de los 2 lados del enlace roto (iones bX e yX con la misma x), así como la detección de iones imonio a baja masa. La masa exacta de los iones producto reduce las posibilidades falsos positivos.

Para una correcta identificación de proteínas, se debe comprobar que los fragmentos mayoritarios del espectro de MS/MS, estén asignados a iones "b_X" (lado N terminal) e "y_X" (lado C terminal), como se puede observar en las figuras 28 y 17 (ésta última en apartado 3.1.1.1); caso contrario, la presencia de iones intensos sin asignar, podría indicar una modificación postraduccional (PTM) de la proteína. En este caso, convendría que el motor de búsqueda buscará todas las posibles modificaciones de las proteínas nativas identificadas, para mirar de asignar ese péptido, a alguna modificación (PMT) de las proteínas nativas identificadas.

¿Porqué fragmentar preferentemente iones con múltiples cargas?.

En LC/MS los péptidos al ionizarse por electrospray (ESI o nano-ESI) proporcionan iones con múltiples cargas; éstos tienen la ventaja de fragmentar dando a menudo 2 fragmentos con carga (detectables en el espectro de MS/MS). En cambio los precursores monocargados, sólo permiten detectar uno de los 2 fragmentos (el que se queda con la carga), el otro se deberá calcular como "pérdida neutra", por la diferencia entre la masa del precursor y la del producto que se queda con la carga. De ahí que, al proporcionar espectros de MS/MS con más información, sea preferible fragmentar péptidos que se hayan cargado con 2 o más protones (z). Éstos se reconocen por la distancia entre los iones del clúster isotópico (1/z).

3.2.3. Secuenciación "De Novo" de Péptidos mediante QTOF.

Como se comentará más en profundidad en otro capítulo, la secuenciación "De Novo" de péptidos consiste en la determinación de la secuencia de aminoácidos que contiene un péptido, a partir del estudio de su espectro de MS/MS. Este tipo de secuenciación, funciona bien para péptidos pequeños (menos de 15 aminoácidos), pero empieza a tener dificultades a partir de 15-20 aminoácidos. Se utiliza para caracterizar péptidos que no han podido ser asignados a ninguna proteína en las búsquedas en bases de datos de proteínas; algún motor de búsqueda (p.e. Spectrum Mill) incorpora la secuenciación "De Novo" entre sus herramientas.

A partir del esquema de fragmentación mostrado en la figura 27, intenta componer la secuencia de aminoácidos que mejor encaja con los iones producto detectados en el espectro de MS/MS (ver figura 29). Las diferencias de masa entre 2 miembros consecutivos de cada una de las series (b e y), indican los residuos de aminoácidos que componen el péptido; por ejemplo m/z b_8 – m/z b_7= 99,068, residuo que corresponde a Val (V). La serie de iones b (en azul) da la composición de izquierda a derecha, empezando por el lado amino terminal (coincidiendo con

la secuencia mostrada en la regla de la figura 29), mientras la serie y (en rojo) la da en sentido contrario, y suele ser más intensa al protonarse el N del enlace roto. **La masa exacta de los espectros de MS/MS proporcionados por un QTOF, será de gran ayuda para reducir la incertidumbre en la asignación de aminoácidos**, aunque no podrá diferenciar entre los isómeros leucina (Le) e isoleucina (iLe) (al tener ambos la misma masa exacta), si podrá diferenciar entre los 0.036Da que diferencian Lys y Gln.

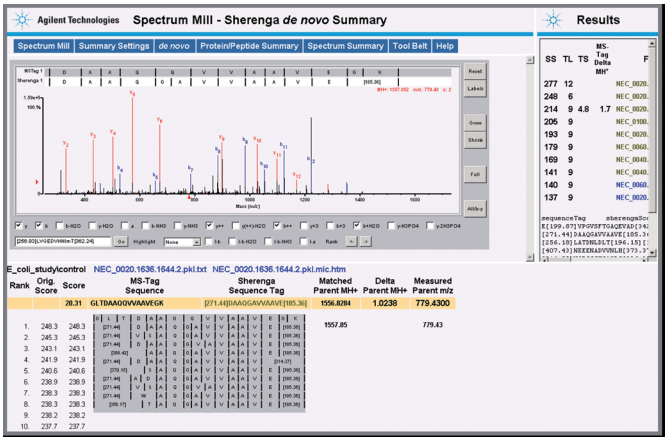


Figura 29.-Ejemplo de Secuenciación "De Novo" a partir de un espectro de MS/MS de un péptido, realizada por software Spectrum Mill. La comparación de la secuencia propuesta con el resto secuencias más probables clasificadas según su puntuación, permitirá evaluar que partes de la secuencia propuesta presentan mayor incertidumbre (ver parte inferior del gráfico).

3.2.4. Posicionamiento en Proteómica Identificativa de los Tiempos de Vuelo con respecto a otros Analizadores de Masa.

Para la identificación de proteínas se requiere de equipos con capacidad MS/MS y a poder ser con masa exacta. Si no se dispone de MS/MS, se requiere disponer como mínimo de masa exacta, y al no tener información del orden de los aminoácidos en la cadena peptídica, se deberá detectar un mínimo de 6-7 péptidos/proteína, para identificar con una mínima fiabilidad; por contra con MS/MS, serán suficientes 2-3 péptidos/proteína; incluso con 1 solo péptido proteotípico sería suficiente.

Aportaciones de la Tecnología QTOF en la Identificación de Proteínas y Posicionamiento "versus" Sistemas Basados en Trampas de Iones:

- **Elevada exactitud de masa, tanto en los iones precursores, como en iones producto**, aspecto que aumenta muy considerablemente el nº de proteínas identificadas y reduce el "False Discovery Rate", como ya se explicaba en el apartado 2.4 y en la tabla 2, que mostraba el fuerte impacto de la masa exacta en el nº de proteínas identificadas con una determinada fiabilidad. Las trampas de iones y los cuadrupolos no son capaces de proporcionar masa exacta, y la lentitud en la obtención de masa exacta de algunos sistemas híbridos basados en trampas de iones, sólo les permite disponer de la masa exacta del ión precursor, pero no de los iones producto.

- **Sensibilidad independiente de la coelución con otros péptidos mayoritarios:** al realizar MS/MS en el espacio (sin acumular iones), los cuadrupolos tienen un rango dinámico de filtrado muy superior al de las trampas de iones; éstas al hacer MS/MS en el tiempo (acumulando iones), pierden sensibilidad para péptidos minoritarios cuando éstos coeluyen con mayoritarios (por la limitación por exceso de repulsiones en el nº de cargas que se pueden acumular). Éste fenómeno no sucede con los QTOF, y aunque éstos tengan, en muestras bien resueltas cromatográficamente, algo menos de sensibilidad para péptidos no muy grandes, la sensibilidad se iguala (o incluso se puede llegar a mejorar) en muestras con coeluciones de péptidos minoritarios y mayoritarios. La

limitación en el nº de cargas a acumular, hace que para el análisis de proteínas intactas y grandes péptidos, con un elevado nº de cargas por ión, los sistemas basados en trampas de iones sean considerablemente menos sensibles que los QTOF.

- **Ausencia "Mínima Masa de Corte" ("Low Mass Cutoff") y Elevado Rango de Masa:** cuando se efectúa MS/MS los QTOF son capaces de registrar todo el rango de masas de iones producto, incluyendo iones de masa pequeña: iones imonio, iones b_x e y_x con x bajas, "reporter ions" cuando se efectúa ITRAQ, entre otros. En cambio, cuando se fragmenta en una trampa, ésta sólo es capaz de retener los iones producto con una m/z superior a 1/3 de la m/z del ión precursor, perdiéndose los pequeños iones indicados anteriormente. En algunos modelos de trampa es posible realizar una colisión pulsada, que permite registrar iones de pequeña masa, pero este modo reduce la sensibilidad a ~1/5 con respecto a la tradicional fragmentación CID (Disociación Inducida por Colisión). También se pierden prestaciones cuando en sistemas híbridos con masa exacta, se utiliza una celda de colisión para fragmentar, en lugar de utilizar la trampa.

- **Masa máxima:** otro aspecto a remarcar es la gran diferencia entre la masa máxima de los QTOF: m/z >~ 10.000 y la de los sistemas basados en trampas de iones e híbridos: m/z < ~ 2.000 (o 4.000 a muy pobre resolución típica ~2-3 m/z). En el análisis de proteínas intactas o péptidos muy grandes, se perdería la parte de la envolvente de distribución de cargas por encima de m/z ~2000, además de una pobre sensibilidad por la gran cantidad de cargas aportadas por cada ión.

- **Elevada Resolución Espectral:** para disponer de masa exacta se requiere de una mínima resolución espectral (típica >10.000), pero superado este umbral, éste es un aspecto menos importante que los puntos anteriores. Al realizarse la identificación de proteínas a partir de los espectros de MS/MS, es muy poco probable encontrarse con 2 péptidos que coeluyan, y que además compartan ión precursor e iones producto con masas isobáricas muy próximas.

¿Se Pueden Utilizar Cuadrupolos o Triple Cuadrupolos en la Identificación de Proteínas?

- **Cuadrupolos simples:** al no disponer de MS/MS ni de masa exacta, **no son adecuados para la identificación de proteínas**, si pueden utilizarse en cambio con electrospray, para determinar la masa de proteínas intactas con una exactitud típica < 0,02%. Ello es posible gracias al uso de softwares de deconvolución de los espectros con múltiples cargas, incluso cuando coeluyen varias proteínas simultáneamente.

- **Triple cuadrupolos:** en modo MS/MS podrían utilizarse, pero tienen el inconveniente de su menor sensibilidad y menor rapidez adquiriendo espectros completos, con respecto a las trampas de iones y los tiempos de vuelo; por ello apenas se utilizan para identificar. Al igual que los cuadrupolos simples, también podrán utilizarse para determinar el peso molecular de mezclas de proteínas intactas.

ESI-QTOF "versus" MALDI-TOF/TOF

- **Las fuentes de iones de electrospray (ESI) (o nanoelectrospray):** generan péptidos cargados con múltiples cargas; ello permite escoger iones precursores con múltiples cargas. Éstos al fragmentar suelen proporcionar 2 fragmentos cargados de los 2 lados de la cadena peptídica, y consecuentemente aparecen reflejados en el espectro de MS/MS. En definitiva los espectros MS/MS con ESI-QTOF proporcionan mayor información estructural que los obtenidos con MALDI-MS/MS.

- **Las fuentes MALDI** ("Matrix Assisted Laser Desorption Ionization"), producen iones monocargados, consecuentemente en sus espectros de MS/MS sólo se reflejará un lado de cadena peptídica, y consecuentemente, proporcionarán menos información sobre la secuencia del péptido, que un espectro MS/MS de ESI de iones con 2 o más cargas.

Además los sistemas ESI-QTOF o ESI-TOF suelen proporcionar una exactitud de masa apreciablemente mejor que la de los sistemas MALDI-TOF/TOF o MALDI-TOF; si además se quiere calcular la masa de las pérdidas neutras su error aumentará más (correspondería a la suma de errores de precursor e ión producto).

Las ventajas de los sistemas MALDI son: su rapidez, el proporcionar espectros de MS más simples (sólo iones con 1 carga) y el no estar limitado el tiempo de medida de un péptido, al ancho de su pico cromatográfico; pero esto a la vez se convierte en inconveniente, al no poderse acoplar en línea un MALDI a un HPLC. Así se pierde la enorme capacidad de simplificación que ofrece la cromatografía. Por ello es cada vez más frecuente, fraccionar previamente las muestras complejas, o bien por geles bidimensionales (separación por tamaño y por pI), o bien con cromatografía líquida mediante el uso de un colector de fracciones del tipo "spotter", que colecta las fracciones directamente sobre placas MALDI.

En los ESI-TOF/QTOF, en el caso de no querer tener la limitación de tiempo del ancho del pico cromatográfico, se puede introducir la muestra en continuo directamente al espectrómetro -sin separación cromatográfica-, mediante una jeringa infusión. Ello permitirá sobre la marcha y en tiempo real, ir realizando y almacenando, los múltiples experimentos que interesen de una manera interactiva.

3.3. Proteómica Cuantitativa.

Los TOF y QTOF actuales ofrecen excelentes prestaciones cuantitativas, aunque no ofrecen la sensibilidad y rapidez que triples cuadrupolos de gama equivalente (en modalidad MRM - ver apartado 3.3.4). A continuación se irán desgranando los distintos tipos de cuantificación, software y modos de adquisición.

Evolución e Importancia de la Proteómica Cuantitativa.

La proteómica cuantitativa cada vez está tomando más peso con respecto a la identificativa; ya no es suficiente saber que proteínas están presentes, sino que hay que conocer cómo cambian sus niveles de expresión, entre los diferentes tipos de población sujetas a estudio. Asimismo, a medida que aumenta el nivel de conocimiento de los proteomas, la proteómica cuantitativa relativa, no específica (o genérica), va migrando hacia una proteómica cuantitativa dirigida ("target"); en ésta ya se cuantifican péptidos concretos. Buena prueba de ello es la creación del SRM/MRM Atlas Humano por parte del grupo de R. Moritz (19:"Development of a Complete Human Peptide and SRMAtlas").

Cuantificación Absoluta y Relativa.

En proteómica la cuantificación suele ser relativa, es decir de unas muestras frente a otras; ello se debe a la dificultad en la disponibilidad de patrones de concentración conocida, para poder crear las curvas de calibrado para realizar una cuantificación absoluta. Las curvas de calibrado se construyen representando varias concentraciones del péptido, "versus" el área o altura del pico del cromatograma de iones específico del péptido. Si se utiliza 1 solo nivel de concentración, el resultado es una recta de calibrado que pasará por el origen de coordenadas (0,0); en definitiva es lo mismo que realizar una proporcionalidad directa entre intensidad de señal y concentración.

Cuantificación Absoluta mediante Dilución Isotópica.

Consiste en añadir a la propia muestra, una cantidad conocida de patrones marcados isotópicamente con ^{13}C y ^{15}N, de los mismos péptidos a cuantificar; éstos se cromatografian con-

juntamente con la muestra a la que se han adicionado (eluirán al mismo tiempo de retención pero con distinta masa); a partir del ratio de respuestas entre el péptido sin marcar (muestra) y el marcado, se podrá calcular la concentración absoluta del péptido, y consecuentemente de la proteína (si el péptido es específico/exclusivo de esa proteína). Hay firmas que ofrecen sintetizar péptidos a medida marcados isotópicamente y en los que 1 de sus aminoácidos está constituido en más de un 98% por ^{13}C y ^{15}N (en lugar de ^{12}C y ^{14}N).

Cuantificación Relativa con "Muestra Promedio" o muestra de referencia.

Una sencilla manera de cuantificar relativamente un conjunto de muestras, es crear una "muestra promedio" mediante un "pool" de alícuotas iguales de todas las muestras, que se utilizará como referencia. A partir de ahí, se crea una curva de calibrado con 1 solo nivel, y se le asigna por ejemplo, concentración 100 a todos los péptidos/proteínas. El resto de muestras se cuantificará a partir de ésta, obteniéndose directamente su concentración relativa, con respecto a esta muestra promedio utilizada como referencia. Otra opción es fijar una muestra de referencia concreta que sea estable para todo el estudio. Esta muestra de referencia utilizada, debe contener todos los péptidos que se quieran cuantificar; sino se deberán crear rectas de calibrado para todos los péptidos de interés utilizando varias muestras de referencia.

Cuantificación Relativa con Etiquetado Isotópico.

Consiste en marcar con distintas etiquetas isotópicas, péptidos de las distintas muestras a comparar. Después éstas se mezclan y analizan conjuntamente en un único análisis. La cuantificación se realizará por el ratio de respuestas de las parejas/tripletes/… de masas, asociadas por una diferencia concreta y exacta de masa; ésta diferencia en masa está asociada a las diferentes etiquetas isotópicas utilizadas. En el caso de cuantificación "targeted" (péptidos concretos), directamente se buscan las masas de interés; pero también admite cuantificación relativa en modo genérico ("non target"), buscando iones que difieran exactamente en determinadas masas. Se utiliza preferentemente en modo MS, y existen diversas técnicas: ICAT, SILAC, $^{18}O/^{18}O$,… que se describen en otro capítulo de este libro.

Cuantificación Relativa con Etiquetado Isobárico Multiplexado.

Se basa en marcar con distintas etiquetas, diversas muestras (p.e. 8 muestras) en el lado amino terminal de sus péptidos. Tiene la particularidad que el péptido, una vez marcado, a pesar de utilizar distintas etiquetas, tendrá la misma masa (etiquetado isobárico); ello permitirá utilizar el mismo ión precursor para las distintas muestras. Estas etiquetas de marcaje han sido diseñadas de manera que al fragmentar, el MS/MS del péptido marcado proporciona unos iones producto específicos de cada uno de las muestras (p.e. ITRAQ "reporter ions" en el rango m/z: 113 a 121); sus intensidades relativas indican la proporción relativa de ese péptido entre las distintas muestras analizadas en la misma inyección. Pandey y Waddell muestran un ejemplo en: (20: "4-Plex iTRAQ Based Quantitative Proteomic Analysis Using an Agilent Accurate-Mass Q-TOF"). La técnica admite tanto la cuantificación relativa de péptidos concretos (adquisición en "targeted MS/MS"), como la cuantificación relativa genérica ("non target") mediante Auto MS/MS ("Data Dependet MS/MS").

Softwares para Cuantificación.

La cuantificación mediante curvas de calibrado se podrá realizar con softwares cuantitativos genéricos, que se suministran con los espectrómetros de masas. Para cuantificaciones relativas "non targeted" utilizando etiquetado isotópico, se deberán utilizar motores de búsqueda o softwares específicos para proteómica, que incluyan esta opción.

Supresión /Exaltación de la Ionización.

Cada vez es mayor la tendencia a utilizar técnicas de cuantificación sin marcaje isotópico, dado su menor coste y mayor simplicidad. Un aspecto importante en la cuantificación, es la evaluación del impacto en la ionización y respuesta del péptido a cuantificar, de las variaciones en concentración de las especies coeluyentes al mismo tiempo de retención; su respuesta puede reducirse (supresión ionización) o amplificarse (exaltación). Por ello el uso de marcaje isotópico

con péptidos que eluyan al mismo tiempo de retención, es de gran ayuda para compensar estas variaciones en la respuesta. En cromatografía líquida el marcaje con ^{13}C suele funcionar mejor que el de deuterio (D); el D produce ligeros desplazamiento de los tiempos de retención debido a la distinta polaridad de los puentes de hidrógeno entre H y D.

Modos de Adquisición para Cuantificación con QTOF.
Para cuantificar se pueden utilizar diversos modos de adquisición, que aquí se introducen y posicionan brevemente, y en los siguientes apartados (3.3.1 y 3.3.2) y en 4.5.4 se desarrollaran un poco más:

- Cuantificación con Adquisición en modo MS: en términos generales es el más utilizado, por su mayor flexibilidad, universalidad y sensibilidad. Se adquirirá guardando en el disco duro los espectros de MS a lo largo de todo el cromatograma; por tanto se dispondrá de la información de todos los compuestos ionizados. Si es necesario, se podrán utilizar elevadas frecuencias de adquisición de datos; habitual hasta 20 espectros/segundo (incluso bastante más en algunos modelos); ello puede ser útil para trabajar en UHPLC con columnas cortas y rellenos de <2 µm de tamaño de partícula que generen cromatogramas muy rápidos, y picos muy finos.

- Cuantificación con Adquisición en modo "targeted MS/MS": en este modo se programa al tiempo de retención de cada precursor, una ventana de tiempo durante la cual se aislará y fragmentará el ión precursor, generando un conjunto de iones producto que darán lugar al espectro de MS/MS con masa exacta. Para detalles del proceso MS/MS ver figura 11 y apartado 2.3; para información sobre la fragmentación peptídica ver figura 27 hacia el final de la sección 3.2.2. La adquisición irá alternando el aislamiento de los distintos precursores que coeluyan al mismo tiempo. El modo "targeted MS/MS" tiene mayor selectividad, pero en general menor rapidez y sensibilidad que el modo MS. Ello se debe a la menor intensidad de los iones producto con respecto a la de los precursores; por ello suele ser más habitual cuantificar en modo MS, extrayendo cromatogramas de iones con muy finas ventanas de masa, que proporcionarán una selectividad equivalente a la de MS/MS, pero con la capacidad de poder cuantificar cualquier compuesto ionizado y además con una elevada frecuencia de adquisición de datos.

  Cuantificación con Adquisición en modo Auto MS/MS: a excepción de cuando se usa etiquetado isobárico multiplexado, para cuantificar no se recomienda utilizar el modo Auto MS/MS (combinación automática de espectros de MS y MS/MS); al variar las condiciones de adquisición dependiendo de lo que el equipo se va encontrando en cada inyección. Si se cuantifica en Auto MS/MS utilizando la dimensión de MS, se deberá asegurar un nº suficientemente elevado de puntos de muestreo a lo largo del pico, para que el perfilado del mismo, no se vea afectado por el cambio en el nº de puntos de muestreo de una muestra a otra. Por ello se debería adquirir un reducido nº de espectros de MS/MS entre cada espectro de MS, y/o asegurar que, en el peor de los casos, se toman al menos 12-15 puntos/pico en la dimensión de MS (8-10 en el caso de trazas).

Optimización Frecuencia de Adquisición y Relación Señal/Ruido.
Una buena cuantificación, además de una buena relación señal/ruido, exige que a lo largo del pico se tome un nº suficiente de puntos de medida; al menos 12-15 puntos/pico, que se reducen a 8-10 en el caso de trazas (mayor frecuencia de adquisición implica peor sensibilidad). Ver apartado 4.5.4 para más información.

3.3.1. Cuantificación mediante LC/MS (modo MS) con masa exacta (TOF o QTOF).

La gran ventaja de los TOF y QTOF en proteómica cuantitativa, es su capacidad de **cuantificar en**

modo MS con masa exacta cualquier péptido ionizado, con una muy elevada selectividad y buena sensibilidad, y sin que "a priori" se tenga que conocer los péptidos que interesan cuantificar (tal como sucede cuando se cuantifica por "targeted MS/MS"). Al adquirir en modo MS con masa exacta, se dispondrá en el disco duro de toda la información necesaria para "a posteriori" poder evaluar en todo momento el nivel de expresión de cualquier compuesto ionizado, sin necesidad de tener que volver a inyectar la muestra. Esta característica permitirá cuantificar las más complejas muestras, aunque coeluyen decenas de compuestos al mismo tiempo de retención.

La cuantificación se efectuará extrayendo del fichero de datos tridimensional (Tiempo de retención /Masa exacta / Intensidad señal) un cromatograma de iones a la masa exacta del péptido, con una muy fina ventana de extracción (unas pocas decenas de mDa); el proceso es fácilmente automatizable y permite extraer automáticamente los cromatogramas de iones de centenares/miles de péptidos; basta conocer su fórmula molecular ($C_xH_yN_zO_sS_t$….) o su masa exacta. Para aumentar la intensidad de señal, se puede extraer un cromatograma de iones que sume las intensidades de las masas más intensas del clúster del péptido en cuestión: contribuciones isotópicas, iones con múltiples cargas,…. Esta opción disminuye algo de selectividad, al aumentar las probabilidades de interferencias de iones de otros péptidos con casi la misma masa exacta; no obstante, estas interferencias son poco probables si se combina separación cromatográfica y resoluciones espectrales superiores a 10.000, capaces de resolver iones que difieran en tan sólo en unas pocas centésimas de masa (m/z).

Cuantificación relativa: este modo de adquisición se podrá utilizar entre otras, con las técnicas de etiquetado isotópico: **ICAT, SILAC, $^{16}O/^{18}O$** que se describen en otro capítulo de este libro. Con estas técnicas, las 2 muestras a cuantificar se tratarán con reactivos marcados con distintos isótopos (ligero/pesado), una vez marcadas isotópicamente se juntan ambas muestras y finalmente se cromatografian conjuntamente. La cuantificación se calcula por la relación de respuestas entre las intensidades de señal de los iones marcados con el isótopo ligero, y los marcados con el isótopo pesado. Cuando interesa la cuantificación masiva de péptidos, algunos motores de búsqueda permiten combinarla con la identificación de proteínas (MS/MS o MS exacta); el motor buscará en los espectros de masas de la dimensión MS (no MS/MS), parejas de iones que difieran exactamente en la diferencia de masa en cuestión (pesado-ligero), y considerará que esta diferencia se reducirá con el inverso del número de cargas de los iones. En la figura 30 puede verse un ejemplo de resultados de proteínas identificadas y cuantificadas relativamente mediante ICAT.

Group (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Mean Peptide Spectral Intensity	ICAT light/heavy (mean)	ICAT light/heavy (std dev)	ICAT light/heavy (# pairs)	Database Accession #	Protein Name
1	14	8	130.27	15	5.61e+007	0.70	0.15	6	4557871	transferrin
2	6	4	54.04	9	3.53e+007	0.95	0.31	3	1351907	bovine serum albumin
3	1	1	20.75	9	1.15e+007	0.00	0.00	0	70561	Myoglobin
4	2	1	15.70	2	5.37e+007	0.81	0.00	1	71826	fibrinogen beta chain
5	1	1	13.77	13	1.45e+007	0.59	0.00	1	115114	Aprotinin
6	2	1	13.73	3	6.47e+007	1.02	0.00	1	129293	ovalbumin
Totals:	26	16								

Figura 30.-Ejemplo de Proteínas Identificadas y Cuantificadas Relativamente mediante ICAT utilizando motor de búsqueda Spectrum Mill. En la 7ª columna se muestra para cada proteína identificada el promedio del ratio de respuestas entre los péptidos marcados con el isótopo ligero y el pesado. La columna 8ª muestra la desviación standard de ese promedio de ratios entre los diversos péptidos marcados de la misma proteína. En la columna 6ª se indica el promedio de la intensidad de señal de los cromatogramas de iones del ión precursor de cada uno de los péptidos identificados de la proteína en cuestión (columna nº 12); estas intensidades definen el código de color que indica el nivel promedio de expresión ("label free") de la proteína en cuestión independientemente de su etiquetado. En las columnas 3 y 9 puede observarse respectivamente para cada proteína el nº de péptidos identificados y el nº de péptidos para los que se ha localizado la pareja ligero/pesado.

3.3.2. Cuantificación mediante LC/MS/MS con masa exacta (QTOF).

Como se comentó anteriormente, el cuantificar en modo MS/MS proporciona una mayor selectividad, y consecuentemente mayor robustez (menor probabilidad de interferencias por otras masas isobáricas), pero en general menor rapidez y sensibilidad dada la menor intensidad de

los iones producto con respecto a la de los precursores; la pérdida de señal que no siempre queda compensada por la reducción del ruido (habitual al hacer MS/MS).

A diferencia de los triple cuadrupolos en modalidad MRM, los QTOF siempre proporcionarán espectros de MS/MS completos con masa exacta, y por tanto se dispondrá de todas las transiciones generadas por el ión precursor para ser utilizadas para cuantificar (p.e. sumando las más intensas) y confirmar la identificación. La cuantificación por LC/MS/MS-QTOF se realiza en el modo "targeted MS/MS", excepto en el caso del etiquetado isobárico multiplexado, que admite también el modo Auto MS/MS, al utilizar para cuantificar el ratio de determinados iones producto ("reporter ions") del mismo ión precursor.

Fragmentación de iones precursores con múltiples cargas: la presencia en un espectro de MS/MS de iones producto con m/z superior a la del precursor, confirma que se ha fragmentado un ión con múltiples cargas; aspecto que contribuye a confirmar que se trata del espectro de un péptido, y no de una pequeña molécula orgánica (no peptídica), de las que habitualmente aparecen a lo largo de la cromatografía; estas suelen generar sólo iones monocargados, y por tanto en sus espectros de MS/MS no aparecen iones por encima del precursor.

Etapas en el Proceso de Cuantificación por MS/MS:

- **Selección del péptido:** de los posibles péptidos para cuantificar una proteína, se deberán escoger, a ser posible, péptidos proteotípicos no muy grandes; es decir péptidos exclusivos de la proteína de interés, y que no puedan provenir de otra (excepto de modificaciones post-traduccionales de la misma). Algunos motores de búsqueda y softwares pueden ayudar a seleccionar este tipo de péptidos. Los péptidos de mayor tamaño suelen ser más específicos, pero en general suelen proporcionar menor sensibilidad (el péptido se divide en un mayor número de especies protonadas).

- **Selección de iones precursores:** de los posibles iones precursores del péptido de interés se escogerá preferentemente, el más intenso de los que presenten múltiples cargas (a menos que presenten intensidades mucho más débiles que los monocargados).

- **Optimización de condiciones de fragmentación:** sin contar con parámetros relacionados con la ionización del péptido (pH fase móvil, flujos, presiones y temperaturas de gas de secado/desolvatación) y captación de iones (voltaje del capilar), esencialmente para cuantificar se deberán optimizar:

  - **El voltaje de cono (o fragmentador):** optimiza la transmisión del ión precursor en la zona donde empieza el vacío. Habitualmente se utiliza el valor de defecto para péptidos que indique el fabricante. Para maximizar la señal del precursor, su optimización suele consistir en irlo aumentando hasta que empiece a fragmentar el precursor (es cuando suele alcanzarse su máxima respuesta).

  - **La energía en la celda de colisión:** optimiza la fragmentación del precursor y formación de iones producto en la celda. Este parámetro es péptido dependiente; la energía óptima dependerá del tamaño del péptido y de la labilidad de sus enlaces. En general se suele ir subiendo hasta que buena parte del precursor ha fragmentado (/desaparecido); no obstante puesto que en una celda de colisión los iones producto también fragmentan, la optimización debe realizarse monitorizando la intensidad individual de los iones producto formados, para optimizar la respuesta de los más intensos.

  Estos parámetros son modelo dependientes, y no se podrán dar pautas, sólo que la energía de colisión es mucho más importante que el voltaje de cono. En términos

generales, especialmente en compuestos de la misma familia o con estructuras semejantes y el mismo número de cargas, la energía de colisión óptima para fragmentar se incrementa linealmente con la m/z del ión precursor. Los fabricantes suelen proporcionar fórmulas, en función de la carga, para calcular la energía de colisión recomendada para utilizar en experimentos Auto MS/MS, a partir de la m/z del ión precursor. Para optimizarla (experimentos "targeted MS/MS") se probarán valores de voltaje alrededor del proporcionado por la fórmula.

- **Selección de los Iones productos**: un QTOF siempre proporciona un espectro completo en el rango de masa seleccionado; por tanto una vez adquiridos los datos, para cuantificar, se podrán utilizar todos aquellos iones producto de intensidad elevada que interesen.

3.3.3. "Screening" masivo de péptidos.

Sin duda una de las capacidades más potentes que ofrece una masa exacta robusta e independiente de la concentración del analito, es la de poder detectar selectivamente cualquier péptido que se haya ionizado. Ello abre la puerta a poder realizador un cribado ("screening") masivo de centenares/miles de péptidos; dependiendo del software que se disponga, este proceso se puede automatizar muy fácilmente, por ejemplo creando una hoja de cálculo con las fórmulas moleculares (o las masas exactas) de los péptidos de interés.

En lugar de ver si están presentes los péptidos de una lista, otra posibilidad es detectar todos los péptidos ionizados mediante herramientas de deconvolución cromatográfica (ver sección "Proceso Deconvolución Mapa 3D" en apartado 3.1.1.2).

3.3.4. Posicionamiento en proteómica cuantitativa de los Tiempos de Vuelo con respecto a otros analizadores de masa.

Todos los espectrómetros de masas modernos permiten cuantificar, de unos a otros variará su sensibilidad, reproducibilidad, rango dinámico, y rapidez para cuantificar muchos compuestos coeluyentes. De todos los espectrómetros, los triple cuadrupolos son los que proporcionan mejores prestaciones cuantitativas.

QTOF "versus" Triple Cuadrupolos:

- **Sensibilidad:** los QTOF son del orden de unas 20 veces menos sensibles que los QqQ en MRM de gama equivalente, pero los QTOF siempre proporcionarán el espectro completo, y en el caso de MS/MS todos los fragmentos del ión precursor (los QqQ sólo los programados). Los QTOF actualmente más sensibles del mercado, con límites de detección de bajos femtogramos, proporcionan niveles de sensibilidad superiores a los de triples cuadrupolos de gama media.

- Rapidez para detectar múltiples compuestos coeluyentes:

  - **En modo MS:** los TOF/QTOF actuales en modo MS son capaces de llegar a frecuencias de adquisición de 20-50Hz, y dada la selectividad y especificidad que puede proporcionar la masa exacta, podrán cuantificar un nº "ilimitado" de compuestos coeluyentes.

  - **En modo MS/MS:** la frecuencia de los QTOF se reduce a 5-30 Hz espectros MS/MS/seg. (dependiendo de los modelos), frecuencia que se ha de dividir por el nº de distintos precursores que coeluyan, y a lo largo del pico cromatográfico se deberán

adquirir un mínimo de 10 a13 ciclos de medida; por el contrario los QqQ en modalidad MRM son capaces de llegar a 100-300Hz, y de cuantificar decenas de compuestos coeluyentes.

- **Rango dinámico y reproducibilidad:** el rango dinámico de los QqQ es de 5-6 órdenes de magnitud, por 3-5 de los TOF/QTOF. La reproducibilidad irá muy ligada a la relación señal ruido:

  RSD% ~ 100 / (S/N) (ruido pico a pico)

  de ahí que a nivel de trazas la mayor sensibilidad de los triples le proporcionen unos mejores coeficientes de variación.

- **QTOF "versus" Trampas e Híbridos con masa exacta:** los QTOF proporcionarán una sensibilidad menos entorno dependiente, y una mayor velocidad de adquisición, tanto en MS como en MS/MS. Adicionalmente su resolución espectral no se verá afectada por la frecuencia de adquisición. En cuanto a la linealidad y reproducibilidad, esta será superior en los QTOF.

3.4. Caracterización de péptidos y proteínas intactas mediante LC/MS con masa exacta (TOF o QTOF).

La caracterización por LC/MS de una proteína intacta conlleva, no sólo el determinar su peso molecular con exactitud, sino sobre todo, entre otros muchos aspectos, determinar:

- Su secuencia de aminoácidos.
- Las diversas isoformas que coexisten conjuntamente con la proteína nativa.
- La localización de modificaciones postraduccionales.
- Si la proteína está glicosilada definir la estructura de los glicanos que la componen.
- Si está formando complejos con otras proteínas y/o otros compuestos químicos.
- …..

En definitiva la completa caracterización de una proteína, requerirá realizar diversos ensayos complementarios para poder finalmente tener una visión detallada de la misma.

3.4.1. Determinación del peso molecular de péptidos y proteínas intactas.

Los múltiples grupos fácilmente protonables de péptidos y proteínas, darán lugar a la formación de iones con múltiples cargas al ionizarse por electrospray. El nº de cargas positivas aumentará con el tamaño del péptido o proteína, y con la reducción del pH. La deconvolución de sus espectros de masas permite calcular su masa exacta con una elevada exactitud (bajas ppm de error), al calcularse con múltiples iones (33: "Accurate-Mass LC/TOF-MS for Molecular Weight Confirmation of Intact Proteins"); ello posibilitará detectar modificaciones postraduccionales y sustituciones de aminoácidos, así como realizar detallados seguimientos de reacciones, en las que intervengan péptidos o proteínas.

Para determinar las masas exactas de péptidos y proteínas, incluso en mezclas complejas, en el mercado existen diversos tipos de algoritmos de deconvolución de espectros de masas que contienen iones con múltiples cargas, unos deconvolucionan espectros puntuales (a un determinado tiempo o promedio de una franja de tiempo), y otros, deconvolucionan los espectros a lo largo del tiempo analizando como suben y bajan las diversas masas a lo largo del cromatograma para agrupar las que corresponden al mismo pico cromatográfico. Existen softwares que permiten combinar diversos de estos algoritmos según el tipo de análisis realizado.

- **Algoritmos de deconvolución espectral de un espectro de masas puntual** (deconvolución 2D: masa – abundancia), permitirán pasar de un espectro que muestra el conjunto de todas las diversas distribuciones de iones con múltiples cargas, de los péptidos o proteínas que coeluyen en una franja de tiempo determinado, a un espectro con carga cero que representará cada proteína o péptido coeluyente, directamente por su peso molecular y abundancia; éste permitirá fácilmente la detección de las diferentes isoformas que coexisten con la forma nativa de la proteína tal como puede observarse en la figura 31. Este tipo de algoritmos se pueden aplicar a mezclas relativamente simples de proteínas o péptidos. Ejemplos de este tipo de algoritmos son los "Maxima Entropia" y el "Resolved Isotope Deconvolution".

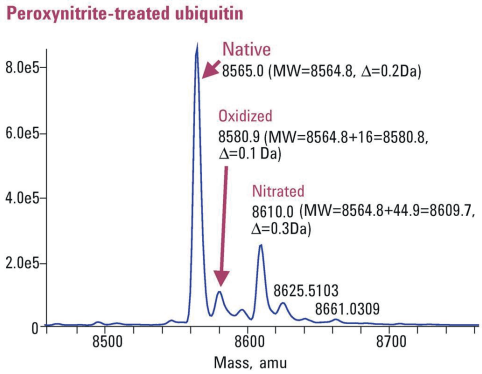


Figura 31.-Resultado del proceso de deconvolución espectral de la proteína intacta Ubiquitina (analizada con un TOF de resolución 10.000); además de la forma nativa se puede observar la presencia de diferentes isoformas incluyendo las formas oxidada y nitrada.

- **Algoritmos de deconvolución de espectros de masas a lo largo del tiempo** (deconvolución 3D: tiempo – masa – abundancia), permitirán la obtención de los espectros individuales (deconvolucionados) de las diferentes proteínas o péptidos aunque estas/os coeluyan con otras. Este tipo de algoritmos se pueden aplicar a mezclas complejas de proteínas o péptidos, especialmente los algoritmos que trabajan con masa exacta. Ejemplos de este tipo de algoritmos son "Molecular Feature Extractor" (MFE), "Large Molecular Feature Extractor" (LMFE), "Find by Chromatogram Deconvolution", "Dissect" y "AMDIS".

El utilizar uno u otro algoritmo, dependerá de la complejidad de la muestra, de si se trata de proteínas o péptidos, de si disponemos o no de masa exacta, y también del software disponible.

- **Algoritmos de Deconvolución Proteínas Intactas:** los denominados de "máxima entropía", requieren previamente especificar el rango de masas esperado, y sirven para mezclas no excesivamente complejas; para ello se suele recurrir a crear varios espectros promedio de diversas franjas a lo largo del cromatograma. Otros algoritmos, como el "MassHunter Large Molecular Feature Extractor" (LMFE) requieren de datos de masa exacta (TOF o QTOF), pero no requieren especificar el rango de masa, puede ser aplicado a complejas mezclas de centenares de proteínas intactas, y permiten rápidamente deconvolucionar directamente el cromatograma completo; éste último (LMFE) suele reducir considerablemente el tiempo de procesado e incrementar el número de proteínas identificadas.

- **Algoritmos de Deconvolución de Péptidos o Proteínas Digeridas:** para péptidos de hasta unos 4KDa y en espectros puntuales de mezclas peptídicas simples, se puede aplicar algoritmos de los denominados "Resolved Isotope Deconvolution"; éstos requieren una buena resolución del perfil de distribución isotópica del péptido (debido a la contribución de los isótopos: ^{13}C, ^{34}S,…). Dependiendo de la resolución del TOF/QTOF (típica entre 10.000 y 40.000) y del número de cargas que adquieran se pueden llegar a resolver los perfiles isotópicos de pesos moleculares de hasta 8-20KDa. En el caso de muestras complejas, y si se dispone de masa exacta, se puede utilizar un algoritmo (de deconvolución del cromatograma): tipo "Molecular Feature Extractor" (MFE) activando las opciones para péptidos (ver más detalles en apartado 4.5.1.2).

3.4.2. Caracterización de la Composición de Proteínas Intactas y determinación de Modificaciones Postraduccionales en proteínas intactas.

Mientras péptidos pequeños (<15 aminoácidos) pueden secuenciarse a partir de su espectro MS/MS, el espectro de MS/MS de proteínas es demasiado complejo para poder ser utilizado para una secuenciación "De novo". Para confirmar la secuencia de aminoácidos de una proteína, se suele recurrir a su digestión enzimática y posterior análisis del correspondiente mapa peptídico. A partir de ahí la estrategia analítica dependerá de:

- A.- es una proteína expresada por un organismo secuenciado, pero se desconoce su posible secuencia.

- B.- se dispone de información de la teórica secuencia (proteínas sintéticas, recombinantes, búsquedas previas..).

Caso A- Proteína expresada por un organismo secuenciado.

Utilizando motores de búsqueda y tal como se describió en el apartado 3.2 (identificación de proteínas), se recurrirá a identificar la proteína realizando búsquedas en bases de datos de proteínas que incluyan al organismo/taxonomía en cuestión. Mediante búsquedas en modo identidad se podrá determinar si la proteína en cuestión es una proteína nativa, y mediante búsquedas en modo homología la asignación de sus posibles isoformas evaluando el impacto de modificaciones en la secuencia nativa. Si la cobertura de la secuencia de la proteína con los datos de LC/MS no es suficientemente elevada (lo ideal sería conseguir el 100%) se podría:

- Mirar de localizar los péptidos perdidos en una adquisición en Auto MS/MS, a partir de la dimensión de MS con masa exacta (extrayendo los cromatogramas de iones con las m/z de los péptidos perdidos); su secuencia se podría confirmar después con un targeted MS/MS a su tiempo de retención.

- Recurrir a la segunda estrategia (caso B) a partir de la secuencia propuesta definida por la búsqueda en la base de datos.

Si la identificación no ha detectado ningún péptido proteo-específico (exclusivo de esa proteína) se deberían evaluar las diversas secuencias de proteínas posibles (caso B).

Caso B- se dispone de información de la teórica secuencia de la proteína.

(Proteínas sintéticas, recombinantes, búsquedas previas,...): existen softwares que permiten definir o importar la secuencia de aminoácidos del péptido o de la proteína, incluyendo modificaciones conocidas; y a partir de la comparación con su digestión "in silico" (teórica), confirmar si los datos experimentales concuerdan plenamente con los teóricos; así se podrán identificar impurezas, variantes no deseadas, modificaciones postraduccionales (aquellos péptidos que no se pueden asignar), y comprobar que se cubre toda la secuencia de la proteína o del péptido.

Detección de Modificaciones Postraduccionales en proteínas intactas:

La detección de modificaciones postraduccionales puede realizarse a través de distintas vías:

- **A partir del espectro de MS deconvolucionado de la proteína intacta:** como ya se comentó en el apartado anterior (3.4.1) y en la figura 31, la deconvolución del espectro de MS de una proteína intacta, puede revelar la presencia de diversas isoformas producto de modificaciones postraduccionales o de su degradación (p.e. oxidaciones). En la figura 32 se observa el espectro de masas obtenido con un TOF de un anticuerpo intacto (IgG4), sin deconvolucionar y deconvolucionado. La envolvente de iones con múltiples cargas de este anticuerpo cubre un rango de m/z de 2000 a 4000 m/z, correspondiente

a un peso molecular (ver espectro deconvolucionado) de 147KDa, con diferencias de 162.1 Da entre las isoformas mayoritarias, correspondientes a distinto nº de hexosas (34: "Antibody Analysis by ESI-TOF LC/MS").

- **A partir de la evaluación de las posibles modificaciones a la secuencia teórica de la proteína intacta:** existen softwares que permiten buscar automáticamente en un fichero de LC/MS, las posibles modificaciones definidas por el usuario, con respecto a una secuencia concreta de una proteína. El software mirará de detectar en el fichero de datos, aquellos péptidos que generarían las modificaciones en cuestión (ver Figura 33).

- **A partir de digestiones de proteínas expresadas por organismos secuenciados:** realizando búsquedas en modo homología en bases de datos de proteínas, como se comentó en el apartado 3.2.2.

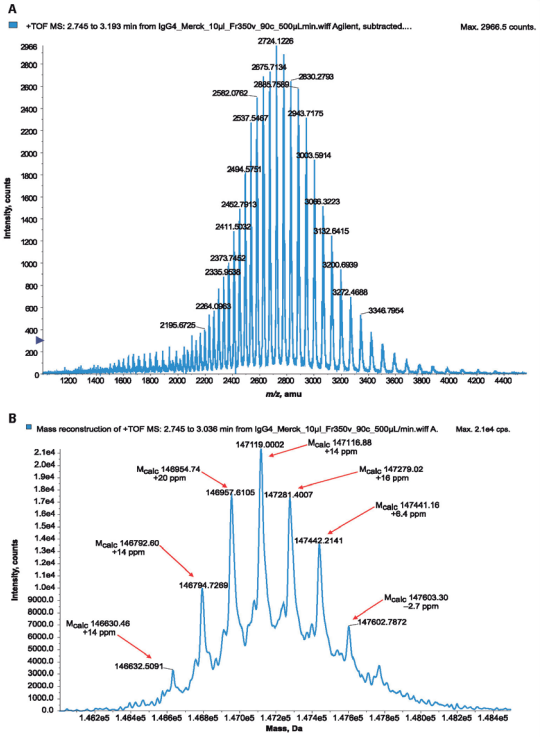


Figura 32.-Resultado del proceso de deconvolución espectral de una proteína intacta. Ejemplo de análisis del anticuerpo intacto IgG4 intacto mediante LC/MS-TOF (Agilent 1200-LC/6210-TOF con columna Poroshell C8 – 300Å). Gráfico A: Espectro de masas (electrospray) sin deconvolucionar de IgG4 intacto. Gráfico B: Espectro Deconvolucionado; éste muestra diferencias de 162.1 Da entre las isoformas mayoritarias correspondientes a distinto nº de hexosas.

- **Detección específica de modificaciones postraduccionales concretas:** en otro capítulo de este libro se presentarán procedimientos muy diversos, que pueden ser de utilidad para reconocer determinadas modificaciones concretas; como por ejemplo la detección de determinadas pérdidas neutras cuando se realiza MS/MS. Aquí simplemente resaltar que existen algunas de ellas (Fosforilaciones y N-Glicosilaciones), cuya

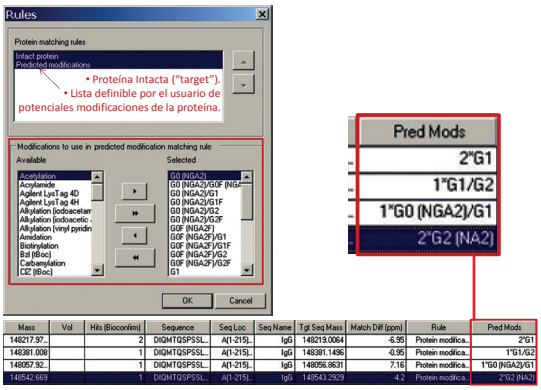


Figura 33.-Ejemplo de definición y resultados de búsqueda automática de modificaciones postraduccionales con software de "Bioconfirmación". El usuario puede seleccionar que modificaciones de la secuencia predefinida de la proteína, quiere que el software busque automáticamente en el fichero de datos; también puede definir nuevas modificaciones. En este ejemplo se han buscado todo un conjunto de glicoformas estándar; se muestran los resultados de la localización de 4 de ellas- ver última columna "Pred Mods".

detección se puede llegar a automatizar mediante el uso de dispositivos de microfluídica, que serán comentados en el capitulo 14 dedicado a la tecnología HPLC-Chip/MS:

- **Fosforilaciones** el uso de una precolumna con óxido de titanio permite preconcentrar y separar, los péptidos fosforilados de los no fosforilados. 1 inyección permitirá obtener 2 cromatogramas:

  - el primero sólo con los péptidos no fosforilados.
  - el segundo con sólo péptidos fosforilados (habitualmente mucho menos concentrados que los no fosforilados dada su muy desfavorable estequiometria); ello permitirá aumentar muy considerablemente el nº de péptidos fosforilados detectados.

- **N-glicosilaciones:** éstas son muy importantes para la caracterización de anticuerpos monoclonales (mAB's - ver final del próximo apartado 3.4.3). Mediante el uso de una precolumna con una enzima inmovilizada, será posible liberar los N-glicanos del mAB, y separarlos mediante una columna de grafito para su caracterización (incluyendo la separación de isómeros que tienen la misma masa exacta).

3.4.3. Caracterización de proteínas recombinantes, anticuerpos monoclonales (mAbs).

En la industria farmacéutica el uso de biofármacos va adquiriendo cada vez mayor importancia; se espera que en el año 2014 representen el 50% de los 1000 fármacos más importantes, proporcionando una buena oportunidad laboral para gente con experiencia en proteómica. Entre las diversas familias de biofármacos se pueden citar:

- Anticuerpos monoclonales (mAB's).
- Proteínas recombinantes, insulina, erytropoeitina (EPO).
- Péptidos como hormonas / antibióticos.
- Citokinas.
- Oligonucleótidos.

De entre todos estos destacan los anticuerpos monoclonales, por su elevada especificidad en sus interacciones con receptores diana, involucrados en una determinada patología; esta especificidad les permite reducir sus efectos colaterales. Se empezaron a aplicar en 1986 para reducir el rechazo en el trasplante de órganos. Hoy en día hay unos 30 aprobados para el tratamiento de distintas enfermedades, y unos 150 en fase de ensayos clínicos, para mejorar terapias en oncología, enfermedades respiratorias, cardiovasculares y oftalmológicas entre otras muchas (35:"Glycosylation as a strategy to improve antibody-based therapeutics". 36: "Monoclonal Antibodies as Innovative Therapeutics". 37: "A New Kind of Cancer Warfare Agent. Next Generation Monoclonal Antibodies").

En el 2011 se les estima un mercado de 26 billones de dólares en aplicaciones clínicas y de diagnóstico. Estas glicoproteínas terapéuticas suelen fabricarse a partir de líneas celulares de mamíferos (hibridomas), después se purifican, concentran y se preparan en una solución tampón adecuada, antes de ser vendidos. Los mAbs a pesar de ser biomoléculas relativamente estables, durante su fabricación, formulación y almacenamiento, pueden sufrir una serie de modificaciones químicas y reacciones de degradación, que requieren métodos fiables y sensibles, para la caracterización de la pureza de la proteína y su integridad estructural. El acoplamiento LC/MS TOF o QTOF proporciona una muy potente y sensible técnica para abordar estos análisis.

Para la caracterización de biofármacos se deberán utilizar diversas técnicas. Además de la secuencia primaria de los aminoácidos que la componen (38: "Faster, More Accurate Characterization of Proteins and Peptides with MassHunter BioConfirm Software"), se deberá controlar la formación de agregados, de productos de descomposición, isoformas inesperadas,…. Por

ejemplo, la formación de agregados puede provocar en el organismo, una respuesta inmunológica con graves consecuencias para el paciente (incluso llegar a provocar su muerte). En el caso de anticuerpos, adicionalmente se deberá analizar la composición y estructura de los N-glicanos que contienen. LC/MS es la herramienta ideal para la caracterización detallada de biofármacos, aunque a nivel de control de calidad rutinario, algunos ensayos pueden ser realizados mediante HPLC con detección UV/VIS. En el caso de LC/MS son de gran ayuda el uso de softwares de "bioconfirmación", que como el descrito en el apartado anterior (3.4.2.- caso B), permitan confirmar si los datos experimentales concuerdan plenamente con los teóricos; así además se podrán identificar impurezas y variantes no deseadas (aquellos péptidos que no se pueden asignar).

Caracterización de Proteínas Recombinantes.

La caracterización de éste tipo de proteínas -"sintetizadas artificialmente" por la expresión de un gen "exógeno" introducido en un organismo que no dispone de él-, se realizará de acuerdo con las pautas y estrategias diversas definidas en el apartado anterior 3.4.2.

Caracterización de Anticuerpos Monoclonales.

Los mAB son proteínas con una estructura y características diferenciales, cuyo control requiere ensayos específicos para abordar su complejidad y gran tamaño. Para un detallado ejemplo de la caracterización de anticuerpos monoclonales puede consultarse la publicación (39: "Primary Characterization of a Monoclonal Antibody Using Agilent HPLC-Chip Accurate-Mass LC/MS Technology).

- Caracterización de Anticuerpos Monoclonales Intactos.
  Un ilustrativo ejemplo es el de la figura 34, ésta muestra el espectro de un anticuerpo intacto sin deconvolucionar y deconvolucionado; en éste último se observan 3 bandas principales, la más abundante correspondiente al anticuerpo con los 2 glicanos, y otras 2 bandas minoritarias para las partes deglicosiladas: una sin glicano (la de menor masa) y otra con 1 solo glicano (la segunda); incluso se muestra que existe una pequeña parte del anticuerpo en la que uno de los 2 glicanos incorpora una unidad de hexosa adicional.

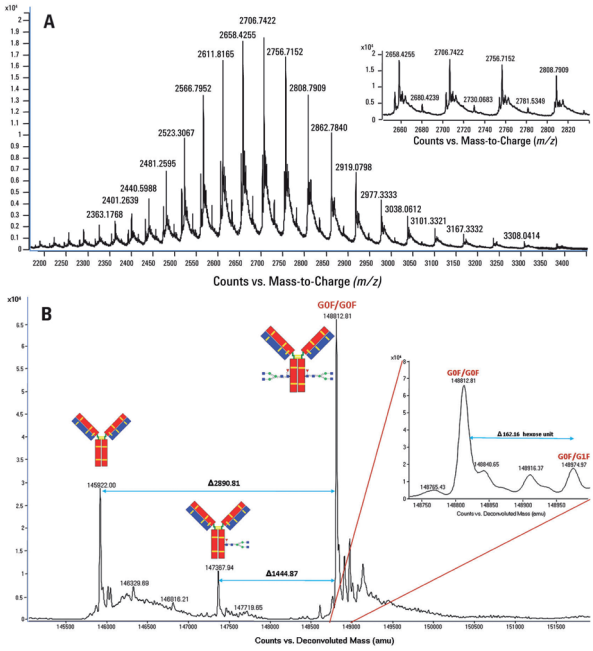


Figura 34.-Análisis de anticuerpo monoclonal intacto mediante Agilent HPLC-Chip/MS-QTOF-6520. Gráfico A: Espectro de masas (electrospray) sin deconvolucionar. Gráfico B: Espectro Deconvolucionado: muestra 3 bandas principales, la más abundante del anticuerpo con los 2 glicanos (148811.95Da; error masa: 5.7 ppm.), y otras 2 bandas minoritarias deglicosiladas: sin glicano (pérdida de 2890.81 Da correspondiente a 2 unidades del glicano G0F), y otra con 1 solo glicano (pérdida de 1444.87 Da correspondiente a 1 unidad del glicano G0F). La ampliación de la zona alrededor de la banda principal muestra que una pequeña parte del anticuerpo tiene uno de los 2 glicanos con una unidad de hexosa adicional (148974.97 Da diferencia 162.16Da) y su composición glicánica pasa a ser G0F/G1F (en lugar de G0F/G0F).

- **Caracterización de la parte NO Glicosilada del mAB:** para un estudio más detallado de la parte no glicosilada del anticuerpo (sin la presencia de parte glicánica – ver estructura de un mAB en figura 35), se puede analizar específicamente la parte proteica, después de llevar a cabo una reacción de deglicosilación del anticuerpo (mediante p.e. PNGase en tampón 20 mM Tris-HCl pH 8.2); así se obtienen espectros sin la variabilidad de masas que aportan las contribuciones de los glicanos del anticuerpo.

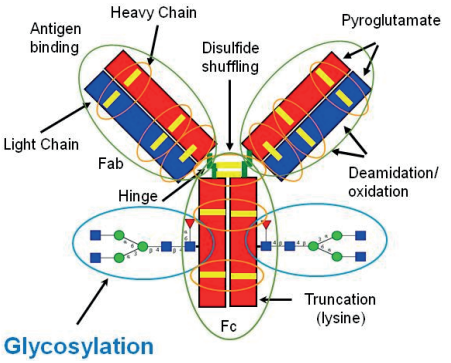


Figura 35.-Estructura de un anticuerpo monoclonal intacto. Se observa la estructura de los 2 grupos glicano unidos en la parte inferior de la cadena proteica pesada. En éste glicano los distintos azúcares están representados por N-acetylglucosamina (cuadrado azul), manosa (circulo verde), fucosa (triángulo rojo). En rectángulos rojos puede observarse la estructura de las 2 cadena proteicas pesadas y en azul las 2 ligeras, unidas mediante 16 puentes disulfuro (rectángulos amarillos).

- **Análisis de las Cadenas Ligeras y Pesadas del anticuerpo:** la reducción del mAB con DTT (p.e. con 25mM DTT a pH 8.6 y 80ºC durante 1h), romperá los puentes disulfuro que unen las 2 cadenas pesadas y las 2 ligeras del anticuerpo, permitiendo al quedar liberadas, la determinación de su peso molecular. En el caso de las 2 cadenas pesadas, éstas continuaran ligadas a su correspondiente glicano (si previamente estaban glicosiladas); si una parte del anticuerpo se había deglicosilado, o tenía un glicano distinto aparecerá reflejado en el correspondiente espectro deconvolucionado. Como comprobación de la masa del anticuerpo, se puede efectuar la siguiente operación:

  Peso Molecular mAB = Masa de las 2 cadenas ligeras + masa de las 2 cadenas pesadas + 32 (pérdida de los 16 puentes disulfuro)

- **Análisis de los Fragmentos Fab** (zona de unión al antígeno) **y FC** (zona unión glicanos)**:** la digestión del anticuerpo con una solución recién activada de papaína (una proteasa de la cisteína), romperá la bisagra del anticuerpo, partiendo la cadena pesada por la mitad y dejando una parte unida a la cadena ligera (figura 36), mientras la otra engloba la unión de las 2 mitades de la cadena pesada que están unidas a los 2 glicanos. Esta digestión se puede realizar, por ejemplo, con una solución de papaína recién activada en un ratio 60 mAB: 1 proteasa – en tampón 50mM fosfato sódico +1mM EDTA + 10mM cisteína a pH=7.0- durante 2 horas a 37ºC (la activación de la papaína se realiza dejándola previamente 15min a 37ºC en el mismo tampón). En este caso la masa del anticuerpo, se puede comprobar mediante: Peso Molecular mAB = Masa de los 2 Fragmentos Fab + masa del Fragmento Fc + 36 (2 moléculas de agua por la rotura de las 2 bisagras).

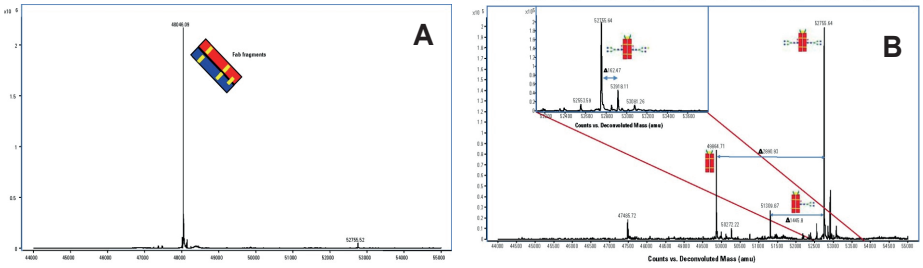


Figura 36.-Espectros deconvolucionados del fragmento Fab (figura A) y fragmentos Fc (figura B) de un anticuerpo monoclonal intacto digerido con papaína (proteasa de la cisteína) que rompe la bisagra del anticuerpo, partiendo la cadena pesada por la mitad.

Caracterización de N-Glicosilaciones.

En el análisis de anticuerpos monoclonales, un aspecto muy importante es la caracterización de la parte glicosilada; ésta tiene un importante efecto en la eficacia, biodisponibilidad y estabilidad del fármaco. Para su caracterización se suele empezar con la liberación del N-glicano de la asparraguina (Asn), por la que está unido a la parte proteica del anticuerpo; ésta acción se realiza mediante el uso de la enzima: PNGase F, y da lugar a la formación de la correspondiente glicosilamina (con un grupo NH2 en el punto de unión con la Asn). A partir de aquí los pasos de hidrólisis y derivatización dependerán del método analítico utilizado (MALDI-TOF, LC/MS, LC-CE con detección fluorescente o amperométrica). Recientemente se ha introducido en el mercado un muy útil dispositivo microfluídico: mAb-Glyco Chip, que mediante el uso de una precolumna con la enzima PNGase F inmovilizada, y diversas precolumnas y columnas (fase reversa y grafito), permite automatizar el proceso, reduciendo enormemente el tiempo de análisis de la muestra (de varias horas/días a unos 10-15 minutos). Si este dispositivo se acopla a un TOF o QTOF, permitirá automatizar la identificación y caracterización de los N-glicanos mediante bases de datos de masa exacta (con opción de tiempos de retención), ver ejemplo en Figura 37. Este tipo de analítica se describirá con mayor detalle en el capítulo 14 dedicado a la tecnología microfluídica: HPLC-Chip/MS o en la publicación (40:"The Agilent mAb-Glyco Chip Kit for rapid and fully automated characterization of N-linked glycans from monoclonal antibodies").

Compound Label	RT	Mass	Name	Formula	Tgt Mass	Diff (ppm)	Isomer(s) Present	Total Volume	Total Vol%
Cpd 28: 2121 0A 1G N	9,589	2092,7583	2121 0A 1G N	C79H132N6O58	2092,7564	0,93		8906	0,3
Cpd 23: 1100 0A 0G N	9,373	1258,4803	1100 0A 0G N	C48H82N4O34	1258,481	-0,58	1258,481	40030	1,5
Cpd 20: 2120 0A 0G N	9,349	1785,6693	2120 0A 0G N	C68H115N5O49	1785,6661	1,83		11972	0,5
Cpd 19: 3100 0A 0G N	9,317	1664,6406	3100 0A 0G N	C64H108N6O44	1664,6398	0,46		10792	0,4
Cpd 17: 2110 0A 0G N	9,305	1623,6155	2110 0A 0G N	C62H105N5O44	1623,6132	1,39		284929	10,8
Cpd 16: Man5 N	9,262	1233,4477	Man5 N	C46H79N3O35	1233,4494	-1,36	1233,4494	134626	5,1
Cpd 15: 2100 0A 0G N	9,250	1461,5628	2100 0A 0G N	C56H95N5O39	1461,5604	1,66		2044974	77,4
Cpd 13: 1100 0A 0G N	9,230	1258,4825	1100 0A 0G N	C48H82N4O34	1258,481	1,15	1258,481	40030	1,5
Cpd 11: 2010 0A 0G N	9,203	1477,5546	2010 0A 0G N	C56H95N5O40	1477,5553	-0,5		5076	0,2
Cpd 10: 1110 0A 0G N	9,198	1420,5339	1110 0A 0G N	C54H92N4O39	1420,5339	0,03		864	0,0
Cpd 9: Man4 N	9,170	1071,3928	Man4 N	C40H69N3O30	1071,3966	-3,55	1071,3966	6937	0,3
Cpd 5: 2000 0A 0G N	9,133	1315,5029	2000 0A 0G N	C50H85N5O35	1315,5025	0,3		59267	2,2
Cpd 4: 1000 0A 0G N	9,103	1112,4235	1000 0A 0G N	C42H72N4O30	1112,4231	0,31		34631	1,3
Cpd 3: Man4 N	9,082	1071,3956	Man4 N	C40H69N3O30	1071,3966	-0,9	1071,3966	6937	0,3
Cpd 1: Man5 N	9,068	1233,4475	Man5 N	C46H79N3O35	1233,4494	-1,52	1233,4494	134626	5,1

Figura 37.-Ejemplo de glicanos identificados en un anticuerpo monoclonal, liberados mediante el uso de la enzima PNGase F y separados mediante una columna de grafito. Este tipo de columna permite la separación de isómeros (remarcados con el mismo color) con igual masa exacta.

4. Aspectos a considerar por parte de operadores de instrumentos de tiempo de vuelo.

Este apartado va dirigido a los usuarios que manejan los equipos, pero puede también resultar de interés a usuarios de servicios de proteómica que quieran conocer algunos detalles sobre aspectos instrumentales relacionados con el análisis de sus muestras, o sobre las posibilidades de los instrumentos utilizados.

Los objetivos de este apartado son:

- Dar algunas breves pautas/ recomendaciones genéricas (los detalles dependerán de cada modelo y marca de equipo), para trabajar en condiciones adecuadas y evitar algunos de los errores habituales por parte de usuarios con poca experiencia.

- Dar a conocer detalles y posibilidades de la tecnología y de softwares que pueden resultar muy útiles en determinadas ocasiones.

4.1. Condiciones de ionización en fuentes Electrospray y nanoElectrospray.

En proteómica mediante LC/MS se utilizan 2 tipos de fuentes de ionización: ESI o nanoESI , intercambiables entre sí (requerirán también intercambiar el HPLC o nanoHPLC). A continuación se explicará cómo seleccionar la más conveniente en función de las necesidades prioritarias del usuario.

Selección Fuentes Electropray (ESI) "versus" Nanoelectrospray (nanoESI).

La diferencia entre ambas radica en la sensibilidad, flujos operativos (y por consiguiente diámetro de las columnas) y robustez proporcionadas. La selección del tipo de fuente de ionización dependerá de las necesidades del usuario:

- **Máxima sensibilidad** en análisis de trazas → fuente nanoelectrospray.
- **Máxima robustez y/o mínimos tiempos de análisis** (elevado nº de muestras a analizar) → fuente electrospray.

La selección de uno u otro tipo de fuente limitará el rango de flujos y por tanto el diámetro de las columnas a utilizar:

- **Nanoelectrospray:** rango de nL/min (típico 200-800 nL/min) → nano-columnas de 50-100 µm de diámetro interno.

- **Electrospray con micronebulizador:** rango de bajos µL/min (típico 1-100 µL/min) → columnas capilares de 180-1.000 µm de diámetro interno.

- **Electrospray con nebulizador estándar:** rango de altos µL/min (típico 50-1.000 µL/min) → columnas estándar de 2.1-4.6 mm de diámetro interno. Existen diversos tipos de fuentes electrospray modificadas que admiten flujos de hasta 2 mL/min; también algunas de antiguas que no admiten más de unos 50-100 µL/min.

Los criterios a considerar a la hora de escoger entre una y otra se basan en:

- **Diámetro columna utilizada:** a diferencia de las fuentes de Ionización Química (APCI - no utilizadas en proteómica), que responden a la cantidad de analito que llega al detector, la ionización por electrospray responde a la concentración de analito, y por tanto, se puede reducir el flujo sin sacrificar sensibilidad. Si se mantiene constante la cantidad absoluta de analito inyectada en columna, **su concentración aumentará proporcionalmente a la reducción de la sección de la columna** (ver tabla 6); por tanto,

Tabla 6.- Muestra la relación entre el diámetro interno de la columna, el inverso de su sección relativa, su capacidad de carga y flujo equivalente; toma como referencia las columnas de 2.1mm (típicas con electrospray). La segunda columna (inverso de la sección), muestra la "teórica"·mejora en sensibilidad que se puede llegar a obtener; a la vez, coincide con la reducción de la capacidad de carga de la columna. La tercera columna indica la cantidad total de muestra que la columna es capaz de separar, manteniendo una buena separación cromatográfica (sin que la fase estacionaria se sature). La última columna indica el flujo típico (rango habitual entre el 50 y 200% de estos valores).

Diámetro Interno Columna	Inverso Sección Relativa (potencial mejora sensibilidad y reducción capacidad de carga)	Capacidad de Carga Columna 15 cm long.	Flujo Equivalente (velocidad lineal constante)
4.6 mm	0,2	5 mg	1,44 mL/min
3 mm	0,5	2 mg	612 µL/min
2.1 mm	1	1 mg	300 µL/min
1 mm	4,4	230 µg	68 µL/min
500 µm	18	60 µg	17 µL/min
300 µm	49	20 µg	6 µL/min
100 µm	441	2µg	681 nL/min
75 µm	784	1 µg	383 nL/min

el trabajar con columnas de diámetro más pequeño proporcionará mayor sensibilidad con este tipo de fuentes de ionización. No obstante, no hay que olvidar que la cantidad de analito que la columna es capaz de separar eficientemente, se reduce en la misma proporción; por tanto se puede perder mucha eficiencia en la separación y sensibilidad, si se inyecta un exceso de muestra, o ésta contiene componentes muy mayoritarios que la saturan; especialmente si éstos coeluyen con minoritarios de interés. Para obtener buenos resultados, es **muy importante no sobrecargar la capacidad de carga de la columna, ni la del sistema de ionización.** Cuanto más compleja y mayor sea la cantidad total de materia introducida en la columna, mayor deberán ser sus dimensiones y las de la precolumna de carga, y más lento deberá ser el gradiente utilizado. Para evitar la sobrecarga (o comprobar que no existe), se puede recurrir a inyectar un menor volumen o diluir la muestra.

Para no perder eficacia adicional en la separación, el volumen muerto (capilares+ conexiones+…) por donde circula la muestra, debe ser inferior al 2% del volumen muerto de la columna (o < 10% del volumen del pico), aspecto que en nanocromatografía no es sencillo de conseguir (volumen muerto columna 75µm d.i. x 15cm ~ 530nL); para reducir el volumen post-columna, el extremo de salida de las nanocolumnas se conecta directamente a la punta emisora sobre la propia fuente. En general la eficiencia en la separación es mayor con columnas de mayor d.i.

Sensibilidad Electrospray (ESI) "versus" nanoelectrospray (nanoESI): la misma cantidad absoluta de muestra inyectada en una columna de 75µm de d.i. con una fuente nanoESI, suele dar unos pocos centenares de veces más sensibilidad que una fuente ESI estándar con una columna de 2.1mm; no obstante, el uso de fuentes ESI mejoradas con focalización por gradiente térmico (p.e. tecnología Jet Stream), y de espectrómetros con captación mejorada de iones (p.e. tecnologías i-Funnel), permiten reducir estas diferencias a muy bajas decenas de veces. A cambio, las fuentes ESI son:

- Mucho más robustas.
- Con HPLC o UHPLC los gradientes llegan mucho antes a columna permitiendo ahorrar varios minutos por inyección.
- La cromatografía es más eficiente que la nanocromatografía.

- **Volumen de muestra a inyectar y configuración del sistema de inyección:** si no se realizan procesos de focalización de la misma en una precolumna, o en la cabeza de la columna de separación, no se debería inyectar más allá del 1-2% del volumen de la columna (5-10 µL para columna de 15cm y 2.1mm d.i. // 5-10nL para 75µm d.i.); aunque dada la menor eficiencia en la separación de las nanocolumnas, éste límite se suele aumentar a una pocas decenas de nanolitros. La alternativa para inyectar mayores volúmenes es utilizar algún sistema de focalización de la muestra:

  - **Volumen de muestra a inyectar y configuración sistema de inyección en nanoESI:** si se desean inyectar varios µL (p.e. 2-5µL) en una nanocolumna (p.e. 75µm de d.i.) a 300nL/miin, se tardarán varios minutos para cargar la nuestra. Por ello será conveniente inyectar en una precolumna concentradora de mayor diámetro interno (p.e. 300µm d.i. y 5mm de longitud -350nL), que admita flujos mayores (p.e. 10-20µL/min), para efectuar la carga de muestra mediante un nano HPLC con una microválvula (2 posiciones y 6 vías), y una bomba de carga adicional (2ª bomba) para empujar rápidamente la muestra a la precolumna, donde ésta se focalizará utilizando un eluyente de carga que retenga a los péptidos en la precoluma. Una vez cargada la muestra, se gira la microválvula para que la muestra sea transferida y separada en la nanocolumna (p.e. 75µ d.i. y 15cm longitud), utilizando la nanobomba (la de separación de la muestra a p.e. 300-600nL/min). Si interesa una buena focalización de la muestra (picos finos), las

conexiones de las bombas y microválvulas, se realizarán de tal manera que la muestra eluya hacia la columna de separación, en sentido contrario al de la carga; si la prioridad es proteger a la columna de separación directamente, se hará en el mismo sentido (suele dar picos más anchos). Si interesan ambas, se hará en sentido contrario, pero se instalará una segunda precolumna justo antes de la columna de separación.

- Volumen de muestra a inyectar y configuración sistema de inyección en ESI: al trabajar a flujos elevados, y no tener el problema de un elevado tiempo de carga de la muestra, es más habitual utilizar configuraciones estándar: 1bomba+inyector+columna, y lo típico sería inyectar 5-10 µL para una columna de 15cm y 2.1mm d.i. Se pueden inyectar volúmenes mayores (p.e. 40µL), siempre y cuando la fase móvil inicial tenga poco poder de elución, y focalice los péptidos menos retenidos. Si se quiere preconcentrar grandes volúmenes, se puede recurrir a la configuración descrita para nano-ESI.

Siempre, pero especialmente cuando se inyecten grandes volúmenes, es muy importante evitar que la muestra precipite, o que se introduzcan partículas en el interior del sistema cromatográfico; tanto más, cuanto menor sea el tamaño de partícula de la columna utilizada. Lo ideal es poder microfiltrar la muestra por 0.22µm o centrifugarla adecuadamente y tomar sobrenadante.

Consideraciones Operatorias Fuente NanoESI.

En las fuentes nanoESI la ionización tiene lugar asistida por voltaje (y no por presión como en electrospray); para ello se utilizan unas muy finas puntas emisoras recubiertas de material conductor, que terminan en un orificio de 10-30 micras dependiendo del rango de nanoflujos que interese (10-1000nL/min). Mediante un contra-electrodo muy próximo (unos pocos mm), y posicionando correctamente la punta emisora (ajuste x-y-z de acuerdo con recomendaciones fabricante), se aplica un voltaje que se va subiendo paulatinamente hasta que se alcanza un valor suficientemente elevado (1-2Kv), para formar una pluma de nebulización suficientemente estable (cono de Taylor), a lo largo de toda la cromatografía. Los valores de voltaje dependerán del flujo, composición del eluyente y estado de la punta (en general proporciones elevadas de agua requerirán mayores voltajes). Estas puntas pueden ser varios materiales (sílice, cuarzo,…) y en general son muy delicadas y de una vida corta; hay que reemplazarlas cada muy pocos días, e incluso en ocasiones al cabo de sólo unas pocas horas. Para evitar que se quemen muy rápidamente, hay que evitar aplicar voltaje sin que circule fase móvil por su interior; asimismo el extremo fino de las puntas no se puede tocar.

El resto de parámetros de la fuente nano-ESI, a excepción de los del nebulizador, siguen a grandes rasgos consideraciones comunes con la fuente ESI (aunque utilizando valores distintos).

Consideraciones Operatorias Fuente Electrospray (ESI).

En las fuentes ESI la nebulización tiene lugar asistida por gas (habitualmente nitrógeno), mediante el uso de un nebulizador. Los diseños de las fuentes y del sistema de entrada de iones son muy variados de unos fabricantes a otros, y pueden llegar a utilizar valores muy distintos; consecuentemente se deberá consultar el manual específico de cada equipo; pero esencialmente sus parámetros de control son:

- Posición del Nebulizador: en algunos modelos hay que hacer un ajuste que puede llegar a ser tridimensional (x-y-z) en función de los flujos de trabajo. En otros, en los que el nebulizador forma un ángulo de 90º con respecto a la entrada del capilar, su posición es fija y no requiere de ajustes.

- Presión o Flujo de Gas del Nebulizador: se deberá ajustar en función del flujo de trabajo. Su valor se suele aumentar con el incremento de flujo.

- **Flujo y Temperatura del Gas de Secado o Desolvatación:** se ajustan en función del flujo y composición de la fase móvil; se requerirán mayores valores a medida que se incrementa el flujo, y la máxima proporción de agua en la fase móvil. Otros equipos requieren adicionalmente definir la temperatura de la fuente, temperatura del capilar, flujos del gas del cono,…

- **Voltaje del Capilar**: es el voltaje que atrae los iones de la polaridad de interés, a la superficie de la gota nebulizada para favorecer su desolvatación, y a la vez, define el gradiente de voltaje para atraer los iones hacia el interior del espectrómetro. Sus valores están en el rango 2-6Kv; en general cuando mayor sea la m/z del analito, mayor es el valor utilizado; en polaridad negativa se deberá trabajar a valores más bajos 2-4Kv.

- **Voltaje de Fragmentador o de Cono:** éste voltaje maximiza la eficiencia en la transmisión de los iones hacia el interior del espectrómetro. Cuanto mayor es la m/z del analito, mayor es en general el valor a utilizar; aunque en algunos equipos es bastante compuesto independiente. Si se aumenta suficientemente, permite fragmentar la molécula antes de entrar en las lentes de focalización del TOF (o en el primer cuadrupolo en el caso del QTOF), para obtener información estructural. Con voltajes elevados la molécula impactará violentamente con el gas presente en la zona del cono ("skimmer"), y fragmentará de una manera similar a como lo haría en una celda de colisión. Este voltaje permitirá obtener información estructural con un TOF ("pseudo MS/MS") o un "pseudo MS/MS/MS" con un QTOF; pero con este tipo de fragmentación inducida por colisión (CID), se fragmentarán todos los compuestos que coeluyan conjuntamente. Algunos autores le llaman "fragmentación en la fuente"; aunque en la mayor parte de equipos, esta fragmentación se realiza en el interior del espectrómetro, en la zona entre la salida del capilar (que conduce los iones al interior del espectrómetro) y el primer cono.

4.2. Sintonizado y Calibrado del instrumento.

El sintonizado de un tiempo de vuelo.
Consiste en el ajuste de los voltajes, frecuencias, radio frecuencias,… en el interior del espectrómetro, para ajustar la resolución espectral, respuesta/ abundancia a lo largo de un amplio rango de masa, exactitud en la asignación de masa, asegurar una buena fidelidad en la proporción de abundancias isotópicas,….. Se realiza bombeando en continuo una solución de compuestos conocidos, con un amplio rango de masas (usualmente entre 100 y 3000 aproximadamente). Suele ser un proceso relativamente automatizado que suele realizar el propio equipo cuando el usuario lo ordena.

El sintonizado se realizará en las 2 polaridades (positiva y negativa). En el caso de QTOF, adicionalmente al TOF, de tanto en tanto, se deberá sintonizar el cuadrupolo; éste requiere una frecuencia de sintonizado mucho menor que los tiempo de vuelo (típico 6-12 meses "versus" 1-2 semanas).

Calibrado del eje de masa: adicionalmente al sintonizado, con los TOF /QTOF se debe realizar un calibrado del eje de masa, éste es muy rápido y se suele realizar a diario; incluso al principio de cada serie de análisis, si se trabaja sin utilizar "masas de referencia interna" adicionadas en continuo o alternantemente. A diferencia del sintonizado, el calibrado de masa no ajusta resolución, intensidad,… simplemente ajusta el eje de m/z.

Opciones rangos de masa: los TOF/QTOF tienen un muy amplio rango de masa (suelen llegar hasta diez o veinte mil de m/z trabajando con reflectrón); aunque éste no suele ser necesario para la gran mayoría de analitos. Para ciclos de medida más rápidos, en las condiciones de calibrado se seleccionará el rango de masa que mejor se ajuste a los compuestos de mayor m/z,

que puedan estar presentes en la muestra (éstos definen el máximo tiempo de vuelo necesario). Así se podrá mejorar la sensibilidad al acumular un mayor nº de vuelos ("transients") en el mismo intervalo de medida, o mantener la sensibilidad pero aumentar el nº de puntos de muestreo a lo largo del pico. Para cuantificar compuestos se requiere un mínimo nº de puntos de medida a lo largo del pico, para perfilarlo con suficiente exactitud:

- 8-10 puntos/pico a nivel de trazas/límites cuantificación (RSD%: 10-15%)
- 12-15 puntos/pico para concentraciones bajas-medias (RSD%: 1-3%)
- 15-20 puntos/pico concentraciones altas (RSD% <1%).

Si se modifica el rango de masas del calibrado, el instrumento se deberá calibrar de nuevo (pero no hace falta volver a sintonizarlo). En cambio, si sólo se modifica el rango de masa de almacenamiento de datos, no hará falta recalibrar el eje de masas. La opción del rango de masa del calibrado, no hay que confundirla con el rango de masa de almacenamiento de datos (igual o inferior al anterior); el de almacenamiento se puede reducir para adaptarlo a la zona de interés y reducir así el tamaño del fichero de datos, pero no afecta al tiempo de vuelo máximo asignado a cada "transient", ni a la frecuencia de adquisición de datos.

Puesta en marcha y parada de los TOF y QTOF: será importante seguir detalladamente el protocolo definido por cada fabricante. Este aspecto cambia considerablemente respecto a cuadrupolos, triple cuadrupolos y trampas de iones. El hecho de operar a un mayor vacio y tener un mayor volumen interno (debido al tubo de vuelo), motiva que los protocolos en el caso de TOF y QTOF incluyan etapas adicionales, que reduzcan la entrada de humedad en el interior del sistema, y faciliten la posterior puesta en marcha para alcanzar el elevado vacío que utilizan; de ahí la importancia de seguir los protocolos de parada y puesta en marcha recomendados.

4.3. Masas de referencia interna para calibrado "on-line" del instrumento.

Para mejorar la exactitud de masa, los TOF/QTOF suelen permitir adicionar en la fuente, idealmente en continuo, una solución que contiene unos pocos compuestos de masa exacta conocida, que se utilizarán como patrón interno para el recalibrado del eje de masas "on-line", espectro a espectro. Para recalibrar se suelen utilizar 2 masas separadas por un mínimo de 500 Daltons/z.

La exactitud actual de los TOF/QTOF permite precisar:

- La masa con exactitudes mejores a la del peso de 1 electrón.
- Los tiempos de vuelo con exactitudes mejor de 1 nanosegundo.

Esta enorme exigencia en la exactitud de la medida, motiva que el uso de masas de referencia interna, permita usualmente reducir a ½-1/3 el error de las medidas. La corrección en continuo sólo se puede utilizar en modo MS, en modo MS/MS se suele realizar adquiriendo un espectro de MS, cada X segundos (definido por el usuario), de ahí que en general el error en modo MS/MS sea del orden de 2-3 veces el del modo MS.

Si se utiliza la opción de corrección utilizando masas de referencia interna, la frecuencia del calibrado del TOF/QTOF se puede reducir; caso contrario sería recomendable realizarla al inicio de cada serie de análisis que requiera una buena exactitud.

Localización de las masas de referencia interna.

Para evitar errores importantes, será clave asegurarse que dentro de la ventana de localización de esas masas de referencia, no aparezca ninguna otra masa de mayor intensidad. Así por ejemplo si hemos seleccionado una ventana de localización de +/- 10ppm y en un momento

del cromatograma aparece otra masa más intensa a 9ppm de la teórica, el equipo tomaría ésta como referencia, y recalibraría ese espectro generando un importante error de masa (de 9ppm en esa zona), muy superior a las 2-4 ppm típicas de los equipos actuales operando en modo patrón externo. Esta mala recalibración interna es más probable a masas bajas, puesto que el fondo ("background") a masas bajas suele ser mucho más abundante que a masas elevadas (más selectivas). Una buena práctica es comprobar al inicio de una serie de análisis, que las masas de referencia se detectan bien y sin presencia de interferencias próximas; especialmente cuando se cambia de fases móviles, columna,…... Si hay interferencias, se puede reducir la ventana de localización y/o incrementar la concentración del compuesto/s de referencia. Estas masas de referencia también se pueden comprobar, revisando un cromatograma ya adquirido, cuando se detecten errores de masa superiores a los habituales.

Recalibración de un espectro mal calibrado.
En algunos software de tratamiento de datos, la recalibración de masa se puede volver a efectuar una vez ya adquiridos los datos. Ello permite corregir una mala recalibración, escogiendo las masas de referencia en un espectro ya adquirido, y recalibrando según interese éste solo, o todos los espectros del fichero de datos.

4.4. Tipos de modos de adquisición en LC/MS-QTOF/TOF.

Los espectros de masas se pueden guardar en 2 formatos:

- **Perfil ("profile"):** guarda el espectro completo con todos los datos (se asemeja a un cromatograma), con una resolución mejor del mDa, y que dependerá de la frecuencia de muestreo del sistema de medida del tiempo de vuelo. Este tipo de fichero es el que proporciona mayor nivel de detalle, pero tiene el inconveniente de consumir una cantidad muy importante de disco duro, que puede llegar incluso a varios centenares de Mb por inyección, y requerir un mayor tiempo de reprocesado para su análisis. Por ello no suele utilizarse para análisis rutinarios, si hay que analizar un considerable número de muestras. Por el contrario, es el modo a utilizar cuando después se quiere deconvolucionar el fichero de datos, utilizando los algoritmos tipo "Maximum Entropy" o "Large Molecular Feature Extractor" (LMFE)

- **"Centroide":** en lugar de guardarse el espectro completo, se integra éste en tiempo real y se simplifica guardándolo en forma de diagrama de líneas. Cada pico de masa se representa por una sola línea de masa, centrada en el pico y con su altura. Si el pico de masa tiene hombros (solapamientos con masas isobáricas no resueltas) u otros picos solapados, estos hombros o picos solapados generarán líneas adicionales. Para simplificar y reducir aún más el tamaño del fichero, se puede definir un umbral (absoluto y/o relativo) por debajo del cual no se guardará ninguna línea de masa. Este tipo de formato consume mucho menos disco duro que el anterior, y es el más utilizado en análisis rutinarios.

Los TOF y QTOF siempre proporcionan un espectro completo; en el caso de los TOF tienen sólo 1 modo de trabajo: MS Scan, en cambio los QTOF ofrecen 3 opciones de adquisición: MS, "targeted MS/MS" y Auto MS/MS que a continuación se describirán.

4.4.1. Modo MS (Espectros MS).

Es el método utilizado en los TOF y también el más utilizado en los QTOF; en esta modalidad el cuadrupolo y la celda de colisión del QTOF, transmiten todos los iones que les entran sin filtrarlos ni fragmentarlos. Es el método a utilizar para:

- Descubrimiento de biomarcadores utilizando la estrategia "metabolómica".

- Screening masivos de péptidos (/proteínas), metabolitos, fármacos,…
- Método muy habitualmente utilizado para cuantificar (utilizando cromatogramas de iones con muy finas ventanas de extracción de masa).
- …..

Para las mismas aplicaciones, también se puede utilizar el modo auto MS/MS, pero en este caso será importante tener a lo largo del pico, un número suficiente de puntos de medida en la dimensión de MS (12-15 puntos // 10 para trazas).

Para la adquisición en modo MS, sólo hay que definir 2 parámetros:

- **Rango masa (mínima-máxima) de almacenamiento**: para reducir el consumo de disco duro, conviene ajustarlo a la zona de interés (incluyendo las masas de referencia), pero sin que exceda la masa máxima seleccionada para el calibrado (si no bastará con recalibrar el eje de masas).

- **Frecuencia de adquisición de datos**: losTOF y QTOF proporcionan una elevada frecuencia de adquisición (en MS típica entre 20 y 50 Hz dependiendo de modelos). Ello permite adquirir un suficiente número de puntos por pico, incluso cuando se utiliza cromatografía muy rápida. No obstante hay que tener en cuenta que al aumentar la frecuencia de adquisición de datos, el número de promedios que se efectúa se reduce y la relación señal/ruido empeora. Si no se va sobrado en sensibilidad, será importante ajustar la frecuencia al valor adecuado (12-15 puntos // 10 para trazas). Para más información ver apartados 3.3 y 4.5.4.

Una útil opción que proporcionan algunos TOF y QTOF, es la de dividir el ciclo de medida en varios espectros independientes (p.e. 4), y así poder por ejemplo:

- **TOF/QTOF: adquirir "simultáneamente" a varios voltajes de fragmentador/cono.** De esta manera, en muestras no muy complejas, en los canales a bajo voltaje de fragmentador, se obtiene la información de los iones precursores que coeluyen; mientras en los canales a elevado voltaje, se obtiene información de sus fragmentos. A partir de ahí simplemente se deconvoluciona la señal para agrupar aquellas masas que a lo largo del tiempo "sube y bajan" conjuntamente, y obtener sus espectros deconvolucionados en el caso de péptidos que no estén perfectamente solapados. Si 2 compuestos coeluyen perfectamente sus fragmentos y precursores no se podrán diferenciar.

- **QTOF en modo MS/MS: adquirir "simultáneamente" a varias energías en la celda de colisión.** Este modo permite adquirir "simultáneamente" los espectros de MS/MS a, por ejemplo, 3 energías: baja-media-alta (p.e. 10-20 y 40eV), para asegurar que con todo tipo de moléculas se obtendrá algún espectro con una fragmentación adecuada.

4.4.2. Modo MS/MS Programado ("targeted MS/MS")

Se utiliza en labores identificativas de compuestos conocidos, proporciona el espectro de los iones producto ("Product Ion spectrum"); requiere conocer la masa y tiempo de retención de los compuestos de interés, para poder programar en el tiempo el aislamiento y fragmentación de sus iones precursores. Se puede utilizar también para cuantificar, pero dependiendo de los compuestos, suele proporcionar menor sensibilidad que cuantificar en modo MS; debido a la menor intensidad de los iones producto con respecto a la de los iones precursores (es habitual ver diferencias de 10 veces). Para tener buenas sensibilidad y estadísticas para el cálculo de la masa exacta, en MS/MS se deberán acumular un mayor número de vuelos ("transients") que en modo MS

Para la adquisición en modo “targeted MS/MS”, hay que definir varios parámetros:

- **Rango de masas independientes para MS/MS y MS:** aunque se trabaje en modo MS/MS, el usuario puede seleccionar la opción de insertar un espectro de MS cada X segundos (definidos por él); así cada pocos segundos se realizará un reajuste del calibrado del eje de masas. Su rango de masas (MS) habrá que ajustarlo al de las masas de recalibración utilizadas. Recordar que al fragmentar iones con múltiples cargas, se obtendrán habitualmente fragmentos a m/z por encima de la del precursor multicargado (fragmentos con menor número de cargas).

- **Frecuencia de adquisición de datos independientes para MS/MS y MS**: dada la menor intensidad de los espectros de MS/MS se suele dedicar más tiempo a adquirir un espectro de MS/MS (menor frecuencia) que uno de MS.

- **Iones precursores:** en el caso de péptidos se prefiere el uso de precursores con múltiples cargas (2-3) al producir espectros de MS/MS con una mayor número de iones producto. En cuantificación se podrán sumar las intensidades de los iones producto más abundantes. En precursores con 1 sola carga, en el MS/MS sólo se detectará 1 de los 2 fragmentos (el otro se puede calcular como una pérdida neutra).

- **Tiempo de retención y ventana de adquisición de los precursores:** cuanto menor sea el número de iones precursores a adquirir simultáneamente, mejor será la sensibilidad y/o menor el total del ciclo del tiempo de medida de los diversos precursores. Los tiempos y ventanas se ajustarán para reducir las coadquisiciones de precursores.

- **Ventana de masa de aislamiento del ión precursor** (“isolation width”)**:** define la resolución espectral que utiliza el cuadrupolo, para aislar en el espacio al ión precursor; aunque los cuadrupolos proporcionan poca resolución espectral, el trabajar con una ventana fina puede ayudar a evitar esporádicamente la fragmentación de algún péptido coeluyente con una m/z muy próxima. Por el contrario, el utilizar ventanas muy finas proporciona menor sensibilidad. Convendrá por tanto ajustarlo en función de las necesidades:

  - **En aplicaciones cuantitativas:** la selectividad se conseguirá con la masa exacta y elevada resolución espectral de los iones producto; a menos que el interferente dé exactamente los mismos iones producto, no debería haber problema desde el punto de vista cuantitativo, y se recomienda utilizar ventanas de ancho medio.

  - **En aplicaciones cualitativas/identificativas:** el escoger una ventana de aislamiento más o menos fina, dependerá de la calidad de la separación cromatográfica, de la complejidad de la muestra y de la concentración relativa entre analitos de interés y compuestos interferentes. Cuanto peor sea la primera, mayor la segunda y menor la tercera, más importante será utilizar finas ventanas que filtren la sustancia interferente. Si ésta es intensa, sus iones producto predominarán en el espectro de MS/MS y pueden dificultar la identificación automática del minoritario péptido de interés.

- **Energía de la Celda de Colisión:** para cada precursor se podrá definir individualmente la que más interese; en algunos equipos se puede definir en función de fórmulas o tablas relacionadas con la m/z del precursor. Para compuestos de la misma familia, la energía de colisión óptima se incrementa linealmente con la m/z.

- **Tiempo de de monitorización de cada fragmentación**: en algunos equipos este tiempo se puede individualizar para cada uno de los precursores; ello permitirá dar más tiempo a compuestos coeluyentes de los que se espera menor respuesta.

4.4.3. Auto MS/MS (“Data Dependent” MS/MS)

Se utiliza para identificar compuestos en muestras “desconocidas”. Combina de una manera automática la adquisición de espectros de MS y MS/MS; el propio equipo decide en tiempo real y en función de lo detectado en los espectros de MS, que iones precursores aísla y fragmenta automáticamente (ver figura 15 en apartado 3.1.1.1). Por defecto empezará seleccionando los precursores más intensos, pero se podrán definir una serie de pautas para proporcionar un cierto nivel de inteligencia al algoritmo de decisión. Los parámetros para su programación son los mismos que en la sección anterior (excepto el tiempo de retención y ventana de adquisición de los precursores), más una serie de parámetros específicos del modo “Data Dependent” MS/MS:

- **Número máximo de precursores por ciclo**: indica al equipo cuantos iones precursores, puede tomar como máximo de cada espectro de MS, para aislar y fragmentar, antes de volver a iniciar un nuevo ciclo, y tomar un nuevo espectro de MS. Las intensidades de los precursores deben superar unos umbrales mínimos, y sus m/z no deben formar parte de las listas de exclusión. Dependiendo de las funciones de control que tenga el equipo, puede convenir limitar este nº para asegurar que se tomen un mínimo de 3 espectros de MS a lo largo del pico, y así asegurar que en algún ciclo se adquirirán los espectros de MS/MS en una zona del pico con buena intensidad.

- **Umbrales de Intensidad mínima para los precursores:** si éstos no superan los umbrales mínimos (absolutos y relativos) de intensidad, el potencial precursor se descartará y no se le hará MS/MS.

Exclusión activa: Número máximo de repeticiones y Tiempo de exclusión.

Define el número máximo de repeticiones del mismo precursor, que se pueden tomar durante un determinado tiempo. Su uso permitirá obtener espectros MS/MS de precursores que no sean los mayoritarios, y así profundizar un poco más, adquiriendo MS/MS de un mayor nº de compuestos coeluyentes. El tiempo de exclusión definirá la ventana de tiempo durante la cual un precursor, que haya alcanzado el nº máximo de repeticiones, pasará a formar parte temporalmente de la lista de iones precursores excluidos (aquellos de los que no se puede obtener su MS/MS). Si el equipo no dispone de la posibilidad de variar automáticamente el tiempo de aislamiento del precursor (para alcanzar una determinada intensidad de señal en el MS/MS), para asegurar que se realicen un mínimo de 3 ciclos de medida a lo largo del pico; el tiempo de exclusión debería ser del orden de 1/3 del ancho del pico cromatográfico en su base. Para diversificar los precursores, puede convenir que no se repitan en el mismo ciclo; ya se repetirán en los siguientes.

Listas de Exclusión y ventana de masa.

Permitirá definir una lista de m/z de iones precursores (y ventana m/z de margen) a excluir para obtener sus espectros de MS/MS. En algunos equipos estas listas se pueden programar en franjas de tiempo. Las listas de exclusión son muy útiles para:

- **Evitar adquirir MS/MS de iones del fondo,** que no pertenecen a la muestra. Para ello es muy útil adquirir un blanco, y después importar la lista de iones detectados y sus tiempos de retención, para programarla como lista de iones a excluir en el tiempo; así se maximiza el tiempo que el equipo dedica a adquirir precursores que realmente pertenecen a la muestra.

- **Evitar adquirir MS/MS de precursores previamente adquiridos:** el añadir a la lista de exclusión (programada en el tiempo), los iones precursores ya previamente adquiridos, permitirá profundizar en el proteoma. A base de sucesivas inyecciones, que cada vez, irán adquiriendo los MS/MS de péptidos coeluyentes que proporcionen menor respuesta.

Lista de Iones preferentes.
Algunos equipos permiten definir un listado de precursores (con la opción de programarlo en el tiempo) que si se detectan, aunque sean minoritarios, tendrán preferencia con respecto a otros de más intensos. Esta opción es muy útil para obtener los espectros de MS/MS en un "screening" dirigido, pero que a la vez permita obtener los espectros de MS/MS de otros compuestos que se detecten.

Preferencias del estado de carga.
Es una opción muy útil para la identificación de proteínas, permitirá dar preferencia o incluso centrarse exclusivamente en iones precursores con múltiples cargas. Proporciona espectros de MS/MS con mayor número de fragmentos que facilitarán considerablemente la secuenciación del péptido y consecuentemente la identificación de la proteína a la que pertenece.

Tiempo de Aislamiento variable en función de la intensidad del precursor.
Es una útil función que variará el tiempo de monitorización de la fragmentación de un precursor (nº de "transients" acumulados), en función de su intensidad. Los iones precursores más intensos se adquirirán más rápidamente (se acumularán un menor nº de vuelos), dedicándose más tiempo a los de menor intensidad. Lo incorporan algunos equipos para acelerar la adquisición y mejorar la sensibilidad de los MS/MS de los precursores menos intensos. Para evitar adquirir espectros MS/MS de baja calidad, algunos equipos permiten definir un umbral mínimo de intensidad prevista para los espectros MS/MS, rechazando aquellos precursores que previsiblemente no sean capaces de proporcionar, en el tiempo máximo de acumulación de "transients" de MS/MS, esa intensidad total (TIC del MS/MS).

Umbral de Pureza del precursor.
Para evitar adquirir MS/MS de precursores que coeluyan, dentro de la ventana de aislamiento del cuadrupolo, con otros "iones interferentes" excesivamente intensos, y que darían lugar a espectros MS/MS mezcla de varios compuestos; algunos equipos permiten definir el mínimo porcentaje de pureza, expresada como intensidad total que el clúster isotópico del precursor aislado, debe tener con respecto a la intensidad total de iones en la ventana de aislamiento del cuadrupolo (ésta se puede variar de 1 a varios Dalton).

Múltiples Energías de Colisión.
Algunos equipos permiten aplicar múltiples energías de colisión al ión precursor aislado para asegurar que alguna proporcione una fragmentación adecuada.

El modo Auto MS/MS permitirá cuantificar en los siguientes casos:

- Mediante cuantificación relativa basada en los ratios de intensidades de iones producto presentes en el mismo espectro de MS/MS, como es el caso de la técnica ITRAQ.

- Utilizando la dimensión de MS, siempre y cuando se haya adquirido un número suficientemente elevado de puntos a lo largo del pico, para que los cambios en la frecuencia de adquisición no afecten al área del mismo.

En cambio NO es adecuado para cuantificar por el área de los cromatogramas de iones producto extraídos de la dimensión de MS/MS, dado el muy reducido número de puntos del que se dispondrá, para el precursor en cuestión.

Otra posibilidad más recientemente introducida de efectuar "MS/MS", es fragmentar a todos los iones sin aislarlos o MS/MS Independiente de los Datos ("MS/MS All ions"); en esta modalidad el cuadrupolo transmite todos los iones sin filtrarlos, y la celda de colisión fragmenta todos los iones utilizando varias energías de colisión. Posteriormente el software de tratamiento de datos deconvolucionará las señales de cada uno de los canales (1 canal = 1 energía colisión), para

agrupar todos aquellos "iones producto" que presentan el mismo perfil cromatográfico y asociarlos con el "ión precursor" que aparezca al mismo tiempo (en el canal sin energía de fragmentación). Para más detalles ver apartado 5.3.

4.5. Posibilidades del Tratamientos de datos.

Un fichero de datos de TOF y QTOF de una muestra compleja, contiene un nivel de información enorme. Será muy importante disponer de un muy potente software que proporcione herramientas para extracción automática de la información útil de los compuestos de interés, y que ésta se pueda obtener de una manera sencilla, rápida y con la mínima intervención del usuario. El adecuado manejo de herramientas como las citadas a continuación, permitirá al usuario ahorrarse muchas horas de trabajo en el procesado de la información procedente de un TOF o QTOF:

- Algoritmos de deconvolución para la detección y extracción de la información todos los compuestos ionizados en la muestra. Este punto se desarrollará en el siguiente apartado.

- Algoritmos de extracción automática de todos los espectros de MS/MS o Auto MS/MS adquiridos de los compuestos ionizados.

- Sistemas de búsqueda masiva de compuestos a partir de su fórmula molecular (/o su masa exacta). Creación de bases de datos de masa exacta/fórmulas moleculares con opción a tiempos de retención.

- Algoritmos de cálculo de las posibles fórmulas moleculares, con capacidad de discernir entre ellas por comparación de los perfiles isotópicos teóricos y experimental; ello será muy útil para identificar compuestos desconocidos o confirmar identificaciones.

- Identificación mediante bibliotecas de espectros de MS/MS, bases de datos de fórmulas moleculares y opcionalmente tiempos de retención.

- Software para cuantificación de alta capacidad, con potentes y agiles herramientas de auto-revisión de los resultados, que alerten al usuario de los datos a revisar manualmente.

Estos módulos de software se pueden complementar con paquetes de software adicionales para aplicaciones más específicas:

- Software Bioinformático para identificación y cuantificación de proteínas: motores de búsqueda para identificación de proteínas nativas y de modificaciones post-traduccionales, algoritmos de secuenciación De Novo de péptidos, herramientas para cuantificación relativa con o sin etiquetado, comparación de niveles de expresión,….

- Software de análisis estadístico multivariante para análisis comparativo de diversas poblaciones de muestras en busca de biomarcadores: técnicas sin supervisión como el análisis de componentes principales (PCA), o supervisadas para clasificar muestras como PLS-DA.

- Software de análisis de rutas metabólicas para evaluar las rutas en las que están involucrados los biomarcadores encontrados, o visualizar cambios de expresión entre diversas poblaciones en las proteínas, péptidos, metabolitos, … involucrados en una determinada ruta metabólica.

- Softwares para estudios de metabolización de un fármaco/compuesto administrado a un organismo; por ejemplo para identificar los metabolitos que se forman a partir de un principio activo.

- …

Para más información sobre las 3 últimas opciones consultar apartado 3.1.1.2.

4.5.1. Determinación automática de compuestos ionizados y cribados masivos.

En un fichero de LC/MS-TOF/QTOF de una muestra biológica adquirido en la modalidad MS, se pueden llegar a detectar miles de compuestos ionizados: proteínas, péptidos, metabolitos,…. Número que va aumentado a medida que la sensibilidad de los instrumentos aumenta. Para extraer toda esta información, y diferenciarla del ruido de fondo, se deberá recurrir a algoritmos de deconvolución que resuelvan espectralmente los continuos solapamientos cromatográficos. Estos algoritmos se pueden clasificar en 2 grupos, los de simple deconvolución (se puede aplicar a todo tipo de MS) y los que adicionalmente agrupan los iones de acuerdo con criterios de "correlación química" (estos requieren masa exacta). Estos algoritmos proporcionarán la masa , el tiempo de retención, el espectro de masas deconvolucionado ("limpio" de los otros compuestos coeluyentes) y cromatogramas de iones específicos de los compuestos en cuestión.

4.5.1.1. Deconvolución a nivel cromatográfico.

Este tipo de algoritmo genérico de deconvolución cromatográfica, es aplicable a todo tipo de espectrómetros de MS (LC/MS, GC/MS con impacto electrónico); analiza a lo largo del cromatograma (/tiempo) como las distintas masas (m/z) "van subiendo y bajando" su intensidad, agrupando todas aquellas que lo hacen conjuntamente (ver figuras 20 y 21 en apartado 3.1.1.2). Ello permitirá resolver espectralmente los solapamientos cromatográficos, siempre y cuando exista una mínima separación cromatográfica entre los compuestos solapados. Si el solapamiento de 2 compuestos es "perfecto", éstos se consideraran como un único compuesto. Algunos nombres comerciales de estos algoritmos: "Mass Hunter Find by Chromatogram Deconvolution", AMDIS, Dissect,….

4.5.1.2. Deconvolución mediante "Molecular Feature Extractor"

El algoritmo "Mass Hunter Molecular Feature Extractor" (MFE) para LC/MS-TOF y QTOF (específico para masa exacta), en una primera etapa agrupa todas aquellas masas exactas que "suben y bajan" conjuntamente, para en una segunda fase agrupar por "correlaciones químicas" entre ellas (contribuciones isotópicas, aductos, multímeros, iones con múltiples cargas, pérdidas neutras,….), todas aquellas masas exactas que "suben y bajan" conjuntamente. Estas correlaciones químicas pueden ser definidas por el propio usuario. La gran ventaja de este algoritmo, y especialmente con muestras complejas, es su capacidad de resolver espectralmente compuestos que incluso estén perfectamente solapados. Es un algoritmo pensado para técnicas de ionización suave, con masa exacta y con poca fragmentación; situación típica en LC/MS con fuentes de ionización a presión atmosférica: electrospray, nano-ESI, APCI, APPI,… En caso de compuestos que fragmenten y su pérdida neutra no se haya especificado, aparecerán reflejados como 2 o más "features" que coinciden exactamente al mismo tiempo, y de las cuales, sólo 1 de ellas presenta las correlaciones típicas del ión "quasi" molecular: aductos (polaridad positiva: sodio, amonio,..//pol. negativa: formiato, acetato,…), multímeros (1M+H, 2M+Na, 3M+H,…) e ión molecular con múltiples cargas.

4.5.1.3. Cribados Masivos y uso de Bases de datos de masa exacta.

Para localizar automáticamente listados de compuestos, una potente herramienta ofrecida por algunos softwares, es la realización de cribados masivos a partir de ficheros de datos adquiridos en modo MS con masa exacta. Para ello bastará crear una base de datos, con las fórmulas moleculares/ masa exacta de los péptidos de interés (en algunos casos pueden crearse directa-

mente a partir de una hoja de cálculo); lo ideal es utilizar la fórmula molecular para poder considerar como criterio confirmativo el "perfil del ratio isotópico". A partir de aquí se ofrecen 2 alternativas:

- Opción A: Extraer e integrar automáticamente los cromatogramas de iones de las moléculas protonadas o desprotonadas, posibles aductos, iones con múltiples cargas,… de todos los compuestos del listado.

- Opción B: utilizar algún software de deconvolución para determinar los compuestos ionizados, y contrastar estos con el listado de interés.

Estos cribados se fundamentan en la selectividad proporcionada por los cromatogramas de iones con muy finas ventanas de extracción, y en la confirmación inicial vía perfil isotópico; la confirmación final convendría hacerla mediante el tiempo de retención de patrones y sus espectros de MS/MS.

4.5.2. Determinación de fórmulas moleculares.

Para la identificación de pequeñas moléculas (también para péptidos pequeños), las etapas para determinar su fórmula molecular son:

- Calcular las posibles fórmulas moleculares a partir de la masa exacta del compuesto (y una pequeña ventana de tolerancia). En el caso de proteómica ésta técnica sólo se puede aplicar a aminoácidos o a péptidos muy pequeños; el número de posibles fórmulas se incrementa enormemente con el incremento de peso molecular; para moléculas de más de mil Daltons el número de posibles fórmulas ya suele ser excesivo, a menos que se puedan descartar un montón de ellas en alguna de las siguientes etapas.

- Descartar fórmulas que no cumplan con el perfil isotópico teórico; asegurándose que el espectro no esté saturado (deformaría el perfil isotópico).

- Si se dispone del espectro de MS/MS del compuesto, se podrán descartar fórmulas adicionales (o confirmar la propuesta) mediante el cálculo de las fórmulas moleculares de los fragmentos; la suma de cada una éstas con su correspondiente pérdida neutra, deberán coincidir con la fórmula del ión precursor.

Una vez se dispone de una (o unas pocas) fórmulas moleculares, para asignarles estructuras se procederá a buscar en bases de datos de posibles estructuras moleculares:

- ChemSpider: http://www.chemspider.com/
- PubChem (NCBI): http://pubchem.ncbi.nlm.nih.gov/search/search.cgi#
- ChemIDplus: http://chem.sis.nlm.nih.gov/chemidplus/ProxyServlet?chemidheavy
- ….

Si se dispone del espectro de MS/MS del compuesto a identificar, existen softwares que permiten "fragmentar teóricamente" las posibles estructuras moleculares encontradas en la etapa anterior, y clasificarlas de acuerdo con su compatibilidad con el espectro de MS/MS.

A partir de ahí la siguiente etapa en la identificación sería el cromatografiar un patrón y obtener su espectro de MS/MS; a partir de la igualdad del tiempo de retención y espectro de MS/MS se podría considerar al pico en cuestión como "inequívocamente" identificado.

En el caso de péptidos y proteínas, en lugar de buscar directamente su composición molecular, se puede recurrir a previamente buscar su composición en aminoácidos, y a partir de ahí su composición molecular (ver siguiente apartado).

4.5.3. Determinación de la composición en aminoácidos de péptidos.

La combinación de 20 aminoácidos esenciales y el que muchos péptidos superan con creces los 1.000Da, genera un elevado número de posibles combinaciones de aminoácidos que encajen con una determinada masa exacta; usualmente muy superior al número de posibles fórmulas moleculares ($C_xH_yN_zO_pS_q$…). Por ello algunos software incorporan toda una serie de herramientas para reducir el número de posibilidades, entre ellas el uso de información del espectro de MS/MS (ver ejemplo en figura 38), que revelen la presencia de ciertos aminoácidos; para ello permiten:

- Poder seleccionar (y también modificar) los aminoácidos a considerar.
- Seleccionar el tipo de ión utilizado del espectro de MS (lón "quasi" Molecular), iones producto en el MS/MS (b,y,…), su número de cargas,…
- El tipo de iones imonio y pérdidas neutras observadas en su espectro de MS/MS.

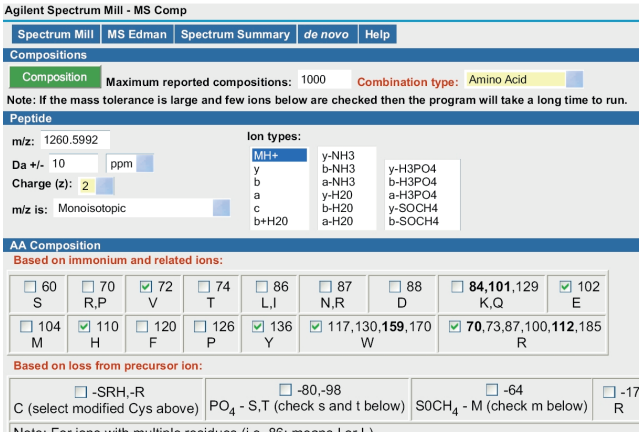


Figura 38: Ejemplo de parte de los parámetros (software Spectrum Mill) para determinación de la composición en aminoácidos de un péptido. Además de búsquedas a partir del péptido protonado, también se puede buscar a partir de datos del espectro de MS/MS, como diferentes tipos de iones y pérdidas neutras, así como de iones imonio y pérdidas observadas. También es posible seleccionar los aminoácidos a considerar (no mostrado en la figura).

Para reducir posibilidades, en el espectro de MS/MS a m/z inferiores a 200, es muy útil detectar los iones imonio, al revelar la presencia de aminoácidos concretos. Si el péptido procede de una digestión tríptica, deberá contener Lys y/o Arg.

Otras alternativas para determinar la composición del péptido por LC/MS con masa exacta son:

- Secuenciar el péptido a partir de su espectro de MS/MS (Secuenciación De Novo); siempre y cuando el péptido no sea muy grande (menos de 15-20 aminoácidos).

- Si el péptido procede de la digestión de una proteína, buscar en bases de datos de proteínas para proceder a su identificación.

4.5.4. Cuantificación en modos MS y MS/MS.

Esta sección resume y suplementa la sección 3.3 sobre Proteómica Cuantitativa. Los espectrómetros de masa exacta basados en tiempo de vuelo (LC/MS-TOF/QTOF), adquieren siempre el espectro completo (en un rango de m/z); la selección de las masas y ventanas de extracción, para crear los cromatogramas de iones utilizados para cuantificar, se realizará a posteriori (excepto la de los precursores en modo targeted MS/MS).

Cuantificación en modo MS (LC/MS-TOF/QTOF).

Dada la universalidad y selectividad del modo MS con masa exacta, y la mayor intensidad de sus iones (con mínima fragmentación); el modo MS suele ser el modo preferido para cuantificar

utilizando cromatogramas de iones específicos, extraídos utilizando muy finas ventanas de extracción de masa (unas pocas centésimas de Dalton, o decenas de ppm).

Cuantificación en modo MS/MS (LC/MS-QTOF).
El modo "targeted MS/MS" proporciona una selectividad adicional pero requiere conocer "a priori", el compuesto, su tiempo de retención, y las condiciones óptimas para aislarlo (en el Q) y fragmentarlo (en celda colisión). La selectividad adicional proporcionada por el MS/MS, no siempre compensa la pérdida de intensidad de los iones producto con respecto a los precursores (múltiples fragmentos "versus" 1 ión no fragmentado), y en bastantes ocasiones el modo MS/MS proporciona menor sensibilidad que el modo MS.

Cuantificación de péptidos y proteínas.
El modo MS puede ser utilizado tanto para cuantificar péptidos como proteínas intactas; en cambio el modo MS/MS puede ser utilizado con péptidos, pero no se recomienda con proteínas intactas, dada la enorme complejidad de sus espectros de MS/MS (gran cantidad de distintos iones producto).

Selección de Masas para Cuantificar.
Se escogerán las masas más intensas del espectro, y con preferencia las de mayor masa (más selectivas); se evitará utilizar masas poco específicas que o comunes a familias de compuestos (excepto para cuantificar precisamente esas familias). Para cuantificar con MS/MS, si los hubiera, se escogerán preferentemente iones producto de elevada intensidad a m/z por encima del precursor; estos son más selectivos (sólo pueden proceder de iones precursores con múltiples cargas). Si los espectros contienen múltiples iones con elevada intensidad, como suele ser el caso de péptidos y sobre todo de proteínas intactas, se puede cuantificar con la suma de los iones más intensos.

Selección de Masas Cualificadoras para Confirmar la Identidad del Compuesto Cuantificado.
Para confirmar que el compuesto que se está cuantificando sea realmente el que interesa y no un interferente, o detectar la presencia de una posible interferencia, es muy útil comprobar que el ratio de las respuestas entre las distintas masas del analito presentes en el espectro, con respecto a la respuesta cuantificadora, se mantiene constante entre patrones y muestras. En otras palabras, sería comprobar que el espectro de MS deconvolucionado o el de MS/MS del analito coinciden entre la muestra y el patrón. Los módulos de análisis cuantitativo incorporan herramientas para realizar automáticamente estos cálculos, y avisar al usuario cuando estos ratios exceden los límites fijados. Estos límites variarán en función de la proporción de las respuestas; en proteómica no hay unos límites pre-establecidos, pero se podrían considerar los que marca la normativa de confirmación de residuos (fármacos, pesticidas,..) en alimentos (ver tabla 7). Se requieren, además de la igualdad de tiempos de retención, al menos 4 iones con masa nominal trabajando en modo MS, o 2 con masa exacta que proporcionen información sobre la estructura de la molécula. Trabajando en modo MS/MS se requiere cualificar 2 transiciones específicas. Los espectros de MS/MS de péptidos dan un número de iones producto muy superior al de las pequeñas moléculas, al trabajar con QTOF con espectros de MS/MS completos, se podrán

Tabla 7.-Tolerancia recomendada para confirmar la identidad del analito cuantificado, mediante el uso del ratio de las transiciones de iones cualificadores.

Intensidad Relativa del Ión Cualificador (% respecto Pico Base)	Tolerancia aceptada en LC/MS para Ratio de Iones Cualificadores
>50%	+/- 20%
20-50%	+/- 25%
10-20%	+/- 30%
≤10%	+/- 50%

escoger todos los iones producto que interesen, para confirmar la identidad del compuesto cuantificado con las mayores garantías.

Selección del Ancho de las Ventanas de Extracción de Cromatogramas de Iones para Cuantificar.
Con datos de adquisición en modo "profile", para definir la ventana de extracción será conveniente tener en cuenta el ancho del pico de masa del patrón, y la robustez de la exactitud de masa; si se escoge una ventana excesivamente fina se puede ganar en selectividad, pero se puede perder reproducibilidad y exactitud en la cuantificación. En equipos con 1-3ppm de exactitud y 10.000-40.000 de resolución, unos rangos típicos de ventanas de extracción para cuantificar estarían en el rango de +/- 10-50ppm de la m/z utilizada. Cuanto mayor sea la exactitud y resolución, más finas serán las ventanas que se pueden utilizar. El ancho del pico de masa al 50% de su altura (pk wd 50% expresado en Daltons/carga) se puede calcular mediante la fórmula:

$$\text{pk wd } 50\% = (m/z) / \text{Resolución (FWHM)}$$

Un péptido de m/z=1.000 con resolución 40.000, generará un pico de masa con un ancho al 50% de su altura de 25 mDa (o 25ppm).

Optimización Frecuencia de Adquisición y Relación Señal/Ruido: una buena cuantificación, exige una buena relación señal/ruido y un nº suficiente de puntos de medida; al menos 12-15 puntos/pico, que se reducen a 8-10 en el caso de trazas. Teóricamente:

- El ruido es inversamente proporcional a la raíz cuadrada del nº de promedios efectuados, es decir disminuye con la reducción de la frecuencia de adquisición. Pasar de adquirir a una frecuencia 100ms/pto. a otra de 400ms/pto. el ruido se reducirá a la ½ (ver tabla 8).

- En los TOF/QTOF que suman "transients", el reducir la frecuencia de adquisición, aumentará la intensidad de la señal (al sumarse un mayor número de "transients").

- Un nº de puntos de muestreo del pico excesivamente bajo, incrementará el error de la medida; para un coeficiente de variación mejor del 3% –desde el punto de vista de perfilado del pico- se requerirán al menos 12 puntos de medida (ver tabla 9).

- Orientativamente el impacto de la relación señal/ruido (S/N), con respecto al coeficiente de variación, se puede estimar por la fórmula:

- **RSD% = 100/ (S/N)** con ruido calculado como pico a pico

- Doblar la relación señal/ruido permitirá reducir a la mitad, su influencia en el coeficiente de variación de la respuesta (ver Tabla 10).

- Para cuantificar péptidos a concentraciones que no den buena relación señal/ruido, será importante ajustar la frecuencia de adquisición entre 8 y15 puntos por pico.

- Para cuantificar picos cromatográficamente solapados (poco usual en MS dada su selectividad), especialmente si están muy desproporcionados, puede convenir doblar el nº de puntos de muestreo, ajustándolo a 25-35 puntos/pico, para no perder resolución cromatográfica.

Tabla 8.- Relación relativa entre, el tiempo de muestreo de la señal (ms promediados) o frecuencia (puntos muestreo /segundo), y el ruido teórico de la línea de base.

Frecuencia de Adquisición: (tiempo de adquisición por punto de medida)	Ruido teórico de la línea de base expresado relativamente
100 ms - 10Hz	1
400 ms – 2.5Hz	1/2
900 ms – 1.1Hz	1/3
2.500 ms – 0.4Hz	1/5
10.000 ms - 0.1Hz	1/10

Tabla 9.- Impacto del perfilado de un pico cromatográfico en su coeficiente de variación: relación entre el coeficiente de variación, y el mínimo número de puntos de medida necesarios a lo largo del pico, para su adecuado perfilado.

RSD% requerido	Puntos/pico necesarios
<1%	Mínimo 15
<1-3%	Mínimo 12
<10-12%	Mínimo 8

Tabla 10.- Impacto del ruido de un pico cromatográfico en su coeficiente de variación: relación entre el coeficiente de variación, y la mínima relación señal/ruido necesaria (ruido calculado como pico a pico).

RSD% requerido	Mín. S/N (p/p)
1%	100
2%	50
5%	20
10%	10

Cuantificación Absoluta mediante Curvas de Calibrado.

Éstas se crearán para cada uno de los analitos de interés, a partir de patrones de concentración conocida. Los patrones suelen prepararse como mezclas de los diferentes compuestos a cuantificar. La matriz de los patrones utilizados para recalibrar las curvas de calibrado, deberá emular de la mejor manera posible, la matriz de las muestras cuantificadas. Así se conseguirá minimizar los cambios en la respuesta, producidos por influencia de la supresión o exaltación de la ionización, cuando se producen cambios importantes en las coeluciones de los analitos de interés con compuestos mayoritarios.

Si no se puede disponer de blancos de matriz de muestra que no contengan los compuestos de interés, o los tengan a muy bajos niveles, una alternativa es utilizar un “pool” de muestras como base (a poder ser con bajos niveles de los compuestos de interés), para crear una curva de calibrado mediante la técnica de adición de estándares.

Cuantificación de muestras mediante técnica de Adición de Estándares.

Es muy útil para crear patrones adecuados cuando no se dispone de buenos blancos de matriz de las muestras. Consiste en adicionar a alícuotas de la muestra a cuantificar, distintas cantidades conocidas de los analitos a cuantificar. A partir de ahí se construye la curva de calibrado (atribuyendo a cada nivel de concentración la cantidad adicionada). Para aquellos analitos cuya concentración en la muestra no sea cero, la correspondiente curva de calibrado (respuesta “versus” concentración), no pasará por el origen de coordenadas (0,0), sino que cortará al eje de concentración, en un punto correspondiente a valores negativos. Esta concentración negativa, es la que corresponderá a la muestra que se pretendía cuantificar. También se puede calcular a partir de la ecuación de la recta de calibrado (calculando C para respuesta 0). Al utilizarse la propia muestra para crear el patrón, es un sistema ideal para eliminar las interferencias cuantitativas por efecto matriz; los inconvenientes: incrementa el número de inyecciones por muestra a cuantificar, y el consumo de patrones.

Cuantificación mediante Curvas de Calibrado utilizando Patrones Internos (ISTD).
Otra alternativa para compensar las interferencias cuantitativas por efecto matriz, es la adición a patrones y muestras de una cantidad conocida de patrones marcados isotópicamente de los analitos a cuantificar; éstos deberán coeluir perfectamente con los analitos no marcados, y se resolverán espectralmente por su distinta masa. Habitualmente éste se adiciona a la misma concentración en muestras y patrones; se recomienda un nivel medio, pero que dé una buena relación señal/ruído (p.e. superior a 300:1 considerando el ruido como "pico a pico" o 5 veces el RMS) para asegurar una óptima cuantificación del ISTD. El ratio de concentraciones analito/ ISTD, ha de ir cambiando con los distintos niveles de concentración; por tanto, NO se podrá preparar por sucesivas diluciones de una solución madre que también contenga el patrón interno (error bastante habitual).

Para usar como patrones internos, hay firmas que ofrecen el poder sintetizar péptidos a medida marcados isotópicamente, y en los que 1 de sus aminoácidos está constituido en más de un 98% por ^{13}C y ^{15}N (en lugar de ^{12}C y ^{14}N).

Cuantificación Relativa mediante Rectas de Calibrado.
Una muy fácil cuantificación relativa con rectas de calibrado, es asignar una concentración arbitraria, por ejemplo 100, a una de las muestras de un nivel de concentración medio, o a un "pool" (mezcla a partes iguales) de las muestras a analizar; después se cuantifica relativamente el resto, a partir de ésta considerada como patrón. Si en lugar de 1 único punto de calibración se quieren utilizar varios, se debería partir de una de las más concentradas y diluirla con "matriz blanca"; si no se dispone de ésta, pueden producirse interferencias cuantitativas por las sucesivas diluciones de la matriz. En este caso, se debería comprobar que el factor de respuesta (ratio respuesta/ concentración) no cambie con la concentración; hay que diferenciar entre los cambios de factor de respuesta por efecto matriz, a los debidos a saturación de la linealidad del electrospray, o de la respuesta del detector.

Cuantificación Relativa mediante Marcaje Isotópico o Etiquetado Isobárico.
Las técnicas de etiquetado isotópico: ICAT, SILAC, $^{16}O/^{18}O$, … o isobárico (p.e. ITRAQ), se describen el apartado 3.3 y más en profundidad en otro capítulo de este libro. Utilizan módulos de software específicos para proteómica, incluidos en algunos casos en los propios motores de búsqueda (ver figura 30 y referencia 20).

5. Últimos Desarrollos Tecnológicos: Cromatografía Líquida / Movilidad Iónica / Cuadrupolo-Tiempo de Vuelo.

Con el objetivo de mejorar la capacidad de resolver muestras de máxima complejidad por LC/ MS, recientemente se han introducido en el mercado distintos dispositivos e instrumentos que combinan la espectrometría de masas con una técnica de separación ortogonal y ultra-rápida, como la espectrometría de movilidad iónica. La combinación de ambas tecnologías proporcionará una dimensión adicional que multiplicará su capacidad de separación (permitiendo resolver incluso isómeros) y mejorará aún más la selectividad proporcionada por el instrumento.

La combinación con cromatografía permitirá disponer de 3 dimensiones de separación (Tiempo – Movilidad Iónica – Masas), cuyas velocidades se acoplan muy bien cuando se utiliza un espectrómetro de masas de Tiempo de Vuelo, por su rapidez en la adquisición de un espectro (microsegundos). La escala de microsegundos permitirá obtener espectros de MS de un ion se-

parado por movilidad iónica en una escala de milisegundos; a su vez ésta permitirá obtener un espectro de movilidad iónica de un pico cromatográfico separado en una escala de segundos.

5.1. Fundamentos y Tipos de Instrumentos.

Mientras en el tubo de vuelo de un espectrómetro de masas (TOF o QTOF) la separación de moléculas ionizadas tiene lugar en una zona con muy elevado vacío, la separación de iones por movilidad iónica tiene lugar en una zona que contiene un gas de amortiguación a una determinada presión. Consecuentemente, mientras en el espectrómetro de masas la velocidad de los iones depende sólo de su densidad de carga (relación masa/carga), en un espectrómetro de movilidad iónica la separación (/velocidad) de los iones también dependerá de su forma y tamaño; además de su masa y carga. Cuanto mayor sea la sección transversal del ión, mayor será el número de colisiones con las moléculas de gas amortiguador y menor su velocidad; así del espectro de IM (movilidad iónica) se obtendrá información relacionada con su sección transversal (forma y tamaño), que complementará la información sobre composición molecular proporcionada por el espectro de MS, y de la estructural (distribución de los átomos y grupos funcionales) proporcionada por el espectro de MS/MS.

Existen diversas opciones para la combinación de espectrometría de masas y movilidad iónica, destacar:

- **Filtros de Movilidad Iónica: FAIMS** (Field Asymmetric waveform Ion Mobility Spectrometry) - también denominada **espectrometría de movilidad diferencial**. El acoplamiento se realiza mediante dispositivos que se introducen a nivel de la fuente de iones, para **filtrar en continuo los iones por diferencias en movilidad iónica** antes de entrar en el interior del espectrómetro de masas. Estos filtros **operan a presión atmosférica y sólo son atravesados por iones con un determinado ratio de movilidad iónica** entre los campos eléctricos de alto y bajo voltaje utilizados. El campo eléctrico se aplica entre 2 electrodos mediante la combinación de una radiofrecuencia (voltaje variable) con una forma de onda asimétrica y un voltaje fijo (corriente continua). La aportación más importante de estos dispositivos es mejorar la selectividad y reducir el ruido de fondo. Su nivel de resolución en movilidad iónica es en general menor que el de los analizadores de tiempo de deriva descritos a continuación.

- **Analizadores de Movilidad Iónica o Tiempo de Deriva** (Drift Time IM): en este caso el analizador de movilidad iónica (tubo de deriva) va situado en el interior del espectrómetro de masas, trabaja a presión reducida, y no se trata de un filtro, sino de un dispositivo que es **atravesado por todos los iones que saldrán del tubo de deriva a distintos tiempos,** proporcionando la movilidad iónica o tiempo de deriva de todos los iones que entran en el analizador. Por ello, antes del analizador se deberá disponer de un sistema de atrapado y retención de los iones que provienen de la fuente en continuo; éste los acumulará durante un cierto tiempo, para liberarlos en forma de pulso de iones cuando el analizador de movilidad iónica ya ha terminado de separar los del ciclo/pulso anterior. En los acoplamientos con espectrómetros de masa de tiempo de vuelo, existen diversos diseños:

  - **Analizador situado entre el cuadrupolo y el tiempo de vuelo (Q-IM-TOF):** utiliza un pequeño tubo de deriva al que se aplica campo eléctrico no uniforme mediante la aplicación de radiofrecuencia con una onda progresiva a lo largo del tubo de deriva. Estos equipos proporcionan un nivel de resolución media (en movilidad iónica), y pueden operar tanto en modo Q-IM-TOF como en modo QTOF clásico.

  - **Analizador situado antes del cuadrupolo (IM-QTOF)**. Este tipo de analizadores de desarrollo muy reciente (41-42), incorporan un gran tubo de deriva al que se aplica un campo eléctrico uniforme de baja intensidad (corriente continua). Proporcionan la ventaja de:

- Una **alta resolución en movilidad iónica**.

- Poder trabajar en modo IM-QTOF, haciendo MS/MS de iones precursores (p.e. isobáricos) separados por movilidad iónica. También permiten trabajar en las distintas modalidades de un QTOF clásico.

- Proporcionan directamente una **exacta medida absoluta de la sección transversal de colisión** (CCS- "collision cross section" o Ω), sin necesidad de calibrar con patrones. Esto se consigue al utilizar un campo eléctrico uniforme de baja intensidad, en lugar de utilizar radiofrecuencias (este tipo de instrumentos requieren calibración con patrones). La medida de la sección transversal se puede utilizar como otro parámetro para corroborar la identificación de la molécula.

- Conserva las características estructurales de las conformaciones moleculares, gracias a utilizar un campo eléctrico uniforme de baja intensidad, en lugar de utilizar radiofrecuencias (la radiofrecuencia eleva la temperatura de los iones en el tubo de deriva, modificando la conformación molecular de las estructuras más lábiles).

- Una **alta sensibilidad** (tanto en modo IM-QTOF como en modo QTOF) cuando se combina con los innovadores diseños de captación y focalización de iones mediante embudos electrodinámicos, descritos en el siguiente apartado.

5.2. Descripción de un Sistema de Movilidad Iónica acoplado con Cuadrupolo Tiempo de Vuelo (LC/IM-QTOF).

En el sistema LC/IM-QTOF descrito a continuación (43), el **analizador de movilidad iónica de alta resolución** (superior a 60 para pequeñas y grandes moléculas) y elevada sensibilidad, consiste en un tubo de deriva ("drift tube") con una longitud de unos 80 cm, al que se aplica un campo eléctrico uniforme de baja intensidad (típico unos 20V/cm). Como gas amortiguador utiliza nitrógeno y el tubo de deriva está situado entre la fuente de ionización y el primer cuadrupolo.

Descripción del proceso de Movilidad Iónica.

El campo eléctrico uniforme de baja intensidad (unos 20V/cm) del interior del tubo de deriva, genera una fuerza de arrastre que acelera y mueve a los iones a través del dispositivo, mientras que las colisiones de estos iones con las moléculas de nitrógeno, generan una fuerza de freno que actúa en contra de la fuerza eléctrica que los mueve. En campos eléctricos de baja intensidad, se alcanza rápidamente un estado de equilibrio entre ambas fuerzas, y los iones empiezan a moverse con velocidad constante (V); ésta es proporcional al campo eléctrico aplicado (E): $V = K \times E$. Donde la constante de proporcionalidad (K) es la movilidad en fase gaseosa del ión; esta constante (K) específica de cada ión, depende de su densidad de carga (masa/carga) y sección, de la temperatura, presión y masa del gas amortiguador, de la longitud del tubo de deriva y del campo eléctrico uniforme (V/cm) aplicado.

En definitiva, la velocidad resultante de los iones en el tubo de deriva dependerá de sus secciones transversales de colisión (en función de su tamaño y forma), carga eléctrica y masa:

- Los iones con múltiples cargas reciben una mayor fuerza debido al campo eléctrico, y se mueven a través del gas amortiguador con más efectividad/velocidad que las cargadas individualmente.

  - En el caso de péptidos y proteínas, ello genera una separación de los mismos en la dimensión de movilidad iónica, según la distribución de su número de cargas.

- Los iones con secciones transversales mayores son más retrasados por su mayor número de colisiones con el nitrógeno del tubo de deriva, y avanzarán más lentamente a través del tubo de deriva.

  - En el caso de péptidos y proteínas, aquellas isoformas que modifiquen su sección transversal, se separarán en el espectro de movilidad iónica; lo mismo pasará con complejos de proteínas con otras proteínas, péptidos o moléculas.

El instrumento utiliza un tubo de deriva en condiciones límite de campo eléctrico de baja intensidad (V/cm), para permitirle generar una precisa información estructural de los compuestos. **En estas condiciones límite, la movilidad iónica depende de la estructura de la molécula y su interacción con el gas amortiguador**, no del campo eléctrico aplicado.

Descripción del Espectrómetro IM-QTOF: El espectrómetro IM-QTOF descrito en la figura 39, muestra siguiendo el recorrido de los iones:

1.- Fuente de ionización Electrospray mejorado (Jet Stream con tecnología i-Funnnel): se encargará de ionizar a las moléculas procedentes del U/HPLC. La fuente puede intercambiarse por otras, como nanoSpray o HPLC/Chip, APCI,.., dependiendo del tipo de cromatografía utilizada o la polaridad de los compuestos que interese ionizar.

2.- Capilar con 6 orificios: dispositivo que mediante un gradiente de campo eléctrico y un diferencial de vacío entre sus extremos, capta a los iones y aísla e independiza las condiciones de ionización de las detección; proporcionando una suave transición de la zona de presión atmosférica (fuente) a la primera etapa de vacío en el interior del espectrómetro de masas.

3.- Embudo Electrodinámico frontal: su función es focalizar y enriquecer los iones de la muestra, eliminando el exceso de gas. Trabaja a alta presión y utiliza su forma y voltajes de radiofrecuencia para focalizar a los iones hacia el centro del embudo, y un voltaje de corriente continua para impulsar a los iones hacia el siguiente embudo de captura ("trapping funnnel").

4.- Embudo Electrodinámico de captura: su función consiste en convertir el haz de iones en continuo que genera la fuente, en un haz pulsante de iones (haz discontinuo) que necesita el analizador de movilidad iónica para poder separar y medir la movilidad iónica de los iones que le entran. Este embudo primero refocaliza y almacena los iones, y luego los libera en paquetes discretos hacia el tubo de deriva. Para ello utiliza un dispositivo de doble rejilla, que maximiza la cantidad de iones que se puede enviar en cada pulso, y minimiza la duración del pulso (para reducir la dispersión de los iones). Combinado con un tubo de unos 80cm proporcionará un analizador de movilidad iónica de alta resolución y muy alta sensibilidad. Este dispositivo de doble rejilla, hace la función de puerta de entrada; retiene y acumula a los iones mediante un pulso de voltaje durante típicamente entre 60 y 100 milisegundos. Posteriormente los inyecta en forma de un pulso muy estrecho hacia el analizador.

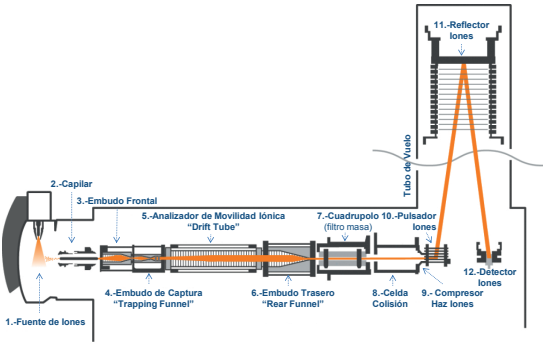


Figura 39: Esquema de un Espectrómetro IM-QTOF de Movilidad Iónica acoplado a un Cuadrupolo-Tiempo de Vuelo con tecnología i-Funnel (Agilent): el analizador de movilidad iónica de alta resolución se encuentra entre la fuente y el cuadrupolo (para poder obtener espectros de MS/MS de iones separados por movilidad). Para proporcionar una elevada sensibilidad, la focalización y captura de iones antes y después del analizador, se realizada mediante la tecnología de embudos electrodinámicos. A partir del cuadrupolo, se encuentran ya los tradicionales componentes de un sistema QTOF: celda de colisión (hexapolar), compresor del haz de iones y reflectrón, entre otros.

5.- Analizador de Movilidad Iónica ("Drift Tube"): su función es separar a los iones según su movilidad cuando son impulsados (según su masa/carga) por un campo eléctrico uniforme de bajo voltaje (corriente continua), y frenados por su impacto con las moléculas de nitrógeno presentes en el tubo de deriva (baja presión), según su sección transversal, tal como ya se explicó al principio de éste apartado en "Descripción del proceso de Movilidad Iónica".

6.- Embudo Electrodinámico Trasero: reorienta los paquetes de iones en un haz más estrecho antes de entrar en la óptica del espectrómetro de masas. La elevada eficiencia de captura de iones del conjunto de embudos electrodinámicos utilizados da como resultado una **sensibilidad de 1000 veces mejor con respecto a otros diseños anteriores sin embudos**.

7.- Cuadrupolo → 8.-Celda de Colisión→ …Tubo Vuelo…. →12.-Detector de Iones. En estas zonas (7 a 12) tienen lugar los tradicionales procesos de un QTOF:

- En modo MS/MS: aislamiento de un precursor en el cuadrupolo y fragmentación en la celda de colisión.

- En modo MS: transmisión en cuadrupolo y celda de colisión de todos los iones entrantes sin filtrarlos ni fragmentarlos.

Después los iones que salen de la celda de colisión se focalizan y se envían a volar (mediante el pulsador de iones) en la zona del tubo de vuelo para poder determinar su relación masa/carga (m/z) con gran exactitud. Todos estos procesos ya fueron explicados con detalle en el apartado 2.1 de este capítulo, y en concreto en la descripción de las zonas 3, 4 y 5.

Este tipo de diseño (IM-QTOF) con el analizador de movilidad situado antes del cuadrupolo, ofrece numerosas modalidades de trabajo, que se resumen en el siguiente apartado.

5.3. Modalidades de Trabajo de un Sistema de Movilidad Iónica acoplado con Cuadrupolo Tiempo de Vuelo (LC/IM-QTOF).

En sistemas de Movilidad Iónica acoplados a tiempo de vuelo, como el descrito en el apartado anterior (con el analizador de movilidad antes del cuadrupolo), se podrá trabajar en los siguientes modos:

- **IM-TOF** (IM-MS)**:** se separan los iones por su movilidad iónica y se mide su masa exacta en el tubo de vuelo (QTOF en modo MS).

- **IM-QTOF** (IM-MS/MS)**:** se separan los iones por su movilidad iónica, se aísla un precursor (en el cuadrupolo), se fragmenta (en la celda de colisión) y se mide la masa exacta de los fragmentos obtenidos en el tubo de vuelo (QTOF en modo MS/MS). En MS/MS existen diversas opciones de submodalidades de adquisición:

  - **MS/MS Dirigido** ("Target MS/MS"): el usuario programa los iones precursores que quiere aislar y fragmentar, su tiempo de retención y margen de tiempo durante el cual se quiere adquirir, la/s energía/s de colisión a aplicar (fija/s o por tablas o fórmulas), entre otros parámetros. Para más información consultar apartado 4.4.2 de este mismo capítulo.

  - **MS/MS No Dirigido,** existen 2 submodalidades:

    - Auto MS/MS o MS/MS Dependiente de los Datos ("Data Dependent MS/MS")**:** es un modo que combina automáticamente espectros de MS y MS/MS; el propio

equipo en función de la intensidad de los iones que encuentra en la dimensión de MS, decide que iones aislar y fragmentar en la dimensión MS/MS. La selección la realiza mediante unos algoritmos que el usuario puede perfilar con criterios de selección y preferencias. Para más información consultar apartado 5.4.3 de este mismo capítulo.

- **MS/MS de Todos los Iones o MS/MS Independiente de los Datos** ("MS/MS All ions"): en esta modalidad **el cuadrupolo transmite todos los iones sin filtrarlos**, y **la celda de colisión fragmenta todos los iones**. Es decir, es un **MS/MS sin aislar ningún precursor**. El proceso se realiza en "paralelo" alternando varias energías de colisión, **incluido un canal sin energía de colisión para monitorizar que "iones precursores" se han fragmentado,** supuestamente, en los canales con energías de colisión elevadas. Posteriormente, el software de tratamiento de datos deconvolucionará las señales de cada uno de los canales (1 canal = 1 energía colisión), para agrupar todos aquellos "iones producto" que presentan el mismo perfil cromatográfico y asociarlos con el "ión precursor" que aparezca al mismo tiempo (en el canal sin energía de fragmentación). Este proceso de deconvolución no dirigido, requiere de una mínima separación de tiempo de retención en los perfiles cromatográficos de los iones a deconvolucionar. Aunque haya solapamientos, se requiere de una mínima separación cromatográfica; de ahí la importante mejora que se obtiene **al simplificarse el nº de compuestos que coeluyen cuando se acopla movilidad iónica con esta opción "MS/MS All ions"**. Además la rapidez del proceso "MS/MS All ions" para fragmentar a todos los precursores, la convierte en una modalidad MS/MS idónea para utilizarla con movilidad iónica. Adicionalmente, si se dispone de bibliotecas de espectros de MS/MS de los compuestos de interés, el uso de éstas permitirá mejorar considerablemente la eficacia del proceso de deconvolución que requiere "MS/MS All ions"; al poder realizarse una "deconvolución dirigida" que es siempre más eficiente que una no dirigida, especialmente para compuestos minoritarios en muestras complejas. El algoritmo de deconvolución dirigido aprovecha la información de las masas que aparecen en los espectros de MS/MS de los "iones precursores", localizados en el canal sin energía de fragmentación, para obtener una mejor y más robusta deconvolución.

- **Modos sin Movilidad Iónica: TOF** (QTOF en modo MS) **o QTOF** (modo MS/MS). Toda la parte correspondiente al analizador de movilidad iónica, opera en unas condiciones en las que focaliza y transmite todos los iones –sin separarlos- hasta el primer cuadrupolo. Si éste y la celda de colisión dejan también pasar todos los iones –sin filtrarlos ni fragmentarlos- se trabajará en modo MS (TOF); caso contrario, se trabajará en modo MS/MS en alguna de las submodalidades descritas en el párrafo anterior.

5.4. Aplicaciones Proteómicas de un Sistema de Movilidad Iónica acoplado con Cuadrupolo Tiempo de Vuelo (LC/IM-QTOF).

La movilidad iónica puede aplicarse tanto al análisis de péptidos como de proteínas; en ambos casos aportará ventajas como se describe en los ejemplos de los apartados 5.4.1. y 5.4.2. Las aportaciones de la aplicación de movilidad iónica acoplada a un QTOF se pueden resumir en:

- Aumentar muy considerablemente la capacidad de resolución de muestras muy complejas, como muchas de las habitualmente analizadas en proteómica.

- Poder resolver isómeros y masas isobáricas muy próximas.

- Llegar a separar isoformas, habituales en proteómica por la complejidad de las proteínas.

- Medir directamente la sección transversal de las moléculas o complejos ionizados, preservando las características estructurales de su conformación molecular. Información que puede ser utilizada para confirmar identificaciones.

- Poder asociar una biomolécula desconocida a una determinada clase de biomoléculas, por sus coordenadas en el gráfico de movilidad iónica "versus" relación masa carga (m/z).

- Disponer de una dimensión adicional para mejorar la medida de la pureza de un pico cromatográfico (mayor capacidad de detección de compuestos coeluyentes).

Aumento de la Capacidad de Resolución.
La posibilidad de disponer de 3 dimensiones de separación (Tiempo - Movilidad Iónica – Masa), **aumentará muy considerablemente la capacidad de resolución de muestras muy complejas**, muy habituales en proteómica, y "simplificará" la muestra que llega al espectrómetro de masas (se verá un ejemplo en la sección 5.4.1.). La capacidad de separación resultante es el producto del poder de resolución de las 3 técnicas acopladas, multiplicado por su factor de ortogonalidad.

Resolución de Isómeros y Masas Isobáricas.
La espectrometría de movilidad iónica se complementa muy bien con la espectrometría de masas, al permitir separar masas isobáricas muy próximas e incluso isómeros; éstos NO se pueden resolver en modo MS (con ningún equipo), al tener exactamente la misma fórmula molecular; aunque en algunos casos sí se pueden resolver por MS/MS.

Llegar a Separar Isoformas.
Si como consecuencia de las diferencias entre ellas, las isoformas presentan una sección transversal suficientemente distinta, se podrán separar en la dimensión de movilidad iónica, aunque tengan la misma masa, o muy semejante. Cuando las isoformas presentan múltiples cargas, éstas se suelen diferenciarse mejor entre las que presentan un menor nº de cargas (las de masas más altas), por su mayor diferencia en tiempo de deriva al viajar a menor velocidad.

Resolver Confórmeros.
Al tratarse de la misma molécula con una distinta distribución espacial de los grupos funcionales, éstos no se podrán resolver por masas pero si se podrán llegar a resolver/detectar por movilidad iónica (se verá un ejemplo en la sección 5.4.2.). Dada la labilidad de muchas de estas estructuras, para preservarlas conviene utilizar un tubo de deriva con campo eléctrico de baja intensidad, para que los iones reciban la mínima energía y no se calienten; también en algunos casos puede convenir utilizar condiciones de baja energía en su focalización y transmisión hasta el tubo de deriva.

Separación por Movilidad Iónica de Iones con Múltiples Cargas.
La movilidad iónica, al igual que sucede con la espectrometría de masas, también proporciona la separación de péptidos /proteínas según su número de cargas, pero a medida que aumentan éstas, la diferencia de tiempo de deriva y resolución entre las distintas cargas se reduce (se verá un ejemplo en la sección 5.4.2.). Consecuentemente, cuando interese diferenciar entre isoformas o masas isobáricas con múltiples cargas, en general, se obtendrá una mejor separación en la zona correspondiente a un menor nº de cargas por su menor velocidad.

Medida Directa de la Sección Transversal de Moléculas o Complejos Ionizados.
La capacidad de preservar las características estructurales de conformaciones moleculares, cuando se utilizan tubos de deriva con campos eléctricos de baja intensidad, permitirá obtener directamente la exacta medida absoluta de la sección transversal de colisión (CCS- "Collision

Cross Section" o Ω), sin necesidad de calibrar con patrones. Esta medida podrá ser utilizada, como una medida más, para corroborar la identificación de péptidos o proteínas intactas, cuando se disponga de sus valores teóricos/bibliográficos; asimismo, podrá dar información sobre la asociación o formación de complejos con otras proteínas o moléculas.

Asociación de Clases de Moléculas.

Cuando se representan distintas clases de familias de moléculas (p.e. péptidos, lípidos, carbohidratos, oligonucleótidos,…) en un gráfico de movilidad iónica "versus" relación masa carga (m/z), se observa que los distintos compuestos de la misma clase, aparecen representados sobre una curva de tendencia que suele diferenciarse de la curva que generan otras familias de compuestos, dependiendo de sus estructuras o características funcionales. Esto permitirá en algunos casos, asociar una biomolécula desconocida a una determinada clase de biomoléculas, por sus coordenadas en el gráfico movilidad iónica "versus" relación masa carga (m/z). Asimismo, dentro de una misma familia, cuando ésta proporciona iones con múltiples cargas, se obtendrán distintas curvas de tendencia según su nº de cargas; a medida que aumenta el nº de cargas, esta diferencia entre curvas disminuirá (al reducirse la diferencia de movilidad iónica y de m/z) al aumentar la velocidad de los iones.

Determinación de la Pureza de un Pico Cromatográfico.

La separación en una escala de milisegundos que proporciona el tubo de deriva, permitirá separar y obtener un espectro de movilidad iónica de picos cromatográficos que tienen varios segundos de ancho. De esta manera cuando varios compuestos (con distinta movilidad) coeluyen en 1 pico cromatográfico se observará que su espectro de movilidad iónica tiene varias bandas. Con moléculas pequeñas/iones con una sola carga, el número de bandas indicará usualmente el nº mínimo de compuestos coeluyentes; no así en el caso de péptidos y proteínas. En el caso de iones con múltiples cargas, aunque el pico corresponda a 1 sólo péptido, se observarán varias bandas de movilidad iónica correspondientes al distinto nº de cargas con que se ha ionizado (siguiendo una línea de tendencia); pero si coeluyen isoformas del péptido u otros péptidos, con movilidades distintas se observarán bandas adicionales. En cuanto a los distintos iones correspondientes al mismo perfil isotópico de un compuesto, éstos mostraran prácticamente la misma movilidad iónica.

Por otro lado, la velocidad de los espectros del TOF (microsegundos), permitirá obtener los espectros de masas de las diferentes bandas observadas en el espectro de movilidad iónica; o incluso por ejemplo, con la modalidad "MS/MS All Ions" obtener los espectros MS/MS de "todos" los péptidos. Todo ello proporciona herramientas adicionales para determinar con mayor exactitud, el número de compuestos coeluyentes en un pico cromatográfico.

5.4.1. Ejemplo de Mejora de la Capacidad Identificadora en un Mapa Peptídico mediante un Sistema de Movilidad Iónica acoplado con Cuadrupolo Tiempo de Vuelo (LC/IM-QTOF).

La ortogonalidad de la separación por movilidad, permitirá aumentar muy considerablemente el nº de péptidos/ proteínas identificados por un IM-QTOF, y superar así a equipos de mayor resolución. La figura 40 muestra los espectros de movilidad iónica y de masas, correspondientes a una franja del cromatograma de una rápida separación (tan sólo 15 minutos) de un proteoma completo, analizado mediante un instrumento LC/IM-QTOF (como el descrito en la figura 39). Esta misma muestra fue también analizada mediante un gradiente mucho más lento (100 minutos en lugar de 15), y con un espectrómetro de masas de mayor resolución, pero sin movilidad iónica (LTQ-FT-MS); éste sólo permitió identificar 1/3 de los péptidos identificados mediante una cromatografía 6 veces más rápida con el sistema LC/IM-QTOF de menor resolución espectral, pero con movilidad iónica.

En el estudio se utilizó un mapa peptídico del proteoma, digerido con tripsina, de plasma de sangre de ratón, a la que se adicionaron 20 péptidos de referencia. El gráfico de la figura 40, muestra la relación m/z en el eje vertical, y el tiempo de deriva (en ms) en el eje horizontal; los distintos colores que se pueden observar en la ampliación de la derecha, se corresponden con la intensidad de la señal (rojo mayor intensidad /amarillo-verde intensidad media / azul menor intensidad de señal). La zona de la derecha corresponde a la ampliación de una pequeña zona del gráfico de la izquierda, una franja espectral con péptidos de baja intensidad de señal; la franja de masas del espectro promedio de la derecha, es 90 veces menos intensa que el espectro de masas de la izquierda. En esta franja del espectro de la derecha, mediante el sistema con movilidad iónica LC/IM-QTOF y una cromatografía de 15 minutos, se identificaron los 10 péptidos remarcados con un circulo blanco en el gráfico de la derecha; el triple de los identificados con un sistema FTMS. Con el sistema LTQ-FT-MS (sin movilidad iónica) y una cromatografía de 100 minutos, sólo se consiguieron identificar 3 de los péptidos; los correspondientes a las m/z de los iones que aparecen en el espectro remarcados con un asterisco rojo.

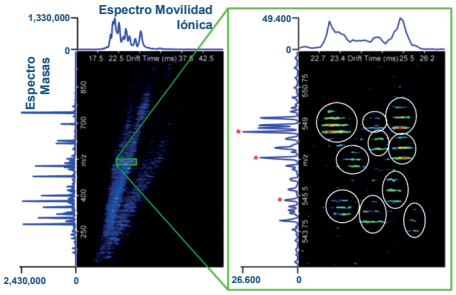


Figura 40: Espectros de movilidad iónica y de masas de un Mapa Peptídico muy complejo con un Espectrómetro IM-QTOF de Movilidad Iónica acoplado a un Cuadrupolo-Tiempo de Vuelo: Los espectros corresponden a una franja del cromatograma de una separación en 15 minutos de un proteoma completo (plasma de sangre de ratón adicionada con 20 péptidos), analizado mediante un instrumento LC/IM-QTOF. La ampliación de una pequeña zona espectral de baja intensidad, muestra la identificación de 10 de los péptidos (marcados con círculos) con el IM-QTOF; la misma muestra analizada con un LTQ-FT-MS de mayor resolución (pero sin movilidad iónica) y con un gradiente de 100 minutos, sólo permitió detectar 3 de los péptidos (aquellos con las m/z remarcadas con un asterisco en rojo).

Como se puede ver con los 10 péptidos remarcados en el gráfico de la derecha de la figura 40, el combinar movilidad iónica con masas, permite resolver péptidos que son difíciles de resolver al solaparse sus m/z. Ello incrementa considerablemente el nº de péptidos que se pueden resolver, incluso con una cromatografía mucho más rápida y un espectrómetro de menor resolución. Los QTOF proporcionan menor resolución espectral que los FTMS, pero los QTOF son más rápidos y permiten su acoplamiento con analizadores de movilidad iónica (sin sacrificar resolución); éste acoplamiento permitirá aumentar el nº de péptidos/ proteínas identificados con respecto al de equipos de mayor resolución.

5.4.2. Ejemplo de Análisis de Proteína Intacta (Ubiquitina) mediante un Sistema de Movilidad Iónica acoplado con Cuadrupolo Tiempo de Vuelo (IM-QTOF).

En la figura 41 se muestran los espectros de movilidad iónica y de masas, obtenidos por **infusión de una solución de la proteína ubiquitina** (sin digerir) en agua/metanol 50/50 con un 0.1% de ácido acético. Se observa como los distintos estados de carga de la ubiquitina, se separan también en la dimensión de movilidad (aunque la presencia de confórmeros la dificulta). En el gráfico central puede observarse la distribución de los distintos estados de carga en los dominios de m/z "versus" movilidad; éstos siguen una curva de tendencia característica.

En el gráfico puede observarse como algunos estados de carga presentan múltiples confórmeros (manchas verticales con distintos máximos de intensidad). El uso de un tubo de deriva con un campo eléctrico de baja intensidad, permite mantener intactas estructuras muy lábiles y facilita la detección de confórmeros. La separación por movilidad de estos confórmeros se aprecia muy bien, en el gráfico superior, cuando se amplía una zona centrada en una franja concreta de masa de extracción, por ejemplo alrededor de m/z 714.7 (iones 7+). Se observa cómo 2 confórmeros se resuelven casi hasta línea de base mediante movilidad, sin que por su espectro de masas se puedan llegar a detectar; al tratarse de la misma molécula, pero con una distinta distribución espacial de los grupos funcionales.

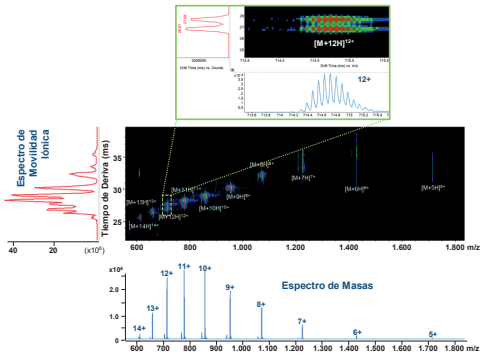


Figura 41: Espectros de movilidad iónica y de masas de una Proteína Intacta (Ubiquitina) con un Espectrómetro IM-QTOF de Movilidad Iónica acoplado a un Cuadrupolo-Tiempo de Vuelo: Los espectros corresponden a la infusión de una solución de la proteína Ubiquitina (sin digerir) en agua/metanol 50/50 con un 0.1% de ácido acético. Los distintos estados de carga de la ubiquitina se separan también en la dimensión de movilidad, y se observa su característica curva de distribución en los dominios de m/z "versus" movilidad. En el gráfico central se detecta como algunos estados de carga presentan múltiples confórmeros (manchas verticales con distintos máximos de movilidad).Al centrar/ampliar una franja de masa concreta, por ejemplo alrededor de m/z 714.7 (7+), se observa muy bien cómo 2 confórmeros se resuelven casi hasta línea de base mediante movilidad, sin que por su espectro de masas se pueda llegar ni siquiera a sospechar de su existencia.

REFERENCIAS.

1. Tswett M. Physico-chemical Studies of Chlorophyll. The Adsorption. Ber Dtsch Botan Ges. 1906; 24: 316–23.

2. Tswett M. Adsorption Analysis and the Chromatographic Method. Application to the Chemistry of Chlorophyll. Ber Dtsch Botan Ges. 1906; 24: 384–93.

3. Stephens WE. A Pulsed Mass Spectrometer with Time Dispersion. Phys Rev. 1946; 69: 691.

4. Wiley WC, McLaren IH. Time-of-Flight Spectrometer with Improved Resolution. Rev Sci Instr. 1955; 26: 1150

5. Wiley WC, inventor; Bendix Aviation Corporation, titular. Mass Spectrometer. US patent nº 2.839.687. 1958 June 17. Disponible en: http://www.freepatentsonline.com/2839687.pdf

6. Mamyrin B A, Shmikk D V, Zagulin V A, et al. The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. Sov Phys JETP. 1973; 37: 45

7. Glish GL, Goeringer DE. Tandem Quadrupole/Time-of-Flight Instrument for Mass Spectrometry/Mass Spectrometry. Anal Chem. 1984; 56: 2291-5. Disponible en: http://pubs.acs.org/doi/pdf/10.1021/ac00277a007

8. Karas M, Bachmann D, Hillenkamp F. Influence of the Wavelength in High Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules. Anal Chem. 1985; 57: 2935-9. Disponible en: http://pubs.acs.org/doi/pdf/10.1021/ac00291a042

9. Fenn JB, Mann M, Whitehouse CM, et al. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. Science. 1989; 246(4926): 64. Disponible en: http://www.sciencemag.org/content/246/4926/64.abstract

10. Dawson JH, Guilhaus M. Orthogonal-acceleration time-of-flight mass spectrometer. Rapid Commun. Mass Spectrom. 1989; 3:155. Disponible en: http://onlinelibrary.wiley.com/doi/10.1002/rcm.1290030511/pdf

11. Appfel JA, Werlich H, Berstch JL, inventors; Hewlett Packard, titular. Orthogonal Ion Sampling for Electrospray LC/MS. US patent nº 5.495.108. 1996 February 27. Disponible en: http://www.freepatentsonline.com/5495108.pdf

12. Lyndon M, Dixon B, Muddiman D, et al. Mass Measurement Accuracy Comparison of Double Focusing Magnetic Sector and Time-of-Flight Mass Analyzers. Poster presentado en conferencia ASMS-2007 y publicado posteriormente en Rapid Commun. Mass Spectrom. 2008; 22: 1563–1566. Disponible en: http://onlinelibrary.wiley.com/doi/10.1002/rcm.3544/pdf

13. Smith R, Kumar D, Lubda D, et al. Making Broad Proteome Protein Measurements in 1-5 min Using High-Speed RPLC Separations and High-Accuracy Mass Measurements. Anal. Chem. 2005; 77: 7763-73. Disponible en: http://pubs.acs.org/doi/pdf/10.1021/ac051257o

14. Fjeldsted J. Time-of-Flight Mass Spectrometry Technical Overview. Agilent Technologies Technical Overview. 2003; nº publicación 5989-0373EN. Disponible en: http://www.chem.agilent.com/Library/technicaloverviews/Public/5989-0373EN%2011-Dec-2003.pdf

15. Agilent. Ion optics innovations for increased sensitivity in hybrid MS systems. Agilent Technologies Technical Overview. 2007; nº publicación 5989-7408. Disponible en: http://www.chem.agilent.com/Library/technicaloverviews/Public/5989-7408EN_HI.pdf

16. Andrews GL, Simons BL, Muddiman DC, et al. Performance Characteristics of a New Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometer. Anal. Chem. 2011; 83: 5442–5446. Disponible en: http://pubs.acs.org/doi/abs/10.1021/ac200812d

17. Horn D, Miller C, Miller B. The Effect of MS/MS Fragment Ion Mass Accuracy on Peptide Identification in Shotgun Proteomics. Agilent Technologies Application Note. 2008; nº publicación 5989-7820EN. Disponible en: http://www.chem.agilent.com/Library/applications/5989-7820EN_LO.pdf

18. Zubarev R, Mann M. On the proper use of mass accuracy in proteomics. Mol Cell Proteomics. 2007; 6(3):377-81. Disponible en: http://www.mcponline.org/content/6/3/377.full

19. Moritz R. Development of a Complete Human Peptide and SRMAtlas. Webinar Simplifying Targeted Protein Quantification with Mass Spec: The SRM Atlas and Multiplexed MRM Protein Assays; 2010 Dic 16; Producido por Genetic Engineering & Biotechnology News; 2010. p. 4-26. Disponible en http://www.genengnews.com/webinars/simplifying-targeted-protein-quantification-with-mass-spec-the-srm-atlas-and-multiplexed-mrm-pr/130/

20. Harsha H., Pandey A, Waddell K, et al. 4-Plex iTRAQ Based Quantitative Proteomic Analysis Using an Agilent Accurate-Mass Q-TOF. Agilent Technologies Application Note. 2008; nº publicación 5990-5881EN. Disponible en: http://www.chem.agilent.com/Library/applications/5990-5881en_lo%20CMS.pdf

21. Schmidt A, Domon B, Aebersold R, et al. An Integrated, Directed Mass Spectrometric Approach for In-depth Characterization of Complex Peptide Mixtures. Mol. Cell Proteomics 7.11, 2138-2150. Disponible en: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2577211/

22. Yanes O, Clark J, Siuzdak G, et al. Metabolic oxidation regulates embryonic stem cell differentiation. Nat Chem Biol. 2010; 6(6): 411–417. Disponible en: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2873061/

23. Coral B, Cifuentes A, Godzien J, et al. Metabolomic Approach with LC-QTOF to Study the Effect of a Nutraceutical Treatment on Urine of Diabetic Rats. J Proteome Res. 2011;10(2):837-44. Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/21087057.

24. Anderson L, Pope M, Miller C, et al. Automation of a SISCAPA Magnetic Bead Workflow for Protein Biomarker Quantitation by Mass Spectrometry. Agilent Technologies Application Note. 2011; nº publicación 5990-7360EN. Disponible en: http://www.chem.agilent.com/Library/applications/5990-7360en_lo%20CMS.pdf

25. Andreson, N.L., T.W. Pearson, et al. (2004) Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). Proteomics. 2004; 3(2): 235-44. Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/15113099.

26. Makarov A. Theory and Practice of Orbitrap Mass Spectrometry. Abstract conferencia invitada IV Congreso de la Sociedad Española de Espectrometría de Masas; 2009 Sept.30-Oct.2; Castellón de la Plana, España. Disponible en: http://www.4rseem.uji.es/abstracts/I1_INVITADA_Makarov.pdf

27. Björhall K, Miliotis T, Davidsson P. Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. Proteomics. 2005;5:307–317. Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/15619298.

28. Dardé VM, Barderas MG, Vivanco F, et al. La depleción de las proteínas mayoritarias del plasma facilita su análisis proteómico en pacientes con síndrome coronario agudo. INVESTIGACIÓN CARDIOVASCULAR. 2006; VOL.9 Nº 1: 37-54. Disponible en: http://www.mapfre.com/ccm/content/documentos/fundacion/salud/revista-cardio/vol9-n1-art4-2D.pdf

29. Tang N, Waddell K, Miller C. Using a Complex Proteomics Standard for Comparison of Protein Identification Workflows. J Biomol Tech. 2010; 21(3 Suppl): S50. Disponible en: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2918202.

30. Hubner NC, Ren S, Mann M. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. Proteomics. 2008; 8(23-24):4862-72. Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/20946814.

31. Hörth P, Miller C, Wenz C. Combining immuno-depletion, protein and peptide OFFGEL electrophoresis for the efficient fractionation of plasma prior to LC-MS analysis. Agilent Technologies Application Note. 2007; nº publicación 5989-6419EN. Disponible en: http://www.chem.agilent.com/Library/applications/5989-6419EN.pdf

32. Lehmann W, Wolschin F, Weckwerth W. Neutral loss-based phosphopeptide recognition: a collection of caveats. J. Proteome Res. 2007, 6 (7): 2866–2873. Disponible en: http://pubs.acs.org/doi/abs/10.1021/pr060573w

33. Tang N. Accurate-Mass LC/TOF-MS for Molecular Weight Confirmation of Intact Proteins. Agilent Technologies Application Note. 2007; nº publicación 5989-7406EN. Disponible en: http://www.chem.agilent.com/Library/applications/5989-7406EN.pdf

34. Chen L, Gieschen A, Lau J, et al. Antibody Analysis by ESI-TOF LC/MS. Agilent Technologies Application Note. 2007; nº publicación 5989-5893EN. Disponible en http://www.chem.agilent.com/Library/applications/5989-5893EN.pdf

35. Jefferis R. Glycosylation as a strategy to improve antibody-based therapeutics, Nature Reviews Drug Discovery. 2009; 8, 226-234. Disponible en: http://www.nature.com/nrd/journal/v8/n3/abs/nrd2804.html

36. Reichert JM. Monoclonal Antibodies as Innovative Therapeutics, Current Pharmaceutical Biotechnology. 2008; 9, 423-430 Disponible en: http://www.benthamscience.com/cpb/openaccessarticles/cpb9-6/0002G.pdf

37. Nyberg KA. A New Kind of Cancer Warfare Agent. Next Generation Monoclonal Antibodies, Oncology business review. November 2007; 36-43 Disponible en: http://www.oncbiz.com/documents/OBR_1107_NGMA.pdf

38. Faster, More Accurate Characterization of Proteins and Peptides with MassHunter BioConfirm Software, Agilent Technologies Technical Note. 2010; nº publicación 5990-5096EN. Disponible en: http://www.chem.agilent.com/Library/technicaloverviews/Public/5990-5096en_lo%20CMS.pdf

39. Miller C, Tang N, Waddell K, et al. Primary Characterization of a Monoclonal Antibody Using Agilent HPLC-Chip Accurate-Mass LC/MS Technology. Agilent Technologies Application Note. 2008; nº publicación 5990-3445EN. Disponible en http://www.chem.agilent.com/Library/applications/5990-3445en_lo%20CMS.pdf

40. Trojer L, Gromadski K, Buckenmaier S. The Agilent mAb-Glyco Chip Kit for rapid and fully automated characterization of N-linked glycans from monoclonal antibodies. Agilent Technologies Technical Overview. 2011; nº publicación 5990-6924EN. Disponible en: http://www.chem.agilent.com/Library/technicaloverviews/Public/5990-6924EN.pdf

41. Ibrahim Y, Danielson W, Smith R, et al. Performance of a New Sensitive LC-IMS-QTOF Platform for Proteomics Measurements. Poster presentado en conferencia ASMS-2011. Disponible en: http://panomics.pnnl.gov/posters/ASMS2011/Ibrahim_PNNL_ASMS2011.pdf

42. Baker E, Danielson W, Smith R, et al. Applying a High Throughput IMS-QTOF MS Platform to Complex Samples for Increased Molecular Coverage. Poster presentado en conferencia ASMS-2013. Disponible en: http://panomics.pnnl.gov/posters/ASMS2013/Baker_PNNL_ASMS2013.pdf

43.Kurulugama R, Imatani K, Taylor L. The Agilent Ion Mobility Q-TOF Mass Spectrometer System. Agilent Technologies Technical Overview. 2013; nº publicación 5991- 3244EN. Disponible en: http://www.chem.agilent.com/Library/technicaloverviews/Public/5991-3244EN.pdf

FTMS Espectrometría de Masas por Transformada de Fourier

11

Michaela Scigelova[1], Bernabé Bodas[2] and Alexander Makarov[1]

[1] Thermo Fisher Scientific, Bremen, Germany;

[2] Consultor, Madrid, Spain.

1. Introducción.

Los espectrómetros de masas han venido siendo utilizados durante largo tiempo en una gran variedad de aplicaciones biológicas. En los últimos años, sin embargo, se ha observado un aumento significativo en el empleo de algunos espectrómetros de masas como de tiempo de vuelo (TOF), de Transformada de Fourier de resonancia ciclotrónica de iones (FTICR), y Orbitrap™, que proporcionan la masa exacta de analitos en un amplio rango de masas.

Los analizadores FTICR y Orbitrap, superan las prestaciones de cualquiera de los espectrómetros de masas comúnmente utilizados, con respecto a la máxima resolución y exactitud de masa alcanzable, de manera rutinaria, incluso para un número pequeño de iones. Ambos espectrómetros de masas comparten ciertas características, como el sistema de detección por corriente imagen y la aplicación de las operaciones matemáticas de la Transformada de Fourier, para la generación de los espectros de masas a partir de las señales transitorias temporales, producidas por la corriente imagen de los iones. Consecuentemente nos referimos a ellos como espectrómetros de masas basados en la Transformada de Fourier (FTMS).

La combinación de FTMS, con la selección previa de iones y dispositivos de fragmentación, y su acoplamiento a cromatografía de líquidos (LC) de fase reversa (RF) representa hoy en día un método altamente eficaz tanto para el análisis de pequeñas moléculas como para Proteómica. Las separaciones LC multidimensionales tienen un importante papel que jugar en aplicaciones proteómicas para reducir la complejidad de la muestra y como buen complemento para el gran rango dinámico de adquisición que ofrecen los instrumentos FTMS.

El presente documento no intenta ser un exhaustivo repaso de las publicaciones sobre la técnica FTMS; intenta ser un material de enseñanza para científicos sin conocimientos profundos de Física, con la finalidad de ayudarles a entender mejor las herramientas de espectrometría de masas, que están llamados a utilizar en su trabajo diario en el laboratorio. Por tanto, sólo los aspectos fundamentales de FTMS, que son de utilidad para este fin, serán presentados. La discusión se centrará en ejemplos, en los que esta tecnología se puede aplicar, y cuáles son las ventajas y los beneficios, así como sus limitaciones, para las aplicaciones prácticas de la Proteómica.

2. La necesidad de la alta resolución y exactitud de masa.

La *exactitud de masa* es el error en la relación entre la *m/z* medida y la *m/z* real, generalmente expresado en ppm. El *poder de resolución de masa* (resolución) es la medida de la capacidad de distinguir entre dos picos ligeramente diferentes en su m/z, entendiendo aquí Dm/z la medida realizada a mitad altura (FWHM) del ión monoisotópico. La linealidad de la detección y la muy alta exactitud en la determinación de la frecuencia son inherentes a la técnica FTMS y permiten alcanzar, un alto poder de resolución, exactitud de masa y rango dinámico. Pero ¿Por qué necesitamos una alta exactitud de masa y una alta resolución?

La ventaja de medir la masa de un compuesto con una adecuada alta exactitud nos puede permitir directamente determinar su composición elemental. La masa exacta por lo tanto actúa como un poderoso "filtro" de utilidad para confirmar la identidad de un compuesto o incluso lleva a la identificación de un desconocido. Esto se ilustra en la Figura 1 mostrando el espectro de fragmentación (MS/MS) del flavonoide quercetina (*m/z* 303). La desviación de masa es de menos de 3 ppm para cualquier fragmento detectado, que, junto con la riqueza del espectro de fragmentación obtenido, nos permiten no solo confirmar la composición elemental del compuesto de partida, sino que nos proporcionan además información útil sobre su estructura.

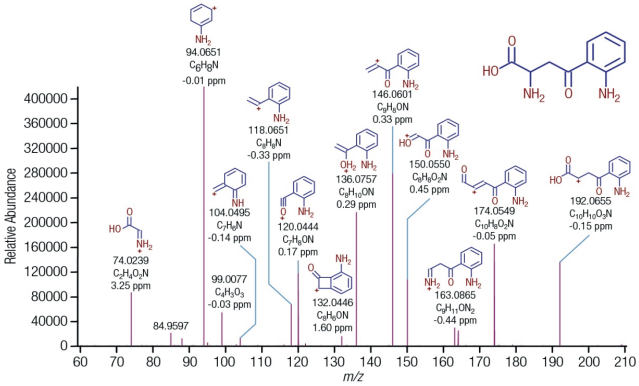


Figura 1. Espectro de fragmentación (MS/MS) de quinurenina (m/z 209). Desviación de masa menor de 2 ppm para cualquier fragmento detectado, que junto con su abundante fragmentación espectral permite confirmar la composición elemental del compuesto de partida, proporcionando a la vez una útil información de su estructura. La asignación de los picos se ha llevado a cabo utilizando el paquete de software Mass Frontier™.

En muchos casos, sin embargo, información adicional además de la masa exacta puede ser necesaria para obtener la adecuada composición elemental, entre otras, la restricción del número de elementos considerados, las leyes químicas de Lewis y Senior o el patrón isotópico.[1,2,3] Otros aspectos de la utilidad de la medida de masa con exactitud, para el análisis de pequeñas moléculas, se incluyen en la aplicación de filtro de masas por defecto[4]. En el contexto de la Proteómica las masas de los precursores se utilizan para restringir las búsquedas en las bases de datos. La Figura 2 ilustra esto en el ejemplo de base de datos de proteínas humanas. Está claro que con una mejor exactitud de masa (menor desviación de masa) el número de posibles candidatos cae significativamente. La exactitud de masa actúa como filtro, lo que se refleja en una posterior evaluación estadística de los resultados con una reducción significativa del número de potenciales falsos positivos, dando como resultado una mayor confianza en la identificación.

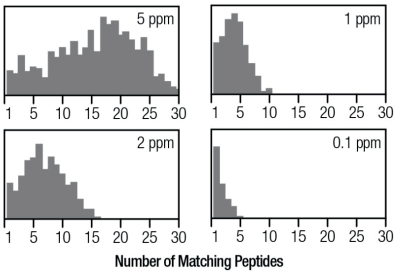


Figura 2. Efecto de la exactitud de masa en la identificación de péptidos, un ejemplo de péptidos tríptico de la base de datos de proteína humana. Histogramas para ventanas de masas de 5, 2, 1 y 0,1 ppm respectivamente, donde el número de péptidos apareados en cada ventana está en el eje X y el de la frecuencia de apariciones en el eje Y. Cortesía de David Fenyo, Rockefeller University.

La otra observación, que se puede hacer del gráfico, es, que incluso con una desviación de masa muy baja de 0,1 o 0,01 ppm (en la actualidad no se pueden realizar en rutina en cualquier espectrómetro de masas) todavía se encontrarán muchos casos en la base de datos humana, donde más de un péptido encaja en la masa seleccionada con su respectiva desviación. Por tanto es necesaria otra información para confirmar la identificación de un péptido además de la masa medida.

La exactitud de masa es, por tanto, un parámetro muy importante en los experimentos de Proteómica. Una incorrecta determinación de la masa puede afectar a las estadísticas de identificación, que serán peores con una ventana de exactitud de masa demasiado grande, o resultaría en la pérdida de identificaciones (falsos negativos) como si la ventana es demasiado pequeña.[5]

Es quizás menos obvio, cual es el papel de la resolución y como la resolución y la exactitud de masa están estrechamente relacionadas. La alta resolución es necesaria, porque no se puede realizar la medida de masa y la cuantificación de un pico con exactitud, si los elementos constitutivos del mismo no están suficientemente resueltos. En este contexto, la elección de la resolución utilizada en el análisis define un nivel de confianza (tolerancia de masa), en el que podemos confiar en nuestra medida de masas. Considerando el ejemplo de un pesticida Pirimicarb (*m/z* 239), el análisis realizado a una resolución de 15.000 medida a media altura (full width at half maximum; FWHM) da una medida de masa de +/- 9 mDa, mientras que con una resolución de 80.000 este intervalo se reduce a +/-1.7 mDa.

Lo que en realidad ocurre es, que en una mezcla compleja muchos componentes de la matriz, así como algunos analitos, pueden coeluir con el compuesto de interés y alguno de ellos puede tener una masa muy similar. La superposición de estos picos de masa muy similar, que coeluyen, modifica la forma del pico del compuesto de interés. Entonces, el pico centroide no se corresponde con su masa exacta. Esto se comprueba con el ejemplo del Pirimicarb, mezclado con otros 115 pesticidas y toxinas dentro de la matriz de pienso para caballos (Figura 3) El panel superior muestra un espectro de masas tomadas en el momento de la elución del Pirimicarb adquirido a una resolución de 15.000. En este caso en particular, las interferencias sin resolver están produciendo una medida de masa del Pirimicarb desviada hacía un valor ligeramente superior con una diferencia de masa de 6,5 ppm. Adquiriendo los datos a más alta resolución se consigue una enorme mejora en la exactitud de la medida de masa (0,3 ppm de desviación de masa).

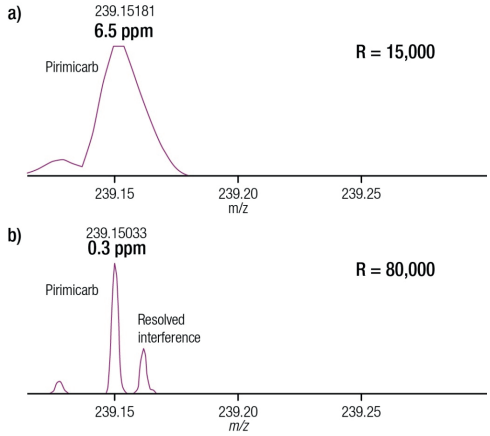


Figura 3. Resultados en alta resolución con mejor exactitud de masa. El pesticida pirimicarb fue medido en una mezcla con otros 115 pesticidas y toxinas de alimentos, en una matriz de pienso para caballos. A, El espectro de masas tomado en el tiempo de la elución del pirimicarb, alcanzó un poder de resolución de 15.000 FWHM. El pico correspondiente al pirimicarb mostró una desviación de masa de 6,5 ppm. B, Conmutando la resolución a 80.000 el pirimicarb se detecta con una desviación de masa de 0,3 ppm. Al mismo tiempo se reveló la presencia de una interferencia, que había sido la responsable de esta relativamente gran desviación de masa observada a más baja resolución. Cortesía de Markus Kellman, Thermo Fisher Scientific.

El efecto de resolución descrito anteriormente, limitando la exactitud de masa alcanzable, tiene consecuencias muy importantes para experimentos de detección rápida y cuantificación. Si, por ejemplo, la ventana de extracción seleccionada es demasiado amplia, para compensar posibles interferencias de la matriz y/o inadecuado poder de resolución, entonces; 1) la exactitud de masa se verá comprometida y 2) las interferencias ocultas contribuirán a la integración del área del pico detectado del compuesto de interés. Existe efectivamente un riesgo serio de

tener un caso de *sobre cuantificación* o incluso un *falso positivo*. Si, por el contario, el usuario establece una ventana de extracción demasiado estrecha, el compuesto con una desviación de masa más alta de la esperada, debido a la presencia de interferencias no resueltas, podría pasar completamente desapercibido, dando lugar al riesgo de dar un *falso negativo*.

En resumen, para una determinada muestra analizada el poder de resolución es un parámetro clave, que afecta a la correcta asignación de la masa de los analitos. La medición de masa exacta solo se puede conseguir, cuando se realiza la medida con una resolución suficientemente alta. También hay importantes consecuencias para screening y cuantificación. En el análisis de mezclas complejas, la resolución afecta a la fiabilidad de la detección de un compuesto y a su cuantificación. En estudios encaminados al descubrimiento de biomarcadores, una mejor exactitud de masa, se traduce en una mejor alineación y cuantificación.[6]

3. Espectrómetros de Masas de Transformada de Fourier.

3.1. Transformada de Fourier

La transformada de Fourier (Fourier Transform; FT) hunde sus raíces en las series de Fourier donde unas complicadas funciones periódicas se pueden escribir como suma de ondas simples, representadas matemáticamente por senos y cosenos. Debido a las propiedades de las funciones seno y coseno, resulta posible recuperar la contribución individual de cada onda en la suma mediante una integral.

En FTMS, las masas están representadas por frecuencias y dado que la frecuencia se puede medir con mucha exactitud, FTMS puede ofrecer una exactitud en la medida de masas potencialmente muy alta. A fin de ilustrar la complejidad de la señal, que uno encuentra en FTMS, vamos a considerar a modo de ejemplo cuatro frecuencias, que representan cuatro diferentes masas: frecuencia "v" de una intensidad 1; frecuencia "2v" de una intensidad 0,5; frecuencia "5v" de intensidad 1,5; y frecuencia "8v" de una intensidad 0,2. Sus respectivas formas de onda se muestran todas juntas en la Figura 4b. Un detector FTMS verá todas estas formas de onda combinadas en una señal similar a la mostrada en la Figura 4b.

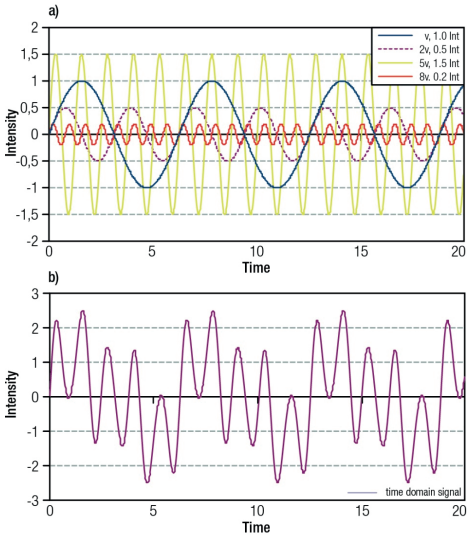


Figura 4. Ilustración de la complejidad de la señal, que encontramos en FTMS. A, La forma de onda para cuatro frecuencias individuales, representando cuatro diferentes masas con sus intensidades correspondientes: frecuencia "v" de una intensidad 1; frecuencia "2v" de una intensidad 0.5; frecuencia "5v" de una intensidad 1.5; y frecuencia "8v" de una intensidad 0.2. B, Las cuatro formas de onda están combinadas en una señal única emulando la señal detectada en el dominio temporal por un FTMS. Cortesía de Anastassios Giannakopulos, Thermo Fisher Scientific.

Sin embargo, el análisis de muestras reales es un asunto mucho más complicado. No es raro en un experimento de Proteómica tener cientos, incluso miles, de especies moleculares presentes en un único espectro de masas. Además cada especie molecular tiene también sus isotopómeros. La complejidad de la señal resultante detectada es por tanto considerable.

3.2. Señal de dominio temporal

En los espectrómetros de masas, que utilizan la FT, los iones son observados (en los platos de detección se induce una corriente de imagen, esta es digitalizada y almacenada en un archivo) durante un periodo de tiempo, que podía extenderse potencialmente hasta algunos segundos. Esta señal de dominio temporal, que contiene todas las frecuencias características de los iones medidos con las intensidades, que se corresponden con la cantidad presente de las especies moleculares ionizadas de la muestra, se procesa con la herramienta matemática de Transformada de Fourier. Esta operación descompone el espectro de frecuencias obtenidas en un complejo (en el sentido matemático, es decir, que contiene una parte real y una parte imaginaria) espectro de masas. La Figura 5 muestra la parte real (conocida como "absorption spectrum") y la parte imaginaria de la solución con sus posiciones mutuales y altura dependientes de fase. Para eliminar esta dependencia de fase se utiliza una raíz cuadrada de la suma de los cuadrados de las partes real e imaginaria ("magnitude spectrum"). El problema con esta simplificación es, que la resolución del "magnitude spectrum" es dos veces menor, de la que podría haber tenido, si la fase pudiera de alguna manera ser conocida. Se han realizado intentos para controlar la fase del movimiento de los iones y este tema se desarrolla más adelante en el apartado 8.

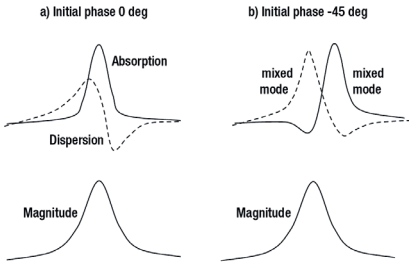


Figura 5. Dependencia de fase de las partes real e imaginaria de la solución transformada de Fourier. A, En la fase del momento inicial igual a cero, las partes real e imaginaria del espectro complejo de frecuencias representa los modos espectrales de absorción pura y dispersión, respectivamente. B, Para cualquier otra fase inicial de dominio temporal los componentes complejos representan una combinación lineal de los modos de absorción y dispersión, y los picos resultantes son asimétricos. Empleando un "Espectro de magnitud" se elimina la dependencia de la fase inicial de dominio temporal con la penalización de la obtención de un espectro con sólo la mitad de resolución de la teóricamente disponible para estos datos en concreto. Cortesía de Robert Malek, Thermo Fisher Scientific.

3.3. Procesamiento de señal avanzado

En la mayoría de aplicaciones de Proteómica, donde el rango dinámico dentro del espectro es de gran importancia, es necesario llevar a cabo un posterior procesado de la señal llamado "apodization" y "zero filling".

La apodization es una transformación de transitorios para suavizar discontinuidades al principio y al final del pico. Se utiliza para eliminar los picos artefactos, adyacentes al pico real, en la base del espectro (Figura 6).

Mediante la eliminación de los pequeños picos de artefactos, el espectro apodizado permite un mayor rango dinámico. Esto tiene un precio: la resolución de un espectro apodizado es aproximadamente la mitad de la resolución del no apodizado. Hay que tener en cuenta, que las prestaciones de los FTICR citados en la literatura se dan a menudo en forma no apodizada, mientras que en el caso del Orbitrap siempre se refiere a la forma apodizada.

"Zero filling" es el tratamiento del espectro, que corresponde a la adición de un intervalo relleno con ceros al transitorio. Esto permite una fácil interpolación entre puntos en el dominio frecuencia y da por tanto una mejor forma del pico.

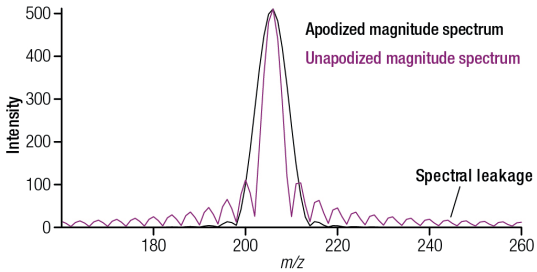


Figura 6. Ejemplo de la simulación de picos espectrales apodizados y no apodizados. Mientras que la apodización mejora el rango dinámico y la apariencia del espectro FTMS, repercute negativamente en su resolución. Cortesía de Robert Malek, Thermo Fisher Scientific.

4. Resonancia ciclotrónica de iones por transformada de Fourier FTICR.

4.1. Principio de operación

Una partícula cargada moviéndose en un campo magnético con un vector de velocidad perpendicular al campo magnético, experimenta una fuerza perpendicular al plano definido por la velocidad y el campo magnético. Esta fuerza, conocida como "Fuerza de Lorentz", provoca un cambio en la dirección del vector de velocidad pero no en su magnitud (Figura 7). Esta fuerza es siempre perpendicular a la dirección del vector de velocidad. Siguiendo el vector de velocidad vemos, que describe un círculo, y como resultado la partícula cargada (un ión) es atrapada por el campo magnético en una trayectoria circular. La frecuencia de esta rotación es característica para cada masa y el campo magnético. En un campo magnético uniforme no depende de las coordenadas y las velocidades iniciales de los iones.

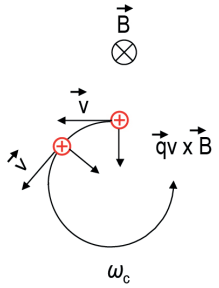


Figura 7. Movimiento del ión dentro de la cámara ICR. La fuerza, que afecta a una partícula cargada en movimiento en un campo magnético homogéneo, hace asumir a la partícula un movimiento circular. Cortesía de Anastassios Giannakopulos, Thermo Fisher Scientific.

Instrumentos, donde los iones son excitados a seguir la trayectoria más larga en presencia de un campo magnético, son los llamados espectrómetros de masas de resonancia ciclotrónica de iones (ICR). La teoría de resonancia ciclotrónica fue desarrollada por Lawrence en 1930, que fue galardonado con el premio Nobel en 1939. Siguieron otros diseños instrumentales, que se utilizaron principalmente para estudios de reacciones ión-molécula.[7] La Transformada de Fourier, introducida por Comisarow y Marshall en 1974, ofrece la ventaja crucial de la detección de todas las frecuencias de forma simultánea, en lugar de una sola frecuencia cada vez.[8]

4.2. Movimiento de los iones dentro de la cámara ICR

La frecuencia angular ciclotrónica en un FTICR, que consta de dos electrodos paralelos continuos y tiene un solo campo magnético, se describe con la siguiente fórmula:

$$\omega_c = \frac{qB}{m} \quad \text{(en radian/s)} \qquad (1)$$

Y como una revolución es igual a 2π radianes:

$$\omega = 2\pi\nu \quad (2)$$

Donde "ν" es la frecuencia de las revoluciones medida en Hz, "ω_c" es la frecuencia angular en radian/s, "q" es el estado de carga que tiene el ión, "m" la masa y "B" la fuerza del campo mag-

nético. Esta fórmula muestra que hay una relación inversamente proporcional entre m/q (relación masa/carga; *m/z*) y la frecuencia orbital observada.

El campo magnético puede confinar solo los iones en la dirección radial y no existe confinamiento en la dirección axial a lo largo del campo magnético. Los campos eléctricos estáticos, con un potencial del orden de unos pocos voltios, se utilizan para confinar los iones en la dirección axial. Estos iones comienzan a ejecutar una oscilación armónica de confinamiento entre los dos electrodos a su frecuencia de confinamiento particular. Las interferencias del campo de confinamiento con el movimiento del ciclotrón de los iones produce desviaciones de la imperturbable marcha del ciclotrón y como resultado reduce ligeramente la frecuencia ciclotrónica. Como la desviación de la frecuencia imperturbable del ciclotrón depende en gran medida del potencial de captura, esta se realiza a los niveles de voltaje más bajos que sean posibles. Finalmente hay también una baja frecuencia de movimiento de exactitud que se conoce como "magnetrón".

Las frecuencias del magnetrón y de confinamiento son habitualmente mucho más bajas, que la frecuencia del ciclotrón. Generalmente no son detectadas, a menos que haya un pequeño desajuste en el eje del campo magnético del instrumento o la amplitud del movimiento de los iones sea tan grande, que llegue a alcanzar las dimensiones de la cámara ICR. Entonces, esas frecuencias se mostrarían como bandas laterales pequeñas de los picos del espectro. Mas detalles sobre principios de FTICR en forma muy didáctica se pueden encontrar en esa referencia.[9]

En conclusión, los iones deben entrar en la cámara ICR a bajas energías cinéticas de modo, que puedan ser confinados con relativamente bajos potenciales de confinamiento. Los iones deben estar situados preferiblemente en el centro de la cámara, para aprovechar al máximo las condiciones de excitación y las dimensiones de la cámara.

4.3. Excitación de los iones en la cámara ICR

Debido a la energía cinética del movimiento térmico, los iones tendrán algún pequeño movimiento ciclotrónico incluso a temperatura ambiente y sin ningún tipo de excitación. Estas condiciones, sin embargo, no generan una imagen detectable en las placas de detección. Los iones necesitan ser coherentemente excitados al radio más grande posible para detectar mejor su señal. Los iones se mantendrán en este radio de giro, después de que la señal de excitación se haya detenido y hasta que su energía se disipe por las colisiones con gas residual, lo que finalmente limita la duración del tiempo de detección. Cuanto mayor sea el tamaño de la molécula y mayor la presión del gas residual dentro del la cámara de ICR, menor será el tiempo disponible para la detección.

Todas las masas deben recibir una excitación similar a fin de evitar la detección dependiente de la masa. Generalmente, se aplica un potencial de RF de excitación desde dos placas situadas en oposición en la cámara de ICR. En FTICR, en lugar de excitar y detectar una masa (frecuencia) cada vez, todo el rango de *m/z* es excitado al mismo tiempo, aplicando generalmente una de las siguientes técnicas (Figura 8):

- *Frequency sweep (chirp) excitation*escanea rápidamente en el rango de frecuencias, que corresponden a las masas que precisan ser excitadas .[10,11] Se puede excitar un amplio rango de *m/z* y se puede lograr una respuesta razonablemente estable. Sin embargo se producen deformidades en ambos extremos del rango de excitación.

- Debido a que la Transformada de Fourier también funciona en sentido inverso, otra forma de producir una forma de onda de excitación es considerar la masa o el rango de masas, que necesitamos excitar, y llevar a cabo una Transformada de Fourier inversa desde el dominio frecuencia al dominio tiempo. La transformación correspondiente se puede almacenar y reproducir para la excitación. Esto se conoce como *stored wave-*

form inverse Fourier transform (SWIFT).[12] Este método ofrece una buena respuesta de excitación.

- Además, las masas no deseadas pueden ser completamente expulsadas de la cámara durante el proceso SWIFT. La forma de onda SWIFT se puede utilizar para el aislamiento de iones antes del análisis MS/MS en la cámara ICR, como se demuestra para el único isotopómero de la ubiquitina (8,6 kDa) y de la anhidrasa carbónica (29 kDa).[13] Otro ejemplo es un análisis de histona H4 de bovina intacta, donde se logró una selectividad de aislamiento de 0,1 *m/z* para el estado de carga 18^+.[14]

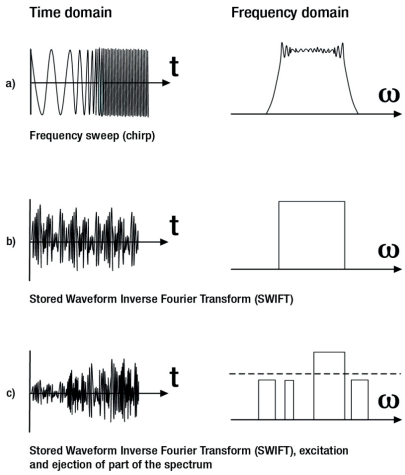


Figura 8. Técnicas aplicadas en FTICR para excitar a un completo rango de masas simultáneamente. A, Un barrido de frecuencias explora el rango de frecuencias correspondiente a las masas, que deben ser excitadas. B, Una forma de onda por transformada inversa de Fourier (SWIFT) genera y registra la señal en dominio tiempo correspondiente a las masas, que han de ser excitadas. Como repitiendo señal grabada previamente, lleva a los iones al estado de excitación. C, El método SWIFT también se puede emplear para extraer por completo las masas no deseadas de la cámara ICR. Cortesía de Anastassios Giannakopulos, Thermo Fisher Scientific.

4.4. Calibración de masa

Como se mencionó anteriormente, los potenciales de confinamiento cambian la simple relación entre la masa y frecuencia en ICR (formula 2) a una forma más complicada:[15]

$$\frac{m}{z} = \frac{A}{\nu_+} + \frac{B}{\nu_+^2} \quad (3)$$

donde "A" y "B" son constantes, que se pueden obtener mediante el ajuste de un conjunto de frecuencias experimentales a valores *m/z.* Añadiendo más parámetros en la ecuación no mejora significativamente el resultado.[16]

La condición limitante, para que la ecuación anterior sea válida, es, que la carga espacial debería ser insignificante, por lo que un número muy pequeño de iones deberían estar presentes. Sin embargo las perturbaciones de campo afectan tanto a los calibrantes como a los iones, y esta ecuación se puede extender también a un gran número de iones siempre y cuando su número en la cámara FTICR se mantenga igual. Por esta razón, los espectrómetros de masas FTICR que tienen control automático del número de iones, que se inyectan en la cámara de ICR, tienen una clara ventaja.[17] Esto es especialmente importante en trabajos de Proteómica, donde el sistema FTICR es a menudo acoplado a un sistema LC y el número de iones puede fluctuar dramáticamente de espectro a espectro.

4.5. Problemas con el potencial de campo

Tanto los campos electrostáticos (es decir el voltaje de confinamiento) como los campos de RF utilizados para la excitación pueden crear problemas para el análisis de la masa. Se necesitan rigurosos requisitos de uniformidad para ambos. Se han propuesto varias soluciones para combatir con el campo de confinamiento axial.[18,19] A finales de los 80, Wang y Marshall utilizaron una malla conectada a tierra, justo frente a los electrodos de confinamiento axial.[20] En este caso, el campo de confinamiento penetrando a través de la red puede tener efecto, solo cuando los iones están muy cerca de los electrodos de confinamiento.

La sencilla idea de utilizar una rejilla en la cámara ICR de los terminales abiertos funciona también para solucionar el problema con la homogeneidad del campo de excitación. Las rejillas se colocan dentro de los electrodos cilíndricos en toda la longitud de la cámara ICR. Las formas de onda de excitación se aplican a estas rejillas, para que el campo de excitación se extienda bien más allá de la región de confinamiento (Figura 9).

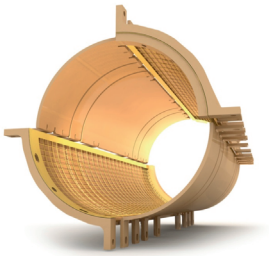


Figura 9. Celda de rejilla utilizada para minimizar las perturbaciones del campo de excitación. La presente gráfica de la cámara de ICR del equipo LTQ FT representa las redes colocadas dentro de los electrodos cilíndricos en toda la longitud de la cámara ICR.

La introducción de la cámara con rejilla ofrece una solución interesante también desde la perspectiva de la exactitud de masa alcanzable en el instrumento. Volviendo a la fórmula de la calibración de masa (fórmula 3 en la sección 4.4) y teniendo en cuenta, que el parámetro "B" de la misma se refiere a los errores de campo eléctrico, se puede observar, que el campo eléctrico produce un error de ΔB de la calibración dando lugar a un error en la asignación de la masa Dm. Para mejorar la exactitud de masa, o se reduce ΔB o se incrementa la frecuencia ciclotrónica mediante un campo magnético mayor. La cámara de rejilla reduce el ΔB y, consecuentemente, el Dm aproximadamente cuatro veces. Si consideramos, que el mismo efecto de Dm podría lograrse mediante el cambio del imán de 7 T por uno de 14 T (incrementando la frecuencia ciclotrónica dos veces), el diseño empleando la cámara con rejilla es una elegante y eficaz solución. Hay, sin embargo también otros métodos para conseguir la homogeneidad del campo de excitación.[21 22 23]

4.6. Efectos del campo magnético en la resolución (imanes de alto campo)

La resolución a baja presión viene dada por[24, 25]

$$\frac{m}{\Delta m} = \frac{1.274 \times 10^{7} z B_{o} T_{aqn}}{m} \quad (4)$$

Donde *m/z* viene dada en unidades por carga elemental. La Figura 10 muestra la resolución alcanzable durante un segundo transitorio para intensidades de campo magnético de 1 a 40 T y para diferentes masas. Los valores de este gráfico son para formas no apodizadas, lo que significa, que la resolución apodizada debería ser aproximadamente la mitad (en la mayoría de los casos no es posible obtener el valor exacto para una resolución apodizada, ya que cada fabricante utiliza un algoritmo de apodización diferente). Una detallada discusión de las características y ventajas de los imanes de alto campo se puede encontrar en Marshall y col.[26] Aquí vamos a resumir algunos de los puntos principales:

- Poder de resolución y exactitud de masa
  La fórmula 4 superior muestra claramente, que el poder de resolución (m/Δm) se incrementará linealmente con el incremento del campo magnético (Figura 10). Las mayores precisiones de masas se pueden conseguir consecuentemente con el aumento del poder de resolución.

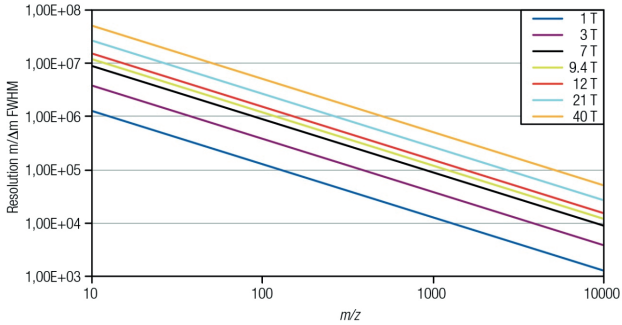


Figura 10. Efecto de la intensidad del campo magnético en la resolución. La resolución alcanzable (FWHM) para un rango de masas dado y para una señal transitoria de 1 segundo, empleando intensidades de campo magnético de 1 a 40 T queda aquí representada. La resolución FTICR representada no es apodizada. Téngase en cuenta que ambos ejes son logarítmicos. Cortesía de Anastassios Giannakopulos, Thermo Fisher Scientific.

- Velocidad de adquisición de datos

La misma fórmula también muestra, que el tiempo necesario para adquirir una señal de dominio temporal para un poder de resolución dado varia con 1/B. Esto puede representar una considerable ganancia de tiempo. En algunos diseños instrumentales esta ventaja no aportará diferencias prácticas, ya que algunas técnicas de fragmentación realizadas en la cámara ICR requieren la introducción de un gas reactivo (para fragmentación o axialización). Esto tiene como resultado una pérdida adicional de tiempo para alcanzar después los niveles de vacío necesarios para la excitación de los iones y su detección.

- Energía cinética

Una mayor energía cinética de los iones con el radio de excitación dado se pueda conseguir con un mayor campo magnético. Altas energías cinéticas pueden ayudar para la disociación por colisión inducida (collision induced dissociation; CID). Por ejemplo, en un imán de 3 T utilizando argón como gas de colisión, un ión de 1.000 Da con un radio de excitación de 1 cm tiene energía cinética de su centro de masa de 1,67 eV, mientras que en un imán de 9,4 T alcanzaría 16,4 eV. Esto producirá una fragmentación más eficiente. Pero como mencionamos anteriormente, para conseguir CID en la cámara ICR, hay que introducir el gas, y la presión en el interior de la cámara aumentará. Para poder continuar después con el análisis se necesita tiempo suficiente para evacuar el gas y alcanzar la baja presión apropiada para la excitación y detección de los iones. Es mucho más práctico realizar CID fuera de la región de ultra alto vacío, en un analizador externo o en una celda de colisión.

- Duración del tiempo de confinamiento

Debido a las colisiones con gas residual en la cámara ICR, los iones pierden energía cinética. Esto significa que el radio ciclotrónico decrece rápidamente con el tiempo y simultáneamente el radio magnetrón se incremente lentamente. El movimiento de los iones se hace más inestable cuando el radio magnetrón alcanza un valor significativamente comparable al radio de la cámara. Cuanto más tiempo tarde en pasar esto, mayor dominio temporal de transitorio se puede lograr y, consecuentemente, mayor resolución se puede alcanzar. El empleo de imanes de mayor intensidad de campo magnético tiene la ventaja, que el tiempo requerido, para que el radio magnetrón del ión alcance el radio de la cámara ICR, se incremente en una dependencia cuadrática de B.

- Fenómenos relacionados con la carga

En una cámara ICR, la existencia de carga concentrada puede ser responsable de la deriva y el ensanchamiento de los picos en el espectro, a veces hasta producir la fusión, en un solo pico, de los señales de dos masas similares. Ese artefacto se conoce como “peak coalescense”. La tendencia para “peak coalescense” disminuye con el aumento de la fuerza del campo magnético. En la práctica esto (y no solo un mero incremento de la resolución y exactitud de masa, que pueden conseguirse por otros medios como, por ejemplo, un mejor diseño de la cámara ICR; mencionado en la sección 4.5) tiende

a ser una de las razones principales para considerar un imán más grande teniendo en mente todos los aspectos negativos, que conlleva esto (por ejemplo, el considerable incremento de los costes y la dificultad de manejo).

Una manifestación diferente del efecto relacionado con la carga puede observarse a bajos niveles de número de iones (<1.000). Los iones se 'evaporan' desde el paquete de iones y forman una "cola de cometa".[27] Después de algún tiempo, dependiendo de la particular geometría de la cámara ICR y de la fuerza del campo magnético, los iones forman un anillo, sin capacidad de inducir corriente de imagen en las placas de detección. Un límite de resolución se alcanza por tanto, debido a una menor duración de la corriente transitoria. La velocidad de extensión de la "cola del cometa" se reducirá a altos campos magnéticos, debido a la disminución de la frecuencia del magnetrón, que es responsable de este efecto.

4.7. Efecto de duración de la corriente transitoria en la resolución

Otro camino para incrementar la resolución, sin utilizar unos grandes imanes, es conseguir la adquisición de corrientes transitorias durante un largo periodo de tiempo. La Figura 11 compara la resolución, que se obtiene con un imán de 7 y 21 T para duraciones de corrientes transitorias de 1 y 5 segundos. Naturalmente, hay un límite porque:

- Largos tiempos de adquisición no resultan siempre prácticos, particularmente en el caso de acoplamientos con LC.

- Hay solo una limitada cantidad de tiempo, en que la corriente transitoria puede ser registrada, antes de que desaparezca debido a las colisiones con el gas residual. Si bien todos los espectrómetros de masas necesitan vacío para el análisis y detección de los iones, los FTMS son más sensibles a la presión que cualquiera de los otros analizadores. Se necesita un vacío de 10^{-9} a 10^{-10} Torr (1 Torr = 133.3 Pa) para conseguir alta resolución, y esta necesidad de alto vacío se hace más estricta para los imanes de más alto campo.

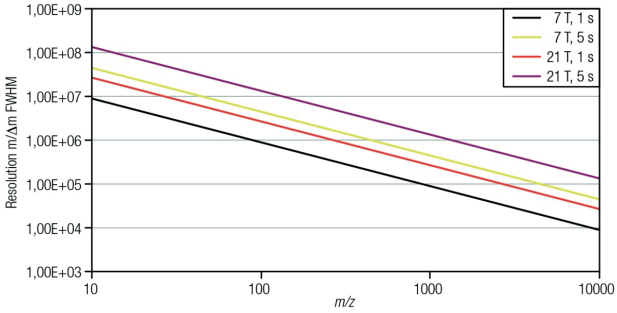


Figura 11. Efecto de la duración de la señal transitoria en la resolución. Se compara la resolución alcanzable (FWHM) para un rango de masas dado con imanes de 7 a 21 T y para una duración de la señal transitoria de 1 a 5 segundos. La resolución mostrada para FTICR no es apodizada, para el analizador Orbitrap es apodizada. Téngase en cuenta, que ambos ejes son logarítmicos. Cortesía de Anastassios Giannakopulos, Thermo Fisher Scientific.

4.8. Fragmentación en la cámara de ICR

La cámara ICR puede confinar diferentes masas, excitarlas hasta alcanzar su mayor energía cinética, llevar a cabo experimentos de fragmentación y reanalizar los productos de fragmentación. Alternativamente, se puede permitir a los iones interaccionar a bajas energías cinéticas, preferiblemente térmicas, con electrones o fotones. Esto hace de la cámara ICR un lugar ideal de reacción, que con una pequeña suma de tiempo añadido, puede ser un espectrómetro de masas en tándem por derecho propio. En este punto tenemos que hacer mención de tres técnicas de fragmentación principales utilizadas en cámaras de ICR y discutir las ventajas y desventajas de cada una de ellas.

Disociación inducida por colisión

Una técnica tradicionalmente utilizada en FTICR es la disociación inducida por colisión, CID. Los iones tienen que entrar en la cámara ICR a una baja energía cinética para poder ser confinados con efectividad por los potenciales de confinamiento a los niveles más bajos posibles. Estas condiciones, sin embargo, no proporcionan ninguna fragmentación. Los iones necesitan ser excitados a un amplio radio para ganar energía cinética. Esta excitación se puede conseguir de forma resonante para cada masa utilizando una de las siguientes técnicas: Excitación por activación de múltiples colisiones (multiple excitation collisional activation; MECA),[28] donde los iones son excitados de manera resonante y después se les permite relajar por colisiones; CID de muy baja energía (very-low-energy collision-induced dissociation; VLE CID),[29] para lo cual se utiliza una deriva de fase de 180º en la forma de onda de la excitación; o la más simple de implementar y al mismo tiempo la técnica más robusta conocida como irradiación sostenida fuera de resonancia (sustained off-resonance irradiation; SORI).[30] En SORI los iones de una masa específica son excitados por un campo de RF, que está un poco fuera de la resonancia con la frecuencia del ciclotrón. Como resultado los iones son alternativamente acelerados y desacelerados con un periodo, que es reciproco a la diferencia de frecuencia.

Todos estos métodos no activan los iones en un solo evento, sino mediante el uso de colisiones múltiples. Ya que se utiliza el gas de colisión introducido a través de una válvula pulsada con apertura de milisegundos, normalmente se produce un retraso de varios segundos para reducir la presión de gas antes de reanudar el análisis FTICR. Utilizando la activación por colisión de gas, los iones absorben la energía en múltiples etapas, y como hay tiempo suficiente para la redistribución de esta energía en la molécula (este proceso se define como “ergodic”), los enlaces más lábiles tienden a romper primero. Como muchas modificaciones postraduccionales son lábiles, incluida la fosforilación, estos enlaces tienden a romper primero. La fragmentación de un péptido no va a continuar, porque el resto del péptido no puede ser más excitado por la frecuencia de excitación original aplicada (la masa del péptido ha cambiado por la pérdida de su modificación y su frecuencia ciclotrónica es ahora diferente).

Disociación por multifotones infrarrojo

Otra estrategia para depositar energía en la molécula es por absorción de fotones. Los iones han sido activados en la cámara ICR utilizando fotones tanto infrarrojo (IR)[31,32] como ultravioleta (UV)[33]. Aunque la disociación por multifotones infra rojos (IRMPD) fue utilizada inicialmente para pequeñas moléculas, pronto se encontró aplicación en la secuencia de biomoléculas[34] Los espectros de fragmentación producidos son similares a los de disociación por colisión inducida, CID, pero con la ventaja añadida de no necesitar pulsos de gas que perjudican las condiciones de vacío. Otra ventaja adicional es, que los iones empiezan el análisis de masa desde el centro de la cámara ICR (lo que simplifica la excitación y maximiza la señal), ya que tienen que estar allí para una exitosa interacción con el haz láser. Sin embargo todavía depende de una interacción, que requiere un tiempo de unos pocos segundos para una exitosa fragmentación.

Disociación por captura de electrones

La disociación por captura de electrones (electron capture dissociation; ECD) ha evolucionado recientemente como un método de activación alternativo, especialmente para secuenciación de péptidos y proteínas. Con ECD, cationes múltiplemente cargados son irradiados con electrones de baja energía producidos por un cátodo emisor colocado detrás de la cámara ICR. La captura de electrones produce un radical catiónico $[M+nH]^{(n-1)+\bullet}$, que puede disociar por medio de una rápida y fácil fragmentación en el enlace N-C de la cadena peptídica, produciendo mayoritariamente iones fragmento de las series “c” y “z”.[35]

ECD tiene algunas notables ventajas; la fragmentación no es dirigida por la protonación del enlace peptídico y muestra naturaleza "no-ergódica" ("no-ergódica" significa que el proceso añade energía interna al precursor más rápidamente de lo que sería la tasa de asignación al azar de energía). Como consecuencia, la fragmentación no tiene que producirse en los enlaces más débiles dentro del compuesto, como sería en el caso de las técnicas CID o IRMPD. Así, las frágiles modificaciones postraduccionales dentro de la secuencia de proteína/péptido quedan preservadas. ECD permite asignar sitios específicos de fosforilación,36,37 *O*- y *N*-glicosilaciones,38 o modificaciones con sulfato.39 ECD es muy prometedora como técnica de disociación complementaria con CID para identificación de proteínas sin ambigüedades, secuenciación *de novo*40 y caracterización detallada de proteínas.41 El método ha demostrado distinguir residuos de leucina e isoleucina42 e incluso diferencias entre péptidos, que contengan un amino ácido D- o L- en su secuencia.43 ECD ha sido también aplicada para el análisis de proteínas intactas,44 proporcionando un buen nivel de fragmentación.

Resumiendo lo anterior, CID e IRMPD inducen la disociación por la excitación vibracional de los precursores, resultando por lo general en la ruptura por el enlace más débil. Dentro de los péptidos, el enlace peptídico tiene la barrera de energía de disociación más baja, dando como resultado el predominio de la formación de iones fragmento del tipo b- e y-. Muchos sustituyentes, tales como modificaciones postraduccionales, tienen a menudo barreras de energía de disociación menores, que las de la ruptura del enlace peptídico. Esto puede dar lugar a espectros de fragmentación más complejos, y potencialmente resultar en la pérdida de información sobre la localización de estos sustituyentes. Las modificaciones postraduccionales se conservan generalmente utilizando ECD como técnica de disociación.

5. Espectrómetro de masas Orbitrap.

El analizador de masas Orbitrap fue descrito por primera vez en el año 2000 y ha alcanzado el estatus de papel protagonista en la técnica de espectrometría de masas. La combinación del analizador de masas Orbitrap con un dispositivo externo de acumulación, como la trampa iónica lineal, permite múltiples niveles de fragmentación (MSn) para la elucidación de la estructura de compuestos y permite el acoplamiento a fuentes de ionización continuas tales como ionización química a presión atmosférica APCI, electrospray ESI o nanoelectrospray nESI. Las prestaciones analíticas (exactitud de masa y resolución) del Orbitrap combinadas con la facilidad de manejo y el tamaño (hay versiones de sobremesa disponibles) pueden abordar un amplio rango de aplicaciones desde identificación y secuenciación hasta el análisis a nivel de trazas de compuestos en matrices complejas, sea en Proteómica, metabolismo de drogas, control de dopaje o la detección de contaminantes en alimentos y medio ambiente.[45,46,47]

5.1. Movimiento de los iones confinados

El espectrómetro de masas Orbitrap, está formado por un electrodo central, en forma de huso, a lo largo su eje, fijado a alto voltaje cuando los iones son confinados, rodeado de otro electrodo en forma de barril que mantiene su potencial a tierra. El electrodo exterior se divide a su vez en dos mitades para formar un "par de electrodos" para la detección de la corriente de imagen (Figura 12). Los electrodos están diseñados de tal manera, que formen una distribución de potencial quadro-logarítmica con muy alta exactitud. Cuando los iones inician su movimiento en el radio y a la energía correcta, se forman trayectorias estables, que combinan la rotación alrededor del electrodo central con oscilaciones a lo largo del eje central, con una forma de espiral

compleja. Es importante significar, que el movimiento axial es completamente independiente del movimiento de rotación.

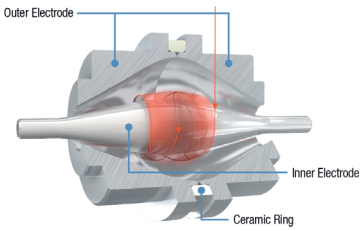


Figura 12. Analizador de masas Orbitrap. Los iones son confinados en un campo electrostático cuadrologarítmico. Un electrodo exterior, que consiste en dos mitades separadas por un material dieléctrico, encierra un electrodo como eje central. La corriente de imagen de los iones, moviéndose en anillos concéntricos a lo largo del electrodo central, se recoge en el electrodo externo.

Los electrodos del espectrómetro de masas Orbitrap crean un campo eléctrico no homogéneo en dos direcciones, axial y radial. Primero, el campo radial E_r atrae los iones hacia el electrodo central. Este campo es más intenso cerca de la zona central del electrodo. Para proporcionar una trayectoria circular, la velocidad tangencial de los iones debe ser ajustada a un valor tal, que la fuerza centrífuga compense la fuerza creada por E_r. En segundo lugar, la fuerza del campo axial E_z está en cero en el plano ecuatorial del analizador Orbitrap, pero aumenta uniformemente en las direcciones opuestas a lo largo del eje z, cuando los dos electrodos coaxiales llegan a estar progresivamente más próximos. Esto significa, que el campo eléctrico axial dirige los iones hacía el ecuador de la trampa con una fuerza proporcional a la distancia del ecuador. Ese campo acelera los iones hacia el ecuador y los iones continúan migrando pasando este punto (punto cero de fuerza de campo) a lo largo del eje z, pero se deceleran, a medida que continúan hacía el extremo opuesto del Orbitrap. Habiendo "gastado" su velocidad axial, los iones se paran y son entonces acelerados de vuelta al ecuador de la trampa por el campo eléctrico simétrico a lo largo del eje z. De esta manera los iones oscilan de forma natural a lo largo del eje z, de un modo análogo al que se mueve un péndulo. Es esta propiedad de campo eléctrico, la que provoca la oscilación armónica masa-dependiente de los iones a lo largo del eje z. Esta oscilación se combina entonces con un movimiento de rotación más complicado.

Las frecuencias rotacionales y las radiales son fuertemente dependientes de la energía inicial impuesta, de la posición y del ángulo de los iones, mientras que la frecuencia de las oscilaciones axiales es completamente independiente de las velocidades iniciales y las coordenadas de los iones. Por tanto, solo la frecuencia de las oscilaciones axiales puede ser utilizada para la determinación de la relación masa/carga *m/z*:

$$\omega = \sqrt{\frac{e}{m/z} \cdot k} \qquad (5)$$

Donde "*k*" es una constante proporcional a la diferencia de potencial entre el electrodo central y los electrodos externos.

5.2. Detección de la corriente de imagen

Las frecuencias de de oscilación axial se pueden medir directamente por la detección de la corriente de imagen producida en los electrodos externos del Orbitrap, después de haber sido amplificada. La ancha banda detectada se procesa por Transformada de Fourier para pasar el registro de la señal en dominio temporal a un espectro de relación masa/carga.[48,49]

La corriente de imagen es amplificada y procesada exactamente de la misma forma, que se hace para FTICR, dando resultados en sensibilidad y relación señal/ruido muy parecidos. Hay, sin embargo, una importante distinción entre los dos sistemas FTMS; la dependencia de la raíz cuadrada originada por la naturaleza del campo electrostático en el analizador Orbitrap provoca un descenso mucho más lento en el poder de resolución observado para los iones con el incremento de *m/z*. Como resultado de esto, el analizador Orbitrap puede, teóricamente, superar las prestaciones del FTICR, en lo que respecta a iones por encima de una cierta relación *m/z* (por lo general, por encima de 1.000 a 2.000 *m/z*) para el mismo tiempo de adquisición. La

figura 13 contrasta la tasa de caída de la resolución teórica de los dos tipos de FTMS, cuando la masa de los iones a observar crece. Para esta comparación se empleó un tiempo de adquisición de 1 segundo. Hay que tener en cuenta, que la resolución citada para el FTICR en esta figura es no-apodizada, mientras que la resolución en el Orbitrap es apodizada (lo que significa, que la resolución apodizada para el FTICR se reduciría a la mitad, de lo que muestra el gráfico).

La sensibilidad de la detección de la corriente de imagen está determinada por el ruido térmico de los componentes electrónicos del preamplificador de corriente de imagen. En el momento presente la mejor solución corresponde experimentalmente a un límite de detección igual a 5-10 cargas elementales por segundo de adquisición.

Es importante señalar, que para los conjuntos de iones una corriente de imagen no cero puede ser detectada solamente, si los iones se están moviendo en fase unos con otros. Esto significa, que los iones de cada *m/z* dentro del Orbitrap deben ser concentrados en paquetes, que tengan una longitud axial menor que la amplitud de oscilación. Es decir, los iones deben oscilar *coherentemente.* Las frecuencias radiales y de rotación no aparecen en el espectro de frecuencias, precisamente a causa de la rápida pérdida de coherencia causada por la dependencia de estas frecuencias de los parámetros iniciales del ión.

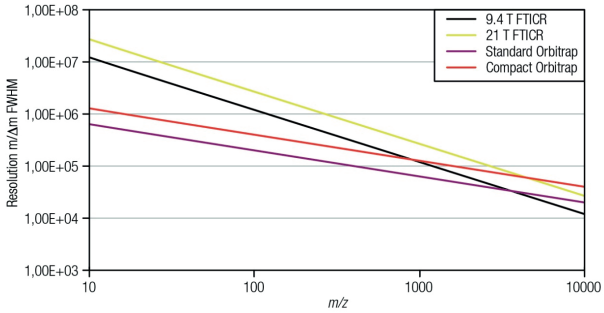


Figura 13. Resolución alcanzable con los analizadores FTICR y Orbitrap. Una señal transitoria con una duración de 1 segundo se ha considerado para FTICR con campos de intensidad de 9.4 y 21 T y dos diferentes diseños del analizador Orbitrap. La resolución, que se observa en el FTICR no es apodizada, mientras que la resolución observada en el analizador Orbitrap es apodizada. En ambos analizadores de FTMS la resolución decrece con el incremento de la masa del analito. Dentro de los parámetros experimentales utilizados en acoplamientos LC/MS para Proteómica (1 segundo de señal transitoria), el sistema Orbitrap puede superar al FTICR, incluso con imanes de gran potencia, para analitos con una gran m/z. Cortesía de Anastassios Giannakopulos, Thermo Fisher Scientific.

5.3. Conseguir los iones en el interior

Un pequeño paquete de iones inyectados desde un multipolo solo de RF entra en el analizador Orbitrap fuera de su plano de simetría a través de un pequeño orificio practicado en uno de los electrodos externos. Cuando los iones entran, experimentan una atracción por el campo del electrodo central que aumenta muy rápidamente con el tiempo. Este campo desvía sus trayectorias en arcos circulares. Los radios de los arcos se hacen progresivamente más pequeños con cada microsegundo que pasa, por lo que los iones no pueden volver al radio de entrada para golpear al electrodo exterior y perderse para su detección.

El componente axial del campo eléctrico es proporcional a la distancia del plano de simetría y así alcanza valores sustanciales en el punto de entrada. Se empieza por llevar a todos los iones hacía el plano de simetría, tan pronto como entran en el campo. Esto da lugar al inicio de la oscilación axial sin ningún tipo de excitación adicional.

Debido a la extracción pulsada del RF multipolar, que envía los iones al Orbitrap, los iones de la misma *m/z* (paquete de iones) entran en el campo de manera prácticamente simultanea. El tamaño axial del paquete de iones se mantiene casi constante y con una buena coherencia de

forma automática, cumpliendo así con este importante requisito previo para una exitosa detección de la corriente de imagen.

Después de que todos los paquetes de iones, en el rango de masas de interés, han entrado en el campo, el incremento del campo eléctrico se detiene y el potencial del electrodo central se estabiliza. Por tanto no se producirá deriva de la frecuencia durante la subsiguiente detección de la corriente de imagen.

Debido a la fuerte dependencia de las frecuencias de rotacionales de la energía de los iones, los ángulos y posiciones iniciales, cada paquete de iones se extiende rápidamente en su coordenada angular y forma un delgado anillo giratorio. Todos los anillos entonces oscilan a lo largo del electrodo central armónicamente, con una frecuencia dependiente solo de la *m/z* de cada ión.

El muy corto tiempo, que los iones están concentrados antes de formar el anillo giratorio, significa que los efectos espacio-carga no tienen tiempo para desarrollarse. Por otra parte, la existencia del electrodo central en el medio del anillo protege de manera efectiva una parte del anillo de la otra. Se puede alcanzar una alta capacidad de carga espacial (millones de iones) antes de que los efectos no deseados en el espectro de masas (es decir, los efectos de agregación, la pérdida de resolución y exactitud de masa) se observen. En comparación, utilizando FTICR, todos los iones de la misma masa permanecen en una línea y se necesita emplear imanes de mayor campo magnético para mitigar los efectos espacio-carga.

5.4. Factores limitantes del poder de resolución

En un analizador comercial Orbitrap (Orbitrap Elite), para alcanzar un poder de resolución nominal (apodizada) de 240.000 FWHM a 400 *m/z* se precisan 768 ms de tiempo de detección. En condiciones ideales, los iones pueden permanecer en el analizador por tiempo indefinido y no habría ningún límite para la resolución alcanzable. Desafortunadamente, las colisiones con el gas residual, así como minúsculas imperfecciones en la fabricación del electrodo, limitan el tiempo, durante el cual la señal puede ser detectada, y por tanto el máximo poder de resolución del analizador.

Las colisiones con el gas residual causan una pérdida de coherencia por la dispersión de los iones. En primer lugar, la pérdida de impulso de iones en las colisiones hace, que el paquete coherente se haga difuso, lo que aumenta las aberraciones y acelera aún más su difusión debido a otros factores como las imperfecciones de campo. En segundo lugar, las colisiones pueden contribuir estimular la fragmentación de iones metaestables, que podría conducir también a una pérdida directa de iones, si posteriormente chocan con el electrodo. Ambos procesos son aleatorios en el tiempo y, por tanto, producen paquetes de iones incoherentes. Estos no pueden ser detectados por su corriente de imagen, aún cuando los iones permanecen estables dentro de la trampa. El tiempo entre colisiones es inversamente proporcional a presión residual dentro de la trampa y a la sección transversal del ión. Para presiones de 10^{-10} Torr, el intervalo de tiempo para colisionar con el gas residual puede variar, desde varios segundos para moléculas pequeñas, hasta <1 segundo para moléculas medianas y proteínas. Hay que recordar, que aunque el tiempo entre colisiones disminuye para las moléculas más grandes (ya que esos tienen la sección transversal de colisión más grande), el porcentaje de pérdida de energía por colisión es menor. Es, porque se necesita, que tengan lugar múltiples colisiones, para que la energía cinética se disipe. Mejorando el nivel del vacío del sistema de masas Orbitrap, se ha podido determinar la resolución isotópica de los iones de proteínas de varias decenas de kDa.[50]

En la práctica, el poder de resolución alcanzable viene limitado por el tiempo característico necesario en separaciones LC (típicamente entre 0.5 y 1 segundo). El poder de resolución está entonces solo limitado por el número de oscilaciones de los iones durante este periodo de tiempo, que es proporcional a las frecuencias de las oscilaciones axiales. Como la frecuencia axial

es inversamente proporcional a la raíz cuadrada de *m/z*, el poder de resolución, en un tiempo limitado, también disminuye como $(m/z)^{-1/2}$.

Debido a la escasa dependencia de la sensibilidad del tiempo de detección, el analizador Orbitrap tiene una importante ventaja para acoplamientos a cromatografía de líquidos: el rango dinámico cae mucho más lentamente con un incremento de la frecuencia de repetición en comparación con otros analizadores de masa exacta (por ejemplo TOF).[51] En este contexto vale la pena mencionar, que el analizador Orbitrap muestra una insignificante pérdida de sensibilidad al aumentar la resolución.

6. Conectando los FTMS a otros espectrómetros. Sistemas híbridos.

FTICR es un sistema MS/MS por derecho propio y puede alcanzar alta resolución en la selección de iones precursores, aunque penalizando con una pérdida de tiempo en el proceso de aislamiento, fragmentación, "*axialización*" y nueva medición, y penalizando en la intensidad del ión precursor por la selección en alta resolución. Pero hay una ventaja significativa, cuando se utiliza un dispositivo de selección del ión precursor, como un cuadrupolo de transmisión y una celda de colisión o una trampa lineal cuadrupolar con capacidad para MS^n. En el caso de la trampa lineal cuadrupolar (LTQ), que incorpora su propio detector, se pueden realizar experimentos MS^n con detección en la trampa, mientras que se realizan las medidas en alta resolución en la FTICR.

La situación es diferente en el caso del analizador Orbitrap. Cuando un ión se descompone en su interior, los fragmentos mantendrán la misma velocidad. Como su energía permanece proporcional a su relación *m/z* individual, sus trayectorias se harán altamente elípticas y la mayoría de los fragmentos golpearán con uno de los electrodos y por tanto se perderán para la detección (con excepción de aquellos que tienen una relación *m/z* similar a la de los iones precursores). Aunque existen soluciones tecnológicas para conseguir la capacidad MS/MS dentro del analizador Orbitrap,[48] el primer sistema comercial introducido en el 2005 se realizó acoplando el analizador Orbitrap a una trampa iónica lineal (LTQ; Figura 14). El sistema híbrido MS resultante tiene capacidad MS^n y puede utilizar múltiples métodos de fragmentación (CID, disociación por alta energía de colisión (HCD), ETD). Es tal la versatilidad del sistema que puede ser tanto bien apreciado para análisis de rutina, como para Proteómica de "top-down". La Figura 15 muestra un espectro de fragmentación ETD de proteína intacta (29 kDa) y la cobertura de secuencia obtenida. La trampa iónica lineal fue elegida debido a su alta sensibilidad, preciso control de la población de iones, ciclos de tiempo cortos y alta capacidad de carga.[52] Una detallada des-

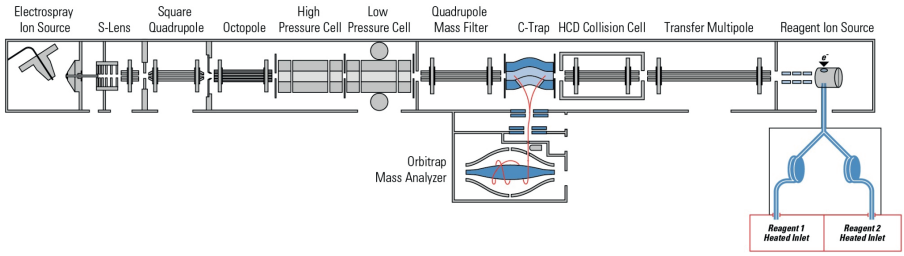


Figura 14. Esquema de un espectrómetro de masas híbrido ion trap-Orbitrap. Las partes más importantes del sistema de un instrumento disponible comercialmente se pueden contemplar en este diagrama. Un modulo opcional para experimentos ETD se puede incorporar para ampliar las posibilidades del equipo.

cripción de las prestaciones del sistema híbrido trampa iónica lineal-Orbitrap (LTQ Orbitrap) se puede encontrar en Makarov y col.[53]

Dependiendo de las necesidades analíticas, los dos analizadores de masas (la parte FTMS y la parte trampa iónica lineal del sistema híbrido) pueden ser utilizados de forma independiente o en conjunto. Los espectros MS^n detectados por cualquiera de los dos analizadores resultan muy similares. La única diferencia viene dada por la resolución y la exactitud de masa de los picos observados. Una gran ventaja de aquellos instrumentos híbridos es su capacidad de realizar un trabajo real en paralelo; la parte inicial de la señal transitoria en el FTICR/Orbitrap sirve para definir la masa de los iones precursores, que la trampa iónica lineal fragmentará, mientras la detección de corriente de imagen en el FTICR/Orbitrap continúa alcanzar la resolución final especificada. La otra ventaja, que el sistema híbrido FTMS ofrece, es la posibilidad de adquirir multiples espectros MS/MS en la trampa iónica lineal por cada espectro de masas adquirido en alta resolución en el FTICR/Orbitrap.

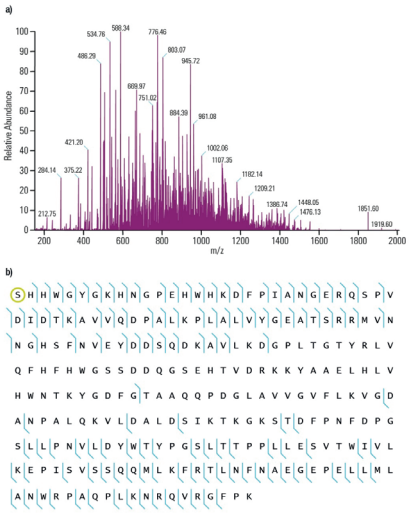


Figura 15. Análisis de una proteína intacta utilizando la técnica de disociación por transferencia de electrones (ETD) en el espectrómetro de masas híbrido ión trap-Orbitrap. A, El espectro de fraqmentación ETD de estado de carga 36^+ de anhidrasa carbónica (29 kDa). Tiempo de reacción ETD extendido (aquí se utilizaron 50 ms) genera fragmentos terminales cortos. B, Interpretación del espectro de ETD destacando una muy alta cobertura de la proteina. Para interpretación de los espectros ETD de alta resolución se ha utilizado el paquete de software ProSight™.

7. Prestaciones y características de los espectrómetros de masas basados en FT.

7.1. Exactitud de masa

Una de las características más codiciadas en un analizador de masas es, sin ninguna duda, su exactitud de masa. FTMS puede proporcionar de manera fiable una exactitud de masa inferior a 1 ppm.[51,54,55,56,57] Este logro se puede alcanzar con la ayuda de ciertos iones del ruido de fondo presentes por doquier. Se ha conseguido una exactitud en la medida de la masa de 100 ppb en una carrera cromatográfica LC/MS, sin necesidad de estándar interno. En caso de péptidos eso permite a menudo una sola composición elemental y por tanto representa la más alta exactitud útil.

La adquisición de espectros de masas en tándem con una exactitud de masa alta es una interesante alternativa al clásico esquema de adquisición, donde los iones se detectan con mucha menor exactitud de masa y resolución en la trampa iónica lineal o el triple cuadrupolo. Una exactitud mayor de masa de los fragmentos detectados añade especificidad a las identificaciones. Esto hace, que sea necesario adquirir un menor número de espectros MS/MS a alta resolución en comparación con la adquisición en un instrumento más rápido pero de resolución unidad, tal como la trampa iónica. Un mayor grado de confianza es, sin embargo, una ventaja decisiva en el caso de péptidos con modificaciones inesperadas.[58]

7.2. Poder de resolución

Muy alta resolución, de hecho, se puede lograr, cuando se registran largas señales transitorias. En un espectrómetro de masas de 9,4 T FTICR, por ejemplo, una señal transitoria de 5 segundos podría producir una resolución de 6.000.000 a 100 *m/z*; 600.000. a 1.000 *m/z;* y 120.000 a 5.000 *m/z*. Hay que tener en cuenta, que las proteínas intactas múltiplemente cargadas aparecen principalmente en el rango de *m/z* 1.000-5.000, lo que hace, que la técnica adecuada para el estudio de las grandes moléculas proteicas. El problema es, que conseguir señales transitorias de más de 5 segundos para las proteínas de gran tamaño, con su correspondiente gran sección de colisión, no es ninguna rutina, incluso a las muy bajas presiones empleadas en FTICR.

Puede ser de utilidad comparar la resolución alcanzada por un FTICR y un Orbitrap para varios valores de *m/z*. Teniendo en cuenta un tiempo de adquisición de 1 segundo, que es bastante típico en los experimentos de Proteómica, ya que coincide con la escala de separación cromatográfica de péptidos y proteínas, la Tabla 1 contrasta las prestaciones de un FTIRC de 9.4 T con las del diseño actual de Orbitrap.[59] El rendimiento del analizador Orbitrap para especies con una *m/z* de 1.000 coincide con el de no-apodizado FTICR de 9.4 T. Para especies con una *m/z* de 5.000 incluso le supera.

Tabla 1. Resolución alcanzada por espectrómetros de masas FTICR y Orbitrap para varios valores de m/z considerando 1 segundo de tiempo de adquisición. FTICR con 9.4 T de fuerza de campo y el analizador Orbitrap se han tenido en cuenta para estos cálculos. Téngase en cuenta que los valores de FTICR no son apodizados mientras que los valores de Orbitrap son apodizados.

m/z	9.4 T FTICR (no-apodizado)	Analizador Orbitrap (apodizado)
100	1,500,000	400,000
1000	120,000	120,000
5000	24,000	56,000

Otro de los retos relacionados con la detección por espectrometría de masas de grandes moléculas es, que cada especie molecular adquiere múltiples estados de carga.

La figura 16 muestra el espectro de ubiquitina adquirido con el analizador Orbitrap. Cada pico en el espectro corresponde a una carga específica, con incremento de estado de carga a la izquierda del espectro y disminución a la derecha del mismo. Cada estado de carga se compone entonces de múltiples picos, representando aductos, muchos de ellos diversos cationes provenientes de las sales y buffers. Y cada uno de ellos tiene a su vez una gran distribución isotópica (una especie de efecto de muñeca rusa). Todo esto significa efectivamente, que la señal inicial de la muestra se ha dividido en un gran número de picos, con el consiguiente impacto negativo sobre la sensibilidad aparente de detección.

Esta característica es inherente a todos los espectrómetros de masas, cuando utilizan ionización por electrospray al producir múltiples iones cargados. Sin embargo, se han publicado excelentes resultados describiendo el análisis de proteínas intactas (anticuerpos

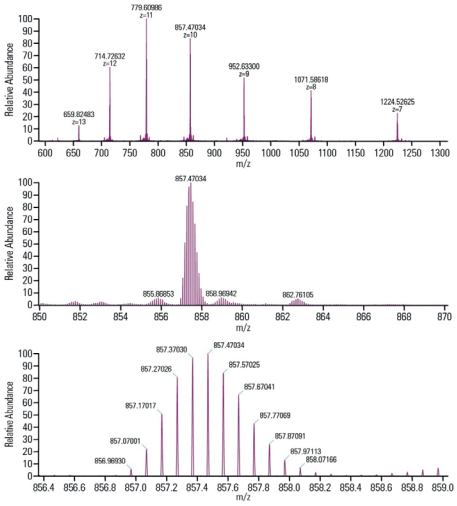


Figura 16. Análisis de proteínas intactas con espectrómetros de masas de alta resolución. La imagen superior muestra la envoltura del estado de carga de la ubiquitina. Una sección del espectro con una especie cargada con 10 cargas está en la mitad de la imagen, así como su posterior ampliación individual en la imagen inferior.

de aproximadamente 150 kDa de peso molecular) utilizando un sistema ESI-Orbitrap, tanto en modo infusión directa, como acoplado a LC.[60] Es pertinente señalar en este punto, que el uso de ionización por desorción asistida por láser (MALDI) no simplifica el asunto. Al detectar por FTMS una especie con *m/z* de 150.000 (MALDI produce en mayoría iones de una sola carga), la adquisición requeriría una señal transitoria extremadamente larga con el fin de alcanzar la resolución demanda. Empleando un detector de tiempo de vuelo (TOF) acoplado a la fuente MALDI no va mucho mejor, porque los iones producidos tendrían que ser acelerados con varias decenas de keV, para que un ión de ese tamaño pudiese producir una señal medible en el detector.[61]

8. Mejora de las prestaciones de los espectrómetros de masas FT.

El desarrollo se centra en obtener un mayor poder de resolución y por tanto una mayor exactitud de masa en un tiempo de adquisición fijo. Una forma de lograrlo es aumentar la frecuencia de las oscilaciones de los iones en el analizador, para que se puedan tomar más oscilaciones por cada relación frecuencia/masa en un tiempo de medida dado. En el analizador Orbitrap, esto se puede lograr disminuyendo la distancia entre el electrodo interior y exterior. Esto proporciona una mayor intensidad de campo para un voltaje dado, que nos lleva a una mayor frecuencia de oscilación. Un poder de resolución de más de 600.000 a 195 *m/z* y la resolución isotópica de proteínas de tamaño medio (aproximadamente 40 kDa) se ha conseguido.[59] Para FTICR, el desarrollo va en la dirección de un aumento de la intensidad del campo magnético (imanes más grandes), la mejora de la homogeneidad del campo eléctrico (tanto DC como RF) dentro de la cámara ICR y la exigencia de un control más preciso de la población de iones dentro de la cámara ICR(como se mencionaba en sección 4.4, 4.5, 4.6, y 4.7).

Para ambos tipos de espectrómetros de masas de FT, la implementación de la corrección de fase en los datos de FTMS resultaría en una mejora de casi el doble en resolución (sección 3.2). El requisito crítico para llevar a cabo este proceso es, que la detección debe incorporar el intervalo de excitación y que los espectros de excitación y detección deben estar temporalmente sincronizados.

En la práctica, esta excitación y detección simultánea es muy difícil de conseguir debido a la saturación del detector. Trabajos de laboratorio de Marshall mostraron como la corrección de fase en FTICR se podrían lograr, aunque no de manera rutinaria.[62,63]

La corrección de fase es mucho más sencilla en el Orbitrap,[64] ya que no hay ningún paso de excitación y el tiempo cero (t=0) es el tiempo de inyección de los iones en el analizador. Se pudo observar una mejora en la resolución de aproximadamente 1,7 veces. Analizando el estado de carga 47^+ de enolasa intacta de levadura (46,64 kDa) se logro la separación completa de sus isotopológos utilizando el analizador Orbitrap estándar (Figura 17).

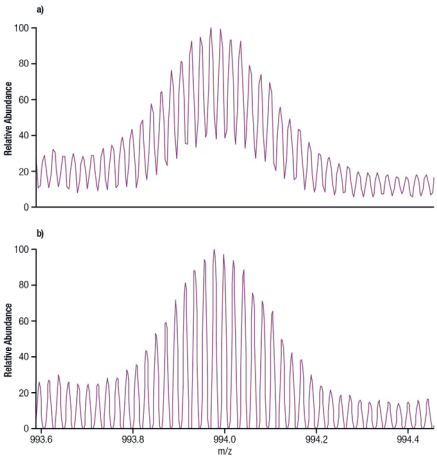


Figura 17. Mejora de la resolución utilizando corrección de fase en el analizador Orbitrap. Estado de carga 47^+ de enolasa de levadura intacta (46,64 kDa) fue detectada en el Orbitrap estándar (con 760 ms de señal transitoria). A, Un experimento sin la corrección de fase alcanza la separación isotópica a la mitad de la altura de los picos. B, El experimento con corrección de fase muestra la separación isotópica a línea de base, lo que mejora la resolución de 1,6 a 1,7 veces. Los espectros han sido realizados con apodizaciones diferentes.

9. Sumario.

FTMS se ha convertido en el "caballo de batalla" de las aplicaciones en Proteómica. La introducción del analizador Orbitrap y su incorporación a los sistemas híbridos, donde destaca la facilidad de manejo y solidez, han llevado la FTMS a la mayoría de las instalaciones de Proteómica. En la actualidad, la resolución del orden de millones se puede conseguir con espectrómetros de masas FT utilizando largas corrientes transitorias. En el campo de Proteómica, la capacidad de resolución ofrecida de más de 100.000 parece resolver la mayoría de los problemas analíticos. El alto rango dinámico de detección dentro de un mismo espectro junto con la posibilidad de medir con gran exactitud las frecuencias, y por lo tanto las masas, incluso a niveles de unos pocos iones, hacen de los instrumentos FTMS un sistema ideal para aplicaciones cromatográficas LC/MS. Los recientes desarrollos instrumentales en el Orbitrap han permitido, que los espectrómetros de masas de sobremesa puedan medir espectros de amplio rango de masas en alta resolución y con una supervisión mínima.

Incrementando las frecuencias de oscilación se podrán mejorar las prestaciones de ambos tipos de instrumentación FTMS. El progreso va en la dirección de imanes campo más alto para FTICR y, de manera similar, campos eléctricos más potentes y/o nuevas geometrías para el analizador Orbitrap. Teniendo en cuenta, que la tecnología Orbitrap está aún en su "infancia", se espera mucho más de su desarrollo. Instrumentos portátiles de alta resolución o nuevos sistemas de geometría híbrida representan algunas de las posibilidades.

AGRADECIMIENTOS.

Los autores desean agradecer a los científicos de Thermo Fisher Scientific, que han proporcionado datos, espectros, así como los gráficos y figuras utilizados en este capítulo: Eugen Damoc, Robert Malek, Anastassios Giannakopulos y Marcus Kellmann. Muchas gracias también a David Fenyo, de Rockefeller University, USA, por permitirnos la utilización de su material en alguna de las figuras. Las otras imágenes son propiedad registrada y su impresión requiere el permiso de Thermo Fisher Scientific.

REFERENCIAS

1. Kind T, Fiehn O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. Bioinformatics 2006; 7: 234-244.

2. Kind T, Fiehn O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *Bioinformatics* 2007; 8: 105-125.

3. Böcker S, Letzel MC, Lipták Z, Pervukhin A. SIRIUS: decomposing isotope patterns for metabolite identification. *Bioinformatics* 2009; 25: 218-224.

4. Zhu MS, Ma L, Zhang HY, Humphreys WG. Detection and structural characterization of glutathione-trapped reactive metabolites using liquid chromatography–high-resolution mass spectrometry and mass defect filtering. *Anal. Chem*. 2007; 79: 8333-8341.

5. Cox J, Mann M. Computational principles of determining and improving mass precision and accuracy for proteome measurements in an orbitrap. *J. Am. Soc. Mass Spectrom*. 2009; 20: 1477-1485.

6. Scheltema RA, Kamleh A, Wildridge D et al. Increasing the mass accuracy of high-resolution LC-MS data using background ions – a case study on the LTQ-Orbitrap. *Proteomics* 2008; 8: 4647-4656.

7. Wobschall D. Ion Cyclotron Resonance Spectrometry. *Rev. Sci. Instrum*. 1965; 36: 466-476.

8. Comisarow MB, Marshall AG. Fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett.* 1974; 25: 282-283.

9. Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* 1998;17: 1–35.

10. Comisarow MB, Marshall AG. Frequency-sweep Fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett*. 1974; 26: 489-490.

11. Marshall AG, Roe DC. Theory of Fourier transform ion cyclotron resonance mass spectroscopy: Response to frequency-sweep excitation. *J. Chem. Phys*. 1980; 73: 1581-1590.

12. Guan S, Marshall AG. Stored waveform inverse Fourier transform (SWIFT) ion excitation in trapped-ion mass spectrometry: Theory and applications. *Int. J. Mass Spectrom. Ion Proc.* 1996; 157-158: 5-37.

13. O'Connor PB, McLafferty FW. High-resolution ion isolation with the ion cyclotron resonance capacitively coupled open cell. *J. Am. Soc. Mass Spectrom.* 1995; 6: 533-535.

14. Guan S, Burlingame AL. High mass selectivity for top-down proteomics by application of SWIFT technology. *J. Am. Soc. Mass Spectrom*. 2010; 21: 455–459.

15. Ledford EB, Rempel DL, Gross ML. Space-charge effects in Fourier-transform mass-spectrometry - mass calibration. *Anal. Chem.* 1984; 56: 2744–2748.

16. Wang M, Marshall AG. Mass shifts induced by negative frequency peaks in linearly polarized Fourier transform ion cyclotron resonance signals. *Int. J. Mass Spectrom. Ion Proc*. 1988; 86: 31-51.

17. Syka JEP, Marto JA, Bai DL et al. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res*. 2004; 3: 621–626.

18. Hanson CD, Castro ME, Kerley EL, Russel DH. Field-corrected ion cell for ion cyclotron resonance. *Anal. Chem*. 1990; 62: 520-526.

19. Caravatti P, Allemann M. The infinity cell: a new trapped-ion cell with radiofrequency covered trapping electrodes for Fourier transform ion cyclotron resonance mass spectrometry. *Org. Mass Spectrom.* 1991; 26: 514-518.

20. Wang M, Marshall AG. A "screened" electrostatic ion trap for enhanced mass resolution, mass accuracy, reproducibility, and upper mass limit in Fourier-transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* 1989; 61: 1288–1293.

21. Wang M, Marshall AG. Elimination of z-ejection in Fourier transform ion cyclotron resonance mass spectrometry by radio frequency electric field shimming. *Anal. Chem*. 1990; 62: 515-520.

22. Beu SC, Laude DA Jr. Elimination of axial ejection during excitation with a capacitively coupled open trapped-ion cell for Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem*. 1992; 64: 177-180.

23. Bruce JE, Anderson GA, Lin CY, Gorshkov M, Rockwood AL, Smith RD. A novel high-performance Fourier transform ion cyclotron resonance cell for improved biopolymer characterization. *J. Mass Spectrom*. 2000; 35: 85-94.

24. Marshall AG. Theoretical signal-to-noise ratio and mass resolution in Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem*. 1979; 51: 1710–1714.

25. Marshall AG, Comisarow MB, Parisold G. Relaxation and spectral line shape in Fourier transform ion resonance spectroscopy. *J. Chem. Phys*. 1979; 71: 4434-4444.

26. Marshall AG, Guan S. Advantages of high magnetic field for FTICR MS. *Rapid Commun. Mass Spectrom*. 1996; 10: 1819-1823.

27. Nikolaev EN, Heeren RMA, Popov AM, Pozdneev AV, Chingin KS. Realistic modeling of ion cloud motion in a Fourier transform ion cyclotron resonance cell by use of a particle-in-cell approach. *Rapid Commun. Mass Spectrom*. 2007; 21: 3527–3546.

28. Lee SA, Jiao CQ, Huang Y, Freiser BS. Multiple excitation collisional activation in Fourier-transform mass spectrometry. *Rapid Commun. Mass Spectrom*. 1993; 7: 819-821.

29. Boering KA, Rolfe J, Brauman JI. Control of ion kinetic energy in ion cyclotron resonance spectrometry: Very-low-energy collision-induced dissociation. *Rapid Commun. Mass Spectrom*. 1992; 6: 303-305.

30. Gauthier JW, Trautman TR, Jacobson DB. Sustained off-resonance irradiation for collision-activated dissociation involving Fourier transform mass spectrometry. Collision-activated dissociation technique that emulates infrared multiphoton dissociation. *Anal. Chim. Acta* 1991; 246: 211-225.

31. Woodin RL, Bomse DS, Beauchamp JL. Multiphoton dissociation of molecules with low power continuous wave infrared laser radiation. *J. Am. Chem. Soc.* 1978; 100: 3248-3250.

32. Peiris DM, Cheeseman MA, Ramanathan R, Eyler JR. Infrared multiple photon dissociation spectra of gaseous ions. *J. Phys. Chem*. 1993; 97: 7839–7843.

33. Williams ER, Furlong JJP, McLafferty FW. Efficiency of Collisionally-activated dissociation and 193-nm photodissociation of peptide ions in Fourier transform mass spectrometry. *J. Am. Soc. Mass Spectrom.* 1990; 1: 288-294.

34. Little DP, Speir JP, Senko MW, O'Cornnor PB, McLafferty FW. Infrared multiphoton dissociation of large multiply-charged Ions for biomolecule sequencing. *Anal. Chem*. 1994; 66: 2809–2815.

35. Zubarev RA, Kelleher NL, McLafferty FW. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* 1998; 120: 3265-3266.

36. Stensballe A, Jensen ON, Olsen JV, Haselmann KF, Zubarev RA. Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun. Mass Spectrom.* 2000; 14: 1793-1800.

37. Shi SD, Hemling ME, Carr SA, Horn DM, Lindh I, McLafferty FW. Phosphopeptide/ phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal. Chem.* 2001; 73: 19-22.

38. Mirgorodskaya E, Roepstorff P, Zubarev RA. Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal. Chem.* 1999; 71: 4431-4436.

39. Kelleher NL, Zubarev RA, Bush K et al. Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. *Anal. Chem.* 1999; 71: 4250-4253.

40. Budnik BA, Olsen JV, Egorov TA et al. De novo sequencing of antimicrobial peptides isolated from the venom glands of the wolf spider *Lycosa singoriensis*. *J. Mass Spectrom.* 2004; 39: 193-201.

41. Kjeldsen F, Haselmann KF, Budnik BA, Sorensen ES, Zubarev RA. Complete characterization of posttranslational modification sites in the bovine milk protein PP3 by tandem mass spectrometry with electron capture dissociation as the last stage. *Anal. Chem.* 2003; 75: 2355-2361.

42. Kjeldsen F, Haselmann KF, Budnik BA, Jensen F, Zubarev RA. Dissociative capture of hot (3-13 eV) electrons by polypeptide polycations: an efficient process accompanied by secondary fragmentation. *Chem. Phys. Lett.* 2002; 356: 201-206.

43. Adams CM, Zubarev RA. Distinguishing and quantifying peptides and proteins containing D-amino acids by tandem mass spectrometry. *Anal. Chem*. 2005; 77: 4571-4580.

45. Ge Y, Lawhorn BG, ElNaggar M et al. Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc*. 2002; 124: 672–678.

46. Han X, Aslanian A, Yates JR. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol*. 2008; 12: 483-490.

47. Mann M, Kelleher NL. Precision proteomics: The case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U.S.A*. 2008; 105: 18132-18138.

48. Scigelova M, Makarov A. Advances in bioanalytical LC–MS using the Orbitrap™ mass analyzer. *Bioanalysis* 2009; 1: 741-754.

49. Perry RH, Cooks RG, Noll RJ. Orbitrap mass spectrometry: Instrumentation, ion motion and applications. *Mass Spectrom. Rev.* 2008; 27: 661-699.

50. Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal. Chem*. 2000; 72: 1156-1162.

51. Denisov E, Strupat K, Makarov AA, Zabrouskov V. Pushing intact protein detection limits of the Orbitrap mass analyzer. En: *Proc. 55th ASMS Conf. on Mass Spectrom. and Allied Topics*, Indianapolis; Am. Soc. Mass Spectrom.; 2007.

52. Makarov A, Denisov E, Lange O, Horning S. Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J. Am. Soc. Mass Spectrom*. 2006; 17: 977-982.

53. Schwartz JC, Senko MW, Syka JEP. A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom*. 2002; 13: 659-669.

54. Makarov A, Denisov E, Kholomeev A et al. Performance evaluation of a hybrid linear ion trap/Orbitrap mass spectrometer. *Anal. Chem*. 2006; 78: 2113-2120.

55. Olsen JV, de Godoy LM, Li G et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* 2005; 4: 2010-2021.

56. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotech*. 2008; 26: 1367-1372.

57. Dernovics M, Lobinski R. Speciation analysis of selenium metabolites in yeast-based food supplements by ICPMS assisted hydrophilic interaction HPLC hybrid linear ion trap/Orbitrap MS^n. *Anal. Chem*. 2008; 80: 3975-3984.

58. Erve JCL, DeMaio W, Talaat RE. Rapid metabolite identification with sub parts-per-million mass accuracy from biological matrices by direct infusion nanoelectrospray ionization after clean-up on a ZipTip and LTQ/Orbitrap mass spectrometry. *Rapid Commun. Mass Spectrom*. 2008; 22: 3015-3026.

59. Scherl A, Shaffer SA, Taylor GK et al. On the benefits of acquiring peptide fragment ions at high measured mass accuracy. *J. Am. Soc. Mass Spectrom*. 2008; 19: 891-901.

60. Makarov A, Denisov E, Lange O. Performance evaluation of a high-field Orbitrap mass analyzer. *J Am. Soc. Mass Spectrom*. 2009; 20: 1391-1396.

61. Bondarenko PV, Second TP, Zabrouskov V, Makarov AA, Zhang Z. Mass measurement and top-down HPLC/MS analysis of intact monoclonal antibodies on a hybrid linear quadrupole ion trap-Orbitrap mass spectrometer. *J. Am. Soc. Mass Spectro*m. 2009; 20: 1415-1424.

62. Kaufmann R, Kirsch D, Rood HA, Spengler B. Secondary-ion generation from large keV molecular primary ions incident on a stainless-steel dynode. *Rapid Commun. Mass Spectrom*. 1992; 6: 98-104.

63. Beu SC, Blakney GT, Quinn JP, Hendrickson CL, Marshall AG. Broadband phase correction of FT-ICR mass spectra via simultaneous excitation and detection. *Anal. Chem*. 2004; 76: 5756-5761.

64. Xian F, Hendrickson CL, Blakney GT, Beu SC, Marshall AG. Automated broadband phase correction of Fourier transform ion cyclotron resonance mass spectra. *Anal. Chem.* 2010; 82: 8807-8812.

65. Lange O, Makarov A, Balschun W, Denisov E. Accelerating spectral acquisition rate of Orbitrap mass spectrometry. *58th ASMS Conf. on Mass Spectrom. & Allied Topics.* Salt Lake City, Utah; Am. Soc. Mass Spectrom.; 2010.

Espectrómetro de masas de trampa iónica cuadrupolar

Michaela Scigelova[1], Bernabé Bodas[2], Jae Schwartz[3]

[1] Thermo Fisher Scientific, Bremen, Germany.

[2] Consultor, Madrid, Spain.

[3] Thermo Fisher Scientific, San Jose, CA.

1. Introducción.

Cuando Wolfgang Paul y su equipo de colaboradores desarrollaron un espectrómetro de masas basado en la estabilidad de paso en un campo alterno (a finales de 1950), nació el filtro de masas cuadrupolar. Al mismo tiempo, también describieron su analogo tridimensional, la trampa iónica cuadrupolar (1). Durante los primeros 30 años de su existencia, la trampa iónica cuadrupolar fue solo una herramienta especial utilizada por un número relativamente reducido de físicos e, incluso menor, de químicos (2). Como los iones permanecen confinados en estos analizadores durante un tiempo comparativamente largo, las trampas iónicas eran ideales para estudios de propiedades fundamentales de los iones. Cuando por fin se descubrió una manera práctica de obtener un espectro de masas de los iones contenidos en la trampa (3), apareció el primer espectrómetro de masas de trampa iónica cuadrupolar en 1984 (4), que puso esta tecnología muy rápidamente en las manos de cientos de químicos analíticos.

Las primeras versiones de este espectrómetro de masas utilizaban el impacto electrónico y la ionización química, que tenían lugar en el interior del analizador y permitían el análisis de muestras en fase gaseosa, incluyendo aquellas que eluían de un cromatógrafo de gases (GC). A principios de los 90, con la aparición de las técnicas de ionización a presión atmosférica, API, junto con la capacidad de inyectar con efectividad los iones en la trampa desde una fuente de iones externa, se desarrollaron los primeros acoplamientos de la cromatografía de líquidos (LC) a la trampa iónica cuadrupolar. Estos sistemas LC acoplados con las trampas iónicas cuadrupolares, LC/MS, emergieron como una poderosa herramienta de trabajo para el análisis cualitativo de muestras biológicas y su utilización se generalizó en todos los campos de la química y la biología (5).

La introducción de la trampa de iones lineal en 2002 (6, 7) se dirigió a solucionar las dos limitaciones principales de la trampa de iones cuadrupolar: los conocidos efectos espacio carga y su baja eficacia de confinamiento. Una eficacia más alta (mejorada más de 10 veces) y el aumento de la capacidad de almacenamiento (más de 20 veces) se convierten directamente en un aumento de la sensibilidad, mejores límites de detección, incremento del rango dinámico y un mayor rendimiento en aplicaciones MS^n. Una ventaja adicional de la configuración de la trampa iónica lineal incluye una mayor eficacia de detección (7).

Todas las mejoras de prestaciones asociadas a la trampa iónica lineal han hecho posible otros dos notables desarrollos en este campo. El primero ha sido la introducción de una novedosa técnica de fragmentación llamada disociación por trasferencia de electrones (ETD) (8). Aunque la disociación inducida por colisión (CID), asociada tradicionalmente con las trampas iónicas, es una técnica altamente efectiva, ETD puede además proporcionar una información de la secuencia superior, especialmente para cierta clase de péptidos (p. e. fosfopéptidos, glicopéptidos o péptidos muy básicos). La técnica ETD, que implica una reacción química de múltiples péptidos positivamente cargados, con especies cargadas negativamente enriquecidas en electrones en el interior de la trampa iónica, puede ser seguida como una reacción química. Estas reacciones secuenciales ión-ión, en combinación con la espectrometría de masas, pueden proporcionar una gran cantidad de información sobre las proteínas analizadas (9).

La idea de acoplar una trampa iónica lineal a un analizador de masas de Trasformada de Fourier, conocido por su alta resolución y medida de la masa exacta de los iones, fue otro avance en este desarrollo (10). El espectrómetro de masas híbrido resultante combina la velocidad, sensibilidad y capacidad MS^n de la trampa iónica lineal, con la exactitud de masa por debajo de la ppm y una resolución superior a 100.000 (definida como ancho total a mitad de la altura, FWHM). La introducción de un híbrido, trampa iónica lineal-analizador Orbitrap™ ha revolucionado, sin ninguna exageración, el campo de la Proteómica, haciendo un sistema de ultra altas prestaciones, sólido y fácil de manejar, accesible a una amplia comunidad de usuarios sin necesidad de ser expertos espectrometristas (11).

La más reciente generación disponible de trampas iónicas lineales (por si solas o como parte de un avanzado sistema híbrido) trata de hacer frente a los incómodos compromisos asociados con el régimen de presión de trabajo en una trampa de iones. Mientras que obtenemos una mayor eficacia en la inyección y disociación de los iones a presiones de operación relativamente más altas, al mismo tiempo, presiones de trabajo más bajas nos permiten alcanzar mayor resolución y mayor velocidad de análisis. Se ha desarrollado una configuración de doble presión, que consiste en dos cámaras de confinamiento separadas y comunicadas por una apertura, permitiendo conseguir dos regiones de presión diferencial para optimizar cada fase del experimento (12, 13).

La base del éxito de las trampas iónicas se ha debido fundamentalmente a sus características, sus eficientes múltiples niveles de fragmentación (MS^n) y su alta sensibilidad para proporcionar precisos espectros de masas en barridos completos. Estas dos cualidades dan como resultado, por ejemplo, la producción de espectros de masas con información de secuencia incluso a bajos niveles de concentración (atomoles) de muestras biológicas, que pueden ser procesadas automáticamente para identificar péptidos o proteínas (14). Dado su pequeño tamaño, relativo bajo costo, modesta presión de trabajo requerida y flexibilidad experimental, las trampas iónicas siguen siendo muy apreciados entre investigadores de diferentes campos de aplicación.

1.1. Hitos en la evolución de las trampas iónicas

1956 Presentación del dispositivo, trampa iónica cuadrupolar, por Wolfgang Paul y colaboradores.

1970's Finnigan Corporation (en la actualidad Thermo Fisher Scientific) desarrolla el detector de trampa iónica.

1983 Finnigan Corporation Introduce en el mercado la primera trampa iónica cuadrupolar.

1984 Stafford y sus compañeros de trabajo publican los detalles del barrido de inestabilidad selectivo de masa.

1990 Primera demostración de experimentos MS/MS en la trampa iónica.

2002 Introducción de la trampa iónica lineal.

2004 Implementación de la disociación por transferencia de electrones (ETD) en la trampa iónica lineal.

2004 Acoplamiento de la trampa iónica lineal a analizadores de resonancia ciclotrónica de iones (LTQ FT).
2005 Acoplamiento de la trampa iónica lineal al analizador Orbitrap (LTQ Orbitrap).
2010 Trampa iónica lineal de doble cámara de presión diferencial.

2. Anatomía y principio de funcionamiento.

Todos los experimentos de espectrometría de masas incluyen la ionización de un analito, su análisis por su relación *m/z* y su detección. Para espectrometría de masas en tándem, MS/MS, se requiere también conseguir la selección de un ión precursor de interés y su fragmentación para conseguir los iones fragmento. A continuación se describen en esta sección brevemente los principios de operación de la trampa de iones cuadrupolar dirigida a lectores con conocimientos básicos de física y de espectrometría de masas. Para alcanzar una mayor profundidad en la descripción de la teoría y la práctica de las trampas iónicas cuadrupolares los lectores deberían revisar respectivamente los trabajos (15-17) y las ediciones especiales de revistas (18), o la publicación de cinco volúmenes coeditada por March y Todd como la máxima autoridad en este tema (19). Para la discusión sobre técnicas de ionización se remite a los lectores al capítulo correspondiente en esta monografía. Puede que sea conveniente en este momento simplemente mencionar, que si bien el impacto electrónico y la ionización química se utilizan predominantemente para los métodos basados en separaciones GC para su acoplamiento a analizadores de trampa iónica cuadrupolar, mientras que las técnicas de ionización a presión atmosférica, tales como electrospray y nano electrospray, son las comúnmente utilizadas para análisis de muestras separadas por LC, como péptidos y proteínas.

2.1. Inyección y acumulación de iones

El análogo de tres dimensiones del filtro de masas cuadrupolar con simetría de rotación, la trampa de iones cuadrupolar, consiste en tres electrodos que generan un campo eléctrico cuadrupolar tridimensional (Figura 1). Dos electrodos en los extremos en forma de tapa están sometidos habitualmente a potencial de tierra con respecto al electrodo en forma de anillo situado entre ellos, al cual electrodo se le aplica un voltaje de radio frecuencia (RF) a menudo en el rango de MHz. Para funcionar como una trampa iónica, la dirección del campo debe seguir cambiando, para que la aceleración de los iones dentro de la dirección radial (dirección r) y su desaceleración en la dirección axial (dirección z) cambie a desaceleración radial y aceleración axial en cada ciclo de radio frecuencia RF (Figura 2). La aplicación de una señal de RF en el rango de MHz a la estructura del electrodo de tres elementos lleva al ión a someterse a oscilaciones estables, siempre que su relación *m/z* esté dentro de los límites apropiados. Los iones de una relación *m/z* dada son estables en la trampa bajo

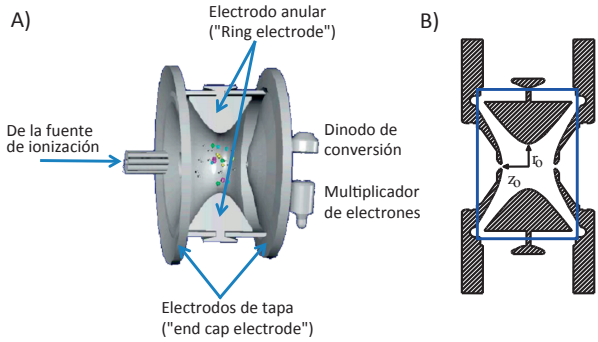


Figura 1. Trampa iónica cuadrupolar de simetría rotacional. A, La trampa iónica cuadrupolar de simetría rotacional consiste en tres electrodos: Dos electrodos externos en forma de tapa, "end cap", y un electrodo en forma de anillo entre ellos, "ring electrode". Los electrodos externos tienen un hueco cada uno, que permite la entrada de iones desde la región de la fuente de ionización y la salida hacia el detector, consistente en un dinodo de conversión y un multiplicador de electrones. B, Diagrama del corte en sección de la trampa iónica.

unas condiciones de operación, que se pueden resumir fácilmente en forma de diagrama de estabilidad expresado en los términos de las coordenadas de Mathieu a_z y q_z (Figura 3) cuyos valores consolidan algunas de las variables operacionales. Se requiere la estabilidad simultánea tanto radial como axial (que son independientes). De otra manera el ión saldría rápidamente del volumen o zona de confinamiento. Cuando los electrodos de los extremos en forma de tapa se mantienen a un potencial cero (a_z=0) el diagrama de estabilidad de masa puede ser simplemente representado como una línea de estabilidad de masa (Figura 4). La amplitud de RF aplicada al electrodo anular, determina entonces el rango de *m/z,* donde los iones permanecen estables en la trampa (véase la fórmula 1):

$$q_z = k \frac{V}{(m/e)} \qquad [1]$$

Donde *V* es el voltaje de la radio frecuencia (RF), *k* es una constante del instrumento y *m/e* es la masa en relación a la carga de un ión en la trampa.

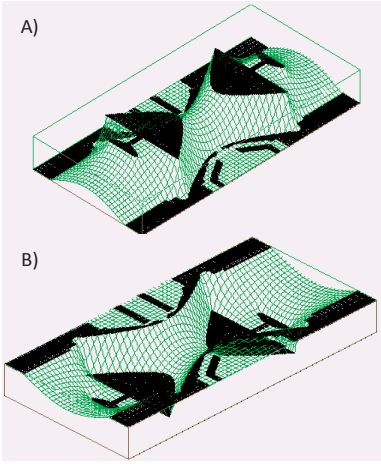


Figura 2. Simulación del potencial de confinamiento en una trampa iónica cuadrupolar de simetría rotacional. La aplicación de una señal de RF en el rango de MHz a una estructura de electrodos de tres elementos produce en los iones unas oscilaciones estables, siempre que su relación con m/z de los iones esté dentro del rango apropiado. La dirección del campo debe seguir cambiando, de modo que la aceleración del ión en la dirección radial y su desaceleración en la dirección z cambie a desaceleración radial y aceleración z durante cada cambio de ciclo de RF. A, Voltaje positivo en el electrodo anular. B, Voltaje negativo en el electrodo anular. Los gráficos han sido tomados del software SIMION®, Scientific Instruments Services, Inc.

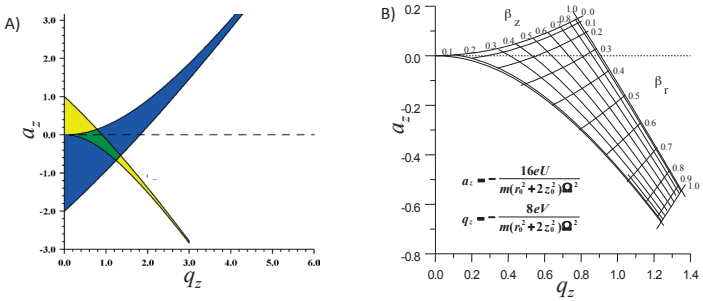


Figura 3. Diagrama de estabilidad de Mathieu. A, En el diagrama puede observarse la zona de confinamiento de los iones, en la dimensión r (área azul) y en la dimensión z (área amarilla). Para que los iones puedan ser confinados, necesitan estar confinados en ambas dimensiones r y z (área verde). B, Detalle de la zona de confinamiento de los iones. Los iones de una relación dada m/z (aquí en la fórmula designada como m/e) pueden variar de posición dentro de la zona. Se pueden mover de un punto a otro, cambiando la corriente continua DC y el voltaje de radiofrecuencia RF aplicado y, sin embargo, permanecen en trayectorias estables. Obsérvese, que las coordenadas a_z y q_z reflejan el potencial DC (U) y el potencial de RF (V) aplicado, respectivamente, la frecuencia angular (ω) de la RF y las dimensiones de la trampa (r_o y z_o). Para más detalles, consultar las referencias 16 y 19.

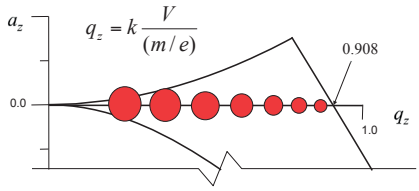


Figura 4. Línea de estabilidad de masa. Manteniendo el potencial de los electrodos de los extremos, "end cap", a tierra (z=0), el diagrama de estabilidad de masa de un ión en la trampa iónica se puede simplificar en una línea de estabilidad de masa. La amplitud de RF aplicada al electrodo anillo determina entonces el rango de cambio de masas, en el que los iones permanecen estables en la trampa (según la fórmula incluida).

Aunque es significativamente más complicado, es a menudo útil pensar en el confinamiento de los iones como un simple movimiento sinusoidal durante el almacenado, con frecuencias de movimiento inversamente proporcionales a su relación *m/z*. El efecto promedio del campo de confinamiento de RF conduce continuamente a los iones hacia el centro del dispositivo. Con la ayuda de un gas de amortiguamiento (se utiliza helio) introducido en la trampa, los iones ocupan un volumen casi esférico en el centro del dispositivo. Las colisiones con el helio disminuyen la energía cinética de los iones („*collisional cooling*"), disminuyendo el tamaño de su nube, dando como resultado tanto una mayor sensibilidad, como un incremento de la resolución del espectro de masas adquirido.

2.2. La obtención de un espectro

2.2.1. Barrido de inestabilidad selectiva de masas

El dispositivo de trampa iónica se vio relativamente poco utilizado como analizador de masas hasta que George Stafford y sus compañeros de trabajo, a principios de 1970 (3), presentaron un método de barrido llamado barrido de inestabilidad selectiva de masas. Este barrido utiliza el concepto, de que, con un cambio apropiado de los voltajes de operación, los iones confinados de una particular relación *m/z* serán inducidos a adoptar trayectorias inestables. De esta manera, aplicando un rango de la amplitud del voltaje (V) de RF, suministrado al electrodo anular, los iones de sucesivamente creciente relación *m/z* pueden ser inducidos a adoptar trayectorias inestables y salir de la trampa iónica a través de un pequeño orificio, para ser detectados utilizando un multiplicador de electrones montado externamente, proporcionando una señal y permitiendo por tanto obtener un espectro de masas.

$$q_z = k \frac{V}{(m/e)}$$

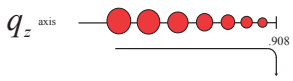


Figura 5. Barrido de inestabilidad selectiva de masa. Un ión con una baja relación m/z alcanza la inestabilidad con una menor amplitud de RF que otro con un valor más alto de m/z. Elevando el voltaje de RF los iones son desplazados a lo largo del eje q_z, hasta que alcanzan el límite de la zona de estabilidad, donde se vuelven inestables y salen de la trampa. Los iones son secuencialmente extraídos en función de su relación m/z.

Una imagen del barrido de inestabilidad selectiva de masas se puede ver en la Figura 5. Un ión con una baja relación *m/z* empieza a ser inestable a una menor amplitud de RF que uno de mayor relación *m/z*. En consecuencia, en un barrido de inestabilidad selectiva de masas los iones son desplazados a lo largo del eje q_z al elevar el voltaje de RF, hasta que llegan a ser inestables en el límite de la zona de estabilidad y salen de la trampa. Por tanto los iones de relación *m/z* creciente son secuencialmente extraídos y pueden ser detectados.

2.2.2. Resonancia de expulsion

El rango de relación masa/carga, que se puede obtener con diferentes tipos de analizadores de masas, viene normalmente dado por sus dimensiones y sus características de funcionamiento. Consecuentemente, el rango de masas pude ampliarse con los cambios apropiados en sus características operacionales, aunque esto generalmente tiene un costo sobre las prestaciones. Para la trampa de iones cuadrupolar trabajando en el modo de inestabilidad selectiva de masa la relación de condiciones operativas relativas al rango de masas es:

$$m/z = 4\, V / r_o^2\, \omega^2\, q_z \qquad [2]$$

Donde *V* es el voltaje de radio frecuencia (RF), ω es su frecuencia angular, r_o es el radio interno de la trampa de iones y q_z es un parámetro dependiente de la masa relativo al punto, en el cual se produce la inestabilidad.

El rango de masas de la trampa de iones puede ser ampliado, por tanto, por alguna de estas cuatro maneras:

- Trabajando a más altos valores de voltaje de RF
- Trabajando a más bajos valores de velocidad angular
- Utilizando trampas iónicas más pequeñas
- Produciendo la inestabilidad a valores inferiores de q_z.

El aumento del voltaje máximo de RF (que ya alcanza 15.000 V pico a pico) no es una propuesta muy atractiva debido a las serias posibilidades de averías eléctricas. La disminución de la frecuencia angular ha sido probada, así como la fabricación de trampas más pequeñas, pero ambas aproximaciones son, tanto instrumentalmente como técnicamente, muy exigentes y desde el punto de vista operativo comprometen otras características básicas, como la resolución y la velocidad de barrido. El camino más práctico para ampliar el rango de masas de la trampa resultó ser el de conseguir, que los iones confinados se vuelvan inestables a un valor q_z distinto, al que corresponde el límite normal de inestabilidad (20). Esto se consigue modulando el movimiento de los iones a la frecuencia elegida, utilizando un campo eléctrico dipolar aplicado a través de los electrodos de los extremos en forma de tapas. Los iones de una relación *m/z* dada, que están en resonancia con la frecuencia aplicada, pueden tomar una energía traslacional y pueden ser expulsados de la trampa. Este proceso se conoce como resonancia de expulsión, "resonant ejection" (21). Mediante la normal exploración de la amplitud del voltaje de RF principal, los iones de diferentes masas adquirirán secuencialmente valores q_z y frecuencias de movimiento hasta alcanzar la resonancia de expulsión. Los iones salen ahora de la trampa en secuencia, ordenados por sus valores de *m/z*, a una amplitud de RF menor de la que sería ordinariamente requerida. Para describir esta situación de una manera ilustrativa, el modo resonancia de expulsión crea un agujero en la región de estabilidad a un valor dado de q_z que corresponde a la frecuencia aplicada a los electrodos de los extremos (Figura 6).

El siguiente texto resume la secuencia de etapas empleadas para obtener el espectro de masas de los iones confinados dentro de la trampa iónica cuadrupolar. Consiste en un barrido lineal de la amplitud del voltaje de RF durante un cierto intervalo de tiempo, mientras se aplica un voltaje de resonancia de extracción adicional a través de los electrodos de los extremos. Los iones son extraídos cuando su frecuencia de movimiento coincide con la señal del voltaje de resonancia aplicado, debido a que la amplitud de RF está cambiando. Esto da como resultado, que los iones son extraídos de la trampa a través de la abertura de los electrodos extremos en orden a su relación *m/z* y son detectados. Dado que existe una relación directa entre la relación *m/z* de un ión y la amplitud de RF, en la que se va a extraer, tomando la respuesta de señal a la rampa de amplitud de RF obtendremos un espectro de masas.

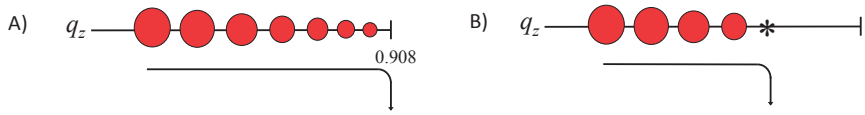


Figura 6. La resonancia de expulsión es una variable del barrido de inestabilidad selectiva de masa. A, Diagrama esquemático del barrido de inestabilidad selectiva de masa. B, Resonancia de expulsión. Utilizando un campo eléctrico dipolar, aplicado a través de los electrodos de los extremos, los iones de una particular relación m/z, en resonancia con la frecuencia aplicada, adquieren una energía traslacional y son expulsados de la trampa. Los iones confinados en la trampa se vuelven entonces inestables a otros valores de q_z (menores), que el que les corresponde por su límite de inestabilidad normal. Manteniendo el campo eléctrico dipolar conectado y haciendo al mismo tiempo un barrido de amplitud de voltaje de RF de manera controlada se consigue, que los iones de diferente masa adquieran secuencialmente valores de q_z, que les confieren entonces la frecuencia adecuada para alcanzar la resonancia de expulsión. Efectivamente, los iones salen de la trampa en secuencia en función de su valor m/z, a una amplitud de RF menor de la normalmente requerida. El modo de resonancia de expulsión crea una especie de hueco en la línea d estabilidad a un valor de q_z, que corresponde a la frecuencia aplicada en los electrodos de los extremos. Esta es una explicación necesariamente demasiado simplificada de la situación, que se produce en las masas a lo largo de zona de linealidad del eje q_z. Para una mayor profundización y detalles técnicos ver la referencia 21.

2.3. Sistema de detección

Un sistema común de detección para las trampas iónicas cuadrupolares consta de un dinodo de conversión y un multiplicador de electrones. El dinodo de conversión es una superficie metálica cóncava, localizada en un ángulo recto al haz de iones. Se aplica un alto potencial al dinodo de conversión, para crear un fuerte campo eléctrico suficiente para atraer los iones hacia su superficie. Para la detección en modo positivo o negativo, el voltaje se ajusta a -15 y +15 kV, respectivamente. Cuando un ión golpea la superficie del dinodo de conversión, se producen partículas secundarias, que serán eventualmente aceleradas en el multiplicador de electrones. Para la detección de iones positivos, el voltaje del dinodo de conversión se fija a -15 kV. Las partículas secundarias producidas incluyen electrones, fragmentos del ión original e iones que proceden de la capa de moléculas acumuladas de la superficie del dinodo. Solo las partículas con carga negativa (electrones e iones negativos) saldrán de la superficie del dinodo y viajarán hacia el multiplicador de electrones, cuando se aplica un potencial de -15 kV para la detección de iones positivos. Todas las partículas con carga positiva volverán a caer a la superficie del dinodo. Para la detección de iones negativos, se aplicará un voltaje de + 15kV al dinodo de conversión y todas las partículas secundarias cargadas positivamente (típicamente H^+ y otros pequeños iones positivos) son acelerados hacia el multiplicador de electrones.

El multiplicador de electrones es un dispositivo, que multiplica la corriente, que le llega por la colisión de los electrones con la superficie de un electrodo interno especialmente diseñado para este proceso de multiplicación de corriente. Las partículas secundarias del dinodo de conversión, por tanto, golpean la superficie de la pared interna del cátodo del multiplicador de electrones con suficiente energía como para liberar electrones. Los electrones extraídos se aceleran aún más, atraídos por el creciente gradiente de potencial positivo. Vuelven a colisionar de nuevo con la superficie interior, produciendo la emisión de más electrones. La corriente resultante es proporcional al número de partículas secundarias, que golpean el multiplicador. La amplificación de corriente proporcionada por un multiplicador de electrones es típicamente del orden de 10^5 a 10^6, para la colisión de un solo ión cargado, que golpea a la entrada del electrodo.

2.4. Manipulación de iones

Aunque prácticamente cualquier compuesto, que pueda ser ionizado, puede ser inyectado en la trampa iónica, y su masa medida utilizando el método de inestabilidad de masa, es la capacidad de manipular las trayectorias de los iones almacenados y realizar experimentos en tándem (MS/MS), lo que permite conseguir una importante información estructural. Hay que seguir dos pasos importantes en la manipulación de iones para realizar experimentos MS/MS y MS^n en la trampa iónica: el aislamiento de los iones de interés y la fragmentación de estos iones.

2.4.1. Aislamiento de iones

Una vez, que los iones han sido inyectados con éxito y están almacenados en la trampa, algún subconjunto de estos iones, que puede ser de especial interés, puede ser aislado para su futura caracterización. Aunque el aislamiento de los iones de una particular relación *m/z* (o rango de valores *m/z*), mientras que todos los demás son expulsados de la trampa, se puede conseguir por muchos diferentes métodos, un método común consiste en aplicar una forma de onda de expulsión de resonancia de multi-frecuencia (22). Esta técnica utiliza el proceso de resonancia de expulsión descrito anteriormente. Sin embargo, las frecuencias de resonancia múltiples se aplican de forma simultánea para expulsar muchos iones de diferente *m/z* simultáneamente excepto aquellas, que corresponden a la relación masa carga de los iones de interés (Figura 7).

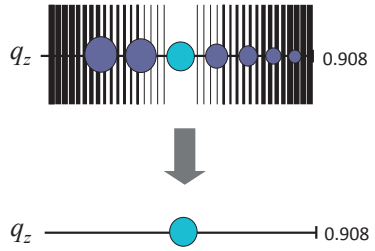


Figura 7. Aislamiento de iones de una relación concreta m/z. Se aplican múltiples frecuencias de resonancia, de manera simultánea, para expulsar todos los iones, a excepción de los iones de interés.

2.4.2. Fragmentación

La fragmentación de los iones en la trampa iónica se consigue más eficazmente por colisiones energéticas con las moléculas de helio presentes en la trampa. Así el helio que se ha utilizado como gas de amortiguamiento para reducir las colisiones por la energía cinética de los iones, sirve también apropiadamente para la disociación de los iones cuando su energía cinética es intencionadamente incrementada. La disociación por colisión ocurre como resultado de una técnica conocida como excitación resonante, "resonant excitation".

2.4.2.1. Excitación resonante

Tal como apuntábamos anteriormente, si los iones confinados se ponen en resonancia aplicando un campo adicional, los iones pueden ganar energía de este campo y ser expulsados de la trampa. Sin embargo, si la amplitud del voltaje complementario se mantiene suficientemente baja, los iones pueden ser cinéticamente excitados pero no expulsados. En este caso la fuerza de amortiguación proporcionada por las colisiones con helio elimina la energía de traslación y convierte una fracción de esa energía en energía interior de los iones. Esta energía interna puede causar la ruptura de los enlaces moleculares. La activación en una trampa iónica se produce así por múltiples colisiones a baja energía con el gas helio de amortiguación y la fragmentación es por tanto conocida como disociación inducida por colisión. El espectro MS/MS del ión precursor seleccionado se consigue entonces por la extracción secuencial de los iones producto, utilizando el modo de operación de inestabilidad selectiva de masa.

El particular, proceso para lograr la disociación inducida por colisión en la trampa iónica se denomina excitación resonante (Figura 8) y se realiza normalmente a un valor de la amplitud de RF principal, que pone al ión precursor de interés una frecuencia de oscilación fija (valor q). Siempre hay un compromiso entre el aumento de la energía traslacional a un nivel, que puede resultar en una expulsión de los iones de la trampa, y por otro lado alcanzar justo la suficiente energía, para que la activación por colisión se produzca sin llegar a la expulsión. En general, la eficiencia de la disociación inducida por colisión en una trampa iónica es notablemente alta, frecuentemente entre el 50 y el 100%, ayudada por el hecho, de que la mayoría de los iones producto formados son retenidos en la trampa y no se pierden, hasta que son extraídos para el análisis de masas.

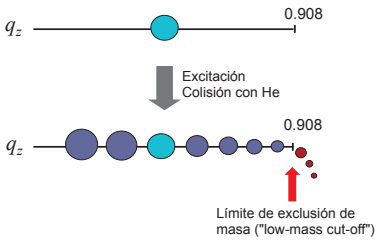


Figura 8. Resonancia de excitación. Los iones confinados pueden ganar energía, cuando entran en resonancia con un campo suplementario. Los iones son cinéticamente excitados, esto produce múltiples colisiones de baja energía con el gas helio de amortiguamiento y, entonces, se produce la fragmentación. La mayoría de los fragmentos quedan retenidos en la trampa. Los fragmentos más pequeños de m/z confinados vienen definidos por la amplitud del campo de excitación de RF aplicado (corte de masas bajas).

2.4.3. Capacidad para MS^n

La alta eficiencia de aislamiento y fragmentación de las trampas iónicas es particularmente atractiva, porque el proceso de fragmentación no tiene porque terminar justo en MS/MS. Los iones confinados pueden ser sometidos a sucesivas etapas de fragmentación (MS^n) secuencialmente en el mismo analizador, proporcionando así una completa información estructural (Figura 9). Se han llegado a realizar experimentos con diez o más etapas secuenciales de aislamiento y disociación por colisión en las trampas iónicas. En experimentos MS^n, la eficiencia, en general, es predominantemente controlada por la fracción sobre el producto total de corriente iónica, que proporciona un ión producto en particular, el cual sirve como ión precursor para la siguiente etapa de disociación. El hecho, de que una población suficientemente grande de iones de la muestra pueda ser confinada en la trampa antes de su disociación para asegurar los niveles adecuados de sensibilidad en la detección, es una de las mayores ventajas de las trampas iónicas.

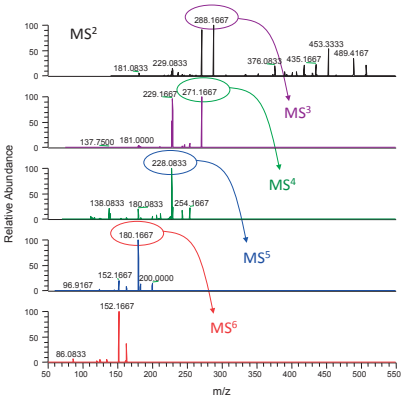


Figura 9. Múltiples niveles de fragmentación, MS^n. Un ión de carga doble de un pequeño péptido (MRFA secuencia de amino ácidos) ha sido sometido a 6 etapas de fragmentación consecutivas en una trampa iónica lineal. Estas detalladas investigaciones permiten el seguimiento preciso de iones fragmento de iones precursores seleccionados, imprescindible para una completa elucidación estructural de compuestos desconocidos.

2.5. Secuencia típica de operación

Asumiendo la comprensión general de los principios básicos de funcionamiento de la trampa iónica, descritos en los apartados anteriores, una secuencia típica de operación podría resumirse de la siguiente manera. En primer lugar, los iones del analito se introducen en la trampa (las trampas más modernas inyectan los iones desde una fuente de ionización externa). Este proceso es un reto importante, ya que los iones deben penetrar en un campo de RF tridimensional y pasan a través de la abertura de inyección situada en el electrodo tapa de uno de los extremos. Como resultado de esto solo una fracción de los iones, que entran, será atrapada. Podemos compensar esta ineficiencia simplemente incrementando el tiempo de acumulación. Así se puede asegurar, que la trampa iónica está llena hasta su capacidad total. Esta capacidad para acumular iones es la característica única de este tipo de analizadores y ayuda a incrementar la sensibilidad y mejorar los límites de detección.

Después de que una población de iones es atrapada, puede ser manipulada de diferentes formas, como ya hemos visto anteriormente. Un ejemplo común es el aislamiento de un ión de una particular *m/z* (o rango de valores *m/z*), mientras que los demás son expulsados de la trampa. Como resultado, el ión aislado puede ser disociado en fragmentos utilizando el proceso de excitación resonante, cuando los iones colisionan energéticamente con los átomos de helio. La secuencia de aislamiento y activación de los iones puede ser realizada durante un número variable de veces, hasta MS^n. Este es otra característica distintiva de los analizadores de trampa iónica. Cuando deseamos obtener el espectro de los iones contenidos en la trampa, iniciamos una rampa de RF a lo largo del voltaje de extracción resonante y los iones son extraídos hacia el detector. Los espectros obtenidos MS/MS o MS^n proporcionan información y características estructurales de los iones precursores aislados, como no puede hacerlo ningún otro analizador.

2.6. Análisis dependiente de los datos („*Data-dependent analysis*“)

El paso de MS a MS/MS o MS^n es relativamente simple y rápido en las trampas iónicas, si lo comparamos con el de otros analizadores. Los métodos de barridos dependientes de los datos obtenidos han sido implementados en las trampas iónicas desde sus principios y se han convertido en esenciales para obtener la máxima información en muestras complejas (23). En una configuración típica de un método de trabajo, el sistema primero adquiere un espectro de masas de un barrido completo. A continuación determina los componentes más intensos del espectro y automáticamente realiza un MS/MS de cada ión precursor, proporcionando información estructural de las N masas de iones seleccionadas. Aquí N es elegido en base a la media de la anchura del pico cromatográfico (tiempo disponible para el análisis de un compuesto dado) y del tiempo total de medida necesario para completar el ciclo de aislamiento y fragmentación. Esto explica, porqué la importancia de la velocidad (tiempo de ciclo) de adquisición, mientras podamos mantener una aceptable calidad de los datos obtenidos, es uno de los parámetros básicos de las prestaciones de los modernos espectrómetros de masas. En el siguiente barrido

MS, el sistema puede ignorar iones, que ya ha sido analizado por MS/MS, y localizar un nuevo número de N iones en el espectro de masas para producir nuevos espectros MS/MS o MS^n.

Se pueden especificar muchos diferentes criterios para activar la adquisición de espectros MS/MS, además de simplemente la intensidad. Por ejemplo, el sistema puede precisar una cierta relación de intensidades entre dos iones que se encuentren en el espectro de masas, para utilizarlo con analitos marcados o etiquetados, y/o los iones deben tener cierto estado de carga, o que deban estar por encima o por debajo de un cierto umbral de señal. Decisiones muy avanzadas se pueden esperar ahora de los modernos sistemas basados en trampa iónica. El análisis de fosfopéptidos proporciona un ilustrativo ejemplo de la utilidad de la toma lógica de decisiones, cuando la fragmentación MS^3 será activada, solo cuando se haya detectado una pérdida neutra especificada del ión precursor. El instrumento monitoriza las diferencias de masa entre el ión precursor seleccionado para su fragmentación y los N iones fragmento más intensos de su espectro MS/MS y activará el modo MS^3, cuando sea detectada la pérdida especificada.

2.7. Ciclo de trabajo y transmisión de iones

La eficiencia global y por tanto la sensibilidad de un espectrómetro de masas viene determinada, en general, por la trasmisión de los iones y por el ciclo de trabajo. El ciclo de trabajo es esencialmente la fracción de tiempo, durante la cual los iones de interés generados en la fuente iónica están siendo manipulados por el analizador de masas. Es inherente a la secuencia de operación descrita anteriormente, que los analizadores de masas de trampa iónica pueden tener un mayor ciclo de trabajo, comparados con otros analizadores de masas, y sin embargo mayor sensibilidad, especialmente en espectros de masas de barrido completo. Esto se debe, a que el espectro completo se obtiene con la misma población de iones, ya sea realizando un barrido completo MS o MS^n. En muchos sistemas de trampa iónica comerciales, el tiempo de acumulación de iones esta automáticamente controlado para optimizar el número de iones analizados y minimizar los efectos espacio/carga, utilizando un procedimiento llamado control de ganancia automático, “automatic gain control” (AGC; 24, 25) Consecuentemente, cuando el nivel de muestra disminuye, el ciclo de trabajo se incrementa. Esto contrasta con los analizadores de masas del tipo haz, como los filtros de masas cuadrupolares, que tienen ciclos de trabajo más altos solo cuando monitorizamos un solo ión seleccionado, pero tienen ciclos de trabajo muy bajos para modos de barrido completo, ya que cada barrido *m/z* necesita nuevos iones desde la fuente iónica.

Como ya se ha indicado, se producen considerables pérdidas de iones durante el llenado de la trampa. Por otro lado, el 100% de los iones son esencialmente expulsados o extraídos de la trampa. Pero debido a que el número de iones que sale por cada uno de los extremos es igual, y a que sólo los iones extraídos por un extremo son detectados (el otro extremo se abre hacia la región de la fuente donde no hay detector), la transmisión se ve reducida por un factor de 2 en esta etapa del análisis de masas. Estas son justo las deficiencias, que los nuevos desarrollos, que a continuación abordaremos, han reducido en gran medida.

3. Recientes desarrollos en tecnología de trampa iónica.

A pesar de que las trampas iónicas cuadrupolares han tenido un gran éxito, su rendimiento tiene algunas limitaciones fundamentales. Las trampas iónicas cuadrupolares tienen un campo de RF, que contiene los iones en las tres dimensiones. La carga total de iones de la trampa puede estar contenida en una nube solo, hasta que, inducida por los efectos espacio/carga,

se produzca la mutua repulsión columbimetrica de los iones. Esto degrada la calidad de los espectros de masa, incluyendo precisión de masa y resolución. Una insuficiencia de las trampas iónicas tradicionales es, que la capacidad de carga está bastante limitada (aproximadamente 1.000 cargas).

Una segunda limitación es la eficacia de confinamiento de los iones inyectados. Como se mencionó anteriormente, los iones inyectados deben penetrar un campo eléctrico de RF, y por tanto los iones serán solo adecuadamente confinados durante una estrecha ventana de tiempo de la oscilación +/- del voltaje de RF.

3.1. Trampa iónica lineal; diseño, prestaciones y características

Las limitaciones fundamentales de las trampas iónicas tradicionales han llevado a los investigadores a estudiar nuevas geometrías de confinamiento. Esto ha conducido por último al desarrollo de una trampa iónica lineal (6, 7). Aquí lineal se refiere al hecho, de que el confinamiento en el campo de RF existe solo en dos dimensiones (x e y, radial), mientras que el confinamiento en la tercera dimensión (z, axial) se lleva a cabo aplicando potenciales de corriente directa (DC). La trampa iónica lineal consiste en cuatro barras paralelas cortadas en tres secciones con una ranura longitudinal en dos barras de la sección central, para la extracción de los iones (Figura 10). El funcionamiento interno de la trampa iónica lineal es esencialmente el mismo que en la trampa tradicional, con la particular diferencia, de que los iones son extraídos radialmente a través de las ranuras practicadas en las barras de los electrodos.

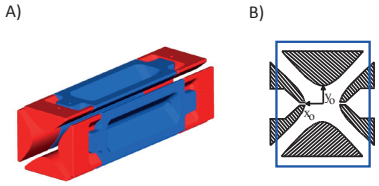


Figura 10. Representación esquemática de una trampa iónica lineal segmentada. A, La trampa iónica lineal consiste en cuatro barras paralelas cortadas en tres secciones (coloreadas). B, Diagrama de un corte en sección de la trampa iónica lineal. Observe las ranuras laterales en las dos secciones de barras centrales para la salida de los iones.

La geometría lineal soluciona, o al menos alivia, los problemas anteriormente mencionados de las trampas iónicas (efectos espacio/carga y eficacia de confinamiento). Como el volumen de confinamiento ha aumentado considerablemente, el dispositivo permite ampliar la forma esférica de la nube de iones a una de forma cilíndrica de longitud variable, dando como resultado una capacidad de carga mayor. La capacidad típica de carga de iones, para experimentos de barridos completos, es de 30.000 (comparados con los aproximadamente 1.000 de las trampas iónicas tradicionales). Se pueden obtener, por tanto, espectros de alta calidad con un solo barrido (sin necesidad de hacer promedio). La mayor capacidad de carga de las trampas iónicas lineales abre la ruta a nuevos análisis de especies altamente cargadas, como grandes proteínas intactas (Figura 11), simplemente porque ahora un número suficiente de iones altamente cargados puede ser analizado en la trampa, asegurando una buena detección y sensibilidad, sin los indeseables efectos de espacio/carga. Debido a que los iones se inyectan en el eje de las x, que no tiene campo de RF, la eficacia de confinamiento se ve también significativamente mejorada. En general,

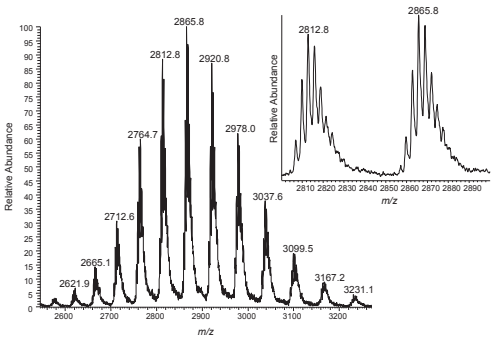


Figura 11. Análisis de proteínas intactas. Análisis por LC-MS de una especie intacta de inmunoglobulina G (1 mg en columna) realizado con un sistema de trampa iónica lineal LTQ Velos (Thermo Fisher Scientific). Se han adquirido más de 250 barridos en menos de 1 minuto durante una carrera de LC. Especies de proteínas con diferentes estados de glicosilación son detalladamente discernibles.

la trampa iónica lineal, incluso con una longitud moderada de aproximadamente 8 cm, proporciona una mejora de eficacia de confinamiento de más de 10 veces e incrementa la capacidad de almacenamiento más de 20 veces (7). Esto se traduce directamente en una mayor sensibilidad, mejores límites de detección, mayor rango dinámico y mejor rendimiento en experimentos MS^n.

Una ventaja adicional de la trampa iónica lineal consiste en una mayor eficacia de detección. Como se mencionó anteriormente, solo la mitad de los iones son detectados durante el proceso de extracción en las trampas iónicas tradicionales, ya que la otra mitad salen de la trampa por la apertura del electrodo del extremo conectado a la fuente iónica (y por tanto no pueden ser detectados). En la trampa iónica lineal, los iones se inyectan de forma ortogonal al eje de detección. Así podemos utilizar dos detectores situados a ambos lados de la trampa para detectar el doble de iones, que podrían ser detectados en las mismas condiciones en una trampa iónica tradicional. Esto incrementa el número de iones detectados por un factor de dos. Por tanto, podemos detectar hasta 40 veces más iones en cada barrido, si recordemos la capacidad de almacenamiento 20 veces mayor utilizando la trampa iónica lineal.

Reduciendo la velocidad de barrido, la trampa iónica lineal tiene capacidad para obtener resolución mayor de 30.000 FWHM (7). Esta resolución ayuda a determinar inequívocamente el estado de carga de los iones, por el espacio entre los picos de sus isótopos, en el espectro de masas (Figura 12) e incluso a separar algunas especies isobáricas.

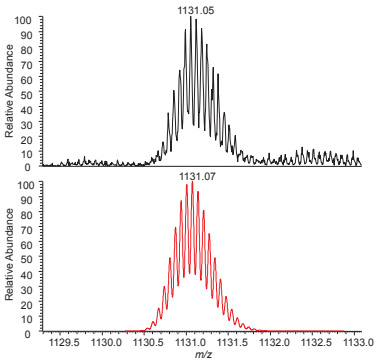


Figura 12. Análisis de alta resolución con la trampa iónica lineal. Los resultados experimentales obtenidos de experimento (imagen superior) y de simulación (imagen inferior), donde se demuestran isotopólogos de mioglobina, con un estado de carga de +15. Estos datos, que se han adquirido en un experimento de infusión utilizando el LTQ Velos, (Thermo Fisher Scientific) muestran una resolución superior a 25.000 FWHM.

3.2. Trampa iónica lineal de doble cámara y presión dual

Las condiciones operativas, tales como la presión de helio, se eligen a menudo para dar un razonable rendimiento al experimento global, pero no pueden ser optimizadas para cada etapa individual del funcionamiento de la secuencia de operación descrita. Por ejemplo, una mayor eficacia en la inyección de iones se obtiene a presiones más altas de trabajo, mientras que presiones de operación más bajas permiten alcanzar mayor resolución y velocidad de barrido. Para optimizar más aún el sistema, se desarrolló una configuración con doble cámara de presión diferencial, que permite tener dos regiones separadas, a diferentes presiones, para cada etapa del experimento (12, 13).

La trampa iónica lineal de doble cámara de presión dual consiste en dos cámaras de confinamiento separadas por una sola lente de apertura, lo cual permite mantener una presión diferencial entre las dos celdas. La trampa de presión dual tiene un total de seis secciones, cada una de ellas con un voltaje DC independiente, permitiendo de esta manera una apropiada manipulación de los iones, incluyendo una eficiente transferencia de los mismos de una cámara a la otra. Se utiliza un voltaje común de RF para todas las secciones, simplificando de esta manera la electrónica del sistema RF. Un diagrama esquemático de un instrumento equipado con la doble cámara de presión dual se puede ver en la Figura 13.

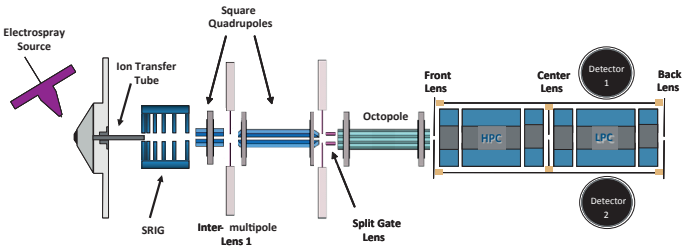


Figura 13. Esquema de un instrumento de trampa iónica lineal de doble cámara y presión dual. La eficacia de confinamiento así como la disociación inducida por colisión viene beneficiada por la mayor presión de helio, mientras que el incremento de la resolución se obtiene, cuando se reduce la presión de gas. La doble cámara de presión dual permite realizar cada etapa experimental en las condiciones óptimas de operación, proporcionando dos zonas de diferente presión en el instrumento. (HPC, cámara de alta presión ~5 mTorr; LPC, cámara de baja presión ~0.34 mTorr).

El efecto de la presión de helio en la eficacia de confinamiento se puede ver en la Figura 14. Los datos indican, que a la presión operativa utilizada previamente (~2.5 mTorr de helio) se obtenía una corriente total de 980 cuentas, que corresponde a una eficacia de confinamiento del ~60%. Elevando la presión operativa de helio por encima de 4 mTorr, la eficacia de confinamiento, que podemos conseguir, es >95%. Incrementando la presión también se consigue un positivo impacto sobre la eficacia de disociación de los iones durante el proceso de disociación inducida por colisión. La cámara de baja presión (operando a 0,34 mTorr) se utiliza para realizar la etapa de análisis de masas e incrementa la resolución de masas (expresada en FWHM) hasta un 30%, a la vez que también incrementa la velocidad de barrido por un factor de 2. Vale la pena señalar, que la resolución puede ser utilizada también para aumentar la velocidad de barrido en otro factor de 2. Así, la velocidad de barrido se puede incrementar hasta por un factor de 4, manteniendo la resolución original.

La configuración de trampa iónica lineal de doble cámara de presión dual resuelve el problema principal de la optimización de la presión de helio. Desacoplando las regiones de presión, utilizadas para la manipulación y análisis de los iones, permite conseguir significativas mejoras, que se pueden utilizar en otros procesos de tal manera, que los límites del rendimiento de la trampa de iones vuelven a experimentar un empuje en todos los aspectos.

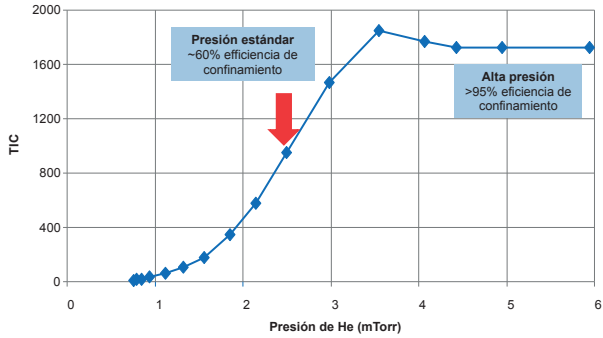


Figura 14. Efecto de la presión de helio en la eficacia de confinamiento en una trampa iónica lineal. A la presión de trabajo previamente utilizada (~2.5 mTorr de helio) se obtenía una corriente total de 980 cuentas, que corresponde a una eficacia de confinamiento de aproximadamente el 60%. Incrementando la presión de trabajo hasta 4 mTorr, se puede alcanzar una eficacia superior al 95%.

3.3. Disociación por transferencia de electrones

La capacidad para realizar estudios espectroscópicos de iones, reacciones ión/molécula y reacciones ión/ión se encontraba entre las prestaciones únicas, que la tecnología de trampa de iones podía ofrecer a los científicos en la primera etapa de la investigación de las trampas. Sin embargo, los esfuerzos se centraron en estudios fundamentales de reacciones en fase gas, convirtiéndose en una relativamente pequeña aplicación de las trampas iónicas, una vez que los espectrómetros de masas fueron controlados por ordenadores comerciales y se hicieron más útiles para realizar novedosos métodos de barridos de masas. Las aplicaciones más prác-

ticas de las trampas iónicas como espectrómetros de masas se generalizaron y robaron el protagonismo a otras aplicaciones. Pero en la actualidad, una combinación inteligente entre la química en fase gas y la espectrometría de masas en tándem ha conducido a la introducción de una novedosa técnica de fragmentación, la disociación por transferencia de electrones (ETD) y reacciones en fase gas ión/ión en la trampa iónica están de nuevo en el centro de atención.

ETD es análoga a la disociación por captura de electrones (ECD), el proceso descrito por Zubarev y colegas de trabajo. Péptidos o proteínas con multiples cargas confinados en la trampa magnética de un espectrómetro de masas de resonancia ciclotrónica de iones por transformada de Fourier, FT ICR, son expuestos a electrones con energías cercanas a energías térmicas (26). Esto lleva al esqueleto peptídico a fragmentarse por un proceso, que no implica redistribución de energía por vibración intramolecular (como en el caso de disociación inducida por colisión). Lo que hace destacable la ECD/ETD es, que la fragmentación se produce a lo largo del esqueleto del péptido de forma independiente de su secuencia, además preservando las modificaciones postraduccionales. El proceso se puede realizar en una escala de tiempo de milisegundos, con eficiencias de conversión de ión precursor a ión producto cercanas al 30%.

Mientras que la técnica de disociación ECD se puede incorporar a los espectrómetros de masas de resonancia ciclotrónica de iones por transformada de Fourier, lo que significa instrumentación de muy alto nivel de precio, no siempre accesible para los laboratorios de Proteómica, la técnica de disociación ETD ha sido desarrollada para su utilización en instrumentación de más bajo coste, totalmente incorporable a los espectrómetros de masas de trampa iónica lineal (8). La necesidad de electrones de energía cercana a energía térmica, que serían extremadamente difícil de manejar en las condiciones del potencial de confinamiento de RF utilizado para mantener los iones en la trampa iónica, se ha conseguido mediante el empleo de aniones con afinidad electrónica suficientemente baja, que pueden funcionar adecuadamente como sustitutos masivos de electrones donantes. El diseño segmentado de la trampa iónica lineal juega un papel fundamental en el proceso de manipulación de los iones (aniones del reactivo y cationes del analito, péptidos y proteínas) asegurando la máxima eficacia de todo el proceso. Un esquema de los pasos implicados en la operación del espectrómetro de masas de trampa iónica lineal para el análisis de la secuencia de un péptido utilizando ETD se puede ver en la Figura 15.

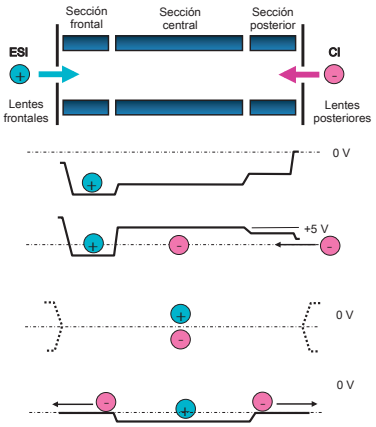


Figura 15. ETD en una trampa lineal segmentada. A, Cationes múltiplemente cargados (ej. péptidos) entran en la trampa lineal desde la fuente de ionización ESI (parte izquierda), mientras el reactivo ETD (aniones de fluoranteno) entra desde la fuente de ionización química situada en la parte posterior de la trampa iónica (parte derecha). B, Los cationes son atrapados y confinados en la sección frontal de la trampa lineal segmentada. C, Los aniones son inyectados. D, Los cationes son introducidos en la sección central, donde se encuentran con los aniones, permitiendo así su mezcla y reacción. E, Los aniones son entonces expulsados y los cationes están listos para ser escaneados fuera de la trampa y detectados.

La metodología ETD facilita el análisis de la secuencia de péptidos por la combinación de la química ión/ión y la de MS/MS. ETD está particularmente bien adaptado para la caracterización de péptidos, que contienen modificaciones postraduccionales, como fosforilaciones o glicosilaciones. Una extensión adicional de este enfoque implica la utilización de otros múltiples reactivos en fase gas para aplicaciones tales como la secuenciación de proteínas intactas (9).

3.4. Híbridos: trampa iónica lineal-espectrómetros de masas de transformada de Fourier

La evolución en el diseño de trampas iónicas lineales no solo ha mejorado las características y el rendimiento de los instrumentos, como el ciclo de trabajo y la eficacia de la transmisión,

sino que también ha abierto el camino a algunos desarrollos revolucionarios utilizando estos sistemas. Una de las mayores ventajas de esta tecnología es, que su geometría mecánica deja la parte posterior de la trampa libre para la extracción de los iones, haciendo a este analizador ideal para su acoplamiento a otros dispositivos, incluyendo otros analizadores de masas en configuraciones híbridas. Se puede incorporar una forma de trampa lineal como se incorpora el tercer cuadrupolo en un equipo de triple cuadrupolo (6).

Otra combinación extraordinariamente potente consiste en una trampa iónica lineal y un espectrómetro de masas por transformada de Fourier. Originalmente se combinó con un espectrómetro de masas de resonancia ciclotrónica de iones por transformada de Fourier (8), pero un sistema híbrido análogo, con un analizador de masas tipo Orbitrap, le siguió poco tiempo después (11). Estos sistemas de masas híbridos no solo mantienen toda la funcionalidad de la trampa iónica lineal, sino que pueden alcanzar precisiones por debajo de la ppm de precisión de masa (27) y resolución superior a 240.000 FWHM en barridos de 1 segundo.

Con unas características y prestaciones verdaderamente notables, su fiabilidad, y, lo más importante, su facilidad de manejo, los sistemas híbridos trampa iónica lineal-Orbitrap han alcanzado una popularidad sin precedentes para los trabajos de Proteómica. Se puede decir sin exagerar que estos nuevos sistemas están produciendo en la actualidad la mayoría de los datos en las aplicaciones proteómicas. Para ver el principio de funcionamiento de los espectrómetros de masas de resonancia ciclotrónica de iones por transformada de Fourier y de los analizadores Orbitrap, puede consultar en el capítulo correspondiente en este libro. Algunas publicaciones recientes proporcionan una perspectiva general de las aplicaciones, que se vienen llevando a cabo con este avanzado tipo de espectrómetros de masas híbridos de trampa iónica lineal combinada con la transformada de Fourier (28-30).

4. Conclusiones.

Las trampas de iones cuadrupolares han evolucionado significativamente en los últimos 30 años, convirtiéndose en la tecnología de referencia para aplicaciones de elucidación estructural en muestras complejas. Dado su pequeño tamaño, relativamente bajo costo, modestas presiones de trabajo requeridas y flexibilidad experimental, estos sistemas se han hecho muy populares. Su principal característica, en la que se basa su alto rendimiento y su éxito, es su capacidad para realizar múltiples niveles de fragmentación (MS^n), junto con su alta sensibilidad para realizar espectros de masas en barridos completos. Cada nueva generación de trampas iónicas desarrollada ha representado un significativo avance tecnológico, aportando soluciones a las limitaciones operativas fundamentales. Otros desarrollos, como los nuevos sistemas híbridos basados en combinaciones de dispositivos, incluyendo técnicas de ionización, se pueden imaginar fácilmente. Abren así las puertas a un futuro lleno de posibilidades, de incluso mejorar los actuales sistemas de trampa iónica para el análisis de muestras complejas en el mundo real.

AGRADECIMIENTOS.

Los autores quieren agradecer especialmente a Robert Malek y Anastassios Giannakopulos por sus útiles comentarios y sugerencias para la confección de este capítulo. Las imágenes son propiedad registrada y su impresión requiere el permiso de Thermo Fisher Scientific.

REFERENCIAS

1. **Paul W, Steinwedel H. 1956; German Patent 944,900.** 1960; US Patent 2,939,925.

2. **March RE, Hughes RJ. coordinadores. Quadrupole storage mass spectrometry.** New York: Wiley Interscience; 1989.

3. **Stafford GC, Kelley PE, Syka JEP, Reynolds WE, Todd JFJ. Recent improvements in and analytical applications of advanced ion trap technology.** Int. J. Mass Spectrom. Ion Processes. 1984; 60: 85-98.

4. **Syka JEP. Practical Aspects of Ion Trap Mass Spectrometry. En: March RE, Todd JFJ, coordinadores. Fundamentals of ion trap mass spectrometry.** Vol. 1. 1ª ed. Boca Raton: CRC Press; 1995. p. 169-205.

5. **Wagner Y, Sickmann A, Meyer HE, Daum G. Multidimensional nano-hplc for analysis of protein complexes.** J. Am. Soc. Mass Spectrom. 2003; 14: 1003-1011.

6. **Hager JW. A new linear ion trap mass spectrometer.** Rapid Commun. Mass Spectrom. 2002; 16: 512-526.

7. **Schwartz JC, Senko MW, Syka JEP. A two-dimensional quadrupole ion trap mass spectrometer.** J. Am. Soc. Mass Spectrom. 2002; 13: 659-669.

8. **Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.** Proc. Natl. Acad. Sci. USA. 2004; 101: 9528-9533.

9. **Coon JJ, Ueberheide B, Syka JEP et al. Protein identification using sequential ion/ion reactions and tandem mass spectrometry.** Proc. Natl. Acad. Sci. USA. 2005; 102: 9463-9468.

10. **Syka JEP, Marto JA, Bai DL et al. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications.** J. Proteome Res. 2004; 3: 621-626.

11. **Makarov A, Denisov E, Kholomeev A et al. Performance evaluation of a hybrid linear ion trap/Orbitrap mass spectrometer.** Anal. Chem. 2006; 78: 2113–2120.

12. **Schwartz JC, Syka JEP, Remes PM, Quarmby ST. A new dual cell linear ion trap configuration for improved quadrupole ion trap performance.** En: 56th Amer. Soc. Mass Spectrom. Annual Conf. on Mass Spectrometry & Allied Topics. Denver, CO; 2008.

13. **Olsen JV, Schwartz JC, Griep-Raming J et al. A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed.** Mol. Cell. Proteomics 2009; 8: 2759-2769.

14. **Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.** J. Am. Soc. Mass Spectrom. 1994; 5: 976-989.

15. **Todd JFJ. Ion trap mass spectrometer – past, presence, and future(?) Mass Spec.** Rev. 1991; 10: 3-52.

16. **Cooks RG, Glish GL, McLuckey SA, Kaiser RE. Ion trap mass spectrometry.** Chem. Eng. News 1991; 69: 26-41.

17. **McLuckey SA, Van Berkel GJ, Goeringer DE, Glish GL. Ion trap mass spectrometry of externally generated ions.** Anal. Chem. 1994; 66: 689-696.

18. **Glish GL, McLuckey SA. coordinadores. Quadrupole Ion Traps. Intl. J. Mass Spectrom.** Ion Processes 1991; 106.

19. **March RE and Todd JFJ. coordinadores.** Practical aspects of ion trap mass spectrometry, Modern spectrometry series, Vol. 1, Fundamental, (1995); Vol. 2, Instrumentation (1995); Vol. 3 Chemical, Biomedical, and Environmental Applications (1995); Vol. 4 Theory and Instrumentation (2010); Vol. 5 Applications of Ion Trapping devices (2010). Boca Raton, FL: CRC Press.

20. **Kaiser RE Jr., Cooks RG; Stafford GC, Syka JEP, Hemberger PH. Operation of a quadrupole ion trap mass spectrometer to achieve high mass/charge ratios.** Int. J. Mass Spectrom. Ion Proc. 1991; 106: 79-115.

21. **Syka JEP, Louris JN. 1988;** US Patent 4,736,101.

22. **Louris JN, Taylor DM. 1993;** US Patent 5,324,939.

23. **Wenner BR, Lynn BC. Factors that affect ion trap data-dependent MS/MS in proteomics.** J. Am. Soc. Mass Spectrom. 2004; 15: 150-157.

24. **Stafford GC, Taylor DM, Bradshaw SC, Syka JEP. En: Proc. 35th Am. Soc. Mass Spectrom. Annual Conf. on Mass Spectrometry & Allied Topics.** Denver, CO; Am. Soc. Mass Spectrom.; 1987. p. 775.

25. **Stafford GC, Taylor DM, Bradshaw SC, Syka JEP. 1992;** US Patent 5,107,109.

26. **Zubarev RA, Kelleher NL, McLafferty FW. Electron capture dissociation of multiply charged protein cations.** A nonergodic process. J. Am. Chem. Soc. 1998; 120: 3265-3266.

27. **Olsen JV, de Godoy LMF, Li G et al. Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap.** 2005; 4: 2010-2021.

28. **Perry RH, Cooks RG, Noll RJ. Orbitrap mass spectrometry: Instrumentation, ion motion and applications.** Mass Spectrom. Rev. 2008; 27: 661–699.

29. **Scigelova M, Makarov A. Advances in bioanalytical LC–MS using the Orbitrap™ mass analyzer.** Bioanalysis 2009; 1: 741–754.

30. **Makarov A, Scigelova M. Coupling liquid chromatography to Orbitrap mass spectrometry.** J. Chrom. A. 2010; 1217: 3938–3945.

El universo de configuraciones de analizadores de masas

13

Juan J. Calvete

Instituo de Biomedicina de Valencia, CSIC.

La espectrometría de masas es una técnica analítica que mide la relación masa/carga de partículas cargadas eléctricamente en fase gaseosa (iones). Para ello, todos los espectrómetros de masas combinan la formación, análisis y detección de los iones. Aunque un espectrómetro de masas puede definirse, pues, como una máquina sofisticada que manipula iones en el espacio y en el tiempo, la gran diversidad de maneras de combinar la separación y la detección de iones da cuenta de la cantidad de diseños de instrumentos existentes en el mercado. El objetivo de este capítulo es poner de relieve la idea de que los principios operacionales de la espectrometría de masas pueden ser entendidos mediante un lenguaje matemático simple, y que la comprensión de los conceptos básicos de la espectrometría de masas es necesaria para utilizar esta técnica versátil con el máximo provecho en el análisis proteómico. La tendencia actual hacia el diseño de sistemas híbridos de analizadores de masas se vislumbra como la clave conceptual y técnica para establecer el elusivo vínculo entre genotipo y fenotipo.

1. Conceptos básicos y ecuaciones que gobiernan las trayectorias de los iones en un analizador de masas.

La utilización de campos eléctrico y magnético para acelerar iones y establecer sus trayectorias en el interior de un espectrómetro de masas es un principio común de todos los analizadores de masa. Los movimientos de los iones en el interior del espectrómetro están definidos por las condiciones del campo y, por tanto, comprender como la relación masa/carga (*m/z*) de un ión puede derivarse mediante espectrometría de masas requiere conocer las ecuaciones que relacionan los parámetros que gobiernan las trayectorias de los iones. Además, dado que no existe un espectrómetro capaz de realizar todas las aplicaciones proteómicas, el conocimiento de las posibilidades analíticas y de las prestaciones de los diferentes sistemas de espectrometría de masas existentes en el mercado, en los servicios de proteómica, o en el laboratorio de un colega, es clave para sacar el máximo partido de la espectrometría de masas aplicada al análisis biológico.

La ecuación fundamental que gobierna la dinámica en el vacío de una partícula cargada sometida a un campo eléctrico combinado con un campo magnético es la Ley de la Fuerza de Lorentz: $F = q(E + vB)$, donde $F = ma$ (segunda Ley de Newton), E es el vector intensidad del

campo eléctrico, B es el vector inducción del campo magnético, q es la carga de la partícula (q = ze, siendo z el número de cargas y e la carga elemental: -1 para el electrón y +1 para el protón), y v es la velocidad expresada como vector. Por tanto, ma = q(E + vB), y m/q = (E + vB)/a. Por paradójico que parezca, un espectrómetro de masas no mide "masas" directamente sino la relación "masa/carga" de partículas con un número de cargas (z) ≥ 1. Desde que en 1897 J.J. Thomson descubriera el electrón [1] midiendo la relación q/m de partículas cargadas negativamente deflectadas mediante campos eléctrico y magnético en el primer "espectrógrafo de masas" consistente en un tubo de rayos catódicos a baja presión [2,3], se han desarrollado diversos métodos para determinar la relación masa/carga de iones en fase gaseosa. Los métodos más comunmente utilizados se basan en: la dispersión de partículas de diferentes momento iónico o energía cinética (instrumentos de sector eléctrico o magnético); la separación basada en la velocidad de los iones (instrumentos de tiempo de vuelo); la transmisión a través de un campo electromagnético (filtro de masas cuadrupolar); o los movimientos periódicos en campos magnético o electromagnético (trampas de iones).

En un espectrómetro de sector magnético (E = 0), la magnitud de la fuerza magnética ejercida sobre la partícula viene dada por $F_m = qvB\sin\beta$, donde β es el ángulo entre la velocidad y el campo magnético. Esto implica que la fuerza magnética sobre una carga estacionaria o en movimiento paralelo al campo magnético (β = 180º; sen180º = 0) es nula. Sustituyendo F_m por la fuerza centrípeta ($F = mv^2/r$), que es perpendicular a B y a v, y que mantiene en trayectoria circular a la partícula, obtenemos $qvB = mv^2/r$, donde m es la masa del ión y r el radio de su trayectoria. Dado que la velocidad del ión está relacionada con el voltaje (V) de aceleración por $qV = 1/2mv^2$, la relación masa/carga (m/q) = $B^2r^2/2V$. Es decir, para un radio de curvatura fijo, un ión de m/q determinada puede aislarse y medirse mediante una combinación adecuada de campo magnético y voltaje de aceleración. La Figura 1 ilustra un ejemplo de derivación de m/z en un espectrómetro de sector magnético.

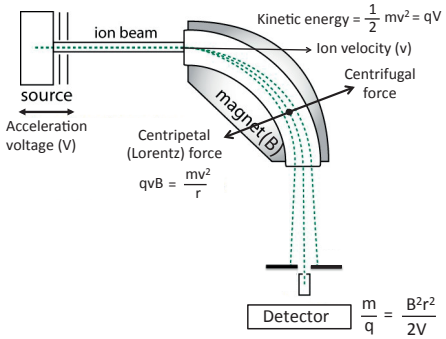


Figura 1. Esquema de la separación de iones de diferente m/q en un espectrómetro de masas de sector magnético. La figura ilustra como se deriva la masa de un ión (en u.m.a.) de un ión detectado en un campo magnético de 1.3 Teslas tras aceleración a 9000 V en el tubo de vuelo de un instrumento de sector magnético de 0.4 m de curvatura. En un campo magnético de intensidad B = 1.3T, un ión de carga q y velocidad v experimenta una fuerza F_B perpendicular a la dirección de B y a la dirección de movimiento del ión igual a Bqv. Dado que F_B está contrarestada por la fuerza centrífuga, $Bqv = mv^2/r$. Sustituyendo v de la ecuación de la energía cinética, $qV = 1/2mv^2$, donde m es la masa del ión en kg (1 u = 1 Da = 1.66×10^{-27} kg), v es la velocidad en m/s, q es la carga del ión (= ze, siendo z el número de cargas y e la carga elemental del electrón, 1.602×10^{-19} C), y V es el voltage de aceleración, tenemos que m(kg) = $zeB^2r^2/2V$ = m(u) = [1 x(1.602×10^{-19} C) x $(1.3T)^2$ x $(0.4\ m)^2$]/[2 x (1.66×10^{-27} kg) x 9000 V)] = 1455.36 Da. Notese que todas las unidades se cancelan dado que T = kg s^{-1} C^{-1} y CV = J = kg m^2/s^2, y por tanto la masa del ión está expresada en unidades de masa unificada (1 u = 1Da).

El cuadrupolo de transmisión consiste en cuatro rodillos metálicos paralelos. Los rodillos opuestos están conectados electricamente con un potential directo (DC) fijo y una corriente alterna de radiofrecuencia (RF) se aplica entre los pares de rodillos. Se puede seleccionar potenciales DC y RF superpuestos de tal manera que sólo un ión de determinada m/z tenga una trayectoria estable a través del cuadrupolo; todos los demás iones chocarán con los rodillos del cuadrupolo y nunca alcanzarán el detector. Ello permite seleccionar y aislar iones de determinada m/z o barrer un rango de valores de m/z mediante variación continua del voltaje aplicado. Las combinaciones de campos DC y RF que resultan en orbitas estables pueden visualizarse en un diagrama de estabilidad como el mostrado en la Figura 2. Las ecuaciones que definen el funcionamiento de un cuadrupolo son soluciones de las ecuaciones de Mathieu [4-6], derivadas originalmente en 1868 por el matemático francés Émile Léonard Mathieu (http://www.physics.drexel.edu/~tim/open/mat/mat.html):

$$a_x = -a_y = 4qU/m^2r_0^2$$

$$q_x = -q_y = 2qV/m^2r_0^2$$

siendo *a* y *q* parámetros relacionados con la magnitud de los potenciales DC (U) y RF (V) que corresponden a trayectorias estables (Figura 2), y r_0 es la distancia al eje central (Z) del cuadrupolo.

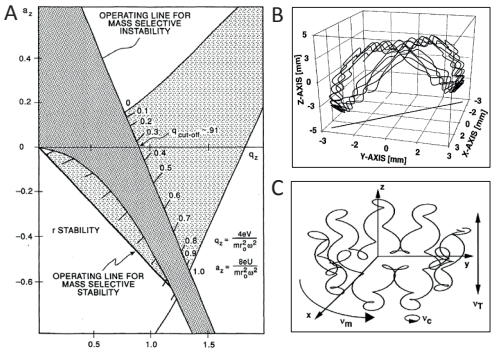


Figura 2. El panel A muestra un diagrama de estabilidad de las trayectorias de iones en función de los voltajes y frecuencias aplicados en una trampa ionica cuadrupolar (trampa de Paul). La figura está tomada de la patente 5399857 "Method and apparatus for trapping ions by increasing trapping voltage during ion introduction" US Patent Issued on March 21, 1995. El panel B muestra la proyección en el plano X-Y de la trayectoria de un ión de m/z 105 confinado en una trampa iónica tridimensional [90]. Panel C, trayectoria de un ión de m/z 2300 en un analizador de resonancia ciclotrónica (trampa de Penning) de 2 pulgadas cúbicas en un campo perfectamente homogéneo de 3T y un voltaje de confinamiento de 10V [91].

En un cuadrupolo de transmisión los iones oscilan en el el plano X-Y con frecuencias que dependen de sus valores de m/z. La trampa iónica cuadrupolar [7] es la versión tridimensional del cuadrupolo de transmisión. Tanto el cuadrupolo como la trampa iónica cuadrupolar fueron ideados por la misma persona, Wolfgang Paul, y la física de ambos analizadores es similar [7, y Capitulos 9 y 12]. En este analizador de masas los iones están sujetos a fuerzas (resultantes del campo RF aplicado) en los tres planos espaciales (X, Y, Z). En la trampa iónica "tridimensional" (3D, trampa de Paul) los iones confinados en un campo cuadrupolar de radio frecuencia experimentan fuerzas que los concentran en un pequeño volúmen en el centro de la trampa, entre un electrodo anular (*ring electrode*) y dos electrodos de tapa (*end-capped electrode*) (Figura 3). Los movimientos de los iones inducidos por el campo eléctrico resultante permite seleccionar qué iones permanecen confinados y cuales son eyectados de la trampa. Los valores de *a* y *q* de las ecuaciones de Matthieu dependen de las dimensiones de la trampa de acuerdo a las siguientes ecuaciones:

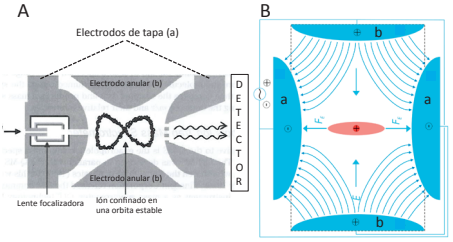


Figura 3. A, Esquema de una trampa de iones tridimensional (3D). El confinamiento de iones se consigue por aplicación de un campo eléctrico oscilante generado los electrodos de tapa y anular de la trampa iónica cuadrupolar (lineas azules en el panel B). En el panel B, un ión positivo está confinado entre los electrodos de la misma polaridad.

$$a_z = -2a_r = 8\ (q/m)\ (U/r_0^2w^2_{RF})$$

$$q_z = -2q_r = 4\ (q/m)\ (V_{RF}/r_0^2w^2_{RF})$$

Los subíndices *z* y *r* indican, respectivamente, movimiento perpendicular a, y entre, los electrodos de tapa de potencial DC = U; V es el potencial RF aplicado al electrodo anular, r_0 es el radio del electrodo anular, y w_{RF} es la frecuencia RF angular. El panel A de la Figura 2 muestra el diagrama de las regiones de estabilidad de una trampa iónica cuadrupolar en función de voltaje y frecuencia aplicados, y el panel B de la misma figura ilustra la trayectoria orbital estable de un ión de m/z 105 confinado en una trampa de iones de Paul.

Una característica única de la trampa de Paul, en comparación con el analizador de masas cuadrupolar, es que todas las etapas de un análisis MS^n (aislamiento y fragmentación del ión precursor, eyección y detección de los iones fragmento) se realizan secuencialmente en el mismo espacio [8,9]. La geometría de confinamiento de la trampa la trampa 3D limita su rango dinámico debido a efectos de carga en el espacio de confinamiento. Otra limitación de la fragmentación de iones en la trampa iónica de Paul es la denominada "regla de 1/3", que alude a que el límite inferior de masa de los iones fragmento confinables en la trampa es de alrededor del 28% de la masa del ión parental. Así, los principios operacionales de la trampa 3D (Fig. 2A) determinan que a valores del campo cuadrupolar $q_z > 0.908$ los iones adquieren trayectorias inestables y son eyectados de, o se pierden por colisión con las paredes de, la trampa. Consecuentemente, a cualquier valor de voltaje RF de confinamiento, hay un valor de m/z por debajo del cual los iones no tienen trayectorias estables. Este valor se denomina LMCO (*low mass cut-off*) y está determinado mediante la ecuación:

$$q_z = kV_{RF}/(m/z)$$

donde V_{RF} es la amplitud del voltaje RF de confinamiento, m/z es la relación masa/carga del ión seleccionado, y k es una constante que depende de las dimensiones internas de la trampa de la trampa iónica y de V_{RF}.

Dado que se requiere un mínimo de amplitud RF para producir la excitación y fragmentación de un ión, q_z debe tener un valor ≥ 0.2. Aumentando el valor de q_z se incrementa incrementan las colisiones energéticas produciendo una fragmentación más efectiva, pero también se incrementa el LMCO. Por tanto, hay que elegir un valor de q_z suficiente para producir una fragmentación eficiente y un valor de LMCO lo más bajo posible. Las trampas iónicas comerciales operan a $q_z = 0.25$. En estas condiciones, el LMCO del ión parental es (0.25/0.908) x 100 = 28%.

La realización de experimentos en tándem utilizando filtros de masa cuadrupolares requiere la combinación lineal de múltiples cuadrupolos. La combinación más utilizada es el triple cuadrupolo (Q1-q2-Q3). El primer triple cuadrupolo se contruyó en la Universidad de Michigan a finales de la década de 1970 [10]. En esta configuración, Q1 y Q3 actúan como filtros de masa mientra que q2, un cuadrupolo que solo trabaja con potencial RF, se utiliza como celda de colisión en la que el ión parental seleccionado en Q1 colisiona a ~30 eV con moléculas neutras de He, Ar, o N_2 a una presión de ~10^{-3} Torr, produciendo su fragmentación. Los iones fragmento son transferidos a Q3, donde se realiza un barrido de todo el rango de masas. Una variante del triple cuadrupolo es la utilización de Q3 como trampa de iones en lugar de en modo normal (RF/DC) [11]. La trampa confina iones radialmente mediante un campo bidimensional de radio frecuencias (de aquí su denominación de "trampa 2D"), y axialmente mediante potenciales eléctricos estáticos a ambos extremos de Q2, efectuándose la eyección de los iones axialmente con una eficacia de detección cinco veces superior a cuando la eyección se realiza radialmente, como es el caso en la trampa iónica cuadrupolar lineal (LTQ) [12]. Al igual que para la trampa de Paul, el parámetro q de Mathieu se define mediante la ecuación:

$$q = 4\ (q/m)\ (V_{RF}/r_0^2 w^2_{RF})$$

Un ejemplo de trampa iónica lineal 2D son las configuraciones QTrap® [12]. Ventajas de la trampa 2D respecto a la trampa 3D son la rapidez de escaneo de los mismos en todo el rango de m/z y la mayor capacidad de confinamiento de iones. Una comparación de las capacidades de confinamiento de iones (N) de las trampas 2D y 3D puede calcularse mediante consideraciones geométricas:

$$N_{2D}/N_{3D} = r_0^2 l/z_0^3,$$

donde r_0 y l son el radio y la longitud del campo de la trampa iónica 2D y z_0 es el radio del campo de la trampa 3D. Para una trampa 2D de 20 cm y una trampa 3D de z_0 = 0.707 cm, el incremento en la capacidad de iones de la trampa 2D respecto a la 3D es de aproximadamente un orden de magnitud [12].

Los principios operacionales del espectrómetro de tiempo de vuelo se establecieron ya en 1946 [13] y el primer instrumento, denominado Velocitron, fue construído un par de años despues [14]. En un instrumento de tiempo de vuelo los iones son acelerados a 1-20 kV. La velocidad que un ión adquiere es proporcional a su relación m/z y, por tanto, m/z puede derivarse del tiempo que el ión tarda en recorrer una distancia conocida (el tubo de vuelo, tipicamente 0.5-2 m) antes de llegar al detector. Así, cuando un ión es acelerado por un voltaje U y entra en el tubo de vuelo, su energía potencial (E_p = zeV = qV) es convertida en energía cinética, $qV=1/2mv^2$. Sustituyendo v = d/t, $t = [d/(2V)^{1/2}](m/q)^{1/2}$ o, dado que la distancia recorrida en el tubo de vuelo (d) y el voltaje de aceleración (U) son constantes instrumentales, $t = k(m/q)^{1/2}$.

La espectrometría de masas de transformada de Fourier de resonancia ciclotrónica de iones (FT-ICR MS, del inglés *Fourier-transform ion cyclotron resonace mass spectrometry*) fue desarrollada en 1974 [15,16]. La teoría de la resonancia ciclotrónica fue desarrollada por Ernest O. Lawrence (Premio Nobel de Física de 1939) [17]. El confinamiento y la detección de iones en una trampa ICR (que Hans G. Dehmelt, Premio Nobel de Física de 1989 compartido con Norman F. Ramsey y Wolfgang Paul, denominó "trampa de Penning" [18]) se basan en la observación de que, en un campo magnético uniforme (B), una particula cargada sometida a un campo eléctrico débil producido por electrodos hiperbólicos experimenta una fuerza de Lorentz que la hace rotar en un plano perpendicular al campo magnético, con una frecuencia ciclotrónica (*f*) inversamente proporcional a su relación m/q [16,19]: *f* = qB/2pm. Esta ecuación suele expresarse en términos de la frecuencia angular, relacionada con la frecuencia ciclotrónica por $f = w_c/2\pi$, y por tanto w_c = qB/m.

En una celda de ICR irradiada con pulsos RF de frecuencia w_c, los iones oscilan con movimientos estables (Fig.2C). Los iones así confinados absorben energía aumentando el radio de sus órbitas e induciendo una carga oscilante en la pared de la celda ICR. La frecuencia de esta carga oscilante es detectada y amplifiada para la determinación de la relación m/z de los iones confinados. Además, debido a que el radio orbital (r) de un ión excitado viene dado por la ecuación $r = (E_0 t_{excite})/2B$, donde E_0 y t son, respectivamente, el campo eléctrico y el tiempo de excitación, y B es la intensidad del campo magnético, r es, por tanto, independiente de la relación m/z del ión. Así, FT-ICR MS permite la detección y medición simultánea de un conjunto de iones de diferentes m/z excitados al mismo radio orbital. A este fenómeno se le denomina "la ventaja de Fellgett" [20], en honor a Peter Berners Fellgett, quién primero describió esta observación en su Tesis Doctoral [21,22]. Dado que la frecuencia de la carga oscilante puede ser medida fácilmente y con gran precisión, FT-ICR MS posee la mayor precisión en la determinación espectrométrica de masas. La resolución (R) de un analizador de masas puede definirse como $m/\Delta m$, siendo Δm la anchura al 50% de la altura de un pico de masa m. En FT-ICR $[m/\Delta m] = [\omega/\Delta w]$. Dado que la frecuencia ciclotrónica de un ión es directamente proporcional a la intensidad del campo magnético, un incremento de éste resulta en un incremento de la resolución. Además, en el límite de presión = 0, $\Delta\omega = 2\pi(1.2060)/T$ [23]. A baja presión, la resolución también puede expresarse como $m/\Delta m = (1.274 \times 10^7\ zB_0T)/m$ [24,25]. Queda claro, pues, que tanto aumentando el tiempo de medida (T) como la intensidad del campo magnético (B) puede incrementarse la resolución de la medida de m/z [26]. Se han logrado R > 1,500.000 utilizando espectrómetros FT-ICR, pero la necesidad de utilizar imanes superconductores a temperaturas cercanas al cero absoluto (He líquido, 4.2 K) imponen un alto coste de mantenimiento de estos equipos.

El nuevo milenio fue testigo del desarrollo por Alexander Makarov de un nuevo analizador de masas, Orbitrap™ [27] basado en el principio de confinamiento orbital en campo electrostáti-

co definido por Kingdon en 1923 [28] y modificado por Knight en 1981 [29]. El Orbitrap es una trampa de iones que utiliza un campo electrostático para confinar iones radialmente alrededor de un electrodo central en forma de huso (*spindle electrode*) [30; ver también capítulo 11]. Los iones confinados se mueven radialmente adelante y atrás a lo largo del eje z del campo eléctrico del electrodo central. La relación m/z se determina a partir de la frecuencia angular (ω, en rad s^{-1}) de las oscilaciones harmónicas del ión confinado a lo largo del eje z del campo eléctrico: $\omega = [(q/m)k]^{1/2}$, siendo k una constante del campo proporcional a la diferencia de potencial entre el electrodo central y los electrodos externos. La ecuación muestra que la frecuencia de la oscilación axial de un ión confinado en el Orbitrap es inversamente proporcional a $(m/q)^{1/2}$, mientras que la frecuencia ciclotrónica (w_c) en FT-ICR MS es inversamente proporcional a m/q. La corriente imagen generada es amplificada y procesada mediante transformada de Fourier para obtener el espectro de masas. El analizador de masas Orbitrap posee una sensibilidad y poder de resolución ($m/\Delta m = (1/2\Delta\omega)(kq/m)^{1/2}$) comparables al analizador FT-ICR.

2. Innovaciones para espectrometría de masas biológica: configuraciones híbridas.

Hacia 1970 la espectrometría de masas representaba un método estándar para el análisis de compuestos orgánicos, pero su aplicación en la investigación biológica estaba obstaculizada por la inexistencia de métodos adecuados para ionizar macromoléculas biológicas. Los métodos actualmente utilizados para la ionización de estas moléculas fueron desarrollados durante la década de 1980: ionización por radiación laser asistida por una matriz (MALDI, *matrix-assisted laser desorption/ionisation*) en 1985-1988 [31-34] y electronebulización (o electrospray, ESI) en 1984-1988 [35,36]. Al mismo tiempo que el desarrollo de métodos de ionización, una serie de innovaciones en la tecnología de los analizadores de masas ampliaron la gama de aplicaciones de la espectrometría de masas en la investigación biológica. Aunque ha habido innovaciones tecnológicas de analizadores de masas individuales (ej. Orbitrap descrito el capítulo 11), la tendencia en espectrometría de masas biológica contemporánea es el diseño de configuraciones híbridas que posean capacidades funcionales adicionales a los analizadores de masas individuales. El desarrollo inicial de configuraciones híbridas estuvo, en gran medida, motivado por la necesidad de realizar experimentos de MS/MS en una ventana de tiempos acotada por la escala temporal de la separación cromatográfica de los analitos. Un experimento de MS/MS se compone de etapas de análisis de iones en todo el rango de m/z, selección del ión precursor, fragmentación del ión precursor, y análisis espectral de los fragmentos. Dependiendo de cómo se realicen estas etapas, las configuraciones en tándem de analizadores de masas pueden clasificarse en dos grupos: sistemas constituídos por instrumentos que realizan las diferentes etapas en analizadores de masa separados fisicamente (*tandem-in-space,* TIS) y configuraciones en las que las diferentes etapas de MS/MS se realizan en el mismo instrumento (*tandem-in-time,* TIT).

La Figura 4 muestra las configuraciones híbridas más comunmente utilizadas en espectrometría de masas biológica, que implican combinaciones de las siguientes clases de analizadores de masas de biomoléculas: tiempo de vuelo (*time-of-flight*, TOF), cuadrupolo (*quadrupole*, Q), trampa iónica de Paul (*three-dimensional ion trap*, 3D-IT, *quadrupole ion trap*, QIT), trampa iónica lineal 2D (*linear ion trap*, LIT), trampa iónica lineal cuadrupolar (*linear trap quadrupole*, LTQ), Orbitrap, y FT-ICR. Las configuraciones del tipo *tandem-in-space* utilizan analizadores de tecnología de haz de iones (*beam-type*), comunmente cuadrupolos y analizadores de tiempo de vuelo. Las configuraciones *tandem-in-time* están asociadas a trampas de iones (trampa iónica cuadrupolar, FT-ICR, Orbitrap). Las primeras configuraciones híbridas sector

magnético-cuadrupolo (BqQ), sector magnético-TOF, sector magnético-QIT [37], y QqTOF, desarrolladas en las décadas de 1980 y 1990, utilizaban instrumentos de haz de iones en serie. Configuraciones híbridas que combinan instrumentos *in-space* e *in-time* (Fig.4) fueron introducidas a finales de los '90 [37]. Las configuraciones más recientemente desarrolladas, en los primeros años del nuevo siglo, utilizan combinaciones de diferentes instrumentos de confinamiento de iones (*in-time-in-time*) y trampa de iones-tiempo de vuelo (*in-time-in-space*) (Figura 4) [37].

Un parámetro importante en el acoplamiento de diversos analizadores de masas para la realización de experimentos de MS/MS es el control de la energía cinética de los iones [38]. Así, los instrumentos de sector magnético y los analizadores de tiempo de vuelo operan con iones a altas energías cinéticas (5-20 keV), mientras que los cuadrupolos y trampas iónicas manejan iones de baja energía cinética (<50 eV). El patrón de fragmentación de un ión depende de la energía interna del ión activado. En MS/MS, el proceso se controla mediante la energía de colisión y el número de colisiones entre el ión y átomos de gas neutro. El tiempo transcurrido entre la activación colisional de un ión y la detección de los fragmentos depende de la energía cinética del ión precursor. Una energía cinética alta favorece disociaciones inducidas por colisión (CID, *collision-induced dissociation*) rápidas, pero lograr una eficiencia de fragmentación alta requiere activar al ión precursor con una energía interna lo suficientemente alta para que la disociación

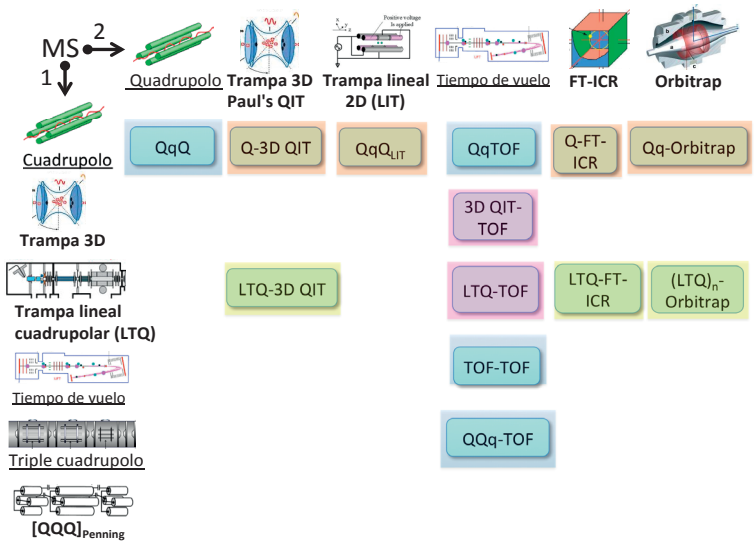


Figura 4. Configuraciones híbridas comunmente utilizadas en espectrometría de masas biomolecular en tandem (MS/MS). Para una perspectiva histórica del desarrollo de sistemas de espectrometría de masas con configuraciones híbridas, consultar la referencia [37]. Los instrumentos de tecnología de haz (beam-type) están subrayados. Las trampas de iones (capaces de realizar por sí mismas experimentos de MS^n, n>2) están resaltadas en negrita. Las configuraciones hibridas in-space/in-space, in-space/in-time, in-time/in-space e in-time/in-time entre instrumentos de tecnología de haz (beam-type) de iones y de confinamiento de iones (trapping-type) están resaltadas, respectivamente, con fondos azul (QqQ [10,92], QqTOF [93], TOF-TOF [94,95], QQqTOF (TripleTOF™) [96]), naranja-marrón (Q-3DQIT [97], QqQ_{LIT} (QTrap™) [12], Q-FTICR [98], Qq-Orbitrap (Q Exactive™) [99]), magenta (3DQIT-TOF [100], LTQ-TOF [101,102]) y amarillo-verde (LTQ-3DQIT [103], LTQ-FTICR [104], LTQ-Orbitrap [40], $(LTQ)_2$-Orbitrap (Orbitrap Velos™) [43]). Nótese que los cuadrupolos, además de utilizarse como analizadores de masas, también se implementan frecuentemente en configuraciones híbridas como focalizadores de iones, celdas de colisión (q, cuadrupolo RF, RF-only quadrupole), o trampas lineales (2D) de iones. También debido a su alta sensibilidad, preciso control de la población de iones, alta resolución, ciclos de tiempo cortos y alta capacidad de carga, se asocian trampas iónicas lineales, QqQ_{LIT} [12] y LTQ [105], a instrumentos de transformada de Fourier. Para una descripción detallada de las prestaciones del sistema híbrido trampa iónica lineal-Orbitrap (LTQ-Orbitrap) y $(LTQ)_2$-Orbitrap Velos™, consultese, respectivamente, las referencias [40,43]. Se han descrito un analizador de masas que consiste en una trampa lineal de tres cuadrupolos de rodillos cilíndricos, sometida axialmente a un fuerte campo magnético estático [106]. Mediante combinación de campos magnético y cuadrupolar eléctrico se establecen las trayectorias de los iones en la trampa, de modo análogo a una trampa de Penning [106,107]. Esta trampa de iones $[QQQ]_{Penning}$ no ha sido, sin embargo, adaptada en ninguna configuración híbrida.

sea eficaz para ser detectada en la ventana temporal de adquisición del espectro MS/MS. La máxima energía cinética de colisión (E_{cin}) que puede ser convertida en energía interna de activación (E_{act}) en una única colisión es:

$$E_{act} = E_{cin}\, m_{gas}/(m_{ion} + m_{gas}),$$

donde m_{ion} y m_{gas} son las masas del ión parental y de los átomos de gas neutro utilizados en la celda de colisión. A bajas energías, el tiempo de interacción entre el ión parental y el átomo de gas colisionante es suficiente para permitir la excitación vibracional del ión. Por el contrario, a energias de colisión de varios keV la interacción entre un ión de masa pocos cientos de Daltons y el átomo de gas colisionante es del orden de 10^{-16}-10^{-15} s. Este rango de tiempo es demasiado pequeño para que la colisión vaya acompañada de vibración (rango temporal ~10^{-14} s). Por otra parte, a baja energía las fragmentaciones en un único enlace químico dominan el espectro mientras que en espectros MS/MS de alta energía las disociaciones de fragmentos son más importantes. Además, los iones (parental y fragmentos) se dispersan tras la colisión. La baja energía de colisión empleada en una cámara de colisión cuadrupolo RF permite relocalizar los iones dispersados en el eje central del cuadrupolo, minimizando la pérdida de iones detectables. Por el contrario, a energías altas de colisión los iones dispersados más de 1º no seran observados en el espectro.

Superados los problemas técnicos, una ventaja evidente de las configuraciones híbridas es que dotan de nuevas posibilidades analíticas al sistema de espectrometría de masas. Por ejemplo, una limitación del acoplamiento de una fuente de ionización por electrospray a una trampa 3D cuadrupolar es que la mayor parte de los iones, producidos continuamente en la fuente de ionización, se pierden mientras la trampa está procesando otros iones, reduciendo pues la eficacia del ciclo de trabajo (*duty cycle*, Dc). El ciclo de trabajo de una configuración ESI-3D-IT se define como $Dc = Tc/(Tc + T_{A,3D})$, siendo Tc el tiempo de relleno de la trampa, y $T_{A,3D}$ el tiempo de análisis requerido por la trampa. Se ha calculado que la eficacia de inyección de iones es tipicamente $\leq$ 5 % de los iones producidos [39]. No es pues de extrañar que muchas de las configuraciones híbridas mostradas en la Fig.4 combinen una primera etapa de acumulación de iones con la posibilidad de realizar experimentos de MS^n y el análisis posterior de alta sensibilidad (TOF), precisión y resolución (FT-ICR, Orbitrap) de la masa de los mismos. Así, el acoplamiento de una trampa lineal cuadrupolar (LTQ) a una trampa 3D incrementa la eficiencia del ciclo de trabajo hasta cerca del 100% [11].

Una limitación de la fragmentación de un ión precursor en un analizador de masas Orbitrap es que los fragmentos adquieren la misma velocidad que el ión precursor. Y, dado que la energía es proporcional a su relación m/z, las trayectorias de iones fragmento cuya m/z sea menor al 30-50% de la m/z del ión precursor se hacen extremadamente elípticas y eventualmente se pierden. Además, los analizadores de masa basados en el confinamiento de iones (intrumentos *in-space*), como las trampas 3D, FT-ICR y Orbitrap, no tienen la capacidad de seleccionar iones utilizando información de MS/MS (ej. *parent ion scan*, *neutral loss scan*). Estos modos de escaneo son esencialmente procesos inversos de un experimento de fragmentación: En Q3 se selecciona un ión específico formado en q2, y Q1 se programa para identificar todos los iones cuya fragmentación produce el ión seleccionado en Q3. Este tipo de operaciones dependientes de datos (*data-dependent scanning*) pueden realizarse en instrumentos de haz de iones. La oportunidad de realizar estos experimentos a alta resolución supuso una de las razones del desarrollo de configuraciones híbridas que utilizan un Orbitrap como detector de masas. El primer sistema híbrido de estas características, el LTQ-Orbitrap™, fue comercializado en 2005 [40]. El confinamiento de los iones previo al análisis en el Orbitrap se realiza en la llamada “C-Trap” [41], que tiene gran capacidad de carga de iones y puede, además, ser utilizada para fragmentaciones adicionales (ej. mediante ETD, *electron transfer dissociation*) [42]. En experimentos dependientes de datos, el Orbitrap es utilizado para identificar al ión precursor, el cual será aislado en la LTQ y transferido y fragmentado en la C-trap. Los iones fragmento son detectados utili-

zando la C-trap o el Orbitrap. Recientemente, las configuraciones LTQ-Orbitrap™ han aumentado con la incorporación de la familia LTQ Velos™ (Velos™, Velos Pro™, Orbitrap Elite™) [43] (Fig.5A), que incluyen dos trampas de iones funcionando a alta y baja presión: la primara trampa confina, con eficiencia >90%, y fragmenta los iones mientras la segunda trampa realiza la detección de los fragmentos con una alta velocidad de barrido. Otra configuración del Orbitrap (Q Exactive™), en la que Q3 de un triple cuadrupolo está sustituído por un Orbitrap (Qq-Orbitrap, Fig.4), combina la rapidez de barrido de un cuadrupolo hiperbólico para la selección de iones con la alta resolución y precisión de la detección de m/z del Orbitrap (R: 140,000 a m/z 200 y <1 ppm de precisión) [44]. Aunque los iones tienen que ser transmitidos en pulsos, el tiempo de carga del Orbitrap es corto, permitiendo la realización de estudios de proteómica cuantitativa en modo MS (SIM, *single ion monitoring*) o MS/MS (PRM, *parallel reaction monitoring*) [45,46].

Al igual que en las configuraciones del Orbitrap, en el sistema híbrido LTQ-FT-ICR, la fragmentación de los iones también se produce en un espacio fisicamente separado de donde se detectan los iones fragmento. Esta separación de funciones permite conciliar la separación cromatográfica de analitos a presión atmosférica (760 Torr) con la realización de experimentos de CID-$(MS)^n$ (que requieren colisiones a una presión de 10^{-6} -10^{-7} Torr) y con la necesidad de mantener el vacío extremo (10^{-10}-10^{-9} Torr) en la trampa ICR, crucial para reducir la colisiones de iones con otras moléculas en el analizador de masas y lograr mediciones de alta resolución. Por otra parte, la limitación, impuesta por la regla de 1/3, de confinar fragmentos de masa ≥ 28% de la masa del ión parental generados mediante CID en trampas de iones (3D cuadrupolar, FT-ICR y Orbitrap) ha sido resuelta utilizando métodos alternativos de fragmentación, *Pulsed Q Collision Induced Dissociation* (PQD) o *Higher energy Collision Dissociation* (HCD), en la trampa de iones LTQ [47-50]. En el método PQD (U. S. Patent 6,949,743 B1 de Thermo Finnigan Llc) [47] se activa el ión precursor con un pulso (t_{pulse} ~ 100 ms-1ms) de q_z alto (Q1= 0.6-0.85), seguido de un tiempo de retardo (t_{delay}) para permitir la fragmentación del ión, y un decremento rápido de q_z a un valor bajo (Q2= 0.015-0.2) pero suficiente para retener todos los fragmentos confinados en la trampa (Fig.2A). A q_z (Q2) de 0.05, el límite de LMCO (*low mass cut-off*) = (0.05/0.908) x 100 = 5.5% de la masa del ión parental. Un parámetro crítico del método PQD es el tiempo de retardo entre Q1 y Q2, ya que q_z debe llevarse a su valor Q2 antes de que un número apreciable de fragmentos de bajo m/z sean expelidos de la trampa. El proceso secuencial PQD requiere una duración que es función de la amplitud del pulso de resonancia de excitación, la geometría de la trampa, la presión del gas de colisión, la amplitud del voltaje RF de confinamiento, así como de la relación m/z y fuerza de los enlaces químicos del ión seleccionado. T_{delay} está típicamente entre 45-500 ms [47]. La implementación de una trampa LTQ operada en modo PQD en las configuraciones híbridas de los analizadores FT-ICR y Orbitrap reduce el límite de masa detectable a ~ m/z 15, posibilita, por tanto, la utilización de estos sistemas híbridos para estudios de proteómica cuantitativa basada en la cuantificación de reporteros iTRAQ (m/z 113-121) [51].

Otra función implementada en algunas configuraciones híbridas es la separación de iones isobáricos por su distinta resistencia a atravesar una cámara de gas (tubo de deriva) a baja presión en un campo eléctrico de baja intensidad. La espectrometría de masas de movilidad iónica (DT-IMMS, del inglés *drift-time ion mobility mass spectrometry*) fue desarrollada en la décadas de 1950 y 1960 por Earl W. McDaniel en el Georgia Institute of Technology [52,53; consultar también el Capítulo 10]. En el tubo de deriva sometido a un campo eléctrico uniforme de baja intensidad, la movilidad de los iones puede tratarse como un fenómeno de "difusión dirigida", cuya velocidad es directamente proporcional a la intensidad del campo eléctrico. La constante de proporcionalidad (K) está relacionada con la sección transversal del ión mediante la fórmula [54]:

$$K = (3q/16N)(2\pi/kT)^{1/2}[(m+M)/mM]^{1/2}(1/\Omega),$$

siendo q la carga del ión; N, la densidad del gas; k, la constante de Boltzmann; T, la temperatura absoluta; m, la masa del gas; M la masa del ión; y Ω, la sección transversal del ión. Así, utilizan-

do DT-IMMS pueden separase iones isobáricos de conformaciones distintas. Celdas de movilidad iónica han sido acopladas a diversos tipos de analizadores de masas: TOF, cuadrupolo, trampa iónica 3D, y FT-ICR [53]. La Fig.5B muestra un esquema del sistema de espectrometría de masas Synapt™ (Waters Corporation), que incluye un cuadrupolo, tubo de separación por movilidad iónica, y un TOF [54].

El análisis por movilidad iónica representa una nueva dimensión analítica de los sistemas de espectrometría de masas, con aplicaciones en imagen por espectrometría de masas (*imaging MS*) [55], y proteómica [56]. Por ejemplo, la separación por movilidad iónica de iones "b" e "y" facilita la interpretación de espectros de fragmentación. Además, en experimentos de proteómica *top-down*, la combinación de fragmentación CID y la separación por movilidad iónica de los fragmentos en función de tamaño-carga y masa-carga facilita la asignación secuencial y la determinación de modificaciones postraduccionales [57].

3. La pregunta del millón: la elección del sistema de espectrometría de masas.

La espectrometría de masas representa una herramienta indispensable en el análisis (bio) químico de la era post-genómica. La tendencia hacia diseños de sistemas de espectrometría de masas cada vez más complejos obedece a la necesidad de lograr nuevas capacidades analíticas para investigar la complejidad espacio-temporal de los organismos vivos. Las nuevas configuraciones, así como la constante mejora de las prestaciones de los sistemas actualmente en el mercado, incide en el coste de los instrumentos y plantea, pues, la importante cuestión de la elección del sistema de espectrometría de masas más adecuado. Desgraciadamente, no existe la máquina ideal y, por tanto, la decisión de cual adquirir debe estar basada en las necesidades de la investigación a realizar. Así, las configuraciones basadas en triple cuadrupolos son especialmente adecuadas para proteómica cuantitativa, al poder operar en modos de monitorización selectiva de iones (SRM, *selected reaction monitoring*; MRM, *multiple reaction monitoring*) [58,59], así como para la identificación de modificaciones post-traduccionales, trabajando en modo *parent-ion scan*. Las configuraciones con trampa iónica 2D sustituyendo a Q3 realizan las funciones clásicas de un triple cuadrupolo y una etapa adicional de MS/MS posterior a la fragmentación en q2. Por su parte, los instrumentos de tiempo de vuelo (TOF) no son instrumentos de barrido: el espectro de masas se genera por medición cuasi-simultánea de los tiempos de vuelo de los iones generados por pulsos de radiación láser. La configuración TOF-TOF combina la detección de iones en TOF_1 con el análisis en TOF_2 de la fragmentación de alta energía de iones seleccionados en TOF_1. El alto rango dinámico y precisión en la medida de m/z (Tabla 1), hace que la configuración TOF-TOF sea especialmente útil para la identificación, rápida y automatizada, de proteínas de organismos cuya secuencia genomica ha sido determinada, mediante combinación de huella peptídica y CID-MS/MS [60; consultar también el Capítulo 19]. Asímismo, la sustitución de un cuadrupolo (Q3) del triple cuadupolo, de gran sensibilidad pero cuya precisión es relativamente baja (Tabla 1), por un TOF de mucha mejor precisión y resolución pero menos sensibilidad (Tabla 1), dota a la configuración resultante (QqTOF, Fig.4) de precisión < 5 ppm sin sacrificio de la sensibilidad en el rango de atomoles. Estas características son de gran importancia en análisis de proteómica dirigida, donde la selección de las transiciones en un fondo peptídico complejo requiere la máxima precisión posible [61].

Las configuraciones que incluyen espectrómetros de transformada de Fourier (FT-ICR y Orbitrap) (Fig.4) ostentan el record de resolución (Table 1). La capacidad de medir con alta precisión la masa de iones precursores y de sus fragmentos de MS/MS aumenta las posibilidades de identificar correcta- y automaticamente las proteínas precursoras utilizando espectros MS/

MS y búsqueda en un banco de datos que incluya la secuencia del genoma de la especie investigada. Por ejemplo, las masas de los aminoácidos lisina (K) y glutamina (Q) difieren en 0.03639 Da. Para distinguir dos péptidos de 1000 Da cuyas secuencias difieren unicamente en un cambio K/Q se requiere una precisión de 36 ppm. La misma sustitución en péptidos de 5000 Da requiere una precisión de 7 ppm. Las trampas FT-ICR requieren imanes superconductores que, al trabajar as temperature de He líquido, son costosos de mantener. El alto poder resolutivo del Orbitrap (Tabla 1), aunque inferior al de FT-ICR, combina gran capacidad de almacenamiento de iones, un rango de m/z hasta ~6000, y un rango dinámico de 10^3, representa una excelente opción para proteómica de alto rendimiento (*high-throughput*) y de alta resolución [62, y Capítulo 11] de organismos con genoma conocido. La proteómica de organismos cuyos genomas no han sido secuenciados requiere un alto grado de secuenciación *de novo* y, por tanto, la calidad de los espectros de fragmentación (presencia de series completas de iones específicos de secuencia, "b" e "y", así como la presencia de iones imonio que informan sobre la composición de aminoácidos del ión precursor) [63] (ver también el Capítulo 18) es un factor principal en la selección del sistema de espectrometría de masas más adecuado.

Además de las prestaciones típicas de los diversos espectrómetros y configuraciones de analizadores de masas, los principales parámetros instrumentales que deben tenerse en cuenta son [64,65]:

i) Precisión. Expresado en partes por millón (ppm), la precisión es igual a [(m/z(exp) - m/z(-theor)) / m/z(theor)] x 10^6) [66];
ii) Resolución. (m/z)/Δ(m/z)$_{50\%}$, masa del ión dividido por la anchura a mitad altura del pico;
iii) Rango inferior y superior de medida de m/z;
iv) Rango dinámico. El rango de m/z en el que la masa de los iones es directamente proporcional a la concentración del analito;
v) Capacidad de realización de experimentos MS/MS o $(MS)^n$;
vi) Sensibilidad. Máxima intensidad de señal m/z respecto al ruido de fondo a (m/z)+1;
vii) Velocidad de adquisición de espectros (*duty cycle*).
viii) Incorporación de modos adicionales de fragmentación de iones (CID, ECD, (*electron capture dissociation*), ETD).
ix) Presupuesto....

La Tabla 1 muestra una comparación no exhaustiva de los parámetros i-vii de algunas de las configuraciones más comunmente utilizadas en proteómica.

Table 1. Figuras de mérito de algunas configuraciones de espectrometría de masas utilizadas en proteómica. Los valores listados son aproximados y pueden variar en diferentes configuraciones híbrida de diferentes frabicantes. Consúltese las especificaciones técnicas de los instrumentos en cuestión. R, modo reflectrón. FWHM, *full width at half maximum*, máxima anchura a mitad altura de pico.

	QQQ	3D-QIT	QQ-qLIT	TOF-TOF	FT-ICR	LTQ-Orbitrap
Rango de m/z	4000	4000	4000	ilimitado 4500 (R)	4000	6000
Precisión	100 ppm	100 ppm	100 ppm	5-10 ppm	<1 ppm	2-5 ppm
Resolución (FWHM)	≥4000	≥4000	≥4000	~25000 (R)	≥1x10^6	≥150000
Rango dinámico	1x10^7	1x10^3	1x10^4	1x10^6	1x10^3	1x10^4
$(MS)^n$	$(MS)^2$	$(MS)^n$	$(MS)^3$	$(MS)^2$	$(MS)^n$	$(MS)^n$
Sensibilidad	atto-fmol	picomol	femtomol	femtomol	1x104	1x104
Velocidad adquisición	~1s	~1s	4/s	200/s	~1s	~1s

4. Comentarios finales: ¿Biología en fase gaseosa?

La creciente implicación de la espectrometría de masas en el análisis biológico, y el consiguiente aumento exponencial del volúmen de negocios de la industria biotecnológica y de las empresas constructores de analizadores de masas, han catalizado la aparición en el mercado de nuevas configuraciones de espectrometría de masas y el desarrollo de diversos métodos de cuantificación de proteínas y proteomas. Aunque ciertamente el nuevo siglo ha sido testigo de innovaciones tecnológicas, la tendencia hacia configuraciones híbridas de analizadores de masas ha dominado los últimos avances en instrumentación [37,38,67]. Los nuevos sistemas de espectrometría de masas representan diversas combinaciones de espectrómetros de tecnología de haz y de tecnología de confinamiento de iones, solventando así limitaciones inherentes a algunos analizadores de masas y dotando al híbrido de capacidades analíticas ausentes en cada uno de los espectrómetros utilizados autonomamente. Las nuevas prestaciones de los sistemas híbridos incluyen ciclos de trabajo más rápidos, mayor sensibilidad y rango de medida, y mejor resolución, precisión y rango dinámico en la determinación de m/z. Estos parámetros son importantes en el análisis de los sistemas biológicos. Sin embargo, otro parámetro esencial para comprender los mecanismos de los procesos vitales, la cuantificación absoluta de los componentes del sistema biológico a lo largo del espacio y del tiempo, sigue eludiendo el análisis espectrométrico. Los métodos de cuantificación utilizados actualmente en estudios proteómicos, tanto si utilizan marcajes isotópicos (O^{16}/O^{18}, ICAT, iTRAQ, SILAC) o contaje de iones (*label-free, spectral counting*) proporcionan estimaciones de la abundancia relativa de los componentes [68-71]. Sin embargo, la comprensión de fenómenos dinámicos, como los mediados por modificaciónes post-traduccionales (PTM, *post-translational modifications*) precisa no solo la identificación de los aminoácidos modificados sino también la estequiometría de la modificación y la cuantificación absoluta de las proteínas modificadas. Ello requiere un combinación de proteómica dirigida (*targeted proteomics*) [58,59] (utilizando estrategias de dilución isotópica de péptidos sintéticos prototípicos marcados isotópicamente como estándares internos y protocolos de cuantificación dirigida (SRM, MRM o PRM, *parallel reaction monitoring*) [72-74]), y análisis proteómico *Top-Down* de las proteoformas presentes en un instante dado en el sistema investigado [75-80]. Tras más de 20 años de tímida biología estructural en fase gaseosa [81], la proteómica Top-Down está ganando adeptos: en 2012 se constituyó el Consortium for Top Down Proteomics (http://www.topdownproteomics.org), aunque las publicaciones en este campo son todavía escasas, particularmente porque la instrumentación necesaria para el aislamiento y la fragmentación dirigida de proteoformas y complejos multiproteicos intactos, así como las herramientas computacionales para la interpretación de los complejos espectros resultantes, son en gran medida experimentales y accesible solo en determinados laboratorios [82-85]. Además, la cuantificación directa y absoluta de proteínas intactas mediante la tecnología actual de espectrometría de masas resulta imposible debido a la altamente variable ionización, transmisión, y eficiencia de detección de los iones [86]. No obstante, recientes avances en instrumentación y nuevos métodos de fragmentación de iones que emplean fotones y electrones están permitiendo la caracterización completa de proteínas que no era posible hace sólo unos pocos años [85]. También el desarrollo de nuevos métodos de ionización de complejos protéicos nativos prometen la identificación y caracterización de proteínas en el rango de 100 kDa a 1 MDa mediante espectrometría de masas [87-89].

La proteómica del siglo XX puede definirse como esencialmente cualitativa. Los resultados de un análisis proteómico típico podían resumirse en largas tablas de identificaciones de las proteínas del sistema analizado pero carentes de cualquier información sobre su abundacia, estequiometría o distribución. Por contra, en la misma época el análisis de la expresión génica se realizaba mediante estrategias cuantitativas (microarrays, PCR cuantitativa). Sin embargo, la abundancia de las proteínas está regulada, en gran medida, por procesos post-transcripcionales y post-translacionales que escapan al análisis genómico y transcriptómico. La proteómica cuantitativa representa conceptualmente y tecnicamente la clave para, analizando la complejidad del proteoma,

establecer el vínculo entre el genotipo y el fenotipo. La incorporación de nuevos avances en las configuraciones de analizadores de masas abrirá sin duda nuevas perspectivas para el análisis de la composición de proteínas y complejos multiproteicos endógenos en los próximos años.

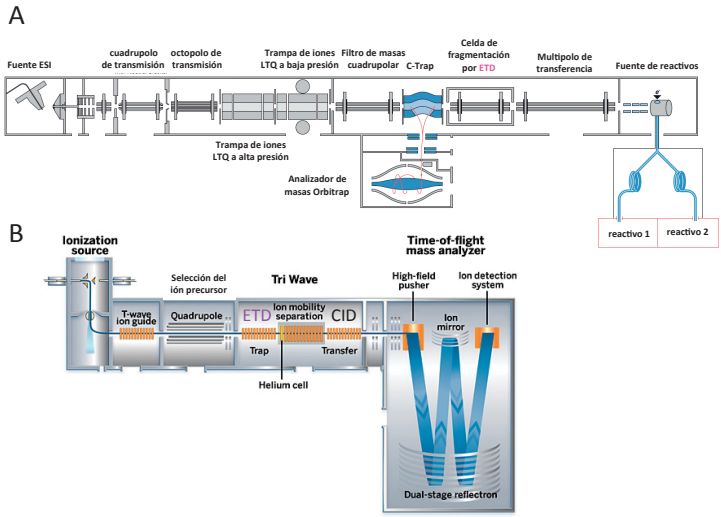


Figura 5. Esquema de los sistemas de espectrometría de masas híbridos LTQ Velos™ (A) y Synapt™ G2 (B). ETD, cámara de fragmentación de un ión precursor por disociación por transferencia de electron (electron-transfer dissociation); CID (collision-induced dissociation), cámara de fragmentación de un ión precursor por disociación inducida por colisión con moléculas de gas inerte.

REFERENCIAS.

1. Thomson, J.J. Cathode rays. *Philosophical Magazine*, 1897, 44, 293-316 (consultar también: http://www-outreach.phy.cam.ac.uk/camphy/electron/electron_index.htm).

2. Wien, K. 100 years of ion beams: Willy Wien's canal rays. *Br J Physics* 1999, 29, 401-414.

3. Falconer, I. Corpuscles to Electrons. In: Buchwald J, Warwick A (eds) *Histories of the Electron*, MIT Press, Cambridge (MA) 2001, pp. 77-100.

4. Miller, P.E., Denton, M.B. The quadrupole mass filter: basic operating concepts. *J Chem Education* 1986, 63, 617-622.

5. Dawson, P.H. Quadrupole mass analyzers: performance, design and some recent applications. *Mass Spectrom Rev* 1986, 5, 1-37.

6. Miller, P.E., Denton, M.B. The quadrupole mass filter: basic operating concepts. *J Chem Education* 1986, 63, 617-622.

7. Paul, W. Electromagnetic Traps for charged and neutral particles. *Rev Mod Phys* 1990, 62, 531-540.

8. Cooks, R.G., Glish, G.L., McLuckey, S.A., et al. Ion trap mass spectrometry. *Chem Eng News* 1991, 69, 26-30.

9. Jonscher, K.R., Yates, J.R. The quadrupole ion trap mass spectrometer-A small solution to a big challenge. *Anal Biochem* 1997, 244, 1-15.

10. Yost, R.A., Enke, C.G. Selected ion fragmentation with a tandem quadrupole mass spectrometer. *J Am Chem Soc* 1978, 100, 2274-2275.

11. Douglas. D.J., Frank, A.J., Mao, D. Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.* 2005, **24**, 1-29.

12. Hager, J.W. A new linear ion trap mass spectrometer. *Rapid Commun. Mass Spectrom.* 2002, 16, 512-526.

13. Stephens, W.E. A Pulsed Mass Spectrometer with Time Dispersion. *Phys. Rev.* 1946, 69, 691.

14. Campana, J.E. Time-of-Flight mass spectrometry: a historical overview. *Instrum. Sci. Technol.* 1987, 16, 1-14.

15. Comisarow. M.B., Marshall, A.G. The early development of Fourier transform ion cyclotron resonance (FT-ICR) spectroscopy. *J. Mass Spectrom.* 1996, 31, 581-585.

16. Marshall, A.G., Hendrickson, C.L., Jackson, G.S. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 1998, 17, 1-35.

17. Lawrence, E. The evolution of the cyclotron. (Nobel lecture December 11, 1951: http://www.nobelprize.org/nobel_prizes/physics/laureates/1939/lawrence-lecture.html)

18. Dehmelt,, H.G. (1989) Experiments with an isolated subatomic particle at rest. Nobel Lecture, December 8, 1989. (http://www.nobelprize.org/nobel_prizes/physics/laureates/1989/dehmelt-lecture.pdf).

19. Marshall, A.G., Hendrickson, C.L. FTICR detection: principles and experimental configurations. Int J Mass Spectrom 2002, 215, 59-75.

20. Fellgett, P. A propos de la theorie du spectrometre interferentielk multiplex. J Phys Radium 1958, 19, 187-191.

21. Fellgett, P.B. *Theory of Infra-Red Sensitivities and its Application to Investigations of Stellar Radiation in the Near Infra-Red* (PhD thesis), 1949.

22. Fellgett, P.B. The nature and origin of multiplex Fourier spectrometry. *Notes Rec. R. Soc*. 2006, 60, 91-93.

23. Marshall, A.G., Hendrickson, C.L. Charge reduction lowers mass resolving power for isotopically resolved electrospray ionization Fourier transform ion cyclotron resonance mass spectra. *Rapid Commun. Mass Spectrom.* 2001, 15, 232-235.

24. Marshall, A.G. Theoretical signal-to-noise ratio and mass resolution in Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* 1979, 51, 1710-1714.

25. Marshall, A.G., Comisarow, M.B., Parisod, G. Relaxation and spectral line shape in Fourier transform ion resonance spectroscopy. *J. Chem. Phys.* 1979, 71, 4434-4444.

26. Vartanian, V.H., Laude, D.A. Optimization of ion cell for FT-ICR MS. *Int J Mass Spectrom. Ion Proc.* 1995; 141, 189-200.

27. Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal. Chem.* 2000, 72, 1156-1162.

28. Kingdon, K.H. A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Phys. Rev.* 1923, 21, 408-418.

29. Knight, R.D. Storage of ions from laser-produced plasmas. *Applied Phys. Lett.* 1981, 38, 221-223.

30. Zubarev, R.A., Makarov, A. Orbitrap mass spectrometry. *Anal. Chem.* 2013, 85, 5288-5296.

31. Karas, M., Bachmann, D., Hillenkamp, F. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal. Chem.* 1985, 57, 2935-2939.

32. Karas, M., Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* 1988, **60**, 2299-2301.

33. Tanaka, K., Waki, H., Ido, Y., et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 1988, 2, 151-153.

34. Tanaka, K. The origin of macromolecule ionization by laser irradiation. Nobel Lecture December 8, 2002 (http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2002/tanaka-lecture.pdf).

35. Fenn, J.B., Mann, M., Meng, C.K., et al. Electrospray ionization for mass spectrometry of large biomolecules. Science 1989, 246, 64-71.

36. Fenn, J.B. Electrospray wing for molecular elephants. Angew. Chem. Int. Ed. Engl. 2003, 42, 3871-3894. Nobel Lecture December 8, 2002 (http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2002/fenn-lecture.pdf).

37. Glish, G.L., Burinsky, D.J. Hybrid mass spectrometers for tandem mass spectrometry. Am. Soc. Mass Spectrom. 2008, 19, 161-172.

38. de Hoffmann, E. Tandem mass spectrometry: a primer. *J. Mass Spectrom.* 1996, 31, 129-137.

39. Le Blanc, J.C., Hager, J.W., Ilisiu, A.M., et al. Unique scanning capabilities of a new hybrid linear ion trap mass spectrometer (QTRAP) used for high sensitivity proteomics applications. *Proteomics* 2003, 3, 859-869.

40. Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., Horning, S. Performance evaluation of a hybrid linear ion trap/Orbitrap mass spectrometer. *Anal. Chem.* 2006, 78, 2113-2120.

41. Makarov, A., Horning, S., Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 2007, 4, 709-712.

42. Syka, J.E., Coon, J.J., Schroeder, M.J., et al. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. USA* 2004, 101, 9528-9533.

43. Olsen, J.V., Schwartz, J.C., Griep-Raming, J., Nielsen, M.L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E.R., Senko, M., Makarov, A., Mann, M., Horning, S. A new instrument for high-speed proteomics: Orbitrap mass analyzer interfaced to a dual linear trap. *Mol Cell. Proteomics* 2009, 8, 2759-2769.

44. Geiger, T., Cox, J., Mann, M. Proteomics on an Orbitrap benchtop mass spectrometer using allion fragmentation. *Mol Cell Proteomics 2010,* 9, 2252 2261.

45. Gallien, S., Duriez, E., Crone. C., et al. Targeted proteomic quantification on quadrupole-Orbitrap mass spectrometer. *Mol. Cell. Proteomics* 2012, 11, 1709-1723.

46. Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J., Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics* 2012, 11, 1475-1488.

47. Schwartz, J.C. U. S. Patent 6,949,743 B1, 2004 (http://www.google.com/patents/US6949743)

48. Schwartz, J. C., Syka, J.E.P., Quarmby, S.T. The 53rd ASMS Conference on Mass Spectrometry and Allied Topics, San Antonio, TX, June 5-9, 2005.

49. Bantscheff, M., Boesche, M., Eberhard, D., Matthieson, T., Sweetman, G., Kuster, B. Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. Mol. Cell Proteomics 2008, 7, 1702-1713.

50. Lössner, C., Blackstock, W., Gunaratne, J. Enhanced performance of pulsed Q collision induced dissociation-based peptide identification on a dual-pressure linear ion trap. J. Am. Soc. Mass Spectrom. 2012, 23, 186-189.

51. Przybylski, C., Jünger, M.A., Aubertin, J., Radvanyi, F., Aebersold, R., Pflieger, D. Quantitative analysis of protein complex constituents and their phosphorylation states on a LTQ-Orbitrap instrument. J. Proteome Res. 2010, 9, 5118-5132.

52. Shvartsburg, A.A., Smith, R.D. Fundamentals of traveling wave ion mobility spectrometry. Anal. Chem. 2008, 80, 9689-9699.

53. Kanu, A.B., Dwivedi, P., Tam, M., et al. Ion mobility-mass spectrometry. J. Mass Spectrom. 2008, 43, 1-22.

54. Pringle, S.D., Giles, K., Wildgoose, J.L., et al. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* 2007, 261, 1-12.

55. Kiss, A., Heeren, R.M.A. Size, weight and position: ion mobility spectrometry and imaging MS combined. *Anal. Bioanal. Chem.* 2011, 399, 2623-2634.

56. McLean, J.A., Ruotolo, B.T., Gillig, K.J., et al. Ion mobility mass spectrometry: a new paradigm for proteomics. *Int. J. Mass Spectrom.* 2005, 240, 301-315.

57. Zinnel, .N.F, Pai, P-J., Russell, D.H. Ion mobility-mass spectrometry (IM-MS) for top-down proteomics: increased dynamic range affords increased sequence coverage. *Anal. Chem.* 2012, 84, 3390-3397.

58. Picotti, P., Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* 2012, 9, 555-566.

59. Carapito, C., Aebersold, R. (Guest editors). Targeted Proteomics. *Proteomics* 2012, 8, 1073-1300.

60. Clauser, K.R., Baker, P., Burlingame, A.L. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* 1999, 71, 2871-2882.

61. Ramanathan, R., Jemal, M., Ramagiri, S., Xia, Y.Q., Humpreys, W.G., Olah, T., Korfmacher, W.A. It is time for a paradigm shift in drug discovery bioanalysis: from SRM to HRMS. *J. Mass Spectrom*. 2011, 46, 595-601.

62. Perry, R.H., Cooks, R.G., Noll, R.J. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrometry Reviews*, 2008, 27, 661-699.

63. Frank, A.M., Savitski, M.M., Nielsen, M.L., Zubarev, R.A., Pevzber, P.A. De novo peptide sequencing and identification with precision mass spectrometry. *J. Proteome Res.* 2007, 6, 114-123.

64. Han, X., Aslanian. A., Yates, III J.R. Mass spectrometry for proteomics. *Curr. Op. Chem. Biol.* 2008, **12**, 483–490.

65. Hart-Smith, G., Blanksby. S.J. Mass analysis. En Barner-Kowollik C, Gruendling T, Falkenhagen J, Weidner S (eds.) *Mass Spectrometry in Polymer Chemistry*, 1st edn. Wiley-VCH Verlag GmbH & Co., 2012, pp 5-32.

66. Brenton, A.G., Godfrey, A.R. Accurate mass measurement: terminology and treatment of data. *J. Am. Soc. Mass Spectrom.* 2010, 21, 1821-1835.

67. Hager, J.W. Recent trends in mass spectrometer development. *Anal. Bioanal. Chem*. 2004, 378, 845-850.

68. Schulze, W.X., Usadel, B. Quantitation in mass-spectrometry-based proteomics. *Annu. Rev. Plant Biol.* 2010, 61, 491-516.

69. Kannaste, O., Suomi, T., Salmi, J., Uusipaikka, E., Nevalainen, O., Corthals, G.L. Cross-correlation of spectral count ranking to validate quantitative proteome measurements. *J. Proteome Res*. 2014, Mar 10. [Epub ahead of print].

70. Gerster, S., Kwon, T., Ludwig, C., Matondo, M., Vogel, C., Marcotte, E.M., Aebersold, R., Bühlmann, P. Statistical approach to protein quantification. *Mol. Cell Proteomics* 2014, 13, 666-677.

71. Chang, C.Y., Sabidó, E., Aebersold, R., Vitek, O. Targeted protein quantification using sparse reference labeling. *Nature Methods* 2014, 11, 301-314.

72. Villanueva, J., Carrascal, M., Abian, J. Isotope dilution mass spectrometry for absolute quantification in proteomics: concepts and strategies. *J. Proteomics* 2014, 96, 184-199.

73. Brun, V., Masselon, C., Garin, J., Dupuis, A. Isotope dilution strategies for absolute quantitative proteomics. *J. Proteomics* 2009, 72, 740-749.

74. Gallien, S., Bourmaud, A., Kim, S.Y., Domon, B. Technical considerations for large-scale parallel reaction monitoring analysis. *J. Proteomics* 2013 Nov 4, doi: 10.1016/j.jprot.2013.10.029. [Epub ahead of print].

75. Kelleher, N.L. Top-down proteomics. *Anal. Chem.* 2004, 76, 197A–203A.

76. Gault, J., Malosse, C., Machata, S., Millien, C., Podglajen, I., Ploy, M.C., Costello, C.E., Duménil, G., Chamot-Rooke, J. Complete posttranslational modification mapping of pathogenic *Neisseria meningitidis* pilins requires top-down mass spectrometry. *Proteomics* 2014 Jan 23. doi: 10.1002/pmic.201300394. [Epub ahead of print]

77. Peng, Y., Chen, X., Zhang, H., Xu, Q., Hacker, T.A., Ge, Y. Top-down targeted proteomics for deep sequencing of tropomyosin isoforms. *J. Proteome Res*. 2013, 12, 187-198.

78. Smith, L.M., Kelleher, N.L. & The Consortium for Top Down Proteomics. Proteoform: a simple term describing protein complexity. *Nature Methods* 2013, 10, 186-187.

79. Lamond, A.I., Uhlen, M., Horning, S., Makarov, A., Robinson, C.V., Serrano, L., Hartl, F.U., Baumeister, W., Werenskiold, A.K., Andersen, J.S., Vorm, O., Linial, M., Aebersold, R., Mann, M. Advancing cell biology through proteomics in space and time (PROSPECTS). *Mol. Cell. Proteomics* 2012, 11, O112.017731.

80. Ahmad, Y., Boisvert, F.M., Lundberg, E., Uhlen, M., Lamond, A.I. Systematic analysis of protein pools, isoforms, and modifications affecting turnover and subcellular localization. *Mol. Cell. Proteomics* 2012, 11, M111.013680.

81. Marcoux, J., Robinson, C.V. Twenty years of gas phase structural biology. *Structure* 2013, 21, 1541-1550.

82. Zhang, H., Cui, W., Gross, M.L. Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies. *FEBS Lett*. 2014, 588*,* 308-317.

83. Li, H., Wolff, J.J., Van Orden, S.L., Loo, J.A. Native top-down electrospray ionization-mass spectrometry of 158 kDa protein complex by high-resolution Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* 2014, 86, 317-320.

84. Liu, X., Hengel, S., Wu, S., Toli , N., Pasa-Toli , L., Pevzner, P.A. Identification of ultramodified proteins using top-down tandem mass spectra. *J. Proteome Res.* 2013, 12, 5830-5838.

85. Rose, C.M., Russell, J.D., Ledvina, A.R., McAlister, G.C., Westphall, M.S., Griep-Raming, J., Schwartz, J.C., Coon, J.J., Syka, J.E. Multipurpose dissociation cell for enhanced ETD of intact protein species. *J. Am. Soc. Mass Spectrom*. 2013, 24, 816-827.

86. Russell, J.D., Scalf, M., Book, A.J., Ladror, D.T., Vierstra, R.D., Smith, L.M., Coon, J.J. Characterization and quantification of intact 26S proteasome proteins by real-time measurement of intrinsic fluorescence prior to top-down mass spectrometry. *PLoS One* 2013, 8, e58157.

87. Schmidt, C., Robinson, C.V. Dynamic protein ligand interactions - insights from MS. *FEBS J*. 2014 Jan 2. doi: 10.1111/febs.12707. [Epub ahead of print].

88. Politis, A., Stengel, F., Hall, Z., Hernández, H., Leitner, A., Walzthoeni, T., Robinson, C.V., Aebersold, R. A mass spectrometry-based hybrid method for structural modeling of protein complexes. *Nature Methods* 2014 Feb 9. doi: 10.1038/nmeth.2841. [Epub ahead of print].

89. Catherman, A.D., Skinner, O.S., Kelleher, N.L. Top Down proteomics: Facts and perspectives. *Biochem. Biophys. Res. Commun*. 2014, Feb 17 [Epub ahead of print] doi: 10.1016/j.bbrc.2014.02.041.

90. Nappi, M., Weil, C., Cleven, C.D., et al. Visual representations of simulated three-dimensional ion trajectories in an ion trap mass spectrometer. *Int J Mass Spectrom Ion Proc*. 1997, 161, 77-85.

91. Schweikhard, L., Alber, G.M., Marshall, A.G. FT/ICR of highly-charged atomic ions. *Phys. Scrip.* 1992, 46, 598-602.

92. Yost, R.A., Enke, C.G. Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Anal. Chem*. 1979, 51, 1251-1264.

93. Glish, G.L., Goeringer, D.E. Tandem quadrupole-time-of-flight instrument for mass-spectrometry/mass-spectrometry. *Anal. Chem*. 1984, 56, 2291–2295.

94. Schey, K., Cooks, R.G. A tandem time-of-flight mass spectrometer for surface-induced dissociation. *Int. J. mass Spectrom. Ion Processes* 1987, 77, 49-61.

95. Medzihradszky, K.F., Campbell, J.M., Baldwin, M.A, Falick, A.M., Juhasz, P., Vestal, M.L., Burlingame, A.L. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* 2000, 72, 552–558.

96. Andrews, G.L., Simons, B.L., Young, J.B., Hawkridge, A.M., Muddiman, D.C. Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal. Chem.* 2011, 83, 5442-5446.

97. Morand, K., Horing, S.R., Cooks, R.G. A tandem quadrupole-ion trap mass spectrometer. *Int. J. Mass Spectrom. Ion Processes* 1991, 105, 13-29.

98. Patrie, S.M., Charlebois, J.P., Kelleher, N.L., Hendrickson, C.L., Quinn, J.P., Marshall, A.G., Mukhopadhyay, B. Construction of a hybrid quadrupole/Fourier transform ion cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa. *J. Am. Soc. Mass Spectrom.* 2004, 15, 1099-1108.

99. Michalski, A., Damoc, E., Hauschild, J.P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., Horning, S. Mass spectrometry-based proteomics using a Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mo. Cell Proteomics 2011, 10, M111.011015.

100. Fountain, S.T., Lee, H., Lubman, D.M. Ion fragmentation activated by matrix-assisted laser desorption/ionization in an ion-trap/reflectron time-of-flight device. *Rapid Commun. Mass Spectrom.* 1994, 8, 407-416.

101. Campbell, J. M., Collings, B. A., Douglas, D. J. A new linear ion trap time-of-flight system with tandem mass spectrometry capabilities. *Rapid Commun. Mass Spectrom.* 1998, 12, 1463-1474.

102. Hashimoto, Y., Hasegawa, H., Waki, I. Dual linear ion trap/orthogonal acceleration time-of-flight mass spectrometer with improved precursor ion selectivity. *Rapid Commun. Mass Spectrom.* 2005, 19, 1485-1491.

103. Cha, B., Blades, M., Douglas, D.J. An interface with a linear quadrupole ion guide for an electrospray-ion trap mass spectrometer system. *Anal. Chem.* 2000, 72, 5647-5654.

104. Syka, J.E., Marto, J.A., Bai, D.L, et al. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* 2004, **3**, 621-626.

105. Schwartz, J.C., Senko, M.W., Syka, J.E. A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom*. 2002, 13, 659-669.

106. Huang, Y., Li, G-Z., Guan, S., Marshall, A.G. A combined linear ion trap for mass spectrometry. *J. Am. Soc. Mass Spectrom*. 1997, 8, 962-969.

107. Van Eijkelenborg, M.A., Storkey, M.E.M., Segal, D.M., Thompson, R.C. Ion dynamics in a novel linear combined trap. *Int. J. Mass Spectrom*. 1999, 188, 155-161.

HPLC-Chip/MS. Flexible y robusta tecnología microfluídica para nanocromatografía/MS

Isidre Masana González[1]

[1] Especialista Productos LC/MS. Agilent Technologies – Barcelona.

1. Introducción.

La necesidad de poder detectar proteínas y péptidos a muy bajas concentraciones, muchas de ellas por debajo de los límites de detección de los instrumentos actuales, fuerza a muchos laboratorios a trabajar con nano-cromatografía y nanoelectrospray, con columnas de típicamente 75 µm de diámetro y unos 200-600nL/min de flujo, a pesar de sus inconvenientes en cuanto a robustez, lentitud y necesidad de ser utilizada por usuarios con apreciable experiencia. Por esa razón, algunos laboratorios prefieren sacrificar algo de sensibilidad y trabajar con cromatografía capilar; utilizando columnas de 180-320µm y trabajando en el rango de µL/min con una fuente electrospray con micronebulizador. Ver tabla 1 para flujos y mejoras teóricas en sensibilidad.

Para solventar las dificultades de la nanocromatografía, en los últimos años se han ido introduciendo en el mercado, distintos dispositivos microfluídicos que abordan la resolución de todas, o parte de estas dificultades. Algunos de estos desarrollos resuelven la corta vida de las puntas del nanoelectrospray, mediante un intercambiador automático, otros, abordan la parte cromatográfica con una nanoválvula que permite interconectar varios dispositivos microfluídicos; como una pre-columna concentradora y la nanocolumna, pero utilizan un nanospray externo. Otros incorporan una punta emisora intercambiable de sílice o acero inoxidable. En este capítulo

Tabla 1. Muestra la relación entre el diámetro interno de la columna, el inverso de su sección relativa, capacidad de carga y flujo equivalente, tomando como referencia las columnas de 2.1mm (típicas con electrospray). La segunda columna (inverso de la sección), muestra la "teórica" mejora en sensibilidad que se puede llegar a obtener; a la vez, coincide con la reducción de la capacidad de carga total de la columna. La tercera columna indica la cantidad total de muestra que la columna es capaz de separar, manteniendo una buena separación cromatográfica, sin que la fase estacionaria se sature. La última columna indica el flujo típico (rango habitual entre el 50 y 200% de estos valores).

Diámetro Interno Columna	Inverso Sección relativa (potencial mejora sensibilidad y reducción capacidad de carga)	Capacidad de Carga Columna L=15 cm longitud	Flujo Equivalente (velocidad lineal equivalente)
4.6 mm	0,2	5 mg	1,44 mL/min
3 mm	0,5	2 mg	612 µL/min
2.1 mm	**1**	**1 mg**	**300 µL/min**
1 mm	4,4	230 µg	68 µL/min
500 µm	18	60 µg	17 µL/min
300 µm	49	20 µg	6 µL/min
100 µm	441	2µg	681 nL/min
75 µm	784	1 µg	383 nL/min

se describirá el primer dispositivo microfluídico que integra en una única pieza, todos los componentes citados, y que recibe el nombre de HPLC-Chip/MS.

La tecnología HPLC/Chip es una tecnología microfluídica, desarrollada a principios de este siglo, para proporcionar la sensibilidad de la nanocromatografía con columnas de 75µm y nanoelectrospray, pero sin sus inconvenientes. Con respecto a la nanocromatografía convencional, mejora muy considerablemente su robustez, sensibilidad y flexibilidad en la automatización de procesos de tratamiento "online" de la muestra. Con respecto a la cromatografía-masas con columnas de 2.1mm y fuentes electrospray convencionales, aporta una mejora en sensibilidad muy importante (ver figura 1); típicamente de más de 100 veces. Las primeras publicaciones sobre el HPLC-Chip aparecieron en el 2004-2005 (1-3) y desde entonces han aumentando enormemente.

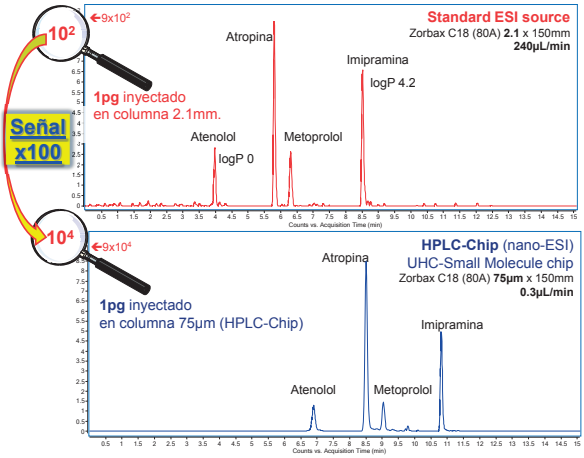


Figura 1. Comparación de sensibilidad entre el HPLC-Chip con nanocolumna 75µm d.i. (figura inferior), y LC-ESI con columna 2.1mm d.i. (figura superior), ambas de 150mm de longitud. En la figura inferior con el HPLC-Chip, la intensidad de señal es 100 veces superior y no se llega a apreciar ruido en la línea de base; se utilizó la misma cantidad de muestra absoluta (1 pg inyectado) y el mismo MS-QqQ.

2. Fundamentos y Características de la tecnología HPLC-Chip/MS.

La tecnología se basa en la creación microcanales en unas láminas de un polímero biocompatible muy inerte (poliimida), y resistente a temperaturas elevadas. Estos se crean mediante ablación por laser. Una vez realizados los microcanales en las diversas láminas que se requieran, éstas se unen mediante un proceso de laminación. La punta del nanoelectrospray está construida del mismo material, al que se le añade un recubrimiento metálico para poderle aplicar el campo eléctrico necesario para producir la pluma de nebulización. Para utilizarlo, el chip se introduce en un dispositivo denominado "Chip Cube", que establece automáticamente todas las conexiones fluídicas con el nano-HPLC, encargado de suministrar los eluyentes y la muestra. El Chip quedará presionado, alineado y herméticamente cerrado con los orificios y superficie que incorpora la nanoválvula del "Chip Cube". Este diseño permite utilizar Chips con diferente número de láminas (/grosor), según la aplicación a utilizar.

El diseño del HPLC-Chip, permite reducir al máximo los volúmenes muertos entre los distintos componentes, y proporciona una enorme flexibilidad al admitir una gran variedad de diseños. Ello incluye diseños a medida del usuario, para automatizar en línea procesos de preparación y medida de la muestra.

La facilidad de operación facilita enormemente el intercambio de Chips, y permite que cada grupo de trabajo/proyecto tenga los suyos específicos/propios. Esta facilidad, contrasta con la fragilidad y delicadez de las conexiones con capilares de sílice fundida, utilizados en nanocromatografía, y que dificultan el cambio continuado de columna y configuraciones instrumentales.

2.1 Componentes.

El sistema HPLC-Chip, a parte del nanocromatógrafo, consta de 3 componentes, fuente, Chip Cube y Chips que a continuación se describen:

2.1.1. Fuente nano-electrospray.

El HPLC-Chip incorpora una típica fuente nanoelectropray con nebulización ortogonal (ángulo de 90º con respecto a la dirección de entrada de los iones al masas), modificada para admitir la inserción en su interior del Chip. En ella, la nebulización de la muestra tiene lugar mediante la aplicación de un intenso campo eléctrico (del orden de 2000 volts), al igual que en el resto de fuentes nanoelectrospray.

El HPLC-Chip también permite utilizar nanocolumnas convencionales; para ello se utilizará un adaptador que posicionará en la nanofuente, las tradicionales micropuntas emisoras para la nanonebulización, y que se conectan a la salida de las nanocolumnas.

Esta nanofuente puede ser fácil y rápidamente intercambiada con otras fuentes de ionización, como ESI, APCI, APPI para trabajar con columnas de mayor diámetro interno, e incluso por una fuente AP-MALDI.

2.1.2. "Chip Cube".

El "Chip Cube" es el dispositivo que se instala encima de la nanofuente para posteriormente poder introducir el Chip. Sus funciones son:

- Posicionar automáticamente el Chip en la posición de nebulización.
- Establecer conexiones fluídicas herméticas con el nano HPLC, para el suministro de los disolventes y la muestra. La conexión entre el "Chip Cube" y el nano HPLC, se realiza mediante el uso de unos muy finos capilares.
- Aplicar los cambios en las conexiones fluídicas que se requieran (programados en el tiempo), para hacer fluir la muestra y los disolventes por donde interese en cada momento. Para ello el "Chip cube" incorpora una doble válvula con dos rotores concéntricos e independientes (de 6 y 10 vías), que permiten automatizar complejos procesos que puedan requerir el uso de múltiples precolumnas y reactores. El "Chip cube incorpora" las partes móviles de la doble nanoválvula.

2.1.3. Chips

El Chip es donde tiene lugar todo el proceso cromatográfico, desde la preconcentración (en precolumnas), hasta la separación de la muestra en nanocolumna de 75µm, e incluso en algunos diseños incorporan reactores con enzimas inmovilizadas. El final de la columna acaba en la punta emisora para la formación del nanospray (asistida por campo eléctrico) e ionización de la muestra.

Como se comentará en el apartado 4.3, los chips se ofrecen en diversos tamaños de precolumna y columna, diversos tipo de configuraciones y rellenos.

2.1.4. Nano-cromatógrafo

El sistema HPLC-Chip necesita de un nanocromatógrafo provisto de 2 sistemas de bombeo y un microinyector:

- **Nanobomba:** trabajando típicamente a 300-600 nL/min. Se encargará de arrastrar a los analitos de la precolumna de preconcentración, a la nanocolumna donde tendrá lugar su separación cromatográfica.

- **Bomba Capilar:** trabajando típicamente a unos pocos µL/min. se encargará arrastrar a la muestra del micro-inyector a la precolumna de preconcentración; donde si las condiciones del eluyente utilizadas y la capacidad de carga son adecuadas, los péptidos de interés quedarán suficientemente retenidos, para su posterior elución hacia a la columna donde tendrá lugar su separación cromatográfica.

- **Microinyector Refrigerado**: se encarga de insertar en el flujo de arrastre suministrado por la bomba capilar, el volumen de muestra a inyectar; habitualmente unos pocos µL, pero puede llegar a reducirse a unas decenas de nL si la elevada concentración de la muestra lo requiere.

En el mercado existen 2 tipos de diseños de nanobombas:

- **Nanobombas o Bombas capilares CON División de Flujo**: suelen consistir en bombas con mezclado a alta presión, que después de la mezcla incorporan un divisor de flujo ("Split") que reduce el flujo, de unos centenares de µL/min a unos pocos µL/min, o nL/min en uno de sus 2 ramales. En algunos diseños (como el utilizado por el HPLC-Chip), estas válvulas de división del flujo son activas y están controladas por un sensor de flujo que mide los nL/min o µL/min que van hacia el inyector-columna. Estas bombas tienen el inconveniente de consumir un poco más de disolvente; unos pocos centenares de µL/min (100mL/8 horas); consumo que se puede regular dependiendo de la rapidez con que se quiere que el gradiente llegue a columna. En cambio aportan un par de importantes ventajas:

  - **Permiten ahorrar tiempo de análisis:** los cambios del gradiente (concentración eluyentes) al crearse y homogeneizarse en una zona de "flujo elevado", llegan en mucho menor tiempo a cabeza de columna. Consecuentemente permite incrementar la productividad del nano-LC/MS.

  - **Pueden utilizarse a altos flujos para trabajar con columnas convencionales:** si se desconecta el divisor de flujo, estas bombas puede utilizarse hasta flujos de unos pocos mL/min para trabajar con columnas de hasta 4.6 mm d.i.; por ejemplo con un colector de fracciones y columnas de inmunoafinidad para separar proteínas minoritarias de mayoritarias (4). De hecho, la configuración utilizada con el HPLC/Chip permite trabajar con columnas desde 50µm hasta 4.6m d.i. (en este caso con fuente ESI).

- **Nanobombas SIN División de Flujo:** suelen consistir en bombas de jeringa que operan directamente a nL/min, y por tanto no requieren efectuar una división de flujo. Tienen la ventaja de reducir a un nivel ínfimo, el consumo de disolvente (ahorran unos 100mL/8h.); pero al generarse y homogeneizarse el gradiente (composición de eluyentes) en una zona a muy bajo flujo (nL/min), suelen retrasar muy considerablemente el tiempo de llegada del gradiente a cabeza de columna, aumentando considerablemente los tiempos de análisis (los primeros péptidos no suelen aparecer hasta transcurridos 10-20min).

Etapas Análisis: en una análisis convencional por HPLC-Chip se puede distinguir 2 fases:

- **Fase de carga de muestra:** la bomba capilar con un eluyente eminente acuoso (típico 2-3% Acetonitrilo en HCOOH 0.1%), arrastra la muestra del microinyector para que sea retenida por la precolumna.

- **Fase de Análisis:** la nanoválvula del Chip gira, y el gradiente de concentración de la nanobomba eluye los péptidos hacia la nanocolumna, donde éstos se separan. Ver video descriptivo del funcionamiento de la tecnología HPLC-Chip/MS en web referencia 5.

Cromatografía Bidimensional (LC-2D): bastará colocar una columna, de por ejemplo, intercambio catiónico (SCX) entre el microinyector y el Chip. El gradiente salino para fraccionar la 1ª dimensión (SCX), se realizará con la bomba capilar; las fracciones se retendrán en la precolumna, de donde el gradiente de concentración de la nanobomba, las eluirá a la nanocolumna para su separación.

2.2. Materiales y Posibilidades de diseño.

El material de base utilizado para la fabricación de los chips es poliimida, un polímero biocompatible, resistente a altas temperaturas y a agentes químicos. Mediante ablación por láser se crean en láminas de poliimida (ver figura 2):

- **Microcanales:** harán las funciones de conductos o de columna según su diámetro.
- **Cámaras:** harán las funciones de pre-columnas de enriquecimiento o de reactores.
- **Filtros:** mantendrán el relleno en su sitio y protegerán al chip de la entrada de partículas.
- **Punta del nanoelectrospray:** se recubre de una película de metal, para poderle aplicar el campo eléctrico necesario para producir la pluma de nebulización.
- ...

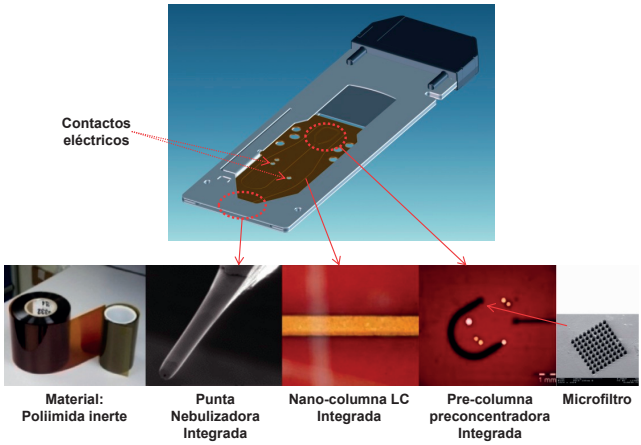


Figura 2. Material de partida y componentes integrados en el Agilent HPLC-Chip/MS: fotos de láminas de poliimida y de diversos componentes creados a partir de su ablación por láser.

Finalmente las diversas láminas que se requieran, se unirán mediante un proceso de laminación para formar un bloque compacto, que se encapsula en un soporte metálico. Este diseño proporciona una enorme flexibilidad permitiendo incluso diseños a medida del propio usuario (ver siguientes secciones 2.3 y 2.4).

La punta emisora del nanoelectrospray del HPLC-Chip está construida del mismo material, bioinerte y resistente a temperatura, que el resto del Chip; por ello, es mucho más resistente que las típicas puntas de cuarzo o sílice, de los nanoelectrosprays tradicionales. Estas puntas tradicionales son muy delicadas y se queman cuando aplicando voltaje, deja de circular fase móvil por su interior. Ello motiva que su vida sea muy corta (habitualmente unos pocos días); otros diseños basados en puntas metálicas son más resistentes, pero tienen el problema de adsorber ciertos péptidos, como por ejemplo fosforilados, generando su pérdida cuando se trabaja a muy bajas concentraciones. El HPLC-Chip solventa todos estos inconvenientes, al ser bioinerte e incluso admitir aplicarle voltaje, cuando no circula fase móvil por su interior; proporcionando una robustez sin igual en el mundo de la nanocromatografía/MS. Los chips suelen "morir", al igual que las columnas convencionales, por la paulatina pérdida de eficacia de la columna, después de centenares de inyecciones de muestras.

2.3. Configuraciones típicas.

La tecnología HPLC-Chip admite múltiples y variados diseños, desde los más sencillos, para efectuar infusión o FIA ("Flow Injection Analysis"), hasta los más sofisticados con reactores enzimáticos; aquí se describirán los más usuales. La configuración más habitual es la esquematizada en el diagrama de flujo de la figura 3. En una primera etapa se carga la muestra en la precolumna de enriquecimiento (ya dentro del HPLC-Chip), y en una segunda el gradiente generado por la nanobomba de separación, arrastra los péptidos a la nanocolumna de separación y al MS (ver figura 4).

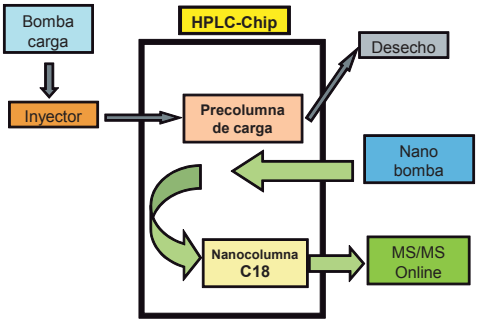


Figura 3. Diagrama de flujo típico en un sistema nano-LC/MS con preconcentración y desalinización de la muestra. La muestra es arrastrada por una bomba de carga (bomba capilar) del microinyector hacia la precolumna de enriquecimiento (ya dentro del HPLC-Chip), donde los péptidos de interés son retenidos (diagrama flujo en gris). Una vez cargada la muestra en la precolumna, la nanoválvula del HPLC-Chip gira y el gradiente generado por la nanobomba, arrastra los péptidos de la precolumna a la nano-columna de separación y al espectrómetro de masas (diagrama flujo en verde).

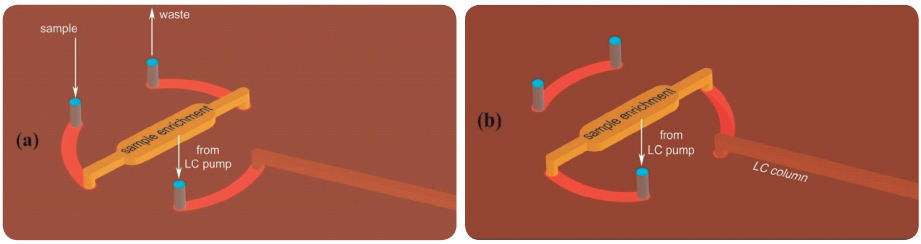


Figura 4. Nanoválvula del HPLC-Chip/MS (A): la muestra procedente del microinyector ("sample" arrastrada por la bomba capilar), se preconcentra en la precolumna de enriquecimiento.

Figura 4. Nanoválvula del HPLC-Chip/MS (B): después del giro de la nanoválvula, el gradiente de elución de la nanobomba ("LC pump") arrastra la muestra preconcentrada, de la precolumna a la columna de separación y al MS. Con la configuración de conexiones mostrada en la figura, la carga y elución de la muestra se efectúan en el mismo sentido; para invertirlas, basta intercambiar las conexiones de nanobomba y nanocolumna.

Otras configuraciones típicas son chips que no incluyen precolumnas (Chip de inyección directa), e incluso sin columna, para simplemente ionizar la muestra tal cual le entra. Las configuraciones genéricas más habituales son:

- **Chips con Precolumna y Columna:** son los más utilizados al combinar la preconcentración/desalización y separación de la muestra. Su diagrama de flujo corresponde al de la figura 3. La fase de carga y elución, se puede configurar para que se efectúen:
  - En sentido contrario: refocalizando los analitos preconcentrados –mejora el perfil de los picos menos retenidos.
  - En el mismo sentido: protegiendo la nanocolumna – posteriormente la precolumna se podrá limpiar en modo backflush (ver figura 4).

  Configuraciones típicas:

  - **Precolumnas de 40 y 160 nL de volumen**: tienen por objeto preconcentrar la muestra atrapando los péptidos de interés (a la vez que la desalinizan). Las habitua-

les son de 40 y 160 nL, pero también se ofrece una versión con una precolumna (o trampa) de 500nL en el caso del UHC-Chip (Chip de Ultra Alta Capacidad), para la retención de moléculas pequeñas muy polares (incluso con log P = 0 – ver pico atenolol en figura 1 (en apartado 1). La selección de una u otra, dependerá de la cantidad de material total que se inyecte en el chip:

- Capacidad carga trampa 40nL: 0.5µg.
- Capacidad carga trampa 160nL: 2µg .

Otro aspecto que influye en la selección, es la polaridad de los péptidos; cuanto más polares sean los péptidos de interés, mayor deberá ser el tamaño de la precolumna para retenerlo.

- **Nanocolumnas de 43 y 150mm de longitud:** su objetivo es separar la muestra antes de entrar en el masas. En la columna de 43mm ésta va en línea recta de la zona de entrada de muestra a la punta de nebulización. La selección de una u otra columna dependerá de la complejidad de la muestra, y resolución cromatográfica necesarias; muestras complejas requerirán las columnas de 150mm, mientras que muestras más simples, como las manchas digeridas de un gel 2D, se podrán analizar con la columna de 43mm de longitud. El tamaño de los microcanales es de 75µm.

- **FIA-Chip:** es el más sencillo de los Chips y se utiliza para ionizar directamente la muestra; simplemente conduce la muestra de la entrada del Chip al espectrómetro donde la ioniza, para realizar infusión o FIA:

  - **Infusión**: permite la entrada en continuo de la misma muestra, y así poder analizarla interactivamente durante varios minutos. Para introducir la muestra se suele utilizar una jeringa externa.

  - **Análisis por Inyección en Flujo (FIA):** la introducción de la muestra se realiza en forma de un único pico, sin separación cromatográfica; utiliza el nano-cromatógrafo como simple sistema de introducción de la muestra al HPLC-Chip/espectrómetro (sin utilizar ninguna columna); aspecto que facilita el análisis automatizado de múltiples muestras (la infusión es un proceso manual).

  - **Chip de Inyección Directa en Columna**: es un chip que no incorpora precolumnas y permite la inyección directa de la muestra en la columna de 150mm y 75µm que incorpora. Es útil cuando interesa analizar péptidos muy pequeños y muy polares, que sean difíciles de retener en una precolumna; o cuando interesa que toda la muestra llegue al masas. Será importante que la muestra esté disuelta en un medio altamente acuoso, que retenga a los péptidos más polares en cabeza de la columna de fase reversa, para evitar picos deformes y muy anchos por el arrastre inducido por el propio disolvente de la muestra; especialmente si se inyectan volúmenes grandes.

La flexibilidad del diseño multicapa del dispositivo, permite innovar automatizando complejos procesos analíticos, mediante la integración de diversas precolumnas y reactores. Por ejemplo el Fosfochip, mediante el uso de 3 precolumnas (ver figura 5), dos de fase reversa y una rellena de dióxido de titanio, permite la preconcentración selectiva de fosfopéptidos, y su análisis por separado del resto de péptidos de la muestra. En el apartado 4.2 se describirá éste Chip con más detalle.

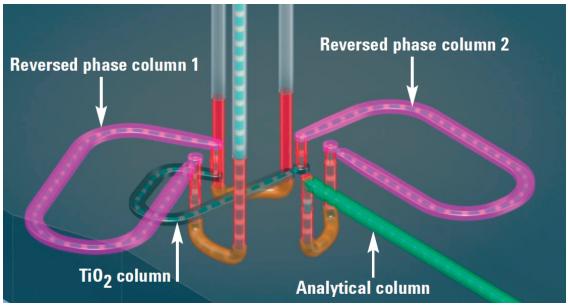


Figura 5. Diseño multicapa del Agilent HPLC-Chip/MS: éste permite integrar múltiples precolumnas, reactores y columnas para automatizar complejos procesos analíticos. En la figura se visualiza la zona de carga de muestra del Fosfochip; éste incorpora 3 precolumnas: 1 de TiO2 y 2 de fase reversa.

2.4. Configuraciones a medida.

La enorme flexibilidad que proporciona el diseño multicapa del Chip, y su fabricación mediante ablación por laser (dirigida por ordenador), permite la fabricación de Chips a medida de las ideas del propio usuario. Ello permitirá llegar a automatizar, muy complejos procesos de análisis y tratamiento de muestras online, diseñados por el propio usuario, e integrados en el mismo dispositivo. Esta facilidad de innovación, constituye una excelente oportunidad para generar novedosas publicaciones, e incluso, llegar a comercializar aquellos diseños que puedan tener un amplio uso.

Las configuraciones a medida se pueden clasificar en 2 grupos:

- **Rellenado con material proporcionado por el usuario, de configuraciones estándar:** el usuario suministra el material con el que le interesa se rellene la configuración de interés.

- **Diseños realizados por el propio usuario:** éste define su diseño: el número y tamaño de las precolumnas, reactores, columnas, filtros,….el orden de interconexión,….En definitiva, el plasmar del papel a un Chip, las ideas innovadoras del usuario para automatizar sofisticados procesos analíticos.

3. Aportaciones de la tecnología HPLC-Chip/MS.

Las aportaciones de la tecnología HPLC-Chip/MS son numerosas, Côté y Thibault detallan algunas de ellas en: "Advantages of using Intelligent Sample Loading with the HPLC-Chip for automated sample enrichment" (6) Las aportaciones más remarcables de la tecnología HPLC-Chip las podríamos resumir en:

- **Robustez y Facilidad de uso**: el usuario no tiene que preocuparse de fugas o volúmenes muertos generados por una mala nano-conexión. El intercambio de Chips se realiza en segundos, y se equilibra en muy pocos minutos, incluyendo la pluma del nanospray. El chip admite muchos centenares de inyecciones sin que tenga que reemplazarse.

- **Mejor Sensibilidad y picos más finos y simétricos**: la poliimida utilizada para la fabricación del Chip, es un material mucho más inerte que la sílice fundida utilizada en las nanocolumnas convencionales; por ello el Chip:

- elimina colas en los picos de péptidos, producidas por su interacción con los silanoles de la pared de sílice de las nanocolumnas clásicas.

- proporciona picos más finos y simétricos, y por tanto, con mayor resolución, altura y sensibilidad.

Hardouin y Caron "Usefulness of an integrated microfluidic device (HPLC-Chip-MS) to enhance confidence in protein identification by proteomics" (7) reportan una mejora de 5 veces en sensibilidad, simplemente por el intercambio de una fuente nanoelectrospray por el HPLC-Chip, utilizando el mismo espectrómetro.

- **Mejor Reproducibilidad en tiempos de retención**: la combinación de un sistema libre de fugas, con la excelente reproducibilidad de un sistema de bombeo que mide y controla en continuo los nL/min que entran en el HPLC-Chip, proporciona una excelente reproducibilidad de los tiempos de retención. Ello permitirá utilizar ventanas de tiempo más finas para la monitorización específica de péptidos ("target MS/MS" en QTOF's o "Dynamic MRM" con QqQ's), aumentando la sensibilidad de los análisis (se podrán utilizar "dwell times" mayores). Una buena reproducibilidad de tiempos, también facilitará tareas de identificación y diferenciación de péptidos por tiempo de retención.

- **Capacidad de preconcentrar, limpiar y desalinizar la muestra en línea:** el uso de un sistema con 2 bombas, y una precolumna conectada a una nanoválvula, permitirá preconcentrar la muestra en una precolumna/trampa, para ganar sensibilidad inyectando volúmenes de varios µL (muy superiores al propio volumen de la nanocolumna). Adicionalmente todas las sales, no retenidas en la precolumna, irán directamente al desecho sin pasar por la nanocolumna, ni ensuciar la fuente nanoelectrospray.

- **Flexibilidad, Rápido y Fácil intercambio de Chips:** facilitará que cada proyecto/grupo de trabajo pueda utilizar sus propios chips.

- **Amplia gama de Chips:** para todo tipo de aplicaciones de lo más diverso, desde moléculas pequeñas: fármacos, metabolitos,…., glicanos,… hasta mapas peptídicos y proteínas intactas. Incluso se ofrecen chips para preconcentrar selectivamente fosfopéptidos, y otros que incluyen reactores enzimáticos para liberar y analizar N-glicosilaciones.

- **Capacidad de automatizar en línea múltiples procesos de preparación y análisis de muestras**: ofrece una atractiva capacidad de innovar y publicar, desarrollando nuevos diseños que automaticen en línea, sofisticados procesos de tratamiento y análisis de muestras. En el siguiente apartado se verán algunos ejemplos.

Todas estas aportaciones permiten aumentar considerablemente la productividad del laboratorio, a la vez que abren la puerta a publicar innovaciones científicas, con el desarrollo de avanzados sistemas de análisis por nano-LC/MS. Por ejemplo, Staples y Zaia en: "Improved hydrophilic interaction chromatography LC/MS of heparinoids using a chip with postcolumn makeup flow" (8); para mejorar la estabilidad del nanospray en polaridad negativa a lo largo de todo el gradiente, adicionan, entre el final de la nanocolumna y la punta del nanospray, un microcanal por el que añaden en continuo 200nL/min de acetonitrilo.

4. Aplicaciones de la tecnología HPLC-Chip/MS.

En este apartado se describirán los Chips recomendados para aplicaciones típicas, y mostrarán algunos ejemplos de aplicación del HPLC-Chip, incluyendo algunas referencias bibliográficas para quienes estén interesados en profundizar.

4.1. Identificación y Cuantificación de Péptidos y Proteínas.
Para éste tipo de aplicación se utilizarán los chips estándar con precolumna y nanocolumna de 300-C18 (/C8), con la combinación de volumen de trampa y longitud de nanocolumna que mejor convenga a las características de la muestra:

- Capacidad carga Precolumnas (incluyendo proteína y otros compuestos presentes en la matriz de la muestra inyectada):
  - 160nL para capacidad carga total <2 µg
  - 40nL para capacidad carga total <0.5 µg

- Longitud nanocolumna:
  - 43mm: muestras sencillas, muestras procedentes de manchas de geles 2D, mezclas de digeridos de varias proteínas,…
  - 150mm: muestras complejas, sin fraccionar, que contengan un número elevado de péptidos,….

- Identificación de proteínas desconocidas: en la fase de descubrimiento de biomarcadores o de identificación de proteínas desconocidas, convendrá utilizar columnas más bien largas para tener una mejor separación cromatográfica, y estar en mejores condiciones para obtener espectros MS/MS de péptidos minoritarios.

- Cuantificación de proteínas: en la fase de validación o cuantificación de proteínas concretas, se podrá reducir la longitud de la columna, para reducir los tiempos de análisis y poder analizar un mayor número de muestras; no obstante, al reducir la longitud de la columna, convendrá evaluar el impacto que la supresión o exaltación de la ionización, pueda tener sobre la sensibilidad en el caso de muestras complejas.

- Análisis rápidos: cuando interesa analizar un gran número de muestras, el chip con trampa de 40nL y columna de 43mm, permitirá análisis muy rápidos:

  - Miller y Glatz (9) con un gradiente en 7min y trabajando a 600nL/min, muestran péptidos con picos muy simétricos de tan sólo 2-3 segundos de ancho al 50% de su altura (ver figura 6).

  - Un acoplamiento 2D con una columna de intercambio BioSCX II situada entre el microinyector y el chip. La rapidez del análisis de la segunda dimensión por fase reversa en el chip, permite poder analizar un mayor número de fracciones de la primera dimensión (SCX), o alternativamente de un sistema de Offgel Electrophoresis, o de manchas de un gel 2D.

- Caracterización de Proteínas Intactas: para caracterizar proteínas intactas utilizando cromatografía, se recomienda el Protein Chip con una trampa de 40nL y una columna de 300-C8 de 43mm de longitud. Cuando no interese efectuar cromatografía, se puede recurrir al FIA-Chip para realizar FIA o infusión como en la publicación de Kilby: "Noncovalently Associated Complexes: Apomyoglobin and Myoglobin Studies Using the HPLC-Chip/MS System Coupled with ESI-TOF Mass Spectrometry" (10).

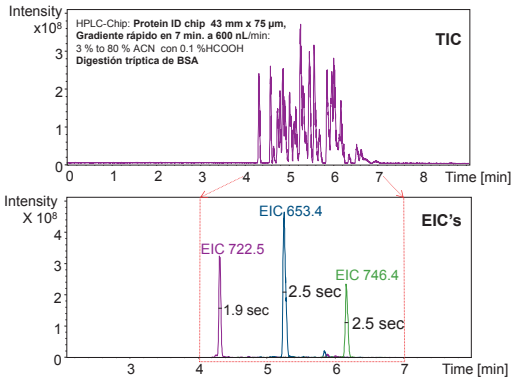


Figura 6. Simetría Péptidos con el HPLC-Chip: análisis rápido de una digestión tríptica de Albúmina de Suero Bovino (BSA) con un "Protein ID chip" de 43mm. Gráfico superior: TIC (Total Ion Chromatogram), éste suma las respuestas de todos los péptidos ionizados; cada punto del cromatograma corresponde a la suma de intensidades de todas las m/z. Gráfico inferior: de la misma inyección, se representan sólo las m/z (EIC) correspondientes a las masas de 3 péptidos (m/z: 722.5, 653.4, 746.4). Los picos de los EIC son simétricos, muy finos (ancho al 50% altura del orden de 2 segundos), y con una buena selectividad; ello facilitará la identificación de un mayor número de proteínas. Las condiciones de elución y dimensiones del chip utilizado proporcionan una excelente rendimiento para la identificación rápida (9 minutos) de proteínas en mezclas poco complejas (p.e. digestiones de manchas geles 2D).

4.2. Preconcentración y Análisis Selectivo de Fosfopéptidos.

Por la importancia y dificultad del análisis de fosforilaciones, consecuencia de su desfavorable estequiometria, es donde el uso de la tecnología HPLC-Chip combinada con el Fosfochip, proporciona un mayor valor añadido. El automatizar en línea, la preconcentración selectiva de los fosfopéptidos (con TiO2), y su análisis por separado del resto de péptidos de la muestra, aumenta enormemente el número de fosfopéptidos identificados; además reduce muy apreciablemente, el esfuerzo que exige el análisis de fosfopéptidos a muy bajas concentraciones, en muestras proteolíticas complejas.

Un buen ejemplo de las aportaciones y posibilidades del Fosfochip, se muestra en el trabajo de Heck y Raijmakers, "Exploring the Human Leukocyte Phosphoproteome Using a Microfluidic Reversed-Phase-TiO2- Reversed-Phase High-Performance Liquid Chromatography Phosphochip Coupled to a Quadrupole Time-of-Flight Mass Spectrometer" (11), donde para caracterizar el fosfoproteoma de leucocitos humanos, combinan un fraccionamiento previo por SCX, seguido del análisis mediante HPLC-Chip/QTOF. Son capaces de identificar 1.012 fosfopéptidos procedentes de 573 proteínas diferentes. En total 11.539 péptidos procedentes de 2.767 proteínas. También evalúan la especificidad y eficiencia del fosfochip, analizando los resultados de la fracción más rica de fosfopéptidos (la correspondiente a péptidos con 1 carga):

- **Especificidad** (capacidad de retener sólo fosfopéptidos): sólo detectan 11 péptidos NO fosforilados (retenidos por el TiO2) frente a más de 600 fosfopéptidos.

- **Eficiencia en la retención** (capacidad de retener todos los fosfopéptidos): sólo detectan 24 fosfopéptidos que no han sido retenidos por el TiO2 (+8 parcialmente retenidos) frente a los más de 600 fosfopéptidos retenidos por el TiO2.

En otras publicaciones (12-13 ésta última en castellano), los mismos autores analizan células de osteosarcoma humano, y comparan los resultados entre el Fosfochip y un Chip estándar de C18 (15 cm, 75 µm Reprosil C18); para ello, una alícuota de una fracción de SCX rica en fosfopéptidos, es analizada con ambos chips; el resultado es espectacular (ver Figura 7):

- Chip C18: 11 fosfopéptidos
- Fosfochip: 116 fosfopéptidos (+4 no retenidos en TiO2)

En el chip convencional los fosfopéptidos coeluyen con otros péptidos (no fosforilados) muy mayoritarios, consecuentemente en un análisis no dirigido (Auto MS/MS), será difícil tomar espectros MS/MS de fosfopéptidos. Por el contrario, con el Fosfochip, éstos eluyen en un cromatograma por separado, en el que sólo aparecen fosfopéptidos; consecuentemente, "todos" los espectros MS/MS adquiridos serán de péptidos fosforilados; de ahí el espectacular incremento en el número de identificaciones. Si además se aumenta x5 la cantidad de muestra analizada, el

número de fosfopéptidos identificados por el motor Spectrum Mill pasa de 116 a 504 fosfopéptidos (+6 no retenidos en TiO2).

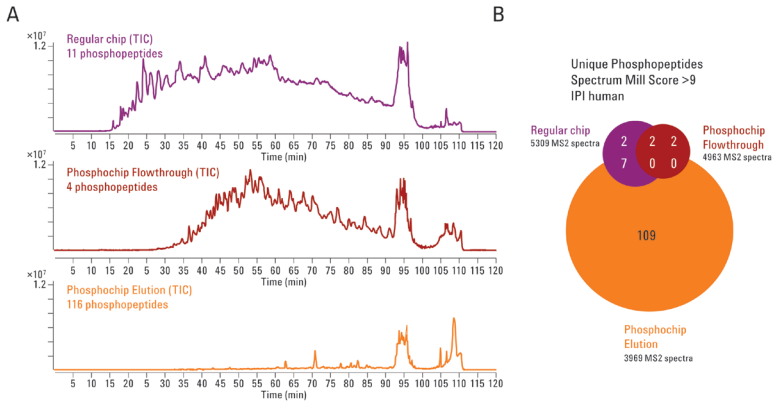


Figura 7. Comparación del rendimiento en la identificación de fosfopéptidos del Agilent Fosfochip "versus" Chip C18. 1.-Cromatograma superior: TIC del Chip C18: se identifican 11 fosfopéptidos. 2.-Cromatograma central: TIC de la fracción NO retenida en TiO2 del Fosfochip: se identifican 4 fosfopéptidos (éste cromatograma corresponde a los péptidos no fosforilados). 3.-Cromatograma inferior: TIC de la fracción retenida en TiO2 del Fosfochip: se identifican 116 fosfopéptidos. Los 3 cromatogramas se representan en la misma escala; puede observarse la desfavorable estequiometria de los péptidos fosforilados; su separación cromatográfica de los péptidos no fosforilados (mucho más concentrados) facilita enormemente su identificación, y aumenta espectacularmente el número de identificaciones.

Aportaciones específicas del Fosfochip:

- Aumento en el número de fosfopéptidos Identificados.
- Automatizar en línea la preconcentración selectiva de fosfopéptidos.
- Analizar por separado péptidos fosforilados (muy minoritarios), de los NO fosforilados (mayoritarios).

Componentes del Fosfochip para realizar la diferentes etapas de preconcentración y análisis (ver **Figura 5 en apartado 2.3**):

- 2 precolumnas de 100nL rellenas con ZORBAX Extend-300-C18.
- 1 precolumna de 40nL rellena con esferas de dióxido de titanio (TiO2).
- 1 nanocolumna de 150mm de Reprosil Pur C18 Aq.

Etapas en un Análisis mediante el Fosfochip. Las etapas para la preconcentración y análisis por separado de péptidos fosforilados y no fosforilados son:

- **Preconcentración inicial de la muestra**: la mezcla de péptidos se inyecta, con la nanoválvula del fosfochip en posición de carga; todos los péptidos quedan retenidos (fosforilados y no fosforilados) en la primera precolumna de fase reversa que encuentran (RP1). Ver en figura 8 la fase de carga de la muestra. Ésta se efectúa a 3 µL/min, utilizando la bomba capilar para arrastrar los péptidos del microinyector al HPLC-Chip, utilizando como fase de arrastre: una mezcla de 0.6% de ácido acético y 2% de ácido fórmico en agua.

- **Elución y análisis de los péptidos NO fosforilados:** la nanoválvula del chip conmuta (Ver figura 8 posición de análisis), y el eluyente procedente de la nanobomba (gradiente agua/acetonitrilo a pH ácido), arrastra y va eluyendo todos los péptidos, de la primera

precolumna (RP1) hacia la segunda precolumna (la de TiO2). En ésta los fosfopéptidos quedan retenidos, mientras los no fosforilados continúan su camino hacia la tercera precolumna (RP2), donde no se retienen, y continúan avanzando hacia la nanocolumna; donde tiene lugar la separación final, antes de ser ionizados en el nanospray que incorpora el fosfochip. Así se obtiene un primer cromatograma con todos los péptidos no fosforilados (que no han sido retenidos en la precolumna de TiO2).

En caso de sólo interesar los péptidos fosforilados, se puede efectuar un rápido gradiente con la bomba capilar, manteniendo la nanoválvula del chip en posición de carga; de esta manera los péptidos NO fosforilados son enviados directamente a desecho y no se analizan; otra alternativa más sencilla, es inyectar 8µL de solvente orgánico para eluirlos de la precolumna donde estaban retenidos.

- **Lavado precolumna (con péptidos fosforilados retenidos):** en esta fase los péptidos no fosforilados ya han sido eluidos del fosfochip, sólo quedan los péptidos fosforilados; éstos se encuentran retenidos en la segunda precolumna (TiO2). Antes de proceder a su elución, la nanoválvula del chip vuelve a la posición de carga, y opcionalmente se puede efectuar un lavado de las precolumnas inyectando 20µL de acetonitrilo al 50% que contiene una serie de aditivos (cítrico, DHB, DMSO); Ver figura 8 posición de carga.

- **Elución de los péptidos Fosforilados del TiO2 y refocalización en C18:** la elución de los fosfopéptidos de la precolumna de TiO2, se realiza manteniendo la nanoválvula del chip en posición de carga, e inyectando 20µL de una solución de una solución tampón básica; ésta arrastra los fosfopéptidos hacia la segunda precolumna de fase reversa (RP2 - figura 8 posición de carga), donde quedan retenidos/refocalizados.

- **Elución y análisis de los péptidos Fosforilados:** la nanoválvula del chip conmuta (figura 8 posición de análisis), y el eluyente procedente de la nanobomba (gradiente agua/acetonitrilo a pH ácido), arrastra y va eluyendo todos los fosfopéptidos retenidos en RP2 hacia la nanocolumna, para su separación e ionización. Así se obtiene un segundo cromatograma con todos los péptidos fosforilados (que inicialmente habían sido retenidos en la precolumna de TiO2).

- **Regeneración de la precolumna de TiO2**: antes de proceder al siguiente análisis, se inyectan 4µL de una solución de regeneración de la precolumna de TiO2, ésta se mantendrá en la precolumna durante unos 3 minutos.

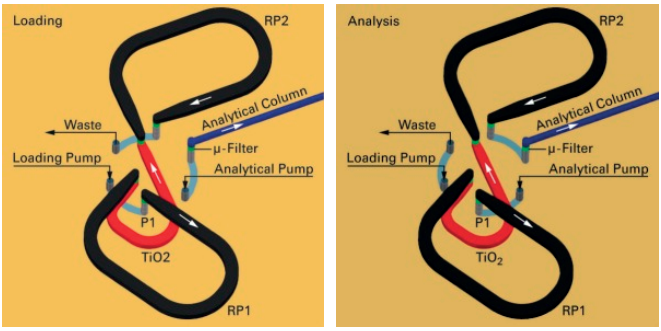


Figura 8. FosfoChip: esquema de la zona de la nanoválvula. Figura izquierda ("Loading") etapas 1, 3 y 4: Fase de carga de muestra y preconcentración inicial. En la etapa 1 la muestra se carga (procedente del microinyector y arrastrada por la bomba capilar de carga), y preconcentra en la primera precolumna de fase reversa que se encuentra (RP1). En las etapas 3 y 4 se procede al lavado y elución de los fosfopéptidos de la precolumna de TiO2 (y su retención en RP2). Figura derecha ("Analysis") etapas 2 y 5: Fase de análisis de muestra. El eluyente de la nanobomba analítica arrastra los péptidos de las 3 precolumnas hacia la nanocolumna de separación (y MS). En la etapa 2 se eluyen todos los péptidos preconcentrados en la primera precolumna (RP1), y se analizan los péptidos no fosforilados. En la etapa 5 se eluyen y analizan todos los fosfopéptidos retenidos en la segunda precolumna de fase reversa (RP2 – situada después de la de TiO2).

Un resumen de las 4 etapas más importantes puede verse en figura 9.

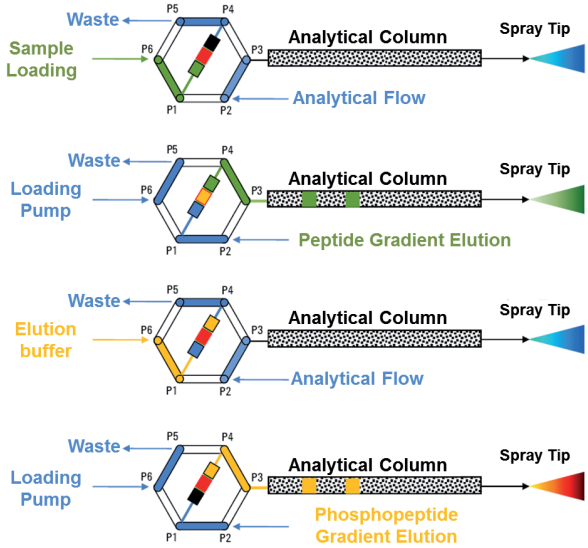


Figura 9. Diagrama FosfoChip: Etapas en el análisis de péptidos fosforilados y no fosforilados. En verde se representa la carga de la muestra y elución de los péptidos no fosforilados. En naranja se presenta la desorción de los fosfopéptidos de la precolumna de dióxido de titanio (en rojo), y su posterior separación en la columna.

Ejemplo de Cuantificación con Fosfochip: Otegui y Sussman utilizan un Fosfochip acoplado a un QqQ-6460 para la cuantificación de diversos fosfopéptidos: "Molecular Characterization of Mutant Arabidopsis Plants with Reduced Plasma Membrane Proton Pump Activity" (14).

4.3. Caracterización de N-Glicosilaciones en Anticuerpos Monoclonales (mAB).

Dada su elevada especificidad, los fármacos basados en anticuerpos monoclonales (mAB), cada vez están adquiriendo más importancia. La compleja estructura de oligosacáridos que incorporan los mAB's, juega un papel muy importante en las características del fármaco; consecuentemente, además de la caracterización de la parte proteica, también será muy importante una buena caracterización de sus N-glicosilaciones.

Los procedimientos utilizados hasta la fecha para la caracterización de las N-glicosilaciones en mAB's, contando el tiempo de preparación de la muestra, requieren entre 3 y 8 horas para su análisis por MALDI-TOF, o entre 2 y 4 días utilizando LC o CE con fluorescencia. Utilizando un nuevo Chip (mAB-Glyco Chip) que incorpora un reactor enzimático, ese tiempo se reduce a 10-15minutos; adicionalmente, utilizando una base de datos de masa exacta y tiempo de retención (opcional), se podrá automatizar la identificación y caracterización de los N-glicanos.

El mAB-Glyco Chip es un desarrollo aún más sofisticado que el del anterior apartado, es el primer chip comercial que hace uso de los 2 rotores independientes (uno interno de 6 vías y otro externo de 10), que incorpora la doble nanoválvula del Chip Cube; los utiliza para manejar el reactor enzimático, precolumna y nanocolumna que incorpora.

¿Puede utilizarse el mAB-Glyco Chip para caracterizar N-glicosilaciones en glicoproteínas?: El mAB-Glyco Chip ha sido desarrollado para anticuerpos, pero puede utilizarse con otro tipo de proteínas, aunque la eficiencia en la liberación de los N-glicanos puede variar apreciablemente dependiendo de la estructura de la proteína; consecuentemente, se debería optimizar el tiempo y las condiciones de la reacción enzimática por un lado, y el gradiente de separación de los glicanos por otro.

Etapas en un Análisis mediante el mAB-Glyco-Chip: Las etapas para la deglicosilación del anticuerpo, retención y análisis de los glicanos liberados son:

- **Inyección y llenado del reactor con la muestra:** la bomba capilar arrastrará a 3µL/min, mediante el tampón de deglicosilación, unos 2 µL del anticuerpo (típico a 0.5-1 mg/mL) del microinyector al rotor externo del Chip-Cube, y llenará de muestra el reactor de 310nL, que contiene la enzima PNGasa F inmovilizada sobre partículas de silica de 5 µm.

- **Proceso de Deglicosilación:** se aísla el reactor, y durante unos 4 minutos se deja que el anticuerpo monoclonal reaccione con la PNGasa F, para deglicosilarlo y liberar los N-glicanos (en forma de glicosilaminas). Mientras tanto, la bomba capilar limpia el circuito de posibles restos de muestra, y la nanobomba continua acondicionando la precolumna y nanocolumna de separación para la próxima etapa.

- **Descripción del proceso:** como se muestra en la (figura 10), los N-glicanos están unidos a las cadenas polipeptídicas a través de residuos de Asparragina, la PNGasa rompe este enlace liberando las correspondientes glicosilaminas; éstas en medio ácido se hidrolizan parcialmente (5-10% en las condiciones utilizadas) dando lugar a la formación de 2 anómeros del glicano libre (isómeros diastereoméricos según la diferente distribución espacial de los hidroxilos) (15); estos 2 anómeros se separan entre sí en la columna de grafito; así cada N-glicano liberado dará lugar a 3 picos: 1 glicosilamina (mayoritaria) y 2 anómeros del glicano libre.

- **Transferencia de los N-glicanos y eliminación de la proteína:** los 2 rotores giran simultáneamente para que la bomba capilar con tampón de deglicosilación, arrastre los N-glicanos del reactor enzimático a la precolumna de 160nL rellena de carbón grafitizado poroso; en ella se retendrán los N-glicanos, mientras la proteína no se retendrá y eluirá a desecho (a menos que recoja con un microcolector de fracciones). Al girar los 2 rotores, la nanobomba quedará directamente conectada la nanocolumna (sin la precolumna) para su acondicionamiento.

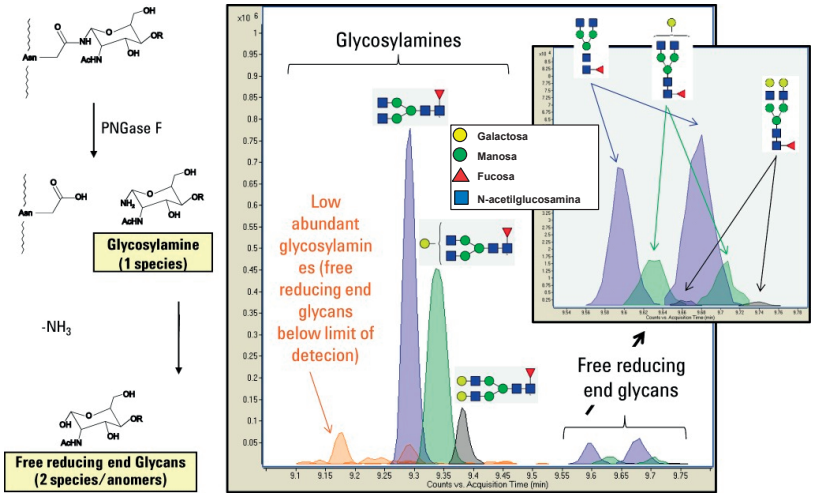


Figura 10. Proceso mAB-Glyco-Chip: los N-glicanos unidos al anticuerpo a través de residuos de asparagina, se liberan por acción de la enzima PNGasa F (inmovilizada en el reactor del Chip), generando las correspondientes glicosilaminas; parte de éstas, en medio ácido pierden amoniaco y dan lugar a la formación de 2 isómeros del glicano libre. Separación cromatográfica: los 2 anómeros se separan entre sí, y de la correspondiente glicosilamina; cada N-glicano liberado da 3 picos (representados con el mismo color): 1 glicosilamina (mayoritaria) y 2 anómeros del glicano libre ("free reducing end glycans"). Los códigos de símbolos de las estructuras glicánicas corresponden a azúcares (monosacáridos) y derivados típicos en N-glicanos procedentes de la IgG.

- **Separación y Análisis de los N-glicanos:** el rotor interno gira para que la nanobomba, mediante un gradiente agua/acetonitrilo en medio ácido (0.1% HCOOH), eluya a 500nL/min los N-glicanos de la precolumna a la nanocolumna de carbón grafitizado poroso de 43mm x 75µm, donde tendrá lugar la separación de las glicosilaminas y de los anómeros de los glicanos libres (figura 10). Mientras dura el análisis la bomba capilar acondiciona el reactor enzimático con el tampón de deglicosilación. Para que no pierda actividad, el reactor enzimático debe mantenerse siempre con tampón de deglicosilación en su interior; cuando no se vaya a utilizarse durante más de 1 día conviene congelar el chip -20ºC con el reactor lleno de tampón. Almacenado en condiciones adecuadas, la vida del mAB-Glyco chip es de varios centenares de inyecciones (16).

- **Análisis de los datos:** los datos son deconvolucionados mediante el algoritmo "Molecular Feature Extractor" para detectar las masas exactas, tiempos de retención e intensidades de señal de todos los glicanos. Estos se identifican y anotan mediante el uso de una base de datos de masa exacta ("mAb-Glyco database" con tiempo de retención opcional). Por último se calcula la intensidad de señal y ratios relativos de cada uno de los glicanos, sumando las señales correspondientes a los diferentes picos de cada especie de N-glicano liberado, aplicando un factor corrector para compensar el diferente nivel de respuesta entre glicosilaminas y glicanos. En la figura 11 puede verse una tabla con un ejemplo de resultados. En la web de la referencia (17) se describen las normas para la anotación de los glicanos

Compound Label	RT	Mass	Name	Formula	Tgt Mass	Diff (ppm)	Isomer(s) Present	Total Volume	Total Vol%
Cpd 28: 2121 0A 1G N	9,589	2092,7583	2121 0A 1G N	C79H132N6O58	2092,7564	0,93		8906	0,3
Cpd 23: 1100 0A 0G N	9,373	1258,4803	1100 0A 0G N	C48H82N4O34	1258,481	-0,58	1258,481	40030	1,5
Cpd 20: 2120 0A 0G N	9,349	1785,6693	2120 0A 0G N	C68H115N5O49	1785,6661	1,83		11972	0,5
Cpd 19: 3100 0A 0G N	9,317	1664,6406	3100 0A 0G N	C64H108N6O44	1664,6398	0,46		10792	0,4
Cpd 17: 2110 0A 0G N	9,305	1623,6155	2110 0A 0G N	C62H105N5O44	1623,6132	1,39		284929	10,8
Cpd 16: Man5 N	9,262	1233,4477	Man5 N	C46H79N3O35	1233,4494	-1,36	1233,4494	134626	5,1
Cpd 15: 2100 0A 0G N	9,250	1461,5628	2100 0A 0G N	C56H95N5O39	1461,5604	1,66		2044974	77,4
Cpd 13: 1100 0A 0G N	9,230	1258,4825	1100 0A 0G N	C48H82N4O34	1258,481	1,15	1258,481	40030	1,5
Cpd 11: 2010 0A 0G N	9,203	1477,5546	2010 0A 0G N	C56H95N5O40	1477,5553	-0,5		5076	0,2
Cpd 10: 1110 0A 0G N	9,198	1420,5339	1110 0A 0G N	C54H92N4O39	1420,5339	0,03		864	0,0
Cpd 9: Man4 N	9,170	1071,3928	Man4 N	C40H69N3O30	1071,3966	-3,55	1071,3966	6937	0,3
Cpd 5: 2000 0A 0G N	9,133	1315,5029	2000 0A 0G N	C50H85N5O35	1315,5025	0,3		59267	2,2
Cpd 4: 1000 0A 0G N	9,103	1112,4235	1000 0A 0G N	C42H72N4O30	1112,4231	0,31		34631	1,3
Cpd 3: Man4 N	9,082	1071,3956	Man4 N	C40H69N3O30	1071,3966	-0,9	1071,3966	6937	0,3
Cpd 1: Man5 N	9,068	1233,4475	Man5 N	C46H79N3O35	1233,4494	-1,52	1233,4494	134626	5,1

Figura 11. Análisis N-glicosilaciones: ejemplo de resultados de la identificación y cuantificación relativa de los N-glicanos procedentes de un anticuerpo monoclonal (mAB). Los glicanos remarcados con el mismo color corresponden a isómeros separados cromatográficamente mediante la nanocolumna de grafito.

Para ampliar información consultar:

- La web descrita en referencia (18) para visualizar un ilustrativo video sobre el funcionamiento del mAB-Glyco-Chip.

- Binum y Killeen en "Characterization of IgG N-Glycans Employing a Microfluidic Chip that Integrates Glycan Cleavage, Sample Purification, LC Separation, and MS Detection" (16) describen con todo detalle el proceso utilizado.

- Wollscheid y Aebersold en "Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins" (19) utilizan un HPLC-Chip/QTOF para el análisis de glicoproteínas con el objeto de fenotipar células mediante el estudio de centenares de puntos de glicosilación de su superficie.

- Masana I, en "Caracterización de Anticuerpos Monoclonales (mAB's) y Biofármacos por HPLC y LC/MS" (20) para ampliar información en herramientas cromatográficas para caracterizar anticuerpos.

4.4. Análisis de Glicanos, Nucleósidos y Moléculas Muy Polares.

Para analizar glicanos directamente, u otras moléculas muy polares que se puedan retener en grafito, como es el caso de los nucleósidos y oligosacáridos entre otros, se puede recurrir a utilizar un PGC-chip con una nanocolumna de 43mm y 75µm de carbón grafitizado poroso (PGC), provisto de una precolumna de preconcentración de 40nL del mismo material, y utilizando los clásicos gradientes de fase reversa: agua/acetonitrilo con 0.1% HCOOH. Por ejemplo:

- Giessing y Kirpekar (21) en "A Nano-Chip-LC/MS^n Based Strategy for Characterization of Modified Nucleosides Using Reduced Porous Graphitic Carbon as a Stationary Phase", utilizan un Chip de 15cm relleno de PGC con una precolumna de 9mm y 160nL, y un gradiente agua/acetonitrilo en medio ácido (0.1% HCOOH) para el análisis de nucleósidos. Incluso por primera vez, para evitar la oxidación de compuestos muy lábiles, se utiliza como agente reductor adicionado a la fase móvil de LC/MS: 1mM de ácido oxálico (previamente acondicionan el chip con una solución 2.5mM).

- Niñonuevo, Grim y Lebrilla utilizan un PGC Chip acoplado a un TOF para el análisis de oligosacáridos en diversas publicaciones; en una de ellas analizan alditoles de mucinas y leche humana (22); en otra en suero humano (23).

4.5. Análisis de pequeñas moléculas: metabolitos, fármacos,…

El HPLC-Chip tiene aplicabilidad a cualquier tipo de molécula ionizable por electrospray, que requiera los máximos niveles de sensibilidad, o cuando interese consumir la mínima cantidad posible para su análisis, como por ejemplo:

- Búsqueda de drogas o fármacos en pelo, donde el compuesto queda estabilizado, y permite un seguimiento de su ingesta a lo largo del tiempo (entre 1 semana y 3 meses); o de fármacos en pequeños animales.

- Flamini y Traldi utilizan un Chip de C18 –QTOF para el análisis de antocianinas en uva "Fast analysis of isobaric grape anthocyanins by Chip-liquid chromatography/ mass spectrometry" (24).

- Buckenmaier y Bonvie en "HPLC-Chip/Triple-Quadrupole MS for quantification of pharmaceuticals in diminishing small volumes of blood" utilizan un chip para el análisis de fármacos en "Dried Blood Spots" (DBS), analizando gotas de sangre secas depositadas en una pequeña cartulina. Éste dispositivo facilita y reduce enormemente los costes de la toma de muestra, su almacenamiento y transporte (25).

REFERENCIAS.

1. Rozing G, van de Goor T, Lauer HH, et al. An experimental study of chromatographic dynamics in open and packed non-cylindrical conduits. J. Sep. Sci. 2004; 27(17-18): 1391-1401. Disponible en: http://onlinelibrary.wiley.com/doi/10.1002/jssc.200401856/pdf

2. Yin H, Killeen K, Brennen R, van de Goor T, et al. Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. Anal Chem. 2005; 77(2):527-33. .Disponible en: http://pubs.acs.org/doi/pdf/10.1021/ac049068d

3. Fortier MH, E, Goodley P, Thibault P, et al. Integrated Microfluidic Device for Mass Spectrometry-Based Proteomics and Its Application to Biomarker Discovery Programs. Anal. Chem. 2005; 77: 1631-1640. Disponible en: http://pubs.acs.org/doi/pdf/10.1021/ac048506d

4. Barret W. Specificity of High-Abundant Protein Depletion Among Different Species with the Agilent Multiple Affinity Removal System. Agilent Technologies Application Note. 2004; nº publicación 5989-0957EN. Disponible en: http://www.chem.agilent.com/Library/applications/5989-0957EN.pdf

5. Video descriptivo del funcionamiento de la tecnología HPLC-Chip/MS. Disponible en: http://www.chem.agilent.com/en-US/Products/Instruments/lc/analytical/systems/1200serieshplc-chipms/pages/gp15389.aspx

6. Gauthier G,Côté L,Thibault P, et al. Advantages of using Intelligent Sample Loading with the HPLC-Chip for automated sample enrichment. Agilent Technologies Application Note. 2006; nº publicación 5989-5222EN. Disponible en: http://www.chem.agilent.com/Library/applications/5989-5222EN.pdf

7. Hardouin J, Duchateau M, Caron M, et al. Usefulness of an integrated microfluidic device (HPLC-Chip-MS) to enhance confidence in protein identification by proteomics. Rapid Commun. Mass Spectrom. 2006; 20: 3236–3244. Disponible en: http://homepage.mac.com/laigle/biomoceti/PDF/Hardouin_06_1.pdf

8. Staples GO, Costello CE, Zaia J, et al. Improved hydrophilic interaction chromatography LC/MS of heparinoids using a chip with postcolumn makeup flow. Anal. Chem. 2010; 82(2):516-22. Disponible en: http://pubs.acs.org/doi/abs/10.1021/ac901706f

9. Vollmer M, Miller C, Glatz B. Performance of the Agilent 1100 HPLC-Chip/MS System. Agilent Technologies Technical Overview. 2005; nº publicación 5989-4148EN. Disponible en: http://www.chem.agilent.com/Library/applications/5989-4148EN.pdf

10. Kilby GW.Noncovalently Associated Complexes: Apomyoglobin and Myoglobin Studies Using the HPLC-Chip/MS System Coupled with ESI-TOF Mass Spectrometry. Agilent Technologies Application Note. 2006; nº publicación 5989-5596EN. Disponible en: http://www.chem.agilent.com/Library/applications/5989-5596EN_LO.pdf

11. Raijmakers R, Mohammed S, Heck A, et al. Exploring the Human Leukocyte Phosphoproteome Using a Microfluidic Reversed-Phase-TiO2- Reversed-Phase High-Performance Liquid Chromatography Phosphochip Coupled to a Quadrupole Time-of-Flight Mass Spectrometer. Anal. Chem. 2010; 82: 824–832. Disponible en: http://pubs.acs.org/doi/abs/10.1021/ac901764g .

12. Raijmakers R, Mohammed S, Heck A. Facilitating Phosphopeptide Analysis Using the Agilent HPLC Phosphochip. Agilent Technologies Application Note. 2009; nº publicación 5990-4098EN. Disponible en: http://www.chem.agilent.com/Library/applications/5990-4098EN.pdf

13. Masana I, Raijmakers R, Heck A, et al. HPLC-FosfoChip una nueva tecnología para el análisis selectivo de Fosfopéptidos mediante LC/MS. Lifescienceslab. 2009; Jul-Ago:30-35. Disponible en: http://www.lifescienceslab.com/es/revistas-digitales/julio-agosto-2009/#/30/

14. Haruta M, Otegui M, Sussman MR, et al. Molecular Characterization of Mutant Arabidopsis Plants with Reduced Plasma Membrane Proton Pump Activity. The Journal of Biological Chemistry. 2010; 285 (23): 17918–17929. Disponible en: http://www.jbc.org/content/285/23/17918.full.pdf+html

15. Rasmussen JR, Davis J, Van Etten RL. Identification and derivatization of (oligosaccharyl)amines obtained by treatment of asparagines linked glycopeptides with N-Glycanase enzyme. J. Am. Chem. Soc. 1992; 114: 1124-1126. Disponible en: http://pubs.acs.org/doi/pdf/10.1021/ja00029a071

16. Bynum MA, Monel CR, Killeen K, et al. Characterization of IgG N-Glycans Employing a Microfluidic Chip that Integrates Glycan Cleavage, Sample Purification, LC Separation, and MS Detection. Anal. Chem. 2009; 81 (21): 8818–8825. Disponible en: http://pubs.acs.org/doi/pdfplus/10.1021/ac901326u

17. Web con pautas sobre la nomenclatura de glicanos: http://www.functionalglycomics.org/static/consortium/Nomenclature.shtml

18. Video ilustrativo sobre el funcionamiento del mAB-Glyco-Chip: http://www.chem.agilent.com/en-US/Products/Instruments/lc/analytical/systems/1200serieshplc-chipms/Pages/mAb_Glyco_video.aspx

19. Wollscheid B, Bausch-Fluck D, Henderson C, O'Brien R, Bibel M, Schiess R, Aebersold R, and Watts JD. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. Nat. Biotech. 2009; 27(4): 378-386. Disponible en: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2829300/pdf/nihms169130.pdf

20. Masana I. Caracterización de Anticuerpos Monoclonales (mAB's) y Biofármacos por HPLC y LC/MS. Lifescienceslab. 2011; Sep-Oct: 98-102. Disponible en: http://www.lifescienceslab.com/es/revistas-digitales/septiembre-octubre-2011/#/6/

21. Giessing AM, Scott LG, Kirpekar F. A Nano-Chip-LC/MS^n Based Strategy for Characterization of Modified Nucleosides Using Reduced Porous Graphitic Carbon as a Stationary Phase. J. Am. Soc. Mass Spectrom. 2011; 22: 1242-1251. Disponible en: http://www.springerlink.com/content/520tj376277r3576/

22. Niñonuevo M, Grimm R, Lebrilla C. Nanoliquid chromatography-mass spectrometry of oligosaccharides employing graphitized carbon chromatography on microchip with a high-accuracy mass analyzer. Electrophoresis. 2005; 26: 3641–3649. Disponible en: http://onlinelibrary.wiley.com/doi/10.1002/elps.200500246/abstract

23. Niñonuevo M, Grimm R, Lebrilla C. Profile of Native N-linked Glycan Structures from Human Serum Using High Performance Liquid Chromatography on a Microfluidic Chip and Time-of-Flight Mass Spectrometry. Proteomics. 2009; 9(7): 1939–1951. Disponible en: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2765869/

24. Flamini R, Rosso M, Traldi P, et al. Fast analysis of isobaric grape anthocyanins by Chip-liquid chromatography/mass spectrometry. Rapid Commun. Mass Spectrom. 2009; 23: 2891–2896. Disponible en: http://onlinelibrary.wiley.com/doi/10.1002/rcm.4199/pdf

25. Buckenmaier S, Emotte C, Bonvie A. HPLC-Chip/Triple-Quadrupole MS for quantification of pharmaceuticals in diminishing small volumes of blood. Agilent Technologies Application Note. 2008; nº publicación 5989-9896EN. Disponible en: http://www.chem.agilent.com/Library/applications/5989-9896EN.pdf

Preparación de la muestra. Fraccionamiento de subproteomas e inmunodepleción de proteínas mayoritarias

15

Joaquín Fernández-Irigoyen[1], Enrique Santamaría[1] y Fernando J. Corrales[2]

[1] Navarrabiomed-Fundación Miguel Servet;

[2] CIMA, Universidad de Navarra

1. Introducción.

El término proteómica fue introducido por primera vez en 1995, aunque el concepto apareció en la biología moderna unos años antes cuando Anderson y Anderson propusieron la creación de un atlas molecular de las proteínas humanas (1). La proteómica es el estudio del proteoma, que se refiere al conjunto de proteínas de un orgánulo, una célula, un tejido o un organismo completo.

El proteoma, podría definirse como el conjunto de proteínas expresadas por un genoma. En este sentido, sería relativamente sencillo deducir la secuencia de todas las proteínas a partir de la secuencia del genoma, aunque este trabajo dejaría aún sin asignación funcional aproximadamente al 30% de las proteínas codificadas. Además, la caracterización del proteoma humano presenta varios niveles adicionales de complejidad. El proteoma es tremendamente dinámico, no sería posible hablar de un proteoma único, a menos que la definición inicial se complemente con parámetros tanto genéticos como ambientales que van a condicionar en gran medida el repertorio de especies proteicas que lo constituyen. La plasticidad del proteoma es esencial para regular con precisión las funciones celulares mediante mecanismos de control complementarios a los meramente transcripcionales que permitan reacciones rápidas y eficaces a estímulos externos y, así, asegurar respuestas adaptativas que garanticen la supervivencia. Existen diversos mecanismos de regulación postranscripcional que hacen necesario revisar, o al menos matizar, el axioma un gen, una proteína. Los tránscritos se pueden procesar de formas alternativas generando especies proteicas diferentes. Además, las proteínas son susceptibles de sufrir modificaciones postraduccionales (se han descrito más de 300 tipos diferentes) que dan lugar a diferentes formas de la proteína codificada por un único gen. Se estima que el número de especies proteicas por gen es, en humanos, entre 6 y 8. Si asumimos que una célula expresa de 5.000 a 10.000 genes, el proteoma de esta célula estaría constituido por 40.000-80.000 proteínas y, en general, el proteoma humano integraría cientos de miles. Finalmente, es difícil imaginarse hoy día una proteína realizando su actividad biológica de forma aislada en el entorno celular, sino de forma coordinada con el resto de biomoléculas, a través de interacciones reversibles.

A pesar de los espectaculares avances tecnológicos que proveen instrumentos con capacidades impensables hace pocos años, la complejidad del proteoma es inabarcable en toda

su extensión, quizás con la excepción de proteomas procarióticos más sencillos. La filosofía de "divide y vencerás" es el camino a seguir para incrementar la profundidad en el estudio del conjunto de proteínas de un sistema biológico. El fraccionamiento del proteoma permite un análisis más eficaz de cada una de las fracciones generadas, incrementa significativamente el rendimiento del análisis y permite la detección de proteínas minoritarias. Existen diversos métodos de fraccionamiento basados en las diferentes propiedades de las proteínas y péptidos, cuya combinación permite incrementar hasta 3-4 órdenes de magnitud el rango dinámico analizable (2). El fraccionamiento subcelular y la eliminación de proteínas mayoritarias, junto con el enriquecimiento de determinados péptidos/proteínas, son métodos comúnmente utilizados por su relativa sencillez y robustez.

2. Preparación de la Muestra.

La preparación de la muestra es uno de los procesos más importantes en el campo de la proteómica, y de ello depende la calidad y relevancia de los datos obtenidos en el experimento. La recogida de la muestra, su transporte y almacenamiento, así como la extracción de proteínas, son procedimientos tan importantes como el posterior análisis mediante espectrometría de masas. Las muestras de proteínas utilizadas en experimentos de proteómica son obtenidas a partir de células en cultivo, tejido, fluidos biológicos, etc (3). Cada uno de estos tipos de muestra contiene otros componentes como lípidos, ácidos nucleicos, azúcares y otras pequeñas moléculas. Por otro lado, las proteínas, a diferencia del ADN y ARN, son moléculas bioquímicamente muy heterogéneas en tamaño, carga, hidrofobicidad, etc. Además, las proteínas pueden sufrir modificaciones postraduccionales. Esta variedad bioquímica dificulta enormemente el desarrollo de un método que permita la extracción y posterior análisis de todas las proteínas con igual eficacia. La utilización de compuestos, como detergentes, sales o determinados tampones durante la ruptura celular y solubilización de las proteínas, pueden interferir durante los siguientes pasos del proceso. Es fundamental, por lo tanto, elaborar procedimientos robustos y eficaces que aseguren el máximo rendimiento de los procesos analíticos a utilizar (4). Los protocolos deben, además, responder a los requerimientos específicos de los diferentes tipos de muestra biológica que se pretende investigar.

- **Tejidos:** En primer lugar, durante la extracción, es importante evitar la mezcla de diferentes entidades biológicas. Por ejemplo, en el caso de muestras tumorales, deben extraerse libres de estroma, sangre, tejido adyacente, etc., y, en la medida de lo posible, contener poblaciones de células tumorales homogéneas (5,6). La muestra extraída es homogeneizada en presencia de un tampón de lisis adecuado y centrifugada para eliminar los ácidos nucleicos y restos de material insoluble. De esta manera, se obtiene un homogenizado de proteínas en el sobrenadante que podrá ser alicuotado y congelado (7). En algunos casos, como las biopsias, las muestras no pueden ser procesadas tras su obtención y son congeladas en nitrógeno líquido y almacenadas a -80ºC hasta su procesamiento (8). Los tejidos vegetales, así como el estudio de microorganismos, presentan ciertas peculiaridades debido a la presencia de pared celular, como se comentará más adelante.

- **Cultivos celulares:** La principal ventaja de trabajar con líneas celulares es su homogeneidad biológica y la posibilidad de controlar sus condiciones de cultivo (9). Para la obtención del extracto proteico, el cultivo celular se lava, centrifuga y, sobre el precipitado celular, se añade el tampón de lisis en el que se extraen las proteínas a través de ciclos de sonicación. Sin embargo, algunos autores han descrito que este método puede introducir cierta variabilidad en las muestras (10). Para evitarlo, el tampón de lisis puede añadirse directamente sobre el cultivo celular (3).

- **Fluidos biológicos:** Los fluidos biológicos son relativamente fáciles de obtener por métodos no invasivos, y por ello se utilizan de manera habitual en el ámbito clínico para el diagnóstico de una gran variedad de patologías. En el caso de la sangre, los métodos de extracción de proteínas dependen del propósito del análisis. El suero es la fracción de la sangre libre de componentes celulares obtenida tras su coagulación y posterior centrifugación. El plasma es la fracción obtenida mediante el mismo proceso, pero añadiendo previamente diferentes anticoagulantes. El principal problema de trabajar con suero y plasma es el amplio rango dinámico de las proteínas presentes en estas muestras (11). Proteínas como la albúmina, las inmunoglobulinas o la haptoglobina representan hasta un 80% de la proteína sérica total, dificultando el estudio de las proteínas que se encuentran en menor concentración. Por ello, se han desarrollado deiversos métodos para deplecionar las proteínas mayoritarias antes de los análisis proteómicos (12). La HUPO (The Human Proteome Organization) ha propuesto unas recomendaciones para la preparación de este tipo de muestras antes de su análisis proteómico. Tras realizar un estudio piloto en diversos laboratorios utilizando varios métodos de obtención de muestra, fraccionamiento y depleción de proteínas mayoritarias, la HUPO recomienda la utilización de plasma en vez de suero, obtenido mediante la adición de EDTA (o citrato) como anticoagulante. Para mejorar la resolución, sensibilidad y reproducibilidad en la identificación de proteínas, la HUPO recomienda un protocolo combinado de depleción de proteínas mayoritarias, fraccionamiento y análisis mediante espectrometría de masas (13). Otro fluido biológico de interés clínico es el líquido cefalorraquídeo. Su composición es muy similar a la del plasma sanguíneo, aunque la cantidad de proteína es menor (14). Los problemas asociados a la preparación de la muestra son similares a los descritos para el plasma o suero sanguíneo, con el añadido de que en el momento de su obtención la muestra puede contaminarse con sangre (15). Existen diversos trabajos en los que se compara y discute los diferentes métodos de obtención y preparación de la muestra, su análisis y la interpretación de resultados (16,17).

La extracción de proteínas implica la ruptura tisular y la lisis celular. Los principales métodos empleados en el campo de la proteómica incluyen la homogeneización mecánica (homogeneizadores de Dounce y Potter-Elvehjem), sonicación y procesos de congelación/descongelación (18). Además, se utilizan reactivos como los detergentes, que han demostrado su eficiencia en los procesos de extracción de proteínas (3). La elección del método depende del tipo y cantidad de muestra inicial y de los procedimientos analíticos que se emplearán en el estudio: electroforesis, métodos cromatográficos y espectrometría de masas.

En el caso de células vegetales, además, hay que tener en cuenta la presencia de pared celular, compuesta principalmente de celulosa y sus derivados. Para romper dicha estructura, se utilizan diversos métodos químicos y físicos, como determinados tampones de lisis, sonicación por ultrasonidos, ciclos de congelación y descongelación y procesos de centrifugación (19). Algunas bacterias y levaduras también presentan pared celular y cápsulas de polisacáridos (20). Aunque ciertas bacterias se lisan en presencia de tampón de lisis, en general se requiere de procesos mecánicos, como la sonicación, o digestiones enzimáticas (3).

La ruptura de las estructuras celulares provoca la liberación de enzimas hidrolíticas, principalmente proteasas, que actúan degradando proteínas. Aunque su actividad se reduce en presencia de agentes desnaturalizantes, es necesaria la utilización de inhibidores de proteasas en los tampones de lisis (18,21). Generalmente se emplean mezclas de inhibidores con un amplio espectro de inhibición como el fluoruro de fenilmetilsulfonilo (PMSF en inglés) o el fluoruro de bencenosulfonilo (AEBSF en inglés).

En general, la lisis celular requiere de la combinación de detergentes y métodos físicos. Existen numerosos libros y manuales de protocolos en los que se presentan diferentes estrategias para la extracción de proteínas a partir de bacterias, levaduras, plantas y tejidos animales (22-

24). Además, hoy en día existen kits comerciales para la extracción de proteínas a partir de numerosos tipos celulares, que incluyen mezclas de detergentes, inhibidores de proteasas y dispositivos mecánicos (3).

3. Extracción y solubilización de proteínas.

Las proteínas presentan una gran heterogeneidad bioquímica y, además, pueden aparecer integradas en la membrana plasmática o formando complejos con otras proteínas o ácidos nucleicos, dificultando la extracción eficaz y simultánea de todas las proteínas presentes en una muestra biológica (3). La solubilidad de las proteínas implica la ruptura de las interacciones que se producen inter e intra proteínas, tales como puentes de hidrógeno, puentes disulfuro, fuerzas de Van der Wals e interacciones iónicas e hidrofóbicas (25,26).

En 1975, O'Farrell propuso el primer tampón de extracción de proteínas (27). Sin embargo, este tampón no era capaz de solubilizar proteínas de membrana, alcalinas o hidrofóbicas (28). El desarrollo de nuevas estrategias de solubilización en los últimos años ha permitido la identificación de proteínas hasta entonces no caracterizadas, como por ejemplo proteínas de membrana (29,30); proteínas con pI extremos (31-33), proteínas de bajo/alto peso molecular (34,35) o proteínas poco abundantes (36). En cualquier caso, no existe un método universal que permita la extracción y solubilización de cualquier tipo de proteína.

Con el objetivo de facilitar la extracción y evitar la formación de modificaciones químicas, agregados proteicos o procesos de precipitación, la solubilización implica la utilización de agentes caotrópicos (urea y/o tiurea), detergentes (CHAPS o Triton), agentes reductores (DTT o TEAB) e inhibidores de proteasas (37). La utilización de estos agentes, junto con un método adecuado de lisis celular y concentración de la muestra, determina la efectividad de la solubilización (3).

Los agentes caotrópicos son sustancias capaces de romper los puentes de hidrógeno intramoleculares o formados entre diferentes cadenas polipeptídicas . A altas concentraciones, son capaces de desnaturalizar y solubilizar la mayor parte de proteínas presentes en una mezcla compleja. Los agentes caotrópicos más utilizados son la urea y el cloruro de guanidinio (a concentraciones entre 5 y 9 M). Cabe destacar otro agente caotrópico, la tiourea. Aunque su estructura es muy similar al de la urea (la única diferencia es la sustitución del átomo de oxígeno por uno de azufre) sus propiedades difieren significativamente. Se ha comprobado que la adición de este compuesto incrementa significativamente la solubilización de proteínas de membrana (38). La urea, con el aumento de la temperatura, puede hidrolizarse dando lugar a cianato e isocianato. La formación de estos aniones induce modificaciones en algunos residuos, como la carbamilación, que resulta en la formación de homocitrulina a partir de lisina. Por ello, la solución no debe calentarse por encima de 37ºC (3).

Los detergentes son moléculas anfipáticas capaces de romper las interacciones hidrofóbicas que se establecen entre las proteínas y/o con otros componentes celulares, como en el caso de las proteínas de membrana. Los detergentes son moléculas anfipáticas que poseen un grupo apolar (compuesto generalmente por una cadena de hidrocarburos) y uno polar. En función de la carga de este último, los detergentes se clasifican en iónicos (como el sodio dodecilsulfato, SDS) o no iónico (como el 3-[(3-colamidopropil) dimetilamonio] propanosulfónico, CHAPS). Los tampones de extracción de proteínas contienen detergentes a concentraciones que varían entre 1 y 4%, valores que se deben ajustar según el tipo de muestra y teniendo en cuenta la posible interferencia con los análisis posteriores. El CHAPS es el detergente más empleado en estudios de proteómica debido a su alta solubilidad en presencia de urea (3).

Los agentes reductores son capaces de romper los puentes disulfuro creados entre residuos de cisteína, facilitando la solubilidad de las proteínas. Los más empleados son el ditiotritol (DTT) y ditioeritritol (DTE) en concentraciones entre 20 y 100 mM. El inconveniente de estos 2 agentes reductores es que, al estar cargados, pueden migrar durante el proceso de isoelectroenfoque creando un gradiente a lo largo del proceso, produciendo la reoxidación de los grupos sulfidrilo y pérdida de solubilidad de algunas proteínas (39). Para evitar este efecto, se emplean como agentes reductores diferentes derivados de fosfinas, como la tributilfosfina o el triscarboxietilfosfina (TCEP) a concentraciones alrededor de 2 mM. La utilización de estos agentes, que no poseen carga, incrementa la solubilidad de las proteínas durante los procesos de isoelectroenfoque. Además, a diferencia del DTT, no interaccionan con la acrilamida o la 4-vinilpiridina (40). El B-mercaptoetanol se utiliza menos debido a que requiere mayores concentraciones que el DTT o TECP y puede producir alteraciones en el proceso de isoelectroenfoque (41). Tras el proceso de reducción, los grupos sulfidrilo reducidos deben ser protegidos frente a nuevos procesos de oxidación. Para ello existen diferentes agentes alquilantes, como el (metiltio)dimetil sulfóxido (MMTS) o la 4-vinilpiridina, aunque el más utilizado es la Iodoacetamida (IAA), ya que su unión a los grupos –SH libres no altera la carga ni el pI de la proteína (3).

Un factor clave en la solubilidad de las proteínas es el pH del tampón empleado. Las proteínas son menos solubles a pH cercanos a su punto isoeléctrico, pI (pH al que una proteína tiene carga neta cero) (42). En función del pI, las proteínas se distribuyen en 2 grandes grupos, cuyos pI corresponden a valores de pH 5,5 y 9,5, aunque existen otros grupos menos numerosos a pH 8 y 12. Pocas proteínas tienen su pI entre 7 y 8, que es el pH fisiológico (43). Los tampones se utilizan para controlar el pH de la solución y protegerla frente a cambios de pH. Los tampones son más efectivos a ± 0.5 unidades de pH de su constante de disociación (pKa). La capacidad de tamponamiento es una medida de la protección frente a cambios de pH y depende de la concentración del tampón, que suele oscilar entre 25-50 mM. Los diferentes tampones se caracterizan por sus valores de pKa, relación pKa/temperatura (ΔpKa/°C), carga, solubilidad y otras características tales como capacidad de unión a iones metálicos, grupos reactivos (por ejemplo, aminas primarias), volatilidad, etc. (44).

Tras la obtención del extracto proteico, a menudo es necesario eliminar componentes de la muestra biológica, como el ADN, y algunos de los reactivos utilizados para la solubilización de las proteínas, como los detergentes, que pueden interferir en los siguientes pasos. El método más empleado es la precipitación selectiva de proteínas utilizando acetona, ácido tricloacético, cloroformo/metanol, etanol, isopropanol o alguno de los kits, basados en los procedimientos anteriormente mencionados, que existen para este propósito en el mercado (45-47). Los solventes orgánicos favorecen la precipitación porque disminuyen la constante dieléctrica de la solución, lo que resulta en interacciones más fuertes entre las cadenas polipeptídicas y, además, remplazan parte del agua que hidrata las moléculas de proteína. El precipitado proteico puede ser resuspendido en el tampón de apropiado para los siguientes pasos y en un volumen menor, lo que permite, además, concentrar la muestra. Es importante tener en cuenta que cualquier proceso de precipitación genera una pérdida parcial de muestra (puede llegar a un 20%, en función del tipo de muestra y el método empleado), por lo que se recomienda volver a cuantificar la proteína tras el proceso.

4. Fraccionamiento de Proteínas y Péptidos.

Una vez obtenido el extracto proteico, el siguiente paso es hidrolizar las proteínas en péptidos. Aunque existen diversos métodos químicos para la hidrólisis de proteínas, se ha generalizado el empleo de enzimas proteolíticas debido a su especificidad, como la tripsina que corta en extremo C terminal de residuos de arginina y lisina o la Lys-C en residuos de lisina. La utilización de estas enzimas tiene varias ventajas. Por un lado, el tamaño de los péptidos generados es más o menos homogéneo, facilitando su análisis mediante espectrometría de masas. Por otro lado, la especificidad en los puntos de corte permite predecir los péptidos generados a partir de una determinada secuencia, lo que facilita enormemente la identificación de péptidos y proteínas a partir de los datos generados mediante espectrometría de masas. Además, todos los péptidos generados tienen un residuo básico en posición C terminal, lo que incrementa su densidad de carga positiva a pH ácido, facilitándose su ionización en los espectrómetros de masas.

Los métodos de fraccionamiento permiten subdividir la complejidad del proteoma que se pretende analizar con el objetivo de alcanzar una mayor cobertura mediante el estudio de las fracciones obtenidas. Los procesos de fraccionamiento pueden realizarse a nivel de proteína o péptido, más uniformes desde un punto de vista fisicoquímico, o combinarse (48). Los métodos tradicionales de separación y purificación de proteínas y péptidos se han desarrollado durante los últimos años como herramientas aplicadas al estudio bioquímico de proteínas específicas (49). Desde sus inicios, la proteómica ha utilizado estos métodos convencionales como herramientas de prefraccionamiento previo a los análisis de identificación masiva de proteínas mediante espectrometría de masas (50), entre los que destacan los métodos cromatográficos y electroforéticos. La combinación de 2 o más métodos (fraccionamientos multidimensionales) y la separación selectiva (mediante uniones de afinidad) han demostrado ser estrategias que aumentan la eficacia de los análisis. El término *Shotgun Proteomics* fue acuñado por Wolters en 2001 (51) y hace referencia a la metodología que incluye la digestión proteolítica de un proteoma, la separación, generalmente cromatográfica, de los péptidos generados y su identificación/cuantificación en base a los parámetros obtenidos mediante espectrometría de masas.

4.1. Técnicas electroforéticas

4.1.1. Geles SDS-PAGE e Isoelectroenfoque

Uno de los métodos más utilizados para la separación de proteínas son los geles SDS-PAGE y el isoelectroenfoque, tratados extensamente en el capítulo 5 de este manual. Sin embargo, cabe destacar el isoelectroenfoque en fase líquida, que permite la separación de proteínas en base a su pI sin necesidad de cargarlas en un gel de acrilamida. De esta manera se evita el proceso de extracción de las proteínas del gel para su caracterización, aumentando el rendimiento del proceso. El primer dispositivo capaz de realizar un isoelectroenfoque en fase líquida fue desarrollado por Bio-Rad (Rotofor), y permite la separación de proteínas en base a su pI a través de un gradiente de pH generado por anfolitos a través de 20 compartimentos separados por pantallas de nilon permeables (28,52).

4.1.2. Electroforesis de flujo libre

La electroforesis de flujo libre es un procedimiento electroforético continuo en ausencia de una fase estacionaria, como un gel de acrilamida. La electroforesis de flujo libre separa partículas cargadas de tamaño muy diverso, desde pequeñas moléculas hasta células intactas, de acuerdo a su carga, movilidad o pI (Figura 1) (53). La muestra se aplica utilizando una bomba peristáltica y se inyecta en una cámara estrecha limitada por 2 láminas de vidrio. Mediante flujo laminar, la muestra se transporta a través de una fina película de medio acuoso formado entre las 2 láminas de vidrio. Estas láminas contienen 2 electrodos que generan un campo eléctrico

perpendicular al flujo laminar. Las partículas cargadas (proteínas, péptidos, orgánulos, fragmentos de membrana o células intactas) son proporcionalmente desviadas en función de su carga, permitiendo la separación y fraccionamiento en el lado opuesto del sistema (54). La electroforesis de flujo libre puede operar de 3 modos diferentes:

- **Isoelectroenfoque preparativo:** Mediante un gradiente de pH perpendicular al flujo laminar, las partículas se separan en base a su pI. Se utiliza para el fraccionamiento de proteínas a alta resolución, con rendimientos cercanos al 98% y cantidades de hasta 50mg a la hora.

- **Electroforesis zonal:** No existe gradiente de pH y las partículas son separadas en base a su carga neta. Se utiliza para la separación rápida de orgánulos, membranas y células intactas.

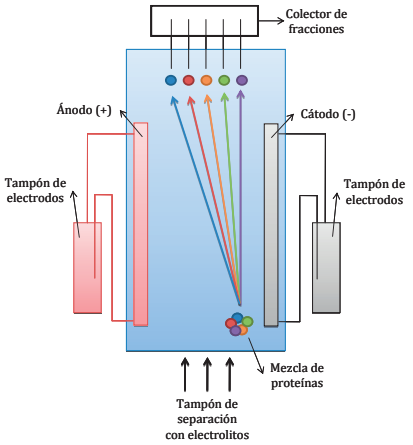


Figura 1. Representación esquemática de los principios básicos que rigen la separación mediante electroforesis de flujo libre.

- **Isotacoforesis:** En este caso se genera un gradiente de pH (como en el isoelectroenfoque preparativo) y, además, el campo eléctrico perpendicular es discontinuo (a diferencia de la electroforesis convencional donde se aplica un campo eléctrico continuo). Se utiliza para la separación de partículas basándose en la movilidad electroforética de las mismas.

4.1.3. Electroforesis Capilar

La electroforesis capilar engloba un conjunto de técnicas que utilizan capilares de sílice fundido en presencia de un campo eléctrico, y se emplea para la separación tanto de proteínas como de péptidos (55). La electroforesis capilar, de manera similar a la electroforesis de flujo libre, funciona en fase líquida y puede operar en modos similares: electroforesis zonal capilar, isoelectroenfoque capilar, e isotacoforesis capilar (56). Las principales ventajas de la utilización de la electroforesis capilar sobre la cromatografía líquida y otros métodos de fraccionamiento electroforético son varias: alta eficiencia de separación, alta resolución, velocidad, bajo coste de los capilares, compatibilidad con un amplio rango de tampones y facilidad para la automatización (56,57).

4.2. Métodos cromatográficos

Desde los inicios de los estudios proteómicos, se han utilizado las aproximaciones clásicas de cromatografía para el fraccionamiento de extractos proteicos en base a diferentes propiedades fisicoquímicas, como la carga neta (cromatografía de intercambio iónico), masa molecular de proteínas y complejos proteicos (exclusión molecular), hidrofobicidad (cromatografía de interacción hidrofóbica) y afinidad por determinados ligandos (cromatografía de afinidad). Todos estos métodos han sido tratados en el tema 4 de este manual.

Entro ellos cabe destacar la cromatografía de afinidad basada en modificaciones postraduccionales. Tales modificaciones son esenciales para la regulación de procesos celulares como la señalización, diferenciación o crecimiento (58,59). Por ejemplo, en el caso de la fosforilación, se han desarrollado columnas de afinidad con cationes metálicos inmovilizados (Fe^{3+}, por ejemplo) que se unen de manera específica a los grupos fosfato de péptidos o proteínas (59). En el caso

de la glicosilación, se utilizan la concanavalina A y/o lectina como ligandos para unir de manera específica glicopéptidos (60).

4.3. Fraccionamiento subcelular

Una alternativa a los métodos clásicos de fraccionamiento de proteínas a partir de extractos celulares es el fraccionamiento subcelular. Este tipo de fraccionamiento, que consiste en la separación de los diferentes orgánulos que componen la célula, tiene 2 grandes ventajas. Por un lado, la complejidad del proteoma de un orgánulo es sensiblemente inferior al de una célula completa (61) y, por otro, la caracterización de las proteínas que componen un determinado orgánulo permite profundizaren su localización subcelular (62).

Hasta hace relativamente pocos años, los estudios de localización subcelular de proteínas se realizaban mediante técnicas de inmunodetección y/o proteínas de fusión (62). Estas metodologías dependen de la disponibilidad de anticuerpos específicos y/o proteínas recombinantes de fusión para la detección mediante Western Blot y microscopía. Aunque estas técnicas han permitido la caracterización de muchos procesos celulares específicos, su capacidad está lejos de cubrir la globalidad del mapa funcional de una célula.

El fraccionamiento de orgánulos implica una lisis celular que permita mantener la integridad física de los diferentes compartimentos subcelulares que la componen. Para ello, existen diferentes protocolos en función del tipo celular y orgánulos de interés (63-65). La separación de los diferentes componentes subcelulares se realiza mediante aproximaciones clásicas, como la centrifugación en gradientes de densidad, basada en propiedades como el tamaño, carga, densidad y otros (66,67).

La aplicación de las técnicas de proteómica, especialmente la espectrometría de masas, en el estudio de subproteomas celulares, se basa en el enriquecimiento de fracciones subcelulares. La Proteómica substractiva, el Perfil de Correlación de Proteínas y la Localización de proteínas en orgánulos mediante marcaje isobárico, son los 3 protocolos mejor establecidos que han permitido profundizar en el estudio de subproteomas celulares.

4.3.1. Métodos clásicos de fraccionamiento Subcelular

4.3.1.1. Centrifugación diferencial

La centrifugación diferencial se basa en la aplicación de centrifugaciones secuenciales a partir de un homogeneizado celular o tisular. Este método permite obtener diferentes orgánulos en 4 fracciones subcelulares: núcleo; mitocondrias, lisosomas y peroxisomas; microsomas (vesículas obtenidas a partir de la ruptura del retículo endoplasmático); y ribosomas y citosol (65). La separación de los diferentes orgánulos depende en la masa, forma y volumen espacial de cada uno de ellos, así como en la densidad del tampón empleado, el tipo del rotor y la velocidad de rotación (68). La principal desventaja de este método es su baja resolución, pudiendo dar lugar al solapamiento de fracciones originando mezclas de diferentes orgánulos con velocidades de sedimentación parecidas (69).

4.3.1.2. Centrifugación en gradiente de densidad

Es el método más eficiente en la separación de componentes subcelulares (64). Se basa en la separación de orgánulos en función de su densidad de flotación. La muestra se dispone por encima o se mezcla con un gradiente de densidad que contiene una concentración muy elevada de sacarosa o de cloruro de cesio. Durante la centrifugación, cada orgánulo se desplaza hasta que su densidad iguala la del medio circulante (punto isopícnico) (70).

La sacarosa es el medio más empleado para la centrifugación en gradiente de densidad, debido a que es biológicamente inerte, dializable y barata. Otros medios utilizados son el Ficoll, el Percoll, el Nycodenz y la Metrizamida. Los gradientes pueden ser autogenerados, como es el

caso del cloruro de cesio, ya que se forman en el propio proceso de centrifugación, o preformados como es el caso de la sacarosa o del percoll. En este último caso la preparación se realiza antes de la centrifugación utilizando un formador de gradientes (71).

Una alternativa es la centrifugación en gradiente discontinuo, en la que se crea un gradiente por tramos en el interior de un tubo al apilar sucesivamente volúmenes homogéneos de soluciones de diferente densidad. En las interfases de las diferentes capas se acumularán las partículas de una determinada densidad. Durante la centrifugación en gradiente discontinuo, los diferentes orgánulos se enriquecen en las diferentes interfases del gradiente mientras que en el gradiente continuo, se distribuyen a lo largo de todo el medio, donde encuentran su punto isopícnico, lo que proporciona una mayor resolución. Sin embargo, la preparación de gradientes continuos tiene la desventaja de la dificultad de su preparación previa y que requiere de aparatos especiales.

4.3.1.3. Fraccionamiento diferencial mediante detergentes

Una alternativa a los métodos de fraccionamiento basados en centrifugaciones seriadas, es el empleo de tampones con diferentes detergentes, cuyas ventajas son reducir las contaminaciones de otros orgánulos y reducir el tiempo de fraccionamiento empleado en las centrifugaciones seriadas (69). Se trata de un método sencillo, capaz de fraccionar proteínas en su estado nativo de acuerdo a 4 localizaciones subcelulares: citosol; membrana celular y de orgánulo; soluble y asociada a DNA; y citoesqueleto. Este método está disponible como kit comercial: ProteoExtract Subcellular Proteome Extraction Kit.

4.3.1.4. Electroforesis de flujo libre

Aunque este método ya ha sido explicado anteriormente en este mismo capítulo, cabe destacar su utilización para el fraccionamiento de componentes subcelulares, en concreto el modo de electroforesis zonal. Las ventajas de utilizar este sistema es que facilita la purificación de los orgánulos intactos, lo que permite realizar estudios funcionales. Por el contrario, una de las principales desventajas es la baja resolución en la separación de componentes con un punto isoeléctrico similar. Sin embargo, la disponibilidad de diferentes soluciones de electrolitos permite poner a punto protocolos para la purificaciónorgánulos específicos (54).

4.3.1.5. Purificación mediante inmunoafinidad

La purificación de orgánulos basada en métodos de inmunoafinidad permite la obtención de fracciones de manera específica y con altos rendimientos en cantidad de proteína (72). Este método se basa en la unión específica entre ligandos (anticuerpos) inmovilizados en un soporte sólido y los orgánulos de interés. Existen modificaciones para el tipo de ligando, y se han utilizado también ligandos químicos, péptidos, aptámeros y otros (49). Las limitaciones de esta técnica están relacionadas con su alto coste, debido al empleo de anticuerpos, y el incremento en la duración de los experimentos requerido para una correcta purificación.

4.3.1.6. Clasificación de orgánulos asistida por fluorescencia

La clasificación de células asistida por fluorescencia (FACS, fluorescence assisted cell sorting) se emplea de manera rutinaria para el enriquecimiento selectivo de poblaciones celulares que poseen marcadores específicos (73,74). El término “Clasificación de orgánulos asistida por fluorescencia” (FAOS, fluorescence assisted organelle sorting) hace referencia a la utilización de la citometría de flujo para la clasificación de estructuras subcelulares y otros componentes (75,76). Para llevar a cabo esta técnica es necesaria una proteína marcadora del orgánulo de interés. Este marcador puede ser una proteína recombinante unida a la proteína fluorescente GFP (77). De manera alternativa, si no se dispone de un marcador específico del orgánulo, el FAOS puede ser utilizado de manera análoga al FACS para la clasificación de células (69). Esta metodología se ha empleado para el fraccionamiento de diferentes orgánulos, como los fagosomas (78), mitocondria (79) y vesículas endocíticas (80).

4.3.2. Estrategias para el estudio del proteoma de orgánulos subcelulares

4.3.2.1. Proteómica Substractiva

La proteómica substractiva surge como herramienta para solventar los problemas de contaminación en los procesos de enriquecimiento. Esta aproximación compara las proteínas presentes en 2 fracciones subcelulares enriquecidas en paralelo a partir del mismo extracto celular. Analizando la complementariedad entre ambas, es posible deducir las proteínas que forman parte de un determinado orgánulo (81). La proteómica substractiva se ha empleado en aquellos casos en los que, sin conseguir un purificación completa del orgánulo, se ha podido enriquecer de manera significativa (61). Mediante el análisis con espectrometría de masas de cada una de las fracciones, se obtienen 2 listas de proteínas, la correspondiente a la fracción de interés y a otra fracción control procesada de igual manera. De la lista de proteínas obtenida, se descartarán aquellas que aparezcan también en la segunda fracción. De esta forma se descartan las proteínas, en teoría, no específicas del orgánulo de interés. Un ejemplo de aplicación es el empleado por Yates y colaboradores, cuyo objetivo era describir las proteínas constituyentes de la envoltura nuclear (82). Para ello prepararon 2 fracciones: una de ellas constituida por envueltas nucleares contaminadas con proteínas de la membrana mitocondrial y la otra por una fracción microsomal, rica en proteínas del retículo endoplasmático, proteínas mitocondriales pero libre de membranas nucleares. Para obtener las proteínas que constituyen de manera específica la envoltura nuclear, sustrajeron las proteínas obtenidas de la fracción microsomal a las obtenidas a partir de la fracción de la envoltura nuclear. De esta manera, los autores identificaron 67 potenciales proteínas constituyentes de la envoltura nuclear, sumándose a las 13 ya previamente conocidas. De todas ellas, 13 fueron confirmadas mediante ensayos de inmunofluorescencia utilizando anticuerpos específicos. Es importante resaltar que esta aproximación no sirve para aquellas proteínas que colocalizan en diferentes compartimentos subcelulares (62). De manera similar, aquellas proteínas que residan en un orgánulo de manera transitoria, serán identificadas en función del momento del ciclo celular en el que se analice la muestra (62).

Una variante de esta aproximación es el empleo de marcajes isotópicos, como ya se ha hecho con iTRAQ (83), SILAC (84) o ICAT (85). Las proteínas de las 2 fracciones a comparar se marcan con isótopos pesados/isobáricos diferentes. Mediante un análisis semicuantitativo, es posible distinguir si una determinada proteína está realmente presente en una de las fracciones o si su presencia se debe a un proceso de contaminación durante el fraccionamiento subcelular. Si la relación de las intensidades de los iones reporteros es grande, la presencia de la proteína en la fracción con menor señal puede deberse a una contaminación durante el proceso de enriquecimiento. Si la relación es cercana a la unidad, la proteína probablemente se encuentra colacalizada en ambas fracciones.

A pesar de la indiscutible utilidad de estas estrategias, existen algunas limitaciones para esta metodología. En primer lugar, los resultados no dependen únicamente de las proteínas identificadas en el orgánulo de interés, sino también de aquellas identificadas en la fracción a comparar. De esta forma, a diferencia del ejemplo expuesto en esta apartado, no siempre está claro qué fracción escoger como control. Esta fracción debe constar de las proteínas que pueden aparecer como contaminantes en la fracción de interés y, además, tiene que poder enriquecerse el compartimento subcelular al que pertenecen. Por otro lado, en el caso de fracciones subcelulares formadas por mezclas complejas de proteínas, no siempre es posible identificar todas ellas. Esto puede dar lugar a la aparición de falsos positivos o negativos, en función de la complejidad de la fracción utilizada como control (81). Finalmente, es importante resaltar que la proteómica substractiva se focaliza únicamente en un orgánulo específico, sin integrar la información de los proteomas de diferentes orgánulos de manera global.

4.3.2.2. Perfil de Correlación de Proteínas

El desarrollo de la espectrometría de masas permite realizar estudios semicuantitativos a partir de los datos obtenidos para la identificación de proteínas. Gracias a ello, es posible comparar la

abundancia relativa de una proteína en diferentes fracciones en base a los datos obtenidos por espectrometría de masas. El Perfil de Correlación de Proteínas (PCP) se basa en la monitorización de la distribución de las proteínas a lo largo de las fracciones obtenidas mediante aproximaciones clásicas, como la centrifugación en gradiente de densidad. Este método se basa en que aquellas proteínas que se localizan en un mismo orgánulo deberían tener el mismo perfil de abundancia a lo largo de las diferentes fracciones. Dichos perfiles se comparan y se agrupan por similitud con marcadores específicos de orgánulos que permiten deducir la localización de cada proteína (Figura 2) (86). La aproximación PCP funciona de manera similar a los métodos tradicionales, en los que se realiza un primer fraccionamiento y se monitoriza la presencia de determinadas proteínas a lo largo de las fracciones, pero en lugar de utilizar anticuerpos para proteínas específicas mediante inmunodetección, las proteínas se cuantifican mediante espectrometría de masas.

En primer lugar, los diferentes orgánulos son separados mediante centrifugación en gradiente de densidad (u otro método). Las diferentes fracciones son analizadas con técnicas de inmunodetección frente a marcadores de orgánulos, proteínas cuya localización ha sido previamente descrita y asociada a un único orgánulo. Cada uno de estos marcadores tendrá un perfil diferente, con un máximo en una fracción concreta que permitirá deducir la distribución de los orgánulos a lo largo del gradiente (86). El perfil de cada uno de estos marcadores se realiza mediante distribuciones X^2 (87). Una vez establecidos los perfiles de cada uno de los orgánulos, se evalúa la distribución de todas las proteínas identificadas mediante espectrometría de masas a lo largo de las diferentes fracciones mediante criterios estadísticos más restrictivos que para los marcadores. De esta manera, para cada proteína identificada se obtendrá un perfil en el que podrá valorarse su localización gracias a la distribución obtenida a lo largo de todo el gradiente. La abundancia de cada proteína en cada una de las fracciones se calcula a partir del área de los respectivos iones en el análisis mediante espectrometría de masas.

La principal ventaja de este método es que, mediante los estudios estadísticos de los perfiles, sólo se tienen en cuenta aquellas proteínas que muestran un comportamiento significativo, evitando aquellos falsos positivos debidos a procesos de copurificación. Por el contrario, se requieren estudios alternativos que permitan la confirmación de los resultados, así como una comparación con bases de datos de localización subcelular (86).

La primera aplicación de este método fue para el estudio del centrosoma (87). Se estableció una distribución a lo largo de un gradiente de densidad compartido por diferentes proteínas cuya localización conocida era el centrosoma. Dicha distribución permitió identificar 23 nuevas proteínas, de las cuales 19 se validaron mediante aproximaciones alternativas. Más tarde, este método se ha extendido para el análisis de múltiples orgánulos de manera simultánea: núcleo, aparato de Golgi, vesículas del retículo endoplasmático, membrana plasmática y mitocondria (86).

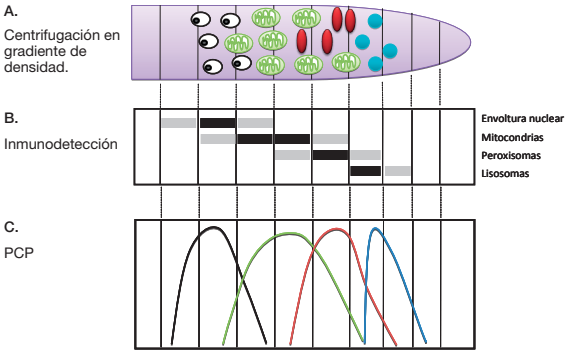


Figura 2. Fundamentos del perfil de correlación de proteínas (PCP). Los extractos que contienen los diferentes orgánulos intactos se separan mediante centrifugación en gradiente de densidad (A), y las diferentes fracciones obtenidas son analizadas mediante técnicas de inmunodetección con anticuerpos específicos de orgánulos (B). Para la PCP, se miden las intensidades de los péptidos trípticos de cada marcador específico de orgánulo mediante espectrometría de masas. De manera similar a los resultados obtenidos por inmunodetección, se generarán los perfiles de distribución de cada orgánulo a lo largo de las fracciones obtenidas mediante centrifugación diferencial (C). Las proteínas pertenecientes a los mismos orgánulos tendrán perfiles semejantes, mientras que las proteínas contaminantes tendrán perfiles diferenciales.

4.3.2.3. Localización de proteínas en orgánulos mediante marcaje isobárico

Esta aproximación es una variante del PCP, en la que se emplean marcajes isotópicos para la cuantificación relativa. La localización de proteínas en orgánulos mediante marcaje isobárico (LOPIT, Localisation of organelle proteins by isotope tagging) utiliza también un gradiente de densidad como punto de partida, además del marcaje iTRAQ para la cuantificación (88). Una vez obtenidas las diferentes fracciones del gradiente, aquellas de interés se tripsinizan y se marcan con los reactivos de iTRAQ. Las fracciones así marcadas se analizan mediante espectrometría de masas. De esta manera se mide la abundancia relativa de cada péptido en cada una de las fracciones. Los péptidos o proteínas se asignan a orgánulos específicos mediante la agrupación de proteínas de localización desconocida con marcadores específicos. Existen diversas métodos para realizar esta asignación, aunque el más utilizado es el análisis de componentes principales (62). El análisis puede ser llevado a cabo con los datos de cuantificación obtenidos o utilizando todos los ratios posibles de los iones reporteros. Además, la asignación de nuevas proteínas puede realizarse mediante un análisis discriminante de mínimos cuadrados parciales (89) utilizando una base de datos curada para cada uno de los orgánulos. Recientemente se ha publicado un protocolo detallado (90) y se ha aplicado a diferentes organismos (91-93). Un aspecto importante a tener en cuenta es que la asignación de proteínas a diferentes orgánulos se basa, de manera inicial, en un reducido número de marcadores conocidos. Sin embargo, la acumulación de datos, combinados de manera apropiada, incrementa la resolución y confianza de los resultados. A menos que los marcadores de partida no estuvieran bien caracterizados y que las asignaciones subsiguientes de proteínas se realizaran sin interpretación biológica, la exactitud de la localización de nuevas proteínas mejora cuando se combinan varios grupos de datos (62).

5. Inmunodepleción de proteínas mayoritarias.

Un caso particular de fraccionamiento es el de los fluidos biológicos en general, y el del plasma y/o suero en particular. El suero contiene una concentración de proteínas cercana a 60 mg/ml, sin embargo el 90% corresponde únicamente a 10 proteínas. Además, el resto de proteínas se encuentran presentes en un amplio rango dinámico de 10 órdenes de magnitud en términos de concentración (94), algo inabordable desde un punto de vista técnico, a pesar de que las tecnologías actuales son capaces de detectar proteínas que se encuentran en concentraciones del rango femtomolar e incluso atomolar. Es importante tener en cuenta que las proteínas menos abundantes, 2-3 órdenes de magnitud por debajo de las mayoritarias no son detectadas por un efecto de apantallamiento (95). Así, los análisis de los proteomas séricos requieren de un proceso de fraccionamiento previo que disminuya el rango dinámico y aumente la concentración de las proteínas minoritarias (96-98).

Los métodos más empleados para el fraccionamiento del plasma están basados en columnas de fase sólida, siendo de tipos muy diversos: intercambio iónico (99,100), quelantes de metales (101), ligandos de afinidad (102), ligandos basados en colorantes (103-105), anticuerpos (102,106) o combinaciones de algunas de ellas (107). Diversas compañías han comercializado diferentes kits de depleción de las proteínas más abundantes (99,100,102,106,108). Los métodos de depleción basados en intercambio iónico o en reactivos como el Cibacron Blue no son específicos de proteína. Sin embargo, los métodos basados en anticuerpos, proteínas, péptidos etc. son más selectivos. Por ejemplo, la utilización de las proteínas bacterianas A y G, que se unen de manera específica a la región Fc de las IgG, se han utilizado con eficacia para la depleción de dichas inmunoglobulinas en muestras de suero y plasma humano (109,110). Pero el método más empleado en la actualidad son las columnas de afinidad con anticuerpos que reconocen a las proteínas más abundantes en el suero, ya que han demostrado ser el método más eficaz, reproducible y selectivo (99,102). Las principales desventajas de este método son

aquellas inherentes al empleo de anticuerpos, esto es, el alto coste y la potencial pérdida de información durante el procesado de la muestra (111).

5.1. Inmunodepleción mediante anticuerpos específicos

El concepto de inmunoafinidad negativa fue introducido por primera vez por Wheatley JB en 1992 (112). Su objetivo era separar la albúmina y la transferrina de un tercer compuesto, IgG, a partir de una mezcla de los 3 componentes. En este caso, los contaminantes eran la albúmina y la transferrina y las proteínas de interés, las IgG. Para ello diseñó una columna en la que inmovilizó anticuerpos frente a las 2 proteínas contaminantes, de manera que el eluido estará compuesto por las proteínas de interés, de ahí el término Cromatografía de inmunoafinidad Negativa.

Basándose en esta idea, Pieper y colaboradores desarrollaron en 2003 la primera columna de inmunoafinidad negativa cuyo objetivo era deplecionar las proteínas de mayor abundancia del plasma (113). Para ello, sintetizaron anticuerpos frente a las proteínas mayoritarias tales como la albúmina, Inmunoglobulinas A y G, Haptoglobina, alpha-1 antitripsina, Hemopexina, Transtiretina, alpha-2 glicoproteína, alpha-1 glicoproteína ácida, alpha-2 macroglobulina y fibrinógeno. Estos anticuerpos se inmovilizaron de forma covalente en una columna a través de proteínas A y G bacterianas, mediante entrecruzamiento. La mezcla de los diferentes anticuerpos no se realizó en proporciones equimolares, sino que se intentó distribuir en las mismas concentraciones en las que aparecen las correspondientes proteínas en plasma. Además, dado el elevado coste asociado a la utilización de anticuerpos, se valoró la posibilidad de reciclar dichas columnas mediante la elución de las proteínas retenidas. Los resultados mostraron que este tipo de columnas basadas en inmunoafinidad eliminan de manera efectiva las proteínas mayoritarias. Para demostrarlo se analizó la fracción no unida a la matriz de la columna, comprobando que la mayor parte de las proteínas más abundantes quedaban retenidas. Dada la elevada concentración de proteína en plasma (alrededor de 60 mg/ml) la muestra es previamente diluida en un tampón que contiene 25 mM fosfato sódico, 150 mM cloruro sódico, 0,01% azida sódica a un pH de 7,2. Para valorar la reutilización de la columna, las proteínas unidas a los diferentes anticuerpos se eluyeron a pH 2,1 en un tampón que contenía 150 mM de cloruro sódico y 5% de ácido acético o glicina y CHAPS. Aunque los siguientes ciclos de depleción con la misma columna mostraron que potencialmente podría ser reutilizada, los análisis de las fracciones no unidas a la matriz mostraron una pérdida de afinidad y selectividad. Sin embargo, y basándose en este mismo trabajo, diferentes compañías han desarrollado diversas columnas de inmunoafinidad para la depleción de proteínas mayoritarias en suero y plasma que permiten una reutilización eficaz de hasta 100 veces (Figura 3). Existen diversos tipos, diferenciándose principalmente en el número y tipo de proteínas deplecionadas (46,104-106,114-122), aunque la mayoría de ellos se centran en la albúmina e IgG (46,105,116,118,122). Durante los últimos años se han desarrollado diferentes columnas de inmunoafinidad para la depleción de las proteínas mayoritarias en plasma humano, difiriendo en el número de proteínas deplecionadas desde 7 (MARS 7 kit de Agilent Technologies), 14 (Seppro IgY14 kit de Sigma Aldrich o el MARS Hu-14 kit de Agilent Technologies) hasta 20 (ProteoPrep20 de Sigma).

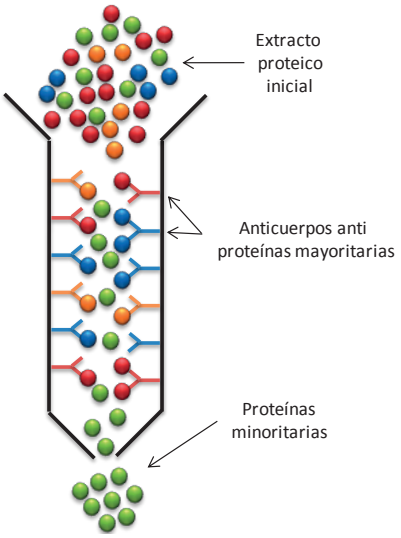


Figura 3. Representación esquemática de la inmunodepleción de proteínas mayoritarias basada en anticuerpos específicos. La mezcla de proteínas se introduce en un dispositivo tipo columna cromatográfica que contiene anticuerpos frente a las proteínas mayoritarias (rojo, naranja y azul). De esta manera, la fracción de proteínas que no se unen (verde) estará compuesta de aquella proteínas que se encuentran en menor proporción.

5.2- ProteoMiner

Un método alternativo para el enriquecimiento de las proteínas minoritarias se basa en el tratamiento de muestras complejas de proteínas con una librería de hexapéptidos unidos a un soporte cromatográfico. Este sistema, desarrollado por Righetti y colaboradores (123), se ha denominado Proteominer y comercializado por BioRad. En teoría, cada hexapéptido se une a un único sitio de reconocimiento de cada proteína. Dado que las proteínas mayoritarias saturan sus ligandos, el exceso de proteína es lavado durante el proceso. Por el contrario, aquellas proteínas que se encuentran en baja concentración se concentran gracias a la unión a sus ligandos correspondientes, consiguiendo disminuir el rango dinámico de la muestra (124).

Las librerías combinatorias de hexapéptidos en fase sólida son sintetizadas y fijadas mediante un espaciador en soportes esféricos de polihidroximetacrilato, a través de la aproximación de Merrifield (125), utilizando un método de recombinación (126,127). Los ligandos se distribuyen homogéneamente en toda la superficie de los soportes esféricos, que a su vez son porosos, pudiendo alcanzar una concentración de 50pmol/ml de esferas. Así, cada estructura esférica contiene millones de copias de un único ligando y, potencialmente, cada esfera contiene un ligando diferente de cualquier otra. Considerando que, para el proceso de síntesis, se utilizan los 20 aminoácidos naturales, la librería de hexapéptidos contiene una población lineal de 206 hexapéptidos, es decir, 64 millones de ligandos diferentes. En principio, en un volumen apropiado de esferas, cabría esperar encontrar un ligando para cualquier proteína presente en un proteoma complejo. La incorporación de aminoácidos no naturales y d-enantiómeros durante el proceso de síntesis generaría una potencial librería capaz de contener un ligando para cada proteína presente en una muestra biológica. (Figura 4)

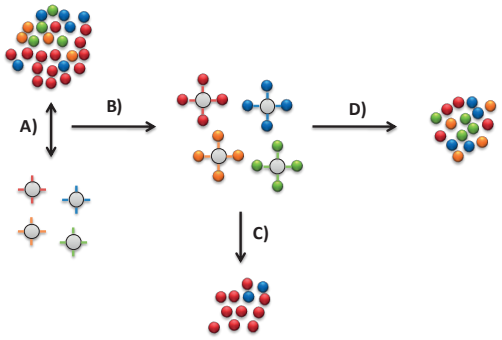


Figura 4. Representación esquemática de ProteoMiner. La mezcla del extracto peptídico y la librería de hexapéptidos (A) resulta en la unión específica péptido-ligando. Cada esfera contiene el mismo número de hexapéptidos, de manera que una vez saturados todos los ligandos, el exceso de péptidos correspondientes es eliminado (C). Finalmente, los péptidos son eluidos de los soportes esféricos, obteniendo una representación de la muestra inicial en la que todos los componentes se encuentran en la misma proporción (D).

Los tipos de interacción que se establecen entre una proteína y un hexapéptido son las clásicas que se establecen en la conformación de macromoléculas, es decir, interacciones débiles que son, al menos, un orden de magnitud menor que las uniones covalentes. Este tipo de interacciones describen cómo los átomos o grupos de átomos son atraídos o repelidos para minimizar la energía de conformación (123) y pueden agruparse en: ion-ion, puentes de hidrógeno, dipolo-dipolo, dispersión e interacciones hidrofóbicas. Todas ellas juegan un papel determinante en el proceso de unión ligando-proteína, y han de tenerse en cuenta en el proceso de elución, paso crucial en la recuperación de la muestra de interés. El proceso de elución puede dividirse en 2 aproximaciones diferentes, en función de si se quiere recuperar toda la muestra en un solo paso (con un eluyente suficientemente fuerte como para destruir todas las uniones ligando-proteína) o una serie de pasos (con eluyentes de fortaleza creciente o mediante mecanismos de elución complementarios). Esta última aproximación tiene la ventaja de permitir un fraccionamiento adicional de la muestra sin utilizar métodos cromatrográficos alternativos.

Algunos estudios han comparado esta tecnología frente a la inmunodepleción de proteínas mayoritarias. En concreto se comparó el sistema ProteoMiner frente a columnas de Inmunodepleción de las 20 proteínas mayoritarias en plasma (95). Los resultados mostraron que ambos

sistemas aportan información complementaria, aunque la inmunodepleción permitió identificar un 25% más de proteínas. Sin embargo, el sistema ProteoMiner tiene la ventaja de que permite obtener una mayor cantidad de proteína que puede ser nuevamente fraccionada mediante cualquier método descrito. Además, tiene la ventaja de ser una aproximación más económica y técnicamente más sencilla (95). Finalmente, cabe señalar que para el empleo de Proteominer en estudios semicuantitativos hay que determinar el volumen de muestra necesario. Por un lado, hay que concentrar la muestra lo suficiente para su detección posterior mediante espectrometría de masas, pero sin llegar a saturar sus ligandos. Por tanto, la elección de un método u otro dependerá de la disponibilidad de muestra inicial y del objetivo final del experimento.

REFERENCIAS.

1. Anderson, N. L., and Anderson, N. G. (1991) *Electrophoresis* 12, 883-906
2. Qian, W. J., Jacobs, J. M., Liu, T., Camp, D. G., 2nd, and Smith, R. D. (2006) *Mol Cell Proteomics* 5, 1727-1744
3. Bodzon-Kulakowska, A., Bierczynska-Krzysik, A., Dylag, T., Drabik, A., Suder, P., Noga, M., Jarzebinska, J., and Silberring, J. (2007) *J Chromatogr B Analyt Technol Biomed Life Sci* 849, 1-31
4. Hagen, J. V. (2008) *Wiley-VCH*
5. Ahmed, N., and Rice, G. E. (2005) *J Chromatogr B Analyt Technol Biomed Life Sci* 815, 39-50
6. Freeman, W. M., and Hemby, S. E. (2004) *Neurochem Res* 29, 1065-1081
7. Steinberg. (2005) *Humana press, totowa*
8. Melle, C., Osterloh, D., Ernst, G., Schimmel, B., Bleul, A., and von Eggeling, F. (2005) *Int J Mol Med* 16, 11-17
9. Kikkawa, R., Fujikawa, M., Yamamoto, T., Hamada, Y., Yamada, H., and Horii, I. (2006) *J Toxicol Sci* 31, 23-34
10. Witzmann, F. A., Clack, J. W., Geiss, K., Hussain, S., Juhl, M. J., Rice, C. M., and Wang, C. (2002) *Electrophoresis* 23, 2223-2232
11. Hortin, G. L., Jortani, S. A., Ritchie, J. C., Jr., Valdes, R., Jr., and Chan, D. W. (2006) *Clinical chemistry* 52, 1218-1222
12. West-Nielsen, M., Hogdall, E. V., Marchiori, E., Hogdall, C. K., Schou, C., and Heegaard, N. H. (2005) *Anal Chem* 77, 5114-5123
13. Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D. W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., Meyer, H., Paik, Y. K., Yoo, J. S., Ping, P., Pounds, J., Adkins, J., Qian, X., Wang, R., Wasinger, V., Wu, C. Y., Zhao, X., Zeng, R., Archakov, A., Tsugita, A., Beer, I., Pandey, A., Pisano, M., Andrews, P., Tammen, H., Speicher, D. W., and Hanash, S. M. (2005) *Proteomics* 5, 3226-3245
14. Aldred, A. R., Brack, C. M., and Schreiber, G. (1995) *Comp Biochem Physiol B Biochem Mol Biol* 111, 1-15
15. Consiglio, A. R., and Lucion, A. B. (2000) *Brain Res Brain Res Protoc* 5, 109-114
16. Sickmann, A., Dormeyer, W., Wortelkamp, S., Woitalla, D., Kuhn, W., and Meyer, H. E. (2002) *J Chromatogr B Analyt Technol Biomed Life Sci* 771, 167-196
17. Yuan, X., and Desiderio, D. M. (2005) *J Chromatogr B Analyt Technol Biomed Life Sci* 815, 179-189
18. Hamdan, and PG, R. (2005) *John Wiley & Sons, Inc.*
19. Islam, N., Lonsdale, M., Upadhyaya, N. M., Higgins, T. J., Hirano, H., and Akhurst, R. (2004) *Proteomics* 4, 1903-1908
20. Cash, P., and Argo, E. (1998) *Methods in Molecular Biology*
21. Rabilloud, T. (1999) *Biochemistry & Biophysics, Springer*
22. Westermeier, R., and Naven, T. (2002) *Wiley-VCH; 1 edition*
23. Kamp, R. M., Calvete, J. J., Choli-Papadopoulou, and (Eds.), T. (2004) *Biochemistry & Biophysics, Springer*
24. Righetti, P. G., Stoyanov, A., and Zhukov, M. Y. (2001) *Elsevier, Amsterdam*
25. Rabilloud, T. (1996) *Electrophoresis* 17, 813-829
26. Wilkins, M. R., Williams, K. L., Appel, R. D., and (Eds.), D. F. H. (1997) *Springer, Verlag, Berlin/Heidelberg/New York*
27. O'Farrell, P. H. (1975) *The Journal of biological chemistry* 250, 4007-4021
28. Righetti, P. G., Castagna, A., Antonioli, P., and Boschetti, E. (2005) *Electrophoresis* 26, 297-319
29. Guillemin, I., Becker, M., Ociepka, K., Friauf, E., and Nothwang, H. G. (2005) *Proteomics* 5, 35-45
30. Bierczynska-Krzysik, A., Kang, S. U., Silberrring, J., and Lubec, G. (2006) *Neurochem Int* 49, 245-255

31. Boyd-Kimball, D., Castegna, A., Sultana, R., Poon, H. F., Petroze, R., Lynn, B. C., Klein, J. B., and Butterfield, D. A. (2005) *Brain Res* 1044, 206-215

32. Bae, S. H., Harris, A. G., Hains, P. G., Chen, H., Garfin, D. E., Hazell, S. L., Paik, Y. K., Walsh, B. J., and Cordwell, S. J. (2003) *Proteomics* 3, 569-579

33. Hoving, S., Gerrits, B., Voshol, H., Muller, D., Roberts, R. C., and van Oostrum, J. (2002) *Proteomics* 2, 127-134

34. Culvenor, J. G., Ilaya, N. T., Ryan, M. T., Canterford, L., Hoke, D. E., Williamson, N. A., McLean, C. A., Masters, C. L., and Evin, G. (2004) *Eur J Biochem* 271, 375-385

35. Fountoulakis, M., Juranville, J. F., Roder, D., Evers, S., Berndt, P., and Langen, H. (1998) *Electrophoresis* 19, 1819-1827

36. Leimgruber, R. M., Malone, J. P., Radabaugh, M. R., LaPorte, M. L., Violand, B. N., and Monahan, J. B. (2002) *Proteomics* 2, 135-144

37. Gorg, A., Weiss, W., and Dunn, M. J. (2004) *Proteomics* 4, 3665-3685

38. Rabilloud, T., Adessi, C., Giraudel, A., and Lunardi, J. (1997) *Electrophoresis* 18, 307-316

39. Herbert, B. (1999) *Electrophoresis* 20, 660-663

40. Herbert, B. R., Molloy, M. P., Gooley, A. A., Walsh, B. J., Bryson, W. G., and Williams, K. L. (1998) *Electrophoresis* 19, 845-851

41. Marshall, T., and Williams, K. M. (1984) *Anal Biochem* 139, 502-505

42. Righetti, P. G., and Gianazza, E. (1980) *J Chromatogr* 184, 415-456

43. Weiller, G. F., Caraux, G., and Sylvester, N. (2004) *Proteomics* 4, 943-949

44. (2010) *GE healthcare, life sciences*

45. Tantipaiboonwong, P., Sinchaikul, S., Sriyam, S., Phutrakul, S., and Chen, S. T. (2005) *Proteomics* 5, 1140-1149

46. Chen, Y. Y., Lin, S. Y., Yeh, Y. Y., Hsiao, H. H., Wu, C. Y., Chen, S. T., and Wang, A. H. (2005) *Electrophoresis* 26, 2117-2127

47. Zellner, M., Winkler, W., Hayden, H., Diestinger, M., Eliasen, M., Gesslbauer, B., Miller, I., Chang, M., Kungl, A., Roth, E., and Oehler, R. (2005) *Electrophoresis* 26, 2481-2489

48. Liebler. (2002) *Humana Press, Inc., Totowa, NJ.*

49. Fang, X., and Zhang, W. W. (2008) *Journal of proteomics* 71, 284-303

50. Isaaq, H. (2001) *Electrophoresis*

51. Wolters, D. A., Washburn, M. P., and Yates, J. R., 3rd. (2001) *Anal Chem* 73, 5683-5690

52. Bier, M. (1998) *Electrophoresis* 19, 1057-1063

53. Appelmans, F., and De Duve, C. (1955) *The Biochemical journal* 59, 426-433

54. Weber, G., and Wildgruber, R. (2008) *Methods in molecular biology (Clifton, N.J* 384, 703-716

55. Simpson, D. C., and Smith, R. D. (2005) *Electrophoresis* 26, 1291-1305

56. Huang, Y. F., Huang, C. C., Hu, C. C., and Chang, H. T. (2006) *Electrophoresis* 27, 3503-3522

57. Kolch, W., Neususs, C., Pelzing, M., and Mischak, H. (2005) *Mass Spectrom Rev* 24, 959-977

58. Mukhopadhyay, D., and Riezman, H. (2007) *Science (New York, N.Y* 315, 201-205

59. Villen, J., and Gygi, S. P. (2008) *Nature protocols* 3, 1630-1638

60. Lewandrowski, U., Moebius, J., Walter, U., and Sickmann, A. (2006) *Mol Cell Proteomics* 5, 226-233

61. Yates, J. R., 3rd, Gilchrist, A., Howell, K. E., and Bergeron, J. J. (2005) *Nature reviews* 6, 702-714

62. Gatto, L., Vizcaino, J. A., Hermjakob, H., Huber, W., and Lilley, K. S. (2010) *Proteomics*

63. Brunet, S., Thibault, P., Gagnon, E., Kearney, P., Bergeron, J. J., and Desjardins, M. (2003) *Trends in cell biology* 13, 629-638

64. Huber, L. A., Pfaller, K., and Vietor, I. (2003) *Circ Res* 92, 962-968

65. Stasyk, T., and Huber, L. A. (2004) *Proteomics* 4, 3704-3716

66. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007) *Anal Bioanal Chem* 389, 1017-1031

67. Wiederhold, E., Veenhoff, L. M., Poolman, B., and Slotboom, D. J. (2010) *Mol Cell Proteomics* 9, 431-445

68. Graham, J. M. (2002) *ScientificWorldJournal* 2, 1607-1609

69. Lee, Y. H., Tan, H. T., and Chung, M. C. (2010) *Proteomics* 10, 3935-3956

70. Rickwood, D. (1984) *IRL Press. Oxford, England*, pp27-42, 45-73, 82-87

71. (2007) *GE healthcare, life sciences*

72. Ackermann, B. L., and Berna, M. J. (2007) *Expert review of proteomics* 4, 175-186

73. Ibrahim, S. F., and van den Engh, G. (2007) *Advances in biochemical engineering/biotechnology* 106, 19-39

74. Givan, A. L. (2004) *Methods in molecular biology (Clifton, N.J* 263, 1-32

75. Bock, G., Steinlein, P., and Huber, L. A. (1997) *Trends in cell biology* 7, 499-503

76. Gauthier, D. J., Sobota, J. A., Ferraro, F., Mains, R. E., and Lazure, C. (2008) *Proteomics* 8, 3848-3861

77. Cao, Z., Li, C., Higginbotham, J. N., Franklin, J. L., Tabb, D. L., Graves-Deal, R., Hill, S., Cheek, K., Jerome, W. G., Lapierre, L. A., Goldenring, J. R., Ham, A. J., and Coffey, R. J. (2008) *Mol Cell Proteomics* 7, 1651-1667

78. Dhandayuthapani, S., Via, L. E., Thomas, C. A., Horowitz, P. M., Deretic, D., and Deretic, V. (1995) *Molecular microbiology* 17, 901-912

79. Cossarizza, A., Ceccarelli, D., and Masini, A. (1996) *Experimental cell research* 222, 84-94

80. Wilson, R. B., and Murphy, R. F. (1989) *Methods in cell biology* 31, 293-317

81. Walther, T. C., and Mann, M. (2010) *The Journal of cell biology* 190, 491-500

82. Schirmer, E. C., Florens, L., Guan, T., Yates, J. R., 3rd, and Gerace, L. (2003) *Science (New York, N.Y* 301, 1380-1382

83. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) *Mol Cell Proteomics* 3, 1154-1169

84. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) *Mol Cell Proteomics* 1, 376-386

85. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) *Nature biotechnology* 17, 994-999

86. Foster, L. J., de Hoog, C. L., Zhang, Y., Zhang, Y., Xie, X., Mootha, V. K., and Mann, M. (2006) *Cell* 125, 187-199

87. Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A., and Mann, M. (2003) *Nature* 426, 570-574

88. Dunkley, T. P., Watson, R., Griffin, J. L., Dupree, P., and Lilley, K. S. (2004) *Mol Cell Proteomics* 3, 1128-1134

89. Karp, N. A., Griffin, J. L., and Lilley, K. S. (2005) *Proteomics* 5, 81-90

90. Sadowski, P. G., Dunkley, T. P., Shadforth, I. P., Dupree, P., Bessant, C., Griffin, J. L., and Lilley, K. S. (2006) *Nature protocols* 1, 1778-1789

91. Dunkley, T. P., Hester, S., Shadforth, I. P., Runions, J., Weimar, T., Hanton, S. L., Griffin, J. L., Bessant, C., Brandizzi, F., Hawes, C., Watson, R. B., Dupree, P., and Lilley, K. S. (2006) *Proceedings of the National Academy of Sciences of the United States of America* 103, 6518-6523

92. Tan, D. J., Dvinge, H., Christoforou, A., Bertone, P., Martinez Arias, A., and Lilley, K. S. (2009) *Journal of proteome research* 8, 2667-2678

93. Hall, S. L., Hester, S., Griffin, J. L., Lilley, K. S., and Jackson, A. P. (2009) *Mol Cell Proteomics* 8, 1295-1305

94. Liumbruno, G., D'Alessandro, A., Grazzini, G., and Zolla, L. (2010) *Journal of proteomics* 73, 483-507

95. Millioni, R., Tolin, S., Puricelli, L., Sbrignadello, S., Fadini, G. P., Tessari, P., and Arrigoni, G. (2011) *PloS one* 6, e19603

96. Jin, W. H., Dai, J., Li, S. J., Xia, Q. C., Zou, H. F., and Zeng, R. (2005) *Journal of proteome research* 4, 613-619

97. Faca, V., Pitteri, S. J., Newcomb, L., Glukhova, V., Phanstiel, D., Krasnoselsky, A., Zhang, Q., Struthers, J., Wang, H., Eng, J., Fitzgibbon, M., McIntosh, M., and Hanash, S. (2007) *Journal of proteome research* 6, 3558-3565

98. Tang, H. Y., Ali-Khan, N., Echan, L. A., Levenkova, N., Rux, J. J., and Speicher, D. W. (2005) *Proteomics* 5, 3329-3342

99. Chromy, B. A., Gonzales, A. D., Perkins, J., Choi, M. W., Corzett, M. H., Chang, B. C., Corzett, C. H., and McCutchen-Maloney, S. L. (2004) *Journal of proteome research* 3, 1120-1127

100. Ramstrom, M., Hagman, C., Mitchell, J. K., Derrick, P. J., Hakansson, P., and Bergquist, J. (2005) *Journal of proteome research* 4, 410-416

101. Zhang, R., Barker, L., Pinchev, D., Marshall, J., Rasamoelisolo, M., Smith, C., Kupchak, P., Kireeva, I., Ingratta, L., and Jackowski, G. (2004) *Proteomics* 4, 244-256

102. Bjorhall, K., Miliotis, T., and Davidsson, P. (2005) *Proteomics* 5, 307-317

103. Lopez, M. F., Mikulskis, A., Kuzdzal, S., Bennett, D. A., Kelly, J., Golenko, E., DiCesare, J., Denoyer, E., Patton, W. F., Ediger, R., Sapp, L., Ziegert, T., Lynch, C., Kramer, S., Whiteley, G. R., Wall, M. R., Mannion, D. P., Della Cioppa, G., Rakitan, J. S., and Wolfe, G. M. (2005) *Clinical chemistry* 51, 1946-1954

104. Zolotarjova, N., Martosella, J., Nicol, G., Bailey, J., Boyes, B. E., and Barrett, W. C. (2005) *Proteomics* 5, 3304-3313

105. Ahmed, N., Barker, G., Oliva, K. T., Hoffmann, P., Riley, C., Reeve, S., Smith, A. I., Kemp, B. E., Quinn, M. A., and Rice, G. E. (2004) *British journal of cancer* 91, 129-140

106. Echan, L. A., Tang, H. Y., Ali-Khan, N., Lee, K., and Speicher, D. W. (2005) *Proteomics* 5, 3292-3303

107. Govorukhina, N. I., Reijmers, T. H., Nyangoma, S. O., van der Zee, A. G., Jansen, R. C., and Bischoff, R. (2006) *Journal of chromatography* 1120, 142-150

108. Huang, L., Harvie, G., Feitelson, J. S., Gramatikoff, K., Herold, D. A., Allen, D. L., Amunngama, R., Hagler, R. A., Pisano, M. R., Zhang, W. W., and Fang, X. (2005) *Proteomics* 5, 3314-3328

109. Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J., Angell, N. H., Smith, R. D., Springer, D. L., and Pounds, J. G. (2002) *Mol Cell Proteomics* 1, 947-955

110. Vesterberg, O., and Anundi, H. (1991) *Appl Theor Electrophor* 2, 159-161

111. Luque-Garcia, J. L., and Neubert, T. A. (2007) *Journal of chromatography* 1153, 259-276

112. Wheatley, J. B. (1992) *J Chromatogr* 603, 273-278

113. Pieper, R., Su, Q., Gatlin, C. L., Huang, S. T., Anderson, N. L., and Steiner, S. (2003) *Proteomics* 3, 422-432

114. Qian, W. J., Kaleta, D. T., Petritis, B. O., Jiang, H., Liu, T., Zhang, X., Mottaz, H. M., Varnum, S. M., Camp, D. G., 2nd, Huang, L., Fang, X., Zhang, W. W., and Smith, R. D. (2008) *Mol Cell Proteomics* 7, 1963-1973

115. Polaskova, V., Kapur, A., Khan, A., Molloy, M. P., and Baker, M. S. (2010) *Electrophoresis* 31, 471-482

116. Colantonio, D. A., Dunkinson, C., Bovenkamp, D. E., and Van Eyk, J. E. (2005) *Proteomics* 5, 3831-3835

117. Roche, S., Tiers, L., Provansal, M., Seveno, M., Piva, M. T., Jouin, P., and Lehmann, S. (2009) *Journal of proteomics* 72, 945-951

118. Govorukhina, N. I., Keizer-Gunnink, A., van der Zee, A. G., de Jong, S., de Bruijn, H. W., and Bischoff, R. (2003) *Journal of chromatography* 1009, 171-178

119. Gong, Y., Li, X., Yang, B., Ying, W., Li, D., Zhang, Y., Dai, S., Cai, Y., Wang, J., He, F., and Qian, X. (2006) *Journal of proteome research* 5, 1379-1387

120. Stempfer, R., Kubicek, M., Lang, I. M., Christa, N., and Gerner, C. (2008) *Electrophoresis* 29, 4316-4323

121. Desrosiers, R. R., Beaulieu, E., Buchanan, M., and Beliveau, R. (2007) *Cell biochemistry and biophysics* 49, 182-195

122. Quero, C., Colome, N., Prieto, M. R., Carrascal, M., Posada, M., Gelpi, E., and Abian, J. (2004) *Proteomics* 4, 303-315

123. Righetti, P. G., Boschetti, E., Lomas, L., and Citterio, A. (2006) *Proteomics* 6, 3980-3992

124. Boschetti, E., and Righetti, P. G. (2008) *Journal of proteomics* 71, 255-264

125. Merrifield, R. B. (1965) *Science (New York, N.Y* 150, 178-185

126. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., and Knapp, R. J. (1991) *Nature* 354, 82-84

127. Furka, A., Sebestyen, F., Asgedom, M., and Dibo, G. (1991) *International journal of peptide and protein research* 37, 487-493

Aislamiento selectivo de Péptidos. Scape

16

Gabriel Padrón Palomares[1], Lázaro Betancourt Nuñez[2], Aniel Sánchez Puentes[3]

[1] Subdirector de Investigaciones Biomédicas del Centro de Ingeniería Genética y Biotecnología;

[2] Departamento de Proteómica del Centro de Ingeniería Genética y Biotecnología;

[3] Investigador. Departamento de Proteómica del Centro de Ingeniería Genética y Biotecnología.

1 Introducción.

John Yates III publicó por primera vez la alternativa a la electroforesis bidimensional para los estudios de proteómica, que consiste en la digestión de una mezcla de proteínas con tripsina (u otra enzima específica), sin previa separación, seguida del análisis por LC-MS-MS. La mezcla de péptidos se separa por cromatografía bidimensional o multidimensional. Este método se conoce como MudPIT (del inglés *multidimensional protein identification technology*) o *shotgun proteomics*[1,2].

La mezcla de péptidos que se obtiene es extremadamente compleja, con más de 100 000 péptidos, y aun el empleo de dos métodos cromatográficos ortogonales no son suficientes para una buena separación, por lo que el número de péptidos que pueden ser secuenciados es reducido, y por tanto se reducen las proteínas identificables, particularmente aquellas poco abundantes.

Una solución a esta situación es simplificar la mezcla inicial con la aplicación de métodos que puedan aislar un subconjunto de péptidos que posean alguna característica común y que sean representativos de las proteínas presentes en la mezcla.

Esta solución está basada en la asunción de que la secuenciación de uno o dos péptidos de una proteína es suficiente para identificarla en las bases de datos.

2. ICAT.

Ruedi Aebersold publicó en 1999[3] el primer método de aislamiento selectivo de péptidos, conocido como ICAT (del inglés *Isotope Coded Affinity Tags*), incluso antes de la publicación del MudPIT, y que fue diseñado para aislar solo los péptidos que contienen cisteínas. El ICAT está basado en la reacción de los grupos tioles con un reactivo que consta de tres grupos funcionales. Uno de estos reacciona con los grupos tioles de las cisteínas, el segundo es biotina para la purificación de los péptidos modificados por cromatografía de afinidad en una matriz de avidina y el tercer grupo es una región que contiene isotopos ligeros o pesados que son utilizados para la cuantificación relativa de las proteínas en dos condiciones diferentes.

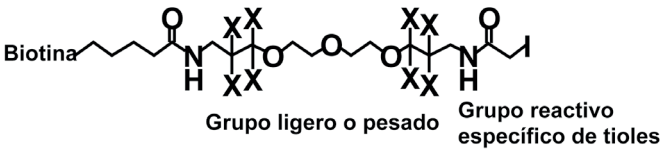
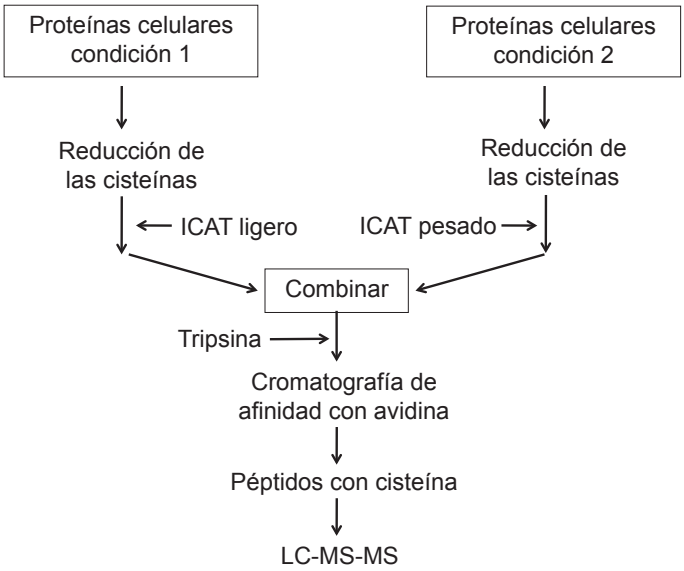


Figura 1. Reactivo ICAT. Las X son hidrógenos en el reactivo ligero y deuterios en el reactivo pesado.

Figura 2. Procedimiento general del método ICAT.



La primera versión del reactivo ICAT pesado tenía 8 deuterios, o sea 8 unidades de masa más que el reactivo ligero (con hidrógenos). Las proteínas se reducen para dejar libre a todas las cisteínas y en una de las dos condiciones se hacen reaccionar con el reactivo ligero, mientras que en la segunda se hacen reaccionar con el reactivo pesado. Posteriormente ambos extractos de proteínas, con las cisteínas modificadas, se mezclan y se digieren con tripsina. Los péptidos que contienen cisteína se aíslan en una columna de avidina y se analizan por LC-MS-MS. El método simplifica considerablemente la mezcla (un promedio de seis péptidos por proteína en humanos y cuatro en levadura) y simultáneamente se puede realizar la cuantificación relativa de los péptidos en las dos condiciones (y por tanto de las proteínas correspondientes).

Con posterioridad se demostró que los péptidos modificados con el reactivo ligero y el pesado no coeluyen exactamente en RP[4] y por tanto no es posible una cuantificación confiable. Existe una segunda versión del reactivo que se basa en ^{13}C y ^{12}C en lugar de ^{2}H y ^{1}H para los reactivos pesados y ligeros, respectivamente[5], además de introducir un sitio de ruptura para eliminar la biotina después de la cromatografía de afinidad y antes del paso de LC-MS-MS.

Aunque el ICAT ha sido empleado en numerosas aplicaciones tiene dos desventajas: no es posible identificar las proteínas que no contengan cisteína en la secuencia y el reactivo es muy costoso.

A partir de la publicación del artículo de Aebersold, aparecieron numerosas publicaciones con métodos para el aislamiento selectivo de péptidos basados en reacciones con aminoácidos poco abundantes como metionina[6], triptófano[7] o con serinas o treoninas en el N-terminal[8]. Otra alternativa es el aislamiento de péptidos que contienen histidinas con el uso de IMAC[9].

3. SCAPE.

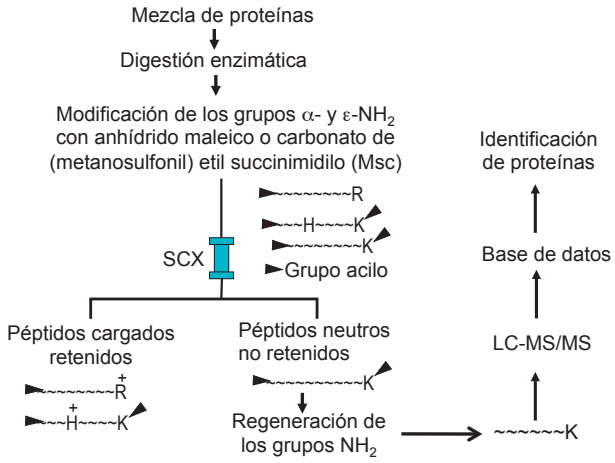
Un procedimiento diferente lo constituye el método SCAPE (del inglés *S*elective *CA*pture of *PE*ptides). Se han publicado tres variantes de este método, que difieren de los anteriores en que se basan en una reacción muy conocida (acilación) con grupos abundantes como los aminos del N-terminal y de la cadena lateral de las lisinas, seguida por el aislamiento selectivo de los péptidos a pH ácido con el uso de la cromatografía de intercambio catiónico[10,11] o de la cromatografía de afinidad[12].

Estas tres variantes se describen a continuación.

3.1. SCAPE: Aislamiento selectivo de péptidos que no contienen arginina ni histidina (péptidos nHnR)[10]

El procedimiento general está resumido en la figura 3.

Figura 3. Procedimiento genera
método nHnR.

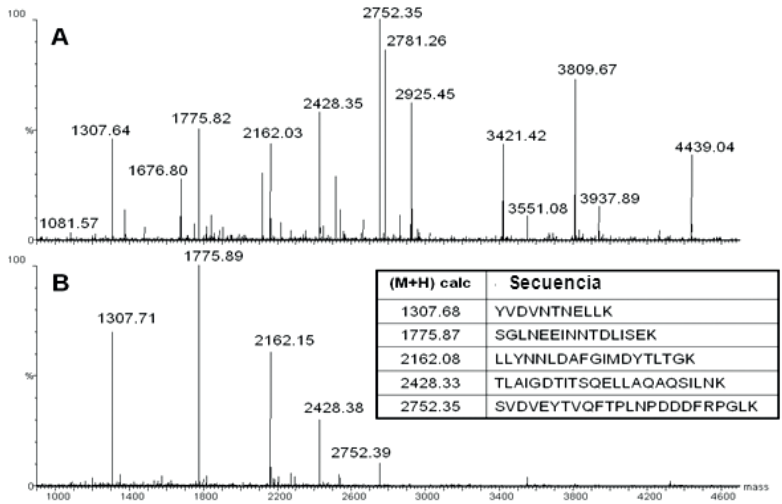


La mezcla de proteínas se digiere con LEP (lisil endopeptidasa) o tripsina y los péptidos obtenidos se hacen reaccionar con anhídrido maleico o carbonato de (metanosulfonil) etil succinimidilo (Msc) para modificar reversiblemente los grupos amino α y ε de los péptidos. A pH 2, los péptidos que contengan arginina o una o más histidinas en la secuencia estarán cargados positivamente, mientras que los péptidos que carezcan de estos aminoácidos en la secuencia serán neutros a ese pH. Los péptidos neutros no serán retenidos al pasar la mezcla por una columna de cromatografía de intercambio catiónico a pH entre 2 y 3 y serán fácilmente separados de los restantes péptidos cargados. Estos péptidos trípticos han sido nombrados nHnR, pues no contienen arginina ni histidina en la secuencia. Posteriormente, el grupo acilo es eliminado por tratamiento con ácido fórmico a alta temperatura en el caso del anhídrido maleico[10] o con NaOH para el Msc[13] y los péptidos son analizados por LC-MS-MS. La regeneración de los grupos amino facilita una mayor eficiencia de ionización de los péptidos.

Al aplicar este método a una proteína pura, la estreptoquinasa recombinante (SK), se logra una simplificación notable de la mezcla de péptidos (figura 4).

En la figura 4 se observan los espectros de masas de los péptidos obtenidos por la digestión de la SK con LEP (4A) y de los péptidos nHnR aislados (4B). De un total de 20 péptidos de más

de 800 Da que se pudieran esperar de una digestión de la SK con LEP, solo se observan cinco, cuyas secuencias se indican. El quinto péptido no es realmente nHnR, pues tiene una arginina en la secuencia y por tanto debió quedar retenido en la columna de intercambio catiónico a pH 2-3. La cercanía a la arginina de tres ácidos aspárticos podría favorecer interacciones intramoleculares que explicarían que este péptido no quedara retenido.



(M+H) calc	Secuencia
1307.68	YVDVNTNELLK
1775.87	SGLNEEINNTDLISEK
2162.08	LLYNNLDAFGIMDYTLTGK
2428.33	TLAIGDTITSQELLAQAQSILNK
2752.35	SVDVEYTVQFTPLNPDDDFRPGLK

Figura 4. Aplicación del método nHnR a la estreptoquinasa. A): Espectro de masas de los péptidos obtenidos por digestión con LEP. B): Espectro de masas de los péptidos nHnR aislados. En el recuadro se muestran las secuencias de estos péptidos.

3.2. SCAPE: Aislamiento selectivo de péptidos con dos o más cargas (péptidos RH)[11]

El procedimiento general está resumido en la figura 5.

El procedimiento es muy similar al anterior y los péptidos se seleccionan por sus diferencias de carga. Sin embargo, en este caso se emplea anhídrido acético para bloquear los grupos amino, pues no es necesaria la eliminación de los grupos acetilo antes del paso de LC-MS-MS. Los péptidos que serán analizados, aunque tendrán el N-terminal bloqueado, contienen al menos dos residuos básicos ionizables.

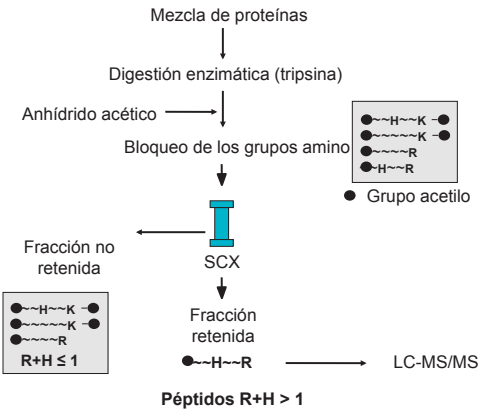


Figura 5. Procedimiento general del método RH.

Al aplicar la mezcla de péptidos a la columna de intercambio catiónico a pH entre 2 y 3 es posible encontrar fácilmente condiciones en las que los péptidos neutros o con una sola carga no sean retenidos, mientras que los de dos o más cargas (argininas + histidinas > 1: péptidos RH) serán retenidos. A continuación los péptidos RH son eluídos y analizados por LC-MS-MS.

De esta forma se logra igualmente una reducción considerable del número de péptidos que serán analizados, como se demuestra con la aplicación del método a la SK recombinante (figura 6).

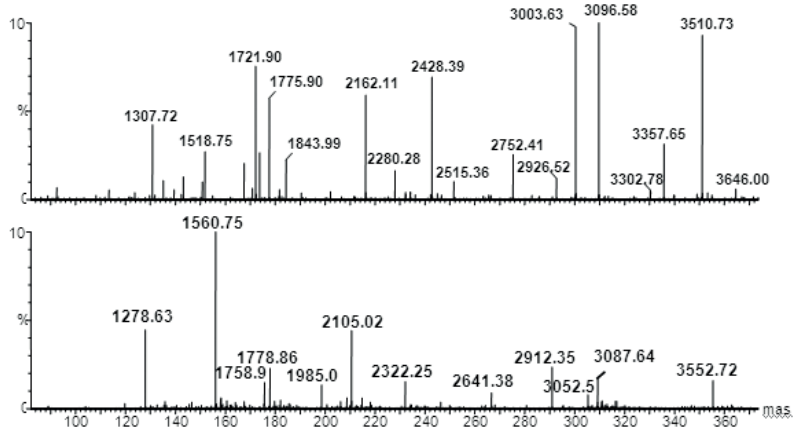


Figura 6. Aplicación del método RH a la estreptoquinasa. Espectro de masas de los péptidos trípticos (arriba). Espectro de masas de los péptidos RH aislados (abajo).

En la tabla 1 se muestran las secuencias de los péptidos RH de la estreptoquinasa que fueron identificados. Se observa una alta selectividad aun cuando aparecen péptidos donde no ocurrieron algunos cortes con la tripsina.

Tabla 1. Secuencias de los péptidos RH de la estreptoquinasa aislados e identificados. Cada péptido tiene un grupo acetilo en el N-terminal y en cada una de las lisinas.

Secuencia	Grupos acetilos	m/z experimental	m/z esperado
$_{211}$T*H*PGYTIYE*R*$_{220}$	1	1 236.63	1 278.61
$_{221}$DSSIVT*H*DNDIF*R*$_{233}$	1	1 560.75	1 560.73
$_{38}$FFEIDLTS*R*PA*H*GGK$_{52}$	2	1 758.90	1 758.89
$_{390}$YTEEE*R*EVYSYL*R*$_{402}$	1	1 778.86	1 778.83
$_{280}$KGEKPYDPFD*RSH*LK$_{294}$	4	1 985.00	1 985.01
$_{373}$*R*PEGENASY*H*LAYDKD*R*$_{389}$	2	2 105.02	2 104.98
$_{123}$DGSVTLPTQPVQEFLLSG*H*V*R*$_{143}$	1	2 322.25	2 322.21
$_{38}$FFEIDLTS*R*PA*H*GGKTEQGLSPK$_{60}$	3	2 641.38	2 641.35
$_{373}$*R*PEGENASY*H*LAYDKD*R*YTEEE*R*$_{395}$	2	2 912.35	2 912.32
$_{365}$IITVYMGK*R*PEGENASY*H*LAYDKD*R*$_{389}$	3	3 052.52	3 052.50
$_{117}$VYFADKDGSVTLPTQPVQEFLLSG*H*V*R*$_{143}$	2	3 087.64	3 087.59
$_{83}$AIQEQLIANV*H*SNDDYFEVIDFASDATITD*R*$_{113}$	1	3 552.72	3 552.68

Durante la reacción con anhídrido acético se observó que también puede ocurrir la acetilación de las tirosinas. Por esta razón es necesario añadir un paso de eliminación selectiva de la O-acetilación[14] con amoníaco al 1 %. La persistencia de esta modificación crearía varias dificultades como el incremento de la complejidad de la muestra, el incremento de la hidrofobicidad de los péptidos con posibles pérdidas en la columna de RP y el aumento de las posibilidades de identificar incorrectamente las proteínas al tener que considerar una modificación adicional. Este tratamiento no reduce el recobrado general de los péptidos RH y demuestra ser muy efectivo, pues no se detectan péptidos O-acetilados después de su aplicación.

3.3. SCAPE: Aislamiento selectivo de péptidos sin lisinas y delimitados por argininas (péptidos RRnK)[12]

Esta tercera variante del SCAPE se basa en la cromatografía de afinidad y no en la diferencia de cargas. El procedimiento general se muestra en la figura 7. El primer paso es una digestión con LEP. La mezcla de péptidos es tratada con biotina para bloquear todos los grupos amino. Los péptidos biotinilados son digeridos con tripsina, que cortará en las argininas y se producirán tres tipos de péptidos: aquellos con lisina en el C-terminal pero sin modificación en el grupo amino del N-terminal, péptidos con arginina en el C-terminal con el N-terminal biotinilado y péptidos con arginina en el C-terminal pero con el N-terminal libre. Estos últimos serán los únicos que no serán retenidos al aplicar esta nueva mezcla a una columna de afinidad con estreptavidina y que en la secuencia de la proteína están flanqueados por argininas (péptidos RRnK).

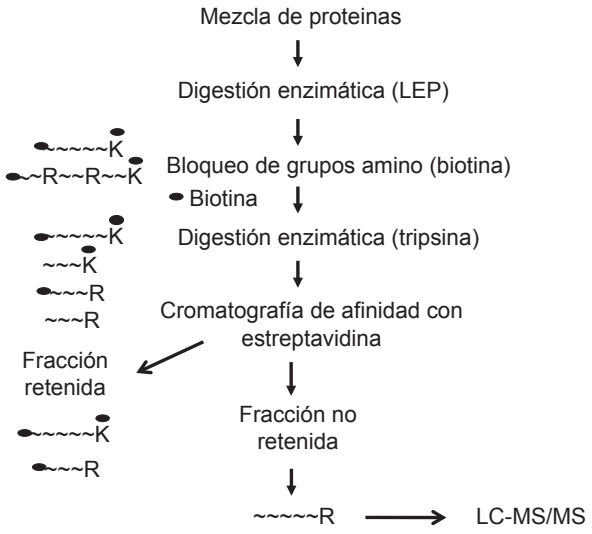


Figura 7. Procedimiento general del método RRnK.

En este procedimiento también ocurre la O-acilación de las tirosinas, por lo que es necesario incluir el paso para su eliminación en medio básico.

La efectividad de este método se demuestra al aplicarlo a la estreptoquinasa: se detectan solo tres péptidos, por lo que se logra muy buena simplificación de la mezcla (figura 8).

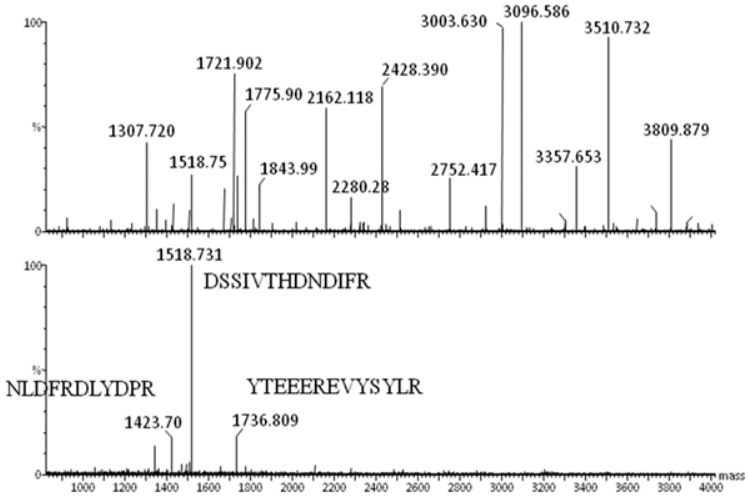


Figura 8. Aplicación del método RRnK a la estreptoquinasa. Espectro de masas de los péptidos trípticos (arriba). Espectro de masas de los péptidos RRnK aislados (abajo).

4. Análisis, complementariedad, ventajas y desventajas de los métodos Scape.

Simplificación de la mezcla de péptidos y cobertura de los proteomas

En la tabla 2 se muestra el estudio *in silico* para evaluar la potencialidad de las tres variantes de SCAPE comparadas con el ICAT frente a varios proteomas. Los cuatro métodos se comportan de forma similar, con un promedio de cuatro péptidos por proteína y una cobertura de alrededor del 85 %, aunque en algunos proteomas se observan mayores diferencias. En humanos el comportamiento es muy parejo. Para otros proteomas se seleccionaría la variante que asegure mayor cobertura.

Tabla 2. Estudio *in silico* del número de proteínas en diferentes organismos y las proteínas identificables esperadas para los métodos de aislamiento selectivo. Se consideran solo péptidos en el rango de 800 a 3 500 Da.

Organismo	# de proteínas	nHnR*	RH*	RRnK*	ICAT *
V. cholerae	3 785	4 / 92	3 / 78	3 / 80	2 / 76
E. coli	4 322	4 / 92	4 / 85	3 / 85	2 / 72
N. meningitidis	1 967	4 / 93	3 / 78	3 / 78	2 / 72
M. tuberculosis	3 875	2 / 76	7 / 95	5 / 95	3 / 77
S. cerevissiae	4 818	10 / 98	3 / 83	4 / 83	4 / 86
H. sapiens	47 531	6 / 92	5 / 88	6 / 88	6 / 92
Promedio	-	**5 / 91**	**4 / 84**	**4 / 85**	**4 / 83**

** Péptidos por proteína / % de proteínas identificables*

Proteómica cuantitativa

El SCAPE puede ser utilizado con la mayoría de los métodos disponibles para la cuantificación relativa de proteínas en diferentes condiciones. En particular, la variante RH permite la comparación simultánea de cuatro condiciones con el uso de anhídrido acético marcado con ^{13}C o con deuterio: $(CH_3\text{-}CO)_2O$, $(^{13}CH_3\text{-}CO)_2O$ (+1), $(^{13}CH_3\text{-}^{13}CO)_2O$ (+2) y $(CD_3\text{-}CO)_2O$ (+3). Para la cuantificación es posible auxiliarse con programas como el ISOTOPICA[15,16] o el QuiXot[17,18].

Complementariedad.

Los tres métodos SCAPE son complementarios, pues aíslan diferentes grupos de péptidos. Esto puede observarse claramente en la figura 9, donde se evalúan *in silico* seis genomas al aplicar los tres métodos por separado, en combinación de dos de ellos o los tres.

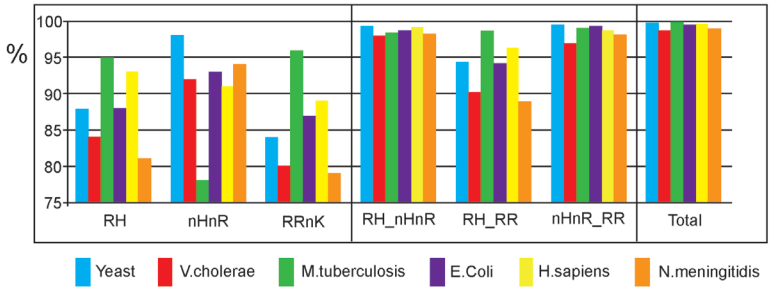


Figura 9. Complementariedad de los métodos SCAPE. Análisis in silico de cada uno de los métodos por separado, en combinación de dos de ellos o los tres en seis genomas.

Hay algunos genomas en los que un solo método podría ser muy eficaz, como por ejemplo el nHnR en levadura o el RRnK en *M. tuberculosis*. Al combinar dos de los métodos se logra más del 95 % de cobertura en casi todos los genomas. La combinación de las tres variantes aseguraría un 100 % de cobertura, sin embargo a costa de un largo procedimiento experimental. Por otra parte la combinación de RH y nHnR o RRnK y nHnR pudieran lograr una cobertura cercana al 100 % sin un excesivo trabajo experimental.

Ventajas y desventajas del SCAPE

Las principales ventajas de las tres variantes de SCAPE son:

- Métodos sencillos basados en reactivos y procedimientos comunes en los laboratorios de proteómica.
- Reducción de la complejidad de la mezcla de péptidos e incremento de la posibilidad de identificar un mayor número de proteínas, particularmente las menos abundantes.
- Posibilidad de cuantificación relativa de dos o más condiciones.
- Son métodos complementarios. La cobertura de cualquier proteoma puede incrementarse a más del 95 % si se combinan dos de las variantes.
- La alta selectividad de estas variantes permite utilizarlas como un constreñimiento adicional para incrementar la confiabilidad de la identificación de las proteínas.
- La principal desventaja radica en que se eliminan numerosos péptidos, lo que reduce considerablemente la posibilidad de identificar las proteínas que no contengan el tipo de péptidos que se intenta aislar o que contengan solo uno o dos de ellos. Además en estas condiciones es muy poco probable poder identificar modificaciones postraduccionales.

5. SCAPE: alternativas para incrementar el número de péptidos analizados.

Teniendo en cuenta la principal desventaja de los métodos de aislamiento selectivo de péptidos, se han diseñado alternativas con el objetivo de aumentar el fraccionamiento de la mezcla total para reducir su complejidad, facilitar la secuenciación de un mayor número de péptidos e incrementar las proteínas identificables en un experimento.

Analizaremos dos de estas alternativas del SCAPE, que están muy relacionadas con los procedimientos de aislamiento selectivo analizados.

Fraccionamiento por carga

La primera de estas alternativas[19] se basa en un método que es idéntico al aislamiento selectivo de los péptidos nHnR hasta el paso del intercambio catiónico a pH 2-3 (figura 10). En este punto se establecen condiciones tales que se puede separar la mezcla de péptidos en tres grupos en dependencia de la carga. Un primer grupo son péptidos neutros que no serán retenidos en la columna a pH 2-3 y que no contienen argininas ni histidinas (nHnR). El segundo grupo son los péptidos que tienen una sola carga, o sea, que contienen una arginina o una histidina y el tercer grupo está constituido por péptidos con una carga mayor de 1 (argininas + histidinas > 1, RH).

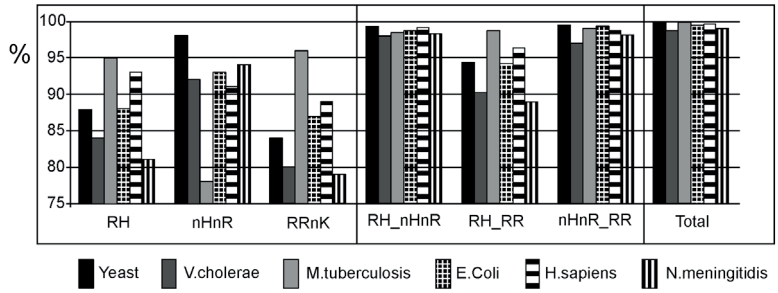
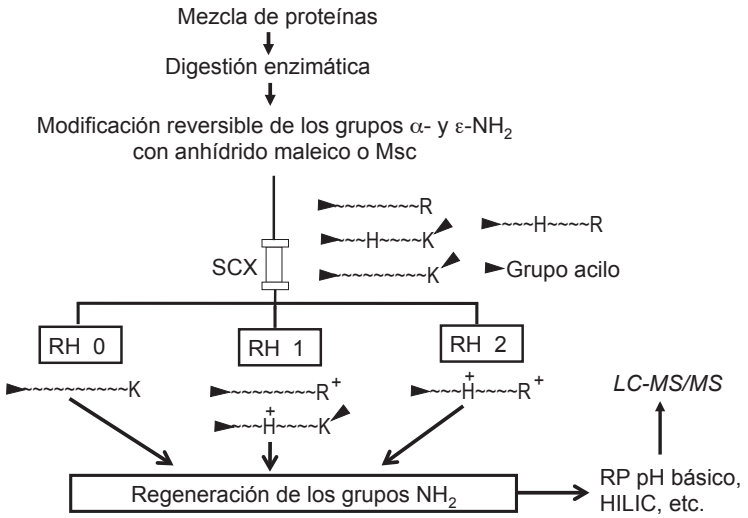


Figura 10. Procedimiento general para la separación de la mezcla de péptidos en tres grupos de acuerdo a la carga: RH0, RH1 y RH2 y posterior separación y análisis de los péptidos de cada uno de los grupos.



Las tres fracciones obtenidas, RH0, RH1 y RH2, son tratadas por separado para regenerar los grupos amino y cada una se separa por cromatografía bidimensional. En la primera dimensión puede usarse fase reversa en medio básico o interacción hidrofílica. Por último cada fracción obtenida en este primer paso cromatográfico se analiza por LC(RP)-MS-MS.

La separación en estos tres grupos es muy reproducible, con cromatogramas muy similares al mostrado en la figura 11 y ocurre con una alta selectividad, usualmente mayor del 98 %[19].

Esta elevada selectividad puede ser utilizada para añadir un constreñimiento que incremente la confiabilidad de la asignación de los péptidos y la identificación de las proteínas.

La segunda alternativa es un método idéntico al aislamiento de péptidos RH (figura 5), pero en este caso no se desecha la fracción no retenida en la columna de intercambio catiónico a pH 2, que es analizada directamente en un espectrómetro MALDI-TOF-TOF[20]. Esta fracción contiene péptidos neutros y péptidos con una sola carga (Arg + His = 1). Al ser analizada con el uso de MALDI como método de ionización, el resultado estará altamente sesgado (prácticamente 100 %) y se identificarán solamente los péptidos con una carga positiva, debido al conocido efecto de supresión en MALDI-MS[21,22]. De hecho el 98 % de los péptidos identificados en esta fracción no retenida tienen una arginina (y por tanto no tienen histidinas).

La inclusión en el análisis de la fracción no retenida puede incrementar en más de un 70 % el número de péptidos y proteínas identificados con un mínimo de trabajo experimental adicional[20].

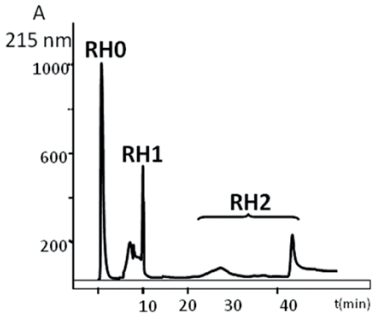


Figura 11. Cromatograma típico de la separación de la mezcla de péptidos en tres grupos de acuerdo a la carga: RH0, RH1 y RH2.

6. Cofradic.

Otro método de aislamiento selectivo de péptidos es el COFRADIC (del inglés *COmbined FRActional Diagonal Chromatography*), desarrollado por el grupo de Joel Vandekerckhove en la Universidad de Ghent, Bélgica.

El COFRADIC[6,23] es una modificación de la llamada cromatografía diagonal de péptidos[24]. Está basado en dos pasos cromatográficos de RP idénticos con una reacción intermedia para modificar específicamente la cadena lateral de un determinado aminoácido, y así alterar el tiempo de retención en RP de los péptidos que contengan al aminoácido modificado. El cambio en el tiempo de retención permite seleccionar estos péptidos, que son posteriormente analizados por LC-MS-MS.

La principal característica del COFRADIC es su versatilidad, pues se puede cambiar la reacción intermedia de manera que modifique diferentes aminoácidos de acuerdo a lo que se desea

analizar. El artículo original[6] describe el uso de la oxidación de la metionina al derivado sulfóxido para analizar solo los péptidos que contienen este aminoácido. El derivado es más hidrofílico y el péptido modificado eluye a menor tiempo de retención.

El método ha sido aplicado también para la identificación de péptidos con cisteínas[25], que se hacen reaccionar con el reactivo de Ellman para formar el enlace con el ácido nitrobenzoico, antes del primer paso de RP. Este grupo hidrofóbico es eliminado después del primer paso cromatográfico y los péptidos con cisteínas son identificados por su corrimiento a menores tiempos de retención.

Otra aplicación ha sido la identificación de los péptidos del N-terminal de las proteínas[26,27]. En este caso los grupos amino libres de las proteínas son primeramente acetilados y las proteínas digeridas con tripsina. Después del primer paso de RP-HPLC, los grupos alfa amino libres de los péptidos internos y del C-terminal son bloqueados usando el ácido 2,4,6-trinitrobenceno-sulfónico (TNBS), dando lugar a derivados marcadamente mas hidrofóbicos. Los péptidos del N-terminal de las proteínas permanecen inalterados y son identificados en el segundo e idéntico paso de RP-HPLC. Estos últimos son colectados para su posterior análisis por LC-MS/MS. El COFRADIC-N-terminal ha sido empleado ampliamente en combinación con técnicas de marcaje isotópico para caracterizar el procesamiento de proteínas en diferentes modelos de muerte celular[28], a los sustratos de enzimas N-acetil transferasas[29] y para aportar nuevos elementos sobre la función de la acetilación en el N-terminal de las proteínas[30].

El COFRADIC también ha sido aplicado para el estudio de modificaciones postraduccionales, tales como fosforilación[31] y glicosilación[32].

El método ha sido adaptado a las exigencias de alto flujo de los experimentos de proteómica al combinar varias fracciones del primer paso de RP, separadas por una ventana de tiempo tal (por ejemplo, 13 minutos en el caso de la oxidación de las metioninas), que aparezcan más allá del corrimiento esperado después de la reacción (de aquí el nombre de *COmbined FRACtional Diagonal Chromatography)*[6,23]. De esta forma se reduce el número de corridas en el segundo paso de RP.

7. Detalles experimentales de uno de los métodos SCAPE (RH).

El método RH puede comparar cuatro condiciones simultáneamente con el uso de cuatro variantes de anhídrido acético: $(CH_3CO)_2O$, $(^{13}CH_3CO)_2O$, $(^{13}CH_3{}^{13}CO)_2O$ y $(CD_3CO)_2O$. Las cisteínas de las proteínas provenientes de cada condición experimental se someten a reducción y alquilación, y las proteínas se digieren en sucesión con LEP y tripsina. Cada una de las cuatro mezclas de péptidos generadas se modifica por separado con uno de los anhídridos. Las posible O-acilaciones que hayan ocurrido se hidrolizan en medio básico. Las cuatro fracciones se mezclan antes de la cromatografía de intercambio catiónico fuerte. La fracción retenida se separa con un gradiente continuo de cloruro de potasio y finalmente los péptidos se analizan por LC-MS/MS.

7.1. Digestiones proteolíticas en solución

Las proteínas se disuelven a una concentración de 5 mg/mL en solución tampón de HEPES 0.5 mol/L, urea 8 mol/L a pH 8.0. Los puentes de disulfuro se reducen por adición de DTT hasta una concentración de 0.01 mol/L durante 4 horas a 37 °C. Las cisteínas libres se bloquean con acrilamida 0.03 mol/L (puede hacerse con ácido yodoacético o con yodoactemida) y se dejan

reaccionar durante 1 hora a temperatura ambiente. Posteriormente se añade LEP hasta obtener una relación enzima:sustrato de 1:200 y se deja reaccionar durante 4 horas a 37 ºC. La solución se diluye cuatro veces con agua, se añade tripsina hasta una relación enzima:sustrato de 1:100 y se mantiene durante 12 horas a 37 ºC. La proteólisis se detiene por adición de aprotinina en una relación 3:1 (masa:masa) con la tripsina.

7.2. Reacciones de acilación en solución

Los ésteres activos: *N*-acetoxi-succinimida, *N*-acetoxi-succinimida (^{13}C), *N*-acetoxi-succinimida ($^{13}C_2$) y *N*-acetoxi-succinimida (D_3), se sintetizan según el protocolo reportado en la literatura[4]

Una vez preparados estos reactivos se pueden almacenar a -20 ºC por largos periodos de tiempo.

Los péptidos trípticos provenientes de cada una de las condiciones comparadas se enfrían en un baño de hielo, y a cada una de las mezclas se adiciona uno de los agentes acilantes (disueltos en 5-10 µL de Dimetilsulfóxido o Dimetilformamida), en una relación molar de 10:1 con respecto a los grupos amino primarios generados durante las digestiones proteolíticas (aproximadamente 1 mg del éster activo por cada 100 µg de proteína, o entre 3 mg y 5 mg para 500 µg de proteína). La reacción ocurre durante 20 minutos a temperatura ambiente. Se hace una segunda adición 1:10 del agente acilante y se deja reaccionar otros 20 minutos, o sea, 40 minutos de reacción total.

7.3. Hidrólisis de las O-acilaciones

A cada una de las mezclas de péptidos acilados se le añade amoniaco al 25 % o trietilamina hasta alcanzar pH 11, y se incuba durante dos horas a 37 °C (para eliminar posibles O-acilaciones, fundamentalmente en las tirosinas). Las soluciones se concentran en el rotoevaporador durante 30 minutos y se acidifican con 2 µL de ácido fórmico.

7.4. Unificación de las cuatro fracciones y desalado

Las cuatro fracciones obtenidas después de la hidrólisis de las O-acilaciones se combinan y los péptidos se desalan mediante el empleo de una columna de fase reversa (o ZipTip).

7.5. Cromatografía de intercambio catiónico

Para la cromatografía de intercambio catiónico se utiliza una columna Biobasic (4.6 x 250 mm, 5 µm, 300 Å, Thermo Electron Corporation, EUA). La composición de las soluciones es: solución A: KH_2PO_4 0.020 mol/L, ACN al 20 %, TFA al 0.05 %; solución B: KH_2PO_4 0.020 mol/L, acetonitrilo al 20 %, TFA al 0.05 %, KCl 0.8 mol/L, ambas a pH 2.7. Los péptidos se separan con el siguiente gradiente:

- 0 % de B durante 20 minutos
- hasta 13 % de B en 30 minutos.
- hasta 36 % de B en 20 minutos.
- hasta 100 % de B en 5 minutos.

El flujo de trabajo es de 0.7 mL/min. Las fracciones se colectan cada dos minutos, comenzando en el minuto 20 y terminando en el minuto 80.

En estas condiciones, con la solución inicial A, los péptidos que no sean RH no se retienen en la columna de intercambio iónico. Así los péptidos que se fraccionan en ese gradiente son todos péptidos RH. Por supuesto la columna de intercambio catiónico fuerte puede ser de otro fabricante y en este caso los flujos y el gradiente deberán ser ajustados.

REFERENCIAS.

1. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001; 19: 242-247.

2. Washburn MP, Ulaszek R, Deciu C, et al. Analysis of Quantitative Proteomic Data Generated via Multidimensional Protein Identification Technology. Anal Chem. 2002; 74: 1650-1657.

3. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999; 17: 994-999.

4. Zhang R, Sioma CS, Wang S, Regnier F. Fractionation of isotopically labeled peptides in quantitative proteomics. Anal Chem. 2001; 73: 5142-9.

5. Yi EC, Li X, Cooke K, et al. Increased quantitative proteome coverage with 13C/12C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. Proteomics 2005; 5: 380-387.

6. Gevaert K, Van Damme J, Goethals M, et al. Chromatographic Isolation of Methionine-containing Peptides for Gel-free Proteome Analysis. Mol Cell Proteomics 2002; 1: 896-903.

7. Kuyama H, Watanabe M, Toda C, Ando E, Tanaka K, Nishimura O. An approach to quantitative proteome analysis by labeling tryptophan residues. Rapid Commun Mass Spectrom. 2003; 17: 1642-1650.

8. Chelius D, Shaler TA. Capture of Peptides with N-Terminal Serine and Threonine: A Sequence-Specific Chemical Method for Peptide Mixture Simplification. Bioconjug Chem. 2003; 14: 205-211.

9. Ren D, Penner NA, Slentz BE, Mirzaei H, Regnier F. Evaluating Immobilized Metal Affinity Chromatography for the Selection of Histidine-Containing Peptides in Comparative Proteomics. J Proteome Res. 2003; 2: 321-329.

10. Betancourt L, Gil J, Besada V, et al. SCAPE: A New Tool for the Selective CApture of PEptides in Protein Identification. J Proteome Res. 2005; 4: 491- 496.

11. Sánchez A, González LJ, Betancourt L, et al. Selective isolation of multiple positively charged peptides for 2-DE-free quantitative proteomics. Proteomics 2006; 6, 4444-4455.

12. Sánchez A, González LJ, Ramos Y, et al. Selective Isolation of Lysine-Free Tryptic Peptides Delimited by Arginine Residues: A New Tool for Proteome Analysis. J Proteome Res. 2006; 5: 1204-1213.

13. Gil J, Betancourt LZ, Sardiñas G, et al. Proteomic study via a non-gel based approach of meningococcal outer membrane vesicle vaccine obtained from strain CU385: a road map for discovering new antigens. Hum Vaccin. 2009; 5: 347-356

14. Zappacosta F, Annan R. N-Terminal Isotope Tagging Strategy for Quantitative Proteomics: Results-Driven Analysis of Protein Abundance Changes. Anal. Chem. 2004; 76: 6618-6627.

15. Fernandez-de-Cossio J, González LJ, Satomi Y et al. Isotopica: a tool for the calculation and viewing of complex isotopic envelopes. Nucleic Acids Research 2004; 32: W674-W678.

16. Fernandez-de-Cossio J, Gonzalez LJ, Satomi Y, et al. Automated interpretation of mass spectra of complex mixtures by matching of isotope peak distributions. Rapid Commun Mass Spectrom. 2004; 18: 2465-2472

17. López-Ferrer D, Ramos-Fernández A, Martínez-Bartolomé S, García-Ruiz P, Vázquez J. Quantitative proteomics using 16O/18O labeling and linear ion trap mass spectrometry. Proteomics 2006; 6 Suppl 1: S4-11.

18. Ramos-Fernández A, López-Ferrer D, Vázquez J. Improved method for differential expression proteomics using trypsin-catalyzed 18O labeling with a correction for labeling efficiency. Mol Cell Proteomics 2007; 6: 1274-86.

19. Betancourt LH, Sánchez A, Pérez Y, et al. Charge state-selective separation of peptides by the reversible modification of amino groups and strong cation-exchange chromatography: evaluation in proteomic studies using peptide-centric database searches. J Prot. 2011 May 9 [Epub ahead of print].

20. Sanchez A, Sun W, Wang L, et al. Selective isolation-detection of two different positively charged peptide groups by strong cation exchange chromatography and matrix-assisted laser desorption ionization/mass spectrometry: application to proteomics studies. Eur J Mass Spectrom. 2010; 16: 693-702.

21. Krause E, Wenschuh H, Jungblut P. The Dominance of Arginine-Containing Peptides in MALDI-Derived Tryptic Mass Fingerprints of Proteins. Anal Chem. 1999; 71: 4160-4165.

22. Brancia F, Oliver S, Gaskell S. Improved matrix-assisted laser desorption/ionization mass spectrometric analysis of tryptic hydrolysates of proteins following guanidination of lysine-containing peptides. Rapid Commun Mass Spectrom. 2000; 14: 2070-2073.

23. Gevaert K, Impens, F, Van Damme P, Ghesquière B, Hanoulle X, Vandekerckhove J. Applications of diagonal chromatography for proteome wide characterization of protein modifications and activity based analyses. FEBS J. 2007; 274: 6277-6289.

24. Brown JR, Hartley BS. Location of disulphide bridges by diagonal paper electrophoresis, the disulphide bridges of bovine chymotrypsinogen A. Biochem J. 1966; 101: 214-228.

25. Gevaert K, Ghesquiere B, Staes A, et al. Reversible labeling of cysteine-containing peptides allows their specific chromatographic isolation for non-gel proteome studies. Proteomics 2004; 4: 897-908.

26. Gevaert K, Goethals M, Martens L. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted n-terminal peptides, Nat Biotechnol. 2003; 21: 566–569.

27. Staes A, Van Damme P, Helsens K, Demol H, Vandekerckhove J, Gevaert K. Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics 2008; 8: 1362–1370.

28. Van Damme P, Martens L, Van Damme J et al. Caspase-specific and nonspecific in vivo protein processing during Fas-induced apoptosis. Nat Methods 2005; 2: 771-777.

29. Arnesen T, Van Damme P, Polevoda B, et al. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. Proc Natl Acad Sci USA 2009; 106: 8157-8162.

30. Van Damme P, Van Damme J, Demol H, Staes A, Vandekerckhove J, Gevaert K. A review of COFRADIC techniques targeting protein N-terminal acetylation. BMC Proc. 2009, 3 Suppl 6: S6

31. Gevaert K, Staes A, Van Damme J, et al. Global phosphoproteome analysis on human HepG2 hepatocytes using reverse-phase diagonal liquid chromatography Proteomics 2005; 5: 3589-3599.

32. Ghesquière B, Van Damme J, Martens L, Vandekerckhove J, Gevaert K. Proteome-wide characterization of N-glycosylation events by diagonal chromatography, J Proteome Res. 2006; 5: 2438–2447.

Mecanismos de fragmentación de iones peptídicos mediante disociación inducida por colisión (CID)

Pablo Martínez-Acedo[1], Fernando J. García-Marqués[1], Estefanía Núñez[1,2], Jesús Vázquez[1,2]

[1] Centro de Biología Molecular Severo Ochoa, CSIC-UAM;

[2] Centro Nacional de Investigaciones Cardiovasculares.

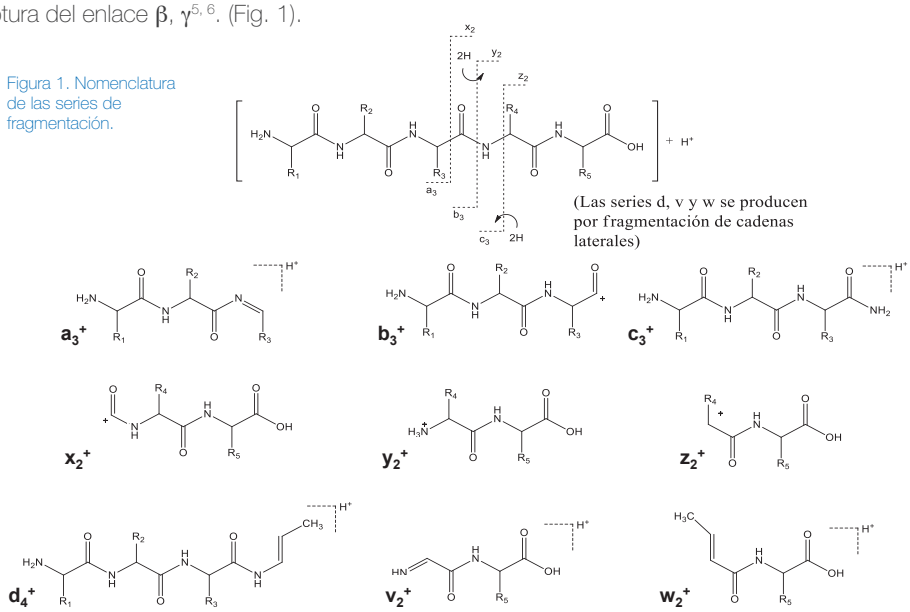
La identificación sistemática y la secuenciación de proteínas mediante espectrometría de masas requieren un alto conocimiento de los mecanismos de fragmentación de péptidos en fase gaseosa. Estos mecanismos se llevan estudiando desde hace algo más de dos décadas y han llevado al establecimiento de unas reglas generales de fragmentación[1, 2]. Este conocimiento de los procesos principales de la fragmentación de péptidos protonados es importante no sólo desde un punto de vista teórico, sino también desde un punto de vista práctico. Los motores de búsqueda utilizados de rutina en Proteómica comparan los espectros MS/MS de los péptidos generados a partir de las proteínas mediante digestión con los espectros teóricos generados a partir de la secuencia de los péptidos *in silico*. El conocimiento preciso de esos mecanismos de fragmentación permite mejorar los motores de búsqueda y aumentar la fiabilidad de la identificación de péptidos. Por otro lado, este conocimiento es fundamental para interpretar los espectros de fragmentación y deducir la secuencia de péptidos cuando no se dispone de información sobre su posible secuencia en las bases de datos ("*secuenciación de novo*").

1. Nomenclatura de las series de fragmentación.

La cadena peptídica está formada por la unión de aminoácidos a través del enlace peptídico entre el grupo carboxilo de un aminoácido y el grupo amino del aminoácido siguiente (Fig.1). En 1984, Roepstorff y Fohlman propusieron una nomenclatura para clasificar los iones fragmento generados por disociación del ión precursor correspondiente a un péptido en un espectrómetro de masas[3], que sería completada por Biemann unos años más tarde[4]. Bajo condiciones de fragmentación de baja energía tipo CID (que se define más adelante en el apartado *Métodos de activación de los iones*), los iones fragmento se generan principalmente mediante una ruptura que tiene lugar a lo largo de la cadena peptídica principal, dando lugar a un fragmento C-terminal y a otro N-terminal. Si la carga original del péptido queda retenida en el fragmento C-terminal, se generan iones fragmento de las series x_n, y_n y z_n dónde n representa el número de aminoácidos que componen el fragmento, y la letra hace referencia al punto de la cadena donde tiene lugar la ruptura (Fig.1). Los símbolos "o ' se utilizan para indicar si hay pérdida o adición de protones en el ión que resulta de la fragmentación: cuando se encuentran a la izquierda

de la letra, el ión fragmento pierde protones, mientras que si se encuentran a la derecha de la letra, el ión fragmento gana protones; así, *y"* significa que tenemos un ión de la serie *y* con 2 H^+ (Peter Roepstorff, comunicación personal). Si por lo contrario, la carga se sitúa en el fragmento N-terminal, se generan los iones fragmento de las series a_n, b_n, y c_n (Fig.1). Los iones de estas series no se generan por pérdida o transferencia de hidrógeno, por lo que los símbolos " o ' no se utilizan. Esta nomenclatura se explica porque, en los comienzos de la aplicación de la espectrometría de masas al análisis de péptidos, el mecanismo de fragmentación se intentaba explicar en términos de la pérdida o transferencia de protones. Los iones más frecuentes son los iones de la serie *b* y de la serie *y*, que se forman precisamente por la ruptura del enlace peptídico. Los iones de las series *c* y *z* se generan por la ruptura entre el carbono α y el grupo amino del aminoácido en posición C-terminal en relación al enlace peptídico, mientras que si la fragmentación se produce entre el grupo carbonilo y el carbono α del aminoácido en posición N-terminal se forman los iones de las series *a* y *x* (Fig.1). Bajo condiciones de CID de alta energía, se generan los iones de las series w_n, v_n y d_n por ruptura en la cadena lateral del aminoácido en posición C-terminal de los fragmentos ya generados. Los iones w_n se forman a partir de los iones z_n por ruptura entre el carbono β y el carbono γ, los iones v_n a partir de los iones de la serie *y* por ruptura del enlace α, β, y los iones de la serie *d* se generan a partir de los iones *(a+1)* por ruptura del enlace β, γ[5, 6]. (Fig. 1).

Figura 1. Nomenclatura de las series de fragmentación.



2. Fragmentación de péptidos mediante CID en un espectrómetro de masas.

La espectrometría de masas en tándem implica la activación de un ión precursor y el análisis de los productos de la fragmentación de ese ión. La activación de los iones es una etapa crucial que define qué tipo de iones se generan y se identifican en un espectro de fragmentación. Se han desarrollado muchas técnicas diferentes para fragmentar péptidos en los espectrómetros de masas; entre ellas, cabe citar la disociación post-fuente (PSD), la disociación inducida por superficie (SID), la disociación por haces de fotones infrarrojos (IRMPD), la disociación por captura de electrones (ECD), la disociación por transferencia de electrones (ETD) y la disociación inducida por colisión con Q pulsante (PQD), siendo la disociación inducida por colisión (CID) la más empleada en los espectrómetros de masas a día de hoy[7]. En este capítulo, nos centraremos fundamentalmente en la técnica CID, que es también la más conocida desde el punto de vista molecular.

La disociación inducida por colisión (*collision-induced dissociation*), también llamada por algunos autores disociación activada por colisión (CAD), es un método que ha ido evolucionando a lo largo de las últimas cuatro décadas, siendo en la actualidad el método de activación más común para fragmentación de péptidos, pudiendo llevarse a cabo en una gran variedad de analizadores[8]. La CID es un proceso que tiene lugar en dos etapas; en la primera, la colisión inelástica entre un ión precursor con alta energía cinética (m_p^+) y un gas inerte (N), generalmente helio o argón, en la celda de colisión produce un aumento de la energía interna del ión precursor (representado por $[m_p^+]^*$ en la Eq.1, donde N' es la forma post-colisión del gas inerte). En la segunda etapa tiene lugar la ruptura de enlaces químicos en el ión precursor activado y la consecuente formación de iones fragmento (proceso esquematizado en la Eq.2, donde m_1^+ y m_n son productos de la fragmentación)[9]. La etapa de activación es por lo general varios órdenes de magnitud más rápida que la etapa de disociación.

$$m_p^+ + N \rightarrow [m_p^+]^* + N' \quad \text{(Eq.1)}$$

$$[m_p^+]^* \rightarrow m_1^+ + m_n \quad \text{(Eq.2)}$$

La fragmentación tipo CID pueden llevarse a cabo de diversas formas, según el instrumento utilizado. Dependiendo de la energía cinética del ión precursor, las técnicas CID se pueden clasificar en dos categorías:

- CID de baja energía (1-100 eV): comúnmente utilizado en cuadrupolos y trampas iónicas tridimensionales o lineales, cuyas celdas de colisión contienen generalmente N_2, Ar ó He. La excitación del ión precursor suele ser vibracional y larga (ms-s) y las colisiones múltiples (excepto en los métodos de excitación por resonancia en las trampas iónicas). El espectro de fragmentación generado suele estar dominado por iones de las series *b* e *y*.

- CID de alta energía (3-10 keV): este método, también denominado HCD, suele usarse en analizadores acoplados a sectores o analizadores tipo TOF u orbitrap, en los cuales los iones precursores tienen una energía cinética más alta y cuya activación es muy rápida (< 10^{-10} s). Se suele usar He como gas de colisión y los espectros de fragmentación generados suelen contener, además de los iones de las series convencionales, iones producidos por fragmentaciones múltiples pertenecientes a las series w_n, v_n y d_n.

3. Mecanismo molecular de la fragmentación por CID.

3.1. Teoría del protón móvil

Entre las diversas teorías que se han propuesto para explicar los procesos de fragmentación de péptidos, la denominada *teoría del protón móvil* es la más aceptada actualmente. Esta teoría asume que en los iones peptídicos formados por los métodos de ionización suave los protones se localizan en los sitios más básicos del péptido, los cuales se sitúan, de acuerdo a su basicidad en fase gaseosa, en la arginina, la lisina, el extremo N-terminal, la histidina y la prolina (Fig.2), en este orden. La activación de los iones puede producir una relocalización de los protones en sitios menos básicos, como los átomos de oxígeno y nitrógeno de los enlaces peptídicos o las cadenas laterales de algunos aminoácidos (Fig.2), que favorece la fragmentación a través de mecanismos de ruptura dirigidos por carga[10]. Se han propuesto otros modelos, como el *mecanismo de competición*[11], que no serán tratados en este capítulo.

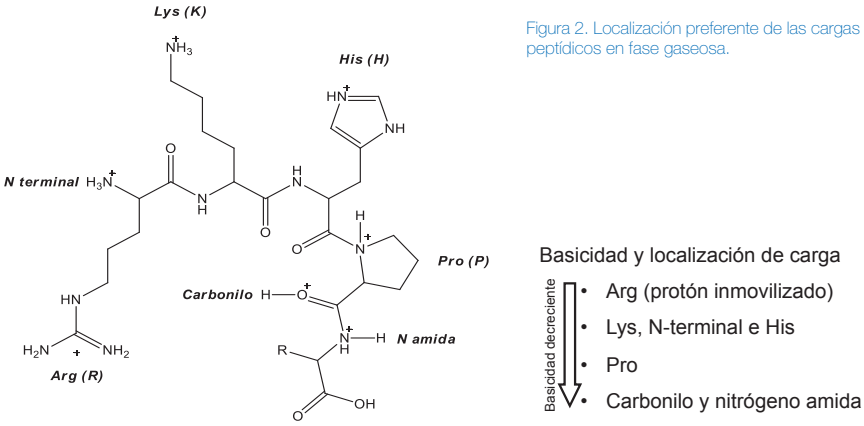


Figura 2. Localización preferente de las cargas en iones peptídicos en fase gaseosa.

3.2. Formación de los iones de la serie *b* y de la serie *y*

Como se ha comentado anteriormente, los iones de la serie *b* y de la serie *y* son los iones fragmento dominantes en los espectros de fragmentación de péptidos protonados generados por CID de baja energía. El mecanismo más aceptado para la formación de los iones de ambas series es común, y está basado en la teoría del protón móvil. Según ella, en los péptidos activados tiene lugar una relocalización de los protones situados en los grupos básicos hacia el nitrógeno amida (Fig.3, arriba a la izquierda) o el oxígeno carbonílico de un enlace peptídico (Fig.3, arriba a la derecha), presentando este último átomo mayor afinidad por los protones[12]. Estas protonaciones debilitan el enlace peptídico, facilitando el ataque nucleófilo del oxígeno del grupo carbonilo del enlace peptídico contiguo (en posición N-terminal) sobre el carbono del grupo carbonilo en el enlace peptídico debilitado, rompiéndose el enlace y formándose un intermediario pentacíclico (Fig.3, abajo a la izquierda) y un ión de la serie *y* (Fig.3, abajo a la derecha) que retiene el protón y mantiene una estructura polipéptidica. Si el protón se transfiere al intermediario pentacícli-

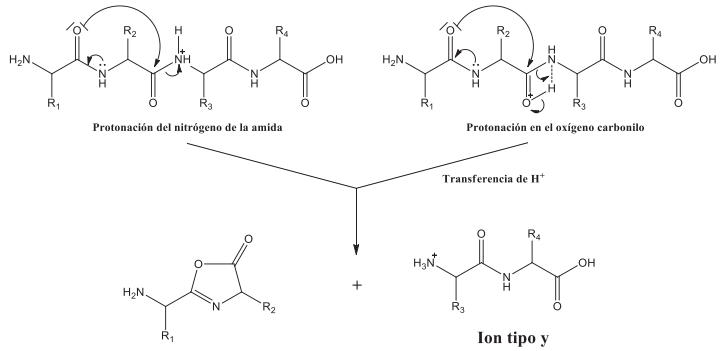


Figura 3. Formación de los iones *y*.

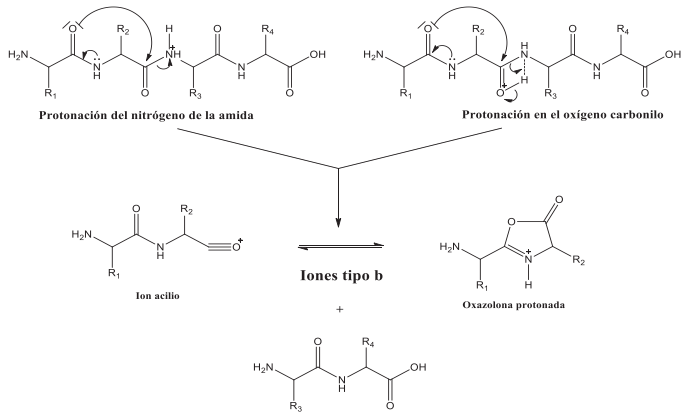


Figura 4. Formación de los iones b.

co durante la ruptura del enlace, se genera un ión de la serie *b* (Fig.4, centro a la derecha), que tiene una estructura de oxazolona protonada, y un péptido que no mantiene el protón (Fig.4, abajo). Se piensa que la estructura cíclica en forma de oxazolona de los iones de la serie *b* está en equilibrio con una estructura de ión acilio algo menos estable (Fig.4, centro a la izquierda), con la que muchos autores representan los iones de esta serie.

3.3. Formación de los iones de la serie *a* y de los iones imonio

Los iones de la serie *b* experimentan una pérdida neutra de CO con mucha facilidad, dando lugar a iones de la serie *a*. Se piensa que esta reacción tiene lugar mediante un movimiento de electrones a través de la estructura del ión *b* en su forma acilio y que finaliza con la pérdida de la molécula CO (Figura 5, centro). En el espectro MS/MS se suelen observar los iones de la serie *b* junto con sus correspondientes iones de la serie *a,* que tienen 28 Da menos. A su vez, los iones de la serie *a* muestran una marcada tendencia a perder un grupo imonio al producirse un ataque nucleófilo del oxígeno del grupo carbonilo del enlace peptídico contiguo (en posición N-terminal) sobre el átomo de carbono del grupo carbonilo C-terminal, produciéndose una ruptura adicional de la cadena peptídica y generándose un nuevo ión de la serie *b* que mantiene la carga (Fig.5, centro a la derecha). Si durante esta ruptura se produce la transferencia del protón al grupo imonio, se produce una oxazolona no cargada y un ión imonio (Fig.5, abajo). Se han propuesto otros mecanismos de formación de los iones imonio que no implican una pérdida neutra de CO y que son característicos de la presencia de ciertos aminoácidos en la secuencia peptídica[13].

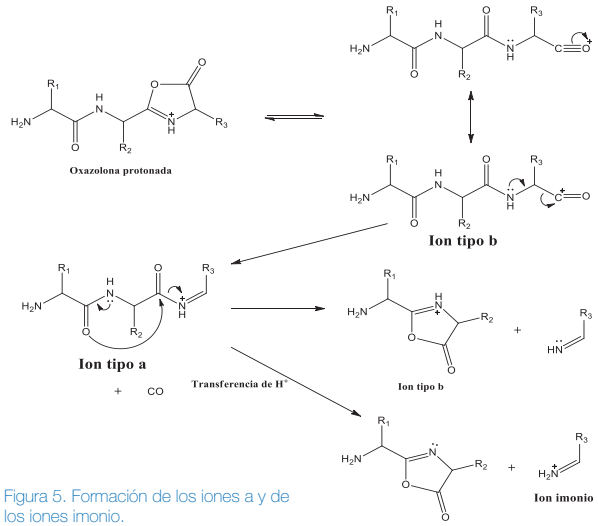


Figura 5. Formación de los iones a y de los iones imonio.

En la Tabla 1 se recogen las masas incrementales de todos los aminoácidos, así como las masas de los correspondientes iones imonio. En la Tabla 2 se resumen las reglas con las que pueden calcularse las masas de los diferentes iones generados a partir de la fragmentación de péptidos, de acuerdo a la estructura de los mismos, detallada en la Fig. 1. En la Tabla 3 se resumen las reglas de fragmentación más comunes, según se explican en los apartados siguientes.

Tabla 1. Masas de los 20 aminoácidos y de sus iones imonio.

Aminoácido	Código	Masa incremental (Da)	Iones Imonio (Da)
Alanina	Ala / A	71.037	44
Arginina	Arg / R	156.101	129
Asparagina	Asn / N	114.043	87
Aspartato	Asp / D	115.027	88
Cisteína	Cys / C	103.009	76
Glutamato	Glu / E	129.042	102
Glutamina	Gln / Q	128.058	101
Glicina	Gly / G	57.021	30
Histidina	His / H	137.059	110
Isoleucina	Ile / I	113.084	86
Leucina	Leu / L	113.084	86
Lisina	Lys / K	128.095	101
Metionina	Met / M	131.040	104
Fenilalanina	Phe / F	147.068	120
Prolina	Pro / P	97.053	70
Serina	Ser / S	87.032	60
Treonina	Thr / T	101.048	74
Triptófano	Trp / W	186.079	159
Tirosina	Tyr / Y	163.063	136
Valina	Val / V	99.068	72

Tabla 2. Reglas para el cálculo de las masas de los iones fragmento.

Fragmento	Masa de los fragmentos (Da)
a_i	Σ masa incremental - 27
b_i	Σ masa incremental + 1
c_i	Σ masa incremental + 18
d_i	Σ masa incremental - cadena lateral - 12
v_i	Σ masa incremental - cadena lateral - 18
w_i	Σ masa incremental - cadena lateral - 17
x_i	Σ masa incremental + 45
y_i	Σ masa incremental + 19
z_i	Σ masa incremental + 2

Tabla 3. Resumen de las reglas de fragmentación por CID.

Resumen de las reglas de fragmentación
• Los iones de las series *b* e *y″* predominan el espectro MS/MS
• La pérdida neutra de CO (-28 Da) de los iones de la serie *b* genera los iones de la serie *a*
• Pérdida neutra de H_2O (-18 Da) favorecida por Cys, Ser, Thr, Glu
• Pérdida neutra de NH_3 (-17 Da) favorecida por Gln, Asn, Lys, Arg
• Los iones imonios en la zona de baja *m/z* son característicos de la presencia de ciertos amino-á
• Pérdida neutra de CH_3SOH (-64 Da) favorecida por Met oxidada
• Pérdida neutra de PO_4H_3 (-98 Da) favorecida por Ser fosforilada y Thr fosforilada, pero no pe
• Ruptura dificultada del enlace Pro-Xxx
• Si un protón móvil está disponible (nº de cargas > nº de Arg)
– Ruptura favorecida del enlace Xxx-Pro
– Ruptura favorecida del enlace His-Xxx
• Si un protón móvil no está disponible (secuestrado por Arg)
– Ruptura favorecida del enlace Asp-Xxx y Glu-Xxx
– Ruptura favorecida del enlace Asp-Pro

4. Fragmentación específica de residuos.

En un estudio publicado en el año 2000, el grupo de Reid[14] examinó 319 espectros MS/MS del digerido de una proteína y observó que el motor de búsqueda SEQUEST fue incapaz de asignar la secuencia de un porcentaje bastante alto (60 %) de espectros de fragmentación, siendo 1/3 de esos espectros péptidos identificados *a posteriori* mediante secuenciación *de novo* o mediante correcta asignación de carga. En otros estudios se han mostrado numerosos patrones de fragmentación atípicos de péptidos en los que aparecen iones que no pertenecen a las series *a/b* e *y*. En base a éstas y a otras observaciones, varios autores han sugerido que los espectros de fragmentación atípicos de péptidos se deben a una fragmentación favorecida por la presencia de ciertos aminoácidos, tales como los ácidos aspártico y glutámico, la prolina y la histidina. Estas fragmentaciones específicas de residuos ocurren cuando el número de protones del ión peptídico es mayor o igual que el número de argininas de su secuencia[15].

4.1. Fragmentación en C-terminal de los residuos ácidos

Numerosos péptidos protonados que contienen ácido aspártico o glutámico tienen un patrón de fragmentación no clásico y sus espectros MS/MS están dominados por iones abundantes de la serie *b* generados en sentido C-terminal respecto a estos aminoácidos[16]. A pesar de que nuestro conocimiento de la fragmentación favorecida por los aminoácidos ácidos ha crecido en los últimos años, los detalles energéticos y cinéticos del mecanismo se desconocen. La implicación del grupo COOH es indudable, pero el mecanismo exacto se desconoce aún. El mecanismo de fragmentación sería independiente de cualquier deslocalización de la carga e implicaría al grupo carboxílico de la cadena lateral de Asp o de Glu, que actúa como donador de hidrógenos (Fig.6). Existen dos teorías principales para detallar el mecanismo. Según Yu et al.[17], la reacción se inicia mediante la transferencia del hidrógeno ácido del grupo carboxílico de la cadena lateral de Asp o de Glu al carbonilo del enlace amida adyacente, seguido de un ataque

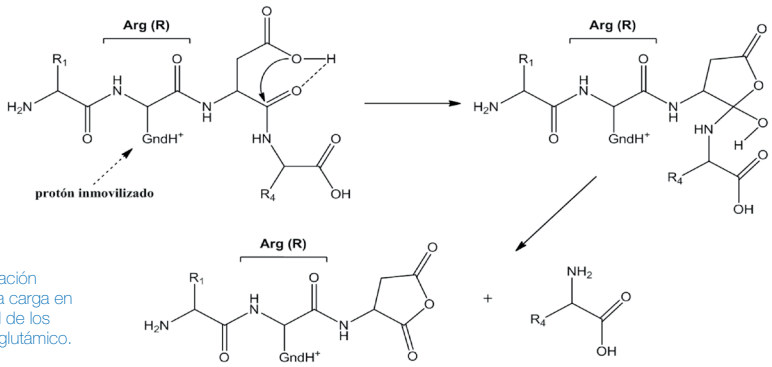


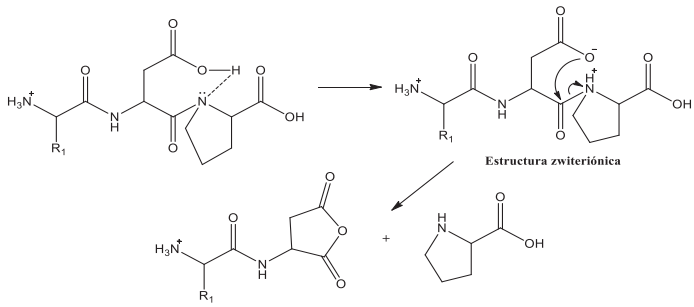
Figura 6. Fragmentación independiente de la carga en posición C-terminal de los ácidos aspártico y glutámico.

nucleófilo sobre el carbono del grupo carbonilo formando un intermediario pentacíclico que se rompe debido al exceso de energía que contiene por la excitación del péptido protonado. Los detractores de este mecanismo alegan que el protón transferido del grupo COOH a la cadena péptica podría "alojarse" en cualquier amida o carboxilo de la cadena peptídica, induciendo fragmentaciones aleatorias a lo largo del péptido. Un mecanismo que no implica movimiento de protones fue propuesto en 2002[18]. La diferencia con el mecanismo anterior es la etapa de iniciación de la reacción: se inicia con la formación de un puente de hidrógeno que debilita el enlace peptídico (Fig.6). El producto final de las reacciones es un anhídrido cíclico con el mismo peso molecular que un fragmento de la serie *b*.

4.2. Fragmentación favorecida del enlace aspártico/prolina

El efecto comentado de la *fragmentación en C-terminal de los residuos ácidos* es particularmente importante en la fragmentación del enlace Asp-Pro, siendo el enlace Asp-Xxx más lábil[17]. Se ha sugerido que el mecanismo de la reacción se inicia a partir del H ácido del aspártico que se transferiría al grupo imida de la prolina[19], formando una estructura zwiteriónica intermediaria (Fig.7). La reacción finaliza con la formación de un anhídrido succínico.

Figura 7. La localización de carga favorece la fragmentación del enlace peptídico aspártico-prolina.



4.3. Fragmentación favorecida en N-terminal de la prolina y desfavorecida en su C-terminal

La presencia de residuos de prolina en un péptido tiene un efecto significativo sobre los fragmentos observados mediante espectrometría de masas en tándem. Este comportamiento inusual, denominado *efecto prolina*[20-24], se ve reflejado en la facilidad para la fragmentación de los péptidos con formación de iones de la series *a*/*b* e *y* en el enlace Xxx-Pro[19]. Como se ilustra en la Fig.8, el efecto prolina en espectrometría de masas de iones positivos es atribuido a la afinidad por los protones del enlace peptídico N-terminal al residuo prolina[25], que, debido a la particular estructura de este aminoácido, que posee una amina secundaria, es superior a la del resto de los enlaces amida del péptido. Por ello, la formación de los iones *b* e *y* mediante ataque nucleófilo del oxígeno del enlace peptídico anterior sobre el grupo carbonilo de la amida protonada está favorecida en este enlace. El ion *y* generado, que posee una prolina protonada en posición N-terminal, se rompe, a su vez, con relativa facilidad, dando lugar a los fragmentos de las llamadas series

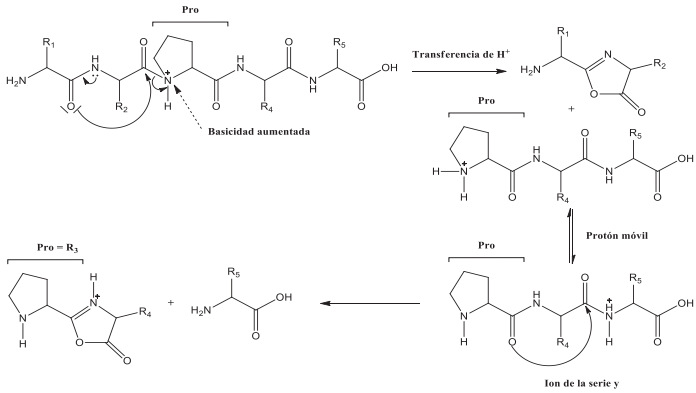


Figura 8. Fragmentación favorecida en N-terminal de la prolina.

internas de la prolina (Fig.8 abajo). Otra característica de los péptidos conteniendo prolina es la baja intensidad o la ausencia de los iones generados por fragmentación en C-terminal de la prolina[19]. Eso se explica por el hecho de que el producto final de la reacción sería un ión bicíclico "tenso", energéticamente desfavorable (Fig.9).

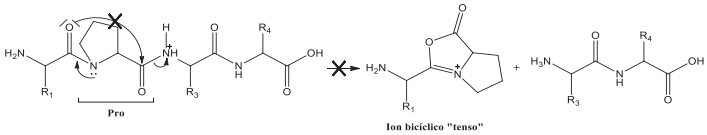


Figura 9. Fragmentación obstaculizada en C-terminal a Prolina.

4.4. Fragmentación favorecida en C-terminal de la histidina

El patrón de fragmentación específico debido a la presencia de histidina es una ruta de fragmentación dirigida por carga, que implica de manera preferente a la cadena lateral de la histidina[10]. Esta fragmentación dirigida por la presencia de histidina tiene lugar cuando el péptido no contiene arginina o, en caso contrario, cuando está doblemente cargado. La primera etapa del mecanismo de la reacción implica la deslocalización del protón unido al grupo imidazol sobre el carbonilo del enlace amida situado en la posición C-terminal contigua a la histidina (Fig.10). La ruptura del enlace peptídico se produce, en este caso, por ataque nucleófilo del nitrógeno de la cadena lateral de la histidina sobre el átomo de carbono del grupo carbonilo protonado. El ión *b* formado es un ión bicíclico estable que no tiene la estructura clásica de una oxazolona[26].

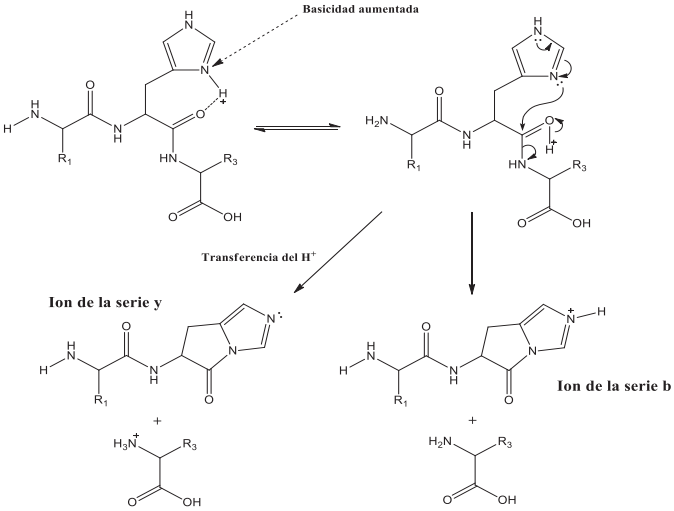


Figura 10. Fragmentación favorecida en C-terminal de la histidina.

5. Casos especiales.

Los espectros de fragmentación de péptidos protonados a menudo contienen iones fragmento debidos a pérdidas neutras de agua, amonio, e incluso del grupo fosfato de los fosfopéptidos. No es extraño encontrarse iones correspondientes al ión precursor que ha perdido H_2O ó NH_3 (18 y 17 Da respectivamente), pero también es frecuente que las pérdidas neutras tengan lugar en iones de las series *b* e *y*. Estas pérdidas neutras muchas veces están favorecidas por la presencia de determinados aminoácidos en la secuencia.

5.1. Pérdida de agua favorecida por serina, treonina y cisteína

La deshidratación debida a la presencia de serina o treonina es una de las mayores causas de pérdida neutra de 18 Da[27]. El mecanismo se inicia con la protonación del oxígeno del grupo hidroxilo de la cadena lateral de Ser o Thr, seguida por el ataque nucleófilo del carbonilo del

enlace amida situado en C-terminal del amino-ácido, formándose un anillo pentacíclico que al perder agua forma una estructura más estable (Fig.11). La presencia de cisteína en la secuencia tiene un efecto similar; en este caso, el ataque nucleófilo sobre el enlace peptídico lo produce el átomo de azufre del grupo tiol[28] (Fig.12)

Figura 11. Pérdida de agua favorecida por serina y treonina.

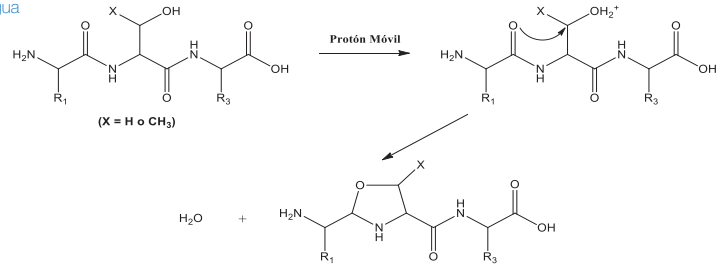
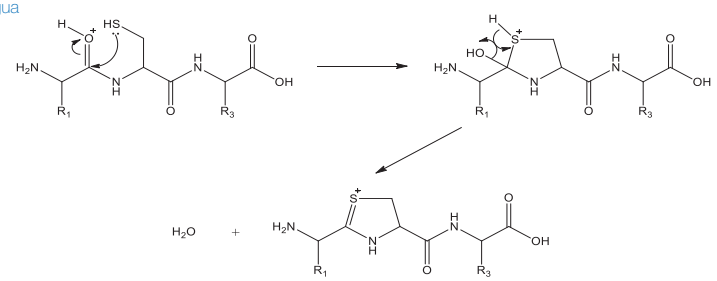


Figura 12. Pérdida de agua favorecida por cisteína.



5.2. Formación de ácido piroglutámico y pérdida de agua a partir de ácido glutámico o de amonio a partir de glutamina

Una de las pérdidas neutras más comunes durante la fragmentación de péptidos es la pérdida de agua o de amonio cuando el extremo N-terminal contiene ácido glutámico o glutamina, respectivamente[29]. La reacción se inicia con la deslocalización del protón del grupo N-terminal hacia la cadena lateral del ácido glutámico o de la glutamina, seguido por un ataque nucleófilo del nitrógeno del mismo grupo N-terminal sobre el carbonilo del grupo ácido o del grupo amino de la cadena lateral, produciéndose la pérdida neutra y un anillo pentacíclico[11] (Fig.13).

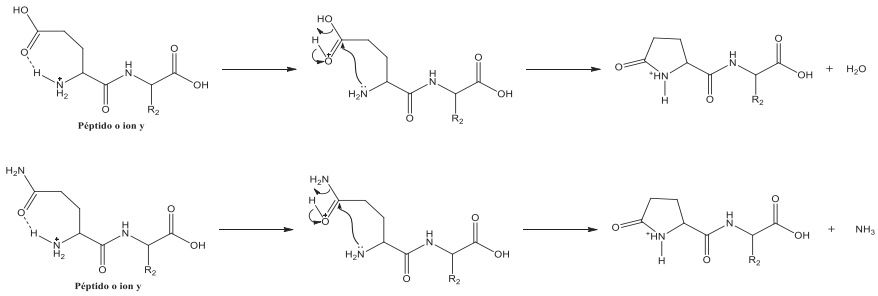


Figura 13. Formación de ácido piroglutámico y pérdida de agua a partir de ácido glutámico o de amonio a partir de glutamina.

5.3. Pérdida de agua favorecida por lisina y arginina

Una característica común de la pérdida de amonio durante la fragmentación de los péptidos es la protonación de la cadena lateral del amino-ácido implicado. En el caso de la pérdida de NH_3 favorecida por Lys o Arg, esa protonación no se debe a una deslocalización de carga ya que los grupos ε-amino y guanidio son sitios de protonación favorecidos (Fig.14). La pérdi-

da de amonio favorecida por lisina se produce probablemente por reacciones de tipo SN_2, llevando a la pérdida del nitrógeno del grupo ε-amino[30]. Si la Lys está en el extremo N-terminal del péptido, el mecanismo más probable de la pérdida de amonio implica el ataque nucleófilo del grupo amino N-terminal sobre la cadena lateral de la Lys, llevando a la formación de ácido pipecólico (Fig.14, izquierda). El ataque nucleófilo del carboxilo en C-terminal de la Lys sobre su cadena lateral también resultaría en la pérdida de NH_3, pero la reacción es cinéticamente desfavorable[30]. Si la Lys se encuentra en medio de la secuencia peptídica, la pérdida de amonio iniciada por el extremo N-terminal es desfavorecida; en este caso la reacción si se produce por el ataque nucleófilo del carboxilo en C-terminal de la Lys sobre el carbono ε de su cadena lateral generando un anillo heptacíclico (Fig.14, derecha). La pérdida de amonio favorecida por Arg se ha estudiado utilizando métodos teóricos y el mecanismo más probable es la substitución de tipo S_N1 en el carbono central ζ del grupo guanidio[31]. El resultado es la formación de un grupo carboimido protonado (Fig.15), estabilizado por el ataque nucleófilo del carbonilo en C-terminal de la Lys para formar un anillo estable.

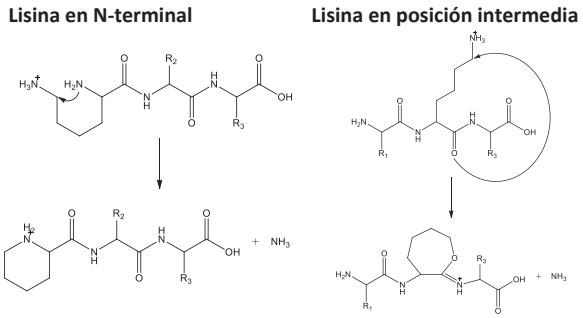


Figura 14. Pérdida de agua favorecida por lisina.

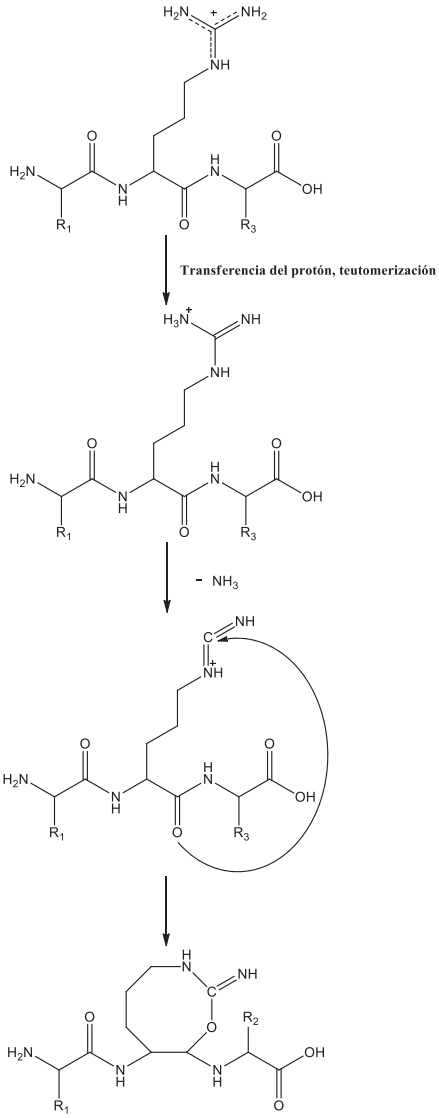


Figura 15. Pérdida de agua favorecida por arginina.

5.4. Oxidación de metionina y pérdida neutra de ácido metanosulfénico

La oxidación de la metionina es una de las modificaciones peptídicas más frecuentes, y tiene lugar de manera espontánea durante el procesamiento de la muestra. La identificación de los péptidos con metionina oxidada es facilitada por la pérdida característica de 64 Da correspondiente a la pérdida neutra de ácido metanosulfénico (CH_3SOH)[32]. Se puede producir por una eliminación 1,2-*cis* (Fig.16, ruta 1), aunque también se puede ser iniciada por una deslocalización del protón del extremo N-terminal a la cadena lateral de la metionina y un ataque nucleófilo del cárbonilo del enlace siguiente o anterior a la metionina sobre la cadena lateral, siendo la eliminación 1,2-*cis* predominante (Fig.16, ruta 2).

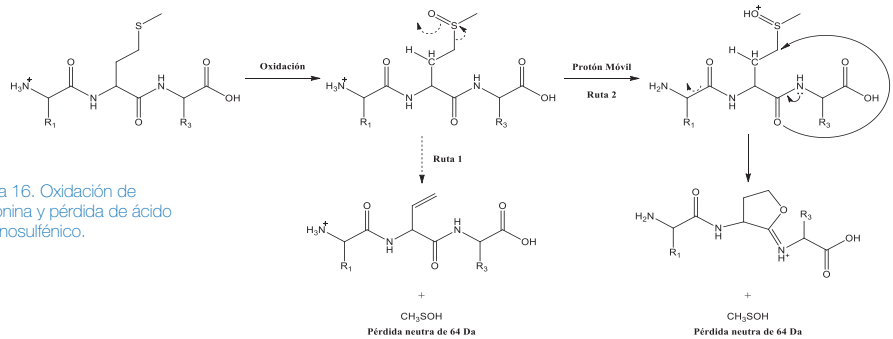


Figura 16. Oxidación de metionina y pérdida de ácido metanosulfénico.

5.5. Formación de iones *b* cíclicos y pérdida neutra de aminoácidos situados en el interior de la secuencia

Se ha observado que no todos los fragmentos peptídicos generados en las trampas iónicas se pueden explicar con las reglas clásicas de fragmentación. Algunos de esos fragmentos anómalos corresponden a la pérdida neutra de aminoácidos internos a partir de iones de la serie *b* y se detectan directamente en el espectro MS/MS en forma de picos relativamente intensos[33]. En nuestro laboratorio hemos propuesto que estos fragmentos atípicos podrían producirse mediante la formación de iones *b* cíclicos, los cuales, a su vez, podrían generarse por dos mecanismos posibles (Fig.17): (1) como se ha comentado anteriormente, los iones de la serie *b* se encuentran en equilibrio entre una forma cíclica (oxazolona) y una forma abierta (ión acilio), de manera que esta última podría formar iones *b* cíclicos de estructura distinta a la oxazolona mediante ataque nucleófilo del nitrógeno del extremo N-terminal del péptido sobre el carbono desapantallado del grupo acilio (ruta I); (2) el nitrógeno del extremo N-terminal podría atacar directamente al mismo átomo de carbono en el grupo amida del péptido intacto, induciéndose directamente la fragmentación y la formación del ión *b* cíclico (ruta II). Esos iones *b* formados por ciclación cabeza-cola pueden reabrirse por enlaces peptídicos diferentes a los iniciales, generando así iones cuya secuencia ha sido reorganizada, y cuya fragmentación libera aminoácidos en los extremos que originalmente estaban en el interior de la secuencia[33].

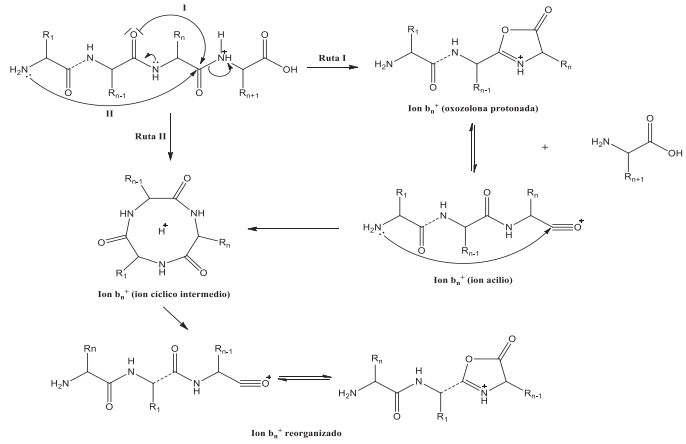


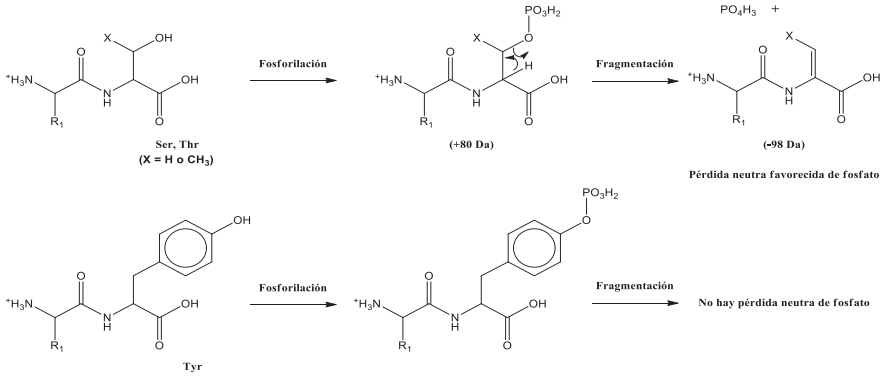
Figura 17. Mecanismo de formación de iones *b* cíclicos y reorganización de los aminoácidos.

5.6. Pérdida de ácido fosfórico en fosfopéptidos

El grupo fosfato de los fosfopéptidos es relativamente lábil y su pérdida por CID es un proceso que necesita poca energía y que compite con la fragmentación de la cadena peptídica. El resultado es un espectro de fragmentación donde suele predominar un pico muy intenso que corresponde al ión precursor que ha perdido 98 Da, que refleja la pérdida neutra de ácido fosfórico (H_3PO_4). La pérdida de H_3PO_4 predomina para los fosfopéptidos que contienen serina fosforilada y en menor proporción para los péptidos con treonina fosforilada[34] , y no tiene lugar en péptidos con fosfotirosina[35]. Se asume que el mecanismo de pérdida neutra del grupo fosfato es una reacción de-eliminación, en la cual el hidrogeno situado en el carbono α del aminoácido

fosforilado se transfiere al grupo fosfato provocando la liberación de H_3PO_4 (Fig.18). Sin embargo, existen otras teorías que implican una deslocalización del protón del carbono α hacía el grupo fosfato seguida de una eliminación de tipo E2 o de una reacción de tipo S_N2[36]. En los péptidos conteniendo tirosina fosforilada el anillo aromático impide la reacción[35] (Fig.18).

Figura 18. Mecanismo de fragmentación de los fosfopéptidos.



REFERENCIAS.

1. Biemann, K. Mass spectrometric methods for protein sequencing. *Anal Chem* 58, 1288A-1300A (1986).

2. Hunt, D.F., Yates, J.R., 3rd, Shabanowitz, J., Winston, S. & Hauer, C.R. Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A* 83, 6233-6237 (1986).

3. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 11, 601 (1984).

4. Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom* 16, 99-111 (1988).

5. Johnson, R.S., Martin, S.A., Biemann, K., Stults, J.T. & Watson, J.T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal Chem* 59, 2621-2625 (1987).

6. Johnson, R.S., Martin, S.A. & Biemann, K. Collision-induced fragmentation of (M + H)+ ions of peptides: side chain specific sequence ions. *Int. J. Mass Spectrom. Ion Proc.* 86, 137–154 (1988).

7. Ahmed, F.E. Utility of mass spectrometry for proteome analysis: part II. Ion-activation methods, statistics, bioinformatics and annotation. *Expert Rev Proteomics* 6, 171-197 (2009).

8. Shukla, A.K. & Futrell, J.H. Tandem mass spectrometry: dissociation of ions by collisional activation. *J Mass Spectrom* 35, 1069-1090 (2000).

9. Sleno, L. & Volmer, D.A. Ion activation methods for tandem mass spectrometry. *J Mass Spectrom* 39, 1091-1112 (2004).

10. Wysocki, V.H., Tsaprailis, G., Smith, L.L. & Breci, L.A. Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom* 35, 1399-1406 (2000).

11. Paizs, B. & Suhai, S. Fragmentation pathways of protonated peptides. *Mass Spectrom Rev* 24, 508-548 (2005).

12. Wu, J. & Lebrilla, C.B. Gas-phase basicities and sites of protonation of glycine oligomers (Glyn; n = 1-5). *Journal of the American Chemical Society* 115, 3270-3275 (1993).

13. Ambihapathy, K., Yalcin, T., Leung, H.W. & Harrison, A.G. Pathways to immonium ions in the fragmentation of protonated peptides. *J Mass Spectrom* 32, 209-215 (1997).

14. Simpson, R.J. et al. Proteomic analysis of the human colon carcinoma cell line (LIM 1215): development of a membrane protein database. *Electrophoresis* 21, 1707-1732 (2000).

15. Tsaprailis, G. et al. Influence of secondary structure on the fragmentation of protonated peptides. *J Am Chem Soc* 121, 5142-5154 (1999).

16. Gu, C., Tsaprailis, G., Breci, L. & Wysocki, V.H. Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of Asp-containing peptides. *Anal Chem* 72, 5804-5813 (2000).

17. Yu, W., Vath, J.E., Huberty, M.C. & Martin, S.A. Identification of the facile gas-phase cleavage of the Asp-Pro and Asp-Xxx peptide bonds in matrix-assisted laser desorption time-of-flight mass spectrometry. *Anal Chem* 65, 3015-3023 (1993).

18. Paizs, B., Suhai, S., Hargittai, B., Hruby, V.J. & Somogyi, A. Ab initio and MS/MS studies on protonated peptides containing basic and acidic amino acid residues: I. Solvated proton vs. salt-bridged structures and the cleavage of the terminal amide bond of protonated RD-NH2. *Int J Mass Spectrom* 219, 203-232 (2002).

19. Breci, L.A., Tabb, D.L., Yates, J.R., 3rd & Wysocki, V.H. Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Anal Chem* 75, 1963-1971 (2003).

20. Hunt, D.F., Yates, J.R., 3rd, Shabanowitz, J., Winston, S. & Hauer, C.R. Protein sequencing by tandem mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 83, 6233-6237 (1986).

21. Loo, J.A., Edmonds, C.G. & Smith, R.D. Tandem mass spectrometry of very large molecules. 2. Dissociation of multiply charged proline-containing proteins from electrospray ionization. *Analytical chemistry* 65, 425-438 (1993).

22. Martin, S.A., Rosenthal, R.S. & Biemann, K. Fast atom bombardment mass spectrometry and tandem mass spectrometry of biologically active peptidoglycan monomers from Neisseria gonorrhoeae. *J Biol Chem* 262, 7514-7522 (1987).

23. Tang, X.J., Thibault, P. & Boyd, R.K. Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation. *Analytical chemistry* 65, 2824-2834 (1993).

24. Vaisar, T. & Urban, J. Probing the proline effect in CID of protonated peptides. *Journal of mass spectrometry : JMS* 31, 1185-1187 (1996).

25. Schwartz, B.L. & Bursey, M.M. Some proline substituent effects in the tandem mass spectrum of protonated pentaalanine. *Biological Mass Spectrometry* 21, 92-96 (1992).

26. Farrugia, J.M., Taverner, T. & O'Hair, R.A.J. Side-chain involvement in the fragmentation reactions of the protonated methyl esters of histidine and its peptides. *Int J Mass Spectrom* 209, 99-112 (2001).

27. Ballard, K.D. & Gaskell, S.J. Dehydration of Peptide [M+H]+ Ions in the Gas Phase. *J Am Soc Mass Spectrom* 4, 477-481 (1993).

28. Reid, G.E., Simpson, R.J. & O'Hair, R.A. Leaving group and gas phase neighboring group effects in the side chain losses from protonated serine and its derivatives. *J Am Soc Mass Spectrom* 11, 1047-1060 (2000).

29. Godugu, B., Neta, P., Simon-Manso, Y. & Stein, S.E. Effect of N-terminal glutamic acid and glutamine on fragmentation of peptide ions. *J Am Soc Mass Spectrom* 21, 1169-1176 (2010).

30. Csonka, I.P., Paizs, B., Lendvay, G. & Suhai, S. Proton mobility and main fragmentation pathways of protonated lysylglycine. *Rapid Commun Mass Spectrom* 15, 1457-1472 (2001).

31. Csonka, I.P., Paizs, B. & Suhai, S. Modeling of the gas-phase ion chemistry of protonated arginine. *J Mass Spectrom* 39, 1025-1035 (2004).

32. O'Hair, R.A.J. & Reid, G.E. Neighboring group versus cis-elimination mechanisms for side chain loss from protonated methionine, methionine sulfoxide and their peptides. *Eur Mass Spectrom* 5, 325-334 (1999).

33. Yague, J. et al. Peptide rearrangement during quadrupole ion trap fragmentation: added complexity to MS/MS spectra. *Anal Chem* 75, 1524-1535 (2003).

34. DeGnore, J.P. & Qin, J. Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J Am Soc Mass Spectrom* 9, 1175-1188 (1998).

35. Boersema, P.J., Mohammed, S. & Heck, A.J. Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom* 44, 861-878 (2009).

36. Palumbo, A.M., Tepe, J.J. & Reid, G.E. Mechanistic insights into the multistage gas-phase fragmentation behavior of phosphoserine-and phosphothreonine-containing peptides. *J Proteome Res* 7, 771-779 (2008).

Secuenciación de novo de péptidos mediante espectrometría de masas: ¿cómo interpretar un espectro MS/MS "a mano"?

Anabel Marina Ramírez[1], Juan J. Calvete[2]

[1] Responsable del Servicio de Proteómica del C.B.M. "Severo Ochoa";

[2] Laboratorio de Venómica Estructural y Funcional, Instituto de Biomedicina de Valencia (CSIC)

1. Introducción.

De novo es una expresión que proviene del latín y significa "nuevo", por lo que comúnmente se define la técnica "*de novo* peptide senquencing" como la obtención de la secuencia de un péptido sin el conocimiento previo de la secuencia de aminoácidos.

La introducción de la secuenciación química de proteínas mediante degradación de Edman [1] en 1950 fue un hito en el desarrollo de la investigación en proteínas. Este método es excelente cuando la disponibilidad de cantidad de péptido es de picomoles. A este nivel, la secuenciación de Edman puede proporcionar la secuencia exacta de aminoácidos, sin ambigüedad, durante un número limitado de ciclos con un tiempo medio por ciclo de 30-50 min. Sin embargo, con bajas cantidades de péptido aparecen "gaps" e incertidumbres en la secuencia.

El número teórico de ciclos útiles (ciclo repetitivo, R) puede calcularse mediante la fórmula:

$$R= [(B/A)^{(1/(NB-NA))}] \times 100$$

donde B es la concentración del aminoácido AA_i en el ciclo NB y A es la concentración del mismo tipo de aminoácido AA_i en el ciclo NA. Por tanto, para calcular R basta con secuenciar el N-terminal de una proteína durante un número suficiente de ciclos para obtener el mismo aminoácido en dos ciclos (NA y NB). Por ejemplo, en la secuencia:

A-H-N-G-Y-S-T-E-E-D-R-K-F-M-Y-V-P..., las concentraciones del aminoácido Y en los ciclos 5 y 15 fueron 135 y 120 picomoles/ciclo, respectivamente. Aplicando la fórmula obtenemos: $R= [(120/135)^{(1/(15-5))}]x100 = [(120/135)^{1/10}]x100 = [(0.888)^{1/10}]x100 = 98\%$. Un rendimiento del 98% en la reacción de Edman implica que en el primer ciclo el 98% de las moléculas secuenciables liberarán su aminácido N-terminal (N1), quedando un 2% de moléculas sin reaccionar. En el segundo ciclo, el 98% de las moléculas acortadas en el primer ciclo, así como el 2% de moléculas intactas, liberarán su aminoácido N-terminal. Así pues, en el ciclo 2 observaremos un rendimiento del (0.98x0.98)x100= 96% de aminoácido 2 (N2). El rendimiento de N3 en el ciclo 3 será (0.98x0.98x0.98)x100= 94%, y así sucesivamente. Por otra parte, los aminoácidos liberados de moléculas "desfasadas" (posición en la secuencia ≠ número de ciclo) contribuirán al

ruido de fondo en ciclos sucesivos. Si consideramos que la degradación de Edman proporciona información útil cuando el rendimiento de nuevo aminoácido secuencial sea superior al 50%, un ciclo repetitivo del 98% permitiría la lectura cierta de 34 resíduos, mientras que para R= 95%, 90% u 88% la secuenciación se extiende, respectivamente, hasta 12, 5, o tan solo 4 ciclos. En la secuenciación de Edman, el rendimiento por ciclo depende de la secuencia, pero también del buen funcionamiento del instrumento. En resúmen, mediante degradación de Edman, en el mejor de los casos, obtendremos una secuencia N-terminal de unos pocos aminoácidos, bien sea de una proteína intacta o de sus péptidos proteolíticos purificados.

El refinamiento gradual de las técnicas de fragmentación de péptidos mediante espectrometría de masas (MS/MS) desde finales de la década de 1950 [2], crearon las bases para la secuenciación de péptidos sin necesidad de previa purificación, proporcionando la obtención, de un modo rápido, de secuencias internas de proteínas mediante el análisis de péptidos proteolíticos [3-5]. La espectrometría de masas en modo MS/MS tiene una mayor sensibilidad además de no necesitar estándares externos para cada aminoácido o variantes de los mismos. Otra ventaja es que la secuenciación MS/MS no se ve afectada por un "bloqueo" del N-terminal, como es el caso de la secuenciación de Edman. Una de las limitaciones de la MS es la dificultad para distinguir masas isobáricas o masas muy próximas como por ejemplo K y Q, si el espectrómetro de masas es de baja exactitud de masa y/o baja resolución.

Durante las dos últimas décadas, la determinación de secuencias de proteínas mediante MS/MS se ha vuelto más y más potente, desarrollo llevado a cabo gracias a las mejoras de las técnicas LC (cromatografía líquida) en combinación con la aparición de la ionización ESI (*electrospray*) [6]. Las ventajas de las técnicas de LC-MS/MS en relación a la velocidad, sensibilidad y aplicabilidad a mezclas complejas de péptidos, han reemplazado, gradualmente, a la degradación de Edman. Como resultado, la proteómica basada en la espectrometría de masas, ha emergido como el método para la identificación de proteínas [7, 8] apoyándose en las bases de datos y empleando motores de búsqueda como MASCOT [9], SEQUEST [10], OMSSA [11], etc.

Si tenemos en cuenta que de todas las combinaciones de secuencias de aminoácidos teóricamente posibles, solamente una pequeña proporción de secuencias de proteínas se encuentra en la naturaleza, una secuencia corta de un péptido es altamente específica para una proteína. Es decir, la secuencia obtenida a partir de un espectro MS/MS de un péptido con alta probabilidad y respaldada por bases de datos permite la identificación de una proteína con un alto nivel de fiabilidad.

El problema aparece cuando la secuencia del/los péptidos no está en ninguna base de datos. A pesar del crecimiento continuo de las bases de datos, la secuenciación *de novo* de péptidos es necesaria en muchas situaciones:

- Organismos cuyos genomas no están secuenciados o lo están solo parcialmente.
- Análisis de variantes de secuencias de proteínas o isoformas obtenidas por "splicing".
- Análisis de péptidos con residuos no proteicos o aminoácidos modificados.

Es necesario abordar la secuenciación *de novo* preparados para enfrentarnos a todos sus retos así como a todas sus ventajas. A pesar de ser científicos, debemos tener "cierta fe" en las secuencias obtenidas *de novo* sin el conocimiento previo de la misma, este es el corazón de la secuenciación *de novo*.

Especialmente cuando empleamos un software, necesitamos tener la suficiente confianza para decir que la secuencia obtenida con mayor puntuación es la más correcta. Para esto es importante testar nuestras habilidades así como las del software para interpretar secuencias, empleando secuencias conocidas como control.

En este capítulo haremos un repaso del estado y los avances actuales en la secuenciación *de novo* terminando con un tutorial de interpretación de un espectro MS/MS en modo manual.

2. Nomenclatura de los iones fragmentos de péptidos.

La introducción en la espectrometría de masas de técnicas de ionización suave ha permitido la generación de iones de péptidos intactos los cuales pueden ser seleccionados y activados con el fin de detectar sus fragmentos. Para el análisis de péptidos, se emplea generalmente el modo positivo que proporciona una mayor riqueza de iones fragmentos específicos de secuencia en los espectros MS/MS. La fragmentación de iones protonados mediante CID (*collision-activated dissociation*) suele darse en el enlace peptídico mediante reacciones inducidas por el protón de acuerdo a la teoría del protón móvil [12].

Los fragmentos peptídicos más comunes observados en colisiones de baja energía (CID) son **a**, **b** e **y"** (Figura 1A). Los iones b contienen el extremo N-terminal y los iones y" contienen el extremo C-terminal. Los fragmentos a, procedentes de los b, se dan con mucha menor frecuencia y abundancia en relación a estos iones b. Los iones a son, a menudo, empleados para el diagnóstico de iones b, frecuentemente se observan parejas a-b en el espectro de fragmentación. Estas parejas están separadas por 28u, la masa para el grupo carbonilo, C=O. La nomenclatura más ampliamente aceptada para la anotación de la secuencia peptídica es la que originalmente propuso Roepstorff and Fohlman [13] y que fue posteriormente modificada por Biemann [14].

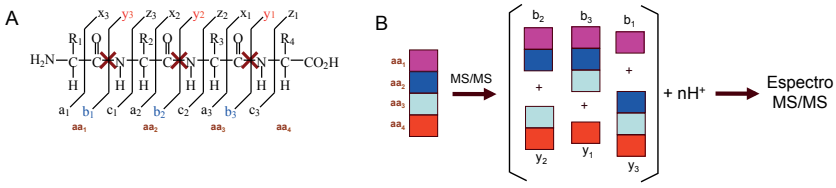


Figura 1: **A**) Nomenclatura de fragmentos de péptidos específicos de secuencia; los iones a-, b-, y c- contienen el extremo N-terminal; los iones x-, y-, y z- contienen el C-terminal. R1, R2, R3 y R4 representan las cadenas laterales de los aminoácidos. Figura 1: **B**) Esquema de los posibles fragmentos específicos de secuencia (b e y) formados por disociación por colisión (CID) de un tetrapéptido. Nótese que todos los fragmentos coexisten en el analizador pero únicamente aquellos que resulten protonados serán visibles en el correspondiente espectro de fragmentación. La intensidad relativa de los iones en el espectro MS/MS reflejará, pues, los sitios de fragmentación preferentes y el reparto entre los diferentes fragmentos peptídicos formados.

Los péptidos no se fragmentan secuencialmente, es decir, el primer evento de fragmentación no comienza en el extremo amino terminal para continuar secuencialmente por el residuo siguiente de la cadena aminoacídica. Los eventos de fragmentación son un poco al azar y no secuenciales. En otras palabras, diferentes moléculas del ión aislado en la cámara de fragmentación se fragmentarán en diferentes posiciones a lo largo de la cadena polipeptídica, generando una población heterogénea de fragmentos (Figura.1B). Sin embargo, solo aquellos fragmentos que retengan los protones que el péptido parental tomó prestados durante el proceso de ionización podrán ser observados en el espectro de masas. Dado que algunas fragmentaciones son preferentes respecto a otras y que el reparto de protones entre los fragmentos complementarios no tiene por qué ser equitativo, los picos de un espectro MS/MS no presentan la misma intensidad relativa (Fig.2).

El conjunto de picos que se observa en un espectro de fragmentación es un reflejo de la población de iones fragmento producidos en la celda de colisión del espectrómetro de masas ordenados de acuerdo a su relación m/z. Es precisamente esta última característica, la ordenación

espectral de los iones fragmento, la que permite deducir la secuencia del péptido parental utilizando la diferencia de masa entre los picos del espectro de fragmentación para asignar eventos de disociación que relacionen unos iones fragmento con otros de su misma serie. Y aunque pudiera parecer que la coexistencia en el mismo espectro de iones **y”** y **b**, complique la interpretación, veremos que cuanto más completas sean las series de iones menor será la posibilidad de interpretaciones erróneas, al permitir leer la secuencia tanto hacia delante (de N-terminal a C-terminal) como hacia atrás.

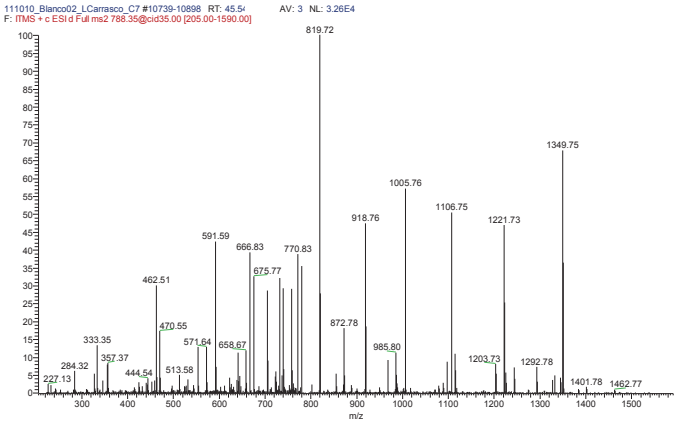


Figura 2. Espectro CID MS/MS del precursor $[M+2H]^{2+}$ 788,35 obtenido en un espectrómetro de masas LTQ-Velos

La serie de iones que se extiende desde el N-terminal es la de los iones b, de tal forma que el fragmento que contiene solo el aminoácido N-terminal sería el b_1, el que contiene los dos primeros aminoácidos N-terminales sería el b_2 y así consecutivamente. De forma similar los iones fragmentos de la serie y”, se extienden desde el extremo C-terminal.

Se puede calcular la masa de cualquier ion b; su valor es la masa del péptido (M) menos OH, o -17:

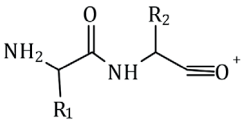
M -17 = ion **b**

bn = (M+1) - 18

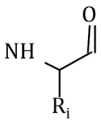
Siendo n = nº total de aminoácidos del péptido.

Ejemplo de ion b:

b = ∑ MRi + 1

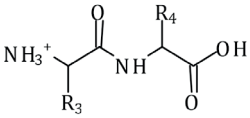


MRi =



Ejemplo de ion y”:

y” = ∑ MRi + 19



Para calcular el valor m/z de un ion y” basta con calcular su $M+H^+$ para cada fragmento.

y”n = M + 1

3. Importancia de la exactitud de masa.

La fiabilidad en la secuencia de un péptido mejora notablemente con una buena exactitud de masa, tanto para los iones fragmento como para el ion precursor. Actualmente existen instrumentos con valores de resolución de masa de 100.000-500.000 y de exactitud de masa en el rango de 2-0,2 ppm (partes por millón), entre ellos el FT-ICR y el OrbiTrap [15].

Una de las ventajas que salta a la vista es la posibilidad de poder discernir entre K y Q (36 mDa de diferencia), entre W y EG (15 mDa de diferencia) y entre pares de aminoácidos con masas muy similares. Esto último es muy importante para averiguar la secuencia de los extremos del péptido ya que, en muchas ocasiones, la única información que nos dan los fragmentos es la de una pareja de aminoácidos. Ello es particularmente notable en la serie b, ya que la formación del b_1 está muy desfavorecida y fundamentalmente se obtiene el b_2. También es crítico cuando en la secuencia hay aminoácidos que no se pueden asignar bien porque la diferencia de masa es de 1 Da (N y D, K/Q y E, I/L y N), si estos aminoácidos están en los extremos y la masa del precursor no tiene una buena exactitud se puede dar una mala asignación de la secuencia.

4. Fragmentación de péptidos mediante CID.

La mayor parte de la secuenciación *de novo* se ha llevado a cabo empleando CID. Mediante este tipo de colisión de baja energía la activación de los iones se realiza con moléculas de gases inertes como el He, N_2 o Ar ya sea en una celda de colisión o en una trampa iónica. La energía traslacional de los iones se convierte, en parte, en energía interna vibracional, la cuál induce la fragmentación peptídica. Para péptidos no modificados, la activación de moléculas protonadas da lugar, principalmente, a fragmentación en el enlace peptídico obteniéndose fragmentos específicos de estructura. La teoría del protón móvil justifica la distribución de los fragmentos observados [16, 17]. Una particularidad del corte en el enlace peptídico de iones protonados de forma múltiple es la obtención de fragmentos b e y" complementarios. Sin embargo, este hecho es más una excepción que una regla ya que la mayoría de los cortes están solo representados por una de las dos series encontrando un leve solapamiento entre ambas.

Distintas situaciones:

A. Péptido tríptico, básico en su extremo C-terminal. Predomina la serie y"

B. Péptido con sitios básicos en la región N-terminal. Predomina la serie b

C. Péptido con aminoácido básico en N y C-terminal. Series b e y" sin apenas solapamiento, con similar intensidad.

Estas reglas "generales" tienen sus particularidades dependiendo de

- **La naturaleza del aminoácido básico** ya que no tienen la misma afinidad por el protón, siendo R el aminoácido que mayor afinidad presenta, seguida de K y H. Por otro lado podemos tener péptidos con aminoácidos básicos a lo largo de la secuencia a la par que en sus extremos, como es el caso de los tríptico con algún "miscleavage". La variabilidad en la presencia de una u otra serie dependerá también de la naturaleza de los residuos básicos. En la práctica no son muy efectivos los cortes cercanos a R o K debido

a que secuestran el protón por lo que los iones correspondientes estarán en una baja proporción. La presencia de dos o más residuos de R contiguos o cercanos da lugar a "huecos" en las series de fragmentación de cualquier tipo debido a la alta basicidad de R.

- **El tipo de celda donde se da la CID,** bien sea una celda de colisión bien una trampa de iones (ya sea tridimensional o lineal)**.** La trampa de iones es un excelente analizador que genera espectros muy reproducibles dominados por la primera generación de fragmentos, es decir, tras la formación de un ion fragmento este no continúa fragmentándose. Se obtienen así espectros con series b e y" bastante solapadas lo que genera información redundante que da más fiabilidad a la secuencia obtenida. La limitación de estos analizadores está en la región baja del espectro, la denominada "regla del tercio" que dice que solo veremos fragmentos mayores que 1/3 del valor m/z del precursor. Este problema se puede solventar realizando subframentación (MS^n), o empleando PQD (pulsed Q-dissociation) [18]. Instrumentos híbridos como el LTQ-Orbitrap incorporan lo que se denomina la celda HCD (*high collision dissociation*) permitiendo la obtención de fragmentos en la región baja del espectro [19]. La trampa iónica es el instrumento ideal (incluso los primeros modelos del mercado) para la distinción entre I y L. El ion imonio de estos aminoácidos (86 u) se puede obtener subfragmentando iones del espectro MS/MS dada la capacidad MS^n de este analizador, un posterior análisis MS/MS de este ion da lugar a un fragmento de 69 m/z en el caso de I pero no de L [20-22].

5. Otras técnicas de fragmentación.

El hecho de que la ionización **MALDI** genere normalmente iones mono-cargados hace que sea un método menos atractivo que la ionización ESI para la secuenciación *de novo*, ya que los espectros de fragmentación obtenidos contienen "a priori" menor información. No obstante, el desarrollo y acoplamiento de analizadores como el TOF-TOF, QTOF y Orbitrap [23, 24] ha mejorado considerablemente la riqueza de los espectros MS/MS en relación a la primera técnica que se introdujo para la detección de fragmentos mediante MALDI-TOF conocida como PSD [25].

De forma relativamente reciente, se han introducido nuevas técnicas de activación para la fragmentación de péptidos, la "Disociación por Captura Electrónica" o ECD [26] y la "Disociación por Trasferencia electrónica" o ETD [27]. Ambas técnicas trasfieren un único electrón de baja energía a un péptido protonado múltiple, normalmente +3 o +4. El corte afecta principalmente al enlace N-C por lo que los fragmentos resultantes son de la serie c o z (Fig. 1). Los espectros obtenidos mediante estas técnicas son idóneos para la caracterización de modificaciones post-traduccionales [28] ya que no se han observado fragmentaciones de cadenas laterales, incluso estructuras lábiles como la fosforilación permanecen intactas con estas técnicas.

6. Modificación química y marcaje isotópico estable en la secuenciación de novo.

El uso de estas técnicas facilitan la secuenciación *de novo* ya que permiten marcar específicamente uno u otro extremo del péptido, por lo que podremos asignar correctamente la serie y" o la serie b.

El marcaje más comúnmente utilizado para el extremo C-terminal se realiza mediante la incorporación de ^{18}O durante la digestión tríptica empleando $H_2{}^{18}O$. De este modo se obtiene un marcaje selectivo de los iones y". La digestión empleando una mezcla de $H_2{}^{16}O$ y $H_2{}^{18}O$ genera iones y" con un patrón isotópico característico [29-31] que facilita su distinción de los iones b no marcados. El marcaje ^{18}O puede ser introducido en un paso posterior tras la digestión [32].

Para marcar el extremo N-terminal se suelen emplear reactivos que modifican su grupo amino primario convirtiéndolo en una amina secundaria o terciaria, de esta forma aumenta su basicidad y, en consecuencia, la tendencia a retener la carga. Esto se traduce en la estabilización de los iones b [33].

7. Secuenciación de novo automática.

El desarrollo de herramientas informáticas para la interpretación de espectros experimenta un desarrollo continuo [34-40]. Muchas de las versiones básicas de estas herramientas se encuentran accesibles en la red. A pesar de que suele ser necesario revisar los resultados que se obtienen manualmente, son programas muy útiles cuando se trabaja con ficheros que contienen un número alto de espectros MS/MS. Normalmente buena parte de la región central de la secuencia se interpreta correctamente, siendo los extremos del péptido los que generan las inconsistencias entre un resultado automático y uno manual. Pero la información proporcionada es muy útil ya que se pueden utilizar otros programas basados en encontrar homologías de secuencias empleando las regiones de los péptidos correctamente interpretadas, de este modo se pueden identificar proteínas homólogas cuyas secuencias sí estén en las bases de datos. Entre estos programas, BLAST (*basic local alignment search tool*) o FASTA (*fast all*) son los más comúnmente utilizados [41-43], es más, algunos *software*s de interpretación automática *de novo* incorporan la búsqueda de homologías mediante BLAST o FASTA para presentar un resultado contrastado con las bases de datos.

Es evidente que para proyectos grandes de secuenciación *de novo* se hace casi imprescindible el apoyo en herramientas bioinformáticas que van a facilitar mucho el trabajo y a disminuir considerablemente el tiempo de interpretación.

8. Interpretando un espectro MS/MS.

8.1. Reglas

Antes de comenzar es importante remarcar que un espectro de fragmentación (MS/MS) es un puzzle al que, normalmente, le faltan piezas (iones fragmento). El éxito de una interpretación *de novo* depende mucho de la "calidad" del espectro que estamos analizando. A este respecto insistiremos en la interpretación de un espectro en particular si es "el espectro" que nos interesa por distintos motivos, por ejemplo el caso de un espectro candidato para un péptido con una modificación postraduccional determinada; en el caso de estar trabajando en la obtención de secuencias de una proteína no presente en base de datos, elegiremos el/los espectros de mayor calidad obtenidos en el análisis.

Asumiendo que el equipo de espectrometría de masas que hayamos empleado trabaje en condiciones óptimas los **factores limitantes** serían los siguientes:

- **Longitud del péptido.** Un péptido muy grande (>3000 Da) generará fragmentos fuera de nuestro rango m/z (para una valor de z ≤ 3), mientras que un péptido muy pequeño (<800 Da) proporcionará una secuencia con un nº de aminoácidos insuficiente para identificar la proteína con fiabilidad.

- **Carga del péptido.** El espectro MS/MS de un pépido con carga 2, $[M+2H]^{2+}$, es el idóneo ya que contiene mayor información, principalmente fragmentos con carga 1, además, tiene menor complejidad en la región baja m/z respecto a un péptido con carga 3 y menor ruido químico que un péptido con carga 1 (donde los espectros son pobres puesto que solo hay una carga que se deslocaliza hacia los enlaces más alejados dando lugar a un bajo nº de fragmentos).

- **Péptido tríptico.** El hecho de que el aminoácido C-terminal sea una K o una R permite que las cargas se deslocalicen desde ambos extremos (N-terminal y C-terminal), si la especie es $[M+2H]^{2+}$, hacia los distintos enlaces peptídicos, lo cual favorece la ruptura potencial de todos los enlaces aumentando así el nº de iones fragmentos (la riqueza del espectro) necesarios para interpretar la secuencia.

- **Secuencia del péptido.** Como veremos posteriormente, existen rupturas favorecidas y desfavorecidas en función de la secuencia, lo que puede mejorar o deteriorar la calidad del espectro.

¿Cómo podemos **mejorar la calidad de un espectro**?:

- **Trabajando en modo off-line.** Este sería el modo óptimo si disponemos de una muestra poco compleja ya que permite optimizar todos los parámetros relacionados con la eficiencia de ionización y de fragmentación en tiempo real hasta obtener la mejor calidad de espectro posible. Otra ventaja es la posibilidad de promediar un nº elevado de espectros MS/MS para el mismo ion precursor lo que aumenta considerablemente la riqueza de fragmentos y mejora mucho la relación señal:ruido. Además, si trabajamos con un analizador de trampa iónica, podremos realizar experimentos MSn, en muchos casos, imprescindibles para la correcta asignación de una secuencia.

- **Trabajando en modo on-line.** Los parámetros de ionización y de fragmentación, en este modo, se optimizan para un péptido "tipo", en la mayor parte de los casos, uno tríptico. Normalmente la energía de colisión está normalizada para adecuarla al valor m/z del precursor. En este modo podemos promediar, después del análisis, todos los espectros MS/MS para un mismo ion precursor que haya realizado la máquina en función del método utilizado. En el caso de un analizador de trampa iónica es posible programar métodos que incluyan MS^n.

Pérdida de Agua y de Amonio [35, 36].

- Los iones b e y" que contengan los aminoácidos R, K, Q y N tienen tendencia a perder amonio, -17. Hablamos entonces de series **b*** e **y***.

- Los iones b e y" que contengan los aminoácidos S, T y E pueden perder agua, -18. En el caso del ácido glutámico se da únicamente en el caso de ser este aminoácido el N-terminal. Hablamos entonces de series **bº** e **yº**.

Intensidad de iones fragmento en el espectro.

- La serie b se ve desfavorecida a la par que se favorece la serie y" en los siguientes casos:

X-P X-G X-H X-K X-R

Nota: Si el protón que se deslocaliza proviene del extremo C-terminal también se observa serie b gracias a la carga que se deslocaliza desde el N-terminal (para un péptido doblemente cargado).

- Fragmentaciones favorecidas:

X-P X-G X-S H-X D-X N-G

En el caso de la existencia de D en la secuencia se da un mecanismo denominado de "carga remota" [44], este implica trasferencia de H^+ de la cadena lateral del D al enlace amida adyacente que luego se rompe favoreciéndose la serie b.

Nota: Como en el punto anterior, la ruptura favorece una serie u otra en función de la avidez por el H^+ de los aa implicados en el enlace, eso no impide que se vea la otra serie cuando el péptido sea una especie +2 o +3.

- Cuando hay una R dentro de la secuencia los iones b* e y* suelen ser más abundantes que el b e y" correspondientes.

X-R-X

- Serie interna de Prolina (también se da con Histidina) [45]. Dado que la ruptura en el enlace X-P está muy favorecida para generar serie y" por la gran afinidad por el H^+ de la prolina, el fragmento y" conteniendo la prolina suele ser el más intenso y mayoritario en el espectro. En este caso se puede generar serie b a partir de este fragmento y". Por ejemplo, para el péptido AFTLPGISNK, el fragmento y" prominente sería: PGISNK que podría dar lugar a los fragmentos de la serie b PGISN, PGIS, PGI, etc. La nomenclatura para estos fragmentos es: (B**x**Y"**y**)**n**, donde **y** sería nº de fragmento y" correspondiente, **x** el nº de fragmento b correspondiente y **n** el nº de aminoácidos del fragmento interno. Pej.- PGISN = $(B_9Y_6)_5$

- Cuando existe una P en la secuencia es común que haya un intercambio de intensidad para los iones b e y". Este hecho se da cuando en la secuencia hay residuos básicos del tipo H, K o R.

Composición de aminoácidos.

Iones Imonio.- Es posible encontrar estos iones en la región baja del espectro pudiendo dar una pista de la composición amino acídica de un péptido. El hecho de que no se vea el ion imonio para un aminoácido no significa que el aminoácido no esté en la secuencia. De hecho estos iones no son comunes en fragmentaciones CID a baja energía.

Estos iones son fragmentos "internos" con un solo residuo que pierden C = O. Por ejemplo, para la lisina cuya masa monoisotópica es de aproximadamente 128 Daltons, su ion imonio sería:

$$MRi + 1 - 28 = 128 + 1 - 28 = 101$$

Masas isobáricas.

- Leucina e Isoleucina tienen masas isobáricas y no pueden diferenciarse empleando colisiones de baja energía. Cuando se observa una diferencia de masas correspondiente a este aminoácido se marca con una X o Lxx, según la nomenclatura de Hunt.

- Lisina y Glutamina tienen masas muy parecidas, 128,09496 y 128,05858 respectivamente. La diferencia de masa (0,03638) puede distinguirse en un espectrómetro de masas con alta exactitud y resolución, no suele ser el caso de los Triple-Cuadrupolos o las Trampas Iónicas. Para estos equipos se puede emplear la estrategia de acetilar la lisina, que incrementa su masa en 42u. Otra opción es asumir que un péptido tríptico tiene más posibilidades de contener en su secuencia una glutamina que una lisina a menos que el aminoácido posterior sea prolina o un aminoácido ácido. Esta asunción hay que tomarla con mucha precaución.

- Hay casos en los que dos residuos tienen la masa cercana o igual a la de un aminoácido, o a la de un aminoácido modificado (Tabla 2).

8.2 Protocolo "paso a paso".

Vamos a ver las etapas para la interpretación de un espectro MS/MS *de novo* utilizando un espectro real. Por los motivos que se han detallado anteriormente el espectro seleccionado es un $[M+2H]^{2+}$ tríptico (los más comunes en proteómica) obtenido en un equipo LTQ-Velos.

El material que vamos a emplear es sencillo, lápiz, goma de borrar, calculadora, tabla de aminoácidos (Tablas 1 y 2) y tabla de pares de aminoácidos (Tabla 3). Todo esto, más un poco de paciencia, hará de la interpretación *de novo* un excelente pasatiempo (en algunos casos, incluso ¡adictivo!).

Tabla 1. Masas monoisotópicas de los aminoácidos más comunes.

AACodes		Mono.	AACodes		Mono.
Gly	G	**57**.021464	Asp	D	**115**.02694
Ala	A	**71**.037114	Gln	Q	**128**.05858
Ser	S	**87**.032029	Lys	K	**128**.09496
Pro	P	**97**.052764	Glu	E	**129**.04259
Val	V	**99**.068414	Met	M	**131**.04048
Thr	T	**101**.04768	His	H	**137**.05891
Cys	C	**103**.00919	Phe	F	**147**.06841
Leu	L	**113**.08406	Arg	R	**156**.10111
Ile	I	**113**.08406	CMC		**161**.01467
Asn	N	**114**.04293	Tyr	Y	**163**.06333
			Trp	W	**186**.07931

Tabla 2. Masas monoisotópicas conflictivas por corresponderse con modificaciones o parejas de aminoácidos.

Amino Acids	Mono		Mono	delta mass
Val	**99**.068414	acetyl-Gly	**99**.032034	0.03638
Leu	**113**.08406	Ile	**113**.08406	0
Lxx	**113**.08406	acetyl-Ala	**113**.047684	0.036376
Asn	**114**.04293	Gly-Gly	**114**.04298	0.000002
Gln	**128**.05858	Lys	**128**.09496	0.03638
Gln	**128**.05858	Gly-Ala	**128**.058578	0.000002
Lys	**128**.09496	Gly-Ala	**128**.058578	0.036382
Glu	**129**.04259	acetyl-Ser	**129**.042599	0.000009
Phe	**147**.06841	Met Sulfoxide	**147**.0354	0.033
Arg	**156**.10111	Val-Gly	**156**.089878	0.011232
Arg	**156**.10111	acetyl-Asn	**156**.0535	0.04761
Tyr	**163**.06333	Met Sulfone	**163**.0303	0.033
Trp	**186**.07931	Ala-Asp	**186**.064054	0.015256
Trp	**186**.07931	Ser-Val	**186**.100443	0.021133
Trp	**186**.07931	Gly-Glu	**186**.064054	0.015256

Tabla 3. Masas de todas las posibles combinaciones de parejas de aminoácidos (AAn) comunes y de iones y_2. Nótese que la masa de un ión b_2 (N-terminal, N) equivale a $(AA_{N1}\text{-}AA_{N2})+H^+$ y que un ión y_2 (C-terminal, C) tiene masa = $(AA_{C2}\text{-}AA_{C1})+18+H^+$, siendo AA una masa incremental de aminoácido listada en la Tabla 1.

	Gly (AA_1-AA_2)	Ala y2	Ser	Pro	Val	Thr	Cys	Ile Leu	Asn	Asp	Lys Gln	Glu	Met	His	Phe	Arg	Tyr	Trp
Gly	114	133 128	144	154	156	158	160	170	171	172	185	186	188	194	204	213	220	243
Ala	147	142	161 158	168	170	172	174	184	185	186	199	200	202	208	218	227	234	257
Ser	163	177	174	193 184	186	188	190	200	201	202	215	216	218	224	234	243	250	273
Pro	173	187	203	194	213 196	198	200	210	211	212	225	226	228	234	244	253	260	283
Val	175	189	205	215	198	217 200	202	212	213	214	227	228	230	236	246	255	262	285
Thr	177	191	207	217	219	202	221 204	214	215	216	229	230	232	238	248	257	264	287
Cys	179	193	209	219	221	223	206	225 216	217	218	231	232	234	240	250	259	266	289
Ile/Leu	189	203	219	229	231	233	235	226	245 227	228	241	242	244	250	260	269	276	299
Asn	190	204	220	230	232	234	236	246	228	247 229	242	243	245	251	261	270	277	300
Asp	191	205	221	231	233	235	237	247	248	231	249 244	245	247	253	263	272	279	302
Lys/Gln	204	218	234	244	246	248	250	260	261	262	256	275 257	259	265	275	284	291	314
Glu	205	219	235	245	247	249	251	261	262	263	276	258	277 260	266	276	285	292	315
Met	207	221	237	247	249	251	253	263	264	265	278	279	262	281 268	278	287	294	317
His	213	227	243	253	255	257	259	269	270	271	284	285	287	274	293 284	293	300	323
Phe	223	237	253	263	265	267	269	279	280	281	294	295	297	303	294	313 303	310	333
Arg	232	246	262	272	274	276	278	288	289	290	303	304	306	312	322	312	331 319	342
Tyr	239	253	269	279	281	283	285	295	296	297	310	311	313	319	329	338	326	345 349
Trp	262	276	292	302	304	306	308	318	319	320	333	334	336	342	352	361	368	372 391

1. Calculamos la masa mono isotópica del precursor mono cargado, $[M+H]^+$

Como se puede ver en el encabezamiento del espectro (Fig. 2), éste se corresponde con un precursor de m/z **788.35** donde z = 2. El "simple" hecho de observar fragmentos con un valor m/z superior a 788.35 ya nos indica que este péptido no puede ser carga 1. Por otro lado no se observan fragmentos superiores al triple de este valor, lo cuál nos dice que no tenemos un espectro carga 3. El aspecto en forma de campana y simétrico es característico de un péptido doblemente cargado, dato que se puede confirmar observando el espectro "zoom-scan", si lo hubiera, o su perfil isotópico en el espectro "full-scan" en un equipo de resolución media.

$$[M+2H]^{2+} \times 2 - 1 = [M+H]^+ = \mathbf{1575{,}7}$$

Un hecho común en un péptido que se fragmenta bien, es no encontrar trazas del precursor (788.35) pero sí es frecuente observar el ion correspondiente a la pérdida de agua (-18/2 = 9) o de amonio (-17/2 = 8,5), en este caso **779,74**. Este punto divide el espectro en dos mitades:

Zona alta.- Todos los fragmentos están a la derecha del ion precursor y son monocargados.

Zona baja.- Todos los fragmentos están a la izquierda del ion precursor y son, mayoritariamente monocargados y, en menor proporción, dicargados.

2. Comenzamos por la zona alta del espectro (Fig. 3 A).

La zona a la derecha del precursor es la más "limpia" del espectro, por lo tanto, la más sencilla de interpretar. Además, solo existe la posibilidad de encontrarnos fragmentos con carga 1.

$[M+H]^+ = y''_n$ (siendo n el nº total de aminoácidos del péptido)
$[M+H]^+ -18 = b_n$

Si tenemos en cuenta que el péptido es tríptico, el aminoácido C-terminal solo puede ser una K o una R:

Buscamos b_n – 128 o b_n – 156:

$$1575{,}7 - 18 - 156 = \mathbf{1401{,}7} \quad Si\ R$$
$$1575{,}7 - 18 - 128 = 1429{,}7 \quad Si\ K$$

Para confirmar la R podríamos buscar el fragmento y''_1:

$$y''_1 = \sum MRi + 19 = 156 + 19 = 175$$

(en el caso de K sería 147). En este ejemplo no podemos verlo puesto que el valor más bajo en el rango de masa es de 200 m/z (Fig. 3 B), este es un problema de los analizadores de trampa iónica y la conocida "regla del tercio". En cualquier caso, señalar que el ion b_1 rara vez se observa.

Figura 3. Secuenciación *de novo* mediante un LTQ-Velos CID del péptido tríptico [I/L][I/L]GADTSVD[I/L]EETGR.
Figura 3A. Fragmentos de la serie b de la zona alta del espectro;

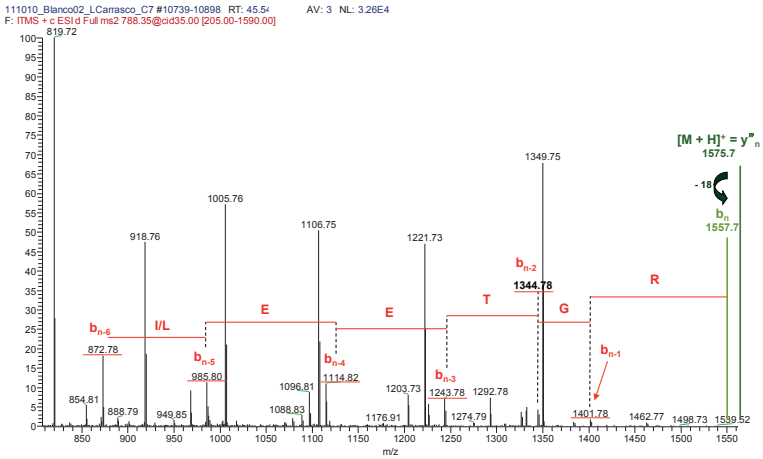
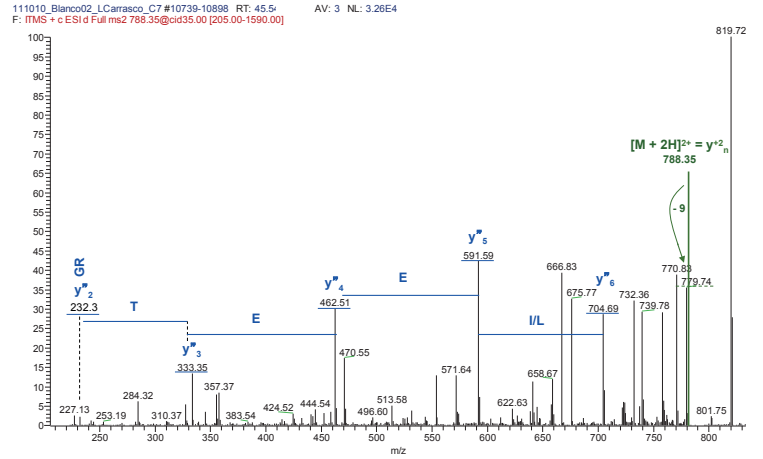


Figura 3B. Fragmentos de la serie y" de la zona baja del espectro.



Para continuar leyendo la serie b tenemos que ir buscando fragmentos hacia la izquierda de 1401.78 (b_{n-1}) hasta encontrar uno cuya diferencia de masa se corresponda con el valor de un aminoácido. De esta forma encontramos **1344,78**: la diferencia de masa es de 57 Da por lo que el antepenúltimo aminoácido de este péptido es una G.

En este momento, en el que ya tenemos asignados dos fragmentos de la serie b, podemos buscar sus iones fragmentos complementarios de la serie y":

$$y'' = [M+H]^+ - b + 1$$
$$b = [M+H]^+ - y'' + 1$$
$$y''_2 = 1575,7 - 1344,7 + 1 = \mathbf{232}$$

Si tenemos en cuenta que:

$$y'' = \sum MRi + 19$$
$$\sum MRi = 232 - 19 = 213$$

Utilizando la Tabla de parejas de aminoácidos (Tabla 3) encontramos que este valor se corresponde con las siguientes combinaciones: (VN) y (GR)

Por lo que confirmamos, interpretando este ion y''_2, la secuencia de los 2 aminoácidos C-terminal.

Este es un ejercicio que iremos haciendo durante toda la interpretación: Confirmar la lectura de la secuencia a partir de una serie buscando los iones complementarios en la otra serie.

Los pares b/y" complementarios, no dan únicamente información acerca de la secuencia, también proporcionan información acerca de la masa molecular ya que la suma de sus masas es equivalente a la masa del precursor. A veces, es tan sencillo como coger una regla, medir la distancia en cm, con el espectro impreso en papel, entre el precursor ($[M+2H]^{2+}$) y un fragmento y", llevar esta distancia al otro lado del espectro y encontramos su b complementario (y viceversa).

Continuamos buscando diferencias de masa que correspondan a un aminoácido desde el último ión b identificado y encontramos una T (101) al fragmento **1243,78** que confirmamos en la zona baja con el y''_3 (**333,35**), después una E (129) al fragmento **1114,82** que confirmamos con el y''_4 (**462,51**), continuamos con otra E (129) al ion **985,8** que confirmamos encontrando y"5 (**591,59**), la secuencia sigue con una ambigüedad, I/L (113) al fragmento **872,78** que confirmamos con el y"6 (**704,69**). En este punto no encontramos ningún fragmento por debajo de 872,78 con un m/z superior al precursor, que tenga la masa de un aminoácido (el más inmediato es 819,72 pero la diferencia de masa es de 53 Da, por lo que es de suponer que este fragmento pertenece a la serie y").

Es importante que al ir buscando fragmentos, utilicemos los más próximos ya que algunas combinaciones de aminoácidos tienen idéntica masa a la de un aminoácido, por ejemplo (G A) = K/Q = 128,05

Lo más sencillo es continuar analizando esta zona alta del espectro y asignar el resto de masas, presuntamente correspondientes a la serie y". Los péptidos trípticos tienden a ser más básicos en el C-terminal por lo que suele ser más prominente la serie y" en esta zona. Comenzamos restando el fragmento más abundante (**1349,75**) del precursor teórico (Fig. 3 C):

1575,7 – 1349,75 = 226

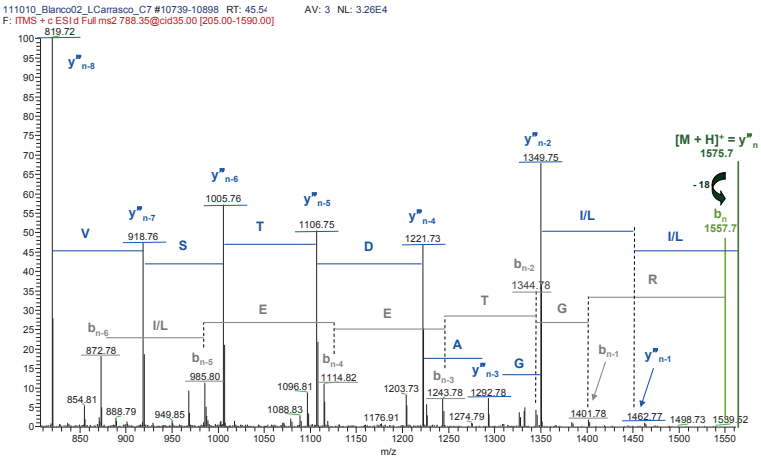
Como podemos ver en la tabla de pesos de aminoácidos (Tabla 1), este valor no se corresponde con ninguno de ellos, y tampoco a una combinación AA-R o AA-K, que podría esperarse si el Δm/z de 226 correspondiera a un ión de la serie y. Sin embargo, un Δm/z de 226 equivale teoricamente a los dipéptido N-terminales (EP, PE) o (I/L I/L), no protonados.

Probamos entonces a encontrar un fragmento que nos pueda decir cuál de estas dos parejas tenemos en el extremo N-terminal, y en qué orden, en el caso de que fuese (EP).

1575,7 – 129 (E) = 1446,7
1575,7 – 97 (P) = 1478,7
1575,7 – 113 (I/L) = **1462,7**

De estas tres masas solo una es visible en el espectro, 1462,77, por lo que tenemos los dos primeros aminoácidos del péptido: I/L I/L.

Fig. 3C. Fragmentos de la serie y" de la zona alta del espectro.



De la misma forma que calculamos anteriormente el ion $y"_2$, vamos a calcular el b_2 (Tabla 3):

$$y" = [M+H]^+ - b + 1$$
$$b = [M+H]^+ - y" + 1$$

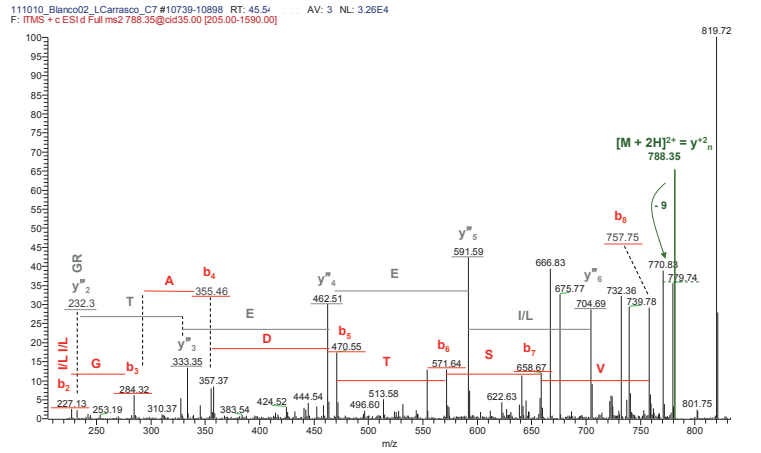
$b_2 = 1575,7 - 1349,75 + 1 =$ ***227***

Teniendo en cuenta que:

$$b = \sum MRi + 1$$

$\sum$ MRi = 227 – 1 = 226, que, como ya sabemos, se corresponde con la pareja I/L I/L, quedando confirmado así el extremo N-terminal (Fig. 3 D).

Fig. 3D. Fragmentos de la serie b de la zona baja del espectro.



En muchos espectros el ion b_2 se confirma ya que aparece junto a él, el ion a_2 (28 Da menor), es lo que se llama la pareja b_2/a_2, característica de la zona baja del espectro. En el ejemplo que estudiamos el rango m/z no nos deja ver este fragmento.

Al igual que hicimos con la serie b, continuamos buscando diferencias de masa que correspondan a un aminoácido desde el último y" identificado, así nos encontramos G al fragmento **1292.78**, que confirmamos en la zona baja del espectro con el b_3 (**284,32**). Seguimos leyendo la secuencia teniendo en cuenta los fragmentos más abundantes de esta región (no asignados a serie b) hasta llegar a **819,72** (leyendo: ADTSV) y confirmando en la zona baja del espectro con la serie b hasta llegar al b_8 (**757,75**). En este punto prácticamente, hemos leído todo el espectro. No necesitamos "interpretar la zona baja" porque ya lo hemos hecho al confirmar las secuencias buscando los iones fragmentos complementarios.

Ahora vamos a "conectar" para completar la secuencia:
* En la figura 3 A terminamos con el fragmento de la serie b, b_{n-6} = 872,78 por lo que podemos calcular cuál es la diferencia de masa con el ion más grande de esta serie que hemos asignado en la figura 3 D que es b_8 = 757,75

872,78 – 757,75 = 115,03 = **D**

* Si este cálculo lo realizamos con la serie y" tenemos que y"n-8 – y"6 (Figs. 3B y 3C):

819,72 – 704,69 = 115,03 = **D**

Por lo que confirmamos el aminoácido que nos falta para obtener la secuencia completa del péptido que queda como sigue:

N-t I/L I/L G A D T S V D I/L E E T G R C-t

Tabla 4. Predicción teórica de series de fragmentación de la secuencia del péptido [I/L][I/L]GADTSVD[I/L]EETGR.

Secuencia: **ILGADTSVDLEETGR**

CARGA +1

		Residuo	Total M+H	series A	series B	series B*	series Bo	series Y"	series Y*	series Yo
1	I	113,08	132,107	86,1	114,1	97,1	96,1	1575,8	1558,8	1557,8
2	L	113,08	245,191	199,2	**227,2**	210,1	209,2	**1462,7**	1445,7	1444,7
3	G	57,02	302,212	256,2	**284,2**	267,2	266,2	**1349,6**	1332,6	1331,6
4	A	71,04	373,249	327,2	**355,2**	338,2	337,2	**1292,6**	1275,6	1274,6
5	D	115,03	488,276	442,3	**470,3**	453,2	452,2	**1221,6**	1204,5	1203,5
6	T	101,05	589,324	543,3	**571,3**	554,3	553,3	**1106,5**	1089,5	1088,5
7	S	87,03	676,356	630,3	**658,3**	641,3	640,3	**1005,5**	988,5	987,5
8	V	99,07	775,424	729,4	**757,4**	740,4	739,4	**918,5**	901,4	900,4
9	D	115,03	890,451	844,4	**872,4**	855,4	854,4	**819,4**	802,4	801,4
10	L	113,08	1003,535	957,5	**985,5**	968,5	967,5	**704,4**	687,3	686,3
11	E	129,04	1132,577	1086,6	**1114,6**	1097,5	1096,5	**591,3**	574,2	573,3
12	E	129,04	1261,619	1215,6	**1243,6**	1226,6	1225,6	**462,2**	445,2	444,2
13	T	101,05	1362,667	1316,6	**1344,6**	1327,6	1326,6	**333,2**	316,2	315,2
14	G	57,02	1419,688	1373,7	**1401,7**	1384,6	1383,6	**232,1**	215,1	214,1
15	R	156,10	1575,789	1529,8	1557,8	1540,7	1539,8	175,1	158,1	157,1

CARGA +2

		Residuo	M+2H	series A	series B	series B*	series Bo	series Y"	series Y*	series Yo
1	I	56,54	66,55	43,5	57,5	49,0	48,5	788,4	779,9	779,4
2	L	56,54	123,10	100,1	114,1	105,6	105,1	731,9	723,3	722,8
3	G	28,51	151,61	128,6	142,6	134,1	133,6	675,3	666,8	666,3
4	A	35,52	187,12	164,1	178,1	169,6	169,1	646,8	638,3	637,8
5	D	57,51	244,64	221,6	235,6	227,1	226,6	611,3	602,8	602,3
6	T	50,52	295,16	272,2	286,2	277,6	277,1	553,8	545,3	544,8
7	S	43,52	338,68	315,7	329,7	321,2	320,7	503,2	494,7	494,2
8	V	49,53	388,21	365,2	379,2	370,7	370,2	459,7	451,2	450,7
9	D	57,51	445,73	422,7	436,7	428,2	427,7	410,2	401,7	401,2
10	L	56,54	502,27	479,3	493,3	484,7	484,2	352,7	344,2	343,7
11	E	64,52	566,79	543,8	557,8	549,3	548,8	296,1	287,6	287,1
12	E	64,52	631,31	608,3	622,3	613,8	613,3	231,6	223,1	222,6
13	T	50,52	681,83	658,8	672,8	664,3	663,8	167,1	158,6	158,1
14	G	28,51	710,34	687,3	701,3	692,8	692,3	116,6	108,1	107,6
15	R	78,05	788,39	765,4	779,4	770,9	770,4	88,1	79,5	79,1

Ahora comprobamos que la suma de las masas de los aminoácidos que forman la secuencia "tentativa" se corresponde con la masa del péptido. Podemos utilizar un programa que calcule este valor así como las series de fragmentación teóricas para esta secuencia, de esta forma confirmaremos nuestras asignaciones. Un programa libremente disponibles en internet es: Fragment Ion Calculator del Institute for Systems Biology (http://db.systemsbiology.net:8080/proteomicsToolkit/FraglonServlet.html). Como se puede observar en el espectro (Fig. 4A) quedan masas sin asignar, estos fragmentos (normalmente de menor intensidad) probablemente se correspondan con pérdidas de agua, de amonio o sean iones doblemente cargados. Con la tabla de predicción de series de fragmentación también podremos asignar estos iones (Tabla 4) o de forma manual calculando si son pérdidas de 18 o 17 Da de fragmentos de las series principales. Para el caso de los iones doblemente cargados podemos asignarlos a su correspondiente carga +1 con el siguiente cálculo sencillo:

$$[M+H]^{+}/2 + 0,5 = [M+2H]^{2+}$$

En la Fig. 4 B se muestra el espectro MS/MS interpretado y confirmado con la ayuda de la tabla 4.

Figura 4. Resultado de la interpretación *de novo* del espectro completo; Panel A, asignación de las series principales b e y"; Panel B, asignación de series secundarias e iones doblemente cargados, (o) pérdida de agua de la serie y", (o) pérdida de agua de la serie b, (2+) especie carga 2 para un fragmento b, (2+) especie carga 2 para un fragmento y".

Figura 4A

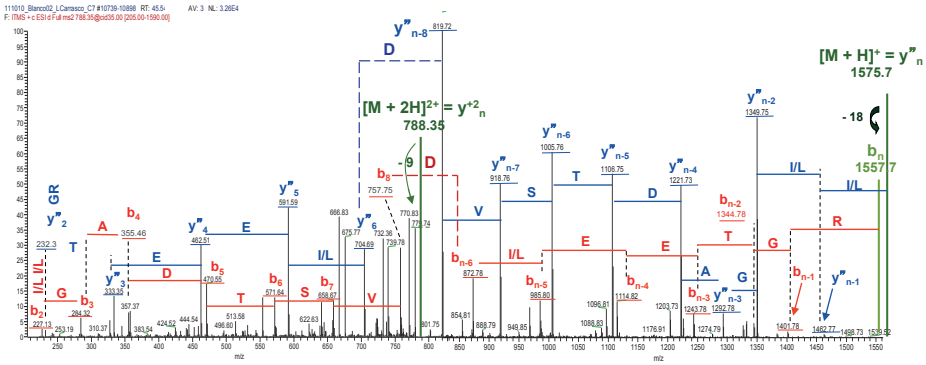
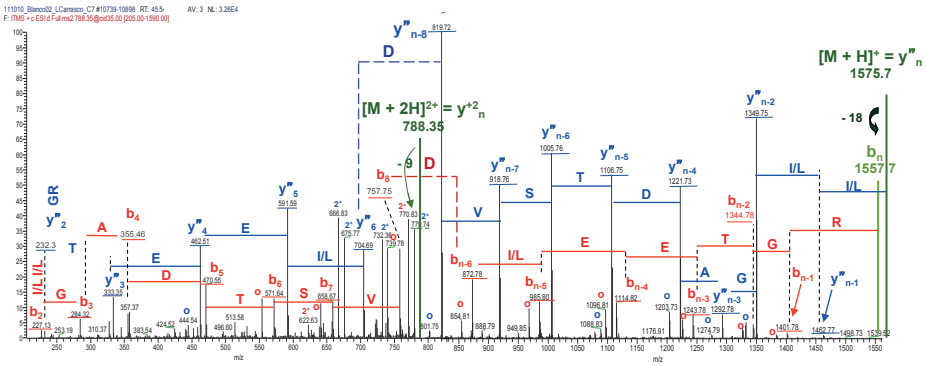


Figura 4B



3. Comentarios.

El espectro que hemos utilizado para aprender a interpretar una secuencia *de novo* no es el más común de los espectros. Normalmente nos encontramos con varias limitaciones, muchas de las cuales ya las hemos ido comentando a lo largo de este capítulo:

- Faltan fragmentos en una, otra o ambas series principales.

- Asignamos regiones de secuencia a partir de una serie que no podemos confirmar con la serie complementaria (ausente).

- "Tramos" de secuencia que se corresponden a parejas de aminoácidos, en algunos casos a varias parejas dependiendo de la diferencia de masa encontrada.

- No se suele observar b_1 (la ruptura del enlace peptídico b_1-b_2 es estéricamente poco probable) ni y''_{n-1} por lo que es difícil secuenciar el extremo N-terminal.

- En ocasiones un valor m/z es común para dos fragmentos de las dos series principales lo cuál dificulta la asignación. (p.ej. $y''_5 = b_6 = 580$)

- Leyendo una serie podemos encontrar una buena asignación a un fragmento de la serie complementaria, esto hace que "crucemos" la secuencia (saltamos de la y" a la b o viceversa).

- Como ya sabemos, hay secuencias que ven favorecida las series con pérdida de agua (-18) o de amonio (-17), en casos, estas series pueden ser más intensas que las principales por lo que dan lugar a confusión.

- Para añadir "más leña al fuego", se puede dar el caso de que dos péptidos con el mismo valor m/z coeluyan y nos encontremos estudiando el espectro de fragmentación "mezcla" de los dos péptidos.

- El ión fragmentado no corresponde a un péptido tríptico, es decir, no termina en K o R. En este caso, como se explicó en el apartado 2, el ión y_1 tendrá una masa = masa incremental del aminoácido (Tabla 1) + 18 + 1 (H^+) (por ejemplo, y_1 de W = 186 + 18 + 1 = 205, Fig.5). La masa de un ión y_1 no tríptico es generalmente distinguible de las masas de iones b. Sólo la masa de y_1 de I/L coincide con b_1 de M (m/z 132) (Tabla 5).

Tabla 5. Masas de los iones b_1 e y_1 de los aminoácidos protéicos más comunes listados en la Tabla 1.

AA	b1	y1
Gly (G)	58	75
Ala (A)	72	90
Ser (S)	88	106
Pro (P)	98	116
Val (V)	100	118
Thr (T)	102	120
Cys (C)	104	122
Ile (I)	114	**132**
Leu (L)	114	**132**
Asn (N)	115	133
Asp (D)	116	134
Gln (Q)	129	147
Lys (K)	129	147
Glu (E)	130	148
Met (M)	**132**	150
His (H)	138	156
Phe (F)	148	166
Arg (R)	157	175
Tyr (Y)	164	182
Trp (W)	187	205

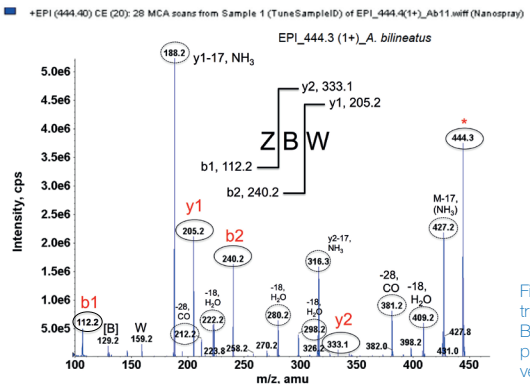


Figura 5. Espectro de fragmentación del tripéptido no-tríptico ZBW (Z= piroglutámico; B= K o Q), un inhibidor endógeno de metaloproteinasas dependientes de Zn^{2+} aislado del veneno de *Agkistrodon bilineatus* [50].

- Algunos portales web que ofrecen tutoriales sobre secuenciación *de novo* son:

  http://www.bioinfor.com/peaks/tutorials/denovo.html
  http://www.ionsource.com/tutorial/spectut/spec1.htm

9. CID-MS/MS de iones peptidicos que contienen enlace(s) disulfuro.

La fragmentación CID-MS/MS de péptidos que contienen enlaces disulfuro intra- o intercatenarios presenta características distintivas respecto a la de péptidos lineales [46]. En particular, la fragmentación inicial del enlace disulfuro (más lábil que el enlace peptídico) sigue un mecanismo asimétrico, por el que una cisteína retiene un sustituyente disulfohidrilo (-S-SH) mientras que la otra cisteína se transforma en deshidroalanina (Fig.6). La posterior fragmentación de la cadena principal genera iones b e y canónicos para fragmentaciones a ambos lados de las cisteínas, mientras que las fragmentaciones que involucran la disociación de enlaces peptídicos entre ambas cisteínas origina iones [b-34 e y+32] o [b-32 e y+34] dependiendo de la ruta de fragmentación asimétrica seguida (Fig.6). La fragmentación de iones peptídicos compuestos por más de una cadena polipeptídica enlazadas entre sí por enlaces disulfuro intercatenarios supone una complicación adicional: el espectro contendrá tantas series de iones b e y canónicas como cadenas polipeptídicas (α, β, γ...) contenga el ion fragmentado y series de iones b-34 e y+32 o b-32 e y+34 cuando la disociación genere fragmentos que incluyan una de las cisteínas (Fig.7).

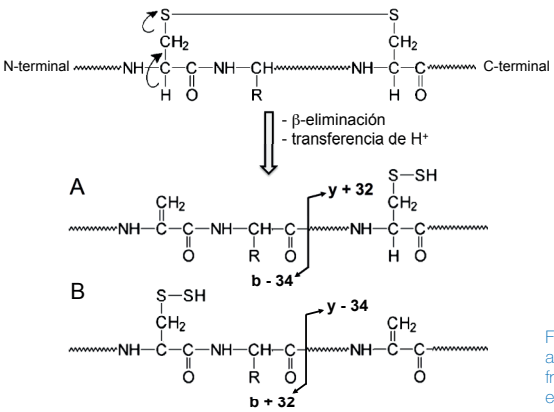


Figura 6. Mecanismo de la fragmentación asimétrica del enlace disulfuro y esquema de fragmentación de la cadena principal que genera iones b-34 e y+32 (A) o b+32 e y-34 (B).

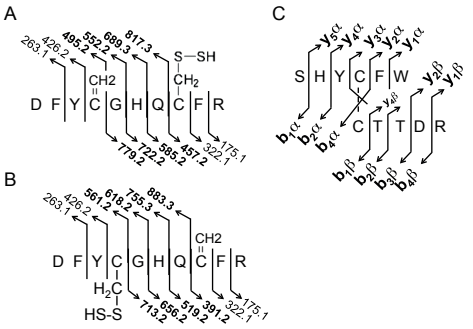


Figura 7. Paneles A y B, esquemas de la fragmentación alternativa de un péptido con enlace disulfuro intracatenario, generando series b-34 e y+32 (A) y b+32 e y-34 (B). Panel C, nomenclatura de los iones generados por CID-MS/MS de péptidos con enlace disulfuro intercatenario. Las series de iones específicos de secuencia (b e y) generadas en la fragmentación de diferentes cadenas polipeptídicas se diferencian mediante letras griegas, α y β.

10. Conclusiones.

A pesar del continuo crecimiento en la secuenciación de genomas completos, la secuenciación *de novo* sigue siendo necesaria. No solo quedan muchos genomas por secuenciar, muchas proteínas sin caracterizar, muchas modificaciones postraduccionales sin describir…, saber interpretar un espectro *de novo* es imprescindible para la persona que trabaja con espectros de fragmentación, aunque se correspondan con secuencias de proteínas presentes en bases de datos. ¿Cómo enfrentarse a un set de datos lleno de "probabilities", "scores", "FDRs", iones "matcheados"…sin saber a simple vista, mirando un espectro, si una asignación es buena o mala? Bien es cierto que estamos en la era de la identificación masiva, de la cuantificación masiva, donde Mann siempre va a la cabeza con el mayor nº de péptidos identificados en una sola carrera…pero nosotros, los que no somos Mann, seguimos trabajando, además, con muestras sencillas (en nº, que no en trabajo..), muestras que contienen una única proteína, la proteína del investigador que quiere caracterizar fosforilaciones, puentes disulfuro, mutaciones, interacciones, actividades enzimáticas a las que hay que poner nombre, etc., etc.

Continuando con la aplicación más evidente, la de identificar proteínas que no están en bases de datos, hay que señalar que no son muchos los trabajos publicados de secuenciación completa de proteínas mediante espectrometría de masas. Para este fin, el abordaje experimental pasa por utilizar varias proteasas (tripsina, quimiotripsina, V8…) para obtener péptidos que solapen en la secuencia, emplear distintos tipos de fragmentación (CID, ETD, ECD) y que la proteína sea pequeña. Hasta la fecha la proteína más grande que se ha secuenciado mediante MS es el citocromo c4, de 21 kDa de peso molecular [47]. Además de para identificar proteínas en base a homologías, la secuenciación *de novo* es muy útil para la obtención de secuencias de proteínas mediante aproximaciones de biología molecular. Por ejemplo, se pueden utilizar las secuencias obtenidas *de novo* para el diseño de "primers" y emplearlos para RACE-PCR (*rapid amplification of cDNA-ends with polymerase chain reaction*) [48, 49]. Finalmente la secuencia completa de la proteína se obtiene mediante secuenciación del DNA.

REFERENCIAS.

1. Edman, P., Method for the determination of the amino acid sequence in peptides. *Acta Chem. Scand.* 1950, *4*, 283–293.

2. Biemann, K., Laying the groundwork for proteomics – Mass spectrometry from 1958 to 1988. *Int. J. Mass Spectrom.* 2007, *259*, 1–7.

3. Biemann, K., Martin, S. A., Mass-spectrometric determination of the amino-acid-sequence of peptides and proteins. *Mass Spectrom. Rev.* 1987, *6*, 1–75.

4. Hunt, D. F., Bone, W. M., Shabanowitz, J., Rhodes, J., Ballard, J. M., Sequence-analysis of oligopeptides by secondary ion-collision activated dissociation mass-spectrometry. *Anal. Chem.* 1981, *53*, 1704–1706.

5. Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S., Hauer, C. R., Protein sequencing by tandem mass-spectrometry. *Proc. Natl. Acad. Sci. USA* 1986, *83*, 6233–6237.

6. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray ionization for mass-spectrometry of large biomolecules. *Science* 1989, *246*, 64–71.

7. Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198–207.

8. Cox, J., Mann, M., Is proteomics the new genomics? *Cell* 2007, *130*, 395–398.

9. Perkins, D. N., Pappin, D. J. C., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551–3567.

10. Yates, J. R., Eng, J. K., Clauser, K. R., Burlingame, A. L., Search of sequence databases with uninterpreted highenergy collision-induced dissociation spectra of peptides. *J. Am. Soc. Mass Spectrom.* 1996, *7*, 1089–1098.

11. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L. *et al.*, Open mass spectrometry search algorithm. *J. Proteome Res.* 2004, *3*, 958–964.

12. Wysocki, V. H., Tsaprailis, G., Smith, L. L., Breci, L. A., Special feature: commentary – mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* 2000, *35*, 1399–1406.

13. Roepstorff, P., Fohlman, J., Proposal for a common nomenclature for sequence ions in mass-spectra of peptides. *Biomed. Mass Spectrom.* 1984, *11*, 601–601.

14. Biemann, K., Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol.* 1990, *193*, 886–887.

15. Marshall, A. G., Hendrickson, C. L., High-resolution mass spectrometers. *Annu. Rev. Anal. Chem.* 2008, *1*, 579–599.

16. Zhang, Z. Q., Prediction of low-energy collision-induced dissociation spectra of peptides. *Anal. Chem.* 2004, *76*, 3908–3922.

17. Steen, H., Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* 2004, *5*, 699–711.

18. Cunningham, C., Glish, G. L., Burinsky, D. J., High amplitude short time excitation: A method to form and detect low mass product ions in a quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* 2006, *17*, 81–84.

19. Olsen, J. V., Macek, B., Lange, O., Makarov, A. *et al.*, Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 2007, *4*, 709–712.

20. Armirotti, A., Millo, E., Damonte, G., How to discriminate between leucine and isoleucine by low energy ESI-TRAP MSn. *J. Am. Soc. Mass Spectrom.* 2007, *18*, 57–63.

21. Hulst, A. G., Kientz, C. E., Differentiation between the isomeric amino acids leucine and isoleucine using lowenergy collision-induced dissociation tandem mass spectrometry. *J. Mass Spectrom.* 1996, *31*, 1188–1190.

22. Nakamura, T., Nagaki, H., Ohki, Y., Kinoshita, T., Differentiation of leucine and isoleucine residues in peptides by consecutive reaction mass-spectrometry. *Anal. Chem.* 1990, *62*, 311–313.

23. Shui, W. Q., Liu, Y. K., Fan, H. Z., Bao, H. M. *et al.*, Enhancing TOF/TOF-based de novo sequencing capability for high throughput protein identification with amino acidcoded mass tagging. *J. Proteome Res.* 2005, *4*, 83–90.

24. Samyn, B., Sergeant, K., Memmi, S., Debyser, G. *et al.*, MALDI-TOF/TOF de novo sequence analysis of 2-D PAGEseparated proteins from Halorhodospira halophila, a bacterium with unsequenced genome. *Electrophoresis* 2006, *27*, 2702–2711.

25. Kaufmann, R., Spengler, B., Lutzenkirchen, F., Mass-spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass-spectrometer using matrix-assisted laser-desorption. *Rapid Commun. Mass. Spectrom.* 1993, *7*, 902–910.

26. Horn, D. M., Zubarev, R. A., McLafferty, F. W., Automated *de novo* sequencing of proteins by tandem high-resolution mass spectrometry. *Proc. Natl. Acad. Sci. USA* 2000, *97*, 10313–10317.

27. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc.Natl. Acad. Sci. USA* 2004, *101*, 9528–9533.

28. Chi, A., Huttenhower, C., Geer, L. Y., Coon, J. J. et al., Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl. Acad. Sci. USA* 2007, *104*, 2193–2198.

29. Shevchenko, A., Chernushevich, I., Ens,W., Standing, K. G. *et al.*, Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass. Spectrom.* 1997, *11*, 1015–1024.

30. Mo, W. J., Takao, T., Shimonishi, Y., Accurate peptide sequencing by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass. Spectrom.* 1997, *11*, 1829–1834.

31. Uttenweiler-Joseph, S., Neubauer, G., Christoforidis, A., Zerial, M., Wilm, M., Automated de novo sequencing of proteins using the differential scanning technique. *Proteomics* 2001, *1*, 668–682.

32. Bantscheff, M., Dumpelfeld, B., Kuster, B., Femtomol sensitivity post-digest O-18 labeling for relative quantification of differential protein complex composition. *Rapid Commun. Mass. Spectrom.* 2004, *18*, 869–876.

33. D. W. Roth, K., Huang, Z., Sadagopan and Watson, J.T., Charge derivatization of peptides for analysis by mass spectrometry. *Mass Spectrometry Reviews* 1998, *17*, 255-274.

34. Fernandez-de-Cossio, J., Gonzalez, J., Satomi, Y., Shima, T. *et al.*, Automated interpretation of low-energy collisioninduced dissociation spectra by SeqMS, a software aid for de novo sequencing by tandem mass spectrometry. *Electrophoresis* 2000, *21*, 1694–1699.

35. Liska, A. J., Shevchenko, A., Combining mass spectrometry with database interrogation strategies in proteomics. *Trac-Trends Anal. Chem.* 2003, *22*, 291–298.

36. Mo, L. J., Dutta, D., Wan, Y. H., Chen, T., MSNovo: A dynamic programming algorithm for de novo peptide sequencing via tandem mass spectrometry. *Anal. Chem.* 2007, *79*, 4870–4878.

37. Pevtsov, S., Fedulova, I., Mirzaei, H., Buck, C., Zhang, X., Performance evaluation of existing de novo sequencing algorithms. *J. Proteome Res.* 2006, *5*, 3018–3028.

38. Pitzer, E., Masselot, A., Colinge, J., Assessing peptide de novo sequencing algorithms performance on large and diverse data sets. *Proteomics* 2007, *7*, 3051–3054.

39. Savitski, M. M., Nielsen, M. L., Kjeldsen, F., Zubarev, R. A., Proteomics-grade de novo sequencing approach. *J. Proteome Res.* 2005, *4*, 2348–2354.

40. Taylor, J. A., Johnson, R. S., Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry. *Anal. Chem.* 2001, *73*, 2594–2604.

41. Shevchenko, A., Sunyaev, S., Loboda, A., Shevehenko, A. *et al.*, Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time of flight mass spectrometry and BLAST homology searching. *Anal. Chem.* 2001, *73*, 1917–1926.

42. Junqueira, M., Spirin, V., Balbuena, T. S., Thomas, H. *et al.*, Protein identification pipeline for the homology-driven proteomics. *J. Proteomics* 2008, *71*, 346–356.

43. Pearson, W. R., Lipman, D. J., Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 1988, *85*, 2444–2448.

44. Kapp, E. A., Schütz, F., Reid, G. E. et al., Mining a Tandem mass Spectrometry Database To Determine the Trends and Global Factors Influencing Peptide Fragmentation. *Anal. Chem.* 2003, *75*, 6251-6264.

45. Schlosser, A., Lehmann, W. D., Patchwork peptide sequencing: extraction of sequence information from accurate mass data of peptide tandem mass spectra recorded at high resolution. *Proteomics* 2002, *2*, 524–533.

46. Mormann, M., Eble, J., Schwöppe, C., Mesters, R.M., Berdel, W.E., Peter-Katalini , J., Pohlentz, G. Fragmentation of intra-peptide and inter-peptide disulfide bonds of proteolytic peptides by nanoESI collision-induced dissociation. *Anal. Bioanal. Chem.*, 2008, 392, 831-838.

47. Branca, R. M. M., Bodo, G., Bagyinka, C., Prokai, L., de novo sequencing of a 21-kDa cytochrome c(4) from Thiocapsa roseopersicina by nanoelectrospray ionization iontrap and Fourier-transform icon-cyclotron resonance mass spectrometry. *J. Mass Spectrom.* 2007, *42*, 1569–1582.

48. Thompson, A. H., Bjourson, A. J., Orr, D. F., Shaw, C., McClean, S., A combined mass spectrometric and cDNA sequencing approach to the isolation and characterization of novel antimicrobial peptides from the skin secretions of *Phyllomedusa hypochondrialis* azurea. *Peptides* 2007, *28*, 1331–1343.

49. Voss, C., Eyol, E., Frank, M., von der Lieth, C. W., Berger, M. R., Identification and characterization of riproximin, a new type II ribosome-inactivating protein with antineoplastic activity from *Ximenia americana. FASEB J.* 2006, *20*, E334–E345.

50. Lomonte, B., Tsai, W.C., Ureña-Diaz, J.M., Sanz, L., Mora-Obando, D., Sánchez, E.E., Fry, B.G., Gutiérrez, J.M., Gibbs, H.L., Sovic, M.G., Calvete, J.J. Venomics of New World pit vipers: Genus-wide comparisons of venom proteomes across Agkistrodon. *J. Proteomics* 2014, 96, 103-116.

Identificación de proteínas mediante huella peptídica y datos de $(MS)^n$

19

Mª Luisa Hernáez[1], Mª Dolores Gutiérrez[1], J. Alberto Medina-Auñon[2], Gloria Alvarez-Llamas[3]

[1] Unidad de Proteómica. Universidad Complutense de Madrid- Fundación Parque Científico de Madrid, Madrid (Spain);

[2] Unidad de Proteómica, Centro Nacional de Biotecnología - CSIC, Madrid (Spain);

[3] Departamento de Inmunología, IIS-Fundación Jiménez Díaz, Madrid (Spain).

1. Identificación de proteínas por espectrometría de masas. Consideraciones generales.

El tema de este capítulo es la identificación y caracterización de proteínas por espectrometría de masas. Se describirá tanto la técnica más sencilla como la huella dactilar peptídica, o simplemente huella peptídica, hasta la identificación masiva de proteínas o de alto rendimiento. Se hará hincapié en los aspectos prácticos como la búsqueda en las bases de datos de secuencia, la elección de la base de datos de secuencia o las estrategias para identificar los picos del espectro de masas no asignados. Se describirán las consideraciones necesarias y los criterios para estimar la fiabilidad de un resultado.

1.1. Antecedentes.

Aunque algunas proteínas pueden tener un alto grado de homología con otras proteínas, consistente en similitud de secuencia, siempre hay alguna o algunas secuencias que son únicas. Si una proteína se puede cortar de una manera predecible, las piezas resultantes o péptidos (con un tamaño y, por tanto, masa específica) formarían una huella única e inconfundible que sería como la huella dactilar para esa proteína. Además, si cada entrada en la base de datos de secuencias de proteínas se pudiera cortar de la misma manera *in silico*, la huella serviría para identificar la proteína. Esta hipótesis será la base para los primeros experimentos de lo que llegaría a ser comúnmente conocido como huella peptídica de masa [1]. Sin embargo, a partir de tres desarrollos adicionales: métodos de ionización molecular más sensibles, incremento del poder de los ordenadores y secuenciación masiva y ampliación de las bases de datos, en 1993 se publicaron por primera vez diferentes trabajos describiendo cómo el mapeo de masa se podía usar para identificar proteínas separadas en gel [2,3,4,5,6].

En 1995 la espectrometría de masas se había convertido en una herramienta integral para determinar la identidad de las proteínas. La degradación automatizada de Edman continuó siendo la principal metodología utilizada para la determinación de la secuencia de las proteínas no anotadas en una base de datos.

M. Mann y M. Wilm [7] presentaron un método basado en la secuenciación de péptidos a partir de espectros de fragmentación o MS/MS, que era un híbrido de la interpretación manual de los espectros de masas o secuenciación de novo y la búsqueda en bases de datos, que denominaron “etiqueta de secuencia de péptido”. Consiste en identificar una proteína en la

base de datos de secuencias aminoacídicas, a partir de la masa del péptido intacto, los datos de una secuencia parcial (etiqueta de secuencia, "sequence tag") extraída de la interpretación manual del espectro de fragmentación del péptido (MS/MS), y la posición de esta etiqueta de secuencia dentro del péptido.

Al mismo tiempo, John Yates *et al.* [8] utilizaron un enfoque muy diferente para identificar la proteína, basado en la correlación entre los espectros de fragmentación experimentales y los espectros de fragmentación teóricos de todos los péptidos trípticos anotados que tengan la misma masa que cada uno de los péptidos precursores experimentales. Su programa, llamado SEQUEST, permitió la identificación de proteínas de una manera completamente automatizada a partir de un conjunto de espectros de masas en tándem (MS/MS). Los enfoques tanto de M. Mann como de J. Yates, dirigidos a la identificación de una proteína en función de los datos de secuencia de un solo péptido, fue la base para la identificación masiva de proteínas años más tarde. Además, la automatización de los distintos pasos de este proceso permitió abordar experimentos de alto rendimiento (shotgun). Es decir, se pueden identificar mezclas muy complejas de proteínas, proteomas o subproteomas, de un modo eficaz y fiable. Una muestra puede generar decenas o incluso cientos de miles de espectros MS/MS, que se pueden buscar en las bases de datos como un solo conjunto de datos.

Este método de identificación proteica también ofrece la posibilidad de conseguir identificaciones útiles a partir de espectros de relativa baja calidad que no darían resultados fiables por sí solos, ni siquiera al ser analizados por el método llamado etiqueta de secuencia. Sin embargo, si existe concordancia para espectros de varios péptidos de la misma proteína, esto puede hacer que estos espectros sean utilizados para obtener un mayor grado de confianza en la correcta identificación de la proteína y que esté presente realmente en la muestra analizada.

1.2. Fundamento de la Identificación de proteínas por huella péptidica de masa ("Peptide Mass Fingerprinting", PMF).

La masa es un atributo molecular característico que puede ayudar a identificar o confirmar la identidad de una molécula. La medida del peso molecular por espectrometría de masas está basada en la producción, separación y detección de moléculas iónicas.

Los espectrómetros de masas permiten medir la masa de las moléculas por la medida de la relación masa/carga (m/z) de los iones analizados con una elevada exactitud (diferencia entre el valor medido y el valor real o teórico), resolución (capacidad para separar dos masas próximas y se define como R= M/ΔM, donde M es la masa de un pico y el ΔM es la diferencia de masa entre dos picos de intensidad parecida adyacentes perfectamente separados o que solapan menos de un 10%) y sensibilidad (cantidad mínima detectada). Si se realiza la digestión con una proteasa específica de una cantidad realmente baja (femtomoles) de una proteína de interés,

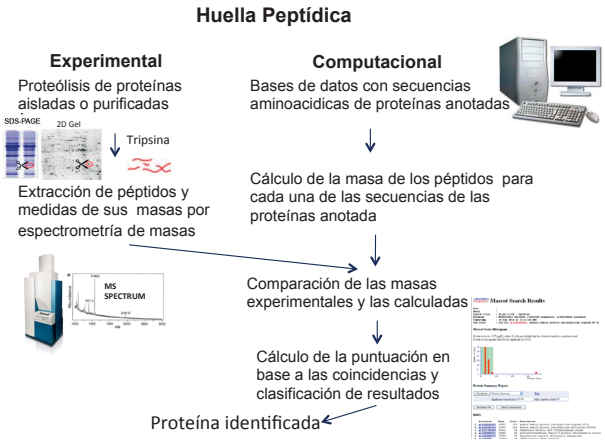


Figura 1. Esquema del proceso de identificación por huella peptídica. Se describen los pasos experimentales y computacionales para obtener y comparar las masas de los péptidos obtenidos después de digerir una proteína separada en gel SDS-PAGE o 2DE.

previamente aislada o purificada, y se mide la masa de los péptidos que se generan en un espectrómetro de masas con la máxima exactitud y resolución posibles, se puede determinar si en las bases de datos de secuencias de proteínas existe alguna proteína que al ser digerida teóricamente con la misma proteasa produzca un conjunto de péptidos con las mismas masas que las medidas experimentalmente.

Esta búsqueda en bases de datos se hace a través de diferentes programas o motores de búsqueda que incorporan algoritmos capaces de hacer la comparación entre las masas teóricas y las experimentales de los péptidos resultantes de digerir las proteínas, asignándole una puntuación a cada proteína en función de las coincidencias y ordenando los resultados de acuerdo a esta puntuación.

Si el número de coincidencias de alguna de las proteínas anotadas es suficiente para que no se deban al azar, podemos considerar la proteína identificada. Un esquema general del proceso se muestra en la Figura 1.

DIGESTIÓN EN GEL.

La identificación proteica por huella peptídica exige la separación o purificación previa de la/las proteína/s de interés. En general, la electroforesis en gel, mono (SDS-PAGE) o bidimensional (2-DE), suele ser la técnica de elección para dicha etapa previa por su gran capacidad de resolución. Posteriormente, es necesario digerir la proteína *in situ*, en el gel, y extraer los péptidos de éste para ser analizados por espectrometría de masas.

La eficacia de la reacción de digestión de la proteína que vamos a analizar resulta tan importante como la elección del espectrómetro de masas adecuado para analizar la muestra; cuanto mayor sea el rendimiento de la digestión, mayor información obtendremos de los péptidos presentes en la misma.

Aspectos básicos y protocolo.

El método de digestión en gel fue establecido por Rosenfeld *et al.* [9]. Los pasos fundamentales se muestran en la Figura 2. Posteriormente se han ido añadiendo numerosas modificaciones dirigidas a aumentar el rendimiento [10], la rapidez, como en el caso de la digestión asistida por microondas que reduce a minutos el tiempo de incubación con la enzima [11] y la reducción en el número de pasos de manipulación de la muestra para minimizar la contaminación o pérdida de rendimiento.

Es importante resaltar que la digestión de la proteína/s a analizar es un paso crítico que condiciona enormemente la cantidad y la calidad de la información que vamos a obtener de los péptidos presentes en la muestra y, por tanto, del éxito de la identificación de proteínas. Su rendimiento es dependiente de la cantidad de proteína de la que se parte, ya que en cada uno de los pasos que se llevan a cabo durante el proceso puede producir perdida de muestra, tanto por adsorción en tubos y puntas de micropipeta como durante el secado a vacío de la muestra o la extracción del gel; esta pérdida puede ser de hasta un 50% [12]. Optimizar y reducir en lo posible los pasos del proceso minimizan la perdida de péptidos. Además, hay que evitar posibles contaminaciones. Esto es tanto más

1. Cortar las bandas o spots de gel y lavar
↓
2. Desteñir y deshidratar
↓
3. Reducir y Alquilar Cisteínas
↓
4. Lavar y cambiar tampón para la digestión
↓
5. Digestión con tripsina (u otra proteasa)
↓
6. Extracción de péptidos
↓
7. Concentración y Desalado de péptidos

Figura 2. Esquema del protocolo de digestión en gel. Incluye los pasos fundamentales del proceso de digestión de una o varias proteínas aisladas en un gel, ya sea mono o bidimensional.

importante cuanto menor sea la cantidad de proteína a analizar, como es el caso de aquellas proteínas previamente separadas por 2-DE y que únicamente se han podido visualizar mediante tinción con plata, o bien proceden de un experimento DIGE (Differential In Gel Electrophoresis). En ambos casos la cantidad presente de proteína es inferior a la observada en tinción por azul de Coomassie, por lo que una identificación satisfactoria podría estar limitada por una digestión ineficaz o una baja sensibilidad en la detección. La tinción con plata, presenta además la dificultad de que interfiere en el proceso de digestión con tripsina, en parte por bloquear algunos de los puntos de corte de la enzima al establecer puentes de metileno en las cadenas laterales de las lisinas, y en parte por dificultar la extracción de péptidos del gel. Estos efectos se incrementan, además, con el tiempo de conservación del gel en el agua previa a la digestión [13], y por tanto es altamente recomendable comenzar la digestión inmediatamente después del proceso de tinción para evitar disminuir la producción de péptidos.

Aunque hay protocolos que demuestran que la tinción con plata de proteínas en gel es compatible con el posterior análisis por espectrometría de masas e identificación de proteínas [14], en la práctica los resultados obtenidos en la identificación de proteínas procedentes de geles teñidos con plata son muy variables, dado su corto rango de linealidad en concentración y las variaciones en el tiempo de revelado, el tiempo de conservación del gel, el uso o no de formaldehido, etc.

Siempre y cuando sea posible contar con cantidad de proteína suficiente, es altamente recomendable usar la tinción de azul de Coomassie en lugar de la tinción con plata, aunque la señal sea mucho más débil. En la Tabla 1 se muestran algunas de las tinciones de proteínas separadas en gel más empleadas, junto con sus límites de detección.

Tabla 1. Tinciones comúnmente empleadas en la visualización de proteínas previamente separadas por electroforesis en gel.

Tinción	t (h)	Límite de detección (ng proteína / banda)	Bibliografía
Coomassie coloidal (G-250) Coomassie R 250	15 12–48	1–16 30–100	Neuhoff et al. 1988, Berggren et al. 2000, Mackintosh et al. 2003, Candiano et al. 2004 Patton 2002, Mackintosh et al. 2003
Plata SyproRuby	0,25 – 3 5,5–12	0,5 0,5–5	Heukeshoven and Dernick 1985, Merril et al. 1986 Berggren et al. 2000, Nishihara and Champion 2002, Lilley and Friedman 2004
Deep Purple	3,5	0,1	Mackintosh et al. 2003
DIGE	0,75	0,025	Marouga et al. 2005

A continuación se muestra un ejemplo de protocolo de digestión con tripsina.

1. Escisión de las manchas proteicas del gel.
Cortar las bandas o manchas, ajustándose al perfil, con una hoja de bisturí o con una punta de micropipeta recortada y depositar cada una en un pocillo de una placa de 96 pocillos o en tubos eppendorf.

2. Lavado.
Realizar 2 lavados con agua Milli-Q durante 10 min.

3. Desteñido (sólo para geles teñidos con solución de plata).
Cubrir las bandas con la solución de desteñido e incubar 10 minutos. Solución de desteñido: 30 mM de ferrocianuro potásico (9,9 mg/ml) y 100 mM de tiosulfato sódico (24,8 mg/ml). Preparar en fresco las soluciones y mezclar 1:1.

4. Lavado.
Con agua Milli-Q durante 5 minutos para retirar la solución de desteñido.

5. Deshidratación.
Añadir acetonitrilo (AcN) hasta que cubra la banda o mancha (20 µl) y esperar 5 minutos para que el gel se deshidrate y libere el colorante. Repetir esta operación dos veces para que el gel quede más blanco.

6. Secar a vacío en centrífuga "speed-vac".

7. Reducción y alquilación.
Cubrir las bandas con una solución 10 mM Ditiotreitol (DTT) en bicarbonato amónico 25 mM (NH_4HCO_3), 30 min a 56°C. Dejar a temperatura ambiente, eliminar el DTT y añadir AcN que se retira rápidamente para incubar con iodoacetamida 55mM (10 mg/ml) en 25 mM NH_4HCO_3 durante 15 min en oscuridad.

8. Lavado.
Incubar 5 minutos con AcN y retirarlo. A continuación añadir NH_4HCO_3 25mM e incubar 5 minutos tras los cuales, sin retirar el sobrenadante, se añade igual volumen de AcN y se incuba 15 minutos. Retirar el sobrenadante.

9. Secar a vacío en centrífuga "speed-vac".
Se eliminan así los restos de solvente obteniéndose un residuo proteico seco.

10. Digestión.
Añadir tripsina en una concentración de 12,5 ng/µl en 25 mM de NH_4HCO_3 y en un volumen que cubra las bandas que ahora están deshidratadas. Incubar en hielo 45 minutos para que se absorba pero sin que llegue a actuar.
Retirar el sobrenadante remanente si lo hubiera y añadir NH_4HCO_3 25 mM hasta cubrir las bandas. Incubar toda la noche a 37°C (o al menos 6 u 8 horas). Tras la digestión agitar con vortex y centrífugar para pasar el sobrenadante a una nueva placa o tubos.

11. Extracción de péptidos.
Cubrir las bandas con 50% AcN /0,1% ácido trifluoroacético (TFA) y sonicar 10 minutos. Pasar el sobrenadante a otro tubo o placa.
Repetir 3 veces esta operación recogiendo todos los sobrenadantes de cada extracción juntos.

12. Concentración de péptidos.
Secar en centrífuga "speed-vac" y resuspender en 5 µl de 0,1% TFA.

13. Análisis por MALDI-TOF.

Consideraciones prácticas

A continuación se describen algunas consideraciones prácticas importantes en cuanto a los pasos más relevantes del protocolo de digestión en gel [15], por sus posibles efectos en el posterior análisis por espectrometría de masas.

El paso de **desteñido** por tratamiento con los agentes reductores tiosulfato sódico y ferrocianuro potásico sólo es necesario y muy importante en los geles teñidos con plata. Está descrito que la eliminación de los iones de plata mejora tanto la eficacia de la digestión como el análisis de masas posterior [16].

En los pasos de **reducción y alquilación**, se lleva a cabo la reducción de los puentes disulfuro entre cisteínas con ditiotreitol (DTT) y la alquilación posterior de los grupos -SH con iodocetamida para bloquear la posible formación de nuevo de estos puentes disulfuro. La cisteína se transforma así en carbamidometilcisteína aumentando en 57 Da su residuo de masa. En este paso la proteína se despliega, haciendo más accesible todos los posibles puntos de corte a la proteasa, para obtener un mayor número de péptidos y por tanto un mayor rendimiento de la digestión proteica. Previo a la adición de la tripsina es necesario eliminar los reactivos anteriores de la pieza de gel, deshidratar con acetonitrilo, y equilibrar con un tampón adecuado para la proteólisis. Para favorecer el acceso de la enzima a la proteína en la red de acrilamida del gel, se seca a vacío antes de añadir la enzima o agente químico elegido.

Es importante que todo el proceso de **digestión** se realice con material estéril (tubos y puntas de pipeta, etc.) no autoclavado y reservado para su uso en proteómica. Deben emplearse reactivos con grado de pureza HPLC, guantes, bata, gorro, etc., para minimizar la contaminación ambiental y la inclusión de proteínas muy abundantes que no son de interés, como las queratinas del pelo, piel, lana, etc., ya que pueden reducir la detectabilidad de las proteínas sujeto del análisis, sobre todo cuando se trata de proteínas de baja abundancia. Esta contaminación se minimizará igualmente al reducir la manipulación previa de la muestra y el número de etapas del protocolo empleado. Existen digestores automáticos, que dispensan y absorben líquidos de forma automatizada, y pueden realizar todo el proceso de digestión en un ambiente limpio protegido con una mampara lo que disminuye las posibles contaminaciones y el tiempo del proceso, aunque como desventaja pueden perder la pieza de gel, si es pequeña, por succión.

También pueden aparecer contaminantes no peptídicos derivados de los reactivos utilizados, como detergentes, matrices (ácido alfa-ciano-4-hidroxicinámico), o polímeros (polietilenglicol) que son ionizables tanto en MALDI, como en ESI aunque suelen ser distinguibles de los péptidos y por tanto no contribuyen a una identificación errónea, aunque sí que aumentan el nivel de ruido y hacen más difícil la detección de los péptidos de interés.

La **elección de la proteasa** o agente químico también es clave para la producción óptima de péptidos adecuados y su posterior análisis. Hay varios aspectos claves a tener en cuenta, como la especificidad de corte, el tamaño medio de los péptidos que se producen y la actividad de la enzima. La tripsina es la enzima más ampliamente utilizada en experimentos de proteómica, ya que su actividad endoproteasa es muy específica. Corta en el extremo carboxiterminal de lisina y arginina, excepto si se siguen de prolina, lo que facilita la comparación con las digestiones teóricas de las proteínas de las bases de datos. En proteínas de mamífero el contenido en arginina es del 6% y en lisina del 5%. Como dato estimativo, por cada 100 aminoácidos hay 11 sitios de corte de tripsina, lo que produce un tamaño medio de los péptidos trípticos de 9-10 aminoácidos. Los péptidos de este tamaño son fáciles de extraer del gel de poliacrilamida donde se encuentra separada la proteína y además presentan un tamaño adecuado para los intervalos de masa en los que los espectrómetros presentan mayor poder resolutivo y exactitud en la medida de la masa.

En el año 2003, Helzen *et al.* [17] realizaron una serie de digestiones teóricas en una base de datos que contenía 20.639 proteínas empleando diferentes agentes proteolíticos, como tripsina (K/R), LysC (K), Bromuro de Cianógeno (M), Hidroxilamina (N-G) y Ácido Diluído (D-P). Los resultados muestran la distribución de los fragmentos obtenidos y su longitud media, y se observa que en el caso de la tripsina produce mayor número de fragmentos y de longitud adecuada (8 aminoácidos de promedio).

Además la tripsina es relativamente barata y muy eficaz y es la enzima que se utiliza de forma rutinaria. En cambio, si conocemos la secuencia de la proteína que queremos identificar o confirmar y no posee lisinas o argininas en número suficiente y/o posición adecuada, podemos elegir otro agente proteolítico o enzima que sea más idóneo. En la Tabla 3 se muestran algunas de las enzimas más usadas en este tipo de aproximación proteómica.

Tabla 3. Proteasas utilizadas en experimentos de proteómica.

	Tipo	Especificidad	Rango de pH
Quimotripsina	Serin	Y, F, W	1.5–8.5
Tripsina	Serin	R, K	7.5–9.0
Glu C	Serin	D, E	7.5–8.5
Lys C	Serin	K	7.5–8.5
Elastasa	Serin	A, V, I, L, G	8.0–9.0
Subtilisina	Serin	Casi todos los AA	7.0–11.0
Proteinase K	Serin	AA hidrofóbicos	7.0
Thrombina	Serin	R	7.5
Faktor X	Serin	I-E-G-R	8.3
Arg C	Cistein	R	7.5–8.5
Papaina	Cistein	R, K, G, H, Y	7.0–9.0
Asp N	Metalo	D (N-terminal)	6.0–8.0
Termolisina	Metalo	AA hidrofóbicos	7.0–9.0
Pepsina	Acida	F, M, L, W	2.0–4.0

Se pueden utilizar otras enzimas o agentes proteolíticos, tales como las que se comentan a continuación. La endoproteasa Lys - C se puede emplear para conseguir una cobertura de secuencia mayor en comparación con la tripsina, debido a que se generan péptidos más largos al romper la secuencia proteica exclusivamente en el extremo C-terminal del aminoácido lisina. La endoproteasa Glu - C (V8) escinde proteínas en el extremo carboxiterminal de ácido glutámico o ácido aspártico dependiente de las condiciones de la reacción. Sin embargo, Glu - C no se usa comúnmente, muy probablemente debido a una mayor autolisis generando muchos péptidos autoproteolíticos. Asp - N y Arg - C son relativamente caras y sólo empleadas para propósitos especiales, cuando se necesite identificar la totalidad de cobertura de la secuencia proteica. El bromuro de cianógeno es un reactivo químico muy efectivo pero tóxico que digiere las proteínas por el extremo carboxiterminal de metioninas con la formación de homoserinlactona. Debido al bajo número de metioninas por proteína, suelen generarse grandes péptidos. Para algunas aplicaciones esto puede ser beneficioso, aunque no es favorable para el análisis por espectrometría de masas. Sin embargo, se puede realizar a continuación una digestión con tripsina, lo que se aconseja para superar este inconveniente [18].

En general, si resulta adecuado y necesario, pueden realizarse sucesivas digestiones con enzimas diferentes.

Una vez concluida la reacción de digestión enzimática, el último paso sería la **extracción de péptidos** del gel. La mayoría de los péptidos, entre el 70-80%, difunden al tampón de digestión durante la incubación con la enzima y pueden recogerse en el sobrenadante, al que llamaremos crudo. Posteriormente, de modo adicional y para aumentar el rendimiento en péptidos, se pueden realizar pasos sucesivos de extracción por incubación con diferentes proporciones de un disolvente orgánico, como acetonitrilo, que favorece la solubilidad de los péptidos más hidrofóbicos. Finalmente, se juntarán todas las fracciones recogidas y se concentrarán los péptidos secando a vacío, evitando llegar a la sequedad total. De este modo se elimina el acetonitrilo que puede interferir en un paso posterior, opcional, de desalado de la solución peptídica antes de su análisis por espectrometría de masas.

En cualquier caso, es importante tener en cuenta que aunque los diferentes pasos de extracción pueden aumentar la obtención de péptidos también suponen un incremento en la

manipulación de la muestra, su posible contaminación y pérdidas. Por tanto, es conveniente minimizar dichas extracciones posteriores, a no ser que sean imprescindibles bien por tener pocos péptidos o bien por estar interesados en la búsqueda de péptidos de naturaleza altamente hidrofóbica que no aparecen en el crudo. En muchos casos, se analiza previamente el crudo por espectrometría de masas y en función de la visualización del espectro MS y/o los resultados de identificación obtenidos se realiza uno o más pasos de extracción.

Por otro lado, en la combinación de los diferentes sobrenadantes pueden concentrarse sales procedentes del gel o de los reactivos utilizados durante el proceso de digestión. Las sales son una de las causas principales de una baja sensibilidad en la detección y relación señal/ruido en los espectros de masas debido a la supresión iónica, sobre todo al emplearse electrospray como fuente de ionización, ya que la ionización por MALDI es más tolerante a impurezas en la muestra. En cualquier caso la limpieza de la muestra, eliminando sales, y concentrando los péptidos puede mejorar la calidad de los espectros MS. Este desalado se puede realizar de diferentes formas dependiendo de los volúmenes, concentraciones y naturaleza de la muestra que vamos a procesar, así como del tipo de análisis y el equipo en el que se va a llevar a cabo. En general, para el análisis por MALDI-TOF de los péptidos procedentes de la digestión de proteínas separadas en gel teñido con azul de Coomassie, se puede obviar este paso de concentración y desalado permitiendo el análisis rutinario de gran número de muestras.

Cuando esta etapa de desalado y concentración es necesaria, generalmente se utilizan microcolumnas de fase reversa (RP), ya sean comerciales embebidas en puntas de micropipeta (por ejemplo, ZipTip (Millipore)) o bien empaquetadas manualmente con resina C18 o poros R2, por ejemplo [19]. Es necesario que los péptidos estén en una disolución acidificada con TFA y libre de disolventes orgánicos, como el acetonitrilo, para que se puedan establecer interacciones hidrofóbicas entre los péptidos y la matriz cromatográfica, que será previamente equilibrada en un paso inicial. Los péptidos serán así retenidos mientras que las sales se eliminarán mediante sucesivos lavados con TFA (0,1%). Los péptidos finalmente se eluyen en un pequeño volumen utilizando un disolvente orgánico (por ejemplo, acetonitrilo al 50%).

Aunque el desalado y concentración de los péptidos en estas microcolumnas es ventajoso para aumentar hasta en 10 veces la señal de los mismos en el espectrómetro de masas, este pretratamiento de la muestra también cuenta con algunas desventajas a tener en cuenta, como el tiempo de procesado y el coste, la pérdida de péptidos hidrofílicos (como los fosfopéptidos) y en general una pérdida de muestra por la adsorción de péptidos largos o muy hidrofóbicos que son fuertemente retenidos y no eluyen.

Una alternativa rápida y sencilla a este proceso de desalado, para los péptidos analizados en MALDI-TOF, es lavar la muestra con agua en la propia placa de MALDI una vez que hemos depositado la muestra y la matriz elegida (generalmente ácido αcyano-4-hidroxicinámico (CHCA)) y han co-cristalizado. Esto es posible gracias a que muchas de las matrices son poco solubles en agua en comparación con las sales y otros surfactantes. El proceso es sencillo, sobre el pocillo con los péptidos/matriz cristalizados se deposita 1 o 2 µl de agua MilliQ, se incuba unos segundos y se retira y se deja secar al aire. Hay que evitar tocar la superficie de la placa o la muestra con la punta de la micropipeta. Con este fin, se ha propuesto también el empleo de ácido fosfórico para mejorar la sensibilidad de la detección (de 3 a 5 veces) [20].

También se han desarrollado placas especiales de MALDI recubiertas de una fina superficie hidrofóbica como teflón, consiguiendo concentrar en una superficie más pequeña toda la muestra. Este modelo permite, además, un lavado en la placa más eficiente, incrementando la señal hasta 10 veces, de manera que han podido llegar a detectarse 100 amoles de péptido [21].

DIGESTIÓN "*IN SILICO*" Y PÉPTIDOS TEÓRICOS.

Existen herramientas informáticas, como MS DIGEST (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest) o PEPTIDE MASS (http://web.expasy.org/peptide_mass/) que permiten realizar la digestión teórica de la secuencia de aminoácidos de una proteína, con una proteasa específica para calcular tanto los péptidos que se producirán como su masa. Esta herramienta, muy sencilla de utilizar, nos permite inspeccionar qué enzima es la más adecuada para realizar la digestión de una proteína concreta o estimar si el rendimiento de la digestión obtenido es adecuado, ya sea para identificar la proteína o bien para calcular y buscar en el espectro MS péptidos con una modificación determinada. En la Figura 3 se muestra la interface web de MSDigest y un ejemplo de los resultados que ofrece.

La secuencia de aminoácidos de la proteína que queremos digerir teóricamente o "*in silico*" con una proteasa específica se puede incluir directamente en el espacio reservado para ello ("user protein sequence") o bien se puede seleccionar la base de datos pública donde está anotada (elegir del desplegable "Database") e incorporar el nº de acceso de acuerdo a las instrucciones que detalla el programa en el recuadro "list of entries". Para realizar la digestión "*in silico*", hay que detallar tanto la proteasa, como una serie de condiciones que están perfectamente descritas en los enlaces correspondientes. Se resaltan los más relevantes: el número de puntos de corte que puede perder la tripsina como máximo en un péptido y las modificaciones, fijas y variables, que se deben tener en cuenta para calcular la masa de los péptidos resultantes. Igualmente, se muestran las condiciones que deben cumplir los péptidos que se forman tras la digestión (recuadradas): el rango de masa, la longitud mínima, diferentes estados de carga, etc. Los resultados incluyen la secuencia de la proteína, el peso molecular y punto isoeléctrico teórico, la composición de aminoácidos, una lista de los valores de m/z, tanto monoisotópica como media, de los péptidos producidos, el lugar que ocupa el péptido en la secuencia de la proteína ("start- end"), el número de puntos de corte perdidos y la secuencia del péptido. Si presenta una modificación aparece el nombre de ésta.

Figura 3. Interface y resultados del Programa MS DIGEST.

ANÁLISIS DE LOS PEPTIDOS POR ESPECTROMETRÍA DE MASAS MALDI-TOF.

Los espectrómetros de masas tipo MALDI-TOF (*Matrix-Assisted Laser Desorption/Ionization - Time-Of-Flight)* son los más utilizados para identificar proteínas por huella peptídica por varias razones: a) son robustos y fáciles de manejar, b) son altamente sensibles, siendo capaces de detectar incluso atomoles de analito, c) cuentan con un elevado poder de resolución, pudiendo medir péptidos de hasta 4.000 Da con una resolución por encima de 20.000, así por ejemplo, un péptido de m/z 2000 Da puede ser separado de otro de m/z 2000,1 Da, y d) son capaces de medir con excelente exactitud la masa, por ejemplo, en un espectro MS internamente calibrado, el error al medir la relación m/z de un péptido de aproximadamente 1000 Da de masa es menor de 10 ppm.

En este tipo de ionización los péptidos toman o pierden un único protón, a diferencia de la fuente de ionización tipo Electrospray en la que los iones presentan múltiples estados de carga (+2,+3, etc.).

En un espectro de masas MALDI-TOF, se representan los valores de m/z de los iones que llegan al detector en el eje de abscisas y en el eje de ordenadas la intensidad de la señal o el porcentaje de intensidad con respecto al pico más intenso.

En la actualidad, estos espectrómetros de masas poseen mecanismos para evitar que iones de la misma relación m/z alcancen a distinto tiempo el detector. Antes del tubo de vuelo se aplica: a) la extracción retardada, que consiste en esperar un tiempo (tiempo de retardo) suficiente para que todos los iones se desorban antes de aplicar el voltaje de aceleración, y b) un gradiente de potencial que corrige la diferencia de velocidad de iones con la misma masa antes de entrar en el tubo vuelo. Al final del tubo de vuelo y antes de llegar al detector, el reflector corrige las posibles diferencias en la velocidad de iones con la misma masa que se hayan producido durante el tiempo de vuelo en el tubo, lo que asegura una gran exactitud de medida, alta resolución y elevada precisión.

En la mayoría de los casos, el software de adquisición de datos del espectrómetro ajusta estos parámetros de forma automática para cada tipo de adquisición (Lineal, MS-Reflector o MS/MS) y pueden optimizarse para la medida de iones en un rango de masa determinado. La intensidad del láser también puede atenuarse, en mayor o menor medida, para aumentar la detección de péptidos menos abundantes, pero sin sobrepasar un límite tal que se produzca un aumento de ruido y pérdida de resolución.

En la Figura 4 se muestra en detalle un pico de un espectro MS, correspondiente al ion peptídico de masa 1562.6960. Se observa la distribución de los picos del péptido en cuanto a sus isótopos de ^{13}C o patrón isotópico y el valor de la masa monoisotópica, que se define como la masa del péptido con los isótopos más ligeros y abundantes en la naturaleza, y gráficamente corresponde al primer pico. Se puede conocer el estado de carga de un ion peptídico por la diferencia de m/z entre de los distintos picos de su distribución isotópica, así, sí la diferencia del valor m/z entre los diferentes isótopos es de 1 Da la carga del péptido es +1, si es de 0,5 Da la carga del péptido es +2, si es de 0,3 Da es +3 y así sucesivamente.

El análisis del digerido en un MALDI-TOF de una proteína previamente separada por electroforesis en gel produce un espectro de masas, en el que aparecen las masas moleculares de todos los péptidos detectados en ese digerido, también llamado huella peptídica de esa proteína. En la Figura 5 se muestra la huella péptidica (espectro MS) de una proteína adquirida en un 4700 MALDI-TOF/TOF de ABSciex.

Preparación de muestra y depósito en placa. Selección de matriz.

Al igual que en el caso de la digestión, es fundamental optimizar, tanto la preparación previa de la muestra, como la adquisición de los espectros en el espectrómetro de masas, para obtener

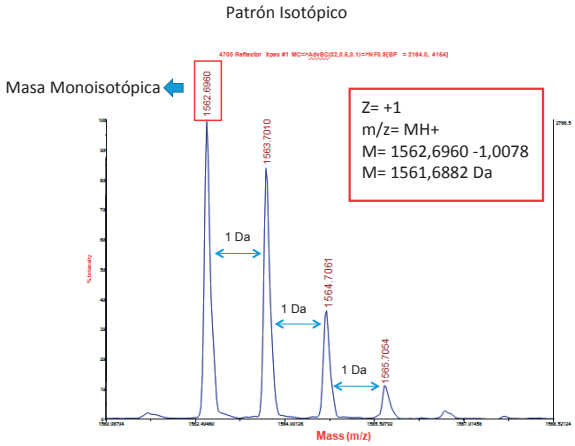


Figura 4. Patrón isotópico de un péptido adquirido en un espectrómetro de masas MALDI-TOF. Se muestra la distribución isotópica del ion peptídico de masa; m/z = 1562,6960, donde se aprecian los picos correspondientes a los diferentes isótopos del péptido, que se diferencian en 1 Da por la sustitución de un ^{12}C por un ^{13}C en la molécula iónica del péptido, dado que la carga del péptido (z) es +1. El primer pico corresponde a la masa monoisotópica del valor m/z o MH+ del péptido compuesto de los isotopos más ligeros y frecuentes, ^{1}H, ^{12}C, ^{14}N, ^{16}O. La masa del péptido M corresponde a la masa, monoisotópica observada menos la masa del protón.

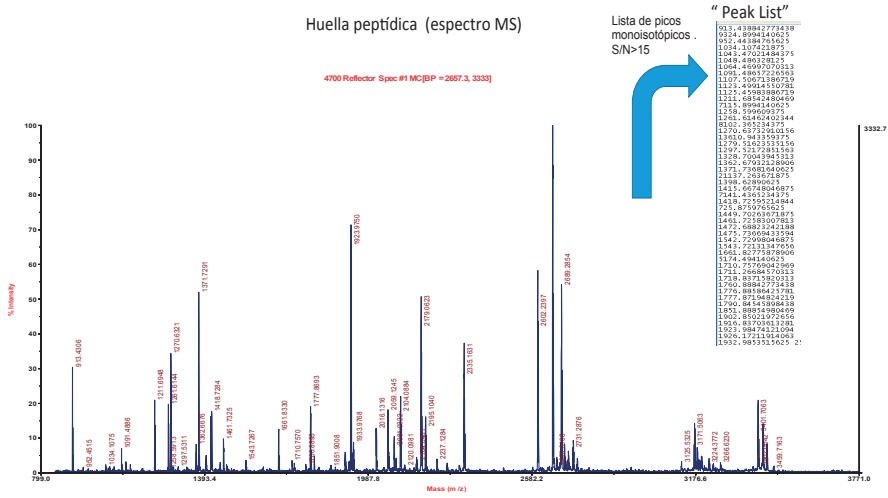


Figura 5

Figura 5. Huella peptídica de la Proteína acídica de cartílago de *Sparatus aurata*, adquirida en un MALDI-TOF. Se seleccionan los valores de m/z que superan una señal/ruido (S/N), en este caso mayor que 15, tal que se consideran los picos perfectamente distinguibles del ruido; éstos conformaran la lista de picos para identificar la proteína en las bases de datos.

espectros de buena calidad que contengan la máxima información posible de los péptidos presentes en la muestra. En la Figura 6 se resumen a modo de esquema las distintas etapas que incluyen la preparación de muestra previa al análisis por MALDI-TOF.

La optimización de la preparación de la muestra incluye la selección del compuesto matriz, su concentración y preparación adecuada. Dos de las matrices más usadas para el análisis de péptidos de peso molecular menor de 10.000 Da son el ácido alfa-ciano-4-hidroxicinámico (en una concentración de 3-10mg/ml en 50% ACN/TFA 0,1%) y el ácido. 2,5-Dihidroxibenzoico, aunque existen otras muchas matrices descritas para otras aplicaciones. En la Tabla 4 se muestra algunas de estas matrices y sus principales aplicaciones.

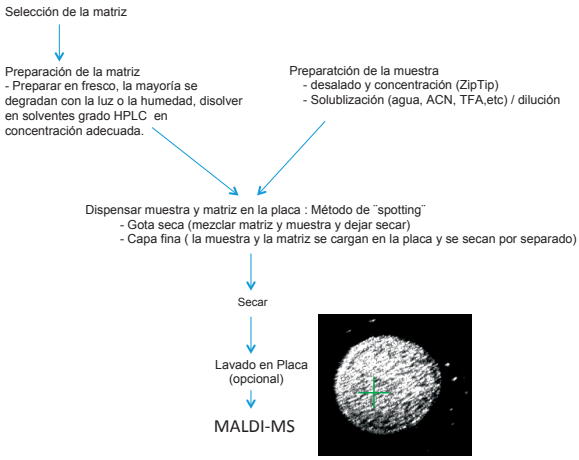


Figura 6. Esquema de la preparación de la muestra para análisis en el espectrómetro de masas MALDI-TOF. Aplicación de la muestra y la matriz en la placa de MALDI-TOF. En la imagen del pocillo de la placa de MALDI se aprecian los cristales de los péptidos y la matriz, en este caso alfa-ciano-4-hidroxicinámico, de forma redondeada y distribuidos homogéneamente por todo el pocillo.

Tabla 4. Selección de matrices empleadas en el análisis peptídico por MALDI-TOF y sus aplicaciones más relevantes.

Matriz	Aplicación
Ácido sinapínico (ácido 3,5-dimetoxi-4-hidroxicinámico)	Péptidos y proteínas > 10,000 Da
Ácido α-ciano-4-hidroxicinámico (CHCA)	Depósito de muestra "gota seca". Péptidos / proteínas <10,000 Da Depósito de muestra "capa fina". Péptidos <-3,000
2,4,6-trihidroxiacetofenona (THAP)	Pequeños oligonucleótidos <3,500 Da, carbohidratos ácidos, glicopéptidos ácidos y compuestos sensibles a ácido
Ácido 3-hidroxipicolínico (3-HPA) en citrato amónico	Oligonucleótidos grandes ~ >3,500 Da
Ácido 2,5-dihidroxibenzoico (2,5-DHB)	Péptidos, carbohidratos neutros o básicos, glicolípidos (iones negativos), polímeros sintéticos polares y no-polares, moléculas pequeñas
Ácido 2,5-dihidroxibenzoico y ácido 5-metoxisalicílico (DHBs)	Péptidos y proteínas > 10,000 Da, proteínas glicosiladas

Existen diferentes métodos para depositar la muestra y la matriz en la placa del MALDI-TOF ('Spotting'). Básicamente se utilizan los dos que se indican a continuación o bien variaciones de éstos:

- Gota seca ('Dried Droplet'), consiste en mezclar la muestra y la matriz en una relación 1:10 antes de depositarlas en la placa o bien, si hay poca cantidad de muestra, mezclarlas en el propio pocillo (1:1) y luego secar.

- Capa Fina ('Thin Layer'), la muestra y la matriz no se mezclan antes de antes de aplicarlos en la placa, primero se aplica la muestra y se deja secar y entonces se aplica la matriz y se deja secar. Esta técnica se utiliza para aumentar la sensibilidad, cuando la concentración de la muestra es <0,1 picom/µl y los péptidos <3.000 Da.

Una vez seca la mezcla, se puede inspeccionar si los cristales de matriz están bien formados y distribuidos. Si no es así, es una indicación para realizar un lavado en placa de la muestra/matriz o bien hacer un desalado de la muestra (ZipTip®) antes de depositarla de nuevo, tal y como se comentó anteriormente.

No existe un método ideal establecido de preparación de la muestra para la identificación de proteínas por huella peptídica, pero sí es claro que tiene un efecto importante en la sensibilidad del análisis por MALDI-TOF y debe ser adaptado al tipo de muestras que analizamos [22]. El número de péptidos detectados en este análisis está altamente condicionado por el empleo de una determinada matriz y/o por la técnica de depósito en placa, pudiendo aumentar considerablemente y por tanto aumentando la probabilidad de identificar la proteína con un elevado nivel de confianza [23].

Adquisición del espectro en MALDI-TOF.

En la mayoría de los espectrómetros MALDI-TOF actuales también es posible procesar el espectro a la vez que se adquiere, obteniendo así espectros calibrados gracias a picos de masas conocidas presentes en el espectro, filtrados con respecto a la eliminación de ruido y aplicada la corrección de la línea base. En los espectros sólo aparecerán las masas monoisotópicas de aquellos picos que superen los criterios de detección seleccionados (peak detection), tales como un umbral de relación señal / ruido (S/N) por encima del cual los picos son perfectamente distinguibles del ruido.

Por tanto, hay que seleccionar adecuadamente:

- 1. El método de adquisición, donde se especifica, básicamente, el método de operación (MS reflector positivo), la intensidad del láser y el rango de masas (m/z) en el que se mide. También deben ajustarse diferentes parámetros como el tiempo de retardo y una serie de voltajes, de forma que se mida en torno a la masa elegida con exactitud y la mayor resolución posible teniendo en cuenta las especificaciones del equipo. El número de pulsos de láser por espectro y el número de espectros que se van a acumular hasta lograr el espectro MS final también son parámetros a tener en cuenta y están entre los más relevantes.

- 2. El método de procesamiento en el que se especifica qué tipo de calibración se realizará (externa, interna o por defecto), los valores de las masas y el número mínimo que se van a usar para calibrar si se trata de una calibración interna, así como, el umbral de S/N que deben superar, que suele ser alto, ya que los picos utilizados para calibrar deben ser claramente detectados. Siempre que sea posible, es recomendable aplicar calibración interna para obtener la máxima exactitud de medida de la masa, requisito necesario en el caso del análisis para la identificación de proteínas por huella peptídica. Generalmente se utilizan para calibrar los espectros, los péptidos presentes en de masa conocida..., procedentes de la autolisis de la tripsina con la que digerimos.

Como resultado del análisis de los péptidos de la muestra por MALDI-TOF, se obtiene un espectro MS de calidad del que se extrae una lista de masas monoisotópicas (MH+) experimentales ("peak list") que se comparará, usando un motor de búsqueda, con las digestiones teóricas de las secuencias aminoacídicas de las proteínas anotadas en la base de datos seleccionada. Para hacer esta comparación, la tolerancia del error permitida debe ser muy pequeña (<50ppm).

En ocasiones, aun siendo el espectro de buena calidad en cuanto al número de picos, intensidad de las señales y resolución y consiguiendo identificar una proteína con una puntación óptima y una cobertura de secuencia elevada, se observa respecto a los péptidos que "no están todos los que son, ni son todos los que están". Es decir, los péptidos correspondientes al % de secuencia de la proteína no cubierta no se encuentran en el espectro. Por otro lado, existe una serie de señales en el espectro que no son asignados a la proteína identificada. Esto tiene distintas explicaciones, algunas de las cuales se detallarán a continuación.

En general, al digerir en gel una proteína con tripsina y analizarla por espectrometría de masas, resulta bastante extraordinario detectar todos los péptidos que componen la misma, tal que se cubra el 100 % de la secuencia. La tripsina puede saltarse algunos de sus puntos de corte,

bien por estar muy próximos, por estar seguidos de prolina o por presentar modificaciones en las lisinas o argininas que impiden la acción de la proteasa. En estas circunstancias pueden generarse péptidos de gran tamaño (>4500 Da) que difícilmente van a conseguir salir del gel por difusión, tanto más cuanto mayor sea su naturaleza hidrofóbica. Por otra parte, estos péptidos quedarían fuera del rango de masas en el que se mide rutinariamente para el análisis de huella peptídica, ya que el rango más usado suele ser de m/z 800 a m/z 4000 Da, aunque péptidos por encima y por debajo de esos valores pueden ser detectados. Por debajo de 800 Da se detectan contaminantes químicos, como iones pertenecientes a la matriz, y péptidos que tienen poca capacidad informativa por el pequeño número de aminoácidos y contribuyen menos a la cobertura de secuencia y a la puntuación para identificar una proteína. Por encima de 4000 Da, como hemos comentado, se dificulta la recuperación de los péptidos del gel de poliacrilamida.

Por otro lado, pueden producirse péptidos generados por cortes inespecíficos de la enzima. Estos péptidos, aunque pertenecen a la proteína identificada, no serán asignados a ésta ya que no cumplen el requisito de la especificidad de corte de la enzima.

En algunos casos pueden realizarse diferentes acciones dirigidas a **¨aumentar los que están y disminuir los que no son¨.** Normalmente esto exige previamente la inspección visual de los espectros por un experto, consume tiempo y requiere el uso de programas específicos y/o herramientas informáticas que permitan diferentes procesamientos, como filtrar la lista de picos, buscar posibles modificaciones, etc.

Para **aumentar los que están,** algunas de las acciones a seguir tienen que ver principalmente con la preparación de la muestra y el análisis por espectrometría de masas. No todos los péptidos se ionizan igual por su estructura primaria y por tanto algunos de ellos serán mucho menos detectables o indetectables. Se puede intentar realizar extracciones con diferentes porcentajes de acetonitrilo después de la digestión, siempre que la cantidad de muestra no sea limitante, para conseguir un mayor número de péptidos y se puede investigar el empleo de otras matrices o métodos de aplicar la muestra en la placa del MALDI-TOF.

Por otro lado, es posible modificar el rango de masas en el que mide el espectrómetro reajustando los parámetros para centrarse en el análisis de determinadas masas de interés, que no hayan sido detectadas.

Otra posibilidad es que el espectro no esté bien calibrado. Podemos usar el concepto de defecto de masa tanto para la inspección relativa a la calibración de un espectro como para discriminar si un pico con un (m/z) determinado es de naturaleza peptídica o no, siempre y cuando el espectrómetro de masas sea capaz de medir con elevada exactitud y resolución la masa. Se define defecto de masa como la diferencia entre la masa exacta de un compuesto y su masa nominal. Así por ejemplo en el caso del carbono, ^{12}C, cuya masa es 12 tomada por referencia, el defecto de masa es 0; para el ^{1}H, de masa 1,0078, es +0,0078; para el ^{14}N, de masa 14,0030, es +0,0030 y -0,0051 para el ^{16}O, de masa 15,9949. Este concepto ha sido utilizado a través de los años en espectrometría de masas y últimamente, cada vez más, gracias a la capacidad creciente de los espectrómetros de masas en la exactitud y alta resolución de la medida de la masa. Generalmente las masas se miden con 4 decimales como mínimo.

El defecto de masa [24] se puede utilizar para filtrar de forma selectiva un gran número de datos MS para seleccionar un compuesto de interés. Según la composición elemental de un péptido y su fórmula general, teniendo en cuenta la masa exacta de cada elemento y su aportación a la masa molecular del péptido, la composición isotópica y la contribución de la masa isotópica relativa a cada uno de los isótopos, las fracciones de masa (es decir, los decimales de la masa molecular) serán diferentes según sea el tamaño del péptido. Así un péptido de masa 200 Da, tendrá una fracción de masa de 1, es decir su masa molecular exacta será 200,1 Da. El incremento será de 0,1 Da de fracción de masa para cada 200 Da de masa molecular. A

partir de 2500 Da esta regla se ajusta menos a la realidad. Esto nos permite, de un vistazo al espectro, sospechar si la calibración es correcta o no. Cuando la cantidad de proteína es muy abundante, los picos pertenecientes a la tripsina pueden no aparecer claramente en el espectro produciendo una calibración interna incorrecta que dará lugar a un mayor error en la medida. Por tanto, la tolerancia permitida al hacer la comparación con la digestión "*in silico*" de las bases de datos es menor que el error en la medida de la masa y en consecuencia estos péptidos no serán asignados a la proteína identificada de la que proceden, a no ser que aumentemos la tolerancia en cierta medida (teniendo en cuenta que aumentamos el riesgo de obtener una identificación falsa) o idealmente recalibraremos el espectro y repetiremos la búsqueda con los nuevos valores de masas.

Otro aspecto a tener en cuenta para mejorar el número de picos asignados del espectro y aumentar así la fiabilidad de los resultados, es una posible elección errónea de la masa monoisotópica, sobretodo en el caso de iones de masa elevada (>3000Da), donde el primer pico, correspondiente a la masa monoisotópica, no suele ser el de mayor intensidad como sucede en general para los picos de masa <2000 Da. La distribución isotópica de un péptido mayor de 2000 Da muestra el primer pico, correspondiente a la masa monoisotópica, menos intenso que el siguiente (molécula del péptido que ha incorporado un ^{13}C, por ser el siguiente isótopo más abundante en la naturaleza) y que el siguiente (molécula del péptido que ha incorporado dos ^{13}C), y posiblemente que el siguiente ya que a medida que aumenta la masa molecular aumenta la probabilidad de incorporar ^{13}C. Como se muestra en la Figura 7, si la masa del péptido es muy elevada, el pico correspondiente a la masa monoisotópica del péptido puede quedar por debajo del umbral de detección que hemos fijado (por ejemplo S/N=40) y sin embargo el pico correspondiente al siguiente isótopo, que pesa 1 Da más, puede superarlo, de manera que este valor sea el elegido erróneamente como masa monoisotópica y será incorporado a la lista de picos para hacer la comparación en las bases de datos. En este caso, no habrá coincidencia con la masa teórica del péptido (el error es de aproximadamente 1 Da), y puede presentar coincidencia de forma errónea con otra proteína. Corregir, una vez realizada la inspección visual del espectro prestando atención, sobre todo, a los Picos con valores de, la elección de las masas monoisotópicas que aparecen en la lista de picos reajustando el umbral de detección, puede permitir asignar ese péptido y por tanto identificar de forma concluyente la proteína o aumentar la confianza del resultado. Ante un péptido con una distribución isotópica dudosa o mal resuelta,

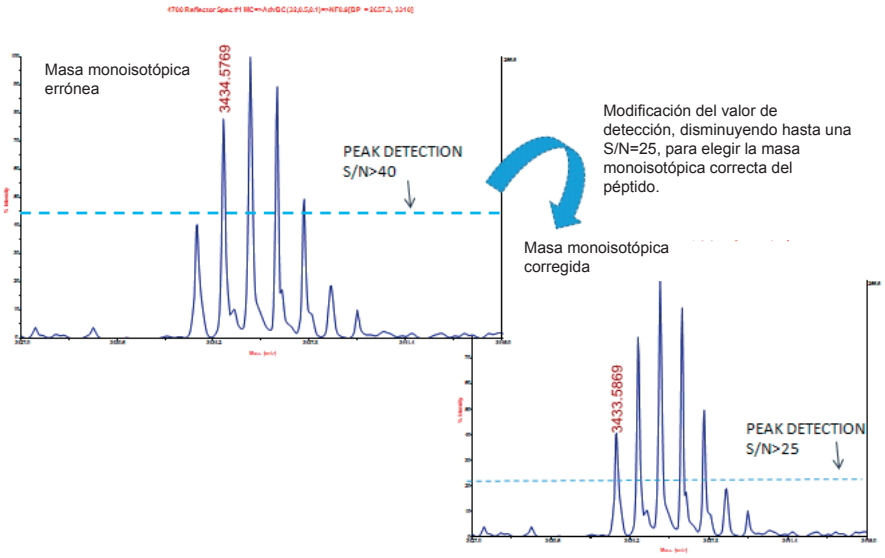


Figura 7. Patrón isotópico del péptido MH+ 3433,58 y elección de la masa monoisotópica según el umbral de detección (S/N) fijado.

es mejor eliminar el pico de la lista, sobre todo si se trata de péptidos de masa alta, ya que si existe coincidencia con la masa teórica, al tener un número mayor de aminoácidos, contribuye en gran medida en la puntuación, dando como resultado una identificación falsa.

Otra posibilidad es que la proteína digerida experimentalmente presente diferencias en parte de la secuencia con respecto a la proteína anotada en la base de datos, debido a sustituciones, mutaciones, o un procesamiento alternativo, de manera que algunos de los péptidos habrán cambiado su masa y la coincidencia entre las masas teóricas y los experimentales.

En el caso de proteínas recombinantes, deben tenerse en cuenta los cambios que incluyan en su secuencia, siendo conveniente crear una base de datos que contenga específicamente la secuencia de esta proteína para obtener una mayor coincidencia y cobertura de secuencia como se muestra en la Figura 8.

Algunos péptidos pueden presentar modificaciones post-traduccionales (PTM) variando su masa, como ocurre en el caso de la oxidación de metioninas. Si existe la sospecha de alguna modificación en concreto, se puede incluir en los parámetros de búsqueda en Peptide Mass o MS Digest, realizando así las digestiones teóricas tales que se incluya la modificación y observar las masas de los péptidos resultantes para comprobar si se encuentran en el espectro. También se pueden emplear diferentes herramientas (muchas de ellas presentes en SIB Bioinformatics Resource Portal http://www.expasy.org/proteomics), como "FindPept tool", que trata de identificar las modificaciones correspondientes a péptidos que están presentes en la muestra, ya sea por una rotura inespecífica de la secuencia, modificaciones químicas artefacto, modificaciones post-traduccionales o cortes autolíticos de la proteasa, de modo que esas señales no han sido correlacionados con ninguna entrada o proteína dada. "Findmod tool", permite predecir modificaciones post-traduccionales (PTM) potenciales y localizar sustituciones de aminoácidos en las secuencias peptídicas.

En la web "The Association of Biomolecular Resource Facilities" se encuentra la herramienta Delta Mass (http://www.abrf.org/ index.cfm/dm.home) que permite obtener un listado de variaciones de masa asociados específicamente a modificaciones determinadas. La herramienta

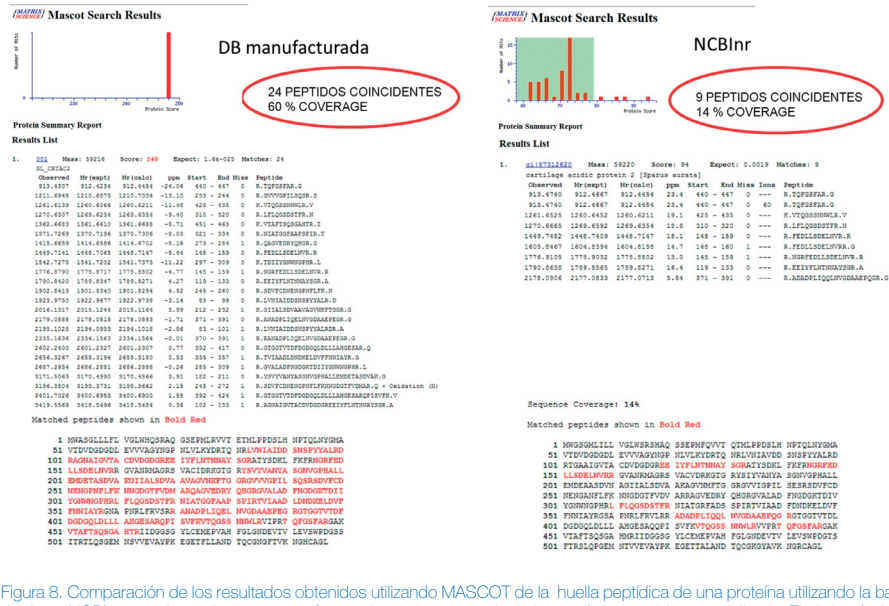


Figura 8. Comparación de los resultados obtenidos utilizando MASCOT de la huella peptídica de una proteína utilizando la base de datos NCBInr y una base de datos manufacturada con la secuencia de la proteína recombinante analizada. El nº de péptidos coincidentes pasa de 9 a 24 y la cobertura de secuencia de 14 al 60 % para la base de datos específica para la proteína.

Unimod (www.unimod.org) presenta similares características. En cualquier caso, la existencia de PTM debe de ser confirmada por espectrometría de masas en tándem (MS/MS).

Por otro lado al hablar de **"disminuir los que no son"** nos referimos a muchas de las señales que no están asignadas a la proteína identificada, es decir, que no forman parte de su huella peptídica y por tanto son contaminantes. Pueden ser iones peptídicos procedentes de otra proteína del mismo organismo objeto del estudio, o bien sustancias químicas procedentes de los reactivos utilizados durante todo el proceso, aunque éstos normalmente no son erróneamente identificados, ya que suelen tener una fracción de masa suficientemente diferente a la de los péptidos y siempre que el espectro esté correctamente calibrado y la exactitud de la masa sea adecuada (la masa de los péptidos tienen un decimal muy diferente a los reactivos contaminantes y por tanto no van a interferir en la identificación). Normalmente estos picos se encuentran por debajo de 1200 Da, por ejemplo en el caso de iones de la matriz ACHC, que por su composición y fórmula su primer decimal suele ser diferente al que le correspondería a esa masa cuando se trata de un péptido (ACHC m/z de 800 a 1000 Da con el primer decimal próximo a 0, péptido m/z 800 su primer decimal debe ser 4 o muy próximo a 4; como 800,37 o 800,48). Otros iones peptídicos contaminantes fácilmente reconocibles son los péptidos de la autolisis de la tripsina y algunas queratinas del pelo, piel, lana, etc. Como ya hemos comentado, estas contaminaciones se producen de manera frecuente cuando la cantidad de muestra que analizamos es pequeña ya desde el gel de partida.

Los picos que se hayan identificado como contaminantes, ya sean de un tipo u otro, es conveniente eliminarlos de la lista de masas experimentales que se van a incluir en la búsqueda contra bases de datos, ya que no contienen información de la molécula de interés. Esto se puede hacer fabricando una "lista de exclusión" que incluya por ejemplo las masas de los péptidos autolíticos de tripsina, queratinas, etc. La mayoría de los espectrómetros de masas poseen herramientas (programas) para importar la lista de masas de cada espectro MS de forma automática, filtrar los picos y enviar la lista a MASCOT u otros motores de búsqueda especificando los parámetros de dicha búsqueda. Estos programas pueden seleccionar del espectro un número de picos determinados, eliminando los que aparecen en la lista de exclusión con una tolerancia de masa específica y en función de su intensidad, inspeccionando cada tramo de la escala m/z hasta llegar al total de picos que queremos incluir. Por ejemplo, podemos importar los valores de m/z de hasta 10 picos cada 200 Da, con una S/N>20, desde un valor de m/z de 850 a 4.000 Da hasta completar una lista de m/z de 100 picos. Garantizamos así tener picos a lo largo de prácticamente todo el rango de masa del espectro MS. Este parámetro se puede modificar según la apariencia de los espectros.

Programas de acceso libre, como "Peak Erazor", http://www.gpmaw.com/html/peakerazor_intro.html, permiten filtrar las listas de picos. Incluso recalibrar las masas si es necesario y crear una lista ERAZOR, similar a la lista de exclusión antes mencionada, en base a las veces que aparecen las mismas masas en muestras diferentes. En la práctica, se incluyen las listas de las masas de los picos (m/z) de los espectros MS de las muestras que queremos analizar y se busca, con la tolerancia especificada (por ejemplo, 100 ppm) aquellos picos que se repiten en dos, tres, cuatro, etc. de las listas, entendiendo por ello que son contaminantes, picos de matriz, de queratinas, etc., y no pertenecen a la proteína de interés. Estos picos conformarán nuestra lista ERAZOR que es guardada con otro nombre y puede ser utilizada como lista de exclusión para eliminar esas masas contaminantes de cada una de las listas de las diferentes muestras que vamos a buscar contra las bases de datos. También existe la opción de emplear la lista de ERAZOR propuesta por defecto en el software, que incluye contaminantes típicamente descritos como tal.

En los apartados anteriores, nos hemos referido principalmente a la identificación de proteínas previamente separadas por electroforesis en gel (2-DE) y en consecuencia únicas en la muestra. Sin embargo, cuando se analizan muestras complejas, incluso cuando han sido previamen-

te fraccionadas por SDS-PAGE, es muy probable que haya más de una proteína en la banda de gel. En este caso, la muestra digerida contendrá los péptidos resultantes de las proteínas presentes en esa banda de gel, que pueden encontrarse en la misma o distinta concentración y por tanto estar representadas por más o menos péptidos. La mezclas de proteínas que sean identificadas por huella peptídica, tanto de forma concluyente como no, deben ser confirmadas mediante un análisis de fragmentación de péptidos por espectrometría de masas en tándem (MS/MS).

1.3. Fundamento de la identificación de proteínas por espectrometría de masas en Tándem (MS/MS).

Este capítulo se centra en los aspectos de la identificación de proteínas, con herramientas bioinformáticas, a partir de datos obtenidos por espectrometría de masas en tándem (MS/MS), en concreto usando las búsquedas de bases de datos. Aunque en este capítulo no se trata, hay otras estrategias para identificar proteínas con datos MS/MS, como la secuenciación *de novo* que consiste en la interpretación manual de los espectros con buena calidad para extraer de ellos una secuencia de aminoácidos, o bien una etiqueta de secuencia, en caso de que sólo se pueda elucidar de forma clara una pequeña parte de la secuencia peptídica que es utilizada también para identificar la proteína en bases de datos.

La principal diferencia respecto a la identificación de proteínas por huella peptídica es que en este caso se identifican secuencias individuales de péptidos, y luego el conjunto de secuencias peptídicas se usa para deducir qué proteínas están presentes en la muestra analizada, por lo que se pueden analizar directamente mezclas de proteínas. Se puede estudiar cada péptido, elucidar su secuencia, por eso puede utilizarse como una técnica confirmatoria de los resultados obtenidos de la identificación de la huella peptídica. Esta técnica es aplicable tanto a muestras procedentes de gel como libres de él, es decir, mezclas de proteínas en solución. De modo general, la/s proteína/s son digeridas enzimáticamente y los péptidos resultantes son separados por uno o más pasos de cromatografía líquida y analizados por espectrometría de masas en tándem (ver Figura 9). En un primer análisis de masas (MS) se determina el valor m/z

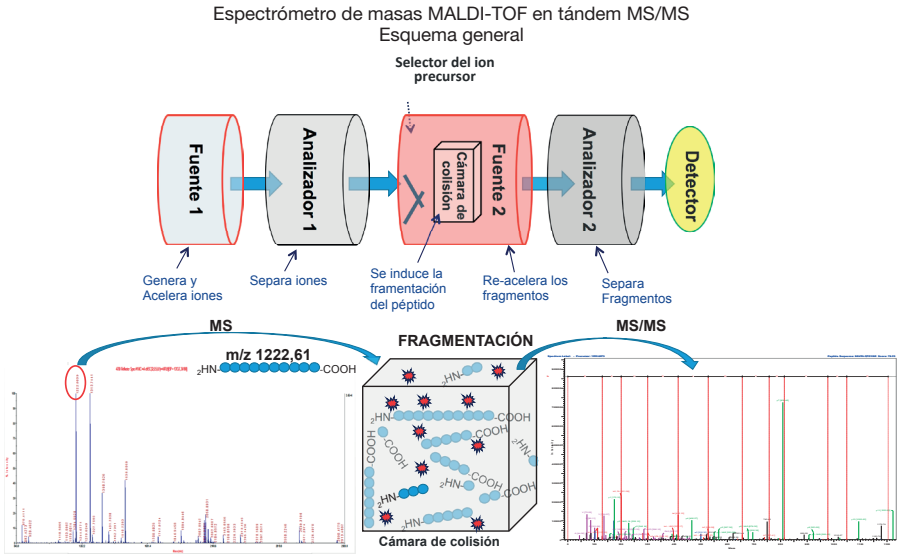


Figura 9. Esquema general de un espectrómetro de masas en tándem (MSMS).

de los péptidos, y algunos de estos son seleccionados de forma individual, aislados y fragmentados. Los fragmentos generados se separan y analizan en un segundo análisis de masas, obteniendo el espectro de fragmentación o MS/MS de cada péptido, que contiene información tanto de la masa del péptido precursor como de la secuencia de aminoácidos de éste. La principal ventaja frente a la huella peptídica es que en este caso obtenemos información de la estructura primaria del péptido.

En la Figura 10 se muestran los dos tipos de aproximaciones proteómicas y el flujo de trabajo típico para la identificación de proteínas por espectrometría de masas.

La aproximación libre de gel tiende a ser más rápida y simple en cuanto a manipulación de la muestra, sin embargo requiere equipamiento sofisticado de LC y EM que requieren de mantenimiento continuo. El paso de digestión en solución es, al igual que en gel, crítico [25] y el que más tiempo consume del proceso.

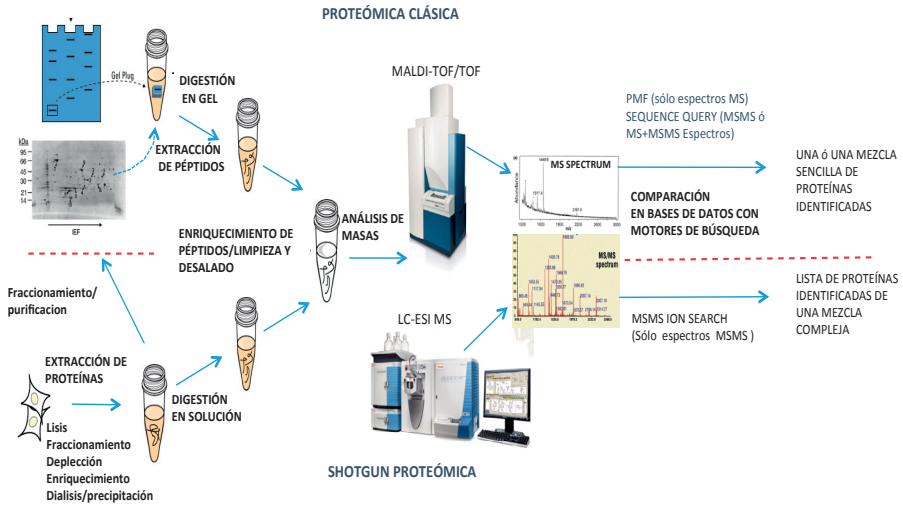


Figura 10. Tipos de aproximaciones proteómicas y flujo de trabajo típico para la identificación de proteínas por espectrometría de masas.

DIGESTIÓN EN SOLUCIÓN.

Básicamente, incluye los mismos pasos que la digestión en gel, comenzando con la solubilización de las proteínas en un tampón adecuado que no afecte ni al proceso de digestión, ni el posterior análisis por EM. Es conveniente evitar detergentes, como SDS o CHAPS incluso a bajas concentraciones, algunos de los reactivos más utilizados son Urea 6-8 M, 20 mM Bicarbonato Amónico (Ambi) o 0,5 M Bicarbonato Trietilamónico (TEAB). A diferencia de la digestión en gel, es necesario cuantificar la proteína que va a ser digerida para estimar la cantidad de enzima que hay que añadir en un ratio eficiente, generalmente de 1/10 a 1/100 (µg tripsina / µg de proteína), ya que éste debe asegurar que la cantidad de enzima es suficiente para digerir la proteína sin encontrarse en exceso, pues éste no puede ser eliminado como ocurre en la digestión en gel, produciendo numerosos productos autolíticos, incluso si se trata de tripsina modificada. Los pasos siguientes, de reducción y alquilación, se realizan en las mismas concentraciones y condiciones parecidas, teniendo la precaución, si el tampón de digestión es Urea, de incubar a una temperatura de 37ºC, en vez de a 56ºC, y aumentando ligeramente el tiempo de incubación durante el tratamiento con DTT, para evitar que ésta se desnaturalice. La incubación con tripsina se realiza a una temperatura de 37ºC, un pH óptimo de 8,5 y normalmente durante 4 a 16 h.

Es altamente conveniente medir el pH antes de añadir la tripsina y corregirlo si fuera necesario añadiendo el tampón adecuado (Ambi o TEAB) hasta alcanzar el pH óptimo. El tampón Tris también puede utilizarse con este fin, aunque hay que tener en cuenta que es incompatible, como otras sales, con el análisis de EM y debe ser previamente eliminado con un ZipTip®.

Existe bastante discusión acerca de cuál es el mejor método para estimar la eficiencia de la digestión. Uno de los más aceptados es determinar el número de péptidos trípticos y la cobertura de secuencia aminoácidica [26] comprobando tanto la eficiencia de la digestión como la de detección, el problema es que este parámetro del porcentaje de cobertura de secuencia podría resultar engañoso ya que depende tanto del análisis como del espectrómetro de masas y de los parámetros de búsqueda empleados al identificar la proteína en las bases de datos. Se puede digerir un proteína conocida (Ej: Citocromo C) como estándar interno para testar la eficiencia de la digestión.

Hay gran número de trabajos en los que se estudia y trata de optimizar el protocolo de digestión, dirigidos a aumentar la eficiencia de la digestión [27,28], reducir el tiempo del proceso [29,30,31], asegurar su reproducibilidad en estudios de cuantificación [32] o incluso a adaptar la digestión al estudio de proteínas específicas como en el caso de la digestión *in vivo* para el análisis del surfoma [33].

A continuación se detalla uno de los protocolos de digestión en solución que obtiene buen rendimiento para proteínas que se encuentran en un rango de concentración de 0,1 a 1,0 mg/mL.

1. **Determinar la concentración de proteína en la solución por el ensayo de Bradford u otro que resulte adecuado.**

2. **Solubilizar y desnaturalizar las proteínas**. Utilizar un volumen de 50 a 100 µl de 50 mM AmBi o Urea 8 M, dependiendo de la concentración de proteína.

3. **Reducción**. Añadir el menor volumen posible (1-5µl) de DDT en Ambi 50 mM, para que éste se encuentre a una concentración final en la muestra de 10 mM e incubar a 37ºC durante 1 h con agitación suave.

4. **Alquilación de cisteínas**. Añadir el menor volumen posible de iodoacetamida en 50 mM de Ambi para alcanzar una concentración final en la muestra de 55 mM e incubar 30 minutos a temperatura ambiente en oscuridad con agitación suave.

5. **Dilución** (**sólo en el caso de que el tampón utilizado sea Urea 8 M**). Diluir añadiendo Ambi 50 mM hasta que la concentración de Urea este por debajo de 2 M para poder llevar a cabo la digestión con tripsina.

6. **Digestión.** Comprobar con tiras de pH, que la muestra se encuentra en un rango de pH de 8,5. De acuerdo con la cantidad de proteína, reconstituir en fresco la tripsina en Ambi 20 mM y añadir tripsina para lograr un ratio 1:10-1:100 tripsina /proteína. Incubar de 4-16 h a 37ºC.

7. **Centrifiugar** para bajar el líquido condensado de la tapa del tubo y añadir un volumen pequeño de 1% TFA para acidificar la muestra (pH<3), deteniendo la reacción de digestión y preparando los péptidos para su posterior desalado en el siguiente paso.

8. **Desalado y concentración de los péptidos**.
Eliminar los resto de sales, Urea y otros contaminantes como restos de detergentes mediante un ZipTip® o poros R2, secar a vacío en centrifuga SpeedVac los péptidos eluidos y resuspenderlos en un volumen adecuado de 0,1% TFA 50% AcN si se van a analizar por EM MALDI-TOF/TOF sin separación previa o bien resuspender en el tampón

de carga de la cromatografía líquida, generalmente 1% de ácido Fórmico (FA) y hasta un 2% de AcN, para su separación y posterior análisis por nano-LC-ESI-MSMS.

9. Análisis por espectrometría de masas en tándem.

Análisis de los péptidos por espectrometría de masas en Tándem.

Cuanto mayor sea la complejidad de la muestra, mayor es la necesidad de separar multidimensionalmente los péptidos antes de su análisis por EM, para obtener una mayor resolución y rendimiento en la identificación de éstos, ya que los espectrómetros de masas sólo tienen capacidad para fragmentar un número limitado de péptidos en una ventana de tiempo determinada, eligiendo los picos más intensos para ser fragmentados, lo que excluye del análisis las proteínas de menor abundancia. Se puede intentar solventar este problema, para conseguir un estudio más profundo del proteoma, utilizando diferentes técnicas de separación previas al análisis por LC-MS/MS e implementando el análisis dependiendo de los datos adquiridos ("Data dependent analysis") para evitar fragmentaciones redundantes y aumentar así el número de péptidos fragmentados. La separación puede producirse a dos niveles, proteínas (previa a la digestión) o péptidos (posterior a la digestión). En el caso de la separación de proteínas puede realizarse un pre-fraccionamiento de la mezcla proteica mediante una separación por SDS-PAGE o, más recientemente, isoelectroenfoque en solución (OFF-GEL). Las fracciones obtenidas serán digeridas en gel o solución, respectivamente, y los péptidos obtenidos se separarán por cromatografía líquida a nano escala, es decir, con un flujo menor de µl/min, de fase reversa (RP-nanoHPLC). Si no ha habido separación previa de proteínas antes de la digestión, la mezcla compleja de péptidos se separa, generalmente, por combinación de distintas técnicas cromatográficas (LC-Intercambio catiónico (SCX) + LC-fase reversa (RP); RP-HPLC a pH básico + RP-HPLC a pH ácido, etc.,) que pueden estar acopladas o no en línea entre ellas y al espectrómetro de masas ("off-line" en los espectrómetros de masas tipo MALDI y "on-line" en los espectrómetros de masas con fuente de ionización ESI).

Típicamente en un experimento LC-MS/MS, se inyectan unos pocos femtomoles de péptidos en solución, que son separados en una columna capilar de fase reversa y eluidos en base a su hidrofobicidad. Son así transferidos con alta eficiencia al espectrómetro de masas donde son ionizados y analizados de acuerdo a su relación m/z.

Como ya hemos dicho, los espectrómetros de masas son la clave en los experimentos de proteómica, por definición un espectrómetro de masas consiste en una fuente de ionización, un analizador de masas que mide la relación m/z de los analitos ionizados y un detector que registra el número de iones a cada valor de m/z.

En general, en el primer analizador se selecciona un péptido (Ión precursor) que será aislado, generalmente, en una cámara y fragmentado por colisión con moléculas de un gas inerte; los iones fragmentos resultantes son separados de acuerdo a su valor m/z en un segundo analizador de masas generando un espectro MS/MS o de fragmentación (Figura 9). Los datos adquiridos contienen información del peso molecular de las especies peptídicas y su abundancia. En el caso de los EM con fuente ESI acoplados a cromatografía, la señal del ion en cada análisis (scan) puede ser sumada y representada a lo largo del tiempo (Ion Current Cromatogram, TIC), siendo esta representación similar a un cromatograma. Desde el TIC, pueden seleccionarse picos para visualizar su espectro de fragmentación (MS/MS).

Dependiendo del tipo de fragmentación y la energía, pueden romperse diferentes enlaces en la cadena peptídica, produciendo diferentes tipos de iones. En la Figura 11 se muestra la nomenclatura de los posibles iones producto generados por fragmentación de un péptido propuesta por Roepstorff P y Fohlman J [34].

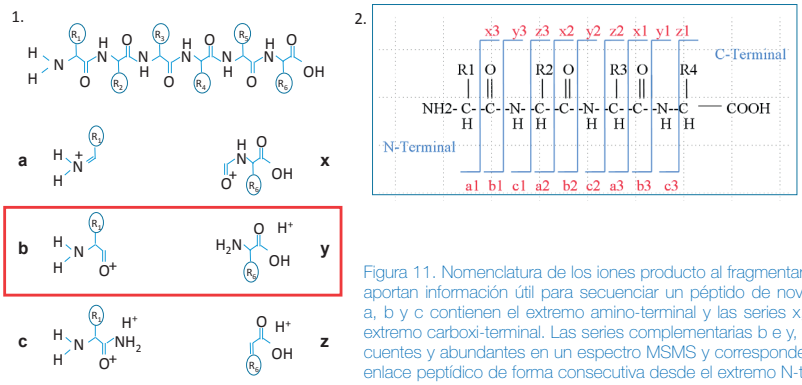


Figura 11. Nomenclatura de los iones producto al fragmentar un péptido que aportan información útil para secuenciar un péptido de novo. 1) Las series a, b y c contienen el extremo amino-terminal y las series x, y, z retienen el extremo carboxi-terminal. Las series complementarias b e y, son las más frecuentes y abundantes en un espectro MSMS y corresponden a la rotura del enlace peptídico de forma consecutiva desde el extremo N-terminal y C- terminal respectivamente. 2) Se muestran las series iónicas complementarias en un péptido teórico constituido por cuatro aminoácidos. Los subíndices que acompaña cada serie iónica se corresponden con la cantidad de residuos aminoacídicos que contiene el ión fragmento.

La técnica de fragmentación más usada y extendida en las últimas décadas, disponible en la mayoría de los espectrómetros de masas en tándem es disociación inducida por colisión ¨Collision Induced Dissociation" (CID). La fragmentación del péptido aislado en la cámara de colisión se produce por colisión con moléculas de un gas inerte como Nitrógeno, Argón o incluso gas atmosférico que es introducido a cierta presión. La transferencia de energía que se produce durante esta colisión induce la fragmentación del ion precursor, dando lugar a fragmentos cargados y neutros, que se aceleran en la segunda fuente y se separan en el segundo analizador de masas. Los m/z de estos iones fragmento son registrados por el detector dando lugar al espectro MS/MS. Los iones más frecuentes producidos en este tipo de fragmentación son los iones de las series ionicas ¨y¨ y ¨b¨, que se producen por rotura del enlace peptídico por el extremo C-terminal y N-terminal respectivamente, reteniendo la carga en el extremo correspondiente en cada tipo de fragmento. No todos las roturas son igualmente favorables, por ejemplo la fragmentación entre el primer y segundo aminoácido por el extremo N-terminal es energéticamente desfavorable, de manera que el ion b1 normalmente no es detectable. Sin embargo, típicamente aparece junto con el ion b2 el ion a2 (b2-28Da; b2-CO) con una intensidad prominente. En un péptido tríptico el primer aminoácido por el extremo C-terminal será lisina o arginina, así la serie iónica ¨y¨ comenzara con las masas y1=147 para la lisina o y1=175 para la arginina. A partir de este y1, el siguiente pico de fragmentación que coincide con el residuo de masa de algún aminoácido será el ion y2 y así sucesivamente; igualmente en la serie complementaria b, a partir del ion b2 aparecerán algunos de los picos correspondientes al resto de la serie. Idealmente la serie completa de los iones ¨y¨ y ¨b¨ confirman la secuencia de aminoácidos completa del péptido, pero en la práctica no todos los picos de las dos series están presentes o son inequívocamente interpretables. Además suelen aparecer otros fragmentos como los internos que corresponden a otro tipo de roturas difícilmente interpretables. A pesar de ello, normalmente parte de la secuencia puede interpretarse con un grado de confianza aceptable.

Si existe una base de datos de las secuencias de las proteínas en estudio, no es necesario extraer información de la secuencia manualmente, sino que directamente se identifican los péptidos por comparación con la base de datos, comparando las fragmentaciones teóricas de los péptidos tríptico que tengan la misma masa que los péptidos fragmentados con los espectros de fragmentación experimentales, de manera que si existen suficientes coincidencias podemos aceptar la secuencia como correcta.

La secuenciación *de novo* de espectros MS/MS para identificar proteínas está restringida a las proteínas cuyas secuencias aminoacídicas no están anotadas en bases de datos o cuando se obtienen espectros de buena calidad que no son identificados por presentar alguna modificación o ser producto de un corte enzimático inespecífico.

En el caso de emplear espectrómetros en tándem con fuente de ionización MALDI, el ion precursor está monocargado (MH+) y también los iones que se producen después de su fragmentación. En los espectrómetros que tienen fuente ESI, los iones precursores están multicargados, siendo el número de cargas positivas igual al número de aminoácidos básicos que incluyan (prolina, histidina, arginina y lisina) más el protón del amino terminal. Aunque en los estudios proteómicos los iones fragmentados con mayor frecuencia son iones con carga +2, los iones producto después de ser fragmentados suelen ser mono y/o doblemente cargados. Estos espectros son más difíciles de interpretar. Como ya se ha comentado antes, podemos conocer el estado de carga de un ion por la diferencia de m/z entre los distintos picos de la distribución isotópica del péptido. Algunos programas pueden convertir el espectro MS/MS donde aparecen iones con diferentes estados de carga a espectros con iones carga +1 (deconvolución), más fácilmente interpretables.

Los otros métodos de fragmentación desarrollados más recientemente "electron capture dissociation" (ECD) [35] y "electron transfer dissociation" (ETD) [36] producen predominantemente iones c y z, están disponibles sólo en algunos espectrómetros como FT y LTQ-OrbiTrap y son empleados, principalmente, en el estudio de modificaciones post-traduccionales, ya que al ser una fragmentación más suave que la que se produce por CID evitan la perdida de la modificación.

Algunos espectrómetros, como los FT y las trampas de iones lineales, tienen además la capacidad para obtener datos de un tercer estado de masa (MS^3) que contribuyen a aumentar el número de péptidos y proteínas identificados. Gracias a la velocidad de adquisición de estos espectrómetros, uno de los iones fragmentos (MS^2) puede ser aislado y sometido a una nueva fragmentación, sus iones fragmento son separados y detectados en un espectro de masas MS^3. La adquisición de este tipo de espectros de forma rutinaria en un experimento de LC-MS/MS puede aportar información útil adicional a la identificación y caracterización de proteínas [37]. Sobre todo en el caso de las proteínas poco abundantes identificadas por un único péptido y en el análisis de fosfopéptidos [38] en los que el MS^3 permite confirmar la modificación y el sitio de fosforilación.

En general, después del análisis por espectrometría de masas en tándem de una mezcla de proteínas tenemos una lista de péptidos fragmentados, desde uno a miles, que han sido caracterizados como picos en base a una serie de parámetros ("peak piking") y que incluye, como mínimo, los datos de la carga y la masa tanto del péptido precursor como de sus fragmentos y las intensidades de estos últimos, en un fichero que será utilizado para hacer la comparación con las bases de datos. Este fichero tiene un formato específico, dependiendo del espectrómetro donde hemos adquirido los datos, que más adelante se comentará, pero en todos los casos incluye los datos antes mencionados para cada péptido, separado de los de otros por las frases "BEGIN IONS" al comienzo y "END IONS" al final.

2. Motores de búsqueda en bancos de datos.

Los motores de búsqueda son las herramientas bioinformáticas que permiten la comparación de los pesos moleculares experimentales con los pesos moleculares teóricos de los analitos y ordenan el resultado de acuerdo con el mayor número de coincidencias.

A partir de cada secuencia de proteínas, es posible calcular los datos de pesos moleculares de las moléculas y así, el proceso de identificación tanto de proteínas como de péptidos se apoya en el uso de bases de datos de secuencias de aminoácidos. Los pasos a seguir se pueden simplificar en los siguientes puntos:

- Obtención de datos de pesos moleculares de péptidos o de sus fragmentos.
- Selección de la base de datos de huellas peptídicas teóricas o fragmentación teórica para cada péptido.
- Comparación de los datos experimentales con los pesos moleculares teóricos.
- Cálculo del número total de péptidos o de iones fragmento coincidentes.
- Cálculo de puntuación según el número de coincidencias.
- Identificación de la proteína o del péptido más probable.

En la actualidad hay muchos motores de búsqueda para identificar proteínas y péptidos basándose tanto en la huella peptídica, como en espectros MS/MS sin interpretar. La diferencia entre ambas aproximaciones radica en el objeto que se persigue en la identificación. Así en el caso de la huella peptídica, de forma iterativa y a partir de cada una las secuencias de las proteínas contenidas en la base de datos, se realizan digestiones *in silico* para obtener el valor m/z de cada uno de los péptidos que constituyen una proteína y así poder identificarla. En el caso del espectro de fragmentación, se realizan fragmentaciones teóricas de las secuencias de péptidos para obtener el valor m/z de cada ion fragmento y así poder identificar un determinado péptido en base al conjunto de los valores m/z de sus fragmentos.

2.1. Motores de búsqueda en identificación por huella peptídica.

Para la identificación de proteínas, los cuatro motores más empleados son: Protein Prospector (UCSF), ProFound (Universidad Rockefeller), Mascot (Matrix Science) y Aldente (GeneBio) (ver Tabla 5). Aunque en principio, la lógica de todos es la misma, existen diferencias en la valoración de cómo se empareja la secuencia de un determinado péptido con la señal (valor m/z) que es detectada en el espectrómetro, haciendo que el resultado de los motores de búsqueda varíe. En todos ellos, y fruto del modelo estadístico que cada uno emplea, la salida consiste en un lista ordenada por puntuación de los diferentes candidatos que, tras repetir el proceso de forma teórica, obtendrían una huella similar a la proporcionada por el espectrómetro. En la actualidad, Mascot es el más empleado y se basa en el sistema de puntuación Mowse [39].

Tabla 5. Motores de búsqueda más empleados

Nombre	URL
Mascot*	http://www.matrixscience.com/
ProteinProspector*	http://prospector.ucsf.edu/prospector/mshome.htm
Profound**	http://prowl.rockefeller.edu/prowl-cgi/profound.exe
Aldente**	http://www.genebio.com/
ProteinPilot	http://www.absciex.com/
Sequest	http://www.thermoscientific.com/
OMSSA	http://pubchem.ncbi.nlm.nih.gov/omssa/
X!Tandem	http://www.thegpm.org/TANDEM/index.html
Phenyx	http://www.genebio.com/
EasyProt	http://easyprot.unige.ch/
VEMS	http://portugene.com/vems.html
FIndPept	http://web.expasy.org/findpept/
InsPecT	http://proteomics.ucsd.edu/Software/Inspect.html
PepFrag	http://prowl.rockefeller.edu/prowl/pepfrag.html
Andromeda	http://www.maxquant.org/
MyriMatch	http://fenchurch.mc.vanderbilt.edu/software.php
MS-GFDB	http://proteomics.ucsd.edu/Software/MSGFDB.html

* Permite búsquedas huella peptídica y MS/MS. ** Sólo permiten búsquedas huella peptídica

2.2. Motores de búsqueda en identificación por espectrometría de masas en tándem.

Es esencial recordar que en la identificación de proteínas por espectrometría de masas en tándem (MS/MS), se identifican péptidos individuales para cada espectro MS/MS, no proteínas. Los motores de búsqueda se utilizan para determinar qué secuencia de péptido en una base de datos de proteínas proporciona la mejor coincidencia.

Cada entrada en la base de datos se digiere, *in silico*, se calculan las masas de los péptidos teóricos según las condiciones de la digestión y para cada péptido se deduce su fragmentación teórica. Es decir, si la masa calculada de un péptido coincide con la masa de un péptido observada o experimental, entonces se comparan espectros MS/MS con bases de datos de fragmentaciones teóricas. Mediante algoritmos, se calcula una puntuación para decidir qué secuencia de péptido de las bases de datos coincide en mayor medida con el espectro experimental.

Los motores de búsqueda que identifican péptidos a partir de sus iones fragmento son un poco más complejos que los descritos para la huella peptídica. Esto se debe a que no sólo se limitan a comparar espectros de fragmentación teóricos y experimentales, sino que además incorporan una funcionalidad extra de inferencia de proteínas [40] para poder agrupar los péptidos identificados con las proteínas con las que se asocian.

Todos estos programas operan con una premisa similar, que es que se conoce la masa del péptido precursor, las masas de los iones fragmento (aunque el tipo de iones generado es desconocido), y la especificidad de la enzima utilizada para generar los péptidos. Estas tres piezas de información pueden ser utilizadas para buscar en bases de datos de secuencias de proteínas. El procedimiento general consiste en un pre-filtro de masas de péptidos, donde se identifican en una base de datos de secuencia todos aquellos péptidos que coinciden con la masa molecular experimental del péptido que se pretende identificar (dentro de una tolerancia especificada). Una restricción opcional adicional es que estas secuencias derivadas de la base de datos coincidan con la diana consenso de la proteasa que se utilizó para producirlos. Dada una lista de secuencias de este tipo, se generan espectros de masas en tándem teóricos que se obtienen utilizando los principios de fragmentación de péptidos conocidos.

En la actualidad hay bastantes motores de búsqueda, comerciales y libres, para poder llevar a cabo la identificación de péptidos a partir de espectros MS/MS. Sin embargo, sólo una parte de estos son los realmente empleados en la rutina habitual del laboratorio. En este sentido, Mascot (www.matrixsicence.com) y Sequest (http://fields.scripps.edu/sequest/) como software propietario y X!Tandem (http://www.thegpm.org/tandem/) como software libre son los paquetes más utilizados en la actualidad. Junto a estos tres habría que resaltar la contribución histórica de *Open Mass Spectrometry Search Algorithm*, OMSSA, aunque lamentablemente en el tiempo que este manual está siendo escrito ha dejado de actualizarse y por tanto sólo representa una opción testimonial para complementar los resultados de otros motores de búsqueda que si se adapten a nuevos objetivos. A lo largo de este capítulo nos centraremos en Mascot y X!Tandem.

Son cuatro los elementos que son continuamente empleados en el proceso de identificación, y hacen referencia a los distintos escalones o estadios que se pueden considerar. En primer lugar tenemos los datos analíticos o espectro MS/MS. Este es una relación de puntos (x,y) que representan la pareja relación m/z e intensidad para cada uno de los iones fragmentos cuya señal ha sido registrada por el detector del espectrómetro. A lo largo de esta sección, nos referiremos de forma genérica con la letra S cuando hablemos de un espectro MS/MS. El segundo es el dato teórico o secuencia del péptido (P). El tercero es la proteína (Pro) de la que se obtienen un determinado conjunto de péptidos tras la digestión enzimática. Por último, la relación base de la identificación es el emparejamiento entre un péptido P y un espectro S, o *PSM: peptide-spectrum match.*

Mascot (http://www.matrixscience.com).
A pesar de ser un software propietario, Mascot permite realizar búsquedas a través de su interfaz pública para aquellos experimentos cuyo tamaño no exceda de 1200 espectros. La interfaz es muy parecida a la empleada para la identificación de proteínas por huella peptídica. Las principales diferencias se limitan a la tolerancia de ajuste de iones fragmentos (MS/MS tol erance.), la selección del instrumento, el tipo de fichero que recopila los espectros, la carga del péptido y también una opción para la cuantificación.

X!Tandem (http://www.thegpm.org/tandem/).
X!Tandem presenta tanto la opción de realizar la búsqueda a través de su interfaz web, como la descarga del software para realizar búsquedas locales. La interfaz web (http://h.thegpm.org/tandem/thegpm_tandem.html) presenta las opciones más habituales como base de datos, tolerancias o PTMs con otras más particulares del sistema como la opción de incorporar los datos a la base de datos GPM.

Esta interfaz representa una opción muy útil para realizar búsquedas aisladas, sin embargo para un uso más cotidiano se recomienda descargar el software e instalarlo localmente. Tras seguir las instrucciones que se facilitan en la página de descarga (http://www.thegpm.org/tandem/instructions.html) se podrán incorporar las bases de datos de secuencias que vayan a emplearse para la identificación de péptidos. La ejecución habitual es en línea de comandos. Esto supone una limitación para aquellos usuarios que no estén habituados a trabajar en modo consola.

Para mejorar este aspecto de la usabilidad de XTandem, se ha desarrollado la interfaz gráfica Searchgui (https://code.google.com/p/searchgui/) que presenta las mismas opciones que la disponible en la web, pero empleando la distribución local con su correspondiente configuración. Este interfaz permite realizar una búsqueda simultánea en X!Tandem y OMSSA, a pesar que este último no siga desarrollándose en la actualidad.

3. Bases de datos de proteínas.

En la actualidad hay multitud de repositorios que son de acceso público y permiten el acceso tanto por el nombre de la proteína como por la llave primaria o identificador interno de la entrada. Ambos, nombre e identificador de la proteína, varían dependiendo de la base de datos consultada, lo que obliga a tener siempre presente la referencia de la fuente de datos que se consulta para determinar con exactitud el ámbito de la información que se asocia a una determinada proteína.

Para las identificaciones basadas en huella peptídica es requisito indispensable contar con alguna base de datos de proteínas que contenga las secuencias de aminoácidos de las proteínas buscadas. Éstas en realidad contienen secuencias de proteínas desnudas, es decir, sin ningún tipo de modificación, y obtenidas en muchos casos directamente de la traducción del gen que sintetiza la proteína.

Aunque existen muchos repositorios, todos bajo la definición genérica de bases de datos de proteínas, no todos almacenan el mismo tipo de información, y varían desde los que almacenan datos referentes a la secuencia de aminoácidos de la proteína junto a anotaciones funcionales y bibliográficas, como SwissProt (http://www.uniprot.org/) o NCBInr (http://www.ncbi.nlm.nih.gov/), a otros que contienen datos relativos a la estructura tridimensional obtenidos por difracción de rayos X, como PDB (http://www.rcsb.org/pdb/home/home.do).

A continuación se presentan algunas de estas bases de datos a modo de ejemplo.

3.1 UniProt KB: UniProt/TrEMBL y UniProt/SwissProt (http://www.uniprot.org).
UniProt (*Universal Protein Resource*) es un catálogo de información que combina las secuencias de las proteínas con anotaciones funcionales y biológicas asociadas con dichas proteínas. UniProt es actualmente mantenida por Instituto Europeo de Bioinformática (EMBL–EBI), el Instituto Suizo de Bioinformática (SIB) y el repositorio de información sobre proteínas (PIR). Dentro de este recurso, es de especial interés para usuarios proteómicos UniProt KB (Protein Knowledge Base) que a su vez se compone de las bases de datos TrEMBL y SwissProt. Ambas están conectadas y son accesibles tanto conjuntamente como por separado, pero difieren en que SwissProt únicamente contiene entradas que muestran una evidencia experimental contrastada y que han sido revisadas por expertos de forma manual.

En la versión de Noviembre de 2013, UniProt/SwissProt contiene 541.762 secuencias, que a su vez corresponden a 192.577.305 aminoácidos y contrastadas en base a un total de 223,949 referencias experimentales. Por su parte, UniProt/TrEMBL contiene 48.180.424 de secuencias que engloban 15.282.737.498 aminoácidos.

Consultando UniProt.
Esta base de datos dispone de una interfaz web que permite la consulta tanto de proteínas aisladas como conjuntos de éstas. El resultado de una consulta sencilla muestra junto con la secuencia de aminoácidos de la proteína detalles experimentales que han sido extraídos tanto de las referencias bibliográficas como de las anotaciones disponibles. Esta información adicional proporciona cierto conocimiento sobre el papel biológico que la proteína en cuestión puede estar desempeñando. Además, toda la información mostrada se puede descargar localmente en distintos formatos de archivos.

Toda esta información es extremadamente valiosa cuando es necesario un análisis detallado de una proteína en concreto, pero para el objetivo de su identificación resultan mucho más valiosas las opciones que proporciona el sistema para poder descargar la secuencia de conjuntos concretos de proteínas. Para estos casos, acudiendo a la opción de recuperar (*Retrieve*) en la interfaz web es posible recuperar toda la información en relación al conjunto de identificadores de proteínas.

Aunque la opción de recuperación por lotes representa una opción más que atractiva para profundizar en las relaciones que se pueden establecer entre un conjunto reducido de proteínas, desde el punto de vista de la identificación masiva de proteínas por espectrometría de masas, resulta más interesante la opción de descargar los proteomas o subproteomas de referencia. Los distintos conjuntos de referencia están disponibles en: http://www.uniprot.org/taxonomy/complete-proteomes

3.2. NextProt (http://www.nextprot.org/).
NextProt es un repositorio de referencia para el caso del proteoma humano. Desarrollado en colaboración entre el Instituto Suizo de Bioinformática y la empresa Ginebra Bioinformática (GeneBio), NextProt es una plataforma que aglutina los recursos de diferentes bases datos con la intención de reunir toda la información de interés que exista en relación a las proteínas humanas. En la versión de Noviembre de 2013 cuenta con 20.140 proteínas humanas relacionadas con 393.716 publicaciones científicas.

Consultando NextProt.
Como en el caso de UniProt, el sistema dispone de una interfaz web que permite realizar consultas. Estas pueden estar centradas en una determinada proteína de interés, pero también está disponible el conjunto completo de las proteínas humanas. Además y en relación al proyecto

de caracterización del proteoma humano (HPP) (http://www.thehpp.org/) están disponibles para su descarga los distintos conjuntos de proteínas según el cromosoma al que están asociados.

3.3. NCBI (http://www.ncbi.nlm.nih.gov/refseq).

Otro de los sistemas de referencia es el dependiente del centro nacional de biotecnología de EE.UU (*National Center for Biotechnology Information, U.S.*) y que para el caso proteómica se centra en el banco de datos RefSeq. Esta base de datos, en la versión de Noviembre de 2013 contiene 36.036.343 proteínas provenientes de 31.646 organismos. Todas las secuencias, tanto de nucleótidos como de aminoácidos, correspondientes a estos organismos están disponibles para su descarga.

Resumidas los tres principales recursos en donde es posible descargar tanto las bases de datos completas como los proteomas de referencia para emplear en la identificación de proteínas y péptidos con la ayuda de los motores de búsqueda, la tabla 6 recoge el volumen y características de dichas bases de datos.

Tabla 6. Estadísticas de los principales repositorios empleados en proteómica.

	UniProt/SP	UniProt/Tr	NextProt	NCBI RefSeq
Número de entradas	541.762	48.180.424	20.140	36.036.343
Número de especies	13041	451756	1 (humano)	31.646
Separada por taxonomías	Si	Si	No*	Si
Revisada	Si (manual)	No	Si	No

* Dispone de separación por cromosomas.

4. Parámetros de búsqueda.

Además de los datos procedentes del análisis por espectrometría de masas, todos los motores de búsqueda requieren información adicional en la forma de "parámetros de búsqueda". Algunos tratan de mantener una sencilla interfaz de usuario, y ofrecer una cierta flexibilidad en cuanto a los parámetros solicitados. Esta sección recoge los parámetros de búsqueda más importantes (ver Figura 12).

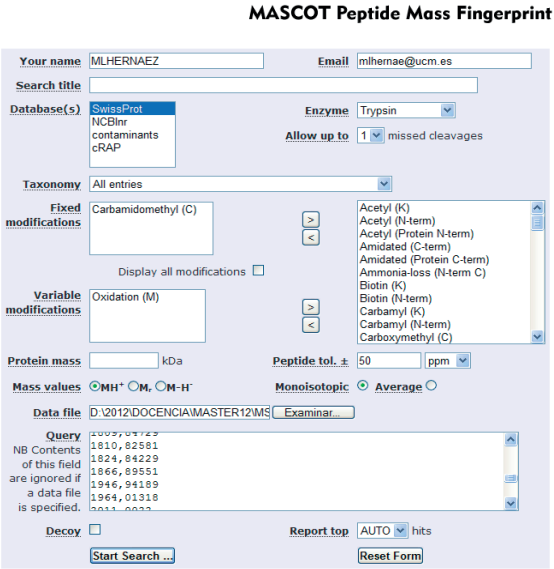


Figura 12. Interfaz del motor de búsqueda MASCOT para identificación de proteínas mediante huella peptídica.

4.1. En identificación por huella peptídica.

4.1.1. Enzima de digestión.

Dado que la mezcla peptídica a analizar es el resultado de la digestión de la proteína con una enzima en particular, será necesario indicar la enzima que se ha empleado y seleccionarla en el buscador. La tripsina es la enzima más ampliamente utilizada en experimentos de proteómica, ya que su actividad endoproteasa es muy específica, rompiendo en el extremo carboxiterminal de lisina y arginina, excepto si va seguido de prolina.

Aunque se recomienda evitar la opción de realizar búsquedas sin definir el tipo de proteasa, a veces interesa realizar búsquedas sin especificidad enzimática, por ejemplo por un motivo confirmatorio de la identificación; así, si la proteína ha sido digerida con tripsina y el espectro coincide con el de un péptido tríptico, se aumenta la confianza en el resultado. En otras ocasiones, puede emplearse para asignar péptidos a la proteína si existen evidencias de que existe proteolisis no específica de modo considerable, probablemente debido a la fragmentación del analito en la fuente del espectrómetro de masas. En estos casos, se puede usar la opción de semi-especificidad de la enzima, es decir, se consiente que uno de los extremos no coincida con los residuos por donde corta la enzima (dicho de otro modo, uno de los extremos de corte no es ni R, ni K, a diferencia de una diana perdida, que tiene una R o K interna, y los dos extremos de corte son R y/o K).

4.1.2. Sitios de corte perdidos (“miss-cleavages”).

Un parámetro que se puede modificar es el número de sitios de corte de la enzima que se permite que se pierdan o, dicho de otro modo, se salten dentro de la secuencia de la proteína. El valor fijado por defecto es 0. Se suele utilizar 1, pero se puede aumentar este valor a 2 si se observa un cierto número de péptidos de mayor tamaño o se trata de proteínas fosforiladas o modificadas de manera que la digestión no se espera que sea totalmente efectiva y se observarían cortes parciales. No es conveniente especificar un número más alto sin una buena razón, ya que cada nivel adicional de dianas perdidas aumenta el número de masas de péptidos que vayan a ser comparadas con los datos experimentales.

4.1.3. Datos experimentales.

Los datos que se envían (“Query”) para la identificación de la proteína mediante el motor de búsqueda correspondiente consisten en un listado con la relación de los valores m/z de cada péptido y/o fragmento, y a veces de las intensidades de estos iones.

Aunque en algunos de los motores de búsqueda las listas de m/z obtenidas de los espectros pueden incluirse directamente en la interfaz junto al resto de parámetros, lo habitual es trabajar con ficheros que contienen estos datos. En este sentido un elemento con el que hay ser cautelosos es el formato de fichero que contiene los datos. Las listas de picos suelen presentarse según los formatos .MGF, .DTA y .PKL desarrollados para Mascot, Sequest y Waters Masslynx, respectivamente. Estos formatos son directamente interpretables por cualquier editor de textos y son relativamente sencillos, ya que esencialmente contienen poco más que pares de relaciones m/z e intensidades. Sin embargo, existen otros formatos (los conocidos como RAW) que son binarios y por tanto sólo interpretables por paquetes de software desarrollados ex profeso por las mismas casas comerciales de los espectrómetros. Para poder trasladar estos últimos formatos a cualquier plataforma analítica que sea distinta a la empleada en la obtención de dichos datos, se han desarrollado una serie de aplicaciones que básicamente traducen la información contenida en los ficheros RAW a los estándares de datos propuestos por el comité internacional encargado de estandarizar los datos y resultados en proteómica, HUPO-PSI (http://www.psidev.info/). En el caso de la espectrometría de masas, el formato estándar recibe el nombre de mzML y engloba tanto los metadatos que describen el proceso de adquisición como las listas de picos que representan a los espectros.

Como consideración general, la detección de las masas se debe realizar después de la calibración del espectro con los estándares internos, para lo cual se pueden emplear las masas de los péptidos autoproteolíticos de la tripsina, así como las de los péptidos añadidos para la propia calibración. Tantos unos como otros se omiten para las búsquedas de bases de datos.

Por supuesto, también debe de especificarse si las masas de los péptidos son monoisotópicas o masas media, y en este último caso, pueden conllevar a una identificación de un falso positivo. En este punto, hay que resaltar que los análisis deben llevarse a cabo en espectrómetros de masas capaces de resolver el cluster de isótopos ^{13}C de cada péptido. Así una distribución isotópica correcta y una relación señal/ruido suficiente son los requisitos mínimos para la definición de la masa de los picos.

4.1.4. Base de datos de secuencias peptídicas.

Existen los siguientes tipos de bases de datos de secuencia de aminoácidos:

- **1. Primarias:** secuencia nucleotídica.
- **2. Secuencia de proteínas:** derivada de la traducción de todas las entradas de la anterior.
- **3. Secuencia de proteínas revisadas (curated):** se construye selectivamente de la de nucleótidos.

Por otro lado, existen 5 factores relevantes en cuanto a la utilidad de las bases de datos:

- 1. Número de entradas que contienen.
- 2. La frecuencia de errores en sus entradas.
- 3. Redundancia de las entradas.
- 4. Presencia de información auxiliar o anotaciones por cada entrada: función, modificaciones post-traduccionales, homologías con otras proteínas, enfermedades asociadas con deficiencias en esa proteína.
- 5. Frecuencia de la actualización de la base de datos.

Para la elección de una base de datos, existen dos enfoques posibles. Una base de datos extensa, con amplias anotaciones de especies, asegurará que el mayor número posible de candidatos sean considerados. Por otra parte, una base de datos más pequeña, específica de la especie producirá un menor número de falsos positivos. En el caso de un análisis proteómico, las dos bases de datos más empleadas son la no redundante del National Center for Biotechnology Information (NCBI nr) y las que dependen de Uniprot (SwissProt y TrEMBL). En la sección dedicada a las bases de datos de proteínas se describen las principales características de éstas [41].

4.1.5. Taxonomía.

Si una base de datos contiene información de taxonomía, la mayoría de los motores de búsqueda pueden utilizar este parámetro para restringir la búsqueda a las entradas de un organismo particular o rango taxonómico. Esto acelera la búsqueda porque, en efecto, hace que la base de datos sea más pequeña. Sin embargo, se pueden seleccionar todas las entradas, sin concretar la especie, y puede ser útil para confirmar el resultado si concuerda con la especie de la que procede la proteína. Por otro lado, si se estudian organismos diferentes de humano, las masas de queratina contaminantes deben ser consideradas, para no asignarlas a proteínas de otra especie.

La limitación de la taxonomía también simplifica el informe de resultados, ya que elimina las proteínas homólogas de otras especies. No es una buena idea especificar una taxonomía muy específica en aquellas búsquedas donde no contemos con el proteoma completo de la especie, ya que puede ser muy útil para identificar por coincidencia en secuencias una proteína de

una especie similar. Esto es especialmente importante para especies poco representadas. Por otro lado, también se recomienda restringir la base de datos de las especies de interés para disminuir el número de posibles coincidencias al azar.

4.1.6. Intervalo de masa y punto isoeléctrico (pI).

Algunos programas de búsqueda, en concreto para huella peptídica, permiten fijar el peso molecular de la proteína intacta como un parámetro que limita la búsqueda a un menor número de proteínas, lo que puede ser un poco arriesgado, y no se recomienda, ya que *in vivo* y *ex vivo* la proteólisis puede producir proteínas con un peso molecular más bajo, diferente de lo que se calcula a partir de la secuencia del gen; sin embargo, podría evitar falsos resultados.

4.1.7. Modificaciones.

Las proteínas y los péptidos pueden ser modificados en cientos de formas. Algunas modificaciones se refieren a la función biológica, pero otras son consecuencia de la manipulación de la muestra. Las bases de datos más completas de modificaciones de proteínas son UniMod (http://www.unimod.org), que se centra en modificaciones relevantes para el análisis por espectrometría de masas, y RESID (http://www.ebi.ac.uk/RESID/), que se concentra en las modificaciones naturales sobre todo después de la traducción [42].

En la búsqueda de base de datos, se pueden seleccionar las modificaciones de dos tipos: fijas y variables. En primer lugar, se pueden seleccionar las modificaciones cuantitativas, generalmente llamadas fijas, estáticas o estequiométricas. Un ejemplo sería la modificación de la cisteínas que se realiza por tratamiento con Iodoacetamida durante la preparación de la muestra para alquilar los grupos sulfidrilo tras la reducción con DTT de los puentes disulfuro y produce la carbamidometilación de las cisteínas. Puesto que todas las cisteínas están modificadas, únicamente conlleva un cambio en la masa de cisteína y por ello no supone gran diferencia en términos de velocidad de búsqueda.

En segundo lugar están las **modificaciones variables**, que son no estequiométricas. Un ejemplo es la fosforilación así como otras modificaciones post-traduccionales. La inclusión de modificaciones variables en una búsqueda se puede realizar con moderación para aminoácidos menos frecuentes, por ejemplo la oxidación de la metionina (+16 Da), pero fácilmente oxidable. Cuando la modificación afecta a residuos más frecuentes como la fosforilación en serina y treonina, no es adecuado incluirlo en la búsqueda si no es objeto del estudio, porque tiene repercusión en un gran número de péptidos y se incrementa el tiempo de la búsqueda y la posibilidad de encontrar falsos positivos. Además, la mayoría de las modificaciones post-traduccionales no afectan a todos los aminoácidos, sino sólo a alguna posición del residuo en el péptido. Por ejemplo, la fosforilación puede afectar sólo a una de las serinas de las muchas que puede contener un péptido determinado. En general, el número de combinaciones y permutaciones aumenta geométricamente con el número de modificaciones que se consideren; una llamada explosión combinatoria. Es por esto que es muy importante ser lo más conservador posible en lo que respecta a la inclusión de modificaciones variables en las búsquedas.

Si no se conoce la modificación, se deberían seleccionar estas tres opciones: la acetilación del extremo amino para las proteínas eucariotas, la oxidación de la metionina y la conversión de glutámico a piroglutámico, ya que son modificaciones variables comunes.

Una tercera clase de modificaciones, hace referencia a los isótopos estables que se usan en experimentos de cuantificación. Algunos péptidos, por ejemplo de las muestras del grupo A), llevarán una etiqueta ligera (o sin etiqueta) y otros péptidos, por ejemplo de las muestras del grupo B), llevarán una etiqueta pesada. Es una modificación variable porque la mayoría de los péptidos se encuentran en las dos muestras a comparar y habrá mezcla de ligeros y pesados.

4.1.8. Tolerancia.

La tolerancia es el parámetro más crítico en la búsqueda y su modificación repercute directamente en los resultados de las identificaciones obtenidos. La exactitud de masa es la diferencia entre el valor de la masa medida experimentalmente (M exp) y el valor teórico (M teórica) de la misma:

absoluta (Da) = [M exp-M teórica]
relativa (%, ppm) = [M exp-M teórica] / M teórica

Ejemplo: 0,05 Da = 0,005 % = 50 ppm para una masa experimental de 1000 Da.

La tolerancia en la medida de la masa molecular (o m/z) del péptido y la ventana de error en los valores de masa experimental de los péptidos, son dependientes del espectrómetro de masas utilizado y la exactitud de la calibración.

Especificar una tolerancia de masa demasiado ajustada es un motivo muy frecuente de no poder conseguir una coincidencia o resultado, pero si el valor de la tolerancia es demasiado laxo se pueden identificar falsos positivos. Los informes generados con los resultados de una búsqueda, por lo general incluyen gráficos que muestran el error cometido para cada valor de masa peptídica detectado (diferencia entre el valor teórico y el experimental), y es importante observar que no hay grandes desviaciones entre los errores, para tener confianza en los péptidos identificados.

En general, es preferible realizar una primera búsqueda con una selección de parámetros conservadora, más estricta, por ejemplo sin sitio de corte perdido o "missclevage" o sin modificaciones variables, y con una tolerancia de masa ajustada. Posteriormente, esta primera búsqueda se puede complementar empleando condiciones más flexibles con el fin de aumentar el número de coincidencias, tal que, por ejemplo, se permitan tolerancias más amplias en masa.

4.2. En identificación por espectrometría de masas en tándem (MS/MS).

Los parámetros descritos para la identificación por huella peptídica se aplican también a las búsquedas MS/MS, pero teniendo en cuenta un parámetro adicional como es la tolerancia de masas de los iones fragmento y algunas consideraciones en otros parámetros (ver Figura 13).

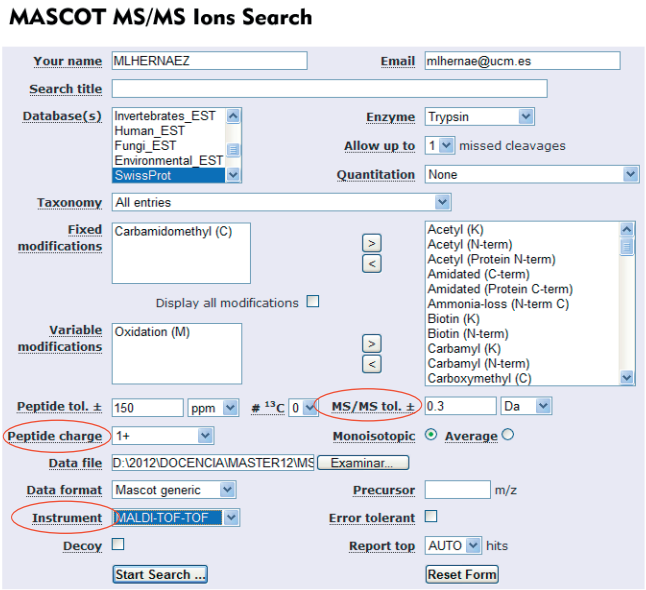


Figura 13. Interfaz del motor de búsqueda MASCOT para identificación de proteínas mediante espectrometría de masas en tándem. (MS/MS)

Enzima de digestión.

Hay ocasiones donde se hace imprescindible realizar la búsqueda sin especificar ninguna proteasa, como es el caso de identificaciones de péptidos endógenos. La desventaja es que esto hace que la base de datos con la que vamos a comparar en la búsqueda sea de 100 a 1000 veces más grande, por lo que se corre el riesgo de perder coincidencias más débiles y por tanto número de identificaciones.

Taxonomía.

No es aconsejable especificar la especie en una búsqueda de MS/MS, ya que si la proteína de la especie en estudio no está anotada en la base de datos, al tener información de la secuencia del péptido/s, se puede identificar esas secuencias en otras especies, donde las proteínas si estén anotadas en las bases de datos. Se trata de secuencias conservadas entre proteínas homólogas de diferentes organismos.

Tolerancia.

La mayoría de los motores de búsqueda admiten tolerancias de masa para los precursores y los fragmentos por separado. Los instrumentos híbridos, en particular, pueden tener muy diferentes precisiones para MS y MS/MS. La tolerancia de masa para los datos de MS/MS se suele expresar en Da o en mmu.

Tolerancias empleadas en análisis MS/MS:

- Equipos de baja resolución: precursor 300 ppm, iones fragmentos 0,6-0,8 Da
- Equipos de alta resolución: precursor 10-100 ppm, iones fragmentos 0,3-0,6 Da

Modificaciones.

En aquellas modificaciones que se relacionan con la función biológica en eucariotas, el lugar donde tiene lugar la modificación dentro de la secuencia proteica puede ser tan importante como la naturaleza de la modificación. Si un péptido contiene más de un residuo susceptible de sufrir una modificación y en algunos casos uno está modificado y en otros casos el otro, puede no ser suficiente con identificar la presencia de una modificación sino que podría ser interesante tener información acerca de la localización de la misma en un residuo particular, dentro del péptido. Es decir hay péptidos con diferentes sitios de modificación. En el resultado de la búsqueda, se muestra la disposición o localización con mayor puntuación y también se pueden mostrar las diferentes localizaciones dejando la interpretación al usuario. Muchas herramientas de software han sido desarrolladas para tratar de identificar y cuantificar las modificación en su localización específica, y a menudo con un enfoque particular en fosforilación [43,44,45,46].

Carga del péptido.

Los valores de masa experimentales que se envían para la identificación de las proteínas son valores de m/z. Para hacer la búsqueda hay que especificar cuál es la carga del ion precursor para poder calcular la masa molecular del precursor.

Para datos obtenidos por ionización tipo MALDI, la carga es +1, ya que los iones están mono-protonados (MH^+).

Para datos adquiridos con ionización por electrospray, se suele seleccionar carga +2, aunque dependiendo de los parámetros del método de adquisición se pueden seleccionar +2,+3 y+4.

Instrumento.

Si se especifica qué tipo de instrumento se ha utilizado para el análisis, el programa determina qué serie de iones fragmento serán considerados en las búsquedas de péptidos.

5. Informe de resultados e interpretación.

El resultado de una búsqueda cuyo objetivo es la identificación de proteínas contiene una gran cantidad de información que resumiremos a continuación. En primer lugar, es importante tener en cuenta que la proteína propuesta como resultado se ajuste a la información de la que disponíamos, tal como el organismo del que procede y la especificidad de la enzima.

5.1. Identificación por huella peptídica.

5.1.1. Péptidos coincidentes y cobertura de secuencia.

Para cada una de las proteínas candidatas, se muestra un listado de los péptidos identificados en base a las masas detectadas del espectro, y su localización dentro de la secuencia de aminoácidos de la proteína. En general, también se puede consultar la digestión teórica que ha generado el propio motor de búsqueda (Figura 14).

Se muestra también la cobertura de secuencia ("sequence coverage") alcanzada para esa proteína, es decir, el porcentaje de la misma que se ha conseguido identificar en función del número de péptidos detectados para cada proteína y la longitud de los mismos. Cuanto más elevado sea este valor mayor es la fiabilidad de la identificación. Se puede considerar como un valor óptimo aquel tal que los péptidos coincidentes cubran al menos un 15-20% de la secuencia de la proteína identificada.

5.1.2. Exactitud en la medida de la masa peptídica.

Se considera que una identificación es óptima cuando la exactitud en la medida de la masa de los péptidos detectados es similar para todos ellos, es decir, la diferencia entre las masas medidas (experimentales) y las calculadas debe ser de la misma magnitud para todos los péptidos identificados.

5.1.3. Concordancia con datos conocidos.

La masa nominal (Mw) y el valor de punto isoeléctrico (pI) teóricos de la proteína identificada deben estar en conformidad con los datos experimentales obtenidos, por ejemplo, de una separación electroforética en gel si se dispone de ellos. Si éste no fuera el caso, se debe de tener en cuenta la presencia de posibles fragmentos de la proteína o la formación de aductos de la misma.

5.1.4. Puntuación.

En el informe de resultados, hay un listado de proteínas candidatas cuyos péptidos en principio habrían generado los valores de masa detectados en el espectro. Dichas proteínas se presentan con diferentes puntuaciones. Algunos motores de búsqueda basan estas puntuaciones en la probabilidad de que las coincidencias entre las masas experimentales y las teóricas que se comparan sean significativas y no ocurran por un evento al azar. Se consideran como candidatos identificados, aquellas entradas que tengan una puntuación por encima de un umbral, que supongan un valor p o p-value<0,05, es decir hay un 95% de probabilidad de que no sea un evento al azar. La puntuación de cada péptido se acompaña de un valor esperado ("expectation value") que define el número de coincidencias con igual o mejor puntación que se esperaría que ocurriera por casualidad en la base de datos utilizada. Para una puntación que está en el umbral de p<0,05, el valor esperado también es 0,05. A mayor puntuación, menor valor esperado.

Por tanto, la puntuación puede servir como primer indicador para saber si la identificación es positiva. En el caso de que ésta sea significativa, otro elemento a valorar es la diferencia entre las

Figura 14. Interfaz del motor de búsqueda MASCOT de resultados de la identificación de proteínas mediante huella peptídica.

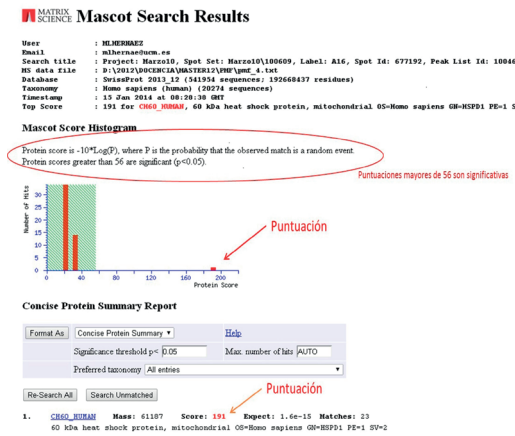


Figura 14 A.

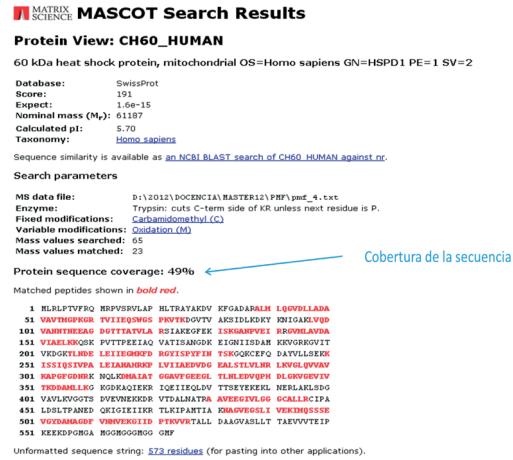


Figura 14 B.

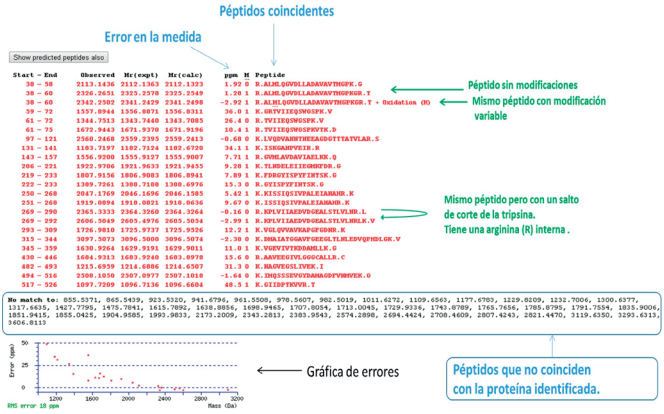


Figura 14 C.

puntuaciones de los distintos candidatos ("delta score"). Si esta diferencia es relevante, la probabilidad de que el emparejamiento con el candidato de mayor puntuación sea por el azar, será más baja que en el caso de que las puntuaciones entre candidatos sean similares (ver Figura 15).

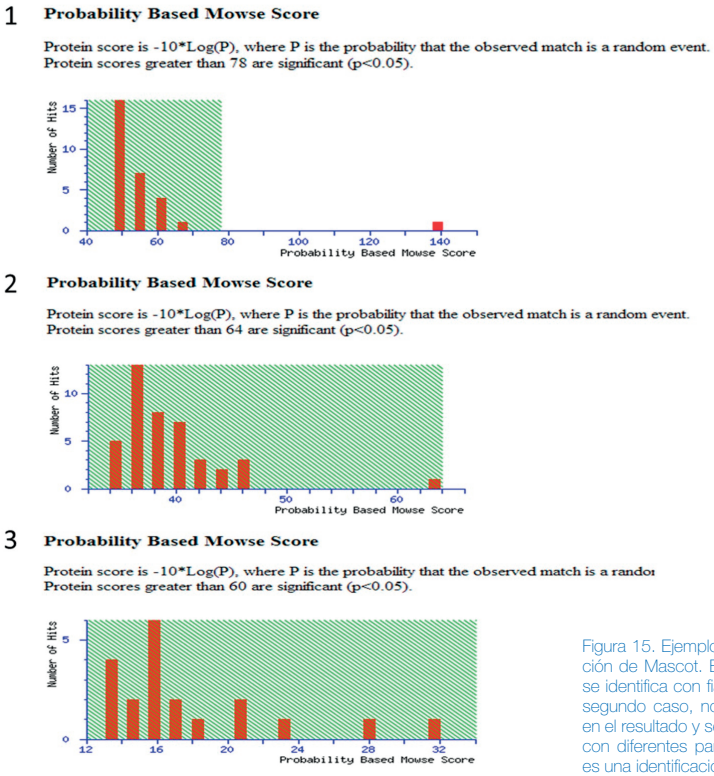


Figura 15. Ejemplo de histogramas de puntuación de Mascot. En el primer caso la proteína se identifica con fiabilidad, pvalue< 0,05. En el segundo caso, no tenemos toda la confianza en el resultado y se debería repetir la búsqueda con diferentes parámetros. En el tercer caso, es una identificación al azar.

5.1.5. Cálculo de la puntuación (score).

Cada motor de búsqueda establece su propio modelo para el cálculo de la puntuación. Actualmente, se emplean diferentes aproximaciones. En el caso de Mascot, se realizan cálculos probabilísticos que atienden a la distribución de las relaciones m/z entre péptidos y proteínas, basado en el algoritmo de MOWSE mientras que ProFound se basa en la aplicación de algoritmos Bayesianos para el análisis de los emparejamientos que se producen [26].

Generalizando podemos decir que el objetivo de un esquema de puntuación es, por un lado, dar la puntuación mayor a la secuencia correcta de la base de datos (aquella que inequívocamente identifica la proteína); por otro lado, también obtener una mayor diferencia entre las puntuaciones de la secuencia correcta y la errónea o al azar. La puntuación está basada en la similitud entre las señales experimentales y los valores teóricos de masa que se obtienen al simular un experimento idéntico con la secuencias de las bases de datos tal que proporcionen las huellas peptídicas teóricas. La concordancia o similitud depende de un conjunto de componentes diferentes y el reto es determinar estas componentes, cómo se deberían medir y qué peso estadístico y relativo tienen entre ellas. Hay que tener en cuenta cómo algunos factores influyen en la puntuación final o “score”:

- La puntuación disminuye al aumentar el tamaño de la base de datos.
- La puntuación aumenta al aumentar el número de masas coincidentes.
- La puntuación disminuye con el aumento neto del número de péptidos teóricos con masas similares a la masa experimental.

5.1.6. Picos no asignados.

Las señales no asignadas en el proceso de identificación por huella peptídica, es decir la diferen-

Figura 16. Interfaz del motor de búsqueda MASCOT de resultados de la identificación de proteínas mediante espectrometría de masas en tándem (MS/MS).

Figura 16 A.

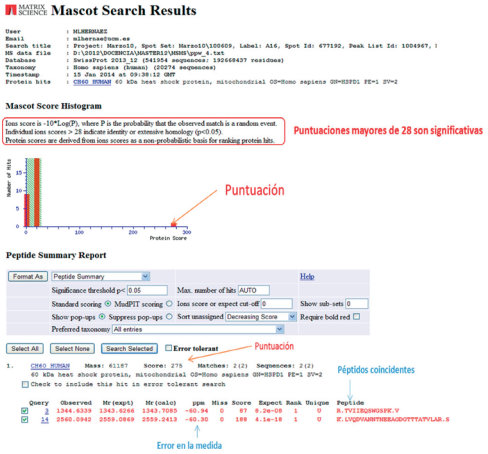
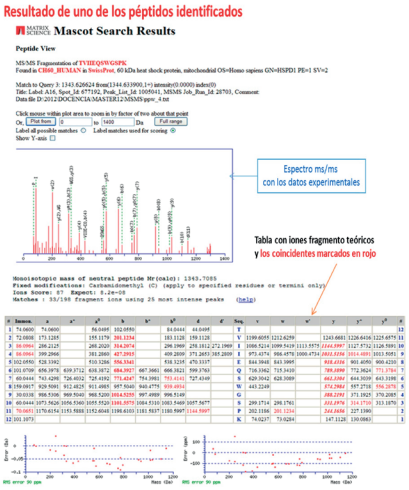


Figura 16 B.

Figura 16 C.



cia entre el número de valores de masa buscado y el número de los valores de masa asignados debe ser tan pequeña como sea posible. Si la diferencia es alta, estos picos no asignados pueden proceder de diferentes fuentes: péptidos de otras proteínas presentes en la muestra porque puede haber mezcla (aunque la proteína proceda se una mancha proteica o spot de un gel bidimensional, a veces puede haber más de una proteína), péptidos de proteínas contaminantes, como las queratinas, de los calibrantes internos, como la tripsina, de la matriz empleada para la ionización en el caso de análisis por MALDI-TOF, péptidos procedentes de una digestión parcial o fragmentos producidos durante el análisis, péptidos con modificaciones post-traduccionales o químicas y péptidos procedentes de una proteína altamente homóloga con la identificada. La identificación de componentes proteicos adicionales se puede llevar a cabo mediante una nueva búsqueda, solo con los datos de los picos no asignados por huella peptídica y/o por MS/MS.

La opción de realizar una segunda búsqueda, es una herramienta útil para reevaluar masas sin asignar.

Si no hay éxito en la identificación de la proteína por huella peptídica puede ser debido a diferentes motivos. Uno de ellos puede ser que la proteína no esté anotada en las bases de datos, que proceda de un organismo cuyo genoma no esté secuenciado. Otra posibilidad es que no

se haya identificado un número suficiente de péptidos para tener una coincidencia significativa para ninguna proteína por el tamaño de la proteína o porque la digestión no haya sido efectiva.

5.2. Identificación por MS/MS.

La ventaja de la espectrometría de masas en tándem respecto a la identificación de proteínas por huella peptídica es principalmente que tenemos información de la secuencia del péptido. Se identifican secuencias individuales de péptidos, y luego el conjunto de secuencias peptídicas se usa para deducir que proteínas están presentes en la muestra analizada. A menos que un péptido sea único de una proteína particular, puede haber alguna ambigüedad en asignar cual es la proteína de la que procede.

La interfaz del motor de búsqueda MASCOT de resultados de la identificación de proteínas mediante espectrometría de masas en tándem. (MS/MS) es similar a la de la huella peptídica y su interpretación también es sencilla y clara. Los resultados se muestran en diferentes formatos y hay enlaces para ver diferente información, tanto de los péptidos como de las proteínas identificadas (ver Figura 16).

Hay que recordar que en Mascot la puntuación se basa en el cálculo de probabilidades de que las coincidencias entre los datos experimentales y los teóricos sean al azar, y en el caso de MS/MS, hay una puntuación para cada péptido (puntuación individual del ion) y una puntuación para cada proteína identificada.

En los resultados se muestra una tabla, donde se resalta en rojo las coincidencias entre los datos experimentales y los iones fragmentos teóricos. Si se especifica el instrumento, se considera la serie de iones más adecuados para ese tipo de equipo (ver Figura 17).

Figura 17. Ejemplo de selección del instrumento y tabla de péptidos teóricos y coincidentes.

Tabla con iones fragmento teóricos y los coincidentes marcados en rojo

Si no se selecciona el instrumento, compara con la series a, b e y

Monoisotopic mass of neutral peptide Mr(calc): 1343.7085
Fixed modifications: Carbamidomethyl (C) (apply to specified residues or termini only)
Ions Score: 72 Expect: 2.9e-06
Matches : 19/56 fragment ions using 25 most intense peaks (help)

#	a	a*	b	b*	Seq.	y	y*	#
1	74.0600		102.0550		T			12
2	173.1285		**201.1234**		V	1243.6681	1226.6416	11
3	286.2125		**314.2074**		I	***1144.5997***	1127.5732	10
4	399.2966		**427.2915**		I	***1031.5156***	1014.4891	9
5	528.3392		**556.3341**		E	***918.4316***	901.4050	8
6	656.3978	639.3712	**684.3927**	667.3661	Q	***789.3890***	772.3624	7
7	743.4298	726.4032	**771.4247**	754.3981	S	***661.3304***	644.3039	6
8	929.5091	912.4825	957.5040	940.4775	W	***574.2984***	557.2718	5
9	986.5306	969.5040	**1014.5255**	997.4989	G	***388.2191***	371.1925	4
10	1073.5626	1056.5360	**1101.5575**	1084.5310	S	***331.1976***	314.1710	3
11	1170.6154	1153.5888	1198.6103	1181.5837	P	***244.1656***	227.1390	2
12					K	147.1128	130.0863	1

Si se especifica el instrumento, MALDI-TOF/TOF además compara con la series d y w.

Monoisotopic mass of neutral peptide Mr(calc): 1343.7085
Fixed modifications: Carbamidomethyl (C) (apply to specified residues or termini only)
Ions Score: 87 Expect: 8.2e-08
Matches : 33/198 fragment ions using 25 most intense peaks (help)

#	Immon.	a	a*	a^0	b	b*	b^0	d	d'	Seq.	v	w	w'	y	y*	y^0	#
1	74.0600	74.0600		56.0495	102.0550		84.0444	44.0495		T							12
2	72.0808	173.1285		155.1179	**201.1234**		183.1128	159.1128		V	1199.6055	1212.6259		1243.6681	1226.6416	1225.6575	11
3	86.0964	286.2125		268.2020	**314.2074**		296.1969	258.1812	272.1969	I	1086.5214	1099.5419	1113.5575	***1144.5997***	1127.5732	1126.5891	10
4	86.0964	399.2966		381.2860	**427.2915**		409.2809	371.2653	385.2809	I	973.4374	986.4578	1000.4734	***1031.5156***	1014.4891	1013.5051	9
5	102.0550	528.3392		510.3286	**556.3341**		538.3235	470.3337		E	844.3948	843.3995		***918.4316***	901.4050	900.4210	8
6	101.0709	656.3978	639.3712	638.3872	**684.3927**	667.3661	666.3821	599.3763		Q	716.3362	715.3410		***789.3890***	772.3624	771.3784	7
7	60.0444	743.4298	726.4032	725.4192	**771.4247**	754.3981	753.4141	727.4349		S	629.3042	628.3089		***661.3304***	644.3039	643.3198	6
8	159.0917	929.5091	912.4825	911.4985	957.5040	940.4775	939.4934			W	443.2249			***574.2984***	557.2718	556.2878	5
9	30.0338	986.5306	969.5040	968.5200	**1014.5255**	997.4989	996.5149			G				***388.2191***	371.1925	370.2085	4
10	60.0444	1073.5626	1056.5360	1055.5520	**1101.5575**	1084.5310	1083.5469	1057.5677		S	299.1714	298.1761		***331.1976***	314.1710	313.1870	3
11	70.0651	1170.6154	1153.5888	1152.6048	1198.6103	1181.5837	1180.5997	1144.5997		P	202.1186	201.1234		***244.1656***	227.1390		2
12	101.1073									K	74.0237	73.0284		147.1128	130.0863		1

5.2.1. X!Tandem: Puntuación en la identificación de péptidos.

A diferencia de Mascot, X!Tandem no dispone de una interpretación rápida y sencilla para in-

terpretar si la identificación de un determinado péptido es válida. En este sentido, X!Tandem se apoya en dos cálculos estadísticos que proporcionan el nivel de confianza que se asocia a un determinado resultado. La primera, p-value, pondera que un emparejamiento entre la secuencia de un determinado péptido (P) con un espectro (S) sea fruto del azar. Dicho de otra forma, el valor p (p-value) fija la probabilidad de que la secuencia de un péptido totalmente aleatoria (*random peptide*) obtenga la misma puntuación, o mejor, que la obtenida entre el emparejamiento de P con S. La segunda, e-value, fija el valor esperado o el número de péptidos que por el azar obtuviesen una puntuación igual, o mejor, que el péptido P en su emparejamiento con el espectro S. En este caso se considera una base de datos con el mismo tamaño que la empleada en la fase de identificación. Ambos valores están relacionados, ya que e-value permite corregir el número de falsos positivos que obtendrían un determinado p-value cuando se realizan búsquedas en bases de datos grandes. Fruto de esta relación, los p-value de péptidos aleatorios con un determinado espectro S se utilizan para calibrar el rango de puntuaciones esperadas (e-value) para el espectro S, incluyendo el emparejamiento real con el péptido P. Sin embargo, si el p-value asociado a P no es el mejor de todo el rango de puntuaciones considerado, la seguridad para considerar dicho emparejamiento como el correcto puede ser bastante pobre.

5.2.2. X!Tandem: Interpretación de resultados.

Una de las desventajas de X!tandem es que los resultados no se ofrecen de una manera visual como si lo hace Mascot. Estos son integrados en un fichero XML cuya descripción está disponible en http://www.thegpm.org/docs/X_series_output_form.pdf. A pesar de que se trata de un formato que proporciona una consistencia tanto sintáctica como semántica, no resulta cómodo para el usuario final cuyo principal interés es saber que proteínas y péptidos se han detectado en un determinado experimento.

Con el fin de proporcionar al usuario una visión más enfocada a este primer objetivo está la herramienta PeptideShaker (https://code.google.com/p/peptide-shaker/) (Figura 18). Entre las ventajas que presenta este software está la de representar gráficamente la salida de búsquedas realizadas tanto por X!Tandem, Mascot y OMSSA. Las salidas, cada una en su formato nativo, son procesadas para poder generar la presentación visual y conjunta de los resultados proporcionados por estos motores de búsqueda.

Figura 18. Salida típica del software PeptideShaker. Presenta de forma agrupada los cuatro elementos básicos que intervienen en la identificación: Proteínas (en la parte superior), Péptidos (en la parte intermedia izquierda), PSM (pare inferior izquierda) y los espectros MS/MS (parte inferior derecha).

PeptideShaker puede ser ejecutado tanto de forma independiente como directamente desde el formulario de Search GUI tras introducir los parámetros correspondientes. En ambos casos habrá que seleccionar el FASTA que contiene la base de datos de secuencias empleada en la identificación, y sólo en el caso de la ejecución independiente se tendrán que introducir todos los ficheros, tanto de espectrometría de masas como de resultados, que queramos analizar.

Otras opciones relativas a la anotación de las proteínas, la visualización de su estructura terciaria o la aparición de modificaciones postraduccionales están disponibles para una mejor interpretación de los resultados obtenidos.

5.2.3. Espectros MS/MS no asignados a una secuencia en las bases de datos.

Generalmente, en experimentos de identificación masiva, la mayoría de los espectros adquiridos no se corresponden con una secuencia presente en la base de datos.

Por supuesto, cuando se trabaja con muestras procedentes de organismos que no están bien representados en cualquier base de datos de secuencias, algunas coincidencias se pueden esperar, ya que no se requieren coincidencias exactas a lo largo de toda la proteína sino en péptidos, que pueden corresponder a secuencias conservadas entre las especies. Sin embargo, incluso para organismos totalmente secuenciados, polimorfismos de la secuencia pueden todavía presentar dificultades en su identificación.

Las modificaciones de los péptidos pueden ocurrir *in vivo* (por ejemplo, fosforilación) o *ex vivo* (por ejemplo, la oxidación de metionina), pero en cualquier caso la masa del precursor observado será diferente de lo que se calcula a partir de una base de datos de secuencia. Dado que el peso molecular del péptido se utiliza como el filtro para derivar las secuencias de los candidatos, un peso molecular no correcto sólo proporcionará secuencias no asignadas con alta fiabilidad.

Sin embargo, a menos que al programa de búsqueda de la base de datos MS/MS se le informe de la modificación específica en estudio, no se producirá la identificación.

Otra razón por la que un espectro podría no coincidir con una entrada de secuencia es que el ion precursor no se deriva de un péptido intacto. Por ejemplo, las búsquedas de bases de datos de péptidos trípticos se suelen hacer con la restricción de que todas las secuencias de candidatos no sólo coinciden con la masa del péptido observada, sino que también se derivan de los cortes por el C -terminal de la lisina y arginina. Si es un fragmento del péptido por fragmentación de los iones de péptidos intactos en la fuente de iones antes de la primera etapa de análisis de masas o producido por otra actividad proteolítica que estaba presente en la muestra antes o después de aislamiento de las proteínas, no se identificará correctamente.

Un problema común es que las masas de péptidos intactos no se calculen correctamente. Si el estado de carga del ion precursor no se determina correctamente o si un pico de isótopos que contiene ^{13}C está erróneamente elegido como el pico ^{12}C monoisotópico, se calculará la masa del péptido intacto incorrectamente.

La explicación más probable para que un espectro no coincida con una entrada en la base de datos de secuencia es que es de calidad insuficiente o el espectro de MS/MS se obtuvo de una contaminación no peptídica. Si los datos se recogen usando disolventes sucios o una columna contaminada, habrá un número de espectros adquiridos para iones precursores no peptídicos. Además, los espectrómetros de masas a veces adquieren el espectro con una intensidad del ion precursor muy baja y se fragmentan picos provenientes del ruido. En las adquisiciones de espectros dependientes de datos, puede ocurrir que se fragmente el precursor en un tiempo temprano de la elución cuando la intensidad del ion precursor también sea baja. Todo esto conduce a la adquisición de espectros de mala calidad que impide que los motores de búsqueda consigan coincidencias con los espectros teóricos [47].

La búsqueda de múltiples pases es la forma más eficiente para encontrar modificaciones menos comunes, incluyendo mutaciones puntuales en la secuencia primaria, y péptidos no específicos. La primera búsqueda es una simple búsqueda en toda la base de datos con condiciones de búsqueda muy restrictivas. Las proteínas candidatas que se encuentran en la primera búsqueda se seleccionan entonces para una búsqueda exhaustiva en el segundo pase. Durante esta segunda búsqueda una larga lista de posibles modificaciones se prueba en serie y las condiciones son más relajadas como la especificidad de la enzima para permitir el encuentro de coincidencias con péptidos no específicos. Debido a que sólo unas cuantas entradas candidatas se han seleccionado en el primer pase, el tiempo de búsqueda no es un problema. Es difícil de aplicar cualquier tipo de tratamiento estadístico de los resultados, ya que las entradas que se buscan han sido pre-seleccionadas.

El uso de las secuencias de péptidos para deducir que las proteínas estaban presentes en la muestra original es sorprendentemente difícil, en experimentos de alto rendimiento o identificación masiva, porque muchas de las secuencias de péptidos en el resultado de búsqueda puede ser asignado a más de una proteína. El principio básico de la inferencia o deducción de proteínas es la creación de una lista mínima de las presuntas proteínas [48].

Aunque en otros capítulos se trata en mayor profundidad la identificación masiva de proteínas o experimentos a gran escala, vamos a exponer algunas diferencias fundamentales con la identificación descrita en este capítulo. Principalmente, cuando se manejan datos correspondientes a cientos o miles de espectros, es necesario y actualmente un requisito indispensable para su publicación, validar la fiabilidad del resultado y estimar el error cometido. Para esta tarea se utilizan bases de datos denominadas "decoy", que consisten en bases de datos que contienen secuencias que no corresponden a priori con proteínas o secuencias al azar, de manera que tenga el mismo tamaño que la base de datos de proteínas contra la que se va a realizar la búsqueda, denominada "target". Se realiza la búsqueda en las dos bases de datos, "decoy" y "target", o en una única que contenga ambas bases de datos. Lo esperado en los resultados de la búsqueda es que no se identifique ningún péptido de la base de datos decoy, por lo que si se encuentra alguna coincidencia se considera que es incorrecta y se puede estimar la tasa de falsos positivos (FDR o "False Discovery Rate"). Se puede estimar esta tasa tanto para proteínas como para péptidos. Los valores que se utilizan son FDR<0,05 para condiciones más relajadas y FDR <0,01 para ser más estrictos con el resultado. En Mascot por ejemplo, se muestran dos resultados de FDR según el sistema de puntuación que sigue, uno para identidad y otro más relajado para homología.

	SwissProt	Decoy	False discovery rate
Peptide matches above identity threshold	3288	46	1.40 %
Peptide matches above homology or identity threshold	3305	47	1.42 %

Select Summary Report

Format As	Select Summary (protein hits)		Help
	Significance threshold p< 0.05	Max. number of hits AUTO	Show Percolator scores ☐
	Standard scoring ○ MudPIT scoring ◉	Ions score or expect cut-off 0	Show sub-sets 0
	Show pop-ups ◉ Suppress pop-ups ○		Require bold red ☐

	SwissProt	Decoy	False discovery rate
Peptide matches above identity threshold	3633	34	0.94 %

Select Summary Report (scores adjusted by Percolator)

Format As	Select Summary (protein hits)		Help
	Significance threshold p< 0.05	Max. number of hits AUTO	Show Percolator scores ☑
	Standard scoring ○ MudPIT scoring ◉	Ions score or expect cut-off 0	Show sub-sets 0
	Show pop-ups ◉ Suppress pop-ups ○		Require bold red ☐

Figura 19. Ejemplo de re-estimación de la tasa de falsos positivos FDR, al aplicar el filtro percolator en MASCOT.

Además hay algunos motores de búsqueda, como MASCOT, que permiten filtrar los resultados una vez obtenidos, por ejemplo mediante la herramienta de "percolator", que consiste en supervisar los resultados y discriminar mejor entre identificaciones correctas e incorrectas, calculando una probabilidad de error posterior a la búsqueda y así se corrige el FDR (ver Figura 19).

Para confirmar los resultados obtenidos, es fundamental aplicar alguna de estas herramientas estadísticas en experimentos con un gran número de datos.

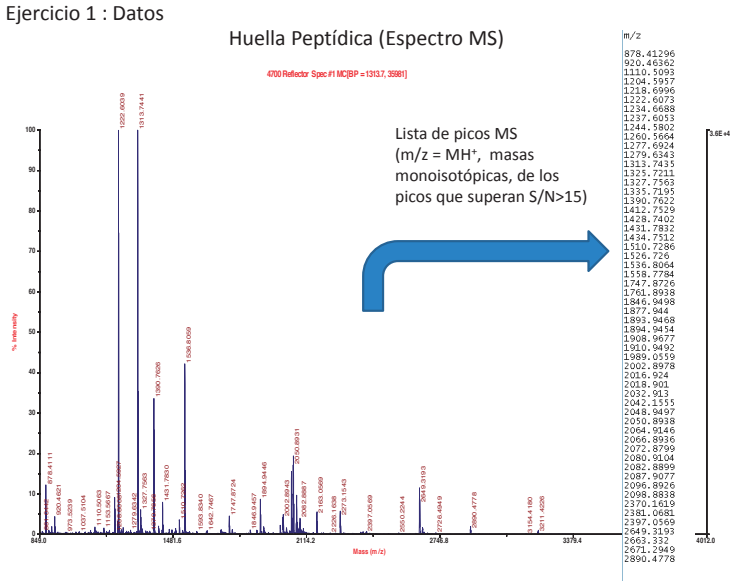
EJERCICIOS EJEMPLO.

EJERCICIO 1. (PMF)

Un extracto proteico de la proteobacteria, *Acidiphilium cryptum*, fue separado en un gel bidimensional y teñido con azul de Coomassie. Uno de los spots, con unos valores experimentales aproximados de peso molecular de 18.000 Da y punto isoeléctrico de 7,6 fue recortado del gel, reducido con DTT y alquilado con iodoacetamida, antes de su digestión con tripsina. El digerido fue analizado en modo Reflector + en un MALDI-TOF y procesado con calibración interna utilizando los péptidos de la tripsina. El fichero de datos del spot digerido contiene una lista de las masas monoisotópicas (MH+) de los péptidos detectados que superan una S/N = 15.

¿Cuáles son los parámetros de búsqueda adecuados: Base de Datos, taxonomía, tolerancia, Nº. de puntos de corte perdidos, etc., para identificar la proteína?

Resultados: La búsqueda se realiza utilizando el motor de búsqueda MASCOT (http://www.matrixscience.com/), con datos de PMF (peptide mass fingerprint) (ver figura "Ejercicio 1: Datos"). Los parámetros adecuados se muestran en la figura "Ejercicio 1: Resultados1" y son:



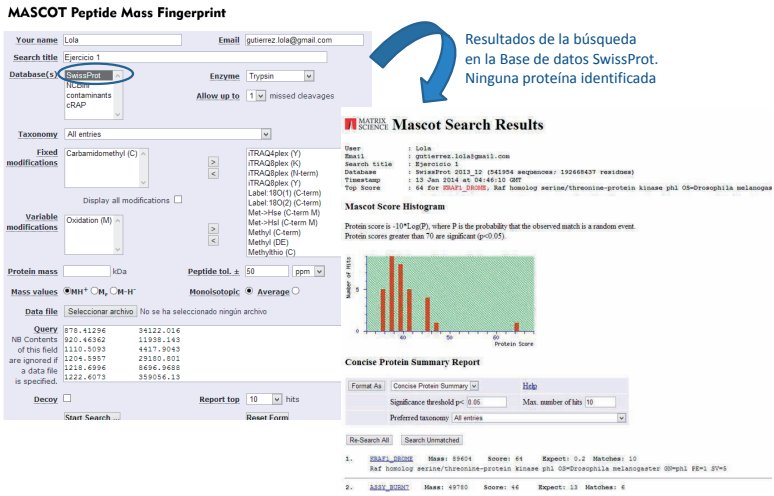
En el paso 1 se selecciona SwissProt como base de datos; sin restricción taxonómica; la enzima utilizada es tripsina, permitiendo hasta 1 salto de corte; la modificación fija de los péptidos, producida por tratamiento con iodacetamida, es la carbamidometilación de las cisteínas (+57 Da) y la modificación variable es la oxidación de las metioninas, dado que durante la preparación

de la muestra las condiciones son oxidantes. La lista de picos incluye valores monoisotópicos de MH+ (carga 1+); la tolerancia para la masa de los péptidos es +/- 50 ppm, puesto que el espectro está calibrado internamente, se podría incluso bajar hasta 10 ppm.

Cuando se realiza la búsqueda en la base de datos SwissProt como primera opción, no se obtiene ningún resultado concluyente (ver “Ejercicio 1: Resultados1”), ya que ninguna proteína supera la puntuación de 70 establecida por MASCOT para que la identificación no se deba al azar con una probabilidad p< 0,05. En esta base de datos hay 3.521 entradas para *Acidiphillium cryptum*, mientras que en NCBI hay 9.607. La proteína por tanto no está anotada en SwissProt.

Ejercicio 1: Resultados1

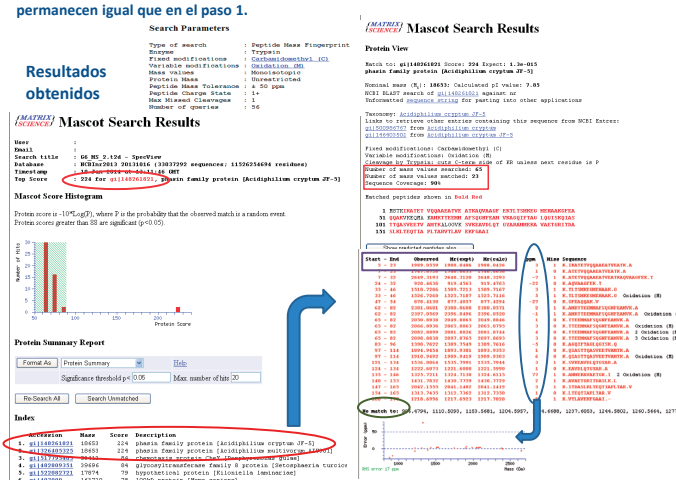
PASO 1: parámetros de búsqueda



El siguiente paso (paso 2), antes de cambiar algún otro parámetro, es repetir la búsqueda utilizando la base de datos NCBInr, que contiene mucha más información aunque no toda ella esté revisada. En este caso sí se obtiene una identificación significativa estadísticamente (ver “Ejercicio 1: Resultados2”).

Ejercicio 1: Resultados2

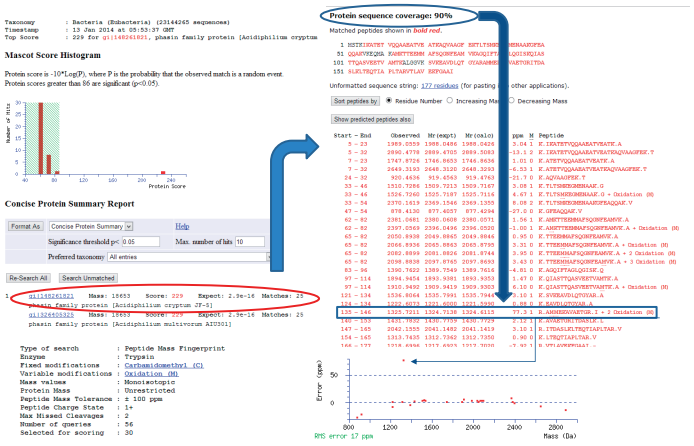
PASO 2: Cambio en los parámetros de búsqueda a la base de datos NCBInr, el resto de parámetros permanecen igual que en el paso 1.



Podemos variar el resto de los parámetros para ver cómo afectan a la puntuación y la cobertura de secuencia de la proteína identificada. En este caso, ni la restricción taxonómica a bacterias, ni el aumento a 2 saltos de corte, mejoran de forma importante ni la puntuación ni la cobertura de secuencia. El aumento en la tolerancia de masa hasta 100ppm, disminuye la puntación y aumenta ligeramente la cobertura hasta el 90%, ya que se incluye en la lista de péptidos coincidentes un nuevo péptido que además de presentar dos modificaciones variables (oxidación de metionina) tiene un error de masa muy por encima del resto de péptidos y claramente discordante para su masa, como se aprecia en el gráfico de distribución del error (ver "Ejercicio 1: resultados3"), por lo que muy probablemente se trata de una falsa coincidencia. De manera que los parámetros de búsqueda más restrictivos, los usados en el segundo paso, resultan los más adecuados por conseguir identificar de forma fiable y concluyente la proteína.

Ejercicio 1: Resultados3

PASO 3: Cambio en los parámetros de búsqueda a Taxonomía restringida a bacteria, nº de misscleavage = y tolerancia de péptido de 100 ppm, el resto permanecen igual que en el paso 1.



¿Qué proteína/s se identifica?, ¿Qué porcentaje de secuencia cubren los péptidos coincidentes?

La proteína identificada es gil148261821, phasin family protein [*Acidiphilium cryptum JF-5*], los péptidos coincidentes, 23 de las 65 masas enviadas, cubren el **90%** de la secuencia de la esta proteína. Como se muestra en la figura "Ejercicio 1: Resultados2.

¿Es un resultado concluyente? ¿Por qué?

Es un resultado concluyente, ya que sin utilizar restricción taxonómica la proteína identificada pertenece al organismo en estudio y sus valores teóricos de Mr y pI se corresponden con los valores experimentales observados en el gel 2D. Además la proteína identificada tiene asignada una puntuación muy elevada (224) y un valor esperado o e-value muy bajo (2.3e-21), que indican que las coincidencias entre las masas medidas y los valores teóricos no se deben al azar.

La mayoría de los picos intensos del espectro se encuentran asignados a esta proteína y prácticamente cubren toda su secuencia.

La búsqueda de los picos no asignados a esta proteína (search unmatched) desde la página de resultados, no identifican ninguna proteína de forma fiable.

¿Se obtiene el mismo resultado utilizando otros motores de búsqueda?

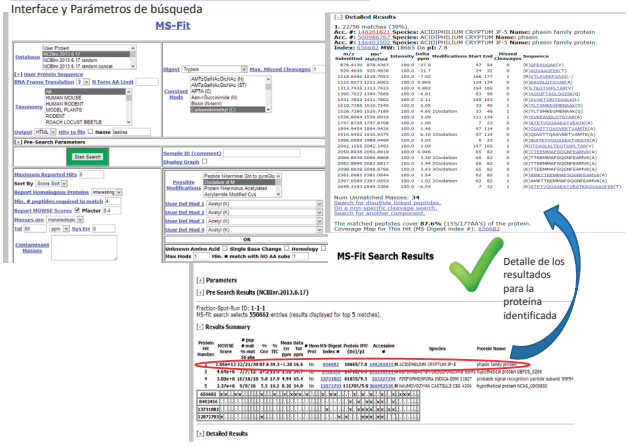
Si, utilizando MSFIT se obtienen los mismos resultados (ver figura "Ejercicio 1: Resultados4") identificándose la misma proteína que en MASCOT, lo que hace el resultado aún más fiable.

Ejercicio 1: Resultados4

PASO 4: Búsqueda de los datos MS en otro motor de búsqueda. MS-Fit

http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msfitstandard

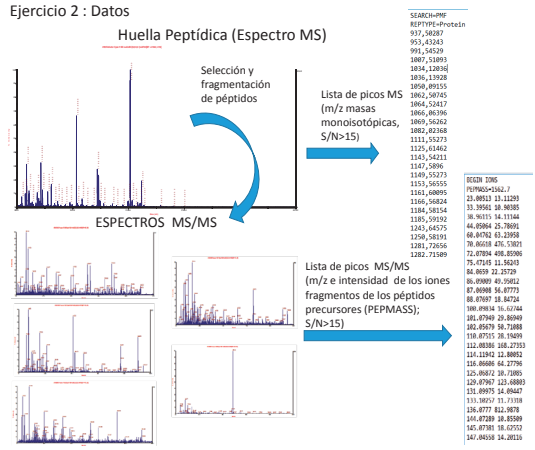
Interface y Parámetros de búsqueda



EJERCICIO 2. (PMF y MS/MS)

Se separa un extracto proteico del hongo oportunista *Candida albicans* en un gel bidimensional, un spot es cortado, reducido, alquilado (con iodacetamida) y digerido con tripsina. El digerido se analiza por MALDI-TOF/TOF en modo reflector positivo obteniendo un espectro MS (huella peptídica) internamente calibrado, del que se seleccionan 5 de los péptidos más intensos y se fragmentan utilizando CID. Se dispone de un fichero de datos de la huella peptídica que contiene una lista de los valores MH+ de los péptidos detectados con una S/N > 15 y un fichero MS/MS, que presenta la masa del péptido precursor y una lista de los MH+ de los fragmentos generados y las intensidades de cada uno de los péptidos fragmentados (espectros MSMS).

Ejercicio 2 : Datos

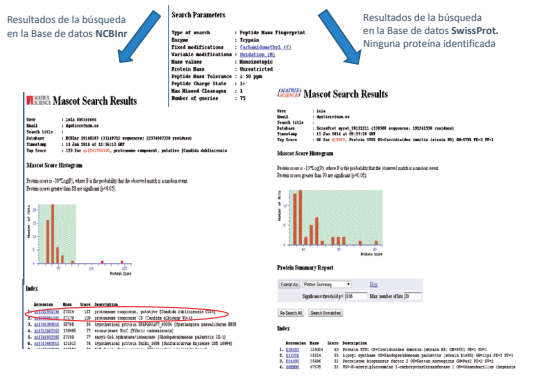


Utilizando los datos necesarios de los que se suministran, identificar la/s proteína/s presentes en el spot.

Utilizando los datos de huella peptídica (MS), y los parámetros de búsqueda que se muestran en la figura "Ejercicio 2: Resultados1" contra la base de datos SwissProt y NCBInr, sólo se obtienen resultados significativos para esta última, en la que se identifica una proteína de *Candida albicans*. En la figura "Ejercicio 2: Resultados2" se muestra el detalle de los péptidos coincidentes (13 de 75) y la cobertura de éstos (38%) para la proteína gil238881595 proteasome component C5 [*Candida albicans*]. Sin embargo, algunos de los picos más intensos que aparecen en el

Ejercicio 2: Resultados1

PASO 1: Identificación por Huella peptídica usando MASCOT



espectro MS, no se encuentran asignados a esta proteína, incluso cuando aumentamos la tolerancia para el péptido a 200 ppm, 2 saltos de corte para la tripsina e incluímos restricción taxonómica a hongos; al tratarse de proteínas de *Candida*, se sigue identificando la misma proteína con el mismo número de péptidos coincidentes.

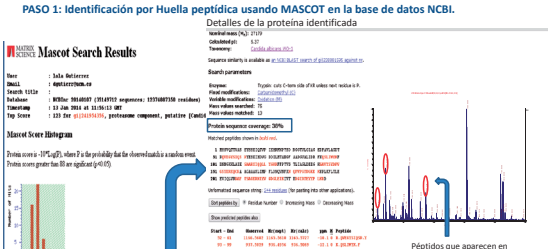
Puesto que disponemos de datos MS/MS, de estos péptidos y de otros 2 más que sí aparecen asignados a la proteína, se emplearán para confirmar la secuencia de estos últimos y tratar de identificar los péptidos que no pertenecen a esta proteína.

Se utiliza la forma de búsqueda Sequence Query de Mascot, donde se pueden incorporar los datos MS/MS para asignarles una secuencia e incluso se puede realizar la búsqueda combinada de datos MS y MS/MS si se dispone de una licencia de MASCOT. En la figura "Ejercicio 2: Resultados3" se muestran los parámetros de búsqueda, que son los mismos que para la huella peptídica a excepción de la tolerancia de masa para el péptido precursor (peptide tol.) que ahora puede ser un poco más laxa (por ej. de 100 ppm) dado que también debe haber coincidencia en la masa de los iones fragmento al hacer la comparación y la tolerancia para éstos es de 0,3-0,6 Da (menor que la resolución isotópica). Los resultados obtenidos en la base de datos SwissProt también aparecen en esta figura e identifican con un único péptido, de los 5 enviados, una proteína de un organismo que no es *C. albicans*, pero también se trata de un hongo y la proteína es de peso molecular similar a la identificada en NCBInr.

Los resultados de esta misma búsqueda seleccionando la base de datos de NCBInr, se muestran en la figura "Ejercicio 2: Resultados4" e

Ejercicio 2: Resultados2

PASO 1: Identificación por Huella peptídica usando MASCOT en la base de datos NCBI.

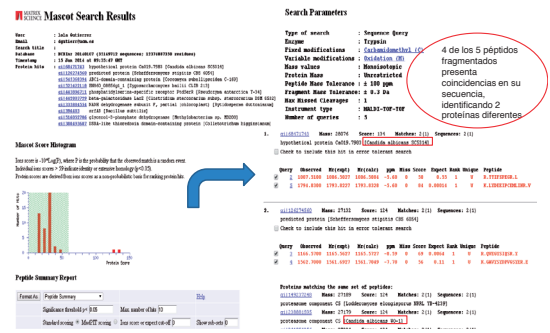


Ejercicio 2: Resultados3

PASO 2: Identificación por fragmentación de péptidos usando MASCOT y la base de datos SwissProt

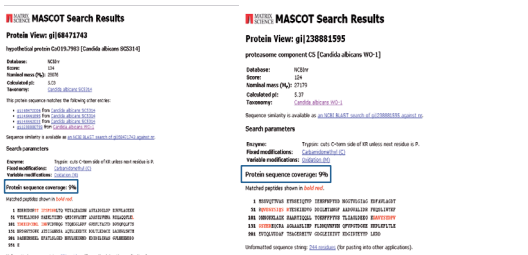
Ejercicio 2: Resultados4

PASO 2: Identificación por fragmentación de péptidos usando MASCOT y la base de datos NCBInr



Ejercicio 2: Resultados5

PASO 3: Cobertura de secuencia de las proteínas identificadas por fragmentación de péptidos en la base de datos NCBInr



identifican dos proteínas de *Candida albicans*, de pesos moleculares similares, cada una de ellas con 2 péptidos asignados.

Ejercicio 2: Resultados6

PASO 4: Cambio de los parámetros de búsqueda (Tolerancia de masa para los fragmentos) en la identificación por MSMS en la base de NCBI

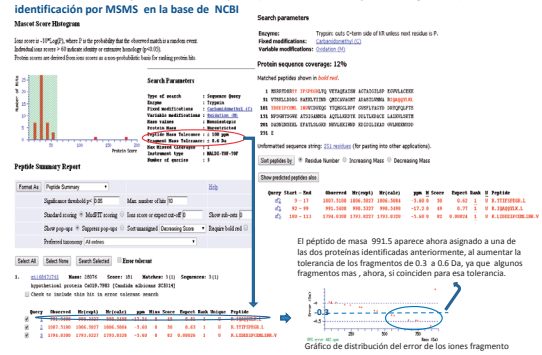


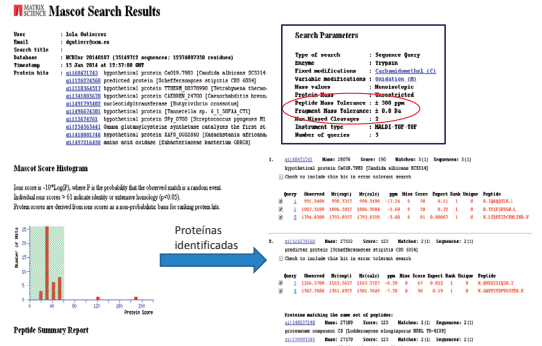
Gráfico de distribución del error de los iones fragmento

¿Tiene algúna influencia en el resultado la base de datos utilizada?

Candida albicans no está anotado completamente en la base de datos SwissProt y hay muchas más proteínas anotadas en NCBInr, como se ha comentado en el apartado anterior. Sólo en NCBInr es posible identificar por huella peptídica de forma estadísticamente significativa al menos una proteína y dos utilizando los datos de espectrometría de masas en tándem. Así que esta base de datos resulta más adecuada para identificar las proteínas en este estudio.

Ejercicio 2: Resultados7

PASO 4: Cambio de los parámetros de búsqueda en la identificación por MSMS en la base de NCBInr



En caso de identificar más de una proteína, ¿Cómo de fiable es el resultado para cada una de ellas. ¿Se puede validar su presencia en el spot o se trata de una contaminación?

Se identifican 2 proteínas, ambas con dos péptidos; al menos uno de ellos supera la puntuación, y la del otro está muy próxima para alcanzar identidad u homología significativa ($p<0,05$), tienen secuencias diferentes y cubren, en los dos casos, el 9% de las secuencia de las proteínas identificadas como se ve en la figura "Ejercicio 2: Resultados5".

Si aumentamos la tolerancia de los iones fragmento a 0,6 Da, el péptido del que también teníamos datos MS/MS y no había sido asignado a ninguna de las 2 proteínas identificadas se asigna a una de ellas, presentando una cobertura de secuencia del 12% (ver "Ejercicio 2: Resultados6"). Incluso si aumentamos de forma excesiva para las condiciones del análisis tanto la tolerancia de la masa del péptido (500 ppm) como la tolerancia de los fragmentos (0,8 Da), se siguen identificando de forma exclusiva y concluyente las mismas dos proteínas, como muestra la figura "Ejercicio 2: Resultados7".

Las dos proteínas identificadas en NCBInr utilizando los datos MS/MS pertenecen a *Candida albicans*, del que proceden las proteínas en estudio y presentan pesos moleculares y puntos isoeléctricos teóricos muy parecidos lo que hace posible que ambas proteínas se encuentren en el mismo spot del gel bidimendional. Se trata de una mezcla de proteínas en el mismo spot, ninguna de ellas es una proteína contaminante.

REFERENCIAS.

1. Henzel WJ, Watanabe C, Stults JT. Protein identification: the origins of peptide mass fingerprinting. J Am Soc Mass Spectrom. 2003,14(9), 931-42.

2. Mann, M., Hojrup, P.,Roepstorff, P. Use of Mass Spectrometric Molecular Weight Information to Identify Proteins in Sequence Databases. Biol. Mass Spectrom. 1993, 22, 338–345.

3. Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. Rapid Identification of Proteins by Peptide Mass Fingerprinting. Current Biol. 1993, 3, 327–332.

4. James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. Protein identification by Mass Profile Fingerprinting. Biochem. Biophys. Res. Commun. 1993, 195, 58–64.

5. Yates, J. R., III; Speicher, S.; Griffin, P. R.; Hunkapiller, T.Peptide Mass Maps: A Highly Informative Approach to Protein Identification. Anal. Biochem. 1993, 214, 397–408.

6. Wong, S. C.; Grimley, C.; Padua, A.; Bourell, J. H.; Henzel,W. J. In Techniques in Protein Chemistry IV; Angeletti, R. H., Ed.; Academic Press: San Diego, 1993, pp 371–378.

7. Mann, M.; Wilm, M. Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags. Anal. Chem. 1994, 66, 4390–4399.

8. Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. J. Am. Soc. Mass Spectrom. 1994, 5, 976–989.

9. Rosenfeld J, Capdevielle J, Guillemot JC, Ferrara P. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. Anal Biochem 1992, 203, 173–179.

10. Kumarathasan P, Mohottalage S, Goegan P, Vincent R. An optimized protein in-gel digest method for reliable proteome characterization by MALDI-TOF-MS analysis. Anal Biochem. 2005, 1, 346(1):85-9.

11. Birendra N. Pramanik, Urooj A. Mirza, Yao Hain Ing, Yan-Hui Liu,1 Peter L. Bartner, Patricia C. Weber, and Ajay K. Bose. Microwave-enhanced enzyme reaction for protein mapping by mass spectrometry: a new approach to protein digestion in minutes. Protein Sci. 2002, 11(11), 2676-87.

12. Stewart II, Thomson T, Figeys D. 18O labeling: a tool for proteomics. Rapid Commun Mass Spectrom 2001, 15, 2456–2465.

13. Richert S, Luche S, Chevallet M, Van Dorsselaer A, Leize-Wagner E, Rabilloud T About the mechanism of interference of silver staining with peptide mass spectrometry. Proteomics 2004, 4, 909–916.

14. Shevchenko A, Wilm M, Vorm O, and MannM. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem. 1996, 68(5), 850-8.

15. Shevchenko A, Tomas H, Havli J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* 2007, 1, 2856 – 2860.

16. Gharahdaghi F, Weinberg CR, Meagher DA, Imai BS, Mische SM. Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: A method for the removal of silver ions to enhance sensitivity. Electrophoresis 1999, 20, 601-605.

17. Helzen and Watenabe. Origin of PMF. J Am Soc Mass Spectrom 2003, 1414, 931–942.

18. Thiede B, Höhenwarter W, Krah A, Mattow J, Schmid M, Schmidt F, Jungblut PR. Peptide mass fingerprinting.Methods. 2005, 35(3), 237-47.

19. J. Gobon et al. Sample purification and preparation technique based on nano-scale reversed phased columns for sensitive analysis of complex peptide mixtures by Matrix assited desorption/ionization mass spectrometry. J. mass spectrom, 1999, 34, 105-116.

20. Smirnov IP, Zhu X, Taylor T, Huang Y, Ross P, Papayanopoulos IA, Martin SA, Pappin DJ. Suppression of alpha-cyano-4-hydroxycinnamic acid matrix clusters and reduction of chemical noise in MALDI-TOF mass spectrometry. Anal Chem. 2004, 76(10), 2958-65.

21. Schuerenberg M, Luebbert C, Eickhoff H, Kalkum M, Lehrach H, Nordhoff E. Prestructured MALDI-MS sample supports. Anal Chem. 2000, 72(15), 343642.

22. Florence Gonnet, Gilles Lemaître, Gilles Waksman and Jeanine Tortajada. MALDI/MS peptide mass fingerprinting for proteome analysis: identification of hydrophobic proteins attached to eucaryote keratinocyte cytoplasmic membrane using different matrices in concert. Proteome Science 2003, 1:2.

23. Neerau D. Padliya, Troy D. Wood Improved peptide mass fingerprinting matches via optimized sample preparation in MALDI mass spectrometry. Analytica Chimica Acta 2008, 627, 162-168.

24. Sleno L. The use of mass defect in modern mass spectrometry. J Mass Spectrom. 2012, 47(2), 226-36. PMID: 22359333.

25. Hanne Kolsrud Hustoft, Helle Malerod, Steven Ray Wilson, Leon Reubsaet, Elsa Lundanes and Tyge Greibrokk (2012). A Critical Review of Trypsin Digestion for LC-MS Based Proteomics, Integrative Proteomics, Dr. Hon-Chiu Leung (Ed.) 73-92).

26. Fei Xu , Wei-Han Wang , Yu-Jing Tan , and Merlin L. Bruening. Facile Trypsin Immobilization in Polymeric Membranes for Rapid, Efficient Protein Digestion. Anal. Chem.: 2010, 82 (24), pp 10045–10051.

27. Proc JL, Kuzyk MA, Hardie DB, Yang J, Smith DS, Jackson AM, Parker CE, Borchers CH. A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin. Journal of Proteome Research 2010, 9(10), 5422-5437.

28. Wiśniewski, J. R., Nagaraj N, Zougman A, Gnad F, Mann M. Brain phosphoproteome obtained by a FASP-based method reveals plasma membrane protein topology. Journal of Proteome Research 2010, 9(6), 3280-3289.

29. Hasan N, Wu HF, Li YH, Nawaz M. Two-step on-particle ionization/enrichment via a washing- and separation-free approach: multifunctional TiO2 nanoparticles as desalting, accelerating, and affinity probes for microwave-assisted tryptic digestion of phosphoproteins in ESI-MS and MALDI-MS: comparison with microscale TiO2. Analytical and Bioanalytical Chemistry 2010, 396(8), 2909-2919.

30. Lin, S., Yao G, Qi D, Li Y, Deng C, Yang P, Zhang X. Fast and Efficient Proteolysis by Microwave-Assisted Protein Digestion Using Trypsin-Immobilized Magnetic Silica Microspheres. Anal. Chem. 2008, 80(10), 3655-3665.

31. Capelo, J. L., Carreira R, Diniz M, Fernandes L, Galesio M, Lodeiro C, Santos HM, Vale G. Overview on modern approaches to speed up protein identification workflows relying on enzymatic cleavage and mass spectrometry- based techniques. Analytica Chimica Acta 2009, 650(2), 151-159.

32. Ileana R. León, Veit Schwämmle, Ole N. Jensen and Richard R. Sprenger. Quantitative assessment of In-solution digestión efficiency identifies optimal protocols for unbiased protein analysis. Mollecular And Cellular Proteomics 2013, 12: 10, 1074/mcp.M112.025585, 2992-3005.

33. Hernáez ML, Ximénez-Embún P, Martínez-Gomariz M, Gutiérrez-Blázquez MD, Nombela C, Gil C. Identification of Candida albicans exposed surface proteins in vivo by a rapid proteomic approach. J Proteomics. 2010, 73(7),1404-9.

34. Roepstorff P y Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom. 1984, 11(11), 601.

35. Ray Bakhtiar Æ Ziqiang Guan. Electron capture dissociation mass spectrometry in characterization of peptides and protein Biotechnol Lett 2006, 28, 1047–1059 DOI 10.1007/s10529-006-9065-z.

36. Frank Sobott, Stephen J. Watt, Julia Smith, Mariola J. Edelmann, Holger B. Kramer, and Benedikt M. Kessler. Comparison of CID Versus ETD Based MS/MS Fragmentation for the Analysis of Protein Ubiquitination. J Am Soc Mass Spectrom 2009, 20, 1652–1659.

37. Peter J. Ulintz, Bernd Bodenmiller, Philip C. Andrews, Ruedi Aebersold, and Alexey I. Nesvizhskii. Investigating MS2/MS3 Matching Statistics a model for coupling consecutive stage mass spectrometry data for increased peptide identification confidences. Molecular & Cellular Proteomics 2007, 6, 71–87.

38. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faergeman, N. J., Mann, M., and Jensen, O. N. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol.Cell.Proteomics 2005, 4, 310–327.

39. Pappin DJ, Hojrup P, Bleasby AJ. Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol. 1993,,3(6), 327-32.

40. Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data — the protein inference problem. Mol Cell Proteomics 2005, 4(10), 1419–40.

41. Kinter, Michael / Sherman, Nicholas E. Protein Sequencing and Identification Using Tandem Mass Spectrometry. Wiley-Interscience Series on Mass Spectrometry. 2000. XVIII, 302.

42. Cottrell JS. Protein identification using MS/MS data. J Proteomics. 2011, 74(10), 1842-51.

43. Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat Biotechnol 2006, 24(10), 1285–92.

44. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P,et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 2006, 127(3), 635–48.

45. Tsur D, Tanner S, Zandi E, Bafna V, Pevzner PA. Identification of post-translational modifications by blind search of mass spectra. Nat Biotechnol 2005, 23(12),1562–7.

46. Tanner S, Payne SH, Dasari S, Shen Z, Wilmarth PA, David LL, et al. Accurate annotation of peptide modifications through unrestrictive database search. J Proteome Res 2008, 7(1),170–81.

47. Johnson RS, Davis MT, Taylor JA, Patterson SD. Informatics for protein identification by mass spectrometry. Methods. 2005, 35(3), 223-36.

48. Claassen M. Inference and validation of protein identifications. Mol Cell Proteomics. 2012, 11(11),1097-104.

Determinación del número de grupos sulfidrilo y de enlaces disulfuro mediante espectrometría de masas

20

Juan José Calvete[1]

[1]Laboratorio de Proteinómica Estructural, Instituto de Biomedicina de Valencia, C.S.I.C.

1. Introducción.

Además de su relevancia funcional en proteínas que actúan mediante mecanismos de oxido-reducción (redox), de todos los aminoácidos proteicos, la cisteína es el único utilizado ampliamente durante la evolución para estabilizar la estructura espacial (terciaria o cuaternaria) de proteínas que, por su tamaño y/o secuencia, carecen de un núcleo hidrofóbico estable o de los elementos de estructura secundaria regular (hélice alfa, hojas beta) formadores de enlaces de hidrógeno (Cheek *et al*., 2006). Por el contrario, los enlaces disulfuro son raros en proteínas termoestables. Los enlaces disulfuro intra- e intermoleculares se forman por oxidación de cisteínas espacialmente cercanas (≤ 5Å entre los Cα) en las estructuras terciaria y cuaternaria, respectivamente. Se acepta generalmente que el efecto estabilizador de un enlace disulfuro es debido a que su formación reduce los grados de libertad conformacional del estado desplegado, reduciendo así la entropía del estado desplegado (Anfisen y Scheraga, 1975; Thornton, 1981). El aporte energético de un enlace disulfuro a la conformación nativa es de 2.5-3.5 kcal/mol. El gran impacto estructural a bajo coste energético destaca la importancia de la ingeniería de enlaces disulfuro como un mecanismo evolutivo de la diversificación estructural (y por tanto, funcional) de proteínas que, como es el caso de muchas toxinas, están sometidas a evolución Darwiniana acelerada (Olivera, 2006; Calvete *et al*., 2003a; Ménez, 2002). En el caso de serpientes y lagartos ponzoñosos, los venenos se originaron en etapas tempranas de su evolución por reclutamiento, y posterior transformación estructural y funcional, de proteínas ordinarias pertenecientes a un número reducido de familias proteicas (Fry *et al*., 2006). Una característica de estas proteínas es su alto contenido en cisteínas, las cuales usualmente se encuentran formando enlaces disulfuro. El número de cisteínas totales, reducidas (grupos sulfidrilo) y oxidadas (formando enlaces disulfuro), es en gran medida específico de cada familia de toxinas (Tabla 1) y tiene, por tanto, un valor taxonómico relevante. En efecto, un serio inconveniente a la hora de caracterizar mediante técnicas proteómicas el conjunto de proteínas presentes en los venenos de serpientes es la total carencia de bases de datos de genomas de reptiles. Una traba añadida es que, a pesar de que la base de datos pública UniProtKB/TrEMBL (v.35.3, 17 de Abril del 2007) (http://us.expasy.org/sprot/) contiene 932 secuencias de proteínas de serpientes ("viper") que representan a todas las familias de proteínas encontradas en los venenos, las estructuras primarias de toxinas de una misma familia varían enormemente debido a su diversificación estructural por evolución acelerada, impidiendo -por lo general- una identificación

basada en el análisis de la huella peptídica mediante espectrometría de masas MALDI-TOF. Se necesita, pues, recurrir a la secuenciación *de novo* mediante MS/MS y a la búsqueda posterior de similitud de secuencia (utilizando BLAST; http://www.ncbi.nlm.nih.gov/BLAST). Sin embargo, el orden de estos dos pasos puede invertirse, para nuestro provecho, si conocemos *a priori* en qué familia proteica englobar a una determinada toxina. Así, en aquellos casos en los que la calidad de los espectros de fragmentación no es suficiente para la secuenciación *de novo* completa de los iones peptídicos, este conocimiento nos permite generar un alineamiento múltiple de todas las proteínas homólogas conocidas que nos sirva entonces bien de guía para extender las secuencias parciales deducidas, bien para realizar un análisis tipo MS-BLAST (Shevchenko *et al*., 2001) (http://dove.embl-heidelberg.de/Blast2/msblast.html) frente a una base de datos propia, específica y reducida.

2. Contando cisteínas y enlaces disulfuro.

A. Proteínas monoméricas. La espectrometría de masas es, sin duda, la técnica idónea, por su sencillez y precisión, para determinar el contenido en cisteínas libres (grupos sulfidrilo, SH) y enlaces disulfuro (cistinas, S-S) de una proteína. Para ello, la muestra (típicamente una fracción cromatográfica eluída de una columna C18 de fase reversa) se divide en tres alícuotas y éstas se secan mediante centrifugación a vacío (SpeedVac) o liofilización. Las muestras se disuelven en 10 μl de tampón 50 mM HEPES, pH 9.0, conteniendo 5M cloruro de guanidina y 1 mM EDTA, y las proteínas se desnaturalizan a 85 ºC durante 15 min. Una de las muestras (A) no será sometida a tratamiento posterior y se utlizará para medir la masa molecular de la proteína nativa (M_{NAT}). A la muestra B se le añade un reactivo alquilante (4-vinilpiridina o yodoacetamida) hasta una concentración final de 10 mM y la mezcla se deja a temperatura ambiente (~25 ºC) y en la oscuridad durante 1 hora. Esta muestra se utilizará para determinar la presencia de grupos sulfidrilo reactivos. Por último, la muestra C se reduce con 10 mM 1,4-ditioeritriol (DTE) durante 15 min. a 80 ºC, y los grupos sulfidrilos se bloquean con un reactivo alquilante (4-vinilpiridina o yodoacetamida)[1] (concentración final de 25 mM) durante 1 hora a temperatura ambiente y en la oscuridad (Henschen, 1986). A continuación, las proteínas de las muestras A, B y C se purifican utilizando pipetas Zip-Tip C18 (Millipore) previamente activadas con una mezcla de acetonitrilo (ACN) al 70% y TFA al 0.1% y equilibradas en TFA al 0.1%. Tras la adsorción a la matriz C18 (5-6 ciclos de pipeteo), las Zip-Tip se lavan 5-6 veces con 0.1% TFA. Las proteínas libres de reactivos se eluyen con 3-5 μl de 70% ACN/0.1% TFA y sus masa moleculares se determinan mediante nanoESI-MS o MALDI-TOF MS utilizando ácido sinapínico (ácido 3,5-dimetoxi-4-hidroxicinámico) saturado en 70% ACN/0.1% TFA como matríz.

El número de grupos sulfidrilo reactivos (N_{SH}) se calcula mediante la ecuación I:

$$N_{SH} = (M_{ALK} - M_{NAT})/M_R \qquad \text{(ecuación I)},$$

donde M_{ALK} representa la masa molecular de la proteína desnaturalizada, pero no reducida, incubada en presencia del reactivo alquilante (muestra B); M_{NAT} es la masa de la proteína nativa (muestra A), y M_R es el incremento de masa debido a la alquilación de un grupo sulfidrilo (105.3 Da para S-piridiletilación, si se utilizó 4-vinilpiridina como agente alquilante, y 57.1 Da por carbamidometilación por yodoacetamida).

[1]. La yodoacetamida y la 4-vinilpiridina, en las concentraciones normalmente empleadas (exceso molar de 5 respecto a la concentración de residuos de cisteína o del reactivo reductor, lo que en la práctica equivale a concentraciones finales de 50-250 mM) son reactivos específicos para grupos sulfidrilo. La solubilidad de 4-VP en tampones acuosos es muy baja y debe emplearse para alquilar proteínas en presencia de cloruro de guanidino 5-6 M. Hay que tener también en cuenta que la pridiletilación aumenta la hidrofobicidad de las proteínas, por lo que, para las medidas espectrométricas, éstas deben ser manejadas en disoluciones conteniendo solventes orgánicos (ej. 70% acetonitrilo/0.1% TFA, que se emplea para desorber proteínas del material C18 de las pipetas Zip-Tip).

El número total de cisteínas se determina entonces mediante la ecuación II:

$$N_{Cys} = [(M_{CM} - M_{ALK})/(M_R + 1)] + N_{SH}$$

(ecuación II),

en la que M_{CM} es la masa de la proteína totalmente reducida y alquilada, y (M_R +1) es el incremento de masa debido a la alquilación de un resíduo de cisteína que previamente a la reducción formaba parte de un enlace disulfuro (106.3 Da o 58.1 Da, dependiendo de que se utilizara, respectivamente, 4-vinilpiridina o yodoacetamida como agente alquilante). En el caso de péptidos, cuya masa molecular puede ser determinada con precisión isotópica (ej. utilizando el reflectrón de un instrumento de tiempo de vuelo o mediante ionización por electrospray), nominalmente bastaría con reducir la muestra para determinar su contenido en S-S (Fig.1). Sin embargo, dado que la reducción de un enlace disulfuro produce un pequeño incremento de masa (+2 Da/S-S), cuando la exactitud de la medida es ≤ 2 Da, se hace necesario amplificar esa diferencia mediante la introducción de un grupo alquilante.

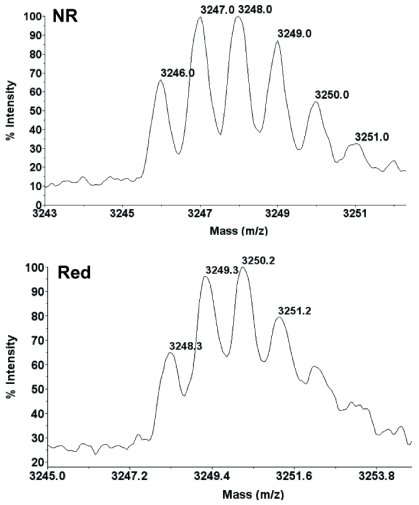


Figura 1. Espectros de masas obtenidos en un instrumento de tiempo de vuelo Voyager DE-Pro™ (Applied Biosystems) en modo reflectror, mostrando la distribución isotópica de los iones. La diferencia de 2 Da entre las masas del péptido nativo (NR) y del reducido (Red) indica la existencia de un S-S intramolecular.

Finalmente, el número de enlaces disulfuro (N_{S-S}) puede calcularse mediante la ecuación III:

$$N_{S-S} = (N_{cys} - N_{SH})/ 2 \quad \text{(ecuación III).}$$

Las unidades de masa en las ecuaciones I-III están expresadas en Daltons (1 Da o unidad de masa atómica unificada (1 u) = 1/12 de la masa del C_{12} = 1/N_A (gramos), siendo N_A el número de Avogadro (1.66053886 ± 0.00000028) x 10^{-24} g).

Para ilustrar lo anteriormente expuesto examinemos la Figura 2. El panel A muestra el espectro de una proteína del veneno de *Sistrurus miliarius barbouri* purificada por HPLC de fase reversa (Juárez *et al*., 2004). La masa de la proteína no varió en presencia de 4-vinilpiridina (M_{NAT} = M_{ALK} = 13980 Da), deduciendose que la proteína no contiene cisteínas libres. Por otra parte, la masa molecular de la proteína reducida y S-piridiletilada (M_{CM}) sufrió un incremento de 15483 - 13980 = 1503 Da, indicando que la proteína posee un total de 1503/106.3 = 14.1 (14) resíduos de cisteína, los cuales, en la proteína nativa, forman 7 enlaces disulfuro. Muy posiblemente se trata de una PLA_2 (Tabla 1).

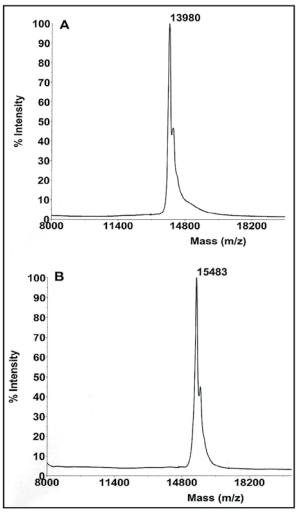


Figura 2. Determinación del número de grupos sulfidrilo y de enlaces disulfuro mediante espectrometría de masas MALDI-TOF utilizando un espectrómetro Voyager DE-Pro™ (Applied Biosystems). Panel A, espectro de masas de la proteína nativa que resultó ser idéntico al de la proteína desnaturalizada, pero no reducida, y tratada con 4-vinilpiridina. Panel B, Masa molecular de la proteína completamente reducida y S-piridiletilada.

B. Proteínas diméricas. El protocolo para cuantificar cisteínas y disulfuros en proteínas monoméricas puede aplicarse a proteínas homodiméricas (A_2) con sólo sustituir M_{CM} por (2 x M_{CM}A) en la ecuacion II:

$$N_{Cys}A_2 = [(2 \times M_{CM}A - M_{ALK})/(M_R + 1)] + N_{SH} \quad \text{(ecuacion IV)},$$

siendo $M_{CM}A$ la masa molecular de la subunidad A reducida y alquilada. La figura 3 muestra el ejemplo de una proteína, cuya masa molecular (M_{NAT} = 14195.9 Da) no varió tras desnaturalización e incubación con 4-vinilpiridina, pero que tras reducción y alquilación produjo una única subunidad de 8162.3 Da. En este caso, N_{Cys} = [(8162.3 x 2) - 14195.9)/106.3] = 20. Y, dado que la proteína nativa no contiene grupos SH, las 20 cisteínas S-piridiletiladas (10 por subunidad) estaban todas formando parte de enlaces disulfuro. Se trata, pues, probablemente, de una disintegrina homodimérica (Tabla 1).

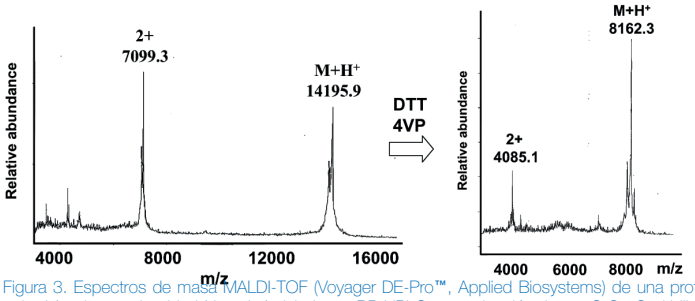


Figura 3. Espectros de masa MALDI-TOF (Voyager DE-Pro™, Applied Biosystems) de una proteína homodimérica nativa (izquierda) y de su subunidad (derecha) aislada por RP-HPLC tras reducción de sus S-S y S-piridiletilación de los grupos sulfidrilo formados.

De manera similar, la cuantificación del número total de resíduos de cisteína de una proteína heterodimérica (LS) exige reemplazar el término M_{CM} de la ecuación II por la suma de las masas de las subunidades mayor (L) y menor (S) reducidas y alquiladas (M_{CM} L + M_{CM}):

$$N_{Cys}LS = [((M_{CM}\,L + M_{CM}\,S) - M_{ALK}) /(M_R + 1)] + N_{SH} \quad \text{(ecuación V)}.$$

La figura 4 muestra los espectros de masas de una proteína de masa molecular $M_{NAT} = M_{ALK}$ = 14733.7 Da (panel central) y de sus subunidades aisladas tras reducción y alquilación con 4-vinilpiridina (4VP, paneles superiores) y con yodoacetamida (IA, paneles inferiores). Resulta evidente que se trata de una proteína heterodimérica y que ninguna de sus subunidades contiene grupos sulfidrilo. Por otra parte, los resultados de alquilación con 4VP o con IA, podemos determinar que el número total de cisteínas (N_{CYS}) del heterodímero LS es: [(8687.1 + 8162.8 - 14733.7)/106.3] = 19.9 y [(8204.6 + 7680.8 - 14733.7)/ 58.1] = 19.8. Además, combinando los dos conjuntos de datos podemos determinar el número de cisteínas presentes en cada subunidad:

$$N_{Cys}L = (M_{VP}L - M_{IA}L)/ (\Delta M_{VP}-\Delta M_{IA}) \quad \text{(ecuación VIa)}$$

$$N_{Cys}S = (M_{VP}S - M_{IA}S)/ (\Delta M_{VP}-\Delta M_{IA}) \quad \text{(equation VIb)},$$

En las ecuaciones VIa y VIb, ($M_{VP}L$ - $M_{IA}L$) y ($M_{VP}S$ - $M_{IA}S$) corresponden, respectivamente, a la diferencia entre las masas alquiladas con 4-vinilpridina y con yodoacetamida de las subunidades mayor y menor, y (ΔM_{VP}-ΔM_{IA}) es la diferencia entre el incremento de masa debido a la alquilación de un grupo tiol con 4-vinilpyridina (ΔM_{VP}) o con yodoacetamida (ΔM_{IA}) (106.3 - 58.1 = 48.2). Utilizando los datos de la figura 4 se deduce que las subunidades L y S de la proteína de M_{NAT} 14733.7 Da contienen, respectivamente, (8687.1 - 8204.6)/ (106.3 - 58.1) = 10 Cys y (8162.8 - 7680.8)/ (106.3 - 58.1) = 10 Cys. Debe tratarse, pues, de una disintegrina heterodimérica (Tabla 1).

Es necesario aclarar que, con independencia de los reactivos alquilantes que se utilicen en un experimento de doble marcaje -como el mostrado en la figura 4- las especies de mayor (o menor) masa molecular corresponden a la misma subunidad. Esta relación de proporcionalidad se

cumple siempre que i) el número de cisteínas de la subunidad mayor sea igual o mayor que el de la subunidad menor ($N_{CYS}L \geq N_{CYS}S$), o ii) en el caso de que $N_{CYS}S > N_{CYS}L$, la diferencia entre las masas moleculares de las subunidades mayor ($M_{NAT}L$) y menor ($M_{NAT}S$) sea superior a la diferencia entre el número de cisteínas de la subunidad menor ($N_{cys}S$) y el número de cisteínas de la subunidad mayor ($N_{cys}L$) multiplicada por el incremento de masa debido a la alquilación de un grupo tiol por el agente alquilante que genera mayor aumento de masa (M_{RH}). Esta segunda condición puede expresarse de manera general mediante la ecuación VII:

$$(M_L - M_S) > (N_{cys}S - N_{cys}L) \times M_{RH} \quad \text{(ecuación VII).}$$

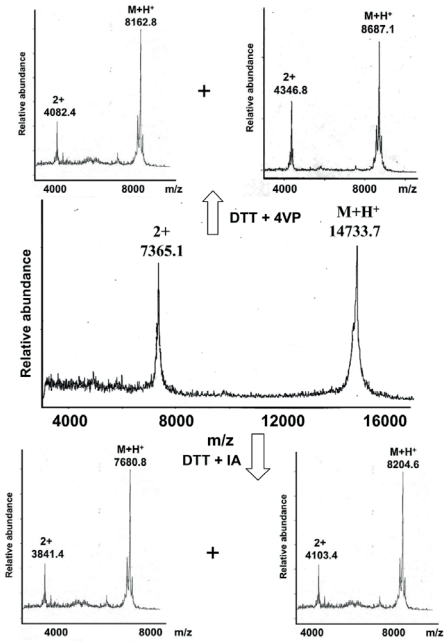


Figura 4. Espectros de masas MALDI-TOF (adquiridos como en las figuras 1 y 2) de una proteína heterodimérica nativa (panel central) y de sus subunidades S (izquierda) y L (derecha) aisladas mediante HPLC de fase reversa tras reducción de sus S-S y S-piridiletilación (paneles superiores) y carbamidometilación (paneles inferiores) de los grupos sulfidrilo generados.

Aunque, tanto en el caso de proteínas homo- y heterodiméricas, los datos relativos a la cantidad de cisteínas de sus subunidades no permiten asignar el número de enlaces S-S intra- e intermoleculares, si que se puede formular de manera general que si una proteína dimérica no contiene grupos -SH, entonces tanto el número de S-S intracatenarios como el de S-S intermoleculares será par si las subunidades contienen un número par (2n) de cisteínas, e impar (2n-1) si el número de cisteínas de cada subunidad es impar. Nótese que no puede darse el caso de que una subunidad contenga 2nCys y la otra (2n-1) Cys.

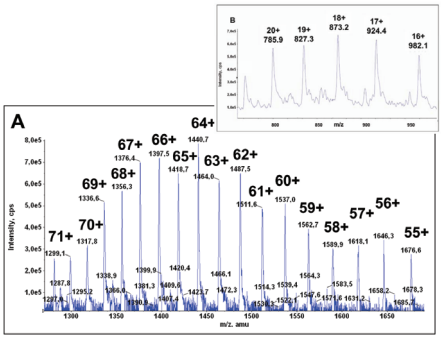
C. Proteínas multiméricas. Conceptualmente, los mismos principios que se aplican para la determinación del contenido de sulfidrilos y enlaces disulfuro en proteínas diméricas son válidos para proteínas n-méricas, si bien, logicamente, hay que modificar las ecuaciones anteriores. Así, el número total de cisteínas de una proteína multimérica (N_{Cys} Mult) puede determinarse mediante una modificación de las ecuaciones II, IV y V:

$$N_{Cys}\text{Mult} = [(\sum M_{CM}\text{ Sub}) - M_{ALK}) / (M_R + 1)] + N_{SH} \quad \text{(ecuación VIII),}$$

donde $\sum M_{CM}$ Sub es el sumatorio de las masas moleculares de todas las subunidades reducidas y alquiladas.

La figura 5 ilustra el protocolo seguido para establecer la estructura cuaternaria de una proteína multimérica (masas moleculares aparentes por SDS-PAGE: M_{NAT}, 94 kDa; M_{SUB}, 13-16 kDa). La masa molecular de la proteína nativa (91151 ± 11 Da) fue determinada mediante ionización por electrospray con un espectrómetro híbrido triple cuadrupolo-trampa iónica lineal (QTrap 2000, Applied Biosystems) (panel A). Mediante la ecuación VIa (o VIb), y partiendo de las masas moleculares determinadas para las subunidades aisladas por HPLC de fase reversa piridiletiladas y carbamidometiladas (tabla), se deduce que cada subunidad contiene un total de 8 cisteínas. Por tanto, las masas moleculares promedio calculadas para las subunidades nativas (con todas las cisteínas formando S-S) (M_{NAT}Sub) son [M_{CM}Sub- (8 x 58.1)]: 16041 Da (SubA), 15906 Da (SubB), 14743 Da (SubC), 14828 Da (SubD) y 14799 Da (SubE). El $\sum M_{NAT}$SubA-E = 76317 Da. La diferencia entre este valor y la masa molecular de la proteína nativa es -14834 Da, in-

dicando que la proteína nativa es hexamérica con una estequiometría cuaternaria $[ABC(D)_2E]$. Este tipo de estructura supramacromolecular es típico de C-lectinas multiméricas formadas por la asociación cíclica de heterodímeros ($\alpha\beta$). Cada dímero contiene 3 S-S intracatenarios, 1 S-S interdimérico y se enlaza a otros dos dímeros mediante dos S-S, uno entre las cisteínas N-terminal (Cys1) de la subunidad α y C-terminal (Cys8) de la subunidad β y otro entre las correspondientes cysteínas C-terminal de la subunidad β y N-terminal de la subunidad α de dímeros vecinos. La proteína 91 kDa puede describirse como un trímero de dímeros de fórmula $[(\alpha\beta)_2]_3$ o $[\alpha\text{-SS-}\beta]_1\text{-Cys}_{Cter}\alpha\text{-Cys}_{Nter}\beta\text{-}[\alpha\text{-SS-}\beta]_2\text{-Cys}_{Cter}\alpha\text{-Cys}_{Nter}\beta\text{-}[\alpha\text{-SS-}\beta]_3\text{-Cys}_{Cter}\alpha\text{-Cys}_{Nter}\beta\text{-}[\alpha\text{-SS-}\beta]_1$ (Calvete *et al.*, 2003b). La asociación cabeza-cola de heterodímeros $\alpha\beta$ es una solución evolutiva sencilla para formar polímeros de complejidad creciente, $(\alpha\beta)n$, que desarrollen nuevas actividades biológicas. Pero esto es harina de otro costal...



IA DTT 4VP

16506	16890
16371	16756
15208	15592
15293	15677
15264	15648

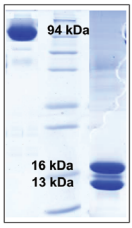


Figura 5. Espectros de masas adquirido mediante ionización por electrospray (ESI) en un espectrómetro QTrap 2000 (Applied Biosystems) de una proteína purificada del veneno de *Bitis nasicornis* (A) y de una de sus subunidades (B). La masa molecular aparente, por SDS-PAGE, de la proteína nativa es 94 kDa y la determinada por ESI-MS es 91151 ± 11 Da. La reducción y alquilación de esta proteína produjo las subunidades listadas en la tabla. El espectro ESI-MS mostrado en el panel B corresponde a la subunidad piridiletilada de 15677 Da.

3. Otras consideraciones.

La metodología descrita ha sido aplicada en nuestro laboratorio esencialmente para caracterizar toxinas de venenos de serpientes e ilustra los conceptos básicos aplicables toericamente a cualquier otra proteína. La limitación práctica es la precisión con que puedan medirse las masas de las diversas especies moleculares. Ello no sólo depende de las características operativas del espectrómetro de masas sino también de las propiedades fisicoquímicas de la proteína. Además, modificaciones postraduccionales como glicosilación (en particular con azúcares ácidos), sulfatación, fosforilación, etc. suelen producir poblaciones proteicas heterogéneas y afectan a las propiedades iónicas y, por tanto, al proceso de ionización de las proteínas. La masa molecular de la proteína cuyo espectro de masas se muestra en la figura 5 pudo ser determinada por ESI-MS con una precisión de unas 120 ppm. Sin embargo, la proteína de la figura 6 (MAT, http://au.expasy.org/uniprot/P13444, M_{av} calculada = 43567 Da, N_{CYS} = 10) no pudo ionizarse mediante electronebulización aunque si utilizando MALDI-TOF-MS. En este caso (Fig.6), se aprecia que los picos son más anchos que los iones de la figura 5 y, por tanto, de menor resolución. No obstante, incluso con una resolución nominal 10 veces inferior a ESI-MS (que puede mejorarse hasta ~ ±200 ppm si se trabaja con un estándar interno y se optimiza el tiempo de extracción de los iones), un valor teórico de 44000 ± 50 Da permite determinar con precisión suficiente la incorporación de un grupo piridiletil (alquilación de un solo grupo sulfidrilo), dado que un incremento de masa de 105.3 es 2.1 veces el error medio. Queda clara, pues, la importancia de amplificar lo más posible la diferencia de masa entre las especies M_{NAT}, M_{ALK} y M_{CM} utilizando reactivos alquilantes que generen un incremento de masa (M_R) mayor que el error experimental de la medida. Aunque los reactivos alquilantes más comunmente utilizados son

yodoacetamida (M_R = 57.1) y 4-vinilpiridina (M_R = 105.3), la variedad de reactivos específicos de grupos tioles es grande, en particular los conjugados a fluoróforos diseñados para estudios de Biología Celular. Un ejemplo interesante es la eosina-5´-yodoacetamida. Este reactivo hidrosolúble (Molecular Probes, Invitrogen: http://probes.invitrogen.com/media/msds/30450.html) produce proteínas fluorescentes y genera un M_R = 830.8. Resulta, por tanto, muy sencillo investigar como paso previo a la cuantificación de N_{SH}, por SDS-PAGE tras el tratamiento con el fluróforo, si una proteína contiene o no grupos sulfidrilo reactivos.

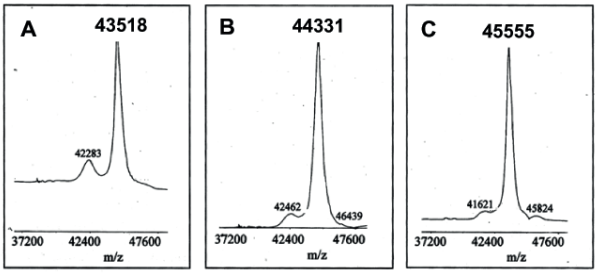


Figura 6. Espectros de masas de la proteína MAT nativa (A), MAT tratada con 4-vinilpiridina en condiciones desnaturalizantes pero no reductoras (B) y MAT reducida y piridiletilada (C) adquiridos mediante MALDI en un espectrómetro Voyager DE-Pro™ (Applied Biosystems) en modo lineal y con 200 nseg de extracción retardada (DE, delayed extraction) de iones. La masa molecular promedio, M_{av}, de MAT fue 49 Da inferior a la M_{av} calculada (43567 Da). No obstante, dado que este error afecta, en principio, por igual a las medidas siguientes efectuadas con la misma calibración, el cálculo de N_{SH} = (44331-43518)/106.3 = 7.6, N_{Cys} = (44555 - 43518)/106.3 = 9.75, y N_{SS} = (9.75-7.6)/2 = 1.07 indicó claramente que 2 de las 10 cisteínas de la molécula estaban formando un enlace disulfuro y las restantes 8 estaban reducidas (Pérez-Sánchez *et al*. 2003).

Tabla 1. Clasificación de toxinas de venenos de serpientes de las familias *Viperidae* (víboras) y *Crotalidae* (crótalos, serpientes de cascabel) en función de su contenido de cisteínas y enlaces disulfuro. [a], enlaces disulfuro intermoleculares; [b], enlaces disulfuro intramoleculares; PLA_2, fosfolipasa tipo A_2; CRISP, proteína rica en cisteínas; SVMP, Zn2+-metaloprotease de tipo PI de veneno de serpiente; DC, fragmento C-terminal de SVMP de tipo PIII (SVMP-PIII) compuesto por un dominio tipo disintegrina y un dominio rico en cisteínas; svVEGF, factor de crecimiento endotelial de veneno de serpiente; LAO, L-aminoácido oxidasa.

Rango de masa molecular (kDa)	1Nº total de:		Familia proteica
	-SH	S-S	
1.6 - 2	-	1	Péptido natriurético
4 - 5	-	3	Miotoxina
	-	4	Disintegrin corta
6 - 8	-	3	Inhibidor tipo Kunitz
	-	5	Subunidad de disintegrina dimérica
	-	6	Disintegrina mediana
10 - 12	1	-	Ohanina
13 - 15	-	2	Cistatina
	-	(2^a + 4^b)	Disintegrina dimérica
	-	7	PLA_2
23 - 33	-	8	CRISP
	1	4	PI-SVMP
	-	6	Proteasa de serina
	-	(1^a + 3^b)	C-lectin $\alpha\beta$
	-	(1^a + 4^b)	svVEGF
	-	13	Fragment DC
46 - 58	-	3	LAO
	1	18	PIII-SVMP

REFERENCIAS.

1. Anfinsen CB, Scheraga HA. 1975. Experimental and theoretical aspects of protein folding. Advances in Protein Chemistry 29: 205–300.

2. Calvete JJ, Moreno-Murciano MP, Theakston RDG, Kisiel DG et al. 2003a. Snake venom disintegrins: novel dimeric disintegrins and structural diversification by disulphide bond engineering. Biochemical Journal 372: 725-734.

3. Calvete JJ, Moreno-Murciano MP, Marcinkiewicz C. 2003b. Venenos de serpientes. Diversificación estructural de proteínas ancestrales. Investigación y Ciencia Abril 2003: 33-35.

4. Cheek S, Krishna S, Grishin NV. 2006. Structural classification of small, disulphide-rich protein domains. Journal of Molecular Biology 359: 215-237.

5. Fry BG, Vidal N, Norman JA, Vonk FJ et al. 2006. Early evolution of the venom system in lizards and snakes. Nature, 439: 584-588.

6. Henschen A. 1986. Analysis of cyst(e)ine residues, disulfide bridges, and sulfhydryl groups in proteins. Advanced Methods in Protein Microsequence Analysis (Wittmann-Liebold et al., eds.), Springer Verlag, Berlin, pp.244-255.

7. Juárez P, Sanz L, Calvete JJ. 2004. Snake venomics: characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. *Proteomics* 4: 327-338.

8. Ménez A (Ed.). 2002. *Perspectives in Molecular Toxinology,* John Wiley & Sons, Ltd.: Chichester, UK.

9. Olivera BM. 2006. Conus peptides: biodiversity-based discovery and exogenomics. Journal of Biological Chemistry 281: 31173-31177.

10. Sánchez-Pérez G, Gasset M, Calvete JJ, Majares MA. 2003. Role of an intrasubunit disulfide in the association state of the cytosolic homo-oligomer methionine adenosyltransferase. *Journal of Biological Chemistry* 278: 7285-7293.

11. Shevchenko A, Sunyaev S, Lobodo A, Shevchenko A *et al*. 2001. Charting the proteomes of organisms with insequenced genomes by MALDI-Quadrupole time-of-flight mass spectrometry and BLAST homology searching. Analytical Chemistry 73: 1917-1926.

12. Thornton JM. 1981. Disulphide bridges in globular proteins. Journal of Molecular Biology 151: 261-287.

Glicómica

Dr. José A. Cremata[1], Dr. Raquel Montesino[2], Dr. Ada Triguero[3], Lic. Gleysin Cabrera[4]

[1] Departamento Química de Carbohidratos. Centro de Ingeniería Genética y Biotecnología. P.O. Box 6162, Habana, CUBA;
[2] Departamento Química de Carbohidratos. Centro de Ingeniería Genética y Biotecnología. P.O. Box 6162, Habana, CUBA;
[3] Departamento Química de Carbohidratos. Centro de Ingeniería Genética y Biotecnología. P.O. Box 6162, Habana, CUBA;
[4] Departamento Química de Carbohidratos. Centro de Ingeniería Genética y Biotecnología. P.O. Box 6162, Habana, CUBA.

1. Elementos básicos sobre glicosilación.

La glicosilación se considera la más abundante de las modificaciones post-traduccionales que sufren las proteínas de células eucariotas como parte del proceso de síntesis y secreción [1]. A diferencia de la síntesis proteica, no existe un molde por el cual se defina el número y posición de cada monosacárido en la cadena naciente. La estructura oligosacarídica va experimentando modificaciones durante su migración por los diferentes compartimentos en la vía secretora. Los factores que determinan la secuencia del oligosacárido son fundamentalmente: las condiciones del microambiente celular, la disponibilidad de donores nucleotídicos de azúcares y sus transportadores, la batería de glicosidasas y glicosiltransferasas y su localización en el sistema de endomembranas, específicamente del Retículo Endoplasmático (RE) y el aparato de Golgi [2]. La unión de las cadenas oligosacarídicas a la proteína no es un proceso aleatorio sino que ocurre en determinadas posiciones, definidas por la secuencia aminoacídica y la conformación espacial que va adquiriendo la proteína durante el proceso de plegamiento.

Se ha demostrado que la parte oligosacarídica tiene gran influencia en varias de las funciones biológicas de las glicoproteínas, como son: inmunogenicidad, solubilidad, reconocimiento, interacción célula-célula y receptor-hormona [3]. Además estos glicanos contribuyen a prolongar el tiempo de vida medio en el plasma sanguíneo [4], así como también pueden contribuir al correcto plegamiento de los polipéptidos durante su proceso biosintético [5-7].

1.1 Composición oligosacarídica de las glicoproteínas.

En las cadenas oligosacarídicas se encuentran fundamentalmente cuatro tipos de azúcares

- **Hexosas:** D-galactosa, D-manosa, D-glucosa.
- **Desoxihexosas:** L-fucosa.
- **Hexosaminas (2-acetamino-2-desoxi-hexosa):** *N*-acetil-D-galactosamina, *N*-acetil-D-glucosamina.
- **Acidos siálicos:** ácido *N*-acetilneuramínico y ácido *N*-glicolilneuramínico.

Como grupo, las glicoproteínas presentan grandes diferencias en su contenido de carbohidratos, que fluctúa entre 1% y 50% aproximadamente de la masa total de la glicoproteína. La unión covalente con el péptido se realiza mediante un enlace glicosídico con el hidroxilo de la cadena

lateral de residuos de serina (Ser), treonina (Thr) en el caso de la *O*-glicosilación o por el grupo NH_2 de la amida de la cadena lateral de la asparagina (Asn) en el caso de la *N*-glicosilación.

1.2 *N*-Glicosilación.

El destino de las glicoproteínas es ser secretadas o formar parte de la superficie celular, aunque algunas proteínas nucleares y citosólicas también están glicosiladas. La glicosilación de una proteína es un proceso biológico complejo, ordenado y no fortuito. No todos los aminoácidos del péptido blanco podrán ser glicosilados, para esto se requiere de una secuencia particular, denominada sitio potencial de glicosilación o "sequon". Este sitio se encuentra formando parte del tripéptido Asn-X-Thr(Ser), donde X puede ser cualquier aminoácido excepto Pro (prolina) [8]. Sin embargo, la presencia de este tripéptido es un elemento necesario pero no suficiente para que ocurra la *N*-glicosilación. Existen otros factores que afectan la transferencia del oligosacárido precursor y que traen como consecuencia que no todos los sitios potenciales de *N*-glicosilación sean finalmente ocupados. Por ejemplo, se ha descrito que existe una relación directa entre la frecuencia de la *N*-glicosilación y la ubicación del sitio potencial a glicosilar en la secuencia primaria de la proteína, siendo esta mayor en sitios potenciales lejanos al extremo carboxílico de la proteína (C-terminal).

La glicosilación comienza en el interior del RE siendo el primer paso la translocación de la proteína al lumen del RE. Simultaneamente ocurre la transferencia de un oligosacárido precursor previamente sintetizado, una parte en la cara citosólica y la otra en el lumen del RE, compuesto de 14 residuos 2 *N*-acetilglucosaminas, 3 glucosas y 9 manosas (Glc3Man-9GlcNAc2), a la Asn del sitio de glicosilación por la acción de la oligosacariltransferasa. Una vez transferido, en el proceso de la "maduración" de la proteína, este oligosacárido sufrirá modificaciones: primero la pérdida de las 3 glucosas y después 1 manosa, formándose el oligosacárido Man8GlcNAc2 listo para salir del RE y transitar por los distintos compartimentos del aparato de Golgi donde ocurrirán otras modificaciones adicionales (Figura 1). Según su estructura los *N*-glicanos se clasifican en: Oligomanosídicos, Complejos o Tipo Lactosamina e Híbridos, estos últimos presentan características estructurales de los dos anteriores.

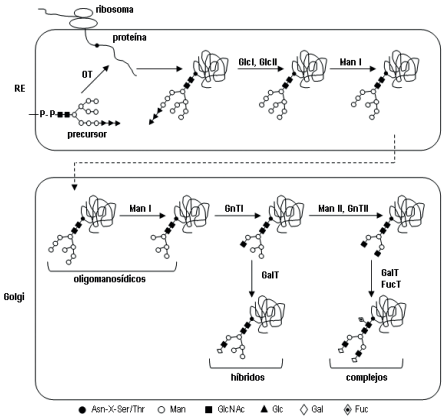


Figura 1. Representación de la biosíntesis de los *N*-glicanos de tipo oligomanosídicos, híbridos y complejos en el Retículo Endoplasmático (RE) y en el Aparato de Golgi. OT, oligosiltransferasa; Glcl, glucosidasa I, Glcll, glucosidasa II; Man I, manosidasa I; Man II, manosidasa II; GnTI, *N*-acetilglucosaminiltransferasa I; GnTII, *N*-acetilglucosaminiltransferasa II; GalT, galactosiltransferasa; FucT, fucosiltransferasa.

Un fenómeno que caracteriza a la glicosilación por su biosíntesis y la gran cantidad de componentes y enzimas que intervienen en el proceso es la microheterogeneidad de las estructuras oligosacarídicas enlazadas a los residuos de aminoácidos, lo que genera una gama amplia de glicoformas para cada glicoproteína. Se entiende por microheterogeneidad las diferentes estructuras que pueden enlazarse a un mismo aminoácido de un sitio de glicosilación determinado. Evidentemente esto será dependiente del organismo, tejido y célula y también de las condiciones microambientales en las que se produjo la síntesis de la glicoproteína [9, 10].

En este mismo sentido la *N*-glicosilación de algunas glicoproteínas puede variar con ciertas enfermedades. Una de las más estudiada ha sido la Artritis reumatoide donde la *N*-glicosilación de las IgGs del suero de individuos enfermos resulta significativamente diferente de la de los individuos controles sanos, con un predominio de la glicoforma biantenaria agalactosilada (y en

general un bajo contenido de Galactosa de sus *N*-glicanos) dado por la deficiencia en la actividad de la Galactosiltransferasa de la población de células B en los enfermos en comparación con los individuos sanos [11]. Otros ejemplos donde la variación de la *N*-glicosilación está directamente asociada con la enfermedad son las conocidas como Desorden Congénito de la Glicosilación (CDG del Inglés) [12]. En este tipo de enfermedad hay un serio compromiso de muchas de las funciones biológicas dando lugar a condiciones clínicas extremadamente severas de los pacientes. Su clasificación varía en dependencia si la deficiencia radica en el ensamblaje del *N*-glicano precursor CDG Tipo I, o posterior a la transferencia al sitio de glicosilación CDG Tipo II [13,14]. También se han observado modificaciones de la *N*-glicosilación y propuestas como biomarcadores de dichas enfermedades, en el caso de inflamación [15], enfermedades del hígado [16] y cáncer [17,18]. No obstante no debe considerarse que la glicosilación sea un proceso aleatorio, por el contrario es altamente reproducible. Cuando ocurren cambios en la glicosilación de glicoproteínas específicas es porque han ocurrido cambios externos o modificaciones del entorno celular. Por tanto, contar con bio(glico)marcadores confiables, reproducibles y sensibles resulta crucial para el diagnóstico y pronóstico de estas enfermedades. De aquí que el "glicoma" del suero total en combinación con la proteómica, a partir de muestras de sangre donde estarían representadas la mayoría de las glicoproteínas séricas provenientes de la síntesis en el hígado o en las células del plasma secretoras de inmunoglobulinas sería una herramienta poderosa en la realización de un diagnóstico adecuado. La identificación de cambios de glicosilación de glicoproteínas asociados a enfermedades puede convertirlas a ellas mismas en los mejores marcadores moleculares. Lograr que este tipo de ensayo altamente especializado esté al alcance de la mayoría de los laboratorios clínicos requiere de llegar a versiones relativamente fáciles de interpretar para que finalmente puedan convertirse en herramientas poderosas en manos de los especialistas.

Tomando en consideración lo anteriormente expuesto y debido al aumento del número de glicoproteínas de uso terapéutico se va tornando imprescindible el perfeccionamiento de los métodos de caracterización de la glicosilación de estas moléculas. Específicamente para los anticuerpos monoclonales las autoridades regulatorias están exigiendo la caracterización estructural detallada de la *N*-glicosilación pues se conoce que estos glicanos afectan la inmunogenicidad, estabilidad y la farmacocinética [19]. Todo esto implica determinar con precisión las estructuras de los oligosacáridos enlazados a la proteína en términos de monosacáridos constituyentes, posición y configuración de los enlaces. Para realizar stos ensayos se requiere de una batería amplia de técnicas analíticas y personal altamente capacitado y entrenado para la interpretación de los resultados.

2. Metodologías empleadas para la caracterización estructural de los N-*glicanos.*

El análisis de la glicosilación de una glicoproteína dada o de mezclas complejas como por ejemplo provenientes del suero total de humanos puede llevarse a cabo bajo diversas estrategias:

- Análisis de los oligosacáridos en la proteína intacta.
- Análisis de los glicopéptidos obtenidos por proteólisis parcial de las glicoproteínas.
- Análisis de los oligosacáridos obtenidos por reacciones de desglicosilación química o enzimática de las glicoproteínas de interés.

En muchas ocasiones la decisión de escoger una estrategia u otra no solo depende de la información que se quiere obtener, o de la cuantitividad de las determinaciones de los *N*-glicanos presentes en la glicoproteína o incluso de la cantidad de material biológico disponible sino

también de la experiencia acumulada en cada una de estas metodologías. La experiencia acumulada durante años en nuestro laboratorio se ha enfocado hacia el análisis y determinación estructural de la población de *N*-glicanos asociados a las glicoproteínas de interés, es por ello que profundizaremos en los métodos actualmente disponibles para acometer este tipo de análisis.

2.1 Reacciones de desglicosilación de glicoproteínas.

Se han descrito varias metodologías para la caracterización de perfiles de *N*-glicosilación basadas, por lo general, en métodos cromatográficos o combinaciones de métodos cromatográficos con espectrometría de masas. Común a todas estas es la liberación química o enzimática, en solución o en banda de gel de poliacrilamida de los oligosacáridos enlazados a los sitios de glicosilación. En la actualidad esta reacción se realiza a nivel picomolar para bandas de gel 1-D y se han realizado modificaciones para poder trabajar a nivel sub-picomolar a partir de bandas 2-D. La hidrólisis del enlace glicosídico entre la proteína y los *N*-glicanos se realiza eficientemente por la acción de la amidasa (péptido *N*-glicosidasa F, PNGasa F). Esta enzima es capaz de romper el enlace entre la Asn y GlcNAc de la mayoría de las glicoporteínas con la excepción de *N*-glicanos que posean un residuo de fucosa α1,3 enlazado a la GlcNAc del núcleo típico de proteínas de plantas e insectos, en esos casos la opción es la PNGasa A. En el caso de la desglicosilación con PNGasa F basta con un simple proceso de desnaturalización de la proteína en estudio. Sin embargo para los casos donde se requiere del uso de la PNGasa A se hace necesario una digestión proteolítica con pepsina o tripsina antes de la reacción de desglicosilación pues la PNGasa A no es capaz de actuar directamente sobre (glico)proteínas intactas. Las desglicosilaciones químicas más difundidas han sido la hidrazinólisis [20] que generan *N*-glicanos reductores similar a los que se generan en desglicosilacion enzimática y la beta eliminación por hidrólisis alcalina en presencia de borohidruro de sodio como agente reductor [21]. Esta última presenta dos grandes problemas, primero que los *N*-glicanos generados son no-reductores y segundo la gran cantidad de productos de degradación que se originan en pH alcalinos. Una opción de reciente desarrollo y aplicación ha sido la adición de amonio (hidróxido/carbonato) al medio de reacción [22,23].

2.2 Marcaje de *N*-glicanos con aminas primarias aromáticas.

La reacción de aminación reductiva de los oligosacáridos con aminas primarias fluorescentes ha permitido un incremento sustancial en la sensibilidad de la detección y una mayor resolución en columnas de cromatografía líquida de alta eficiencia (HPLC) de fase normal. Entre las más empleadas se encuentran la 2-aminobenzamida (2AB) [24,25] y la 2-aminopiridina (2-AP) [26,27]. Tanto la preparación de la solución de marcaje con la amina fluorescente como la reacción en sí son fáciles de ejecutar y de gran estabilidad. Se recomienda la preparación de una solución madre de marcaje que se conserva a -20C de la que se preparan alícuotas que son las que se emplean en la reacción. Los *N*-glicanos una vez marcados con el fluoróforo resultan estables en extremo aún cuando la estructura contenga ácidos siálicos terminales. Otra gran virtud de estas reacciones es que son cuantitativas y en relación *N*-glicano:fluoróforo 1:1 pues la aminación reductiva solo ocurrirá por el terminal reductor del oligosacárido.

2.3 Separación de *N*-glicanos marcados (2AB) en cromatografía de interacción hidrofílica.

La separación de los derivados 2AB se realiza por cromatografía de fase normal en columna Amida 80 con soportes modificados con grupos amidas que permite el análisis de los oligosacáridos neutros y cargados (cromatografía de interacción hidrofílica, HILIC de la sigla en Inglés). La asignación de las estructuras se realiza según los valores de tiempo de retención expresados en unidades de glucosa (GU). Para el cálculo de los GUs de los *N*-glicanos, se toma como referencia una mezcla de oligómeros de glucosa (isomalto-oligosacáridos) de un hidrolizado de dextrana (Figura 2). Por consecuencia, es una medida de la afinidad de cada *N*-glicano por la matriz de la columna y está relacionado con la hidrofilicidad de los oligosacáridos [24]. Inicialmente los cromatogramas resultaban muy extensos con programas de separación de hasta

180 min, en la actualidad se han desarrollado versiones de 120 y 60 min usando el mismo tipo de columna con partículas de 5 μm [28] incluso programas de separación aún menores (30 min) para columnas de partículas de 3 μm. Estos cambios implicaron modificaciones en los programas de cálculo de los valores de GU pues el perfil de separación y resolución del hidrolizado de dextrana varía con respecto al método inicial de 180 min (Figura 3).

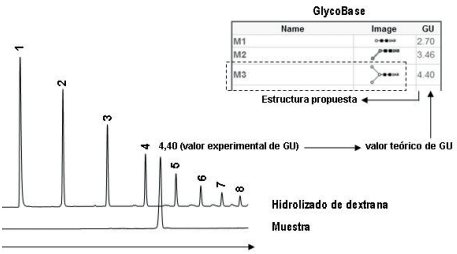


Figura 2. Procedimiento para el cálculo de los valores de GU experimentales de las fracciones de *N*-glicanos y asignación de estructuras empleando la GlycoBase.

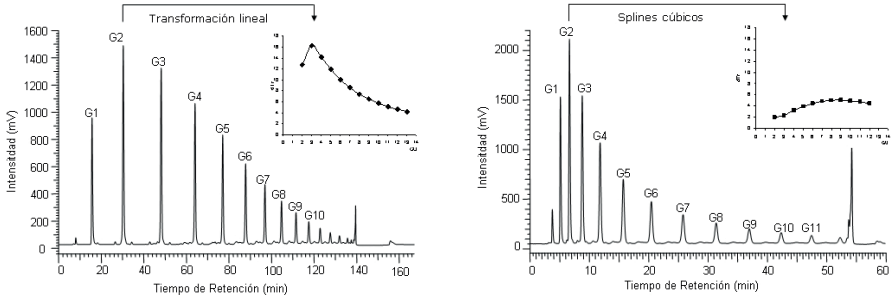


Figura 3. Variación de los tiempos de elución de la serie de isomalto-oligosacáridos empleados como estándares externos para el cálculo de GU en dependencia del tamaño de la partícula de la columna de fase normal Amida 80. Cromatograma superior muestra separación en columna de 5 μm con gradiente largo de 180 min y transformación lineal para cálculo de GU, cromatograma inferior muestra separación en columna de 5 μm con gradiente de 60 min y transformación splines cúbicos para cálculo de GU.

Los valores de GUs no representan un parámetro absoluto, son reproducibles, comparables y predecibles (contribución de los diferentes monosacáridos y su posición relativa dentro del *N*-glicano), siempre que se mantengan las condiciones de separación en que fueron determinados originalmente. Por ejemplo la variación de la temperatura durante la separación en HPLC, afecta significativamente el tiempo de retención, el valor de GU y por consiguiente la asignación de la estructura que se realiza al comparar con los valores informados en la GlycoBase del "*Dublin-Oxford Glycobiology Laboratory*" (http://glycobase.ucd.ie/cgibin/profile_upload.cgi). Esta GlycoBase contiene más de 400 *N*-glicanos con sus correspondientes valores de GU. Los valores de GUs informados en la Glycobase fueron determinados de forma experimental y las estructuras de los *N*-glicanos están debidamente verificadas por digestiones con exoglicosidasas y espectrometría de masas. La representación esquemática de las estructuras en la GlycoBase emplea los elementos y simbología descritos en la Figura 4 [29], simbología que se utiliza en este documento.

La nomenclatura de las estructuras utiliza las siguientes reglas:

- La estructura núcleo de la *N*-glicosilación $Man_3GlcNAc_2$ se representa como M3 (Figura 4B).
- La fucosilación del núcleo se representa con F y entre paréntesis se señala la posición del enlace, por ejemplo si es α1,6Fuc se representa como F(6) (Figura 4C) y se ubica siempre al inicio.
- La xilosilación se representa como X (Figura 4D).
- La GlcNAc que se enlaza al núcleo se representa como A y se indica a continuación cuántos residuos de este monosacárido están presentes. La estructura M3 deja de escribirse cuando el *N*-glicano tiene dos o más GlcNAc unidos al núcleo (Figura 4C).

- La Gal se representa con la letra G. Se indica delante la rama del *N*-glicano a la que se une (si no están todas ocupadas) y después el tipo de enlace y el número de residuos en la molécula, por ejemplo un residuo de β1,3Gal unido a la rama 1-3 se representa como [3]G(3)1 (Figura 4D).

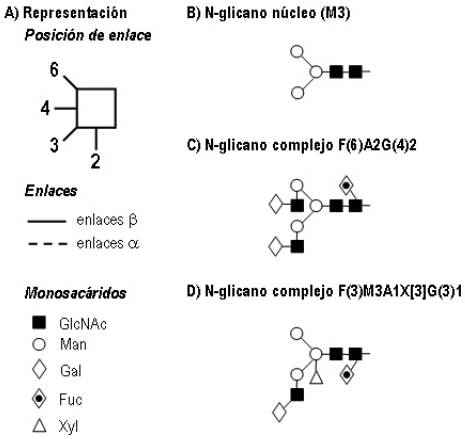


Figura 4. Representación esquemática de las estructuras en la GlycoBase.
A) Elementos y simbología empleadas para la representación de los *N*-glicanos;
B) Representación esquemática y nomenclatura del *N*-glicano núcleo;
C) Representación esquemática y nomenclatura de un *N*-glicano típico de mamíferos;
D) Representación esquemática y nomenclatura de un *N*-glicano típico de plantas.

2.4 Cromatografía de intercambio aniónico débil para simplificar la complejidad de las muestras de *N*-glicanos.

Cuando nos enfrentamos a muestras complejas como pudieran ser las provenientes del suero humano, estaremos en presencia de multiplicidad de estructuras de *N*-glicanos neutros o cargados. Según el principio de separación la cromatografía de fase normal tipo HILIC no es capaz de discriminar entre *N*-glicanos neutros y cargados pues la elución ocurre en función de la hidrofilicidad total del oligosacárido en cuestión, es decir que en una misma zona del cromatograma pueden estar eluyendo compuestos neutros y cargados. Para simplificar esta complejidad se han introducido pasos de fraccionamiento en función del número de cargas de los *N*-glicanos a partir de la separación en columnas de DEAE-HPLC o equivalentes lo que comúnmente se ha denominado WAX-chromatography del término en Inglés. Como estándar de la separación por grupos de *N*-glicanos se ha utilizado la mezcla de *N*-glicanos provenientes de la fetuina, lo que permite calibrar la separación en oligosacáridos neutros, mono, di, tri y tetrasialilados (Figura 5A). Cada fracción de intercambio iónico se analiza nuevamente en HPLC-Amida 80 lo que permite una simplificación del análisis de mezclas complejas de *N*-glicanos neutros y sialilados (Figura 5B).

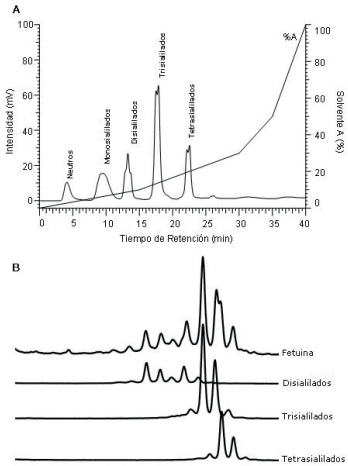


Figura 5. Separación de *N*-glicanos de fetuina bovina en columna de intercambio aniónico DEAE-HPLC, 100 x 7.6 mm. A) Los solventes empleados fueron A: formiato de amonio 0.5M pH 9 y B: Metanol 10% en gradiente según: condición inicial 0% A; t=15 min 10% A; t=30 min 30%A; t=35 min 50% A; t=40 min 100% A; t= 45-50 min 0% A. Los picos del cromatograma representan las fracciones de *N*-glicanos neutros, mono, di, tri, y tetrasialilados. B) Las fracciones correspondientes a los *N*-glicanos di, tri y tetrasialilados se analizaron en cromatografía de fase normal HPLC-Amida 80 como se muestra en la Figura 3.

2.5 Otros métodos de perfil cromatográfico.

Otro método de perfil cromatográfico conocido como mapa bi-tridimensional (2D-3D) de derivados de 2-aminopiridina (2-AP) se basa (2D) en la utilización de dos columnas cromatográficas esencialmente octadecil (ODS)-sílica (interacción hidrofóbica) y una columna amida (interacción hidrofílica) [30]. El análisis de mezclas de oligosacáridos neutros y cargados mediante estas dos separaciones puede generar resultados ambiguos que se resuelve cuando se incorpora una tercera columna (3D) basada en intercambio iónico (dietilaminoetil o DEAE) para separar las fracciones en cuanto a su carga [31]. Con este método se obtiene una adecuada resolución de los oligosacáridos pero tiene la desventaja de requerir de tres corridas cromatográficas en HPLC.

También se ha descrito un método de perfil electroforético basado en la derivatización de los *N*-glicanos por aminación reductiva con el fluoróforo ácido 8-amino-1,3,6-naftalen trisulfónico (ANTS) para su separación en electroforesis (gel o capilar) comercialmente conocido como electroforesis de carbohidratos asistida por fluoróforo (FACE) por su sigla en Inglés. Posteriormente se desarrolló un sistema que combina la separación en HPLC (columna amino-NH_2) y electroforesis. Se obtiene como resultado un mapa bidimensional de tiempo de retención (HPLC) vs movilidad electroforética (FACE) que posee un alto carácter predictivo [32]. Sin embargo, estos derivados ANTS tienen limitaciones en cuanto a la ionización en espectrometría de masas tanto por ionización por electronebulización (ESI) como por láser asistido por matriz (MALDI).

2.6 Secuenciación de *N*-glicanos por digestiones secuenciales con exoglicosidasas.

La información obtenida a partir de los métodos de perfil cromatográfico se complementa con digestiones con exoglicosidasas específicas [33]. El amplio conocimiento que se tiene en la actualidad sobre la especificidad de las exoglicosidasas permite determinar, de manera directa, la naturaleza de los monosacáridos removidos y el tipo de enlace. Inicialmente se hacía necesario aislar los picos cromatográficos que serían sometidos a las digestiones con las mezclas de exoglicosidasas, sin embargo en la actualidad es práctica común realizar las digestiones sobre la mezcla total de *N*-glicanos aislados de la muestra en estudio y su posterior análisis en HPLC. A parir de la comparación de los valores de GU, de cada fracción antes y después de digestión con exoglicosidasas y conociendo la especificidad de las enzimas empleadas permite arribar a conclusiones acerca de los monosacáridos removidos según su enlace y configuración (Figura 6). El éxito de esta metodología en gran medida radica en el correcto diseño de las mezclas de exoglicosidasas y su orden secuencial a la hora de realizar las digestiones. Las enzimas más empleadas son las que aparecen en la Tabla 1.

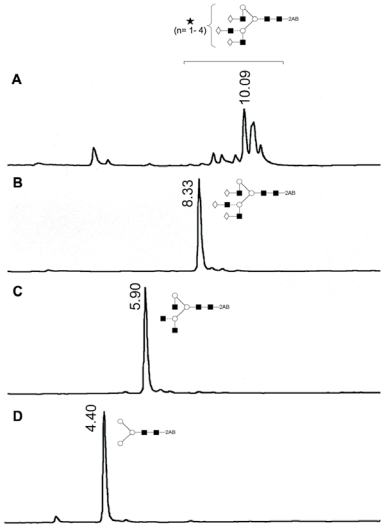


Figura 6. Digestión secuencial de la mezcla total de *N*-glicanos de fetuina bovina. (A) Mezcla total de *N*-glicanos de fetuina bovina sin digerir. (B) Mezcla total de *N*-glicanos digerida con ABS: sialidasa. (C) Mezcla total de *N*-glicanos digerida con ABS+BTG: sialidasa + betagalactosidasa. (D) Mezcla total de *N*-glicanos digerida con ABS+BTG+SPH: sialidasa + betagalactosidasa + *Streptococcus pneumoniae* -*N*-acetylhexaminidase. La digestión con ABS genera un pico de GU correspondiente a la triantena trigalactosilada, ABS+BTG genera un pico de GU correspondiente a la triantena agalactosilada y ABS+BTG+SPH genera un pico correspondiente al núcleo del *N*-glicano.

Tabla 1. Exoglicosidasas más empleadas, fuente, tampón de reacción y concentración final en la reacción.

Enzima	Tampón	% volumen final
ABS Arthrobacter sialidase	50 mM fosfato de sodio pH 6.0	10%
BKF Bovine kidney alpha-fucosidase	100 mM citrato de sodio pH 6.0	10%
BTG Bovine testis beta-galactosidase	100 mM citrato-fosfato de sodio pH 5.0	20%
SPH Streptococcus pneumoniae -N-acetyl-hexaminidase	100 mM citrato-fosfato de sodio pH6.0	20%
JBM Jack bean alpha-mannosidase	100 mM acetato de sodio, 2 mM Zn, pH 5.0	40% + 40% después de 18 hrs de incubación

Adicionalmente, la interpretación de los resultados de las digestiones con exoglicosidasas por espectrometría de masas se ha convertido en una herramienta muy útil en la dilucidación estructural de los *N*-glicanos presentes en una mezcla compleja [34].

2.7 Composición monosacarídica y estructura a partir de espectrometría de masas.

La espectrometría de masas mediante ionización por láser asistida por matriz (MALDI-MS) y electronebulización (ESI-MS), es de los métodos analíticos el más utilizado en el análisis de glicosilación pues brinda información sobre la masa del oligosacárido, su composición en monosacáridos (hexosas, hexosaminas y deoxihexosas) y además detalles estructuras de la molécula. MALDI-MS con analizador de tiempo de vuelo (TOF) brinda información relacionada con la composición de una mezcla de *N*-glicanos liberados de una glicoproteína en cuestión o de mezcla de ellas. También es aplicable tanto a mezcla de *N*-glicanos libres o marcados con fluoróforos aunque la práctica experimental se inclina a favor del análisis de los *N*-glicanos no marcados. Comúnmente la medición se realiza en modo positivo donde se favorece la formación de los iones $[M+Na]^+$. Si de la mezcla en análisis se conoce o presume la presencia de ácidos siálicos terminales en alguno de los *N*-glicanos se recomienda realizar la medición en modo negativo o la derivatización por permetilación de todos los grupos hidroxilos o esterificación de los grupos carboxilos de la muestra para evadir la inestabilidad del ácido siálico durante la ionización y evitar las fragmentaciones que hacen más compleja la interpretación de los resultados. ESI-MS aparentemente es otra opción para el análisis de mezclas de *N*-glicanos aún para aquellos de menor masa, sin embargo es común observar la co-existencia de iones simple, doble y triple cargados en el mismo espectro lo que realmente dificulta el análisis. En modo negativo para *N*-glicanos neutros la formación de aductos de aniones en ESI-MS es de gran utilidad ($[M+Cl]^-$, $[M+NO_3]^-$, $[M+H_3PO_4]^-$), predominando la formación de iones del tipo $[M+nH]^{n-}$ para los N-glicanos sialilados. Las fragmentaciones de estos aductos en MS/MS brindan muchísima más información que la que se obtiene en MALDI por PSD (post-source decay de la sigla en Inglés).

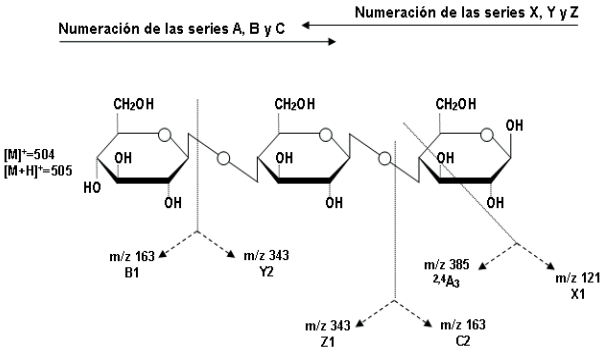


Figura 7. Representación de las fragmentaciones en espectrometría de masas de los *N*-glicanos. Nomenclatura empleada para identificar las fragmentaciones, descrita por Domon y Costello (1988) [37].

Las fragmentaciones de los carbohidratos en espectrometría de masas involucran la ruptura de uno o varios enlaces entre los monosacáridos adyacentes y también diversas fragmentaciones de los anillos tanto en el extremo reductor como no reductor, las que proporcionan información estructural de la molécula involucrada. La ruptura del enlace glicosídico brinda información de la secuencia y el número de ramificaciones, pero no de la posición del enlace. Mientras, las fragmentaciones que involucran la ruptura del anillo son muy útiles para determinar la posición del enlace entre monosacáridos adyacentes. La nomenclatura empleada para identificar las fragmentaciones fue descrita por Domon y Costello [35] (Figura 7A).

Existen dos series de iones A-B-C y X-Y-Z. Los iones X se generan como resultado de la ruptura de un anillo y contienen el extremo reductor del *N*-glicano, los iones Y y Z se forman por la ruptura de enlaces glicosídicos. Los iones fragmentos que se forman por estas mismas rupturas y que contienen el terminal no reductor se designan como A (ruptura del anillo), B y C (ruptura de enlaces glicosídicos). Para identificar la posición del residuo donde ocurre la escisión se utilizan subíndices y supraíndices para indicar los enlaces que se rompen en caso de que ocurra ruptura del anillo.

Los espectros de los *N*-glicanos no derivatizados aunque son aparentemente más difíciles de interpretar, los rendimientos de los iones fragmentos suelen ser más elevados debido a la fácil migración de los átomos de hidrógeno de los grupos hidroxilos durante la ruptura los enlaces glicosídicos (entre monosacáridos) [36]. La ionización de los *N*-glicanos por MALDI-MS produce predominantemente iones $[M+Na]^+$ [37], que bajo condiciones de baja energía, generan productos de la ruptura de enlaces glicosídicos y una baja proporción de iones producidos por rupturas de los anillos de azúcares [38]. Estas últimas fragmentaciones son esenciales para determinar el tipo de enlace entre los monosacáridos.

En ESI-MS se producen varios tipos de iones dependiendo de las condiciones de las fuentes de ionización y del empleo de aditivos en los solventes. Las fragmentaciones MS/MS de iones negativos son una potente metodología para el análisis estructural de los *N*-glicanos neutros y sialilados debido a que cada característica estructural genera un ion fragmento específico generalmente producido por mecanismos que no compiten entre sí, como ocurre en la disociación de iones positivos. Las fragmentaciones que se inducen por la adición de iones negativos producen abundantes rupturas de anillos, información importante para conocer las características estructurales de los *N*-glicanos que son difíciles de identificar por técnicas tradicionales [39]. Los *N*-glicanos generan fragmentos del tipo $^{2,4}A_R$ de alta intensidad, a partir de la GlcNAc del extremo reductor e involucra la pérdida de 161 unidades de masa, que no debe confundirse con la masa de una hexosa (162 unidades) [40,41] (Figura 8). Si el *N*-glicano en estudio presenta un sustituyente en el átomo de carbono de la posición 3 entonces esta fragmentación no ocurre pues no es posible la estabilización de la carga negativa localizada en el oxígeno del grupo hidroxilo correspondiente a esta posición y la ionización predominante pasa a ser la del ion $^{2,4}A_{R-1}$. Aunque el método de fragmentación de iones negativos es compatible

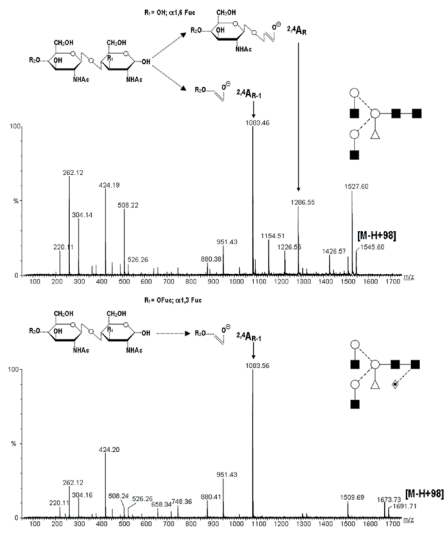


Figura 8. Mecanismo propuesto para la formación del ión $^{2,4}A_R$ [M-161]. En la parte superior se muestra el espectro MS-MS donde aparecen tanto $^{2,4}A_R$ como $^{2,4}A_{R-1}$. En la parte inferior se muestra el espectro MS-MS donde no ocurre la fragmentación $^{2,4}A_R$. La diferencia observada corresponde con la presencia o no de una Fucosa unida al grupo hidroxilo del carbono 3 de la GlcNAc del extremo reductor.

con el análisis de *N*-glicanos derivatizados en el extremo reductor, con fluoróforos neutros como 2AB empleados para la detección cromatográfica, estos derivados presentan el mencionado efecto perjudicial en el análisis. Cuando el extremo reductor del *N*-glicano está sustituido, como pudiera ser por la formación de derivados 2AB, el anillo de la primera GlcNAc se abre durante la reacción de derivatización lo que evita el reordenamiento de los electrones de las estructuras cíclicas. Esta fragmentación es muy útil para confirmar la naturaleza de los aductos de iones negativos y la presencia o ausencia de la fucosa enlazada a la GlcNAc del extremo reductor del *N*-glicano y la posición del enlace Fuc-GlcNAc [23, 41, 42].

El registro de los espectros de masas MALDI o ESI requiere de mayores niveles de pureza que la que normalmente se requieren en HPLC. En cuanto a los límites de detección las mediciones en HPLC de derivados fluorescentes sobrepasa con creces los de espectrometría de masas pero en términos de resolución la espectrometría de masas es muy superior al HPLC. Solamente en casos de compuestos isobáricos se pudieran alcanzar mejores resoluciones en HPLC. En términos de pureza para lograr cristalizaciones eficientes y por ende mejores ionizaciones MALDI es menos exigente que ESI. No obstante se ha podido determinar que algunos componentes generalmente presentes en los procedimientos analíticos tales como iones fosfato y Tris deben estar en concentraciones finales por debajo de 20 mM y el Glicerol no debe superar el 1% (algunas preparaciones de PNGasa F vienen en presencia de Glicerol para garantizar una mayor estabilidad por lo que se recomienda utilizar las preparaciones libres de Glicerol). El SDS interfiere grandemente por lo que el límite aceptable es de 0.01%, otros detergentes también deben estar en el rango 0.1-0.01%.

Para obtener espectros ESI con calidad satisfactoria las purezas requeridas son aún más exigentes y los límites de presencia de sales y detergentes son aún menores que en MALDI. Los espectros ESI de los *N*-glicanos se registran en mezclas de agua:metanol o acetonitrilo. Para favorecer la formación de los iones $[M+H]^+$ o $[M+Na]^+$ en modo positivo se recomienda la adición de ácido fórmico (0.1-1.0%) o concentraciones de sales de metales alcalinos 50 mM respectivamente. Por el contrario para la formación de iones $[M-H]^-$ se requiere de la adición de hidróxido de amonio, si se prefiere la formación de otros aductos estos se logran por la adición de sales de aniones como cloruro, nitrato y fosfato entre otros.

La remoción de detergentes y sales provenientes de los tampones de reacción o de las bandas de gel de poliacrilamida, si la reacción de desglicosilación se llevó a cabo en el gel, se convierte en uno de los pasos esenciales para lograr espectros de masas de buena calidad. Se han ensayado varios métodos pero la purificación en columnas de carbón grafitada parece ser la de mayor eficiencia aunque no se excluye el uso de mini-columnas de fase invertida (C18) o la combinación de ambas. Es recomendable la purificación de los *N*-glicanos no marcados pues la pérdida en las columnas de carbón grafitada es mínima, en esta cromatografía los *N*-glicanos más cargados eluirán a mayores concentraciones de acetonitrilo y ácido trifluoracético. En el caso de las mini-columnas de C18 es aplicable tanto para marcados o no marcados con buenos porcientos de recuperación. Finalmente como procedimiento para eliminar residuos de sales e incluso para favorecer la formación de los aductos deseados se recomienda la diálisis en membranas de alta hidrofobicidad como es el caso del Nafión 117 suministrado por Sigma-Aldrich, precisamente por esta alta hidrofobicidad de la membrana no se recomienda para *N*-glicanos marcados con 2AB. Este procedimiento está diseñado para microdiálisis de escasos microlitros sobre un pequeño pedazo de membrana que flota sobre una gota de la solución dializante, por tanto se recomienda como último paso antes del registro de los espectros de masas ya sea MALDI- o ESI-MS [43].

REFERENCIAS.

1. Herscovics A., Orlean P. Glycoprotein biosynthesis in yeast. FASEB J. 1993; 7, 540-550.

2. Roth J. Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. Chem Rev. 2002; 102, 285-303.

3. Kobata A. Structure and function of the sugar chains of glycoproteins. Eur. J. Biochem. 1992; 209, 483-501.

4. Bishop. LA, Nguyen TV, Schofield PR. Both of the beta-subunit carbohydrate residues of follicle-stimulating hormone determine the metabolic clearance rate and in vivo potency. Endocrinology 1995; 136, 2635-2640.

5. Gething MJ, Sambrook J. Protein folding in the cell. Nature 1992; 355, 33-45.

6. Parodi A. Protein glucosylation and its role in protein folding. Ann Rev Biochem. 2000; 69, 69-93.

7. Helenius A. Intracellular Functions of N-Linked Glycans. Science 2001; 291, 2364-2369.

8. Ziegler FD, Maley F, Trimble RB. Characterization of the glycosylation sites in yeast external invertase. J. Biol. Chem. 1988; 263, 6986-6992.

9. Imperiali B, O'Connor SE. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. Curr. Opin. Chem. Biol. 1999; 3, 643–649.

10. Goochee CF, Monica T. Environmental effects on protein glycosylation. Biotechnology 1990; 8, 421-427.

11. Parekh RB, Dwek RA, Sutton BJ, Fernandes DL, Leung A, Stanworth D, Rademacher TW, Mizuochi T, et al. (1985) Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. Nature 316, 452-457.

12. Jaeken J. Congenital disorders of glycosylation (CDG): update and new developments. J Inherit Metab Dis. 2004; 27, 423-426.

13. Freeze HH, Aeb M. Altered glycan structures: the molecular basis of congenital disorders of glycosylation. Curr Opin Struct Biol. 2005; 15, 490-498.

14. Butler M, Quelhas D, Critchley AJ, Carchon H, Hebestreit HF, Hibbert RG, et al. Detailed glycan analysis of serum glycoproteins of patients with congenital disorders of glycosylation indicates the specific defective glycan processing step and provides an insight into pathogenesis. Glycobiology 2003; 13, 601-622.

15. Gornik O, Lauc G. Glycosylation of serum proteins in inflammatory diseases. Dis Markers 2008; 25, 267-278.

16. Comunale MA, Lowman M, Long RE, Krakover J, Philip R, Seeholzer S, Evans AA, et al. Proteomic analysis of serum associated fucosylated glycoproteins in the development of primary hepatocellular carcinoma. J Proteome Res. 2006; 5, 308-315.

17. Hakomori S. Tumor-associated carbohydrate antigens defining tumor malignancy: basis for development of anti-cancer vaccines. Adv Exp Med Biol. 2001; 491 369-402.

18. Peracaula R, Barrabes S, Sarrats A, Rudd PM, de Llorens R. Altered glycosylation in tumours focused to cancer diagnosis. Dis Markers 2008; 25, 207-218.

19. Raju TS. Terminal sugars of Fc glycans influence antibody effector functions of IgGs. Curr Opin Immunol. 2008; 20, 471-478.

20. Merry AH, Neville DC, Royle L, Matthews B, Harvey DJ, Dwek RA, Rudd PM. Recovery of intact 2-aminobenzamide-labeled O-glycans released from glycoproteins by hydrazinolysis. Anal Biochem. 2002; 304, 91-99.

21. Carlson DM. Oligosaccharides isolated from pig submaxillary mucin. J. Biol. Chem. 1966; 241, 2984–2986.

22. Huang Y, Mechref Y, Novotny MV. Microscale nonreductive release of O-linked glycans for subsequent analysis through MALDI mass spectrometry and capillary electrophoresis. Anal Chem. 2001; 73, 6063-6069.

23. Triguero A, Cabrera G, Royle L, Harvey DJ, Rudd PM, Dwek RA, et al. Chemical and enzymatic N-glycan release comparison for N-glycan profiling of monoclonal antibodies expressed in plants. Anal Biochem 2010; 400, 173-183.

24. Guile GR, Rudd PM, Wing DR, Prime SB, Dwek RA. A rapid high-resolution high-performance liquid chromatographic method for separating glycan mixtures and analyzing oligosaccharide profiles. Anal Biochem. 1996; 240, 210-226.

25. Rudd PM, Colominas C, Royle L, Murphy N, Hart E, Merry AH, et al. A high-performance liquid chromatography based strategy for rapid, sensitive sequencing of N-linked oligosaccharide modifications to proteins in sodium dodecyl sulphate polyacrylamide electrophoresis gel bands. Proteomics 2001; 1, 285-294.

26. Takahashi N, Nakagawa H, Fujikawa K, Kawamura Y, Tomiya N. Three-dimensional elution mapping of pyridylaminated N-linked neutral and sialyl oligosaccharides. Anal Biochem. 1995; 226, 139-146.

27. Tomiya N, Takahashi N. Contribution of component monosaccharides to the coordinates of neutral and sialyl pyridylaminated N-glycans on a two-dimensional sugar map. Anal Biochem. 1998; 264, 204-210.

28. Royle L, Campbell MP, Radcliffe CM, White DM, Harvey DJ, Abrahams JL, et al. HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. Anal Biochem. 2008; 376, 1-12.

29. Harvey DJ, Merry AH, Royle L, Campbell MP, Dwek RA, Rudd PM. Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. Proteomics 2009; 9, 3796-3801.

30. Takahashi N, Wada Y, Awaya J, Kurono M, Tomiya N. Two–elution map of GalNAc-containing N-linked oligasaccharides. Anal. Biochem. 1993; 208, 96-109.

31. Takahashi N. Three-dimensional of N-Linked oligosaccharides using anion-axchange, hydrophobic and hydrophilic interaction modes of high-performance liquid chromatography. *J. Chromat.* 1996; 760, 217-225.

32. Quintero O, Montesino R, Cremata JA. Two-dimensional mapping of 8-Amine 1,3,6-Naphthalen Trisulfonic Acid derivates of N-linked neutral and sialyloligosaccharides. Anal. Biochem. 1998; 256, 23-32.

33. Rudd PM, Guile GR, Kuster B, Harvey DJ, Opdenakker G, Dwek RA. Oligosaccharides sequencing technology. Nature 1997; 338, 205-207.

34. Harvey DJ. Identification of protein-bound carbohydrates by mass spectrometry. Proteomics 2001; 1, 311-328.

35. Domon B, Costello C.E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. Glycoconjugate J. 1988; 5, 397-409.

36. Harvey DJ. Collision-induced fragmentation of underivatized N-linked carbohydrates ionized by electrospray. J Mass Spectrom. 2000; 35, 1178-1190.

37. Harvey DJ. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. Mass Spectrom Rev. 1999; 18, 349-450.

38. Orlando R., Bush CA, Fenselau C. Structural analysis of oligosaccharides by tandem mass spectrometry: Collisional activation of sodium adducts ions. Biomedical and Environmental Mass Spectrometry. 1990; 19, 747-754.

39. Wheeler SF, Harvey DJ. Negative ion mass spectrometry of sialylated carbohydrates: discrimination of N-acetylneuraminic acid linkages by MALDI-TOF and ESI-TOF mass spectrometry. Anal Chem. 2000; 72, 5027-5039.

40. Harvey DJ. Fragmentation of negative ions from carbohydrates: part 2. Fragmentation of high-mannose N-linked glycans. J Am Soc Mass Spectrom. 2005a; 16, 631-646.

41. Harvey DJ. Fragmentation of negative ions from carbohydrates: part 3. Fragmentation of hybrid and complex N-linked glycans. J Am Soc Mass Spectrom. 2005b; 16, 647-659.

42. Séveno M, Cabrera G, Triguero A, Burel C, Leprince J, Rihouey C, et al. Plant N-glycan profiling of minute amount of material, Anal. Biochem. 2008; 379, 66-72.

43. Börsen KO, Mohr MD, Widmer HM. Ion Exchange and purification of carbohydrates on a Nafion membrane as a new sample pretreatment for Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. Rap. Com. Mass Spect. 1995; 9, 1031-1034.

Fosfoproteómica. Técnicas de enriquecimiento de fosfopéptidos. Métodos espectrométricos para la identificación de fosfopéptidos

Montserrat Carrascal, Joaquin Abian y Marina Gay[1]

[1] Laboratorio de Proteómica CSIC/UAB, ProteoRed, Instituto de Investigaciones Biomédicas de Barcelona. Consejo Superior de Investigaciones Científicas. Barcelona. Spain.

1.- Introducción.

La fosforilación de proteínas es uno de los principales mecanismos para la regulación de la función celular en eucariotas. Una compleja red de quinasas y fosfatasas controla prácticamente todos los aspectos de la vida celular mediante la fosforilación o defosforilación selectiva de las proteínas celulares [1, 2]. Se estima que entre un 2 y un 5% del genoma humano codifica este tipo de enzimas, existiendo, solo para quinasas, más de 500 genes diferentes [3]. Consecuentemente, la fosforilación proteica es la modificación postraduccional (PTM) reversible más frecuente. En el proteoma de mamíferos se han predicho más de 100.000 puntos de fosforilación potenciales [4] y se estima que entre el 30-50% de las proteínas expresadas en células eucariotas pueden estar fosforiladas en un momento dado.

Se conocen nueve posibles dianas de fosforilación en los seres vivos: tirosina, serina, treonina, histidina, ácido aspártico, ácido glutámico y, menos frecuentemente, lisina, arginina y cisteína (O-fosforilación Ser/Thr/Tyr, N-fosforilación His/Lys, acyl-fosforilación Glu/Asp y S-fosforilación Cys, Figura 1). En células eucariotas, sin embargo, la fosforilación se produce casi exclusivamente en los aminoácidos Ser, Thr y Tyr, si bien con grandes diferencias en su abundancia relativa (1800:200:1, respectivamente en vertebrados [5]).

Los procesos de fosforilación-defosforilación proteica determinan las características funcionales de muchas proteínas y juegan un papel primordial en mecanismos como la transducción y amplificación de señales en la célula o la regulación del ciclo celular. Esta modificación postraduccional puede modular la actividad de una proteína induciendo cambios conformacionales que exponen sus sitios activos o bien determinar la vida media de ésta. En otros casos, pueden ser el primer paso para el reclutamiento de proteínas en complejos proteicos funcionales o para dirigir su localización subcelular.

Los procesos de fosforilación-defosforilación no actúan simplemente como interruptores de encendido/apagado de la función proteica. La utilización de diferentes puntos de fosforilación por la misma o distintas quinasas/fosfatasas es un mecanismo común para incrementar su capacidad de regulación. Además, una misma enzima puede fosforilar proteínas diferentes activando unas e inactivando otras. Estos mecanismos permiten el control de diferentes actividades en una

misma proteína así como la integración de diferentes señales en la célula permitiendo que agonistas diferentes muestren comportamientos sinérgicos o antagónicos en función de la situación [6].

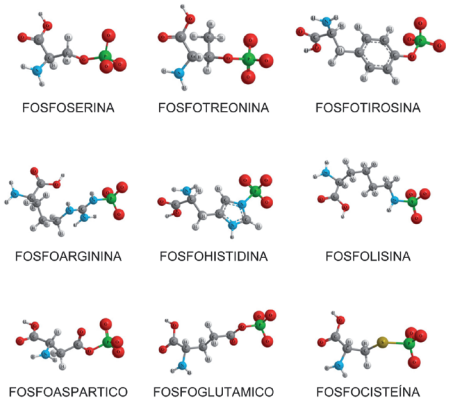


Figura 1: Estructuras químicas de los aminoácidos fosforilados. La modificación más frecuente en eucariotas es la O-fosforilación aunque también se han detectado N-fosforilaciones en Lys, His y Arg [86]. La inestabilidad de la N-fosforilación ha hecho pensar que su proporción ha sido subestimada. Algunos estudios indican que la fosforilación en His puede ser de hasta el 6% del total si bien su detección es difícil debido a su rápida hidrólisis en condi-ciones ácidas [87].

El estado de fosforilación de las proteínas viene determinado por la actividad conjunta de fosfatasas y quinasas que actúan en un equilibrio altamente regulado y cuya alteración es la base de diversas patologías [7, 8]. Muchos cánceres tienen en común la presencia de tirosin-quinasas (PTKs) sobreexpresadas o mutadas a formas constitutivamente activas [9], siendo la oncología la indicación de todos los fármacos inhibidores de quinasas aprobados. Actualmente, existe además un número creciente de moléculas que se están probando en ensayos clínicos de enfermedades autoinmunes como la artritis reumatoide o la psoriasis [10]. Un fármaco que ya se encuentra en un estadio muy avanzado de la fase clínica es el Tideglusib, un inhibidor de la enzima GSK3. que actúa sobre los dos principales procesos causantes de la enfermedad de Alzheimer: la hiperfosforilación de la proteína tau y la agregación del péptido beta amiloide [11].

La fosfoproteómica es la rama de la proteómica que aborda el estudio de este conjunto de proteínas fosforiladas o fosfoproteoma. Los objetivos principales de la fosfoproteómica incluyen la identificación de fosfoproteínas y sus puntos de fosforilación, la identificación de las quinasas y fosfatasas implicadas en su regulación y, en último término, la determinación de las consecuencias funcionales de la fosforilación en el contexto del sistema biológico estudiado.

Las principales dificultades en el análisis global del fosfoproteoma derivan de la baja concentración de sus componentes y de la complejidad resultante de los diversos grados de fosforilación en los que una misma especie proteica puede encontrarse en un momento dado en la célula. Además, estas modificaciones pueden ser temporales y variar de forma rápida afectando solo a una pequeña población de todas las proteínas de una misma especie molecular. Estas características exigen la utilización de procedimientos altamente sensibles, específicos y selectivos para el análisis del fosfoproteoma.

La técnica de referencia para el análisis global del fosfoproteoma fue durante muchos años la electroforesis en gel bidimensional (2DE) combinada con el radiomarcaje con ^{32}P o ^{33}P o la inmunodetección. Siguiendo esta estrategia, la cuantificación de los cambios en los componentes del fosfoproteoma se lleva a cabo mediante programas de análisis de imagen sobre las correspondientes autoradiografías o Western-blots. Las fosfoproteínas detectadas en los geles pueden caracterizarse posteriormente mediante su recorte del gel, su digestión *in situ* mediante enzimas específicas y el análisis de los péptidos de digestión mediante la secuenciación automática de Edman o, más recientemente, por espectrometría de masas (MS) mediante procedimientos de mapeo de masas peptídicas (PMF) y de secuenciación por espectrometría de masas en tándem (MS^n).

Estos procedimientos tienen el inconveniente de que la localización de las proteínas fosforiladas en el gel se basa en métodos de visualización altamente específicos que no detectan la presencia de otras proteínas, quizás más abundantes, que colocalizan en el gel y que interfieren

o impiden la identificación de la proteína diana. Para evitar estas interferencias y aumentar la concentración de las moléculas fosforiladas de forma que sean accesibles al análisis espectrométrico, se recurre a métodos de purificación por afinidad de los extractos proteicos.

Debido a las dificultades expuestas, y a pesar de que se calcula que existen más de 100.000 puntos de fosforilación en el proteoma humano, solo una pequeña parte de éstos han sido descritos, la mayoría, muy recientemente [5, 12-15]. El rápido avance conseguido en los últimos años ha estado muy ligado a la mejora en las características de las tecnologías disponibles, especialmente al continuo desarrollo de los sistemas de análisis de masa (analizadores de trampa lineal, Orbitrap, FT-ICR) y de métodos de secuenciación (fragmentación por ETD/ECD) cada vez más sensibles y con mayor resolución [16, 17]. Probablemente, el primer repertorio de puntos de fosforilación humanos descrito a gran escala corresponde al del fosfoproteoma de células HeLa elaborado por el grupo de Mann [5] incluido en la base de datos (DB) Phosida [18] y que contiene alrededor de 6.000 puntos de fosforilación experimentales. Existen además otras DBs de proteínas que incluyen datos sobre fosforilación y que son accesibles a través de la red como PhosphoEML (43.000 puntos de fosforilación de eucariotas, [19]), Phosphosite (240.000 en vertebrados, www.phosphosite.org), PhosphoPep (que inicialmente contenía datos de Drosophila y que se ha ampliado con secuencias humanas, de *S.cerevisae* y de *C.elegans*, [20]), HPRD (8.160 puntos de fosforilación en proteínas humanas) y LymPHOS (8.300 en células T primarias humanas, [21]). Aparte de estas DBs fosfoproteómicas, las principales DBs generales, como Uniprot o PPD (*plasma proteome database*), contienen también anotaciones sobre modificaciones postraduccionales incluyendo datos empíricos extraídos de la literatura y puntos potenciales de fosforilación.

2.- Métodos de enriquecimiento de fosfopéptidos y fosfoproteínas.

Como ya se ha comentado, la caracterización del fosfoproteoma presenta diversas dificultades derivadas de la baja concentración relativa de sus componentes, de la coexistencia de formas fosforiladas y no fosforiladas de una misma molécula con proporciones muy bajas de las primeras y de la complejidad de los perfiles de fosforilación de una misma proteína donde pueden existir distintos sitios de fosforilación con diversos grados de modificación.

Con objeto de aumentar la concentración de las moléculas fosforiladas se utilizan diversos métodos de enriquecimiento. Los métodos más conocidos incluyen la inmunoafinidad, la cromatografía de afinidad y la cromatografía líquida, esta última empleando principalmente soportes para intercambio iónico o para cromatografía hidrofílica (HILIC). Se han desarrollado además diversos métodos que utilizan reacciones químicas para la fijación específica y reversible de los fosfopéptidos a superficies (bolas, polímeros) que permiten su posterior aislamiento. Otros procedimientos de purificación como la precipitación con fostato de calcio [22] o el COFRADIC [23] también se han utilizado para el aislamiento de péptidos fosforilados si bien son métodos menos extendidos.

2.1. Métodos basados en inmunoafinidad.

Los métodos basados en inmunoafinidad han sido clásicamente los más utilizados para enriquecer fosfoproteínas o fosfopéptidos a partir de extractos proteicos. Estos métodos utilizan anticuerpos anti-fosfotirosina, fosfoserina o fosfotreonina, (anti-pTyr, -pSer y -pThr, respectivamente) para enriquecer fosfoproteínas y fosfopéptidos de forma muy específica. Estos anticuerpos se pueden inmovilizar previamente a un soporte sólido, como agarosa o bolas poliméricas (método directo), o se pueden añadir solubles a mezclas complejas de proteínas

o de péptidos de digestión para ser capturados posteriormente mediante su unión a proteína A/G o a otros anticuerpos ligados a un soporte (método indirecto). Tras la inmunoprecipitación, el extracto enriquecido en fosfoproteínas puede ser separado utilizando electroforesis en gel como paso previo a su digestión en gel o puede ser digerido directamente con un enzima específico. En ambos casos, los péptidos trípticos se analizan por espectrometría de masas para su caracterización.

El uso de anticuerpos anti-pTyr se describió hace ya casi 30 años [24] y se han realizado numerosos estudios fosfoproteómicos a gran escala utilizando esta aproximación [25-27]. La utilización de esta estrategia para la purificación de fosfopéptidos y la determinación directa de sus puntos de fosforilación, fue reportada por el grupo de Vandekerckhove en 1999 en un trabajo en el que describieron los puntos de fosforilación inducidos in-vitro por c-Scr en la Gelsolina [28]. Actualmente la utilización de anticuerpos anti-pTyr es el método utilizado para la caracterización de la fosforilación en tirosinas mediante técnicas proteómicas.

Si bien se dispone de anticuerpos anti-pTyr comerciales muy eficientes y de aplicación general desde hace tiempo, este no ha sido el caso de los correspondientes anticuerpos anti-pSer y anti-pThr. Probablemente debido a su baja inmunogeneicidad, los anticuerpos generados contra estos fosfoaminoácidos mostraban una afinidad baja y dependiente de secuencia. Aunque se ha descrito la utilización de anticuerpos contra proteínas específicas (p.e. las PKB, PKC [29] y GSK3 [30]) para la caracterización de nuevos substratos de fosforilación en células aisladas aprovechando las reacciones cruzadas de estos anticuerpos, la utilidad de estas estrategias para la purificación genérica de péptidos o proteínas conteniendo pSer o pThr es muy limitada.

La no disponibilidad de anticuerpos específicos de fosfoaminoácido retrasó de esta forma el estudio de estas fosfoproteínas y de sus correspondientes quinasas y fosfatasas. Más recientemente se ha demostrado que es posible obtener anticuerpos genéricos utilizando librerías peptídicas degeneradas como inmunógenos [31] y hoy en día se dispone de anticuerpos comerciales específicos de fosfoserina y fosfotreonina que han mostrado ser útiles para su detección y purificación en estudios proteómicos [27]. Estas estrategias sin embargo no son actualmente de uso rutinario en el análisis de fosfoproteómica probablemente debido al éxito alcanzado por técnicas alternativas de purificación, como la cromatografía de afinidad por metales, desarrolladas y comercializadas paralelamente a las primeras.

2.2. IMAC (Immobilized Metal Affinity Chromatography).

El IMAC fue introducido por Porath et al en 1975 para la purificación de proteínas con colas de histidina [32] y es actualmente la técnica más utilizada para el enriquecimiento de fosfopéptidos y fosfoproteínas. La técnica desarrollada por Porath se basaba en la capacidad de la histidina para formar complejos de coordinación con metales de transición divalentes como Ni^{+2}, Co^{+2}, Zn^{+2} o Mn^{+2} lo que permitía purificar de forma específica proteínas ricas en este aminoácido o a las que se había añadido una secuencia de poli-histidina. Más recientemente, se observó que metales como el Fe^{+3}, Ga^{+3} y Al^{+3} mostraban una fuerte unión a los grupos fosfato de los fosfopéptidos, lo que llevó al desarrollo de diversos soportes para cromatografía de afinidad que permitían el aislamiento selectivo de éstos a partir de mezclas complejas.

En la cromatografía IMAC el metal se encuentra inmovilizado formando complejos por coordinación parcial con quelantes polidentados que a su vez se encuentran unidos a un soporte sólido. Los principales quelantes utilizados son los ácidos nitrilotriacético (NTA), etlilendiaminotriacético (tris(carboxymethyl)ethylenediamine, TED) y iminodiacético (IDA). La capacidad de unión de estos compuestos depende el número de puntos de coordinación libres, siendo las resinas IDA las que tienen una mayor capacidad (Figura 2). La agarosa fue el primer material utilizado como so-porte sólido para fijar estos quelantes, ya que es hidrofílica, mecánica y químicamente estable y relativamente inerte. Otros soportes utilizados son la sílice porosa y no porosa y las partícu-

las basadas en polímeros sintéticos como la celulosa, el poliestireno/divinilbenceno y el po-lihidroximetacrilato.

En la cromatografía IMAC las fosfoproteínas y fosfopéptidos se unen a la fase estacionaria por interacciones electrostáticas del grupo fosfato con los iones metálicos. Estos compuestos pueden ser eluidos posteriormente mediante un gradiente de sal o de pH. El método IMAC, sin embargo, no es específico de compuestos fosforilados. Otros péptidos, particularmente aquellos que contienen cadenas de aminoácidos ácidos (aspártico y glutámico) u otros dadores de electrones como la histidina, son también retenidos por este soporte interfiriendo posteriormente con la detección espectrométrica. Este problema se puede minimizar realizando una esterificación de los grupos carboxilo antes de la purificación [33]. El tipo de quelante y el ión metálico utilizado en el método IMAC determinan en gran manera el comportamiento del soporte [34]. Otros factores que afectan la eficiencia general del método IMAC incluyen el grado de fosforilación del péptido, el pH y temperatura de las solucio-nes y la concentración de sales y péptidos en la muestra.

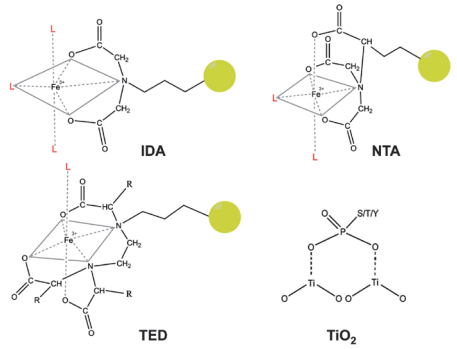


Figura 2: Estructura química de diferentes fases IMAC (IDA, NTA y TED) y del óxido de titanio (TiO_2). IDA: ácido iminodiacético; NTA: ácido nitrilotriacético; TED: etilendiaminotriacético. El NTA es un quelante tetradentado de forma que el complejo metal-NTA, deja 2 de los 6 puntos de coordinación del metal libres para interaccionar con el grupo fosfato. Por su parte el TED es un compuesto pentadentado que deja únicamente un punto de coordinación libre por lo que las resinas basadas en TED tienen una capacidad de unión menor que las de NTA. Finalmente, las resinas basadas en IDA, un quelante tridentado que deja disponibles tres de los puntos de coordinación del metal, son las que proveen de una capacidad de carga mayor.

El IMAC se ha utilizado con éxito en numerosos estudios para el enriquecimiento de fosfopéptidos utilizando columnas cargadas, principalmente, con Fe^{+3} y Ga^{+3}, y tanto de forma off-line como on-line con otras etapas cromatográficas y el análisis espectrométrico. El tipo de adsorbente utilizado ha variado con el tiempo. Los primeros adsorbentes disponibles fueron del tipo NTA o IDA-sílice. La resina más utilizada clásicamente ha sido la resina POROS MC de Perceptive, que es de tipo IDA. Más recientemente la resina Phos-Select, una resina tipo NTA comercializada por Sigma, ha reemplazado a las anteriores en la mayoría de los trabajos. Esta resina viene precar-gada, requiere un tratamiento mínimo y presenta mejor rendimiento y mayor especificidad que las anteriores. La resina Phos-Select no requiere además procedimientos de esterificación previa para eliminar las interferencias de péptidos ácidos. Uno de los primeros trabajos que utilizó este soporte permitió caracterizar más de 5.600 puntos de fosforilación a partir de 10 mg de hígado de ratón y análisis mediante técnicas de *shotgun* proteomics de los fosfopéptidos aislados [15].

En el caso de muestras procedentes de digestión en gel o de fracciones de cromatografía de intercambio iónico la purificación IMAC suele hacerse *off-line* a la separación cromatográfica de sus componentes y al análisis espectrométrico. Para ello se utilizan pequeñas columnas (3-5 mm de longitud) construidas sobre una punta de pipeta plástica. La separación se realiza pasando la solución de péptidos a través de estas columnas a baja presión o bien incubando la muestra con la fase. Una vez lavado el soporte, el material retenido se eluye en un medio muy básico (pH 10-11) y que exige acidificar inmediatamente el eluato para evitar la pérdida del fosfato en Ser y Thr.

En las estrategias *on-line*, el eluato de la columna IMAC pasa directamente a una columna o precolumna de fase reversa donde los péptidos se retienen, se lavan y se cromatografían. De esta forma se minimizan los problemas de pérdida de material que tienen lugar cuando el procedimiento se realiza en diversos pasos. La columna de fase reversa puede estar a su vez conectada a un colector de fracciones o, más comúnmente, a un espectrómetro de masas.

Debido a los grupos fosfato adicionales, algunos fosfopéptidos retenidos en la columna de IMAC son de una polaridad tan alta que no son retenidos en las columnas o precolumnas de fase reversa utilizadas posteriormente para su desalado y/o separación disminuyendo por tanto el rendimiento del método y creando un sesgo en el tipo de péptidos caracterizados en un análisis. Para evitar este problema se han utilizado columnas rellenas de polvo de grafito, un soporte que permite retener péptidos hidrofílicos [35].

El método IMAC se ha aplicado asimismo a la purificación de fosfoproteínas completas. En 2002 y 2003, dos compañías (Quiagen y Clontech) comercializaron columnas basadas en cromatografía de afinidad por metales cuya utilización se ha descrito en numerosos trabajos. Por ejemplo, Puente y cols. compararon, mediante análisis 2DE y espectrometría de masas, el fosfoproteoma de células madre embrionarias de ratón diferenciadas y sin diferenciar, utilizando columnas Quiagen para el enriquecimiento inicial de las proteínas [36]. Otro trabajo más reciente, en este caso con columnas PMAC de Clontech, es el estudio del efecto del inhibidor rapamicina en células cancerosas marcadas mediante el método SILAC. EN este trabajo se combina el enriquecimiento en fosfoproteínas con la columna PMAC con el enriquecimiento posterior de los fosfopéptidos de digestion mediante IMAC. El análisis MudPit de los fosfopéptidos purificados permitió identificar 251 puntos de fosforilación afectados por la rapamicina [37].

2.3. MOAC *(Metal Oxide Affinity Chromatography).*

En 1997, Ikeguchi et al describieron por primera vez el uso de un óxido de metal, el óxido de titanio (TiO_2), para la purificación de aminoácidos fosforilados [38]. La utilidad de esta técnica para la purificación de péptidos fosforilados en estudios de proteómica fue demostrada en 2004 por Pinkse y col. [39]. Estos autores establecieron un sistema cromatográfico bidimensional que utilizaba partículas Titansphere (GL Science) en la primera dimensión y un soporte de fase reversa en la segunda. Los fosfopéptidos se separaban de los péptidos no modificados mediante su absorción en la columna de TiO_2 en condiciones ácidas. Una vez desorbidos de la columna mediante un tren de eluyente de pH básico, los fosfopéptidos se separaban en una columna de fase reversa dispuesta en línea y acoplada a su vez a un sistema de espectrometría de masas en tándem con interfase de electrospray (ESI-MS/MS). Mediante esta estrategia estos autores pudieron identificar ocho puntos de fosforilación en la quinasa PKG [39].

El TiO_2 tiene una alta estabilidad química y térmica y unas propiedades excepcionales como intercambiador de iones anfótero. A pH ácido actúa como un intercambiador de aniones por lo que resulta apropiado para la purificación de fosfopéptidos.

Para la purificación de la muestra, el proceso de carga y lavado del material en la columna se lleva a cabo utilizando tampones que contienen ácidos como el TFA, el 2,5-dihidroxibenzoico (DHB), el glicólico o el láctico. En estas condiciones de pH bajo, los residuos ácidos se encuentran protonados minimizándose la adsorción de péptidos no fosforilados. Estos ácidos orgánicos, además, desplazan los péptidos no fosforilados unidos inespecíficamente al soporte, respetando las interacciones entre el TiO_2 y el fosfato, mucho más fuertes (Figura 3). Estos factores contribuyen a hacer la purificación mediante soportes de TiO_2 más selectiva que la de IMAC.

El ácido utilizado y su concentración óptima en el tampón de carga dependen de la aplicación en particular y de la complejidad de la muestra. El DHB, por ejemplo, se utiliza principalmente para la purificación de péptidos previa a su análisis por MALDI a concentraciones entre 20-40 mg/ml, mientras que el ácido glicólico, se utiliza frecuentemente a concentraciones de 70 mg/ml para la preparación de muestras en análisis por LC/MS. La elución de los fosfopéptidos se realiza a pH alcalino existiendo, igual que en el IMAC, una buena compatibilidad con la fase reversa. Los valores de pKa y pKb para el TiO_2 son de 4.4 y 7.7, respectivamente, [40] por lo que a alto pH (normalmente se utiliza una solución de amoníaco a pH 11.5) este material se vuelve negativo y permite la elución de los fosfopéptidos.

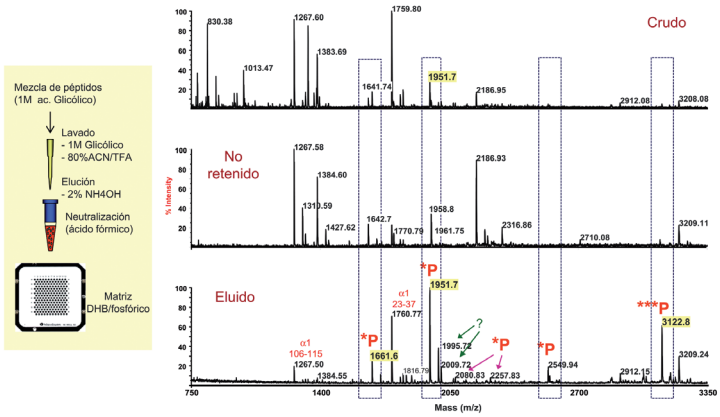


Figura 3: Espectros de MALDI-TOF de un digerido tríptico de caseína bovina antes (Crudo) y después del enriquecimiento en fosfopéptidos utilizando TiO_2 (Eluato). Para el enriquecimiento, la muestra se disuelve en ácido glicólico 1M y se pasa a través en una minicolumna previamente cargada con TiO_2. Tras los pertinentes lavados con ácido glicólico y con agua, los fosfopéptidos retenidos se eluyen con hidróxido amónico a pH 11.2.

En condiciones óptimas el TiO_2 provee de enriquecimientos de hasta el 90% [41]. Utilizando esta metodología en combinación con un marcaje SILAC, Olsen et al identificaron cerca de 6.600 puntos de fosforilación en 2.244 proteínas de células HeLa y cuantificaron sus cambios tras el tratamiento con el factor de crecimiento epitelial [5].

Los soportes de TiO_2 presentan una selectividad diferente a la de los de IMAC de forma que estas técnicas generan repertorios peptídicos complementarios [42-44]. En especial, el IMAC permite una mayor recuperación de péptidos múltiplemente fosforilados en comparación con el TiO_2 [45]. Esta diferencia ha sido la base de un nuevo método de purificación, denominado SIMAC [46], que mediante la combinación de la cromatografía IMAC y TiO_2 permite separar péptidos monofosforilados y múltiplemente fosforilados en muestras biológicas complejas. En un estudio en células madre mesenquimales humanas, esta estrategia combinada, comparada con una única etapa de purificación con TiO_2, permitió duplicar el número de fosfopéptidos identificados.

El TiO_2, junto con las fases Phos-Select, es actualmente el método más utilizado en los estudios fosfoproteómicos a gran escala. Las ventajas del TiO_2 residen en su alta selectividad, su tolerancia a las sales que se utilizan habitualmente en los procedimientos de preparación de muestras y a que es accesible para cualquier grupo de investigación ya que no requiere un equipamiento especial. Existen otras fases MOAC como ZrO2 [47], Al(OH)3 [48] o Nb2O5 [49] que, si bien se ha demostrado que pueden dar resultados complementarios al TiO_2, no son tan utilizadas en los estudios fosfoproteómicos.

El uso de los óxidos de metal para la purificación de fosfoproteínas completas en estudios de fosfoproteómica es menos frecuente. Wolschin y col. describieron el comportamiento de un soporte basado en hidróxido de aluminio Al(OH)3, indicando que este método es más selectivo, tiene una mejor relación coste-eficiencia y es más fácil de aplicar que los que utilizan las columnas comerciales disponibles actualmente [50].

En general todos estos procedimientos basados en IMAC o MOAC requieren el uso de cromatografía en columna y múltiples pasos en la preparación de la muestra que pueden complicar el análisis y disminuir la eficiencia en la recuperación de los péptidos fosforilados. Existen plataformas alternativas que superan estos inconvenientes como las bolas magnéticas o las placas de MALDI con superficie recubierta por estos materiales [51]. Estos procedimientos son cada día de uso más habitual, especialmente para el análisis de fosfopéptidos en muestras de baja complejidad. Por otro lado, Agilent ha comercializado un chip de tecnología microfluídica que com

bina el enriquecimiento de fosfopéptidos mediante TiO_2 con la posterior separación mediante fase reversa de los eluatos. Esta plataforma integra la columna de enriquecimiento, la columna analítica, todas las conexiones y la punta del microelectrospray en un chip reutilizable [52].

2.4. Métodos basados en modificaciones químicas de los fosfopéptidos.

Se han desarrollado diversas estrategias para la purificación de fosfopéptidos (y fosfoproteínas) basadas en la funcionalización de estas moléculas con grupos químicos que permiten su captura específica de forma directa (unión a un soporte sólido) o indirecta (*tags* de afinidad). Todos estos procedimientos implican generalmente alguna reacción de protección de los grupos reactivos del péptido (p.e los de cisteínas y metioninas) seguida de varias etapas de modificación que pueden implicar la pérdida del grupo fosfato en el derivado final producido.

Los procedimientos que transcurren con la pérdida de ión fosfato utilizan una etapa en medio básico fuerte que produce la beta-eliminación de este grupo de pSer y pThr. Este proceso da lugar a un grupo etileno al que, mediante una adición de Michael, puede unirse un reactivo nucleófilo como el etanoditiol. El derivado resultante corresponde a una molécula donde el fosfato original se ha substituido por un tioéter conteniendo un grupo tiol libre que puede utilizarse a su vez para unir moléculas de afinidad como la biotina [53] o para su unión a un soporte sólido [54]. Estas y otras estrategias análogas desarrolladas utilizando reactivos diferentes del etanoditiol [51] tienen el inconveniente de que no permiten recuperar el fosfopéptido en su estado original y al implicar la beta-eliminación del grupo fosfato de los aminoácidos no son aplicables a la captura de fosfotirosinas .

Aebersold y cols han desarrollado diversos procedimientos que utilizan una reacción de condensación catalizada por carbodiimida/imidazol para la formación de derivados fosforoamidato de los fosfopéptidos que permiten ligarlos directa o indirectamente a un soporte sólido (PAC, *phosphoroamidate chemistry*). Los fosfopéptidos purificados pueden posteriormente recuperarse mediante hidrólisis ácida del material. Los primeros protocolos desarrollados im-plicaban la derivatización del grupo fosfato del fosfopéptido con cistamina como paso intermedio. El derivado tiólico formado se inmovilizaba por unión covalente a bolas de vidrio funcionalizadas con grupos iodoacetilo [55]. Otra estrategia consistía en utilizar la reacción de esterificación para unir directamente los grupos fosfato a un dendrímero que actuaba de soporte sólido [56]. Más recientemente este grupo desarrolló protocolos optimizados para la unión de los péptidos a bolas de vidrio que reducían el número de etapas de reacción y que mejoraban la sensibilidad de la técnica. En un estudio del fosfoproteoma de Drosophila, este procedimiento demostró una especificidad y eficiencia similar a la obtenida mediante IMAC y muy superior a la del TiO_2. En este estudio se mostró además la diferente selectividad de estos métodos que producían colecciones de péptidos con coincidencias inferiores al 35% entre PAC e IMAC o el TiO_2 [42]. Esta característica permitió, mediante el uso paralelo de es-tos tres métodos de enriquecimiento junto con espectrómetros de masas de alta resolución (FT-ICR), la caracterización de 10.000 puntos de fosforilación únicos de alta confianza en el proteoma de Drosophila [57].

2.5. Métodos cromatográficos: SCX y HILIC.

En la cromatografía de intercambio iónico fuerte (SCX, Strong Cation Exchange) la retención de los péptidos aumenta con el grado de carga positiva de estos. A pH 2.7 un péptido tríptico suele tener una carga neta mínima de +2 debido al grupo amino en el N-terminal y a la presencia de lisina o arginina en el C-terminal. La adición de un fosfato a esta secuencia reduce la carga neta una unidad debido a la carga negativa del fosfato de forma que la forma fosforilada presentará una menor afinidad por la fase SCX y eluirá a tiempos de retención inferiores a la correspondiente secuencia no fosforilada. De esta forma, en una separación por SCX de un digerido tríptico, los péptidos fosforilados se concentran en las primeras fracciones cromatográ-

ficas, ricas en péptidos con carga +1, mientras que la mayor fracción de los péptidos trípticos eluye en fracciones posteriores [58]. Esta característica ha permitido utilizar la cromatografía SCX para el enriquecimiento de péptidos fosforilados, siendo frecuente su uso en estudios de caracterización a gran escala como un primer paso de enriquecimiento antes de la purificación con IMAC y/o TiO_2 [15, 44, 59].

La cromatografía por interacción hidrofílica (HILIC) es, como la cromatografía en fase reversa, una cromatografía de partición (absorción) en la que contrariamente a la anterior los componentes eluyen en orden de polaridad creciente. En la cromatografía HILIC, los péptidos se cargan en la columna en un disolvente orgánico y se eluyen con un gradiente de disolvente acuoso. Debido al incremento en polaridad conferido por la presencia de grupos fosfato en el péptido, esta técnica puede utilizarse para la obtención de fracciones enriquecidas en péptidos fosforilados [60]. La cromatografía HILIC permite purificar péptidos múltiplemente fosforilados, con carga neta negativa o nula, que no se retienen en las columnas de SCX. Las limitaciones de este método vienen dadas por la baja solubilidad de los péptidos de alto peso molecular en el disolvente orgánico y las dificultades para eluir péptidos múltiplemente fosforilados que pueden interaccionar muy fuertemente con la fase estacionaria.

3. Métodos espectrométricos para la identificación de fosfopéptidos.

Generalmente, la identificación de fosfoproteínas y la caracterización de sus puntos de fosforilación por espectrometría de masas, se lleva a cabo sobre extractos derivados de la digestión enzimática de estas proteínas previamente enriquecidos en fosfopéptidos mediante los métodos descritos en el apartado anterior. La caracterización de fosfopéptidos mediante MS viene dificultada por la baja abundancia de estos compuestos, su baja respuesta en el análisis MS y por patrones de fragmentación con información estructural limitada y que en ocasiones no es suficiente para su identificación [61].

En el caso de muestras con pocos componentes (péptidos o proteínas aisladas) es posible utilizar técnicas de MALDI-TOF para la caracterización de fosfopéptidos; sin embargo, en el caso de muestras más complejas, las estrategias más utilizadas consisten en la combinación de una separación por cromatografía líquida mono o bidimensional de los fosfopéptidos y su análisis mediante espectrometría de masas en tándem (LC-MS/MS) utilizando diferentes modos de fragmentación y barrido.

3.1.- MALDI-TOF y tratamiento con fosfatasas.

Esta técnica se ha utilizado con éxito para la determinación de puntos de fosforilación en proteínas puras o en mezclas poco complejas. Después de la digestión de la muestra el peso molecular de los péptidos se analiza por MALDI-TOF, generalmente tras un enriquecimiento previo en fosfopéptidos. Seguidamente se realiza un tratamiento con fosfatasa, que se puede realizar en la misma placa de MALDI, y se realiza un nuevo análisis de masa. Las señales de péptidos fosforilados se detectan entonces por sufrir un cambio de -80 unidades de masa entre los dos análisis [62]. Estos péptidos posteriormente se pueden fragmentar mediante MALDI-TOF/TOF para determinar la secuencia peptídica y el punto de fosforilación.

El uso de una matriz de MALDI adecuada para el análisis de péptidos fosforilados es un factor crítico en este procedimiento. La matriz más utilizada para este propósito es el DHB, que minimiza la rotura de los fosfopéptidos. El uso de coadyuvantes, como el ácido fosfórico incrementa

además la respuesta de estos compuestos en el análisis espectrométrico [63]. El THPA/DAHC (2,4,6-*trihydroxyacetophenone/diammonium hydrogen citrate*) es otra matriz que ha dado buenos resultados en cuanto a su especificidad por los péptidos fosforilados [64].

3.2 - Fragmentación por CID.

El método de referencia en proteómica para la caracterización de péptidos y fosfopéptidos es su secuenciación a partir de sus espectros de fragmentación obtenidos mediante espectrometría de masas en tándem. El método más utilizado para llevar a cabo esta fragmentación es el denominado de disociación inducida por colisión (CID). En este proceso los iones del péptido o iones precursores sufren colisiones con un gas inerte como el helio, lo que motiva un aumento de la energía interna del péptido y su posterior fragmentación para dar lugar a los denominados iones producto y a pérdidas neutras. Esta fragmentación puede ocurrir en cualquiera de los enlaces de la cadena lineal generando iones fragmento de diferentes tipos. Las series de fragmentos de tipo a, b y c contienen el extremo amino terminal del precursor mientras que las de tipo x, y, y z contienen el carboxilo terminal [65]. Pueden formarse asimismo fragmentos iónicos derivados de la rotura de dos enlaces denominados fragmentos internos que incluyen los iones imonio. Además, en el caso de los fosfopéptidos, el enlace fosfato es un enlace débil que se rompe fácilmente para originar pérdidas neutras (-98, H_3PO_4 y -80 HPO_3), o iones cargados característicos (79-, eliminación del grupo fosfato, y 216-, ión imonio de pTyr) que pueden utilizarse como marcadores para la identificación del fosfopéptido. La pérdida de H_3PO_4 es mayoritaria en péptidos que contienen pSer mientras que en secuencias con pThr es menos extensa y puede estar acompañada de una pérdida de HPO_3. Los péptidos que contiene pTyr no suelen mostrar señales derivadas de perdidas neutras de fosfato y cuando éstas se observan son casi siempre debidas a la pérdida de HPO_3.

La fragmentación por CID puede llevarse a cabo tanto en sistemas de analizadores en tándem convencionales (triples cuadrupolos, TOF/TOF) o híbridos (Q-TOF, Q-TRAP) como en sistemas de atrapamiento de iones (ion trap, ICR, Orbitrap). El modo de análisis de los iones producidos por CID depende del sistema de analizadores utilizado. Los modos clásicos de barrido en un espectrómetro de masas en tándem son los de iones producto, de iones precursores y de pérdidas neutras (Figura 4). Cualquiera de estos modos puede aplicarse al análisis de péptidos fosforilados si bien los más frecuentemente utilizados son los dos primeros.

3.2.1.- Espectrometría de masas de iones producto.

El patrón de fragmentación de un péptido analizado mediante MS/MS depende de la energía de colisión, de la presión y el tipo de gas de colisión y de la carga del péptido. Este patrón es además característico de la secuencia aminoacídica específica del péptido, de forma que esta última puede deducirse del análisis del correspondiente espectro MS/MS de iones producto.

Como ya se indicó anteriormente, en el proceso de fragmentación pSer y pTyr pueden perder fácilmente el grupo fosfato a través de una reacción de beta-eliminación. Como consecuencia, el patrón de fragmentación de los péptidos que contienen pSer o pThr suele estar dominado por una señal derivada de la pérdida de ácido fosfórico a partir del ión precursor (diferencia de -98 Da). La intensidad de esta señal depende del tipo de analizador (en las trampas iónicas es más pronunciada), de la carga del péptido (los iones +2 muestran una mayor pérdida que los iones +3) y de su secuencia [66, 67]. A menudo esta pérdida es mayoritaria, sobretodo en el caso de secuencias que contiene pSer, y va acompañada de una fragmentación muy baja de la cadena del péptido. Además del ión precursor, los iones fragmento del péptido pueden mostrar señales asociadas derivadas de la perdida de fosfato, lo que aumenta la complejidad del espectro y complica la elucidación de la secuencia y del punto exacto de fosforilación. Si bien la pérdida de -98 también se ha observado en algunas secuencias con pTyr, en esos ca-sos no se trata de una señal dominante y probablemente su origen es también diferente ya que en este aminoácido

no es posible una beta-eliminación de fosfato. Así, la señal se atribuye a la pérdida conjunta de HPO3 del residuo fosforilado y H2O de otro residuo o bien a la transferencia del grupo HPO3 a un residuo de ácido aspártico seguida de una pérdida de H3PO4 [68].

En la secuenciación de los fosfopéptidos a partir de sus espectros de fragmentación la localización de pSer se evidencia por la detección de diferencias de 160 uma entre dos señales fragmento, o de 69 uma que corresponde al producto que ha sufrido la beta eliminación, la dehidroalanina. De forma similar, la localización de las pThr se realiza mediante la detección de pérdidas de 181 o del correspondiente derivado a –83, el ácido dehidroamino-2-butirico. Las pTyr se localizan por la diferencia de masa 243 (163 + 80) (Figura 4).

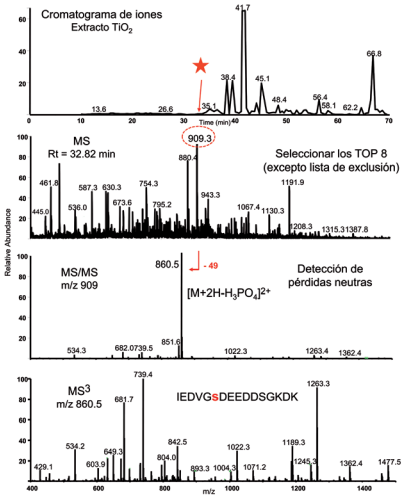


Figura 4: Modos de barrido específicos en un instrumento de triple cuadrupolo para la detección de péptidos fosforilados.

La espectrometría de masas en tándem múltiple se utiliza para identificar de forma muy eficiente fosforilaciones en pSer y pThr. Este tipo de aná-lisis se lleva a cabo generalmente en instrumentos de trampa iónica que permiten realizar varios ciclos consecutivos de aislamiento, excitación y expulsión de los iones, seleccionando en cada ciclo un ión precursor entre los iones producto que se producen en el ciclo previo. Mediante esta estrategia es posible fragmentar los iones derivados de la pérdida de fosfato que dominan los espectros de fragmentación de los fosfopéptidos. De esta forma, aunque el espectro MS2 puede no contener suficiente información de secuencia para la identificación, el análisis MS3 de estos iones suele generar espectros ricos en señales que si lo permiten. Este proceso, que requiere un análisis 'data-dependent' basado en la detección de pérdidas neutras a partir del ion protonado del péptido, es una de las técnicas más comúnmente utilizadas en fosfoproteómica. En la práctica el instrumento se programa para detectar en el espectro de fragmentación MS2 iones fragmentos derivados de la pérdida de 98, 49 o 32.5 uma (correspondientes a pérdidas de fosfato a partir de iones mono, doble y triplemente cargados respectivamente). Si el fragmento detectado está entre los de más intensidad del espectro, el instrumento lleva a cabo un análisis MS3 tomando este ion como precursor (Figura 5).

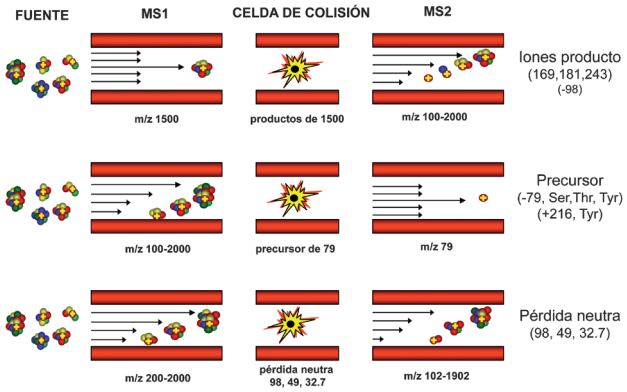


Figura 5: Detección de fosfopéptidos en un espectrómetro de masas de trampa iónica. (Thermo LTQ). En primer lugar se realiza un barrido completo para detectar los iones que están llegando al espectrómetro en ese momento. Seguidamente se seleccionan y fragmentan los ocho iones más abundantes y, en el caso de detectar una pérdida neutra de fosfato (98, 49 o 32.7) en alguno de los 5 iones más abundantes, se realiza un barrido de MS^3. No debe confundirse este modo de barrio con el barrido de pérdidas neutras que se realiza en espectrómetros de masas de doble analizador como el triple cuadrupolo. En este caso, únicamente se detecta la pérdida neutra para realizar un posterior análisis mediante MS^3, mientras que en el caso del barrido de pérdida neutras está pérdida se monitoriza durante todo el análisis.

Esta estrategia puede utilizarse también cuando se tiene cierta información acerca de los posibles puntos de fosforilación existentes. La denominada *"Hipotesis-drive-stage mass spectrometry"* se ha utilizado para mapear los puntos de fosforilación de proteínas conocidas. Según esta estrategia, inicialmente se determinan todos los sitios potenciales de fosforilación de la proteína en cuestión y se calculan las masas de los iones y fragmentos esperables. Con esta información, se programan ciclos de barrido MS2 y MS3 con objeto de confirmar o descartar cada uno de los puntos de fosforilación [69].

Recientemente, Schroider y col. introdujeron un nuevo método de fragmentación para trampas iónicas denominado MSA o pseudoMS3 (MSA, Multi Stage Activation). A diferencia de los métodos descritos anteriormente, que utilizan dos etapas consecutivas de análisis en tándem, en MSA los iones fragmentos se obtienen en una única etapa. Así, los iones derivados de pérdidas neutras son fragmentados mientras el ion precursor se encuentra en la trampa. El resultado equivale a una combinación de los espectros MS2 y MS3 que en muchos casos mejora la identificación [70]. Algunos ejemplos de la aplicación de esta metodología son el análisis del fosfoproteoma del Trypanosoma cruzi [71] o el estudio cuantitativo mediante SILAC del efecto de diferentes inhibidores de quinasas sobre células HeLa [72] .

3.2.2. Espectrometría de masas de iones precursores o PIS.

El análisis PIS es un método de barrido muy selectivo utilizado frecuentemente para el análisis de familias de compuestos con grupos químicos comunes en su estructura. Este tipo de análisis se lleva a cabo en instrumentos de dos analizadores en tándem como el triple cuadrupolo o el Q-TOF. En cada barrido PIS, el primer analizador transfiere secuencialmente todo los iones formados en la fuente a la cámara de colisión donde se fragmentan, mientras que el segundo analizador se encarga de monitorizar un fragmento específico, característico de la familia de moléculas analizadas (Figura 4).

Este modo de trabajo se ha utilizado en proteómica para el análisis de modificaciones postraduccionales. En el caso de la fosfoproteómica, esta metodología se aplica a la detección de péptidos fosforilados en pTyr ya que este modo de trabajo implica la monitorización del ión imonio del resto fosfoaminoacídico y por tanto requiere que el enlace fosfato sea resistente al análisis espectrométrico. El ión monitorizado es el ión imonio a 216.043 Da. Es muy importante trabajar con una buena resolución y precisión en esta región de bajo peso molecular ya que existen otros iones de tipo b como la com-binación NT y GS con masas de 216.098 y 216.069 respectivamente y que podrían gene-rar falsos positivos. Otras aplicaciones de esta técnica en el análisis de modificaciones postra-duccionales son la detección de nitración en tirosina (ión imonio específico a m/z 181.06) y la acetilación de lisinas (ión imonio a m/z 143.1 o m/z 126.1).

El PIS se puede llevar a cabo también en el modo de ionización de iones negativos aunque su aplicación en fosfoproteómica ha sido rara. En este modo, la fragmentación de los fosfopéptidos origina una señal a m/z 79 de-bida al ión fosfato y que suele ser muy abun-dante. Esta técnica es muy selectiva y permite detectar péptidos fosforilados en Ser y Thr en mezclas complejas (Figura 6). Dado que la elu-cidación de la secuencia mediante

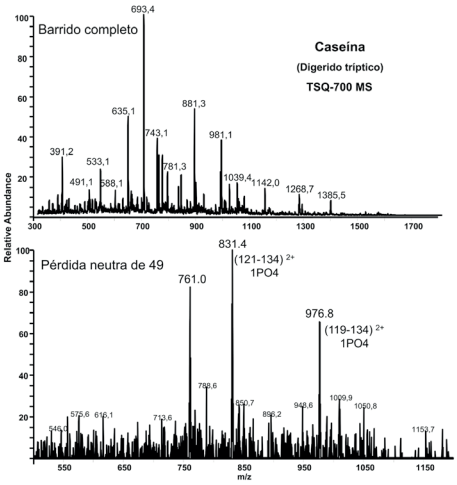


Figura 6: Espectros de barrido completo y de PIS de un digerido tríptico de caseína utilizando un sistema de cuadrupolo TSQ7000 con interfase de nanospray.

fragmentación del péptido en modo negativo no es muy eficiente, en la práctica suele utilizarse un ciclo automático de barrido de dos etapas en el que inicialmente se detecta la señal a m/z 79 en modo negativo y seguidamente se obtiene el espectro de fragmentación MS/MS en modo de iones positivos.

La sensibilidad y selectividad en un análisis PIS puede mejorarse utilizando derivados químicos de los fosfopéptidos capaces de producir iones fragmentos específicos de alta abundancia. Por ejemplo, Arrigoni y col utilizan una reacción de beta eliminación y de adición de Michael en un solo paso para derivatizar la pSer y pThr, sustituyendo el grupo negativamente cargado del fosfato por un grupo positivamente cargado S-etilpiridil que aumenta notablemente la intensidad de los iones precursores formados. En condiciones de fragmentación suaves, este derivado genera un único fragmento de alta abundancia a m/z 106 que constituye un marcador muy específico de fosforilación [73].

3.3.- Otros métodos de fragmentación: ETD y ECD.

Los denominados ECD (Electron Capture Dissociation) y ETD (Electron Transfer Dissociation) son métodos de fragmentación que implican la transferencia directa o indirecta de electrones a los iones del péptido. El método ECD, desarrollado por el grupo de McLafferty [74], implica la captura de un electrón de baja energía por un catión precursor múltiplemente cargado mientras que en el caso del ETD, desarrollado por el grupo de Hunt [75], se transfiere un electrón desde un anión radical reactivo al catión del péptido precursor. Ambos son métodos de disociación dirigidos por radicales que inducen la rotura en el enlace N-Cα del péptido para dar lugar predominantemente a iones producto de tipo c y z. A diferencia del CID, estos métodos respetan los lábiles enlaces fosfato de los fosfopéptidos, lo que, en el caso de los fosfopéptidos, permite una asignación precisa del punto de fosforilación. En la identificación de fosfopéptidos, la fragmentación mediante ETD/ECD produce sus mejores resultados en el análisis de precursores de alto peso molecular y con estado de carga mayor de 2. En comparación, el CID es mejor para péptidos doblemente cargados o con una relación m/z baja. Debido a estas características, ETD/ECD y CID son técnicas complementarias que pueden utilizarse de forma combinada para aumentar el número de sitios de fosforilación identificados en un estudio [68, 76].

4. Análisis *shotgun*.

Ficarro et al describieron en 2002 el primer análisis global de un fosfoproteoma, el de Saccharomyces cerevisiae. En este trabajo se utilizó una columna IMAC para el enriquecimento de los fosfopéptidos en línea con una columna capilar de fase reversa de 360 µm de diámetro interno acoplada a un espectrómetro de trampa iónica LCQ. Mediante este sistema y a partir de 500 µg de muestra de partida, estos autores describieron 383 puntos de fosforilación en la levadura [33].

Esta estrategia, denominada de análisis shotgun (ver capítulo 28), con diferentes variantes, ha sido la utilizada en la mayoría de los estudios posteriores relacionados con el análisis global del fosfoprotema (Figura 7). En un procedimiento estándar, la muestra o muestras se digieren con un enzima específico, generalmente tripsina, y la mezcla se somete a los diferentes procesos de purificación con objeto de enriquecer los extractos en péptidos fosforilados. Seguidamente, las mezclas enriquecidas se inyectan en un sistema cromatográfico acoplado on-line con un sistema de espectrometría de masas en tándem a través de una interfase de microelectrospray. Los sistemas más utilizados usan columnas con un diámetro interno de 75-250 µm y una longi-tud de 5-25 cm trabajando a unos flujos de entre 200 nL/min y 2 µL/min. En la mayor parte de los trabajos publicados, el espectrómetro de masas se programa para realizar un

primer barrido para detectar los iones de cada uno de los compo-nentes que eluyen de la columna y seguidamente cada uno de los iones precursores se fragmenta me-diante MS/MS obteniéndose la información necesaria para la secuenciación. La utilización de diferentes algoritmos de búsqueda como Sequest, Mascot, Ommsa o Phenyx permite la caracterización de la se-cuencia mediante búsqueda en bases de datos. La utilización de varios de estos motores de búsqueda en paralelo incrementa la confianza y el número de secuencias identificadas, sobre todo en el caso de los fosfopéptidos [59].

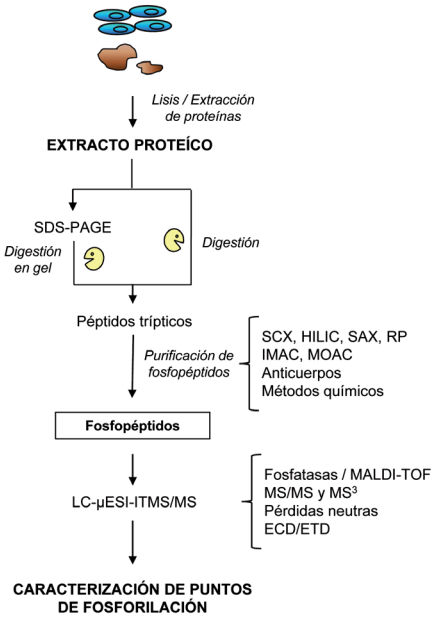


Figura 7: Flujo de trabajo en un análisis cualitativo global del fosfoproteoma.

Las estrategias de barrido citadas, junto con el pos-terior análisis bioinformático de los espectros de frag-mentación permiten identificar de forma eficiente los péptidos fosforilados en estas muestras así como iden-tificar con alta fiabilidad sus secuencias aminoacídicas. La determinación exacta del punto de fosforilación en las secuencias con más de un sitio posible puede ser más complicada, ya que frecuentemente los corres-pondientes espectros de fragmentación suelen aportar una información ambigua. Por este motivo, generalmente se hace necesario la validación de la localización del punto de fosforilación, bien manualmente, bien utilizando herramientas bioinformáticas que evalúan de forma automática los espectros de fragmentación y reasignan el punto de modificación además de generar una puntua-ción que permite discernir si dicho punto de fosforilación es de alta confianza o no [43, 58].

5. Cuantificación de puntos de fosforilación.

La cuantificación relativa o absoluta de los componentes del fosfoproteoma y el análisis de sus cambios bajo diferentes condiciones (p.e. tratamiento con fármacos, estrés, activación celular) aportan una información crítica sin la cual no sería posible entender el funciona-miento de las complejas y dinámicas redes moleculares reguladas por los procesos de fosforilación.

Actualmente se dispone de diversas estrategias para el análisis cuantitativo de péptidos me-diante espectrometría de masas que pueden aplicarse al estudio del fosfoproteoma [77]. Si bien algunos estudios han utilizado técnicas label-free para este propósito [78], la mayoría de estas estrategias utilizan algún tipo de marcaje isotópico para la cuantificación.

Las técnicas más utilizadas actualmente son el marcaje isotópico de las proteínas celulares mediante incorporación metabólica de los isótopos a partir de medios de cultivo especiales (SI-LAC) y el marcaje mediante derivatización química de los péptidos de digestión con los reactivos iTRAQ (Isotope Tagging Reagent for Relative and Absolute Quantification, Applied Biosystems) o TMT (Tandem Mass Tag, Thermo Scientific). Esta última estrategia es especialmente útil en el caso de cultivos primarios o análisis de fluidos biológicos, casos en los que las técnicas de marcaje metabólico como el SILAC no se pueden aplicar.

En la cuantificación mediante SILAC la muestra analizada corresponde a una mezcla de las muestras a comparar, que a su vez derivan de células incubadas en medios con diferente composición isotópica. En los espectros de masas, cada péptido del digerido de la muestra se detecta en forma de un multiplete de señales separadas entre sí en función del número de isotópos pesados que cada péptido tiene incorporados. Las intensidades relativas de las señales en estos multipletes determinan las cantidades relativas del péptido en cada una de las muestras. Ya en el año 2004 el grupo de Gygi utilizó esta metodología para comparar los niveles de expresión de proteínas en fracciones microsomales de células de cáncer de próstata con diferente grado de metástasis, identificando un total de 60 proteínas con niveles más altos en las células más metastásicas y 22 proteínas con niveles inferiores [79]. En otro estudio Gu y colaboradores identificaron cerca de 400 puntos de fosforilación que variaban en respuesta a la inhibición del gen FLT3, gen muy relacionado con la aparición de la leucemia mieloide aguda y una diana muy interesante para el desarrollo de terapias frente a esta enfermedad [14]. El grupo de Mann realizó un estudio con marcaje triple SILAC para cuantificar puntos de fosforilación en células HeLa, tratadas con factor de crecimiento epidérmico con y sin adición de inhibidores de quinasas e identificando más de 6.000 puntos de fosforilación [72]. Una revisión de la utilización del SILAC en fosfoproteómica se encuentra en la referencia [80].

En el caso del marcaje iTRAQ o TMT, los péptidos resultantes de la digestión de cada muestra se marcan con los reactivos correspondientes, se mezclan y se realiza la cuantificación mediante espectrometría de masas en tándem utilizando una serie de iones reportero de bajo peso molecular característicos de cada reactivo. En este caso, las moléculas de péptidos con una misma secuencia aminoacídica presentan la misma masa independientemente del reactivo de marcaje, es decir, independientemente de la muestra de procedencia. De esta forma, para cada péptido existe un único precursor a seleccionar en el espectrómetro de masas, determinándose la proporción de cada uno de los marcadores en el espectro de fragmentación. Este procedimiento provee a estos métodos de una sensibilidad y especificidad superior a la del SILAC. La monitorización de los iones reportero, de bajo peso molecular, fue inicialmente un inconveniente para la utilización de estos métodos de marcaje con espectrómetros de trampa iónica, muy extendidos en los laboratorios de proteómica pero que presentaban limitaciones en el valor inferior del rango de masas utilizable. La reciente aparición de los modos de fragmentación PQD y HCD en los sistemas de trampa iónica y Orbitrap, ha permitido la utilización de estos analizadores en la cuantificación mediante iTRAQ o TMT y su aplicación al estudio cuantitativo del fosfoproteoma. Utilizando un sistema Orbitrap con fragmentación HCD se ha reportado un rango dinámico de dos órdenes de magnitud para la cuantificación de fosfopéptidos en el estudio de la leucemia mieloide aguda [81]. Una revisión de este tema se puede encontrar en la referencia [82].

Otros métodos de marcaje como el método clásico de metilación con CD3OD han sido utilizados en diversos estudios aunque de forma más minoritaria. Con frecuencia el marcaje con CD3O se ha utilizado combinado con procedimientos de purificación por IMAC [83] o con aquellos basados en la modificación química de los fosfopéptidos descritos en un apartado anterior [55, 56], y que requieren la protección de los grupos carboxilo.

Muchos de los trabajos de fosfoproteómica cuantitativa utilizan modos de barrido (como el denominado data-dependent) que intentan identificar el mayor número de compuestos posibles. Cuando se requiere cuantificar un número limitado de fosfoproteínas o fosfopéptidos conocidos, el análisis espectrométrico puede llevarse a cabo mediante la técnica denominada de monitorización de transiciones seleccionadas (SRM, Selected Reaction Monitoring, ver capítulo 9). En el análisis por SRM se programa un ciclo repetido de barridos específicos de cada uno de los péptidos de interés. En cada barrido se selecciona y fragmenta un ión del péptido en cuestión (ion precursor) y se monitorizan uno o más de sus iones producto. Un ejemplo de este método es el estudio de los cambios en 226 puntos de fosforilación a 7 tiempos tras la estimulación con EGF que permitió la cuantificación reproducible en 4 réplicas del 88% de los péptidos monitorizados (frente al 34% en el modo de barrido convencional) [84].

También se han realizado estudios de cuantificación absoluta utilizando péptidos marcados isotópicamente como patrón interno de cada una de los fosfopéptidos en estudio (péptidos AQUA). Mediante esta aproximación se han cuantificado dos puntos de fosforilación implicados en la actividad de las CDKs (Thr14 y Tyr15) en células Jurkat en diferentes estados y en apoptosis [85].

6. Conclusiones y perspectiva.

La determinación global del perfil de fosfoproteínas celulares y la caracterización de sus puntos de modificación es uno de los retos más exigentes de la proteómica actual. El desarrollo de métodos de enriquecimiento eficaces y la disponibilidad de nuevas técnicas de fragmentación y de instrumentos cada vez más sensibles ha mejorado cualitativa y cuantitativamente la información obtenible en estos análisis. De esta forma, la fosfoproteómica está jugando un papel de relevancia en el estudio de los mecanismos moleculares de la función celular.

En la última década, los inhibidores de quinasas han sido una de las principales familias de moléculas utilizadas en el desarrollo de fármacos, principalmente en cáncer. Durante este periodo, solo en los EEUU, se registraron más de 10.000 patentes de inhibidores de este tipo considerándose a las quinasas como la principal diana terapéutica del siglo XXI. En este contexto, la fosfoproteómica promete convertirse en una herramienta clave en la caracterización de las dianas de estos inhibidores, de las vías implicadas y por tanto de los posibles efectos secundarios de estos fármacos.

REFERENCIAS.

1. Hunter, T., *Signaling - 2000 and beyond.* Cell 2000. 1000: p. 113-127.

2. Pawson, T. and J.D. Scott, *Protein phosphorylation in signaling-50 years and counting.* Trends. Biochem. Sci., 2005. 30: p. 286-290.

3. Johnson, S.A. and T. Hunter, *Kinomics: Methods for deciphering the kinome.* Nat. Methods, 2005. 2: p. 17-25.

4. Reinders, J. and A. Sickmann, *State-of-the-art in phosphoproteomics.* Proteomics, 2005. 5: p. 4052-4061.

5. Olsen, J., B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, et. al., *Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks.* Cell, 2006. 127: p. 635–648.

6. Cohen, P., *The regulation of protein function by multisite phosphorylation- a 25 year update.* TIBS, 2000. 25(596).

7. Hunter, T., *The phosphorylation of proteins on tyrosine: Its role in cell growth and disease.* Philos.Trans. R. Soc.London B.Biol.Sci, 1998. 353: p. 583-605.

8. Schlessinger, J., *Cell signalling by receptor tyrosine kinases.* Cell, 2000. 103: p. 211-225.

9. Harsha, H. and A. Pandey, *Phosphoproteomics in cancer.* Molecular Oncology, 2010. 4: p. 482-495.

10. Kontzias, A., A. Laurence, M. Gadina and J.J. O'Shea, *Kinase inhibitors in the treatment of immune-mediated disease.* F1000 Med Rep, 2012. 4: p. 5.

11. Dominguez, J.M., A. Fuertes, L. Orozco, M. del Monte-Millan, E. Delgado and M. Medina, *Evidence for irreversible inhibition of glycogen synthase kinase-3beta by tideglusib.* J Biol Chem, 2012. 287(2): p. 893-904.

12. Zhai, B., J. Villen, S.A. Beausoleil, J. Mintseris and S.P. Gygi, *Phosphoproteome analysis of Drosophila melanogaster embryos.* J. Proteome Res., 2008. 7(4): p. 1675-1682.

13. Mayya, V., D. Lundgren, S. Hwang, K. Rezaul, Wu, J. Eng, et al., *Quantitative Phosphoproteomic Analysis of T Cell Receptor Signaling Reveals System-Wide Modulation of Protein-Protein Interactions.* Science Signaling, 2009. 2: p. 1-16.

14. Gu, T., J. Nardone, Y. Wang, M. Loriaux, J. Villén, S. Beausoleil, et. al., *Survey of activated FLT3 signaling in leukemia.* PLoS One, 2011. 6: p. e19169.

15. Villen, J., S.A. Beausoleil, S.A. Gerber and S.P. Gygi, *Large-scale phosphorylation analysis of mouse liver.* Proc. Natl. Acad. Sci., 2007. 104(5): p. 1488-1493.

16. Bakalarski, C.E., W. Haas, N.E. Dephoure and S.P. Gygi, *The effects of mass accuracy, data acquisition speed, and search algorithm choice on peptide identification rates in phosphoproteomics.* Anal. Bioanal. Chem., 2007. 389: p. 1409-1419.

17. Molina, H., D.M. Horn, N. Tang, S. Mathivanan and A. Pandey, *Global proteomic profiling of phosphopeptides using electron transfers dissociation tandem mass spectrometry.* Proc. Natl. Acad. Sci., 2007. 104(7): p. 2199-2204.

18. Gnad, F., S. Ren, J. Cox, J. Olsen, B. Macek, M. Oroshi, et al., *PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites.* Genome Biol., 2007. 8(11): p. R250.

19. Diella, F., C.M. Gould, C. Chica, A. Via and T.J. Gibson, *Phospho.ELM: a database of phosphorylation sites - update 2008.* Nucleis Acids Res., 2007. 36: p. D240-D244.

20. Bodenmiller, B., D. Campbell, B. Gerrits, H. Lam, M. Jovanovic, P. Picotti, et al., *PhosphoPep--a database of protein phosphorylation sites in model organisms.* Nat. Biotechnol., 2008. 26(12): p. 1339-40.

21. Ovelleiro, D., M. Carrascal, V. Casas and J. Abian, *LymPHOS: Design of a phosphosite database of primary human T cells.* Proteomics, 2009. 9: p. 3741-3751.

22. Zhang X, Y.J., Jensen ON, Roepstorff P., *Highly Efficient Phosphopeptide Enrichment by Calcium Phosphate Precipitation Combined with Subsequent IMAC Enrichment.* Mol Cell Proteomics., 2007. 6: p. 2032-2042.

23. Gevaert, K. and J. Vandekerckhove, *Reverse-phase diagonal chromatography for phosphoproteome research.* Methods Mol Biol, 2009. 527: p. 219-27, ix.

24. Frackelton, A., Jr, A. Ross and H. Eisen, *Characterization and use of monoclonal antibodies for isolation of phosphotyrosyl proteins from retrovirus-transformed cells and growth factor-stimulated cells.* Mol Cell Biol, 1983. 3: p. 1343-1352.

25. Boersema, P., L. Foong, V. Ding, D. Lemeer, B. van Breukelen, R. Philp, et al., *In depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immuno-affinity purification and stable isotope dimethyl labeling.* Mol Cell Proteomics, 2009.

26. Ficarro, S., A. Salomon, L. Brill, D. Mason, M. Stettler-Gill, A. Brock, et al., *Automated immobilized metal affinity chromatography/nano-liquid chromatography/electrospray ionization mass spectrometry platform for profiling protein phosphorylation sites.* Rapid Commun Mass Spectrom, 2005. 19: p. 57-71.

27. Nguyen, V., L. Cao, J. Lin, N. Hung, N. Ritz, K. Yu, et al., *A New Approach for Quantitative Phosphoproteomic Dissection of Signaling Pathways Applied to T Cell Receptor Activation.* Mol Cell Proteomics., 2009. 8: p. 2418-2431.

28. De Corte, V., H. Demol, M. Goethals, J. Van Damme, J. Gettemans and J. Vandekerckhove, *Identification of Tyr438 as the major in vitro s-Scr y actu phosphorylation site in human gelsolin: a mass spectrometric approach.* 1999.

29. Zhang, H., X. Zha, Y. Tan, P. Hornbeck, A.J. Mastrangelo, D.R. Alessi, et al., *Phosphoprotein analysis using anibodies broadly reactive against phosphorylated motifs.* J.Biol. Chem., 2002. 277: p. 39379-39387.

30. Astoul, E., A. Laurence, N. Totty, S. Beer, D. Alexander and D. Cantrell, *Approaches to define antigen receptor-induced serine kinase transduction pathways.* J.Biol.Chem, 2003. 278: p. 9267-9275.

31. J.Comb, M., *Production of motif.specific and context-independent antibodies using peptide libraries as antigens.* 2002: United States.

32. Porath, E., J. Carlsson, I. Olsson and G. Belfrage, *Metal chelate affinity chromatography, a new approach to protein fractionation.* Nature, 1975. 258: p. 2.

33. Ficarro, S., M. McCleland, P. Stukenberg, D. Burke, M. Ross, J. Shabanowitz, et al., *Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae.* Nat Biotechnol, 2002. 20: p. 301-305.

34. Nuhse, T.S., A. Stensballe, O.N. Jensen and S.C. Peck, *Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry.* Mol Cell Proteomics, 2003. 2(11): p. 1234-43.

35. Larsen, M.R., M.E. Graham, P.J. Robinson and P. Roepstorff, *Improved detection of hydrophilic phosphopeptides using graphite powder microcolumns and mass spectrometry: evidence for in vivo doubly phosphorylated dynamin I and dynamin III.* Mol Cell Proteomics, 2004. 3(5): p. 456-65.

36. Puente, L.G., D.J. Borris, J.F. Carriere, J.F. Kelly and L.A. Megeney, *Identification of candidate regulators of embryonic stem cell differentiation by comparative phosphoprotein affinity profiling.* Mol Cell Proteomics, 2006. 5(1): p. 57-67.

37. Chen, R.Q., Q.K. Yang, B.W. Lu, W. Yi, G. Cantin, Y.L. Chen, et al., *CDC25B mediates rapamycin-induced oncogenic responses in cancer cells.* Cancer Res, 2009. 69(6): p. 2663-8.

38. Ikeguchi, Y. and H. Nakamura, *Determination of organic phosphates by column-switching high performance anion-exchange chromatography using on-line preconcentration on Titania.* Anal Sci, 1997. 13(479-483).

39. Pinkse, M.W.H., P.M. Uitto, M.J. Hilhorst, B. Ooms and A.J.R. Heck, *Selective isolation at the fentomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns.* Anal Chem, 2004. 76: p. 2925-3943.

40. Koizumi, Y. and M. Taya, *Kinetic evaluation of biocidal activity of titanium dioxide against phage MS2 considering interaction between the vage and photocatalyst particles.* Biochem Eng J, 2002. 12: p. 107-116.

41. Larsen, M.R., T.E. Thingholm, O.N. Jensen, P. Roepstorff and T.J.D. Jørgensen, *Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns.* Mol. Cell. Proteomics, 2006. 4(873-876).

42. Bodenmiller, B., L.N. Mueller, M. Mueller, B. Domon and R. Aebersold, *Reproducible isolation of distinct, overlapping segments of the phosphoproteome.* Nature Methods, 2007. 4(3): p. 231-237.

43. Carrascal, M., D. Ovelleiro, V. Casas, M. Gay and J. Abian, *Phosphorylation analysis of primary human T lymphocytes using sequential IMAC and titanium oxide enrichment.* J.Proteome Res., 2008. 7: p. 5167-5176.

44. Dephoure, N., C. Zhou, J. Villen, S. Beausoleil, C. Bakalarski, E. Elledge, et al., *A quantitative atlas of mitotic phosphorylation.* PNAS, 2008. 105: p. 10762-10767.

45. Jensen, S. and M. Larsen, *Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques.* Rapid Commun Mass Spectrom., 2007. 21: p. 3635-3645.

46. Thingholm, T.E., O.N. Jensen, P.J. Robinson and M.R. Larsen, *SIMAC - A phosphoproteomic strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptide.* Molecular Systems Biology, 2008.

47. Zhou, H., R. Tian, M. Ye, S. SXu, S. Feng, C. Pan, *et al., Highly specific enrichment of phosphopeptides by zirconium dioxide nanoparticles for phosphoproteome analysis.* Electrophoresis, 2007. 28: p. 2201-2215.

48. Wang, Y., W. Chen, J. Wu, Y. Guo and X. Xia, *Highly efficient and selective enrichment of phosphopeptides using porous anodic aluminia membrane for MALDI-TOF MS analysis.* J Am Soc Mass Spectrom, 2007. 18: p. 1387-1395.

49. Ficarro, S., J. Parikh, N. Blank and J. Marto, *Niobium(V) oxide (Nb2O5): Application to phosphoproteomics.* Anal Chem, 2008. 80: p. 4606-4613.

50. Colby, T., H. Röhrig, A. Harzen and J. Schmidt, *Modified metal-oxide affinity enrichment combined with 2D-PAGE and analysis of phosphoproteomes.* Methods Mol Biol., 2011. 779: p. 273-286.

51. Dunn, J.D., G.E. Reid and M.L. Bruening, *Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry.* Mass Spectrom Rev, 2010. 29(1): p. 29-54.

52. Raijmakers, R., K. Kraiczek, A.P. de Jong, S. *Mohammed and A.J. Heck, Exploring the human leukocyte phosphoproteome using a microfluidic reversed-phase-TiO2-reversed-phase high-performance liquid chromatography phosphochip coupled to a quadrupole time-of-flight mass spectrometer.* Anal Chem, 2010. 82(3): p. 824-32.

53. Oda, Y., T. Nagasu and B.T. Chait, *Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome.* Nat Biotechnol, 2001. 19(4): p. 379-82.

54. McLachlin, D.T. and B.T. Chait, *Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides.* Anal Chem, 2003. 75(24): p. 6826-36.

55. Zhou, H., J.D. Watts and R. Aebersold, *A systematic approach to the analysis of protein phosphorylation.* Nat Biotechnol, 2001. 19(4): p. 375-8.

56. Tao, W.A., B. Wollscheid, R. O'Brien, J.K. Eng, S.-J. Li, B. B, *et al., Quantitative phosphoproteome analysis usign a dendrimer conjugation chemistry and tandem mass spectrometry.* Nature Methods, 2005. 2(8): p. 591-598.

57. Bodenmiller, B., J. Malmstrom, B. Gerrits, D. Campbell, H.L. H, A. Schmidt, *et al., PhosphoPep - a phosphoproteome resource for systems biology research in Drosophila Kc167 cells.* Molecular Systems Biology, 2007. 3: p. 1-11.

58. Beausoleil, S.A., J. Villen, S.A. Gerber, J. Rush and S.P. Gysi, *A probability-based approach for high-throughput protein phosphorylation analysis and site localization.* Nat. Biotech., 2004. 24(10): p. 1285-1292.

59. Carrascal, M., M. Gay, D. Ovelleiro, V. Casas, E. Gelpí and J. Abian, *Characterization of the Human Plasma Phosphoproteome Using Linear Ion Trap Mass Spectrometry and Multiple Search Engines.* J. Proteom. Res., 2010. 9: p. 876–884.

60. McNulty, D. and R. Annan, *Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection.* Mol Cell Proteomics, 2008. 7: p. 971-980.

61. DeGnore, J.P. and J. Qin, *Fragmentation of phosphopeptides in an ion trap mass spectrometer.* J. Am. Soc. Mass Spectrom, 1998. 9(11): p. 1175-1188.

62. Liao, P.C., J. Leykam, P.C. Andrews, D.A. Gage and J. Allison, *An approach to locate phosphorylation sites in phosphoprotein: mass mapping by combining specific enzymatic degradation with matrix assisted laser desorption/ionization mass spectrometry.* Anal Biochem, 1994. 219: p. 9-20.

63. Kjellström, S. and O.N. Jensen, *Phosphoric acid as a matrix additive for MALDI MS analysis of phosphopeptides and phosphoproteins.* Anal Chem, 2004. 76: p. 5109-5117.

64. Hou, J., Z. Xie, P. Xue, Z. Cui, X. Chen, J. Li, et. al., *Enhanced MALDI-TOF MS analysis of phosphopeptides using an optimized DHAP/DAHC matrix.* J Biomed Biotechnol, 2010. 2010: p. 759690.

65. Roepstorff, P., *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomed.Mass Spectrom., 1984. 11: p. 601.

66. Martin, D.B., J.K. Eng, A.I. Nesvizhskii, A. Gemmill and R. Aebersold, *Investigation of neutral loss during collision-induced dissociation of peptide ions.* Anal. Chem., 2005. 77: p. 4870-4882.

67. Jin, W.H., J. Dai, H. Zhou, Q.C. Xia and et. al., *Phosphoproteome analysis of mouse liver using immobilized metal affinity purification and linear ion trap mass spectrometry.* Rapid. Commun.Mass Spectrom., 2004. 18: p. 2169-2176.

68. Boersema, P.J., S. Mohammed and A.J. Heck, *Phosphopeptide fragmentation and analysis by mass spectrometry.* J.Mass Spectrom., 2009. 44: p. 861-878.

69. Chang, E.J., V. Archambault, D.T. McLachlin, A.N. Krutchinsky and B.T. Chait, *Analysis of protein phosphorylation by hypothesis-driven multiple-stage mass spectrometry.* Anal Chem, 2004. 76(15): p. 4472-83.

70. Ulintz, P.J., A.K. Yocum, B. Bodenmiller, R. Aebersold, P.C. Andrews and A.I. Nesvizhskii, *Comparison of MS2-Only, MSA, and MS2/MS3 Methodologies for Phosphopeptide Identification.* J. Proteome Res., 2009. 8: p. 887-899.

71. Nakayasu, E.S., M.R. Gaynor, T.J. Sobreira, J.A. Ross and I.C. Almeida, *Phosphoproteomic analysis of the human pathogen Trypanosoma cruzi at the epimastigote stage.* Proteomics, 2009. 9: p. 3489–506.

72. Pan, C., J.V. Olsen, H. Daub and M. Mann, *Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics.* Mol Cell Proteomics, 2009. 8(12): p. 2796-808.

73. Arrigoni, G., S. Resjo, F. Levander, R. Nilsson, E. Degerman, M. Quadroni, et al., *Chemical derivatization of phosphoserine and phosphothreonine containing peptides to increase sensitivity for MALDI-based analysis and for selectivity of MS/MS analysis.* Proteomics, 2006. 6(3): p. 757-66.

74. Zubarev, R.A., N.L. Kelleher and F.W. McLafferty, *Electron capture dissociation ofmultiply charged protein cations. A nonergodic process.* J. Am. Chem. Soc., 1998. 120: p. 3265-3266.

75. Syka, J., J. Coon, M. Schroeder, J. Shabanowitz and D. Hunt, *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.* Proc Natl Acad Sci U S A., 2004. 101: p. 9528-9533.

76. Sweet, S.M.M., C.M. Nailey and D.L. Cunningham, *Large escale localization of protein phosphorylation by use of Electron Capture Dissociation Mass Spectrometry.* Mol.Cell. Proteomics, 2009. 8: p. 904-912.

77. Macek, B., M. Mann and J.V. Olsen, *Global and Site-Specific Quantitative Phosphoproteomics: Principles and Applications.* Annu. Rev. Pharmacol. Toxicol., 2009. 49: p. 100-221.

78. Soderblom, E.J., M. Philipp, J.W. Thompson, M.G. Caron and M.A. Moseley, *Quantitative label-free phosphoproteomics strategy for multifaceted experimental designs.* Anal Chem, 2011. 83(10): p. 3758-64.

79. Everley, P., J. Krijgsveld, B. Zetter and S. Gygi, *Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research.* Mol Cell Proteomics., 2004. 3: p. 729-735.

80. Pimienta, G., R. Chaerkady and A. Pandey, *SILAC for global phosphoproteomic analysis.* Methods Mol Biol, 2009. 527: p. 107-16, x.

81. Zhang, Y., S.B. Ficarro, S. Li and J.A. Marto, *Optimized Orbitrap HCD for quantitative analysis of phosphopeptides.* J Am Soc Mass Spectrom, 2009. 20(8): p. 1425-34.

82. Jones, A. and T. Nühse, *Phosphoproteomics using iTRAQ.* Methods Mol Biol, 2011. 779: p. 287-302.

83. Salomon, A.R., S.B. Ficarro, L.M. Brill, A. Brinker, Q.T. Phung, C. Ericson, et al., *Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry.* Proc. Natl. Acad. Sci., 2003. 100(2): p. 443-448.

84. Wolf-Yadlin A, H.S., Lauffenburger DA, White FM., *Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks.* Proc Natl Acad Sci U S A., 2007. 104: p. 5860-5865.

85. Mayya, V., K. Rezual, L. Wu, M. Fong and D. Han, *Absolute Quantitifacion of Multisite Phosphorylation by Selective Reaction Monitoring Mass Spectrometry.* Mol Cell Proteomics, 2006. 5: p. 1146-1157.

86. Puttick, J., E. Baker and L. Delbaere, *Histidine phosphorylation in biological systems.* Biochim Biophys Acta, 2008. 1784: p. 100-105.

87. Kleinnijenhuis, A., F. Kjeldsen, B. Kallipolitis, K. Haselmann and O. Jensen, *Analysis of histidine phosphorylation using tandem MS and ion-electron reactions.* Anal Chem, 2007. 79: p. 7450-7456.

Proteómica redox: análisis de modificaciones postraduccionales oxidativas

Antonio Martínez-Ruiz[1], Brian McDonagh[2], José Antonio Bárcena[3]

[1] Investigador del Sistema Nacional de Salud, Servicio de Inmunología, Hospital Universitario de La Princesa, Instituto de Investigación Sanitaria Princesa (IP), Madrid, España;

[2] Investigador, Departamento de Bioquímica y Biología Molecular, Universidad de Córdoba, España. Department of Musculoskeletal Biology, Institute of Ageing and Chronic Disease, Faculty of Health & Life Sciences, Universidad de Liverpool, Reino Unido;

[3] Catedrático, Departamento de Bioquímica y Biología Molecular, Universidad de Córdoba, España.

1. Introducción.

En este capítulo describimos la "proteómica redox"; esto es, el estudio de las modificaciones postraduccionales (PTM, del inglés post-translational modification) oxidativas de las proteínas mediante metodologías proteómicas. Sin embargo, esta definición incluye muchas modifica-ciones posibles, y en realidad nos vamos a centrar en aquellas que se estudian más frecuen-temente, principalmente porque se les asigna papeles relevantes en la funcionalidad de las proteínas. La modificación por formación de puentes disulfuro entre residuos de cisteína, que entra en la definición que hemos dado, ya ha sido descrita en otro capítulo.

Al plantearse un estudio de este tipo, hay que ver qué **objetivos** se quieren perseguir o priorizar, entre otros:

- Detección de modificaciones redox en el proteoma. ¿Qué modificación, o grupo de modificaciones, se quiere estudiar? ¿Se quiere identificar en detalle alguna modificación concreta?
- Identificación de las proteínas modificadas.
- Localización de los aminoácidos concretos modificados dentro de cada proteína.
- Cuantificación del grado de modificación de cada proteína/aminoácido.
- Definir los cambios temporales y funcionales del estado redox de cada proteína en función del tiempo, estímulos, desarrollo, etc.
- Relacionar la modificación de una proteína con su función biológica.

También se pueden plantear al menos dos formas de aproximarse a este tipo de estudios:

Mera búsqueda sin hipótesis preconcebidas de las proteínas que sufren modificacio-nes, que sería lo realmente "proteómico".

Proteómica dirigida a partir de una hipótesis de partida para estudios en proteínas indi-viduales predefinidas.

Según lo que se quiera estudiar, y por supuesto según el tipo de modificación, la forma de trabajar será distinta. En general, se puede decir que para estudiar modificaciones postraduccionales hay que introducir metodologías específicas para detectar o identificar la modificación en algún punto del flujo de trabajo de un estudio proteómico. En el caso de las modificaciones redox, aunque todas son covalentes, presentan características químicas muy diferentes entre sí, por lo que para distintas modificaciones hay que aplicar en ocasiones estrategias bastante distintas.

2. Efectos biológicos de las especies reactivas de oxígeno (ROS) y de nitrógeno (RNS).

Las modificaciones redox de las proteínas son consecuencia de la vida aeróbica, que da lugar a la presencia de especies reactivas de oxígeno (ROS, del inglés *reactive oxygen species*) y de nitrógeno (RNS, del inglés *reactive nitrogen species*), cuya producción y características no vamos a ver aquí en profundidad (1). La modificación de las proteínas puede modificar su función, de ahí su importancia. Conviene establecer dos ideas que se manejan en el campo:

- ***ROS y RNS*** pueden **dañar** a las biomoléculas. El efecto lo producen las especies radicálicas (monoelectrónicas) ($\cdot OH$, $\cdot NO_2$) o un exceso de todas ellas. El exceso no se da normalmente porque entran en juego los sistemas defensivos antioxidantes celulares, como superóxido dismutasa (SOD), catalasa, peroxirredoxinas, glutatión (GSH), etc. Sólo cuando hay un desequilibrio entre la producción y la destrucción de ROS/RNS es cuando se da el auténtico estrés oxidativo agudo y se producen **efectos patológicos** en el organismo.

  Hay un número de enfermedades en las que se ha demostrado una relación entre estrés oxidativo y patología: aterosclerosis, cáncer, envejecimiento, inflamación, enfermedades neurodenegenerativas, etc.

- Sin embargo, cada vez tiene más fundamento la idea de que las especies radicálicas se eliminan con gran eficiencia y las que predominan son las de 2 electrones, H_2O_2, $ONOO^-$; el óxido nítrico (NO), aunque es un radical, se puede añadir a estas dos especies. Su reactividad no es tan alta como su nombre indica y son relativamente **estables**; pueden atravesar las membranas (el peroxinitrito tiene carga pero puede hacerlo a través de canales iónicos); se pueden producir de forma endógena tras la activación de receptores y son metabolizadas por enzimas específicas. Por todo ello pueden actuar como **segundos mensajeros** y las PTM que producen se estudian cada vez más como mediadores en la **señalización celular** (1).

3. ¿Qué tipo de modificaciones redox sufren las proteínas?.

Podemos clasificarlas en dos grupos:

- **Irreversibles:** modificaciones que se considera que no son reversibles en ambientes bioquímicos, por ser muy estables y no disponer de sistemas enzimáticos para su reversión. La modificación se elimina por degradación de la proteína. Se suelen asociar

a daño de las proteínas y se han considerado habitualmente como modificaciones inespecíficas, estudiándose a menudo el grado de modificación como un indicador de estrés oxidativo general. Más recientemente se va viendo que sí que puede haber selectividad y especificidad en los residuos y proteínas afectadas, lo que puede afectar en mayor o menor medida a perder la función biológica o, en algunos casos, a ganar nuevas funciones.

- **Carbonilación.** Formación de grupos carbonilo (C=O) en proteínas, bien por oxidación de cadenas laterales (de p.ej. Arg, Lys y Pro), bien por adición de grupos aldehídos reactivos (en Cys, Lys, His).

- **Nitración de tirosinas.** Incorporación de un grupo nitro (-NO2) formando residuos de 3-nitrotirosina.

- **Oxidación de cisteínas a sulfínico y sulfónico (-SO2H y -SO3H).** Aunque se consi-deran irreversibles, se ha descrito una sulfirredoxina que puede revertir sulfínicos de proteínas específicas a tioles, con gasto de ATP.

- **Reversibles:** modificaciones que sí se revierten en sistemas bioquímicos, bien por enzimas específicas o por otros mecanismos. La reversibilidad hace que puedan ser útiles como mecanismos de regulación de la función de proteínas, participando en transducción de señales, o como mecanismos de protección para evitar mayores grados de oxidación.

- **Oxidación de metioninas a sulfóxido.** Reversible específicamente por metionina sulfóxido reductasa.

- **Oxidación de cisteínas a sulfénico (-SOH).** Reversible por tioles, incluyendo el glutatión.

- **Formación de puentes disulfuro en proteínas.** (-S-S-R, donde R es otra cisteina proteica). Reversible por otros tioles y proteínas como las tiorredoxinas (Trx). Se desarrolla en otro capítulo de este libro.

- **S-tiolación.** Formación de un "disulfuro mixto" (-S-S-R, donde R es un tiol de baja masa molecular). Si el tiol es glutatión (-S-SG) se habla de **S-glutationilación**, y se revierte por varias enzimas, sobre todo glutarredoxinas (Grx).

- **S-nitrosación o S-nitrosilación de cisteínas.** Formación de tionitrito o nitrosotiol (-S-N=O). Reversible por luz, por algunos cationes metálicos, por tioles como glutatión y por tiorredoxina.

La estabilidad de cada modificación es importante para su detección proteómica. Así, las modificaciones irreversibles, más estables, son por lo general más fáciles de estudiar. En el caso de las reversibles, es muy importante en la extracción conseguir inactivar los sistemas de reversión, "congelando" el estado redox del proteoma para poder estudiar las modificaciones presentes en un momento dado en un sistema biológico. En el caso de las modificaciones más lábiles, como la S-nitrosilación, es importante además tomar precauciones para evitar su desaparición espontánea (por ejemplo, manejar las muestras en oscuridad debido a que es fotolábil, y añadir quelantes de cationes metálicos).

4. Análisis de las proteínas dañadas oxidativamente: carbonilación proteica.

Los derivados carbonílicos se forman por el ataque oxidativo directo catalizado por metales (MCO, metal-catalysed oxidation) sobre las cadenas laterales de los aminoácidos **prolina, arginina, lisina y treonina.** Además, otros aminoácidos como histidina o cisteína también pueden presentarse carbonilados porque reaccionen incorporando compuestos carbonílicos (carbohidratos o lípidos, incluyendo productos finales de glicación/lipoxidación avanzadas). Los productos de las reacciones de carbonilación cuantitativamente más importantes son el **semialdehído de glutamato** derivado de Arg y Pro, y **semialdehído aminoadípico** derivado de Lys.

La carbonilación, a diferencia de otras modificaciones redox como oxidación de Met o formación de disulfuros de Cys, es un proceso **irreversible**. Por tanto, la célula debe deshacerse de las proteínas carboniladas mediante su degradación.

Se han desarrollado métodos para la detección y cuantificación de grupos carbonilos con distintos grados de sensibilidad, aunque la mayoría se basa en la derivatización del grupo carbonilo con **2,4-dinitrofenilhidrazina (DNPH)** seguida de inmunodetección con anticuerpos específicos para el grupo dinitrofenilo (DNP) añadido (revisado más exhaustivamente en (2, 3)).

En estudios proteómicos, se emplea frecuentemente una metodología clásica de geles 2D, que consiste en:

- Derivatización de la muestra con DNPH.
- Electroforesis 2D y tinción con Coomassie u otro reactivo de proteína total.
- Western blot con anti-DNP para localizar las proteínas carboniladas.
- Comparación ("matching") de las imágenes del gel y de la membrana (proteína total y modificación postraduccional) mediante análisis de imagen.
- Picado de los spots positivos en el gel.
- Identificación de las proteínas mediante MALDI-TOF.

La misma metodología se puede aplicar a la determinación de **nitrotirosina** y de **glutationilación** de cisteínas, modificaciones para las que también existen anticuerpos específicos.

La técnica tiene los **inconvenientes** de la proteómica basada en geles, a lo que se suma la dificultad de la comparación entre imágenes tomadas en un gel y en la membrana. Recientemente, se ha introducido el uso de hidrazinas con fluoróforos acoplados, lo que simplifica el proceso ya que permite ver por fluorescencia la derivatización en el propio gel (4). Además, permite comparar mucho más fácilmente las imágenes correspondientes a la proteína total y a la modificación, ya que se pueden tomar simultáneamente en el gel, aumentando así la exactitud del procedimiento.

También recientemente se ha desarrollado un protocolo con aproximación "shotgun", es decir, centrado en el estudio de los péptidos mediante LC-MS. Se emplean etiquetas isobáricas que permiten cuantificar el grado de carbonilación e incluso "mapear" los aminoácidos modificados (5).

Comentario: El grado general de carbonilación es una buena medida del estrés oxidativo y del daño general a las proteínas. Por eso, a menudo se ha considerado que la identidad de las proteínas dañadas tenía un interés relativo, porque parecía que la carbonilación sería esencialmente inespecífica ya que se detectaban principalmente las proteínas más abundantes (esto es un sesgo común a otros estudios basados en detección de modificación por 2DE) y aquellas que "trabajan" con metales. Sin embargo, en estudios recientes se ha visto que este último aspecto puede ser precisamente una fuente de especificidad, obteniéndose resultados interesantes en

relación con enfermedades neurodegenerativas. Por ejemplo, en un modelo celular relacionado con la ataxia de Friedrich, la alteración del metabolismo de hierro conduce a una oxidación selectiva de proteínas que unen magnesio al sustituirse el magnesio por hierro, produciéndose oxidaciones que fácilmente pueden afectar al centro activo de enzimas concretas (6).

5. La cisteína como diana de oxidación reversible.

La cisteína es un aminoácido poco frecuente en el proteoma de células eucarióticas (≈1,42%) y su forma aniónica, el tiolato ($R\text{-}S^-$), es el núcleofilo más potente de las proteínas, por lo que resulta muy apropiado para llevar a cabo funciones estructurales, catalíticas y reguladoras. Dada su reactividad, es susceptible de oxidación por H_2O_2, $ONOO^-$, N_2O_3, $\cdot NO_2$, dando lugar a modificaciones de tioles con distintos grados de oxidación reversible o irreversible, tal como se ha descrito en el apartado 3 (ver Figura 1) La oxidación reversible puede aliviar los efectos del estrés oxidativo, proteger a los tioles frente a la oxidación irreversible y se puede emplear para regular la actividad de las proteínas afectadas (Figura 2).

En este sentido, los denominados "interruptores tiólicos (*"thiol switches"*) o Cys susceptibles de oxidación reversible, proporcionan un mecanismo flexible de control para un buen número de procesos celulares (7). Para empezar, el tiol puede funcionar como un interruptor en el sitio activo. Son muchas las proteínas que tienen una Cys en el sitio activo y pueden regular su actividad mediante la formación de modificaciones reversibles. Las interacciones macromolecularestambién se pueden controlar por tioles originando entrecruzamientos de proteínas y pueden proporcionar un mecanismo para la agregación y el tráfico de proteínas. Finalmente, la modulación redox de cisteínas fuera del sitio activo puede actuar como un reostato que afecte a la actividad de

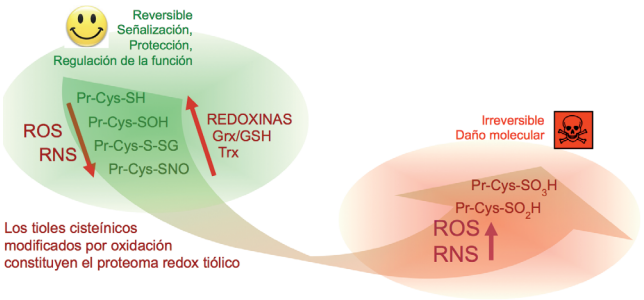


Figura 1: Efectos de ROS y RNS sobre cisteínas en proteínas mediante modificaciones postraduccionales reversibles e irreversibles.

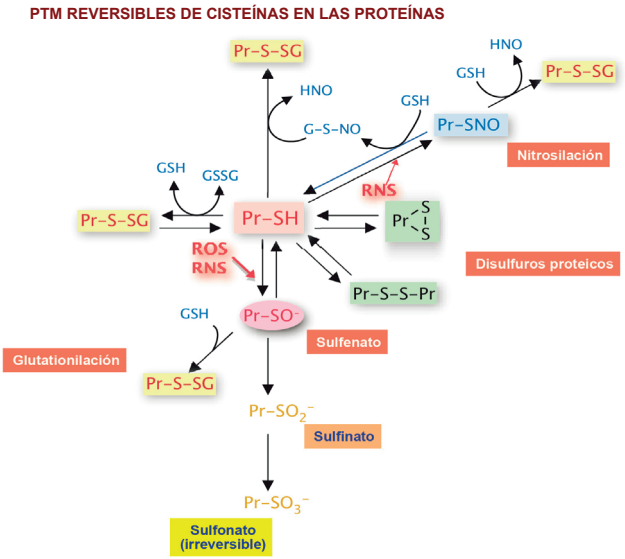


Figura 2: Esquema de las relaciones entre diversas modificaciones postraduccionales re-versibles en cisteínas de proteínas. GSH, glutatión reducido; GSSG, disulfuro de glutatión; GSNO, S-nitrosoglutatión.

as enzimas mediante formación de disulfuros e incluso con la unión de metales que a veces son necesarios para la actividad, como los complejos de hierro- azufre (Fe/S).

En resumen, en determinadas condiciones las ROS y RNS pueden **dañar** a las macromoléculas, pero en otras ocasiones, sin que se produzca estrés oxidativo, pueden tener **efectos** reguladores mediante oxidación reversible de **cisteínas**. En todos los casos, debido a la propia reactividad química y la lógica regulatoria, las modificaciones redox de cisteínas no se revierten espontáneamente cuando se recupera la homeostasis redox, sino que se regeneran con la ayuda de una familia de proteínas, las denominadas **"redoxinas" (Trx, Grx)** mencionadas en el apartado 3. Estas redoxinas están especializadas en la reducción de las formas oxidadas de Cys de vuelta a su estado normal, lo que muestra que estas PTM redox de Cys se utilizan por la célula para modular su funcionamiento como respuesta a situaciones de cambios redox en el entorno.

En presencia de elevados niveles de ROS/RNS, la oxidación de Cys puede ser más profunda y llegar a dañar irreversiblemente a las proteínas, que finalmente deben ser eliminadas de la circulación.

El conjunto de proteínas con cisteínas modificadas por oxidantes constituye el **"proteoma redox tiólico"**, cuyo análisis por métodos proteómicos se desarrolla en el siguiente apartado.

6. Análisis del proteoma redox tiólico.

El estudio directo de las distintas modificaciones oxidativas que pueden sufrir las cisteínas es complicado por varias razones. Al buscar frente a bases de datos espectros de MS o MS/MS estableciendo como modificaciones variables todos los tipos de oxidación de cisteínas se complica enormemente la búsqueda. Pero además algunas de estas modificaciones no se pueden detectar fácilmente en los espectros de masas. Por ejemplo, la S-nitrosilación se rompe con la ionización MALDI habitual, y es muy lábil para detectarla en cantidad suficiente en proteomas enteros analizados por LC-ESI-MS/MS (se ha detectado por esta técnica, pero estudiando proteínas individuales). La S-glutationilación es también difícil de detectar pues el propio glutatión incorporado es un tripéptido que se fragmenta. La oxidación de sulfénico fácilmente evoluciona durante el tratamiento de la muestra a sulfínico o sulfónico. Además, durante el propio procesamiento se pueden producir intercambios entre las oxidaciones de unas cisteínas y otras, alterando la situación que había en la muestra que se quiere estudiar.

Por eso, y sin perjuicio de usar métodos específicos que aprovechen determinadas características de cada modificación, lo más habitual es derivatizar las cisteínas utilizando reactivos para tioles que permiten a su vez marcar estos residuos para detectarlos o purificarlos mejor (para una revisión en profundidad del proteoma redox tiólico se puede consultar una excelente y exhaustiva revisión reciente (8)).

En general, la estrategia más extendida, tanto para aproximaciones con geles como sin geles, consta de 4 pasos (Figura 3):

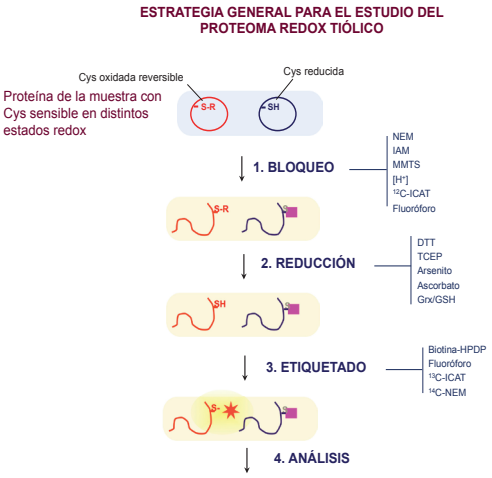


Figura 3: Estratagia general de derivatización química para el estudio del proteoma redox tiólico.

- **Bloqueo** de los tioles libres (el agente bloqueante se puede emplear simultáneamente como marcador de las Cys reducidas).

- **Reducción** de Cys oxidadas reversiblemente.

- **Marcaje** de los "nuevos" tioles (si se marcaron inicialmente las Cys reducidas durante la etapa de bloqueo, éste será el "segundo" marcaje).

- **Análisis:** visualización, purificación, identificación, localización de las Cys modificadas, cuantificación.

1) **Bloqueo:** Se trata de bloquear los tioles libres para que no reaccionen posteriormente en el paso de marcaje. Es importante conseguir un bloqueo completo, aún a costa de perder grupos modificados, para evitar la aparición de falsos positivos. Se utilizan compuestos con grupos reactivos para tioles como por ejemplo (figura 4):

- **Yodoacetamida (IAM):** la reacción es una alquilación por sustitución. Es lenta y muy dependiente del pH. A pH alto es más rápida pero puede reaccionar también con grupos amino.

- **N-etilmaleimida (NEM):** reacción de alquilación por adición. También depende del pH, y es más rápida a pH mayor, pero por encima de pH 8,0 puede dar mucha reacción con grupos amino.

- **Metilmetanotiosulfonato (MMTS):** forma un disulfuro con la cisteína, en una reacción favorecida por utilizar el sulfonato como grupo saliente. Es más específico de cisteínas, pero no se puede utilizar cuando luego se reducen puentes disulfuro.

Para evitar que el estado redox de las cisteínas cambie respecto a la situación que se quiere estudiar, y la oxidación accidental durante los primeros pasos de manipulación de las muestras, es importante intentar "congelar" las reacciones entre tioles. En muchos casos se utiliza una acidificación rápida, por ejemplo, con ácido tricloroacético (TCA), consiguiendo a la vez la precipitación de las proteínas; esto evita el intercambio de disulfuros, que se favorece a pH mayor por la formación de tiolatos, pero paradójicamente puede favorecer sin embargo la nitrosilación si hay nitrito en la muestra. También se puede aprovechar que la NEM difunde a través de membranas y reacciona rápidamente para iniciar el bloqueo al incluirla en la solución de extracción, por ejemplo en cultivos de células de mamífero.

- **Derivatización con reactivos marca-dos:** Existen derivados de IAM y NEM que son fluorescentes, o llevan una etiqueta de biotina. Cuando se usan estos reacti-vos como bloqueantes, las Cys reducidas en la muestra original quedan marcadas, lo que permite su posterior detección es-pectrofluorimétrica, o su captura por afinidad, o la cuantificación por diferencias

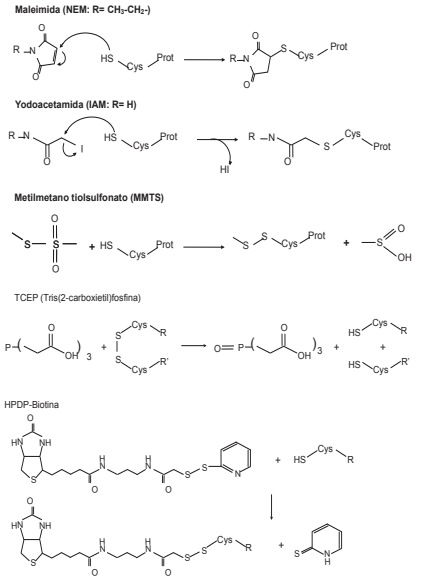


Figura 4: Esquemas de los distintos reactivos que se utilizan en los pasos del método de derivatización, con las reacciones que producen. Con la colaboración de Alicia Izquierdo-Álvarez.

en el espectro de masas. En este caso, sin embargo, es necesario controlar bien la reacción para evitar marcaje inespecífico (si el reactivo está en exceso) o que no haya un bloqueo completo (si el reactivo no se añade en cantidad suficiente).

En condiciones nativas, hay cisteínas que reaccionan mal, pues están enterradas en el interior de la proteína y darían lugar a un bloqueo incompleto. Por ello conviene hacer este paso en condiciones desnaturalizantes, añadiendo detergentes (p. ej. SDS) o una concentración suficiente de urea.

2) Reducción: es una etapa crítica. Se emplean varios agentes reductores de tioles, cada uno de los cuales tiene sus ventajas e inconvenientes. Una de las cuestiones críticas es que los agentes reductores pueden dar reacción cruzada y competir con agentes usados para la detección de los tioles proteicos en etapas posteriores del proceso, por lo que se hace necesario eliminar su exceso.

Según el reactivo utilizado, y en las condiciones adecuadas, se puede estudiar selectivamente una modificación en concreto, o varias a la vez.:

- **Ditiotreitol (DTT):** es un reactivo tiólico capaz de reducir de forma general las distintas oxidaciones que se consideran "reversibles". Frente a otros reactivos tiólicos (como el 2-mercaptoetanol), tiene la ventaja de que forma un disulfuro intramolecular, que desplaza el equilibrio de la reacción favoreciendo la reducción de los tioles con los que reacciona. Tiene el problema de su eliminación, ya que siempre deja "restos", siendo necesario quitarlo de en medio para la tercera etapa de marcaje en la mayor parte de los casos.

- **Tris(2-carboxietil)fosfina (TCEP):** es una buena alternativa al DTT; las tri-alquil fosfinas son potentes reductores de disulfuros orgánicos en condiciones acuosas suaves y cuantitativas (Figura 4). La fuerza del enlace P=O del producto oxidado hace la reacción totalmente irreversible. Se puede emplear a concentraciones mucho menores que el DTT, lo que es una ventaja a la hora de considerar la reacción posterior de marcaje.

- **Arsenito sódico:** reduce selectivamente los ácidos sulfénicos (-SOH). Normalmente los sulfenatos no suelen resistir intactos hasta este paso, por lo que este reactivo no es operativo aquí. Una alternativa para detectar el ácido sulfénico en cisteínas es marcarlos directamente con **dimedona** antes del bloqueo (9).

- **Ascorbato:** el reactivo empleado para reducir selectivamente nitrosotioles (-SNO), desde el protocolo de "biotin switch" descrito por Jaffrey (10). Se ha discutido sobre la especificidad de este reactivo cuando no se usa en las condiciones óptimas (11). No es necesaria su eliminación, por lo que el paso de marcaje se puede hacer simultáneamente.

- **Glutarredoxina con glutatión reducido (Grx/GSH):** Para dirigir la reducción selectivamente hacia aquellas cisteínas que son sustratos in vivo de la Grx, como las glutationiladas. Puesto que la misma Grx es una proteína, la reacción se tiene que hacer en condiciones que no sean desnaturalizantes (12).

3) Marcaje o etiquetado: El marcaje depende de lo que se vaya a hacer después. Se usan generalmente reactivos con los mismos (o similares) grupos funcionales utilizados en el paso de bloqueo (maleimida, yodoacetamida, etc.), conjugados con la etiqueta de elección.

Una de las etiquetas más usadas es la **biotina**, puesto que es relativamente pequeña, fácil de detectar y útil para purificar las proteínas marcadas, como veremos a continuación. Se puede incorporar de forma irreversible o reversible (por ejemplo, con un reactivo que forme enlaces disul-furo, o con reactivos que incluyan algún enlace lábil en medio ácido). En al análisis habrá

que tener en cuenta que algunas proteínas están biotiniladas endógenamente, aunque no en cisteínas.

Otra posibilidad es el marcaje con **fluoróforos** para visualizar las proteínas marcadas, por ejemplo en un gel de electroforesis. A la hora de elegir fluoróforos para electroforesis, además de considerar que no interfieran ópticamente entre ellos o con otro reactivo, hay que tener en cuenta si modifican la movilidad de la proteína; por ejemplo, los fluoróforos cargados desplazan la proteína en isoelectroenfoque (hay una revisión reciente en (13)).

También se ha empleado el marcaje radiactivo, que permite una detección muy sensible y sin diferencias químicas, aunque con las desventajas del manejo de radiactividad.

4) Análisis:

A) Una primera posibilidad es purificar el **"proteoma redox tiólico"** mediante captura por afinidad de las proteínas enteras, o mejor aún, de los péptidos concretos que se marcan (**"peptidoma redox tiólico"**) (Figura 5) (14). Este segundo método tiene la ventaja de que además de identificar la proteína, se localiza la Cys concreta modificada. En ambos casos, el método de purificación más difundido es el de la **avidina/biotina**. Es decir, se etiquetan los nuevos tioles con un reactivo de biotina y se purifican mediante unión por afinidad a una matriz conjugada con avidina (en una de sus múltiples variantes). Presenta la ventaja de la fuerte interacción entre la avidina y la biotina, que permite realizar lavados muy selectivos. Pero también tiene la desventaja de que puede ser muy difícil la elución de las proteínas o péptidos unidos. La utilización de reactivos con biotina que se puedan romper dejando la biotina unida a la avidina es una ventaja para la elución. Entre estos reactivos, la **HPDP-biotina** es muy conveniente, pues se une a la proteína/péptido mediante un enlace disulfuro sensible a los reactivos de tioles habituales (Figura 4). Por ello, se puede despegar de la resina de avidina con 2-mercaptoetanol o con DTT, rindiendo el péptido intacto. Esto facilita la interpretación de los espectros de MS/MS posteriormente, pues los péptidos no llevan biotina, que de otro modo daría lugar a fragmentaciones complejas.

También se han utilizado con éxito matrices activadas con grupos que reaccionen directamente con tioles (por ejemplo, tiol-Sepharose activada, ATS) uniéndose directamente a las proteínas o péptidos sin nece-sidad de etiquetas in-termedias (15, 16). En este caso, se juntan el marcaje y el análisis por purificación, simplificando el proceso. Tiene la desventaja de que no se puede monitorizar el marcaje.

En estudios específicos para la búsqueda de dianas de Trx en plantas y cianobacterias, se ha empleado una estrategia basada en la captura de proteínas diana mediante cromatografía de afinidad con Trx monotióli-

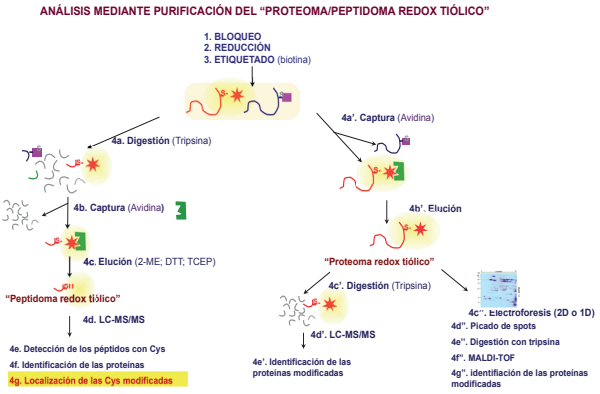


Figura 5: Análisis del proteoma redox tiólico mediante biotinilación y purificación del proteoma marcado. Se muestran distintos tipos de análisis: purificación del peptidoma redox tiólico y análisis por LC-MS/MS, purificación del proteoma redox tiólico y análisis por LC-MS/MS, purificación del proteoma redox tiólico y análisis por electroforesis.

ca inmovilizada. Esta forma artificial de Trx carece de una de las dos Cys del sitio activo y al no poder completar la reacción de reducción de los disulfuros de las proteínas sustrato presentes en el extracto, éstas quedan atrapadas en mitad del ciclo catalítico y unidas a la Trx por un disulfuro. Las proteínas diana se eluyen después con DTT o 2-mercaptoetanol y se analizan (8).

En todos estos casos el análisis se lleva a cabo por las estrategias habituales en proteómica, principalmente:

Si se ha purificado el "peptidoma redox", se analizan los péptidos mediante LC-MS/MS sin el empleo de geles ("shotgun") (Figura 5). La digestión se lleva a cabo tras el marcaje con biotina, pero antes del enriquecimiento en matriz. Lo importante de esta estrategia basada en péptidos es que la mezcla que se analiza en el espectrómetro es poco compleja (teóricamente, sólo contiene aquellos péptidos con Cys que estaban oxidados o reducidos, según la derivatización que se haya hecho) y que se localiza la Cys exacta que ha sufrido la modificación redox.

Si se ha purificado el "proteoma redox tiólico" se pueden analizar las proteínas purificadas mediante LC-MS/MS ("shotgun", aunque en este caso deberían estar todos los péptidos de las proteínas derivatizadas, y no sólo los que contienen Cys), o bien mediante separación en gel (2DE, pero también 1DE, si la mezcla es poco compleja), análisis diferencial, digestión e identificación por huella peptídica y/o fragmentación de péptidos (Figura 5).

- B) Un segundo gran tipo de análisis es la detección de las proteínas marcadas sin purificación (Figura 6). Para eso, se emplean principalmente técnicas de **geles 2DE** clásicas, como ya se ha comentado para el caso de la carbonilación. Cuando las muestras se han marcado con fluoróforo(s) se pueden visualizar directamente en el gel mediante detec-ción **fluorescente**, y se puede co-detectar también la cantidad de proteína en cada punto utilizando un reactivo fluorescente para proteína total (p. ej. Sypro Ruby) (13, 17, 18). El marcaje de **biotina** se puede detectar mediante transferencia a membrana y detección con avidina, como un Western blot. Si el marcaje se ha incorporado mediante enlace di-sulfuro, hay que tener la precaución de no utilizar reductores de tioles en la electroforesis.

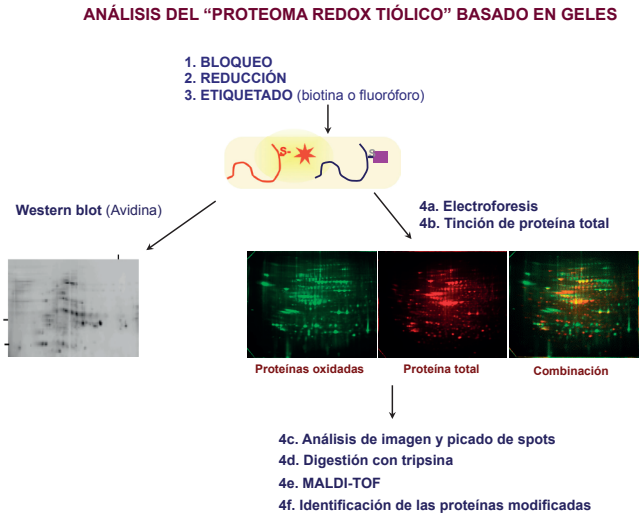


Figura 6: Análisis del proteoma redox tió-lico mediante marcaje (fluorescente o por biotina) de cisteínas oxidadas y electrofo-resis bidimensional. En este caso, el mar-caje indirecto de las cisteínas, combinado con el marcaje de proteína total, permite obtener variaciones en la oxidación de cis-teínas, controlando la cantidad de proteína en cada punto. Experimentos fluorescen-tes realizados por Elena Ramos y Alicia Izquierdo-Álvarez.

5) Comentarios:

- A) En todos los métodos descritos hasta aquí se analiza cada muestra por separado. Para comparar muestras (por ejemplo, control y problema) se deben hacer suficientes (al menos 3) réplicas de cada análisis con el fin de compensar las variabilidades biológica y metodológica y obtener valores diferenciales estadísticamente significativos.

- B) Cuando se detecta un determinado péptido o proteína en una muestra significa que la Cys estaba oxidada, pero cuando no se detecta puede significar varias cosas:

  - En el caso ideal, la Cys está reducida y queda oculta tras la etapa de bloqueo.

  - La Cys está sobreoxidada (irreversible) y no se puede recuperar en la etapa de reducción.

  - La proteína no está presente en la muestra porque la célula no la expresa o porque sus niveles están por debajo del umbral de detección de la técnica empleada. Los cambios en la cantidad de proteínas se pueden analizar simultáneamente en el método "GELSILOX" (ver más adelante), o en métodos basados en detección fluorescente en geles.

  - No se ve por la propia limitación en las técnicas de LC-MS, aunque no quiere decir que no esté. Por ejemplo, en el análisis del peptidoma redox tiólico, puede ocurrir que el péptido no se ionice bien.

- C) La principal limitación de los métodos descritos es que **no se puede cuantificar** el grado de oxidación. Esto es de importancia crucial y se están desarrollando protocolos adaptados para cuantificar el grado de oxidación de cada proteína modificada, entre los cuales se puede citar el GELSILOX (ver más abajo).

- D) Al tratarse de un método no cuantitativo, una excesiva sensibilidad del espectrómetro de masas puede llegar a ser contraproducente, pues es sabido que muchas proteínas no se encuentran en estado reducido al 100% en condiciones normales, sino que están parcialmente oxidadas hasta un cierto nivel. En el caso de proteínas abundantes, se detectaría el péptido oxidado tanto en el control como en una muestra problema sin establecer diferencias, aunque el grado de oxidación puede ser significativamente mayor en la muestra problema y tener consecuencias funcionales. Por tanto, con esta aproximación experimental, se debe llegar a un compromiso de sensibilidad que no detecte los niveles basales de péptidos oxidados, o recurrir a métodos de cuantificación entre muestras (ver abajo).

- E) En el método del peptidoma redox tiólico la identificación de la proteína se hace a partir de un único péptido, por lo que es crucial la calidad de los espectros y un alto "score" del motor de búsqueda. Por esta misma razón, la detección depende de que el péptido que porta la Cys modificada sea susceptible de análisis por MS/MS (ionización, tamaño, hidrofobicidad, hidrofilicidad).

- F) Los protocolos contienen muchas etapas, por lo que se debe ser extremadamente escrupuloso para conseguir una alta reproducibilidad entre muestras y normalizar en las etapas críticas. La cantidad total de proteína debe igualarse entre muestras rigurosamente.

A pesar de todas estas limitaciones, estos métodos tienen un enorme **potencial de escrutinio** y han permitido descubrir numerosas proteínas diana de regulación redox.

7. Análisis cuantitativo de las modificaciones redox tiólicas.

Las estrategias que hemos descrito se han aplicado a menudo con el objeto de identificar qué modificaciones postraduccionales se dan en muestras concretas, lo que se puede considerar análisis cualitativo. Pero la incorporación de los métodos de proteómica cuantitativa "general" a la proteómica redox tiene varios aspectos interesantes, pues según cómo se haga la aplicación de las distintas estrategias, se pueden aprovechar para distintos tipos de cuantificación. En proteómica general, se cuantifica la cantidad de una proteína, o de una forma proteica, entre varias muestras. En nuestro caso, sin embargo, tenemos varias variables a cuantificar:

- cantidad de la proteína (o forma proteica) = P
- cantidad de la forma oxidada de la proteína = Ox
- cantidad de la forma reducida de la proteína = Re

(y esto si consideramos sólo un tipo de modificación oxidativa, o todas en conjunto), además de las relaciones entre ellas:

- forma reducida normalizada por cantidad de proteína = Re/P
- ídem para la forma oxidada = Ox/P
- proporción de oxidada frente a reducida = Ox/Re

Y todo esto, comparando entre distintas muestras… Por tanto, está claro que la cuantificación es algo más complicada en proteómica redox. Sin embargo, recientemente se han hecho varios avances prácticos en el campo, asociados al esquema de derivatización que hemos descrito, y que se resumen a continuación.

En las metodologías de **2-DE**, se ha aprovechado el **marcaje radiactivo** para detectar y cuantificar la proteína oxidada (bloqueo con IAM "normal", reducción de todos los tioles, marcaje con NEM radiactiva), y normalizar con respecto a la cantidad de proteína total en cada punto, medida por señal de Coomassie (19). Se obtiene así una ratio Ox/P (que, al no estar calibrada, no se puede interpretar en términos de número de Cys oxidadas por molécula de proteína).

La metodología DIGE de electroforesis cuantitativa diferencial mediante **marcaje fluorescente** se ha aprovechado en varias formas en proteómica redox. Hay que tener en cuenta que el DIGE "de saturación" aprovecha dos fluoróforos acoplados a maleimida que reaccionan con cisteínas, lo que invita a usarlos en proteómica redox para ver diferencias en cisteínas reducidas u oxidadas. Así, se pueden marcar de forma "directa" las cisteínas reducidas, o de forma "indirecta" las oxidadas (Figura 7). En ambos casos, la señal de proteína modificada se puede relacionar con la de

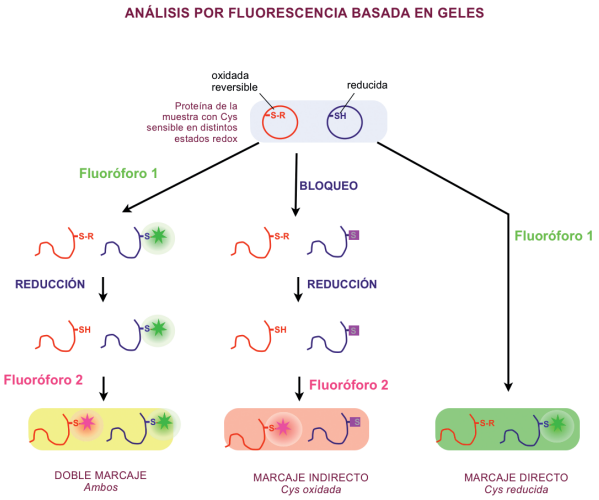


Figura 7: Esquema de tipos de marcaje con fluoróforos (equivalente para otros marcajes): marcaje directo, que detecta las cisteínas reducidas; marcaje indirecto, tras bloqueo y reducción, que detecta las cisteínas oxidadas, y marca-je doble, combinando ambos.

la proteína total en cada punto (Figura 6), obteniendo ratios Re/P u Ox/P. Cuando se utilizan dos fluoróforos para marcar muestras distintas, se puede hacer la cuantificación relativa entre muestras, corriéndolas en el mismo gel (con o sin estándar interno). Sin embargo, en este caso no se obtiene la señal de proteína total para cada punto en cada muestra, por lo que no es tan ventajoso como en proteómica general.

En algunos trabajos también se han aprovechado dos fluoróforos distintos para marcar las cisteínas reducidas y las oxidadas (Figura 7), en un doble marcaje por cada muestra, que permite obtener una ratio Ox/Re y compararla entre muestras distintas. Sin embargo, en este tipo de protocolo hay que asegurarse de que se están marcando adecuadamente ambas poblaciones de cisteínas, y no es fácil optimizarlo. Las concentraciones que se suelen usar de agente alquilante en la etapa de bloqueo son muy grandes, para asegurar un bloqueo completo, aunque haya parte de reacción inespecífica; sin embargo, si el bloqueante es un fluoróforo, el bloqueo total suele ir asociado con marcaje inespecífico, y el marcaje específico suele ir asociado a bloqueo incompleto.

Las estrategias de análisis cuantitativo por **marcaje isotópico** también se han aprovechado de varias formas en proteómica redox. Los reactivos ICAT, tan usados en proteómica general (ver capítulo de "Proteómica cuantitativa"), son derivados de yodoacetamida, por lo que también son una invitación a la proteómica redox tiólica. Así, se pueden comparar cuantitativamente cisteínas reducidas u oxidadas en dos muestras distintas derivatizando de forma directa o indirecta con cada reactivo ICAT una muestra. La gran ventaja de esta aproximación es que la misma metodología ICAT incluye la purificación de los péptidos derivatizados en cisteínas mediante avidina/biotina.

Se han utilizado etiquetas isobáricas (iTRAQ) para la cuantificación diferencial del estado redox de las Cys en dos muestras diferentes en un único experimento (20). El método da buenos resultados para el análisis de una proteína modelo in vitro, pero todavía está en fase de optimización para su aplicación a muestras complejas en estudios masivos del proteoma redox tiólico. La metodología iTRAQ sí se ha aplicado para estudiar la cinética de desnitrosilación de proteínas en líneas celulares en cultivo (16).

También se pueden usar otros reactivos, como el marcaje con 18O, si se acopla este marcaje con el proceso de purificación de péptidos oxidados o reducidos (por biotina/avidina o por resina activada para tioles).

Los reactivos ICAT también se han aprovechado para marcar sucesivamente las proteínas reducidas y oxidadas de una sola muestra, obteniendo una ratio Ox/Re, en un método que se ha llamado **"OxICAT"** (Figura 8) (21). La muestra se bloquea en condiciones desnaturalizantes con el ICAT ligero (12C) y en la etapa de marcaje se incuba con el ICAT pesado (13C). De esta forma, tanto las cisteínas reducidas originalmente como las oxidadas llevan la misma etiqueta, sólo que las primeras llevan la etiqueta "ligera" y las segundas la etiqueta "pesada". Luego se juntan y se analizan en la misma carrera de LC-MS en donde se

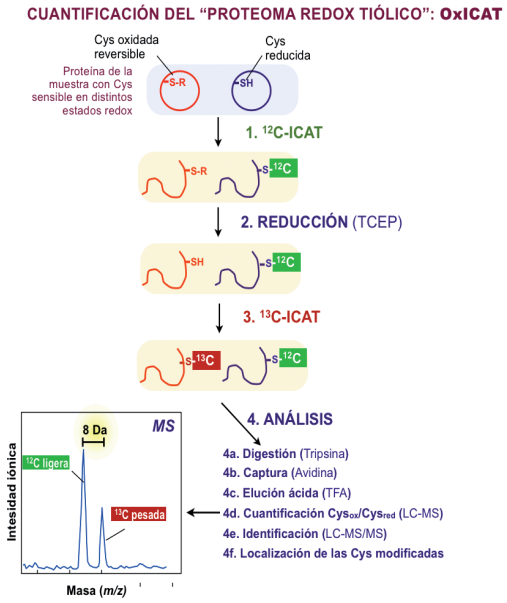


Figura 8: Esquema de cuantificación del proteoma redox tiólico mediante la técnica "oxICAT" de marcaje diferencial de las cisteínas reducidas y oxidadas en una misma muestra.

detectan como pares de iones parentales que difieren en unas pocas unidades de masa. Las intensidades de la señal (área de los picos) se comparan y se obtiene una ratio que debe ser equivalente a la ratio de péptido oxidado y reducido en la muestra original.

El método se ha empleado para la cuantificación del estado redox de proteínas individuales in vitro, como prueba de concepto, demostrando su validez. Sin embargo, un análisis "high-throughput" en extractos de E.coli determinó sólo ratios comprendidas entre 1,6 y 2,3 para las proteínas identificadas en células sometidas a estrés oxidativo comparadas con células control (21). La limitación es parecida a la que hemos descrito para el marcaje simultáneo con dos fluoróforos, y el protocolo, según nuestra experiencia, parece difícil de adaptar para otros tipos de células.

Recientemente, también se ha desarrollado un método, que se ha denominado **"Gelsilox"** (22), por el que se aprovecha el análisis masivo de péptidos mediante LC-MS para hacer una cuantificación relativa (entre dos muestras) de proteína total, cisteínas reducidas y cisteínas oxidadas (Figura 9). En las reacciones, este método tienen una gran ventaja, pues se utilizan reactivos asequibles, que se pueden añadir a gran concentración tanto para bloqueo como para marcaje (NEM e IAM respectivamente, aunque se pueden cambiar), asegurándose así la eficacia de ambos procesos. La otra gran ventaja es aprovechar el análisis masivo haciendo varias fracciones del peptidoma, para mejorar la calidad de las cuantificaciones mediante el adecuado análisis estadístico. Básicamente, el método consiste en: bloqueo con NEM (durante la extracción), reducción con DTT y marcaje con IAM (estos dos pasos se hacen en matriz de acrilamida), digestión de ambas muestras y marcaje de una de ellas con 18O. Ambos digeridos se mezclan para proceder a la separación en fracciones que se inyectan en LC-MS/MS. Los péptidos que no tienen Cys se aprovechan para determinar la cantidad relativa de cada proteína, mientras que los péptidos derivatizados con NEM o con IAM se aprovechan para cuantificar la variación en la forma reducida u oxidada de cada péptido con Cys (22).

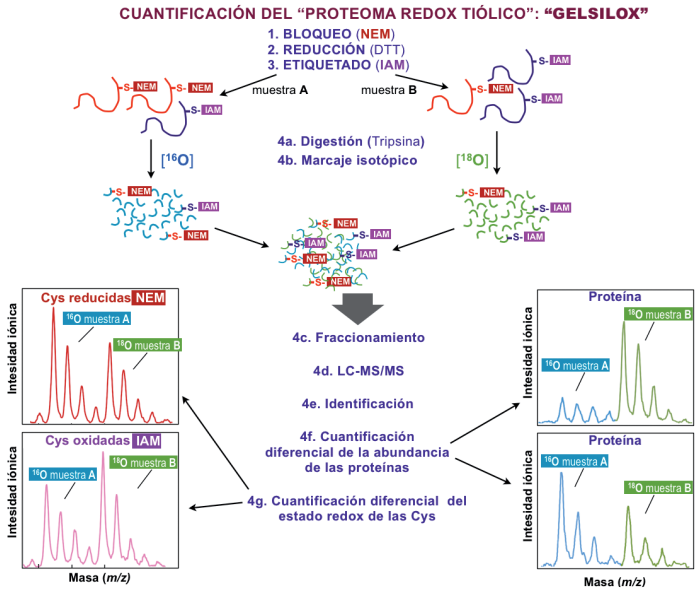


Figura 9: Estrategia de cuantificación simultánea del proteoma total y del proteoma redox tiólico seguida en la técnica GELSILOX, que permite el marcaje simultáneo de cisteínas reducidas y oxidadas, analizando cuantitativamente mediante LC-MS ambas poblaciones de péptidos derivatizados, así como los péptidos sin Cys, que permiten cuantificar la proteína total presente en dos muestras. Adaptada de un esquema de Pablo Martínez-Acedo y Jesús Vázquez.

REFERENCIAS.

1. D'Autréaux B, Toledano MB. ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis. Nat Rev Mol Cell Biol. 2007; 8: 813-24.

2. Irazusta V, Moreno-Cermeño A, Ros J, Tamarit J. Estrategias proteómicas para el análisis de grupos carbonilo como marcadores de daño oxidativo en proteínas. Proteómica. 2008; 2: 51-8.

3. Irazusta V, Moreno-Cermeño A, Cabiscol E, Tamarit J, Ros J. Proteomic strategies for the analysis of carbonyl groups on proteins. Curr Protein Pept Sci. 2010; 11: 652-8.

4. Tamarit J, de Hoogh A, Obis E, Alsina D, Cabiscol E, Ros J. Analysis of oxidative stress-induced protein carbonylation using fluorescent hydrazides. J Proteomics. 2012; 75: 3778-88.

5. Kim J-H, Sedlak M, Gao Q, Riley CP, Regnier FE, Adamec J. Dynamics of Protein Damage in Yeast Frataxin Mutant Exposed to Oxidative Stress. OMICS. 2010; 14: 689-99.

6. Irazusta V, Moreno-Cermeño A, Cabiscol E, Ros J, Tamarit J. Major targets of iron-induced protein oxidative damage in frataxin-deficient yeasts are magnesium-binding proteins. Free Radic Biol Med. 2008; 44: 1712-23.

7. Klomsiri C, Karplus PA, Poole LB. Cysteine-Based Redox Switches in Enzymes. Antioxid Redox Signal. 2010; 4: 1065-77.

8. Lindahl M, Mata-Cabana A, Kieselbach T. The disulfide proteome and other reactive cysteine proteomes: analysis and functional significance. Antioxid Redox Signal. 2011; 14: 2581-642.

9. Nelson KJ, Klomsiri C, Codreanu SG, Soito L, Liebler DC, Rogers LC, et al. Use of dimedone-based chemical probes for sulfenic acid detection methods to visualize and identify labeled proteins. Methods Enzymol. 2010; 473: 95-115.

10. Jaffrey SR, Erdjument-Bromage H, Ferris CD, Tempst P, Snyder SH. Protein S-nitrosylation: a physiological signal for neuronal nitric oxide. Nat Cell Biol. 2001; 3: 193-7.

11. Forrester MT, Foster MW, Benhar M, Stamler JS. Detection of protein S-nitrosylation with the biotin-switch technique. Free Radic Biol Med. 2009; 46: 119-26.

12. Lind C, Gerdes R, Hamnell Y, Schuppe-Koistinen I, von Löwenhielm HB, Holmgren A, et al. Identification of S-glutathionylated cellular proteins during oxidative stress and constitutive metabolism by affinity purification and proteomic analysis. Arch Biochem Biophys. 2002; 406: 229-40.

13. Izquierdo-Álvarez A, Martínez-Ruiz A. Thiol redox proteomics seen with fluorescent eyes: the detection of cysteine oxidative modifications by fluorescence derivatization and 2-DE. J Proteomics. 2011; 75: 329-38.

14. McDonagh B, Ogueta S, Lasarte G, Padilla CA, Bárcena JA. Shotgun redox proteomics identifies specifically modified cysteines in key metabolic enzymes under oxidative stress in Saccharomyces cerevisiae. Journal of Proteomics. 2009; 72: 677-89.

15. Hu W, Tedesco S, Mcdonagh B, Bárcena JA, Keane C, Sheehan D. Selection of thiol- and disulfide-containing proteins of Escherichia coli on activated thiol-Sepharose. Analytical Biochemistry. 2010; 398: 245-53.

16. Forrester MT, Thompson JW, Foster MW, Nogueira L, Moseley MA, Stamler JS. Proteomic analysis of S-nitrosylation and denitrosylation by resin-assisted capture. Nat Biotechnol. 2009; 27: 557-9.

17. Tello D, Tarín C, Ahicart P, Bretón-Romero R, Lamas S, Martínez-Ruiz A. A "fluorescence switch" technique increases the sensitivity of proteomic detection and identification of S-nitrosylated proteins. Proteomics. 2009; 9: 5359-70.

18. Izquierdo-Álvarez A, Ramos E, Villanueva J, Hernansanz-Agustín P, Fernández-Rodríguez R, Tello D, et al. Differential redox proteomics allows identification of proteins reversibly oxidized at cysteine residues in endothelial cells in response to acute hypoxia. J Proteomics. 2012; 75: 5449-62.

19. Le Moan N, Clement G, Le Maout S, Tacnet F, Toledano MB. The Saccharomyces cerevisiae proteome of oxidized protein thiols: contrasted functions for the thioredoxin and glutathione pathways. J Biol Chem. 2006; 281: 10420-30.

20. McDonagh B, Martinez-Acedo P, Vazquez J, Padilla CA, Sheehan D, Barcena JA. Application of iTRAQ Reagents to Relatively Quantify the Reversible Redox State of Cysteine Residues. Int J Proteomics. 2012; 2012: 514847.

21. Leichert LI, Gehrke F, Gudiseva HV, Blackwell T, Ilbert M, Walker AK, et al. Quantifying changes in the thiol redox proteome upon oxidative stress in vivo. Proc Natl Acad Sci U S A. 2008; 105: 8197-202.

22. Martínez-Acedo P, Núñez E, Sánchez-Gómez FJ, Moreno M, Ramos E, Izquierdo-Álvarez A, et al. A novel strategy for global analysis of the dynamic thiol redox proteome. Mol Cell Proteomics. 2012; 11: 800-13.

Proteómica de Organismos Huérfanos

24

María-Luz Valero; Laura Cantero[1]

[1] Laboratorio de Proteómica, Centro de Investigación Principe Felipe. Valencia, España.

1. Introducción.

En el estudio de los sistemas biológicos, las aproximaciones experimentales basadas en el análisis de ARNm (ácido ribonucleico mensajero) son muy potentes y están altamente automatizadas. Aun así, es importante reconocer que pueden existir discrepancias entre el mensajero (transcrito) y su efector final (proteína madura). Y puesto que la mayoría de funciones biológicas en la célula son realizadas por proteínas y no por ARNm, los perfiles de transcripción no proporcionan siempre información adecuada para la descripción de un sistema biológico [1]. En este contexto, resulta evidente que mecanismos de control post-transcripcional como la regulación del "splicing" y del transporte de los ARNm tienen importantes efectos en los sistemas biológicos, igual que las vidas medias de los ARNm y los mecanismos prost-traduccionales de modificación de proteínas o el tráfico intracelular de las mismas [2-4]. Estudios de expresión en procariotas y en organismos eucariotas más o menos complejos (Figura 1, [5]), incluyendo un gran número de muestras humanas, han corroborado este hecho y revelado una pobre correlación entre la cantidad de ARNm presente en una muestra y la abundancia tanto del nivel de ARNm transcrito como de proteína [6-13]. Adicionalmente, estudios proteómicos realizados en la bacteria *Pseudomonas fluorescens* han permitido identificar proteínas correspondientes a regiones génicas catalogadas como no codificantes, algunas proteínas identificadas en el estudio son antisentido para genes predichos, otras intergénicas y una se localiza en la misma dirección que un gen anotado pero en una pauta de lectura diferente [14]. Todo lo cual demuestra que para la descripción global de un

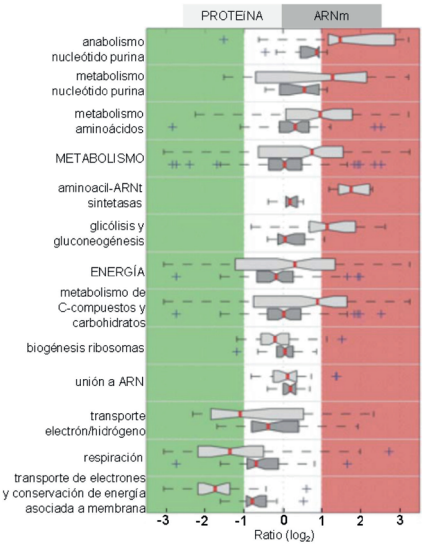


Figura 1. Diagrama de caja de relaciones de ARNm y proteína para varias categorías funcionales determinadas por de Groot MJ, et al.[5]. Las proporciones de proteína y ARNm están representadas por las cajas gris claro y gris oscuro respectivamente. Los límites de un cuadro indican los valores inferiores y superiores del cuartil, la línea representa la mediana. Las líneas discontinuas que se extienden desde los extremos de las cajas muestran la medida del resto de los datos. Se indica por "+" valores fuera de los rangos establecidos Las muescas de la caja representan una estimación sólida de la incertidumbre acerca de la mediana para la comparación de las cajas. Así, las cajas cuyas muescas no se sobreponen indican que las medianas de los dos grupos difieren con una significación del 95%.

sistema biológico es necesaria una caracterización del mismo a nivel de proteína. Para ello son imprescindibles técnicas sensibles y de alto rendimiento, motivo por el que se ha generalizado el uso de la proteómica para la caracterización global de los sistemas biológicos a nivel de proteína.

El impulso de la proteómica derivado de este hecho ha sido asistido por un aumento, en los últimos años, de la efectividad de las técnicas utilizadas [15] tanto a nivel de separación como de identificación y caracterización, cualitativa y cuantitativa, de proteínas. Así, se ha venido confiando en técnicas proteómicas para la obtención de información completa de los sistemas biológicos con notable éxito en muchos casos [16-19]. La proteómica ha permitido, incluso, obtener información detallada de algunas modificaciones post-traduccionales de gran relevancia biológica entre las que destaca la fosforilación de proteínas [20-25].

A pesar de los muchos logros obtenidos, lo cierto es que la proteómica no ha satisfecho totalmente las expectativas generadas durante sus inicios [26-28]. Una de las razones fundamentales de ello es que los estudios de proteómica global solo han podido realizarse en organismos modelo, entendiéndose organismo modelo en el campo de la proteómica como aquel bien descrito a nivel de proteína. Este hecho ha excluido o minimizado el uso de las técnicas proteómicas en campos de gran relevancia científica y económica donde no se trabaja con estos organismos. Y el problema se agrava para aquellos organismos mal representados en las bases de datos utilizadas habitualmente en los análisis proteómicos [29]. A estos organismos los llamaremos en adelante organismos huérfanos.

En este contexto es fácil imaginar que existen campos donde estudios proteómicos podrían haber proporcionado información muy útil como la parasitología [30-32]; la ecología [33-35], la ecotoxicología [36, 37], la agricultura [38-41], la acuicultura [42, 43], la alimentación [43-45] o la biología de poblaciones [46] por citar algunos ejemplos. Todas estas áreas de estudio tienen en común que habitualmente no se trabaja con organismos modelo y en muchos casos, además, se estudian organismos huérfanos.

Tabla 1. Ejemplos de utilización de técnicas proteómicas en el estudio de organismos no modelo.

Área de Estudio	Referencia bibliográfica	Organismo estudiado	Número de Proteínas (1)
Parasitología	Bellafiore S, et al. 2008	Meloidogyne incognita	168
	Robinson MW, et al. 2009	Fasciola hepatica	833
	Chuan J, et al. 2010	Schistosoma japonicum	24289
		Schistosoma mansoni	27680
		Brugia malayi	23318
Ecología	Parker R, et al. 2010	Apis Mellifera	15544
	Apraiz I, et al. 2006	Mytilus edulis	3088
	Ripley J, et al. 2008	Micropterus dolomieu	134
Ecotoxicología	Bundy JG, et al. 2008	Lumbricus rubellus	110
	Hogstrand C, et al. 2004	Oncorhynchus mykiss	6480
	Khan FR, et al. 2010	Gammarus pulex	12
Agricultura	Amaralj RS, et al. 2010	Saccharum officinarum	1133
	Acín P, et al. 2009	Sesamia nonagrioides	25
	Barker RL, et al. 2009	Zea mays	106513
	Forné I, et al. 2009	Review Article	
Acuicultura	Varó I, et al. 2010	Sparus aurata	557
	Yan F, et al. 2010	Scylla serrata	125
	Li P, et al. 2010	Acipenser baerii	329
Alimentación	Wu R, et al. 2009	Lactobacillus casei Zhang	5698
	Xu YJ, et al. 2009	Sus scrofa	36050
	Kussmann M, et al. 2010	Review article	
Biología de Poblaciones	González EG, et al. 2010	Merluccius merluccius	150
Biología evolutiva	Buggiotti L, et al. 2010	Ficedula albicollis	367
		Ficedula hypoleuca	302

(1) Número de proteínas del organismo en NCBI a 13/12/2010.

A pesar de que la proteómica global no es fácilmente abordable en estos casos, en los ejemplos citados en la tabla 1 [29-36, 38, 40-51] puede comprobarse que las técnicas de proteómica han demostrado su capacidad para ayudar a los investigadores a hacer frente a un creciente

número de cuestiones biológicas en este tipo de organismos. Lo que ha determinado que se sigan haciendo esfuerzos para optimizar la información obtenida en los experimentos proteómicos de organismos huérfanos.

En este capítulo intentaremos clarificar dónde radican los problemas para el análisis proteómico de los organismos huérfanos y procuraremos posibles soluciones a estos problemas.

2. Origen de la dificultad de los análisis proteómicos de organismos huérfanos.

Aunque existen ejemplos de proteómica 'Top-down' aplicada a organismos no modelo y/o no secuenciados [52-54] la mayoría de estos estudios de proteómica se abordan por la estrategia conocida como proteómica "Bottom-up". En esta aproximación la identificación de las proteínas se realiza a través del análisis por MS (espectrometría de masas) y/o MS/MS (espectrometría de masas en tándem) de los péptidos resultantes de digerir las proteínas con un enzima proteolítico, habitualmente tripsina.

La estrategia 'Bottom-up' se compone de una serie de etapas consecutivas que podrían resumirse brevemente de la siguiente forma (Figura2):

Etapa 1. Obtención de la muestra de proteínas a analizar, debe garantizar que el producto obtenido es representativo del proteoma en estudio.

Etapa 2. Opcional, consiste en un fraccionamiento de las proteínas de la mezcla para reducir su complejidad.

Etapa 3. Digestión de las proteínas de la muestra con un enzima proteolítico.

Etapa 4. Separación de los péptidos proteolíticos obtenidos en la etapa anterior antes de su análisis por MS y/o MS/MS. Puede constar de un único paso de separación o de varios consecutivos y en la mayoría de los análisis suele acabar con una separación cromatográfica acoplada a un espectrómetro de masas.

Etapa 5. Obtención de información a nivel de masa o de masa y fragmentación de los distintos péptidos proteolíticos.

Etapa 6. Análisis de la información de MS o MS/MS obtenida para identificar y caracterizar las distintas proteínas presentes en la muestra de partida. Habitualmente se realiza consultando bases de datos de proteínas con distintos algoritmos de búsqueda.

En este esquema de análisis la dificultad de trabajar con organismos huérfanos radica fundamentalmente en la última etapa, la de identificación de la proteína a partir de los datos obtenidos de MS-MS/MS, debido al bajo número de proteínas anotadas de estos organismos en las bases de datos. Pero también puede ser más compleja la etapa inicial ya que suele ser más difícil encontrar protocolos bien establecidos de preparación de las muestras para estos organismos.

Para comprobar este último punto de forma sencilla basta con hacer una búsqueda rápida de protocolos de preparación de muestra para proteómica en PubMed (http://www.ncbi.nlm.nih.gov/pubmed). Si centramos la búsqueda en proteómica de muestras humanas encontramos

277 entradas, que se reducen a 41 para proteómica de plantas, a 17 para proteómica de alimentos y a 5 en el caso de patógenos.

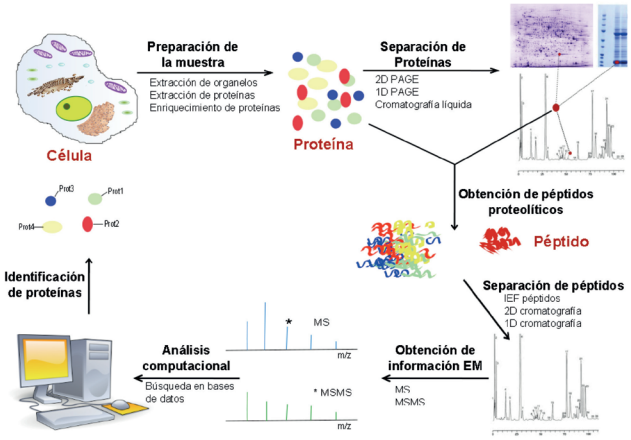


Figura 2. Flujo de trabajo de un experimento de proteómica "Bottom-up". La adquisición de información de MS-MS/MS se realiza a nivel de péptido.

2.1. Preparación de la muestra.

La importancia de los protocolos de extracción de proteína en un experimento de proteómica es evidente, ya que solo si se puede extraer y solubilizar una proteína, existe la posibilidad de identificarla.

Un ejemplo ilustrativo (Figura 3) de este hecho es la comparación, descrita recientemente en la literatura, de los perfiles proteicos 2DPAGE (electroforesis en gel de poliacrilamida de dos dimensiones) de proteínas de la lombriz *Eisenia fetida* obtenidas por distintos métodos [55]. Puede observarse que cada uno de los métodos de obtención de la muestra proporciona un subproteoma diferente. Obviamente el análisis de cada una de estas preparaciones dará resultados diferentes a nivel de proteína.

Si bien la obtención de proteínas es siempre importante, es fundamental en el caso de muestras de plantas y de hongos [38, 39]. Para plantas, un ejemplo de la importancia de la extracción de la muestra se ha descrito recientemente en la aplicación de la proteómica al estudio de la caña de azúcar. En este caso la obtención de proteína es extremadamente dificultosa debido a la naturaleza del tallo, muy rígido y fibroso y por la presencia en el mismo de enzimas de oxidación, compuestos fenólicos y niveles extremos de hidratos de carbono [38].

En el caso de los hongos filamentosos lo que dificulta la obtención de proteína es la presencia de una pared celular que, además de extraordinariamente robusta, constituye la mayoría de la masa celular [39].

Puesto que la mayoría de estudios proteómicos de plantas y hongos se hacen con organismos huérfanos, el problema de preparar la muestra agrava el problema inherente a la identificación de proteínas de estos organismos.

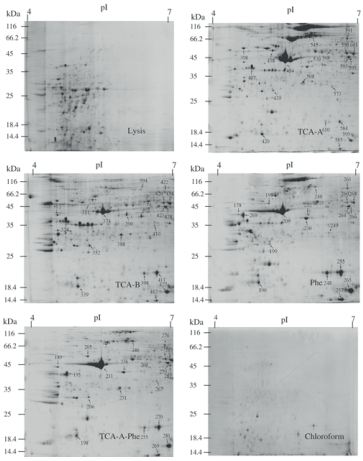


Figura 3. Análisis 2DPAGE de proteínas de extracto total de la lombriz E. fetida, extraído usando distintos protocolos de lisis; TCA-A, TCA-B, Phe, TCA-A_Phe y cloroformo. Las proteínas se gradiente de pH lineal de 4 a 7 en la primera dimensión y se visualizan usando tinción Coomassie coloidal ([55] Wang X, et al. Protein extraction from the earthworm Eisenia fetida for 2-DE. Proteomics 2010; 10:1095-9. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.)

A pesar de que no es objetivo de este capítulo una descripción detallada de metodologías de preparación de muestra, hemos querido citar estos dos ejemplos de la dificultad para obtener proteomas de organismos huérfanos representativos. Aunque, resulta evidente que problemas similares se pueden presentar trabajando con otros tipos de muestra.

Una vez que se ha conseguido una buena preparación de la muestra, el protocolo de trabajo es independiente del origen de la misma y es solo en la etapa de asignación de la información de MS y MS/MS obtenida, donde surgen las verdaderas dificultades de trabajar con organismos mal representados en bases de datos.

2.2. Análisis de la información de espectrometría de masas.

En proteómica 'Bottom-up' las dos principales fuentes de información que se registran en el análisis de MS para cada péptido son la masa del ión precursor y el espectro de fragmentación del mismo. La combinación de estos datos se utiliza para determinar la secuencia del péptido por comparación con los péptidos teóricos de todas las proteínas existentes en una base de datos y así identificar la proteína de la que el péptido se deriva. Habitualmente la masa del ión precursor es utilizada por el motor de búsqueda para generar una lista de péptidos candidatos, entre los de la base de datos, cuyos espectros de fragmentación teóricos se comparan con el espectro obtenido. El péptido con el mejor ajuste de los espectros de fragmentación determina la proteína de procedencia [56, 57].

Aunque este es un método robusto y útil de identificación de proteínas su eficacia no es total. Así, cuando los resultados de una búsqueda en bases de datos con un conjunto de datos de LC-MS/MS (LC: cromatografía líquida) son revisados, se puede comprobar que numerosos espectros de fragmentación han quedado sin asignar.

Suponiendo que un determinado espectro de fragmentación contiene información adecuada, es decir, presenta un número razonable de iones fragmento con buena relación señal/ruido, hay varias causas comunes por las que no es posible asignarlo a una secuencia peptídica.

Detallamos algunas de estas causas a continuación:

(A) baja especificidad del enzima utilizado.
(B) incorrecta determinación de la carga del precursor.
(C) subestimación del error de medición de masa.
(D) modificaciones post-traduccionales y químicas insospechadas.
(E) ausencia de la secuencia del péptido en la base de datos explorada.

Algunos de estos problemas (A, B, C) pueden evaluarse con el análisis de una muestra control y son fáciles de solucionar por ensayo y error. Sin embargo, los dos últimos plantean un desafío particularmente difícil. Para el caso de que existan modificaciones inesperadas (químicas, post-traduccionales, o sustitución puntual de aminoácidos) respecto a la secuencia nativa del péptido, en la actualidad, la mayoría de algoritmos de búsqueda permiten explorar esta posibilidad [58-60] minimizando la penalización por lo que se denomina "explosión del espacio de búsqueda", es decir, el aumento del número de candidatos a considerar. Pero, en el caso de que la secuencia peptídica no se encuentre en las bases de datos es necesario utilizar otras aproximaciones para la asignación de su espectro.

El abordaje clásico en esta situación es la secuenciación *de novo*. En él la secuencia de aminoácidos de un péptido se deduce directamente a partir de su espectro de fragmentación. Y, a partir de la secuencia del péptido se intenta identificar la proteína de la que este procede. Actualmente, además de la secuenciación *de novo* pueden plantearse en paralelo estrategias

alternativas como combinar la información de MS y MS/MS para proteínas individuales, o aumentar la información disponible en la base de datos utilizada para la identificación.

En todos los abordajes descritos anteriormente se hace necesario disponer de espectros de fragmentación de buena calidad (suficientemente informativos) para obtener una identificación. Pero aun en el caso de no disponer de espectros informativos existe la posibilidad de obtener una identificación positiva. Para ello es necesario mejorar los espectros de fragmentación utilizando herramientas químicas. Posteriormente se intenta otra vez la identificación, a partir de los nuevos espectros, por cualquiera de las aproximaciones disponibles.

A continuación detallaremos las alternativas descritas para la identificación de proteínas a partir de datos de MS y MS/MS de algunos de sus péptidos proteolíticos en el caso en que dicha proteína no se encuentre en las bases de datos. (Figura 4)

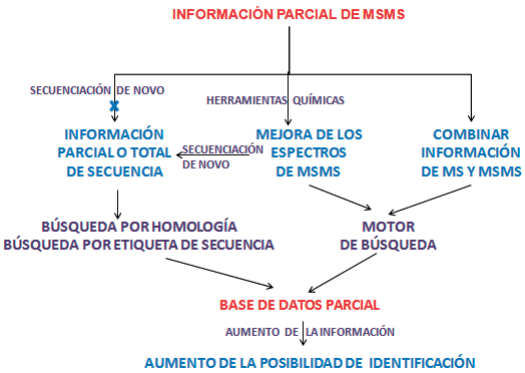


Figura 4. Esquema de las posibles herramientas para solucionar la dificultad de identificar proteínas de organismos huérfanos. En rojo se señalan las posibles causas de una identificación negativa. En azul se indican posibles soluciones y en morado aproximaciones para conseguir estas soluciones.

3. Posibles soluciones a la dificultad de los análisis proteómicos de organismos huérfanos.

3.1. Secuenciación de *novo*.

Por secuenciación *de novo* se entiende en proteómica la deducción de la secuencia, parcial o total, de aminoácidos de un péptido a partir de su espectro de fragmentación. Aunque la interpretación de un espectro MS/MS de calidad puede realizarse manualmente, es un proceso lento y tedioso. Por lo que, en general, se utilizan aproximaciones computacionales para derivar la secuencia del péptido fragmentado [61-63]. Pero, incluso asistida por análisis computacional, la secuenciación *de novo* no suele ser una tarea fácil ni rápida.

Existen dos problemas principales a la hora de secuenciar *de novo*. En primer lugar cada fragmentación puede producir dos tipos de iones, lo que significa que los dos extremos del péptido han de considerarse simultáneamente. En segundo lugar se desconoce a que tipo pertenece un ión dado, por lo que un mismo pico puede ajustarse a un tipo, a dos tipos o a ningún tipo de ión. Así, en general, el método directo de hacer “crecer” la secuencia de un extremo a otro no funciona.

En este contexto, se han desarrollado diferentes métodos para la secuenciación automática *de novo* de péptidos utilizando sus espectros de MS/MS. Algunos algoritmos de secuenciación tratan de predecir la secuencia total de aminoácidos del péptido desde su espectro de fragmentación nativo y su masa [63-65]. Estos algoritmos necesitan que toda la secuencia

se pueda deducir del patrón de fragmentación. Generan una secuencia que corresponde a la masa completa del péptido y no consideran que las regiones estables de este pueden no haberse fragmentado, de manera que su información de secuencia no está contenida en el espectro. Dado que no existe información adicional, químicamente codificada, sobre que tipos son los iones detectados, es difícil reconocer si el espectro analizado pertenece a un péptido que se fragmenta según las reglas de fragmentación utilizadas para construir el algoritmo de interpretación, o si pertenece a un péptido que produce un patrón de fragmentación anómalo. Este hecho puede dar lugar a interpretaciones de secuencia equivocadas [66].

Algunos programas intentan subsanar esta deficiencia. PepNovo [67] por ejemplo, utiliza una red probabilística para modelar los fenómenos de fragmentación de péptidos en un espectrómetro de masas. Además, utiliza una hipótesis de cociente de semejanza para determinar si es más probable que los picos observados en el espectro de masas hayan sido producidos bajo este modelo de fragmentación, que según un modelo probabilístico que trata la aparición de picos como un evento aleatorio.

De forma alternativa varios programas informáticos de secuenciación *de novo* usan la teoría de gráficos [61, 68-70]. Estos, en primer lugar, convierten el espectro en un gráfico acíclico dirigido o "gráfico de espectro" donde un nodo en el gráfico corresponde a un pico de masa del espectro y dos nodos están conectados por una arista si la diferencia de masa entre ellos corresponde a la masa de un residuo de aminoácido (AA). En este caso, la arista se nombra como el correspondiente AA. Posteriormente el programa intenta encontrar un camino que conecte los extremos N (amino) y C (carboxilo) terminales y que conecte todos los vértices correspondientes a una serie de iones [71]. Esta aproximación puede, en algunos casos, proporcionar simultáneamente caminos no deseados con múltiples vértices de la gráfica asociados a los mismos iones. El problema se reduce en este caso a encontrar el camino antisimétrico más largo (o de mejor puntuación) en un gráfico dirigido [72]. Un camino es antisimétrico si incluye al menos un par de nodos complementarios, es decir cuya suma es la masa del péptido fragmentado (ver figura 5). Programas de secuenciación *de novo* basados en esta aproximación son Lutefisk [69] y Sherenga [63] entre otros.

Otra forma diferente de secuenciar es la que utiliza el programa PEAKS de Bioinformatics Solutions [61, 73] donde se computa, a través de un sofisticado algoritmo dinámico, la mejor secuencia de entre las combinaciones de aminoácidos con peso molecular igual al del péptido fragmentado. Se consideran para la masa del precursor los 10000 mejores candidatos con todas las posibles combinaciones de aminoácidos. Para cada cálculo se estiman iones *a*, *b*, *c*, *x*, *y* así como *b*/*y*-17/18. La asunción básica del modelo es que cuantos más picos abundantes, del espectro de fragmentación, se asignen a esos iones mayor probabilidad tiene una secuencia de ser la correcta. Para cada masa *m* se calcula primero la recompensa/penalización de que haya un ión *y* (o *b*) con masa *m*. Si existe un pico cercano a *m,* la recompensa es igual a la abundancia logarítmica del pico multiplicada por un factor que refleja el error de la masa entre *m* y el pico y por otro que refleja la coexistencia de iones *x*, *y*-H2O, *y*-NH3 (o *a*, *c*, *b*-H2O, *b*-NH3). Si no existe un pico cercano a *m* la penalización es un valor constante y negativo. El problema se reduce finalmente a encontrar una secuencia tal que sus iones *b*/*y* maximicen esta recompensa a sus valores de masa.

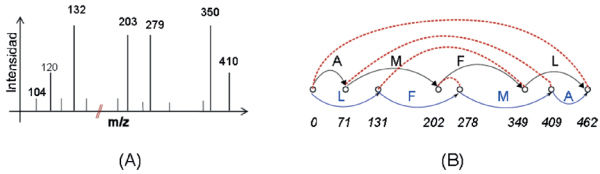


Figura 5. Esquema de secuenciación por teoría de gráficos.
(A) Espectro de fragmentación teórico del péptido AMFL
(B) El correspondiente gráfico de espectro extendido, las líneas punteadas sin dirección conectan vértices complementarios.

Esta aproximación goza de gran aceptación actualmente y proporciona, en general, buenos resultados [74].

A pesar de la existencia, como hemos comentado, de numerosas herramientas de secuenciación *de novo* el desarrollo de algoritmos sigue en vigor y adaptándose a las nuevas necesidades de las técnicas. Por ejemplo, recientemente, se ha descrito Vonode [62] un programa especialmente diseñado para información procedente de espectrómetros de masas de alta resolución.

Si se desea disponer de una información más completa de la secuenciación *de novo* existen revisiones recientes en la literatura como la de Seidler [75].

Sea cual sea el algoritmo utilizado para la secuenciación, siempre es aconsejable revisar manualmente las asignaciones que este ha generado. Algo que puede ser muy laborioso. Aun así, la utilidad de estas herramientas es evidente especialmente en la extracción de los espectros informativos de entre todos los disponibles.

Independientemente del método de secuenciación *de novo* utilizado, al final del proceso solo se dispone de información, total o parcial, de uno o varios péptidos de una proteína, pero se desconoce la proteína de la que se han derivado. De manera que, una vez se ha acabado el proceso de secuenciación y en función de la información de secuencia deducida, será necesario utilizar una aproximación u otra hasta llegar a identificar la proteína de interés como se muestra en la figura 6.

Si la secuencia obtenida *de novo* es completa, es necesario encontrar en las bases de datos alguna proteína con homología de secuencia con la secuencia deducida. Para lo que es necesario utilizar herramientas bioinformáticas específicas (Figura 6, A). Si no se encuentra ninguna proteína homóloga en las bases de datos puede utilizarse la secuencia de aminoácidos obtenida para diseñar cebadores degenerados y clonar el gen correspondiente (Figura 6, D) pudiendo de esta forma identificar la proteína de interés.

Si en la etapa de secuenciación *de novo* la información de secuencia obtenida es parcial, sigue siendo posible identificar una proteína utilizando métodos de etiqueta de secuencia (Figura 6, B). Y finalmente, si los espectros de fragmentación obtenidos no nos han permitido obtener información de secuencia podríamos, tal y como se ha comentado, utilizar herramientas químicas para obtener espectros de calidad a partir de los cuales si sea posible deducir una secuencia de AA (Figura 6, C).

Así en la secuenciación *de novo* la interpretación de un espectro no es la etapa final, sino el inicio de otras hasta llegar a la identificación de una proteína. Cuanto mayor sea la información de secuencia deducida menor será, en general, el esfuerzo necesario para identificar la proteína de interés.

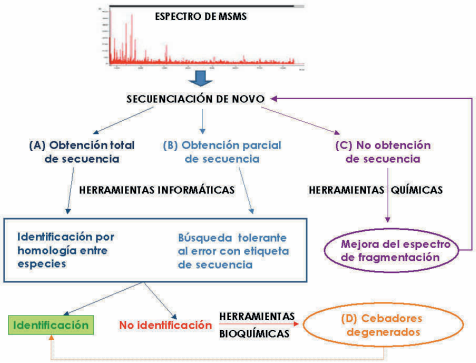


Figura 6. Flujo de trabajo desde la secuencia de novo hasta la identificación de la proteína de interés. Dependiendo de la información de secuencia deducida será necesario aplicar unas aproximaciones u otras para identificar una proteína.

A continuación comentaremos más ampliamente las distintas aproximaciones descritas.

A. Identificación de proteínas por homología entre especies.

Tal y como se ha comentado, una vez que se ha obtenido la secuencia completa de un péptido es necesario conocer a que proteína pertenece o podría estar relacionado. Para ello se han desarrollado multitud de herramientas informáticas algunas de las cuales se listan en la tabla 2 [76-84]. De entre ellas, las herramientas más utilizadas son probablemente BLAST (Basic Local Alignment Search Tool) [76], o su secuela Gapped BLAST [80], junto con FASTA (Fast protein and nucleic acid sequence comparison) [85].

Tabla 2. Herramientas de búsqueda por homología en bases de datos.

Herramientas de búsqueda por homología en bases de datos			
Descripción	**Tipo de secuencia***	**Autores**	**Año**
BLAST			
Búsqueda local con una rápida k-tupla heurística	Ambos	Altschul SF, et al.	1990
FASTA			
Búsqueda local con una rápida k-tupla heurística, más lenta y sensible que BLAST	Ambos	Lipman DJ, et al.	1985
SSEARCH			
Búsqueda Smith-Waterman, lenta pero más sensible que FASTA	Ambos	Pearson WR	1991
GGSERACH/GLSEARCH			
GGSEARCH implementa un alineamiento tipo Needleman-Wunsch pero usa penalizaciones por huecos refinadas. GLSEARCH es más apropiado para búsquedas globales.	Proteína	Pearson WR	1995
PSI-BLAST			
Búsqueda local con especificidad de posición	Proteína	Altschul SF, et al.	1997
HMMER			
Búsqueda local y global con modelo de perfil Hidden Markov	Ambos	Sonnhammer, EL et al.	1998
SAM			
Búsqueda local y global con modelo de perfil Hidden Markov	Ambos	Karplus K, et al.	1999
HHpred/HHsearch			
Comparación por pares con modelo de perfil Hidden Markov, muy sensible, pero solo búsquedas en bases de datos de alineamiento.	Proteína	Söding J	2005
IDF			
Frecuencia de documento inversa	Ambos		
CS-BLAST			
BLAST específico de contexo de secuencia, más sensible que BLAST, FASTA, y SSEARCH. Version iterativa de CSI-BLAST específica de posición, más sensible que PSI-BLAST	Proteína	Biegert A, et al.	2009

***Tipo de secuencia:** Proteína o nucleótido

BLAST es un programa informático de alineamiento de secuencias de tipo local, tanto de secuencias de ADN (ácido desoxirribonucleico) como de proteínas. El programa es capaz de comparar una secuencia problema con una gran cantidad de secuencias que se encuentran en una base de datos y encontrar las secuencias con mayor parecido a la secuencia problema. Es importante mencionar que BLAST usa un algoritmo heurístico por lo que no puede garantizar que ha encontrado la solución correcta. Sin embargo, BLAST es capaz de calcular la significación de sus resultados, y proporcionar un parámetro para juzgar la calidad de los mismos.

Gapped BLAST o BLAST 2.0 [80] es una mejora del algoritmo original de BLAST, que contempla la existencia de pequeñas inserciones o eliminaciones en las secuencias que se están comparando, permitiendo así alinear uno o varios nucleótidos o aminoácidos con huecos vacíos llamados “gaps”. Además agrega dos parámetros al algoritmo, uno es la penalización que se da en la puntuación por alinear un nucleótido o aminoácido con un hueco y el otro es una penalización por extender un hueco preexistente. Siempre se considera más “costoso” abrir un nuevo hueco que expandir uno existente. Actualmente es la versión de BLAST en uso.

FASTA fue descrito por primera vez como FASTP (Fast protein sequence comparison) [77] y fue diseñado originalmente para buscar similitudes de secuencias proteicas. FASTA, descrito en 1988, agregó la posibilidad de hacer búsquedas de ADN contra ADN, ADN contra proteínas y un programa más sofisticado de aleatorización para evaluar la significación estadística de los resultados. El paquete actual de FASTA incluye programas para búsquedas del tipo proteína/proteína, ADN/ADN, proteína/ADN traducido (con cambios en la pauta de lectura) y búsqueda ordenada y desordenada de péptidos. Las versiones recientes utilizan un algoritmo para tener en cuenta errores de desplazamiento de la pauta de lectura, pues las búsquedas que traducen

las seis pautas suelen tener problemas cuando comparan datos de secuencia de proteínas con los nucleótidos. Además de los métodos de búsqueda heurísticos rápidos, el paquete FASTA provee SSEARCH, una implementación del algoritmo Smith-Waterman. Uno de los principales objetivos del paquete es el cálculo estadístico de las similitudes para que sea sencillo juzgar si un alineamiento ha ocurrido por azar o si se puede usar para inferir homología.

Respecto al resto de herramientas cabe destacar que los métodos más recientes de alineamiento suelen ser más sensibles que los antiguos, pero son necesarias estimaciones estadísticas más precisas para poder explotar al máximo sus prestaciones [83].

En general, a la hora de valorar los resultados de alineamiento de secuencia hemos de tener en cuenta varios factores. En primer lugar, la significación estadística de los alineamientos es dependiente de la longitud de la secuencia y los segmentos de secuencia obtenidos, por espectrometría de masas de péptidos trípticos, son generalmente cortos, habitualmente de 10 aminoácidos o menos. Como consecuencia, los algoritmos de búsqueda a menudo no logran encontrar homólogos de forma estadísticamente significativa. En segundo lugar, la localización y cobertura de los péptidos respecto a la proteína homóloga es muy variable y pueden representar a regiones de baja o nula similitud al homólogo en la base de datos. Por lo tanto, una baja significación estadística en el alineamiento puede significar la ausencia de homólogos en la base de datos o que los péptidos secuenciados representan regiones mal conservadas, incluso de homólogos bastante cercanos. La incapacidad de distinguir entre estas dos posibilidades podría dar lugar a una interpretación errónea respecto a la novedad de la proteína de interés.

Con el objetivo de minimizar este problema se han desarrollado numerosas herramientas para una identificación más efectiva de proteínas homólogas a otras descritas en bases de datos. En concreto, muchos grupos han utilizado la coherencia entre los resultados, de búsqueda en bases de datos, para todos los péptidos disponibles de una proteína única para desarrollar mejores aproximaciones de identificación de homólogos lejanos. A continuación se describen algunas herramientas que han destacado sobre otras y se han utilizado con éxito en numerosos estudios de proteómica (ver ejemplos en la tabla 3; [71, 86-102]

FASTS. FASTS [92] es una modificación de FASTA que permite enviar múltiples secuencias cortas de una misma proteína en una búsqueda única y usa un sistema de puntuación basado en la probabilidad de alineamiento en vez de en la similitud. FASTS requiere solo 15-20 residuos en tres o cuatro péptidos para identificar robustamente homólogos con 50% o más de identidad de secuencia.

MS-Shotgun. MS_Shotgun [88] es una modificación de Shotgun [103] que utiliza secuencias peptídicas individuales en las búsquedas y, posteriormente, determina la congruencia entre los resultados obtenidos determinando que proteínas han sido identificadas por más de una secuencia peptídica.

CIDentify. Se trata de una versión modificada de FASTA que usa múltiples secuencias candidatas generadas por Lutefisk [69, 71] y puede considerar las peculiaridades y ambigüedades de las secuencias obtenidas por secuenciación *de novo,* a partir de datos de MS/MS, como por ejemplo equivalencias de masa (*m*) de algunos residuos con otros o combinaciones de otros (ejemplo: *m*Trp=*m*Gly+*m*Glu).

MultiBLAST Suite. Es una herramienta que facilita la generación y presentación de múltiples permutaciones y concatenaciones de secuencias peptídicas a Gapped BLAST. Para ello genera un único fichero de texto que contiene todos los fragmentos de secuencia deducidos a partir de datos de una proteína única [88].

MS BLAST. MS BLAST [104] es una herramienta de búsqueda por semejanza de secuencia que utiliza candidatos de secuencia redundantes, degenerados y parcialmente inexactos obtenidos de la interpretación de espectros de fragmentación.

MS-Homology. MS-Homology [105] está integrado en Protein Prospector y puede tomar datos de MS/MS como combinación de etiquetas de aminoácidos y masas para encontrar péptidos homólogos o idénticos en la base de datos. La búsqueda se hace, en paralelo, con datos de varios espectros siendo necesario un número mínimo de péptidos que encajen en una proteína para que esta se identifique significativamente.

Tabla 3. Herramientas de búsqueda por homología múltiple en bases de datos.

Ejemplos de utilización de herramientas de búsqueda por homología en bases de datos			
Título	**Revista**	**Autores**	**Año**
MS-Shotgun			
Protein identification by spectral networks analysis	Methods Mol Biol	Bandeira N	2011
Beyond Edman Degradation: Automated De novo Protein Sequencing of Monoclonal Antibodies	Nat Biotechnol	Bandeira N, et al.	2008
Functional assignment of the 20 S proteasome from Trypanosoma brucei using mass spectrometry and new bioinformatics approaches	J Biol Chem	Huang L, et al.	2001
FASTS			
Mixed peptide sequencing and the FASTF/FASTS algorithms	Methods Enzymol	Ray R, et al.	2003
Strategies for proteomics with incompletely characterized genomes: the proteome of Bos taurus serum.	Electrophoresis	Wait R, et al.	2002
MALDI-TOF/TOF de novo sequence analysis of 2-D PAGE-separated proteins from Halorhodospira halophila, a bacterium with unsequenced genome	Electrophoresis	Samyn B, et al.	2006
Getting more from less: algorithms for rapid protein identification with multiple short peptide sequences	Mol Cell Proteomics	Mackey AJ, et al..	2002
CIDentify			
Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry	Anal Chem	Taylor JA, et al.	2001
A hybrid method for peptide identification using integer linear optimization, local database search, and quadrupole time-of-flight or OrbiTrap tandem mass spectrometry	J Proteome Res	DiMaggio PA Jr et al.	2008
MultiBLAST Suite			
Functional assignment of the 20 S proteasome from Trypanosoma brucei using mass spectrometry and new bioinformatics approaches	J Biol Chem	Huang L, et al.	2001
MS BLAST			
Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated de novo sequencing	Proteomics	Waridel P, et al.	2007
Homology-based functional proteomics by mass spectrometry: application to the Xenopus microtubule-associated proteome	Proteomics	Liska AJ, et al.	2004
MALDI-TOF/TOF de novo sequence analysis of 2-D PAGE-separated proteins from Halorhodospira halophila, a bacterium with unsequenced genome	Electrophoresis	Samyn B, et al.	2006
A workflow to increase the detection rate of proteins from unsequenced organisms in high-throughput proteomics experiments	Proteomics	Grossmann J, et al.	2007
The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches	Mol Cell Proteomics	Habermann B, et al.	2004
Identifying secreted proteins of Marssonina brunnea by degenerate PCR	Proteomics	Cheng Q, et al.	2010
MS-Homology			
Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: I. How much of the data is theoretically interpretable by search	Mol Cell Proteomics	Chalkley RJ, et al.	2005
MALDI-TOF/TOF de novo sequence analysis of 2-D PAGE-separated proteins from Halorhodospira halophila, a bacterium with unsequenced genome	Electrophoresis	Samyn B, et al.	2006
Defining parameters for homology-tolerant database searching	J Biomol Tech	Kayser JP, et al.	2004
OpenSea			
High-throughput identification of proteins and unanticipated sequence modifications using a mass-based alignment algorithm for MS/MS de novo sequencing results	Anal Chem	Searle BC, et al.	2004
Identification of protein modifications using MS/MS de novo sequencing and the OpenSea alignment algorithm.	J Proteome Res	Searle BC, et al.	2005

OpenSea. En esta aproximación las secuencias *de novo* y las de las bases de datos se interpretan como masas de residuos, y son las masas y no los códigos de aminoácidos los que se comparan. Para proporcionar mayor flexibilidad las masas de alinean en grupos lo que permite resolver muchos errores de secuenciación *de novo* como la presencia de residuos, o combinaciones de residuos, isobáricos [101].

Como consecuencia del esfuerzo realizado en el desarrollo de estas herramientas, actualmente, el flujo de trabajo de asignación de secuencia e identificación de proteínas por homología se encuentra en muchos casos automatizado, y se denomina identificación de proteínas por homología entre especies. Por ejemplo, la figura 7 ilustra como SPIDER [106] integrado en PEAKS (Bioinformatics solutions) reconstruye una secuencia correcta a partir de una secuencia *de novo* y un péptido homólogo de forma totalmente automatizada.

Con el tiempo se han consolidado algunas de las herramientas desarrolladas sobre el resto. Por ejemplo destaca MS BLAST, desarrollada por Shevchenko et al. [104], con numerosos ejemplos de su utilidad descritos en la literatura (tabla 3). Pero, la decisión de utilizar unas u otras, de entre todas las herramientas disponibles, viene impuesta frecuentemente por los recursos bioinformáticos disponibles en un laboratorio. De forma que la mayoría de investigadores suele utilizar aquellas herramientas integradas (o fácilmente integrables) en los paquetes informáticos de sus espectrómetros de masas.

Figura 7. Se ilustra como la herramienta Spider [106] integrada en Peaks Explorer (Bioinformatics Solutions) deduce la secuencia correcta desde una secuencia de novo y un péptido homólogo de las bases de datos.
http://www.bioinformaticssolutions.com/products/peaks/spider.php

B. Identificación de proteínas por búsqueda tolerante al error con etiqueta de secuencia.
A pesar de los muchos esfuerzos realizados en la última década, los avances en la secuenciación *de novo* de péptidos han sido lentos y solo el 30-45% de todos los péptidos son correctamente reconstruidos [67]. La secuencia exacta y total de AA de un péptido puede ser una meta inalcanzable a partir de algunos espectros, pero existen aproximaciones que aprovechan la información parcial deducida de los espectros para identificar proteínas con gran fiabilidad. Esto es así porque, aun en el caso de que solo sea posible derivar parte de la secuencia de un péptido desde su espectro de fragmentación, suele ser posible encontrar la proteína de la que se deriva tal y como demostraron Mann et al. [107]. La aproximación demostrada por ellos se conoce como búsqueda a partir de "etiqueta de secuencia" o "sequence tag" en inglés.

En ella la secuencia parcial obtenida divide al péptido en tres regiones (ver figura 8). Una región central secuenciada y una región a cada extremo caracterizada por un valor de masa adicional (m1 y m3). La construcción m1-secuencia-m3 se denomina "etiqueta de secuencia de péptido" y es un identificador muy específico del péptido. La inclusión de la información de las masas de los extremos hace que la información total sea hasta un millón de veces más discriminante que solo la información parcial de secuencia.

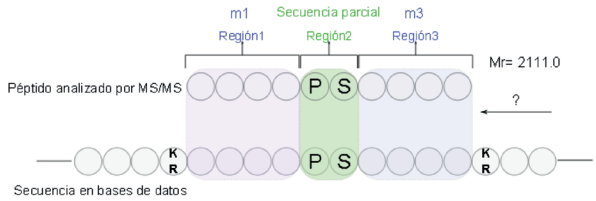


Figura 8. Principio de emparejamiento de la etiqueta de secuencia de un péptido a una secuencia propuesta. La cadena de aminoácidos superior representa la secuencia obtenida por MS/MS, y la cadena inferior representa los aminoácidos en la secuencia de la base de datos con la que se compara la etiqueta. La secuencia parcial divide al péptido en tres regiones. La masa adicional m1 de los residuos en la región 1 junto con el N-terminal es un criterio para inferir la secuencia de origen, igual que lo es la masa adicional en la región 3, m3. Además se requiere que el péptido cumpla las reglas de corte del enzima proteolítico, marcado como KR para la tripsina. La flecha con interrogante indica que las dos direcciones de búsqueda han de considerarse (Adaptada Mann M, et al. 1994 [107]).

La aproximación original ha sido optimizada en herramientas como MS_Profile [108] de Pevzner PA, et al. Y ha sido, con o sin modificaciones, implementada en numerosos paquetes de identificación de proteínas como Peaks (Bioinformatics Solutions Inc.), Mascot (MatrixScience), Protein Prospector (UCSF, University of California. San Francisco), etc.

Finalmente, los recientes avances en esta aproximación han permitido a las herramientas de identificación basadas en etiquetas de secuencia, como por ejemplo InsPecT [109] y Paragon [110], ser órdenes de magnitud más rápidas que las herramientas tradicionales de búsqueda en bases de datos [111] como Sequest [112] y Mascot [113].

En el contexto de identificación de proteínas de organismos huérfanos esta estrategia también ha demostrado su utilidad, pues ha sido específicamente optimizada para la identificación de proteínas cuya secuencia no es conocida, pero posee homología con la de otras proteínas de las bases de datos. Se define esta aproximación como búsqueda "tolerante al error" a partir de "etiqueta de secuencia". En este caso a cada parte de la etiqueta de secuencia se le tolera un error. Como consecuencia de ello la búsqueda puede producir un elevado número de identificaciones con baja significación estadística, pero la consideración simultánea de múltiples etiquetas de secuencia que provienen de diferentes espectros MS/MS puede proporcionar identificaciones estadísticamente significativas.

Ejemplos de su uso son los métodos MultiTag [114, 115] y GutenTag [116] que permiten asignar significación estadística al alineamiento de secuencia de múltiples etiquetas de secuencia con "tolerancia al error" con una proteína de la base de datos.

Conviene destacar que motores de búsqueda tan populares como MASCOT también permiten esta forma de búsqueda lo que hace que pueda utilizarse en la mayoría de laboratorios de proteómica.

Con las herramientas descritas es posible identificar proteínas de organismos huérfanos a partir de espectros de MS/MS poco informativos siempre que pueda deducirse de ellos como mínimo una etiqueta de secuencia. Sin embargo, esto no es útil para el caso de espectros de MS/MS no canónicos o de los cuales no es posible deducir una etiqueta de secuencia de longitud mínima. En estos casos podremos utilizar herramientas químicas para mejorar los espectros y obtener así mayor información de secuencia.

C. Mejora de espectros de fragmentación.

Como ya se ha comentado, durante la fragmentación de un péptido en un espectrómetro de masas es posible que algunas regiones del mismo, más estables, no se fragmenten, de manera que su información de secuencia no está contenida en el espectro. Además, en este y otros casos es posible que un péptido se fragmente según un patrón de fragmentación anómalo, lo que dificultaría la interpretación del espectro obtenido [66].

Para cambiar el comportamiento de fragmentación de los péptidos Pappin et al. propusieron introducir un grupo básico en el extremo N-terminal de los mismos [117, 118]. De esa forma los iones *b* están más favorecidos y la serie *y* suele ser más larga que en el péptido no modificado. La aparición simultánea de las dos series favorece la interpretación del espectro. Alternativamente, se han descrito aproximaciones de modificación química de los péptidos para identificar algunos tipos de ión fragmento por el desplazamiento de masa asociado a la modificación [75]. En algunos casos esto se ha abordado por esterificación de todos los grupos ácido, incluido el C-terminal, del péptido y en otros introduciendo un isótopo de ^{18}O en el extremo del péptido tras la digestión tríptica [66, 119, 120].

Todas estas estrategias suelen requerir varias etapas de manipulación de la muestra. En cada una de estas etapas es posible que el rendimiento no sea cuantitativo y que se generen subproductos por reacciones químicas no deseadas. Por ello estos métodos, comparados con los convencionales, implican el uso de una mayor cantidad de proteína, no siempre disponible, y pueden dar lugar a mezclas más complejas. Debido a esta mayor dificultad en la preparación y análisis de la muestra, de entre todas las herramientas químicas descritas en la literatura, solo vamos a comentar la derivatización con SPITC (4-sulfofenilisotiocianato) porque de ella hay numerosos ejemplos de utilización exitosa en la literatura y ha sido ampliamente utilizada por nosotros mismos [121].

El fenilisotiocianato (reactivo de Edman's o PITC por su nombre en inglés) se ha utilizado tradicionalmente en química de proteínas para la secuenciación de Edman. PITC reacciona con los grupos amino de los péptidos o proteínas para formar sus derivados feniltiocarbamilados. En el caso de los grupos α-amino se da un reordenamiento posterior, con el grupo carbonilo del primer enlace peptídico, que ocasiona la escisión del aminoácido N-terminal. El mecanismo que tiene lugar es pues similar al de formación de iones tipo *b* durante la fragmentación de péptidos en un espectrómetro de masas [122].

Puesto que, como se ha descrito, la presencia de una carga negativa en el grupo introducido mejora los espectros de MS/MS [123], es un derivado de PITC, el reactivo SPITC, el que se ha venido utilizando en proteómica. Aunque su utilización se ha restringido, fundamentalmente, a espectrometría MALDITOF (Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight) ya que su efecto es solo significativo en péptidos monocargados.

Los péptidos monocargados requieren una energía mucho mayor para su fragmentación que los doble o triplemente cargados. De acuerdo con la teoría de la movilidad de protón [124] esto

es así porque, en péptidos dicargados, un protón está localizado en un residuo básico (Lys o Arg) mientras que el otro puede desplazarse a lo largo del esqueleto peptídico facilitando la fragmentación [125]. En el caso de los péptidos con una carga, en los que el único protón está anclado al residuo básico, es necesaria mayor energía para su fragmentación. Para ionizar péptidos con una carga negativa fija, como los derivados con SPITC, son necesarios dos protones, lo que los hace más parecidos a los péptidos con dos cargas en la fragmentación. La carga negativa en N-terminal proporciona una ventaja adicional, los fragmentos N-terminales que retendrían una carga (iones *b*) son neutros en los péptidos derivados, y por lo tanto no se detectan, simplificándose considerablemente el espectro de fragmentación al contener solo iones *y*, lo que lo hace más fácilmente interpretable (ver figura 9).

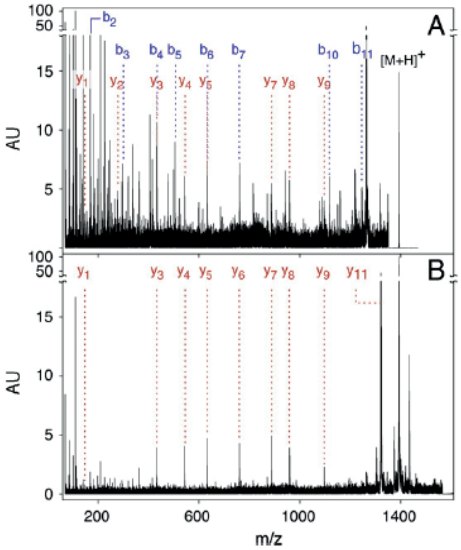


Figura 9. Secuenciación de novo usando SPITC. Espectro MALDI TOF/TOF de un péptido de Lys-C de Cyclin A2 antes (A) y después (B) de modificarse con SPITC [121]. La fragmentación del péptido modificado ocasiona la pérdida del grupo introducido y la formación del ión yn-1 (y11 en este caso). Habitualmente estos iones son los más intensos, pero además el espectro se simplifica y la secuencia puede derivarse fácilmente a partir de los iones y.

Por eso la metodología es más útil con ionización MALDI, aunque la optimización de la fragmentación de los péptidos en los modernos espectrómetros MALDI TOF/TOF hace que no sea tan necesario el uso de la derivatización en este caso, sí lo es para la espectrometría MALDI PSD [125].

En este contexto, la derivatización con SPITC suele utilizarse para el análisis de proteínas individuales, habitualmente separadas por 2D PAGE, y es posible encontrar numerosos ejemplos de la mejora que proporciona, especialmente en el análisis de proteínas de organismos huérfanos. Por ejemplo, se ha utilizado con éxito en el análisis de proteínas de fresa [126] de un experimento DIGE donde se consiguió identificar un 58% de las muestras a pesar de que no se disponía ni de secuencias genómicas. De forma parecida, la derivatización permitió a Oehlers et al. identificar la práctica totalidad de las proteínas de interés de un experimento DIGE de medaka (Oryzias latipes) [127]. Y ha sido también descrito un incremento de hasta el 78% en las identificaciones de proteínas de un nematodo, sin genoma conocido, separadas por geles 2D PAGE [128].

Además de asociado a espectrometría MALDI PDS o MALDI TOF/TOF la derivatización se ha utilizado en combinación con otras técnicas en experimentos de SRM por MALDI QqQLit [129] donde, previamente, se realiza una separación de los péptidos por cromatografía líquida. Este trabajo demuestra que el uso de la metodología no está restringido a proteínas individuales e incrementa notablemente su utilidad al poderse aplicar a mezclas.

D. Identificación de proteínas mediante el diseño de cebadores degenerados a partir de la información de secuencia obtenida.

Existen casos en los que no ha sido posible la identificación de la proteína desconocida a partir de la de secuenciación deducida, ni mediante búsquedas por “homología de secuencia”, ni por “etiqueta de secuencia tolerante al error”, debido a la baja homología de la secuencia deducida y las secuencias presentes en las bases de datos. En estos casos es posible la identificación de forma alternativa, utilizando las secuencias peptídicas obtenidas, de la interpretación de los espectros de fragmentación, para abordar el diseño de cebadores degenerados con el fin de clonar y secuenciar el gen de la proteína desconocida utilizando métodos basados en PCR

(Polymerase chain reaction). Para ello es necesaria la traducción previa de la secuencia del péptido a la secuencia genética de acuerdo con el código genético universal.

Obviamente esta estrategia no es fácil y su éxito depende de múltiples factores e inconvenientes que hay que vencer a la hora de abordarla y detallamos a continuación. En primer lugar, la secuencia peptídica interpretada del espectro de masas es a menudo parcialmente inexacta ya que, en algunos casos, no es posible distinguir entre aminoácidos (o combinaciones de AA) isobáricos o de masa similar. Por otro lado, las secuencias obtenidas son generalmente cortas, lo que limita la selección de los péptidos candidatos para el diseño de los cebadores. Y finalmente, la disposición de la secuencia peptídica interpretada respecto a la proteína es desconocida, lo que implica según el caso, realizar el diseño bidireccional de los cebadores para cada péptido y utilizar distintas combinaciones de cebadores en la amplificación del gen.

Estas dificultades encarecen el experimento y lo hacen mucho más laborioso, por lo que no son muchos los trabajos que utilizan este tipo de estrategia. Sin embargo, encontramos en la literatura algunos ejemplos en los que ha funcionado con éxito esta aplicación, empleando diferentes estrategias que han culminado con la secuenciación parcial del gen y la identificación de la proteína desconocida.

En un primer trabajo, como el de Salekdeh et al., se consiguió identificar una proteína de arroz, no presente en las bases de datos, que se encontraba fuertemente sobreexpresada en condiciones de sequía frente a condiciones de riego en un estudio de expresión diferencial realizado con geles bidimensionales. A partir de un extracto de ARN del tejido vegetal y mediante cebadores degenerados diseñados a partir de los datos de MS, se consiguió amplificar, clonar y secuenciar el ADNc del gen de la proteína desconocida. Al realizar un BLAST de la secuencia del ADNc contra las bases de datos de ADN se reveló la homología de la proteína con una familia de proteínas implicadas en la respuesta de algunas plantas a señales medioambientales [130].

En otro trabajo se identificó la proteína de un patógeno no secuenciado que infecta la planta del girasol causando una enfermedad de elevado coste económico [131]. Para dicha identificación se realizó una búsqueda por homología de secuencia encontrándose homología con una proteína de otro organismo estrechamente relacionado con el patógeno. Los péptidos secuenciados a partir de los datos de MS/MS fueron posteriormente verificados mediante el diseño de cebadores degenerados, que permitieron amplificar el gen mediante RACE-PCR (Rapid amplification of cDNA-end) e identificar una nueva proteína elicitora en el patógeno.

En el caso del trabajo realizado por Cheng [98] se realizó un estudió proteómico del secretoma de un hongo patógeno, no secuenciado, causante de una enfermedad en el chopo. Muchas de las proteínas de interés fueron identificadas por búsqueda por homología de secuencia, pero para aquellas proteínas sin identificación positiva, se recurrió a una novedosa estrategia. Se utilizó un método de PCR degenerada: CODEHOP (Consensus-degenerate hybrid oligonucleotide primers, [132]) combinado con RACE-PCR. El diseño de los cebadores CODEHOP con una doble región, “core region” (degenerada) y “clamp region” (consenso), se basó en la correcta interpretación de las secuencias de los péptidos identificados por MS/MS. Los fragmentos de ADNc amplificados y secuenciados permitieron identificar estas proteínas del secretoma específicas de la esta especie.

Los tres ejemplos descritos demuestran la utilidad de la técnica, a pesar de las dificultades mencionadas, cuando el interés de la proteína en estudio justifica el esfuerzo que requiere esta aproximación.

3.2. Optimización de las búsquedas en bases de datos.

Hasta este momento hemos descrito la identificación de proteínas no presentes en bases de

datos a partir de la información, total o parcial, deducida de su secuencia por comparación con otras proteínas existentes en bases de datos. Como hemos visto resulta evidente que, en todos los casos, el éxito en la identificación de proteínas dependerá de la información contenida en las bases de datos. Si la información disponible es muy pobre para el organismo en estudio, o se estudian proteínas muy divergentes, el resultado puede ser una identificación negativa a pesar del esfuerzo realizado. De hecho, en estudios de evaluación de metodologías de identificación por secuenciación *de novo* y búsqueda de homología entre especies se ha determinado que el porcentaje de falsos positivos, falsos negativos y verdaderos positivos está relacionado con el porcentaje de identidad entre la secuencia compuesta para la búsqueda y la secuencia encontrada en la base de datos [97]. Otros estudios realizados usando por un lado métodos de identificación en bases de datos a partir de información de MS-MS/MS y por otro lado, secuenciación de novo y búsqueda por homología han concluido que ambos tipos de búsqueda han de realizarse en paralelo y sus resultados han de compararse [100]. Cuando esto se hace los resultados no coinciden en un porcentaje significativo de las identificaciones (16% en el estudio citado). Pero cuando coinciden, la búsqueda directa en bases de datos proporciona mayor porcentaje de cobertura de secuencia de la proteína; es decir, información adicional de la proteína identificada y mejor confianza en la identificación. Además, es fácilmente comprensible, por el esfuerzo que supone, que la aproximación de secuenciación *de novo* y búsqueda por homología de secuencia no se utilice rutinariamente en experimentos de proteómica global.

Por todo lo expuesto anteriormente resulta evidente que la optimización de las búsquedas en bases de datos permitiría optimizar la identificación de las proteínas de interés de organismos huérfanos. Para mejorar los resultados de estas búsquedas podemos, por un lado, aumentar la información disponible en las bases de datos para el organismo en estudio. Y por otro lado, en los casos en que sea posible, podemos combinar la información de MS y MS/MS en una misma búsqueda.

Respecto a este último punto nos gustaría destacar, por su utilidad, la eficacia de los instrumentos MALDI-TOFTOF en el análisis de proteínas individuales de organismos huérfanos [133].

Estos instrumentos permiten, con elevada sensibilidad, adquirir toda la información de huella peptídica de una proteína además de espectros de fragmentación de algunos de sus péptidos (3 a 10 precursores habitualmente). Esta información, combinada, ha permitido en muchos casos obtener identificaciones de organismos mal representados en bases de datos [74, 134, 135]. Los datos de huella peptídica toleran mejor la homología de secuencia al contener información de la totalidad de la proteína y no solo de un péptido o unos pocos péptidos de la misma [136, 137]. De esta manera es más posible encontrar regiones conservadas de la proteína e identificarla. Además se dispone de información de MS/MS que puede asignarse a un péptido de dicha proteína, o ser utilizada para secuenciación *de novo* en caso de ser necesario.

Independientemente de que utilicemos información MS-MS/MS o solo MS/MS en las búsquedas, cuanto mayor sea la información que contienen las bases de datos mayor será el número de proteínas identificadas y mejor la significación estadística de la identificación, sea cual sea la aproximación utilizada. Por ello en los últimos años ha ido aumentando el número de ejemplos en los que la información de MS obtenida en un experimento se contrasta directamente con los espectros teóricos obtenidos *in silico* a partir de bases de datos genómicos, de secuencias de ESTs (Expresed sequence tags) o incluso de ADN [138]. Estas bases suelen contener mayor información que las bases de proteínas especialmente para organismos huérfanos.

Actualmente, como consecuencia de estos trabajos, se dispone de diferentes algoritmos de identificación, tipos de bases de datos y repositorios [56, 57] para optimizar la búsqueda de proteínas en bases de datos.

Si decidimos aumentar la información disponible para un organismo de interés utilizando bases de datos genómicas hemos de tener en cuenta que, comparado con las secuencias de genes, el uso de ESTs como fuente de secuencias peptídicas elimina las conjeturas que participan en la búsqueda de la secuencia de codificación y la determinación de su estructura intrón-exón [139] y deja solo la cuestión del marco de traducción sin resolver. Es por ello que, en general, se recomienda el uso de bases de datos de ESTs específicas de especie si la información a nivel de proteína ha resultado deficiente. El uso de ESTs es además aconsejable ya que la rapidez con la que se generan, por secuenciación de ADNc generado a partir de ARNm, hace que la información disponible aumente rápidamente.

A pesar de que utilizar ESTs en las búsquedas no es nuevo (Yates et al, 1995; Neubauer et al, 1998) una serie de aspectos han debido optimizarse hasta la implantación general de su uso. Por un lado, su gran tamaño inicial que penaliza el tiempo de búsqueda que suele aumentar linealmente con el tamaño de la base de datos utilizada [140]. Por otro lado, la propensión de estas bases a contener errores derivados de los métodos de secuenciación de ADN de un solo paso.

Con el tiempo se han eliminado errores y repeticiones lo que, finalmente, ha hecho que puedan usarse de forma rutinaria con la mayoría de algoritmos de búsqueda.

Normalmente las secuencias de nucleótidos son traducidos, por los motores de búsqueda, "sobre la marcha" en seis pautas de lectura y todas las secuencias de aminoácidos resultantes se tienen en cuenta para asignar los espectros de fragmentación obtenidos. De esta forma se usan en motores de búsqueda de uso general como MASCOT o Sequest [141]. Para aquellos motores de búsqueda que no permitan utilizar directamente las bases de datos de ESTs existen herramientas informáticas que traducen los ESTs a secuencias de aminoácidos y generan ficheros adecuados para la mayoría de algoritmos [142-144].

Respecto a las fuentes de secuencias de ESTs [145], podríamos destacar NCBI (National Center for Biotechnology Information, del National Institutes of Health, USA) que produce y mantiene la base de datos abierta de GenBank [146]. GenBank y sus colaboradores reciben secuencias producidas en laboratorios de todo el mundo de más de 100,000 organismos diferentes (ver algunos ejemplos en la tabla 4). Y la información crece a un ritmo exponencial, doblándose aproximadamente cada 18 meses. A través de NCBI las secuencias de los diferentes organismos son sencillas de obtener y tienen todas la misma estructura, lo que facilita la creación de bases en uso para un determinado motor de búsqueda: Ya que, las reglas que permitan definir una base de datos para un organismo valdrán para el resto. En el caso de MASCOT, por ejemplo, las "parse rules" fijadas en la definición de la base de datos serán las mismas para todas las bases.

Un inconveniente de estas bases, debido a la forma en que los ESTs se secuencian, es que distintos "expressed sequence tags" suelen ser secuencias parciales del mismo ARNm de un organismo. En un esfuerzo por reducir el número de etiquetas de secuencia generadas durante el análisis de descubrimiento de genes, diferentes grupos han combinado los ESTs en contigs. Un contig, de acuerdo con Staden, es un conjunto de segmentos de ADN de una fuente genética única que se solapan [147]. Por ello el uso de bases de datos de contigs puede ser mejor que el de ESTs al haberse disminuido la redundancia.

Ejemplos de fuentes de contigs son TIGR gene indices [148]; Unigene [149] y STACK [150].

En el caso de que en estos repositorios generales no se encuentre información ni de ESTs ni de contigs de algún organismo existen muchos otros accesibles [151-154], aunque pueden ser más o menos difícil poner en uso una base de datos a partir de la información que contienen. La dificultad radica, fundamentalmente, en la sintaxis utilizada para la identificación y descrip-

ción de cada contig (también es aplicable a las bases de ESTs) y en la presencia de repeticiones, importantes limitaciones para algunos motores de búsqueda como MASCOT.

Como hemos visto, en muchos casos, disponer de una base de datos genómica activa suele requerir más tiempo que para proteínas. Además, las búsquedas también requieren mayor tiempo de computación que con las bases de datos convencionales [155]. Probablemente sea esto la causa de que esta aproximación no sea habitualmente usada como primera opción en la identificación de proteínas de organismos huérfanos. A pesar de su evidente utilidad para identificar estas proteínas. Pero en aquellos casos en los que la información a nivel de proteína es despreciable el esfuerzo está justificado ampliamente por los resultados, tal y como puede comprobarse en los ejemplos detallados en la tabla 5 [32, 141, 144, 156-167].

Tabla 4. Número de ESTs disponibles en NCBI para algunos organismos. Comparación con el número de proteínas disponibles en el mismo repositorio.

Organismo	Número de ESTs (1)	% ESTs	Número de Proteínas (1)	% Proteínas
Homo sapiens	8,314,469	100	558988	100
Zea mays	2,019,105	24	112005	20
Sus scrofa	1,621,012	19	36575	7
Arabidopsis thaliana	1,529,700	18	183852	33
Danio rerio	1,481,936	18	70889	13
Xenopus	1,271,375	15	38881	7
Oryza sativa	1,250,954	15	535540	96
Ciona intestinalis	1,205,674	15	16035	3
Rattus norvegicus	1,162,136	14	109938	20
Gallus gallus	600,423	7	36959	7
Salmo salar	498,212	6	17133	3
Pinus taeda	328,662	4	2350	0
Aedes aegypti	301,596	4	35787	6
Solanum lycopersicum	297,104	4	5950	1
Citrus sinensis	213,830	3	671	0
Schistosoma mansoni	205,892	2	27689	5
Daphnia pulex	152,659	2	31425	6
Schistosoma japonicum	103,725	1	24289	4
Apis mellifera	78,274	1	15544	3
Sparus aurata	67,673	1	557	0
Aspergillus niger	46,938	1	26304	5
Oncorhynchus tshawytscha	14,173	0	6480	1
Gammarus pulex	12,345	0	12	0
Mytilus edulis	5,036	0	3088	1
Fasciola hepatica	1,677	0	833	0
Sonneratia caseolaris	249	0	262	0
Saccharum officinarum	105	0	1133	0
Lepidodinium chlorophorum	49	0	18	0

(1) información disponible en NCBI a 010111. % respecto al número de entradas para homo sapiens
En negrita se han remarcado organismos estudiados en los ejemplos citados en la Tabla 1

4. Conclusiones.

La dificultad de identificar proteínas de organismos mal representados en bases de datos de proteínas radica básicamente en la falta de secuencias proteicas disponibles de los mismos. Para intentar optimizar el porcentaje de información de MS y/o MS/MS asignada en un experimento de proteómica con este tipo de organismos pueden abordarse distintas aproximaciones. Por un lado, encontramos la secuenciación *de novo* a partir de los datos de MS/MS obtenidos y posterior identificación de la proteína por homología con otras.

Tabla 5. Ejemplos de utilización de bases de datos genómicos para identificar proteínas de organismos mal representados en bases de datos de proteínas.

Autores	Revista	Año	Organismo en estudio	Proteínas identificadas		
				PROTEÍNAS(1)	ESTs(1)	TOTAL
Lei Z, et al.	Photochemistry	2011	Medicago truncatula	338	357	384
Jones BM, et al.	Mar. Biotechnol	2010	Emiliania huxleyi	33	80	99
Bai S, et al.	Journal of Experimental Botany	2010	Petuniaxhybrida	73	92	133
Luster DG, et al.	Proteomics	2010	Phakopsora pachyrhizi	127	266	393
Schokraie E, et al.	PLoS One	2010	Milnesium tardigradum	71	83	144
Mary J, et al.	Proteomics	2010	Alvinella pompejana	n.p.	19	19
Vicent B, et al.	BMC Genomics	2010	Chelonus inanitus	n.p.	29	29
Ju JW, et al.	Proteomics	2009	Clonorchis sinensis	n.p.	62	62
Bellafiore S, et al.	PLoS Pathogens	2008	Meloidogyne incognita	333	383	486
Chongsatja PO, et al.	Proteomics	2007	Penaeus vannamei	23	28	32
Laugesen S, et al.	Int Jour Mass Spectrom	2007	Hordeum vulgare	12	14	14
Carpentier SC, et al.	Proteomics	2007	Musa spp.	99	79	123
Incamps A, et al.	Biotechnol Bioeng	2005	Medicago sativa	57	70	87
Kwon KH, et al.	Proteomics	2003	Panax Ginseng	76	110	176
Nam MH, et al.	Proteomics	2003	Panax Ginseng	81	117	147

(1) Base de datos utilizada en la identificación, n,p, Dato no proporcionado

Por otro lado, es posible mejorar la información disponible en bases de datos, para el organismo en estudio, a fin de optimizar la aproximación de búsqueda directa en bases de datos a partir de la información de MS-MS/MS. Numerosos estudios realizados recomiendan que ambas estrategias se realicen en paralelo y posteriormente se comparen los resultados obtenidos.

En ambos casos es posible que sea necesario derivatizar químicamente los péptidos para obtener información de MS/MS útil para la identificación.

En cualquier caso el esfuerzo a realizar para identificar este tipo de proteínas por técnicas proteómicas excede, con mucho, lo rutinario. Pero, en general, la información que estos estudios proporcionan justifican sobradamente dicho esfuerzo.

REFERENCIAS.

1. Carpentier SC, Coemans B, Podevin N, et al. Functional genomics in a non-model crop: transcriptomics or proteomics? Physiol Plant. 2008;133:117-30.

2. Mata J, Marguerat S, Bahler J. Post-transcriptional control of gene expression: a genome-wide perspective. Trends Biochem Sci. 2005;30:506-14.

3. Oono Y, Seki M, Nanjo T, et al. Monitoring expression profiles of Arabidopsis gene expression during rehydration process after dehydration using ca 7000 full-length cDNA microarray. Plant J. 2003;34:868-87.

4. Higashi Y, Hirai MY, Fujiwara T, et al. Proteomic and transcriptomic analysis of Arabidopsis seeds: molecular evidence for successive processing of seed proteins and its implication in the stress response to sulfur nutrition. Plant J. 2006;48:557-71.

5. de Groot MJ, Daran-Lapujade P, van Breukelen B, et al. Quantitative proteomics and transcriptomics of anaerobic and aerobic yeast cultures reveals post-transcriptional regulation of key cellular processes. Microbiology. 2007;153:3864-78.

6. Corbin RW, Paliy O, Yang F, et al. Toward a protein profile of Escherichia coli: comparison to its transcription profile. Proc Natl Acad Sci U S A. 2003;100:9232-7.

7. Greenbaum D, Colangelo C, Williams K, et al. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol. 2003;4:117.

8. Gygi SP, Rochon Y, Franza BR, et al. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999;19:1720-30.

9. Griffin TJ, Gygi SP, Ideker T, et al. Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. Mol Cell Proteomics. 2002;1:323-33.

10. Lee PS, Shaw LB, Choe LH, et al. Insights into the relation between mrna and protein expression patterns: II. Experimental observations in Escherichia coli. Biotechnol Bioeng. 2003;84:834-41.

11. Mehra A, Lee KH, Hatzimanikatis V. Insights into the relation between mRNA and protein expression patterns: I. Theoretical considerations. Biotechnol Bioeng. 2003;84:822-33.

12. MacKay VL, Li X, Flory MR, et al. Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone. Mol Cell Proteomics. 2004;3:478-89.

13. Tian Q, Stepaniants SB, Mao M, et al. Integrated genomic and proteomic analyses of gene expression in Mammalian cells. Mol Cell Proteomics. 2004;3:960-9.

14. Kim W, Silby MW, Purvine SO, et al. Proteomic detection of non-annotated protein-coding genes in Pseudomonas fluorescens Pf0-1. PLoS One. 2009;4:e8455.

15. Wolters DA, Washburn MP, Yates JR, 3rd. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001;73:5683-90.

16. Washburn MP, Wolters D, Yates JR, 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19:242-7.

17. Peng J, Elias JE, Thoreen CC, et al. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res. 2003;2:43-50.

18. Shevchenko A, Jensen ON, Podtelejnikov AV, et al. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc Natl Acad Sci U S A. 1996;93:14440-5.

19. Baerenfaller K, Grossmann J, Grobei MA, et al. Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. Science. 2008;320:938-41.

20. Gruhler A, Olsen JV, Mohammed S, et al. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol Cell Proteomics. 2005;4:310-27.

21. Li X, Gerber SA, Rudner AD, et al. Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae. J Proteome Res. 2007;6:1190-7.

22. Bodenmiller B, Mueller LN, Mueller M, et al. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. Nat Methods. 2007;4:231-7.

23. Young NL, DiMaggio PA, Plazas-Mayorca MD, et al. High throughput characterization of combinatorial histone codes. Mol Cell Proteomics. 2009;8:2266-84.

24. Garcia BA, Pesavento JJ, Mizzen CA, et al. Pervasive combinatorial modification of histone H3 in human cells. Nat Methods. 2007;4:487-9.

25. Villen J, Beausoleil SA, Gerber SA, et al. Large-scale phosphorylation analysis of mouse liver. Proc Natl Acad Sci U S A. 2007;104:1488-93.

26. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422:198-207.

27. Marcotte EM. Measuring the dynamics of the proteome. Genome Res. 2001;11:191-3.

28. Byrum S, Montgomery CO, Nicholas RW, et al. The promise of bone cancer proteomics. Ann N Y Acad Sci. 2010;1192:222-9.

29. Buggiotti L, Primmer CR, Kouvonen P, et al. Identification of differentially expressed proteins in Ficedula flycatchers. Proteomics. 2008;8:2150-3.

30. Robinson MW, Menon R, Donnelly SM, et al. An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen Fasciola hepatica: proteins associated with invasion and infection of the mammalian host. Mol Cell Proteomics. 2009;8:1891-907.

31. Chuan J, Feng Z, Brindley PJ, et al. Our wormy world genomics, proteomics and transcriptomics in East and southeast Asia. Adv Parasitol. 2010;73:327-71.

32. Bellafiore S, Shen Z, Rosso MN, et al. Direct identification of the Meloidogyne incognita secretome reveals proteins with host cell reprogramming potential. PLoS Pathog. 2008;4:e1000192.

33. Parker R, Melathopoulos AP, White R, et al. Ecological adaptation of diverse honey bee (Apis mellifera) populations. PLoS One. 2010;5:e11096.

34. Apraiz I, Mi J, Cristobal S. Identification of proteomic signatures of exposure to marine pollutants in mussels (Mytilus edulis). Mol Cell Proteomics. 2006;5:1274-85.

35. Ripley J, Iwanowicz L, Blazer V, et al. Utilization of protein expression profiles as indicators of environmental impairment of smallmouth bass (Micropterus dolomieu) from the Shenandoah River, Virginia, USA. Environ Toxicol Chem. 2008;27:1756-67.

36. Bundy JG, Sidhu JK, Rana F, et al. 'Systems toxicology' approach identifies coordinated metabolic responses to copper in a terrestrial non-model invertebrate, the earthworm Lumbricus rubellus. BMC Biol. 2008;6:25.

37. Calzolai L, Ansorge W, Calabrese E, et al. Transcriptomics and proteomics. Applications to ecotoxicology. Comp Biochem Physiol Part D Genomics Proteomics. 2007;2:245-9.

38. Amalraj RS, Selvaraj N, Veluswamy GK, et al. Sugarcane proteomics: establishment of a protein extraction method for 2-DE in stalk tissues and initiation of sugarcane proteome reference map. Electrophoresis. 2010;31:1959-74.

39. Gonzalez-Fernandez R, Prats E, Jorrin-Novo JV. Proteomics of plant pathogenic fungi. J Biomed Biotechnol. 2010;2010:932527.

40. Acin P, Carrascal M, Abian J, et al. Expression of differential antennal proteins in males and females of an important crop pest, Sesamia nonagrioides. Insect Biochem Mol Biol. 2009;39:11-9.

41. Baker RL, Brown RL, Chen ZY, et al. A maize lectin-like protein with antifungal activity against Aspergillus flavus. J Food Prot. 2009;72:120-7.

42. Forne I, Abian J, Cerda J. Fish proteome analysis: model organisms and non-sequenced species. Proteomics. 2010;10:858-72.

43. Varo I, Rigos G, Navarro JC, et al. Effect of ivermectin on the liver of gilthead sea bream Sparus aurata: A proteomic approach. Chemosphere. 2010;80:570-7.

44. Wu R, Wang W, Yu D, et al. Proteomics analysis of Lactobacillus casei Zhang, a new probiotic bacterium isolated from traditional home-made koumiss in Inner Mongolia of China. Mol Cell Proteomics. 2009;8:2321-38.

45. Xu YJ, Jin ML, Wang LJ, et al. Differential proteome analysis of porcine skeletal muscles between Meishan and Large White. J Anim Sci. 2009;87:2519-27.

46. Gonzalez EG, Krey G, Espineira M, et al. Population Proteomics of the European Hake (Merluccius merluccius). J Proteome Res. 2010;9:6392-404.

47. Hogstrand C, Balesaria S, Glover CN. Application of genomics and proteomics for study of the integrated response to zinc exposure in a non-model fish species, the rainbow trout. Comp Biochem Physiol B Biochem Mol Biol. 2002;133:523-35.

48. Khan FR, Bury NR, Hogstrand C. Cadmium bound to metal rich granules and exoskeleton from Gammarus pulex causes increased gut lipid peroxidation in zebrafish following single dietary exposure. Aquat Toxicol. 2010;96:124-9.

49. Yan F, Zhang Y, Jiang R, et al. Identification and agglutination properties of hemocyanin from the mud crab (Scylla serrata). Fish Shellfish Immunol. 2010.

50. Li P, Hulak M, Rodina M, et al. Comparative protein profiles: Potential molecular markers from spermatozoa of Acipenseriformes (Chondrostei, Pisces). Comp Biochem Physiol Part D Genomics Proteomics. 2010;5:302-7.

51. Kussmann M, Panchaud A, Affolter M. Proteomics in nutrition: status quo and outlook for biomarkers and bioactives. J Proteome Res. 2010;9:4876-87.

52. Wynne C, Fenselau C, Demirev PA, et al. Top-down identification of protein biomarkers in bacteria with unsequenced genomes. Anal Chem. 2009;81:9633-42.

53. Tsai YS, Scherl A, Shaw JL, et al. Precursor ion independent algorithm for top-down shotgun proteomics. J Am Soc Mass Spectrom. 2009;20:2154-66.

54. Renkonen R, Renkonen J, Joenvaara S, et al. Allergens are transported through the respiratory epithelium. Expert Rev Clin Immunol. 2010;6:55-9.

55. Wang X, Chang L, Wang G, et al. Protein extraction from the earthworm Eisenia fetida for 2-DE. Proteomics. 2010;10:1095-9.

56. Domon B, Aebersold R. Mass spectrometry and protein analysis. Science. 2006;312:212-7.

57. Nesvizhskii AI, Vitek O, Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat Methods. 2007;4:787-97.

58. Ahrne E, Muller M, Lisacek F. Unrestricted identification of modified proteins using MS/MS. Proteomics. 2010;10:671-86.

59. Creasy DM, Cottrell JS. Error tolerant searching of uninterpreted tandem mass spectrometry data. Proteomics. 2002;2:1426-34.

60. Pevzner PA, Dancik V, Tang CL. Mutation-tolerant protein identification by mass spectrometry. J Comput Biol. 2000;7:777-87.

61. Ma B, Zhang K, Hendrie C, et al. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom. 2003;17:2337-42.

62. Pan C, Park BH, McDonald WH, et al. A high-throughput de novo sequencing approach for shotgun proteomics using high-resolution tandem mass spectrometry. BMC Bioinformatics. 2010;11:118.

63. Dancik V, Addona TA, Clauser KR, et al. De novo peptide sequencing via tandem mass spectrometry. J Comput Biol. 1999;6:327-42.

64. Zidarov D, Thibault P, Evans MJ, et al. Determination of the primary structure of peptides using fast atom bombardment mass spectrometry. Biomed Environ Mass Spectrom. 1990;19:13-26.

65. Sakurai T, Matsuo T, Matsuda H, et al. PAAS 3: A computer program to determine probable sequence of peptides from mass spectrometric data. Biomed. Mass. Spectrom. 1984;11:3.

66. Uttenweiler-Joseph S, Neubauer G, Christoforidis S, et al. Automated de novo sequencing of proteins using the differential scanning technique. Proteomics. 2001;1:668-82.

67. Frank A, Pevzner P. PepNovo: de novo peptide sequencing via probabilistic network modeling. Anal Chem. 2005;77:964-73.

68. West DB. Introduction to Graph Theory. Prentice Hall Inc.,Englewood Cliffs, NJ. Upper Saddle River, NJ,. 1996.

69. Taylor JA, Johnson RS. Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom. 1997;11:1067-75.

70. Bartels C. Fast algorithm for peptide sequencing by mass spectroscopy. Biological Mass Spectrometry. 1990;19:368.

71. Taylor JA, Johnson RS. Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry. Anal Chem. 2001;73:2594-604.

72. Cormen TH, Leiserson CE, Rivest RL. Introduction to Algorithms. MIT Press, Cambridge, Massachusetts. 1990.

73. Liu C, Song Y, Yan B, et al. Fast de novo peptide sequencing and spectral alignment via tree decomposition. Pac Symp Biocomput. 2006;255-66.

74. Tannu NS, Hemby SE. De novo protein sequence analysis of Macaca mulatta. BMC Genomics. 2007;8:270.

75. Seidler J, Zinn N, Boehm ME, et al. De novo sequencing of peptides by MS/MS. Proteomics. 2010;10:634-49.

76. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. J Mol Biol. 1990;215:403-10.

77. Lipman DJ, Pearson WR. Rapid and sensitive protein similarity searches. Science. 1985;227:1435-41.

78. Pearson WR. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. Genomics. 1991;11:635-50.

79. Pearson WR. Comparison of methods for searching protein sequence databases. Protein Sci. 1995;4:1145-60.

80. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389-402.

81. Sonnhammer EL, Eddy SR, Birney E, et al. Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res. 1998;26:320-2.

82. Karplus K, Barrett C, Cline M, et al. Predicting protein structure using only sequence information. Proteins. 1999;Suppl 3:121-5.

83. Soding J. Protein homology detection by HMM-HMM comparison. Bioinformatics. 2005;21:951-60.

84. Biegert A, Soding J. Sequence context-specific profiles for homology searching. Proc Natl Acad Sci U S A. 2009;106:3770-5.

85. Pearson WR. Empirical statistical estimates for sequence similarity searches. J Mol Biol. 1998;276:71-84.

86. Bandeira N. Protein identification by spectral networks analysis. Methods Mol Biol. 2011;694:151-68.

87. Bandeira N, Pham V, Pevzner P, et al. Automated de novo protein sequencing of monoclonal antibodies. Nat Biotechnol. 2008;26:1336-8.

88. Huang L, Jacob RJ, Pegg SC, et al. Functional assignment of the 20 S proteasome from Trypanosoma brucei using mass spectrometry and new bioinformatics approaches. J Biol Chem. 2001;276:28327-39.

89. Ray R, Haystead TA. Mixed peptide sequencing and the FASTF/FASTS algorithms. Methods Enzymol. 2003; 366:84-95.

90. Wait R, Miller I, Eberini I, et al. Strategies for proteomics with incompletely characterized genomes: the proteome of Bos taurus serum. Electrophoresis. 2002;23:3418-27.

91. Samyn B, Sergeant K, Memmi S, et al. MALDI-TOF/TOF de novo sequence analysis of 2-D PAGE-separated proteins from Halorhodospira halophila, a bacterium with unsequenced genome. Electrophoresis. 2006;27:2702-11.

92. Mackey AJ, Haystead TA, Pearson WR. Getting more from less: algorithms for rapid protein identification with multiple short peptide sequences. Mol Cell Proteomics. 2002;1:139-47.

93. DiMaggio PA, Jr., Floudas CA, Lu B, et al. A hybrid method for peptide identification using integer linear optimization, local database search, and quadrupole time-of-flight or OrbiTrap tandem mass spectrometry. J Proteome Res. 2008;7:1584-93.

94. Waridel P, Frank A, Thomas H, et al. Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated de novo sequencing. Proteomics. 2007;7:2318-29.

95. Liska AJ, Shevchenko A, Pick U, et al. Enhanced photosynthesis and redox energy production contribute to salinity tolerance in Dunaliella as revealed by homology-based proteomics. Plant Physiol. 2004;136:2806-17.

96. Grossmann J, Fischer B, Baerenfaller K, et al. A workflow to increase the detection rate of proteins from unsequenced organisms in high-throughput proteomics experiments. Proteomics. 2007;7:4245-54.

97. Habermann B, Oegema J, Sunyaev S, et al. The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches. Mol Cell Proteomics. 2004;3:238-49.

98. Cheng Q, Cao Y, Jiang C, et al. Identifying secreted proteins of Marssonina brunnea by degenerate PCR. Proteomics. 2010;10:2406-17.

99. Chalkley RJ, Baker PR, Huang L, et al. Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: II. New developments in Protein Prospector allow for reliable and comprehensive automatic analysis of large datasets. Mol Cell Proteomics. 2005;4:1194-204.

100. Kayser JP, Vallet JL, Cerny RL. Defining parameters for homology-tolerant database searching. J Biomol Tech. 2004;15:285-95.

101. Searle BC, Dasari S, Turner M, et al. High-throughput identification of proteins and unanticipated sequence modifications using a mass-based alignment algorithm for MS/MS de novo sequencing results. Anal Chem. 2004;76:2220-30.

102. Searle BC, Dasari S, Wilmarth PA, et al. Identification of protein modifications using MS/MS de novo sequencing and the OpenSea alignment algorithm. J Proteome Res. 2005;4:546-54.

103. Pegg SC, Babbitt PC. Shotgun: getting more from sequence similarity searches. Bioinformatics. 1999;15:729-40.

104. Shevchenko A, Sunyaev S, Loboda A, et al. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. Anal Chem. 2001;73:1917-26.

105. Chalkley RJ, Hansen KC, Baldwin MA. Bioinformatic methods to exploit mass spectrometric data for proteomic applications. Methods Enzymol. 2005;402:289-312.

106. Han Y, Ma B, Zhang K. SPIDER: software for protein identification from sequence tags with de novo sequencing error. J Bioinform Comput Biol. 2005;3:697-716.

107. Mann M, Wilm M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal Chem. 1994;66:4390-9.

108. Kim S, Bandeira N, Pevzner PA. Spectral profiles, a novel representation of tandem mass spectra and their applications for de novo peptide sequencing and identification. Mol Cell Proteomics. 2009;8:1391-400.

109. Tanner S, Shu H, Frank A, et al. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. Anal Chem. 2005;77:4626-39.

110. Shilov IV, Seymour SL, Patel AA, et al. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. Mol Cell Proteomics. 2007;6:1638-55.

111. Gupta N, Bark SJ, Lu WD, et al. Mass spectrometry-based neuropeptidomics of secretory vesicles from human adrenal medullary pheochromocytoma reveals novel peptide products of prohormone processing. J Proteome Res. 2010;9:5065-75.

112. Yates JR, 3rd, Eng JK, McCormack AL, et al. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem. 1994;67:1426-36.

113. Perkins DN, Pappin DJ, Creasy DM, et al. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999;20:3551-67.

114. Liska AJ, Popov AV, Sunyaev S, et al. Homology-based functional proteomics by mass spectrometry: application to the Xenopus microtubule-associated proteome. Proteomics. 2004;4:2707-21.

115. Sunyaev S, Liska AJ, Golod A, et al. MultiTag: multiple error-tolerant sequence tag search for the sequence-similarity identification of proteins by mass spectrometry. Anal Chem. 2003;75:1307-15.

116. Tabb DL, Saraf A, Yates JR, 3rd. GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. Anal Chem. 2003;75:6415-21.

117. Kondo H, Rabouille C, Newman R, et al. p47 is a cofactor for p97-mediated membrane fusion. Nature. 1997;388:75-8.

118. Boersema PJ, Taouatas N, Altelaar AF, et al. Straightforward and de novo peptide sequencing by MALDI-MS/MS using a Lys-N metalloendopeptidase. Mol Cell Proteomics. 2009;8:650-60.

119. Wilm M, Shevchenko A, Houthaeve T, et al. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature. 1996;379:466-9.

120. Shevchenko A, Chernushevich I, Ens W, et al. Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. Rapid Commun Mass Spectrom. 1997;11:1015-24.

121. Garcia-Murria MA, Valero MA, Sanchez Del Pino MM. Simple chemical tools to expand the range of proteomics applications. J Proteomics. 2011;74:137-50.

122. Summerfield SG, Bolgar MS, Gaskell SJ. Promotion and Stabilization of b_1 ions in Peptide Phenythiocarbamoyl Derivatives: Analogies with Condensed-phase Chemistry. Journal of Mass Spectrometry. 1997;32:225-31.

123. Roth KD, Huang ZH, Sadagopan N, et al. Charge derivatization of peptides for analysis by mass spectrometry. Mass Spectrom Rev. 1998;17:255-74.

124. Dongre AR, Jones JL, Somogyi A, et al. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: evidence for the mobile proton model. J Am Chem Soc. 1996;118:9.

125. Keough T, Youngquist RS, Lacey MP. Sulfonic acid derivatives for peptide sequencing by MALDI MS. Anal Chem. 2003;75:156A-65A.

126. Hjerno K, Alm R, Canback B, et al. Down-regulation of the strawberry Bet v 1-homologous allergen in concert with the flavonoid biosynthesis pathway in colorless strawberry mutant. Proteomics. 2006;6:1574-87.

127. Oehlers LP, Perez AN, Walter RB. Detection of hypoxia-related proteins in medaka (Oryzias latipes) brain tissue by difference gel electrophoresis and de novo sequencing of 4-sulfophenyl isothiocyanate-derivatized peptides by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Comp Biochem Physiol C Toxicol Pharmacol. 2007;145:120-33.

128. Leon IR, Neves-Ferreira AG, Valente RH, et al. Improved protein identification efficiency by mass spectrometry using N-terminal chemical derivatization of peptides from Angiostrongylus costaricensis, a nematode with unknown genome. J Mass Spectrom. 2007;42:1363-74.

129. Lesur A, Varesio E, Hopfgartner G. Protein quantification by MALDI-selected reaction monitoring mass spectrometry using sulfonate derivatized peptides. Anal Chem. 2010;82:5227-37.

130. Salekdeh GH, Siopongco J, Wade LJ, et al. Proteomic analysis of rice leaves during drought stress and recovery. Proteomics. 2002;2:1131-45.

131. Jung S, Fladerer C, Braendle F, et al. Identification of a novel Plasmopara halstedii elicitor protein combining de novo peptide sequencing algorithms and RACE-PCR. Proteome Sci. 2010;8:24.

132. Staheli JP, Ryan JT, Bruce AG, et al. Consensus-degenerate hybrid oligonucleotide primers (CODEHOPs) for the detection of novel viruses in non-human primates. Methods. 2009;49:32-41.

133. Wang W, Sun J, Nimtz M, et al. Protein identification from two-dimensional gel electrophoresis analysis of Klebsiella pneumoniae by combined use of mass spectrometry data and raw genome sequences. Proteome Sci. 2003;1:6.

134. Wu J, Lin Q, Lim TK, et al. White spot syndrome virus proteins and differentially expressed host proteins identified in shrimp epithelium by shotgun proteomics and cleavable isotope-coded affinity tag. J Virol. 2007;81:11681-9.

135. Wang W, Meng B, Chen W, et al. A proteomic study on postdiapaused embryonic development of brine shrimp (Artemia franciscana). Proteomics. 2007;7:3580-91.

136. Van Leene J, Stals H, Eeckhout D, et al. A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana. Mol Cell Proteomics. 2007;6:1226-38.

137. Suckau D, Resemann A, Schuerenberg M, et al. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. Anal Bioanal Chem. 2003;376:952-65.

138. Wright JC, Sugden D, Francis-McIntyre S, et al. Exploiting proteomic data for genome annotation and gene model validation in Aspergillus niger. BMC Genomics. 2009;10:61.

139. Allmer J, Markert C, Stauber EJ, et al. A new approach that allows identification of intron-split peptides from mass spectrometric data in genomic databases. FEBS Lett. 2004;562:202-6.

140. Edwards NJ. Novel peptide identification from tandem mass spectra using ESTs and sequence database compression. Mol Syst Biol. 2007;3:102.

141. Kwon KH, Kim M, Kim JY, et al. Efficiency improvement of peptide identification for an organism without complete genome sequence, using expressed sequence tag database and tandem mass spectral data. Proteomics. 2003;3:2305-9.

142. de Souza GA, Arntzen MO, Wiker HG. MSMSpdbb: providing protein databases of closely related organisms to improve proteomic characterization of prokaryotic microbes. Bioinformatics. 2010;26:698-9.

143. de Souza GA, Arntzen MO, Fortuin S, et al. Proteogenomic Analysis of Polymorphisms and Gene Annotation Divergences in Prokaryotes using a Clustered Mass Spectrometry-Friendly Database. Mol Cell Proteomics. 2011;10:M110 002527.

144. Jones BM, Edwards RJ, Skipp PJ, et al. Shotgun Proteomic Analysis of Emiliania huxleyi, a Marine Phytoplankton Species of Major Biogeochemical Importance. Mar Biotechnol (NY). 2010.

145. Boguski MS, Lowe TM, Tolstoshev CM. dbEST--database for "expressed sequence tags". Nat Genet. 1993;4:332-3.

146. Benson DA, Karsch-Mizrachi I, Lipman DJ, et al. GenBank. Nucleic Acids Res. 2009;37:D26-31.

147. Staden R. A strategy of DNA sequencing employing computer programs. Nucleic Acids Res. 1979;6:2601-10.

148. Lee Y, Tsai J, Sunkara S, et al. The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. Nucleic Acids Res. 2005;33:D71-4.

149. Stanton JA, Macgregor AB, Green DP. Identifying tissue-enriched gene expression in mouse tissues using the NIH UniGene database. Appl Bioinformatics. 2003;2:S65-73.

150. Christoffels A, van Gelder A, Greyling G, et al. STACK: Sequence Tag Alignment and Consensus Knowledgebase. Nucleic Acids Res. 2001;29:234-8.

151. Grimplet J, Cramer GR, Dickerson JA, et al. VitisNet: "Omics" integration through grapevine molecular networks. PLoS One. 2009;4:e8365.

152. Bernal A, Ear U, Kyrpides N. Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. Nucleic Acids Res. 2001;29:126-7.

153. Noda H, Kawai S, Koizumi Y, et al. Annotated ESTs from various tissues of the brown planthopper Nilaparvata lugens: a genomic resource for studying agricultural pests. BMC Genomics. 2008;9:117.

154. Liu F, Chen P, Cui SJ, et al. SjTPdb: integrated transcriptome and proteome database and analysis platform for Schistosoma japonicum. BMC Genomics. 2008;9:304.

155. Ning K, Fermin D, Nesvizhskii AI. Computational analysis of unassigned high-quality MS/MS spectra in proteomic data sets. Proteomics. 2010;10:2712-8.

156. Lei Z, Dai X, Watson BS, et al. A legume specific protein database (LegProt) improves the number of identified peptides, confidence scores and overall protein identification success rates for legume proteomics. Phytochemistry. 2011.

157. Bai S, Willard B, Chapin LJ, et al. Proteomic analysis of pollination-induced corolla senescence in petunia. J Exp Bot. 2010;61:1089-109.

158. Luster DG, McMahon MB, Carter ML, et al. Proteomic analysis of germinating urediniospores of Phakopsora pachyrhizi, causal agent of Asian soybean rust. Proteomics. 2010;10:3549-57.

159. Schokraie E, Hotz-Wagenblatt A, Warnken U, et al. Proteomic analysis of tardigrades: towards a better understanding of molecular mechanisms by anhydrobiotic organisms. PLoS One. 2010;5:e9502.

160. Mary J, Rogniaux H, Rees JF, et al. Response of Alvinella pompejana to variable oxygen stress: a proteomic approach. Proteomics. 2010;10:2250-8.

161. Vincent B, Kaeslin M, Roth T, et al. The venom composition of the parasitic wasp Chelonus inanitus resolved by combined expressed sequence tags analysis and proteomic approach. BMC Genomics. 2010;11:693.

162. Ju JW, Joo HN, Lee MR, et al. Identification of a serodiagnostic antigen, legumain, by immunoproteomic analysis of excretory-secretory products of Clonorchis sinensis adult worms. Proteomics. 2009;9:3066-78.

163. Chongsatja PO, Bourchookarn A, Lo CF, et al. Proteomic analysis of differentially expressed proteins in Penaeus vannamei hemocytes upon Taura syndrome virus infection. Proteomics. 2007;7:3592-601.

164. Laugesen S, Bak-Jensen KS, Hägglund P, et al. Barley peroxidase isozymes Expression and post-translational modification in mature seeds as identified by two-dimensional gel electrophoresis and mass spectrometry. International Journal of Mass Spectrometry. 2007;268:9.

165. Carpentier SC, Witters E, Laukens K, et al. Banana (Musa spp.) as a model to study the meristem proteome: acclimation to osmotic stress. Proteomics. 2007;7:92-105.

166. Incamps A, Hely-Joly F, Chagvardieff P, et al. Industrial process proteomics: alfalfa protein patterns during wet fractionation processing. Biotechnol Bioeng. 2005;91:447-59.

167. Nam MH, Heo EJ, Kim JY, et al. Proteome analysis of the responses of Panax ginseng C. A. Meyer leaves to high light: use of electrospray ionization quadrupole-time of flight mass spectrometry and expressed sequence tag data. Proteomics. 2003;3:2351-67.

168. Chalkley RJ, Baker PR, Hansen KC, et al. Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: I. How much of the data is theoretically interpretable by search engines? Mol Cell Proteomics. 2005;4:1189-93.

Proteómica en hongos fitopatógenos

25

Dra. Raquel González Fernández[1,2], Dr. José Valero Galván[1,2], Prof. Jesús V. Jorrín Novo[1]

[1] Departamento de Ciencias Químico Biológicas, Instituto de Ciencias Biomédicas, Universidad Autónoma de Ciudad Juárez, Chihuahua, México;

[2] Grupo de Bioquímica y Proteómica Vegetal y Agroforestal, Departamento de Bioquímica y Biología Molecular, Universidad de Córdoba, Spain.

1. Introdución.

Los hongos constituyen un amplio y heterogéneo grupo de organismos eucariotas. La mayoría de las especies de hongos son saprófitas. Sin embargo, un número de ellas son parásitos de las plantas, debido a que necesitan un hospedador para completar su ciclo biológico. Sólo un número limitado de especies fúngicas son capaces de penetrar e invadir los tejidos del hospedador, evitando el reconocimiento y la respuesta de defensa de las plantas para obtener nutrientes de éstas, y causar enfermedades que pueden ocasionar a veces la muerte del hospedador [1]. Los hongos fitopatógenos tienen gran interés agronómico, ya que causan significativas pérdidas tanto en el rendimiento pre-cosecha como en el almacenamiento post-cosecha [2]. Dichos hongos poseen una gran versatilidad en sus mecanismos de infección ya que pueden afectar a diferentes plantas hospedadoras y a diferentes tejidos vegetales (por ejemplo, hojas, flores, raíces y frutos).

Los hongos fitopatógenos poseen un ciclo de vida complicado, reproduciéndose tanto de una forma asexual como sexual, y con unas etapas del ciclo que implican la formación de diferentes estructuras de infección, vegetativas y reproductivas [3]. La interacción entre el hongo y la planta depende de su reconocimiento mutuo, la señalización y la expresión de factores de patogenicidad y virulencia por parte del hongo, mientras que por parte de la planta, depende de la existencia de mecanismos de defensa. Esto resulta en interacciones compatibles (susceptibilidad) o incompatibles (en plantas no hospedadoras, o con resistencia basal o específica de la planta hospedadora). Existe un gran número de mecanismos y moléculas que contribuyen a la patogenicidad o virulencia fúngica, entre los que se incluyen proteínas que degradan la pared celular vegetal [4], proteínas inhibidoras [5] y micotoxinas [69]. Por tanto, el conocimiento del ciclo de vida del patógeno y de los factores virulencia es fundamental para el diseño de estrategias eficaces de protección de cultivos, las cuales incluyen el desarrollo de genotipos vegetales resistentes obtenidos mediante la reproducción vegetal clásica [10] o ingeniería genética [11], el desarrollo de fungicidas [12] o el uso de estrategias de control biológico [13].

Los genes que causan patogenicidad se definen como aquellos genes necesarios para el desarrollo de la enfermedad, pero que no son esenciales para el patógenos para completar su ciclo de vida *in vitro* [14]. El estudio de los productos de estos genes es de gran relevancia para utilizarse como potenciales dianas para el desarrollo de fungicidas para el control de la enfermedad, y que pueden descubrirse mediante el análisis de las secuencias de los genomas, transcriptomas y proteomas. Recientemente se ha creado una base de datos de interacciones

entre patógenos y hospedadores ("Pathogen-Host Interactions", http://www.phi-base.org) [15, 16]. Esta base de datos cataloga los genes de patogenecidad, virulencia y efectores verificados experimentalmente de hongos, oomicetos y bacterias patógenas que infectan a hospedadores humanos, animales, plantas, insectos, peces y hongos.

Se han realizado diversos estudios sobre los hongos fitopatógenos y sus interacciones con las plantas mediante la genética clásica, la biología celular y la bioquímica, así como utilizando modernas y holísticas técnicas ómicas de alto rendimiento [14, 17-21]. Además, se han desarrollado herramientas bioinformáticas específicas, por ejemplo, para precedir productos génicos involucrados en las interacciones hongo patogénico-planta [22]. En los últimos años, los estudios de hongos fitopatógenos han experimentado un notable aumento debido al creciente número de proyectos de secuenciación de genomas de hongos de interés agronómico, lo que ha facilitado enormemente la posibilidad de identificar las proteínas de interés. Se puede consultar los proyectos de secuenciación terminados o en marcha en las siguientes páginas web: '*the Broad Institute Database*', http://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/fungal-genome-initiative; '*the Phytopathogenic Fungi and Oomycete EST Database*', COGEME, http://cogeme.ex.ac.uk. Así, por ejemplo, para *Botrytis cinerea* se han secuenciado los genomas de dos cepas silvestres aisladas de dos hospedadores diferentes: B05.10 ('*Botrytis cinerea Sequencing Project, Broad Institute of Harvard and MIT*': http://www.broadinstitute.org/annotation/genome/botrytis_cinerea/MultiHome.html; [23]) y T4 ('*Botrytis cinerea genome project*, URGI, Genoscope': http://www.genoscope.cns.fr/spip/-Botrytis-cinerea-.html

Al igual que la genómica y la transcriptómica, la proteómica ha evolucionado para incorporar técnicas de alto rendimiento y protocolos que permiten un análisis más rápido de un gran número de proteinas [24, 25]. Así la proteómica constituye una herramienta eficaz para proporcionar información importante sobre la patogenicidad y los factores de virulencia relacionados con ésta, a fin de abrir nuevas posibilidades para el diagnóstico de las enfermedades, y desarrollar estrategias de protección de cultivos eficientes y que sean además respetuosas con el medio ambiente [19-21]. A diferencia de otros sistemas biológicos, como en las levaduras [26] y en humanos [27], el potencial de la proteómica en investigación de hongos patógenos de plantas está lejos de ser plenamente explotado, como pone de manifiesto el bajo número de dichas especies bajo investigación con dicha técnica, la baja cobertura del proteoma de las especies investigadas, y el uso casi exclusivo de las técnicas clásicas de primera generación que se basan en electroforesis bidimensional (2-DE) acoplada a la espectrometría de masas (MS) (revisado recientemente en [21]). Aún así, todo esto, junto con la posibilidad de obtener mutantes '*knockout*' o de silenciar genes [18], está contribuyendo a la compresión de la patogenicidad, virulencia y agresividad de los hongos fitopatógenos.

2. Flujo de trabajo en proteómica adaptado a hongos fitopatógenos.

En general, el flujo de trabajo de un análisis proteómico incluye, entre otros, los siguientes pasos (Figura 1): diseño experimental, muestreo, preparación de las muestras (liofilización, pulverización), extracción de las proteínas (precipitación y limpieza), separación de las proteínas, análisis mediante MS, identificación de las proteínas, análisis estadístico de los datos, cuantificación, y análisis, manejo y almacenamiento de los datos [3, 21]. El protocolo más adecuado a utilizar depende y debe ser optimizado para el sistema biológico (que en nuestro caso son hongos fitopatógenos), es decir, especie de hongo, material de partida (micelio, proteínas secretadas, esclerocios, esporas), así como los objetivos de la investigación (descriptivo, comparativo, PTMs, interacciones entre proteínas, proteómica dirigida).

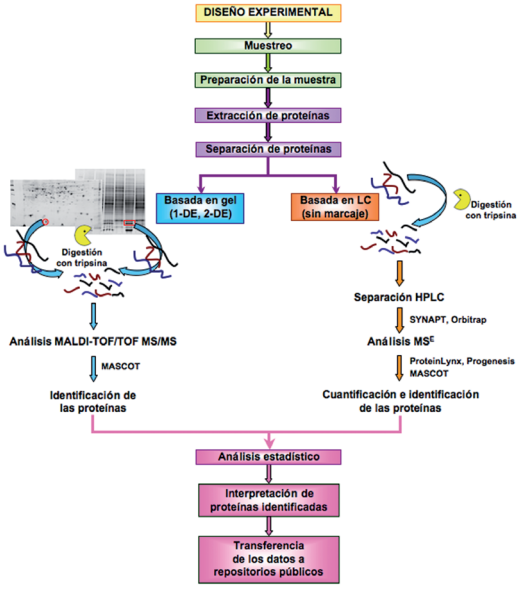


Figura 1. Esquema del flujo de trabajo del análisis proteómico. Se representan las etapas del flujo de trabajo utilizado en nuestro grupo para *B. cinerea.* En dicho esquema se muestran también las diferentes técnicas de separación de proteínas que utilizamos para una mayor covertura del proteoma de dicho hongo: separación basada en gel (*'gel-based'*) (1-DE y 2-DE) y separación basada en LC sin marcaje (*'gel-free/label-free'*).

2.1. Diseño experimental.

Un buen diseño experimental es crucial para el éxito de cualquier experimento en proteómica [28, 29]. Para ello, el establecimiento de un número adecuado de réplicas es esencial. Este número debe establecerse teniendo en cuenta el carácter dinámico del proteoma, y un buen número de réplicas permitirá una correcta interpretación de los resultados y una segura asignación de cualquier proteína al grupo de aquellas que son variables [30, 31]. También es muy importante la realización de un estudio de la variabilidad analítica y biológica. La variabilidad analítica examina tanto los métodos utilizados en las diferentes etapas del experimento (la extracción de proteínas, isoelectroenfoque, electroforesis, tinción-decoloración de los geles) como la exactitud de los equipos y programas informáticos utilizados en la adquisición y análisis de imágenes y datos. Por otra parte, la variabilidad biológica examina las diferencias entre muestras biológicas independientes. Así, por ejemplo, en los estudios realizados para caracterizar el perfil de proteínas de la cepa silvestre 2100 de *B. cinerea*, se determinó la variabilidad analítica y biológica a partir de 64 manchas proteicas de tres réplicas de geles 2-DE de la misma muestra para la variabiliadad analítica y de geles 2-DE de tres muestras independientes para la variabilidad biológica, analizándose los coeficientes de varianza (CV). Así se obtuvo una variabilidad analítica del 16,1% y una variabilidad biológica del 37,5 %. La variabilidad analítica fue similar a la obtenida para bacterias, plantas, ratones o humanos, y la biológica fue mayor que la conseguida para otros organismos cultivados en condiciones controladas [32]. El uso de más de seis repeticiones biológicas no redujo significativamente el error estándar, por lo que esta cifra podría ser óptima para la comparativa en experimentos proteómicos. Ya que normalmente esto no es factible, la mayoría de los trabajos publicados utilizan sólo tres o cuatro réplicas biológicas. De la misma manera, se realizó un estudio de la variabilidad biológica de seis cepas silvestres de *B. cinerea* (B05.10, T4, 2100, 2850, 2996 y 20518) (Tabla 1) analizando los CV de 75 manchas proteicas escogidas aleatoriamente [33]. Los CV variaron entre el 29,3% para la cepa B05.10 y el 42,7% para la 20518. Para el caso del CV de la cepa 2100 fue de 35,0%,

Tabla 1. Descripción de las cepas de B. cinerea estudias en Gonzalez-Fernandez et al. (2013) [33].

Cepas	Descripción
B05.10	Obtenida de un aislado de *Vitis* tras un tratamiento con benomilo [34, 125]
T4	Aislada de plantas de tomate crecidas en un invernadero (Francia) [126]
2100a	Aislada de hojas de *Vicia faba* (Reino Unido)
2850a	Aislada de bulbos de azafrán (Spain)
2996a	Aislada de planta de haba (Reino Unido)
20518a	Aislada de uvas podridas (España)

*Estas cepas se obtuvieron en la Colección Española de Cultivos Tipo (CECT; http://www.cect.org/).

porcentaje similar al obtenido en el estudio previo realizado por Fernández-Acero *et al.* [32]. Era de esperar que la cepa B05.10 tuviera una variabilidad biológica más baja, ya que esta cepa fue obtenida tras un tratamiento con benomilo de un aislado colectado en *Vitis vinifera* (cepa SAS56) para conseguir una cepa más estable [34].

2.2. Rotura de la pared celular.

La mayoría de los hongos fitopatógenos son hongos filamentosos. Este tipo de hongos pueden considerarse, al igual que las plantas, como material biológico recalcitrante, por lo que la preparación de las muestras es un paso crítico. La ruptura celular es difícil debido a la presencia de una pared celular robusta, formada principalmente por quitina, y que constituye la mayor parte de la masa celular [35]. Para superar esta dificultad, existen varios métodos como la lisis mecánica, utilizando perlas de vidrio [36, 37] o molinos celulares [38], o la lisis por medio de ultrasonidos [39]. Se ha comprobado que estos métodos son más eficientes que aquellos basados en la extracción química o enzimática [40]. Shimizu y Wariishi [41] emplearon un método alternativo para evitar la dificultad de romper la pared celular del hongo mediante la generación de protoplastos de *Tyromyces palustris*, obteniendo así un mejor patrón proteíco bidimensional a partir de protoplastos que de las células intactas. Sin embargo, el método más ampliamente utilizado para la ruptura celular es mecánico y consiste en pulverizar el micelio en nitrógeno líquido usando un mortero [28, 33, 42].

2.3. Extracción de proteínas.

2.3.1. Micelio.

La obtención de extractos de proteínas es también un paso crucial para el análisis proteómico [43]. El protocolo más utilizado para la precipitación de las proteínas de extractos de material fúngico utiliza soluciones que contienen disolventes orgánicos, como el ácido tricloroacético (TCA), seguido de la solubilización del precipitado en una solución apropiada. Este método minimiza la degradación o modificación de proteínas durante el proceso, evita la actividad de las proteasas [44] y se eliminan compuestos como sales, polisacáridos, polifenoles, ácido nucleicos y lípidos, que puede interferir en el isoelectroenfoque (IEF) [45]. Sin embargo, el tratamiento con TCA complica la posterior solubilización de proteínas, en especial de las proteínas hidrofóbicas. Este problema se superó en parte mediante el uso de sustancias caotrópicas (como, la urea y la tiourea) [46], de nuevos detergentes zwitteriónicos (como el 3-[(3-colamidopropil) dimetilamonio]-1-propanosulfonato (CHAPS) o la sulfobetaína 3-10 (SB 3-10)) [47-50], o con el uso de un breve tratamiento con hidróxido de sodio [44], para provocar un aumento de la resolución y capacidad de los geles 2-DE. Otros métodos de extracción de proteínas han introducido el uso de una solución de extracción ácida para reducir el rayado (en inglés '*streaking*') producido en los geles 2-DE a causa de la pared celular del hongo [51], así como de un tampón fosfato para solubilizar las proteínas antes de la precipitación [52, 53]. En un estudio reciente de nuestro grupo en el que se estudió el proteoma de seis cepas silvestres de *B. cinerea* (Tabla 1), el rendimiento en la extracción de proteína varió de 6,3 (cepa 2996) a 14,6 mg/g de micelio liofilizado (cepa B05.10), dependiendo de la cepa. En el caso de la cepa 2100, el rendimiento de proteína fue mayor que el obtenido con otros protocolos (6,0 mg/g de micelio liofilizado mediante la solubilización de las proteínas con un tampón fosfato y su precipitación mediante TCA/acetona, a 8,7 mg/g micelio liofilizado [33]). Estos datos coincidieron con los resultados mostrados en un estudio posterior de los mismos autores [32], donde el uso combinado de la precipitación de proteínas mediante TCA/acetona y su extracción posterior con fenol proporcionó una definición mejor de las proteínas en los geles, debido a que reduce el rayado y se detectan un mayor número de manchas proteicas. El mayor rendimiento en la extracción de proteína obtenido para la cepa B05.10 podría deberse tanto al método de extracción utilizado como a las características intrínsecas propias del micelio entre las diferentes cepas (por ejemplo, a la facilidad para romper la pared celular de los hongos y a los compuestos fúngicos intracelulares que

pueden afectar el rendimiento de extracción) [28]. Sin embargo, esta cepa también pudiera mostrar una mayor biosíntesis y acumulación de proteínas que las otras cepas en las condiciones de cultivo del estudio. Una alternativa o complemento a la utilización del TCA es el uso de la polivilnilpolipirrolidona (PVPP) que se añade en la extracción con fenol. La PVPP es un polímero insoluble que fija de manera selectiva los polifenoles limpiando la muestra de dichos compuestos [54]. También se pueden eliminar los compuestos que pueden interferir en el IEF (aniones TCA, polisacáridos, compuestos fenólicos), introduciendo varias etapas de bajo voltaje (150, 250, 500 y 1000 V) previas al enfoque [55].

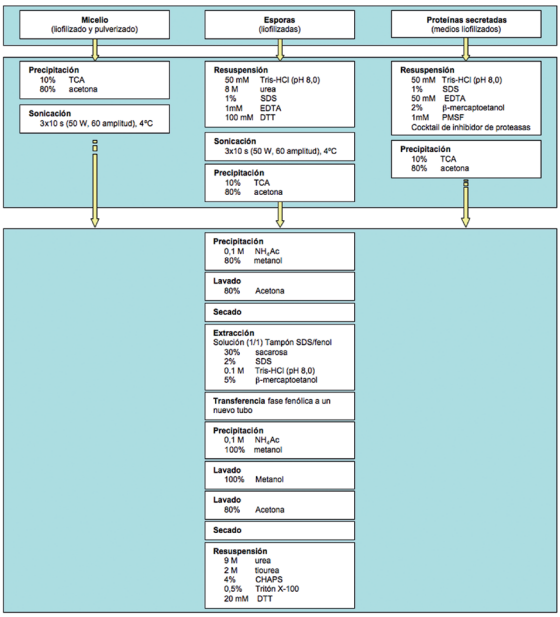


Figura 2. Esquema de los protocolos de extracción de proteína. Se representan los protocolos esquematizados que se utilizan en nuestro grupo, dependiendo del material de partida de *B. cinerea*. Para micelio y proteínas secretadas se utilizó el método combinado de extracción de TCA/acetona-fenol con algunas modificaciones [123, 124] y para los conidios una extracción con tampón más el método de precipitación anteriormente mencionado [39, 123, 124].

En la figura 2 se puede observar un esquema de las etapas utilizadas en nuestro laboratorio para la extracción de proteínas de micelio, esporas y proteínas secretadas en *B. cinerea* [21]. Asímismo, recientemente se ha publicado los protocolos detallados utilizados por nuestro grupo [28, 43].

2.3.2. Esporas.

En el caso de las esporas, existen escasos trabajos que utilicen este material y la mayoría de ellos utilizan un método mecánico como es el uso del mortero y la posterior homogenización con un tampón de extracción [5658]. En un estudio realizado con *Blumeria graminis* f. sp. *hordei*, las esporas se rompieron con una pequeña cantidad de arena en un mortero y a continuación se añadió un tampón de extracción para solubilizar las proteínas [57]. En otro estudio realizado en *Magnaporthe oryzae*, las esporas se rompieron también utilizando un mortero y se homogenizaron con un tampón de extracción que contienía Mg/CHAPS entre otros compuestos [56]. Además, se han optimizado protocolos alternativos para la extracción de proteínas de esporas utilizando condiciones ácidas y el uso de ultrasonidos para producir la rotura de la pared celular [39].

2.3.3. Proteínas secretadas.

Para la obtención de las proteínas secretadas se necesitan protocolos especiales, debido a que en los medios existe una concentración muy baja de proteínas, a veces por debajo del límite de detección de los métodos colorimétricos de cuantificación (Bradford, Lowry o BCA), y a la presencia de polisacáridos, material mucilaginoso, sales y metabolitos (ácidos orgánicos de bajo peso molecular, ácidos grasos, fenoles, quinonas y otros compuestos aromáticos) que son también secretados por el hongo. La presencia de estos compuestos extracelulares puede afectar a los métodos de cuantificación de proteína anteriormente mencionados, provocando una fuerte sobreestimación de la cantidad de proteína total [59]. Esta determinación también puede verse afectada por la alta concentración de ciertos reactivos en las soluciones de solubilización (como la urea, tiourea o DTT) que pueden interferir en la medición espectrofotométrica, produciendo una sobreestimación de la cantidad total de proteína en la que, dependiendo del

método, las diferencias varían en dos órdenes de magnitud [55]. Así la comparación de los diferentes métodos de precipitación de proteínas ha demostrado su limitada aplicabilidad al analizar el secretoma completo en hongos [33, 55, 60-62].

2.4. Separación de proteínas.

En investigación de hongos fitopatógenos, la proteómica basada en gel es la principal técnica usada en la separación de proteínas (Tabla 2). A pesar de su simplicidad, la electroforesis monodimensional (1-DE) puede ser una técnica válida para proporcionar información, especialmente en el caso de la proteómica comparativa cuando el número de muestras a analizar es elevado. Así, mediante el uso de 1-DE es posible distinguir entre fenotipos de diferentes cepas silvestres de *B. cinerea* e identificar proteínas que participan en los mecanismos de patogenicidad (Figura 3) [33]. Con el software adecuado, la 1-DE es una técnica sencilla y fiable para obtener perfiles proteicos y es especialmente útil en el caso de las proteínas hidrofóbicas y de bajo peso molecular [63]. Además, la combinación de la 1-DE y la cromatografía líquida (LC) (donde tras separar las proteínas en el gel, se cortan las bandas, se digieren las proteínas con tripsina y los péptidos resultantes se separan mediante LC), y el análisis MS/MS es capaz de proporcionar una mayor cobertura de las proteínas de las muestras. Por otro lado, la 1-DE es un buen método para la obtención de resultados preliminares antes de comenzar con el análisis 2-DE [33, 64-68].

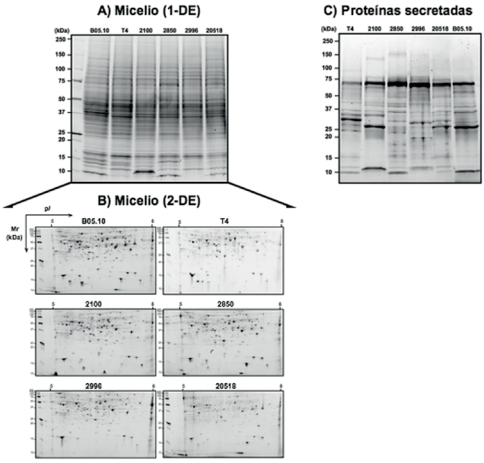


Figura 3. Perfil de proteínas (A) y mapas 2-DE (B) para micelio y perfil de proteínas secretadas (C) de seis cepas silvestres de *B. cinerea* (Tabla 1). Para la 1-DE y 2-DE, se utilizaron geles *'Criterion Stain Free™ precast Tris–HCl'*, con gradiente lineal de 4–20% y de 8–16% respectivamente (Bio-Rad). Se cargaron 15 µg y 50 µg de proteínas de micelio para los geles 1-DE y 2-DE respectivamente. En la 2-DE, se utilizaron tiras de IPG de pH 5–8 lineales de 11 cm. En el caso de la 1-DE de las proteínas secretadas se cargaron, 1,5 µg de proteínas.

La 2-DE es la plataforma dominante en investigación de hongos fitopatógenos (Tabla 2). En el estudio del proteoma del micelio de seis cepas silvestres de *B. cinerea* (Tabla 1) mediante 2-DE, se resolvieron entre 431±26 manchas proteicas para la cepa 2996 y 557±22 para la B05.10 en un rango de p*I* de 5 a 8 [33]. De las 674 manchas proteicas consistentes y diferentes detectadas en las tres réplicas de todas las cepas, 187 fueron comunes a todas ellas. La cepa B05.10 fue la más divergente, presentando 44 manchas proteicas específicas de la cepa (un 8,6% del total) y la cepa 2996 la menos con sólo 7 manchas proteicas específicas (un 1,7% del total).

Tabla 2. Lista de hongos fitopatógenos que han sido estudiados mediante un enfoque proteómico.

HONGO	MATERIAL FÚNGICO	COMENTARIO	FLUJO DE TRABAJO EXPERIMENTAL	NÚMERO DE PROTEÍNAS IDENTIFICADAS	REFERENCIA
Aspergillus ssp.	Esporas	Caracterización y tipado de esporas intactas de 20 especies de *Aspergillus* mediante MS. Optimización de la extracción de proteínas.	1-DE, MALDI-TOF MS	11	[39]
Aspergillus flavus	Secretoma	Identificación de proteínas secretadas de *A. flavus* crecido en medio líquido que contiene flavonoides (3-O-glucósido) y en medio líquido rico (patata-dextrosa).	1-DE/2-DE, LC-MS/MS	22	[66]
	Secretoma	Identificación de proteínas secretadas de *A. flavus* crecido en medio líquido con tres fuentes de carbono diferentes (flavonoides (3-O-glucósido) glucosa y patata-dextrosa).	1-DE/2-DE, MALDI-TOF-MS	51	[65]
Blumeria graminis f.sp. hordei	Esporas	Descripción del proteoma de las esporas	2-DE, MALDI-TOF/ TOF-MS/MS	123	[57]
	Haustorios	Descripción del proteoma de los haustorios.	LC-MS/MS	204	[127]
	Micelio esporulante, haustorios in planta	Descripción del proteoma del micelio.	1-DE, nESI-LC-MS/MS	71	[128]

Botrytis cinerea	Micelio	Descripción del proteoma del micelio	2-DE, MALDI-TOF/TOF-MS/MS, LC-MS/MS	22	[53]
	Micelio	Análisis comparativo de dos cepas silvestres que difieren en virulencia y en la producción de toxinas.	2-DE, MALDI-TOF/TOF-MS/MS, LC-MS/MS	8	[52]
	Micelio	Análisis del proteoma intracelular durante el proceso de degradación de celulosa.	2-DE, MALDI-TOF/TOF-MS/MS	306	[32]
	Secretoma	Análisis del secretoma usando medio mínimo suplementado con extractos de tomate maduro (totalmente rojo), fresas maduras o de hojas de Arabidopsis.	1-DE, LC MS/MS	89	[93]
	Secretoma	Estudio del impacto de el grado de esterificación de la pectina en las enzimas secretadas.	1-DE, LC MS/MS	131	[94]
	Secretoma	Análisis del secretoma usando medio mínimo suplementado con varias fuentes de carbono y elicitores vegetales.	2-DE, MALDI-TOF/TOF-MS/MS	56	[61]
	Secretoma	Análisis del secretoma temprano utilizando medio mínimo enriquecido con compuestos vegetales de bajo peso molecular (tomate, fresa o kiwi).	2-DE, MALDI-TOF MS and 1-DE, LC-MS/MS	105	[62]
	Secretoma	Estudio del efecto del pH sobre el secretoma.	2-DE, MALDI-TOF/TOF-MS/MS	21	[119]
	Secretoma	Estudio del efecto de los metales pesados en las proteínas secretadas.	2-DE, MALDI-TOF/TOF-MS/MS	116	[118]
	Micelio	Evaluación de la técnica de LC-MS^E sin marcaje para la identificación y cuantificación de proteínas de micelio de las cepas silvestres B05.10 and T4.	LC-MSE	264	[42]
	Micelio y secretoma	Análisis proteómico de micelio y del secretoma de seis cepas silvestres.	1-DE, 2-DE, MALDI-TOF/TOF-MS/MS, LC-MS^E	148	[33]
Curvularia lunata	Micelio	Análisis del proteoma de seis aislados diferentes para la relacionar el mapa proteico con la virulencia.	2-DE, MALDI-TOF/TOF-MS/MS	20	[67]
	Micelio	Análisis proteómico y de hibridación supresiva sustractiva para detectar los cambios en los niveles de proteínas y genes en *C. lunata* en la interacción con poblaciones resistentes de maíz.	2-DE, MALDI-TOF/TOF-MS/MS	16	[129]
Fusarium graminearum	Secretoma	Análisis del secretoma del hongo crecido en glucosa y en lúpulo de pared celular.	1-DE/2-DE, LC-MS/MS	84	[68]
	Secretoma	Análisis de proteínas secretadas por el hongo crecido in vitro (medios suplementados con diez fuentes de carbono diferentes) e in planta (espigas de trigo).	1-DE, LC-MS/MS	289	[130]
	Micelio	Estudio de las diferencias en el perfil de proteínas a través del tiempo producido mediante la inducción de la producción in vitro de tricotecenos.	2-DE, LC-MS/MS and iTRAQ	435	[91]
	Micelio	Estudio del efecto provocado por el virus FgV-DK21 en el desarrollo y la virulencia del hongo.	2-DE, LC-MS/MS	23	[131]
	Micelio	Análisis del fosfoproteoma.	SCX, IMAC, LC-MS/MS	1496	[132]
	Secretoma	Estudio del secretoma bajo condiciones de crecimiento que simulaban el ambiente nutricional in planta.	2-DE, MALDI-TOF/TOF-MS/MS	69	[133]
Fusarium oxysporum	Micelio	Estudio del efecto antagonista/patogénico de las cepas silvestres de MSA 35 de *F. oxysporum* en presencia de la planta hospedadora *Lactuca sativa*.	2-DE, MALDI-TOF/TOF-MS/MS	46	[134]
Leptosphaeria maculans	Secretoma	Optimización de los protocolos para la extracción, concentración y resolución de proteínas secretadas por el hongo.	1-DE, liquid-phase IEF, 2-DE		[135]
Magnaporthe oryzae	Esporas	Estudio de los efectos producidos por la delección de un regulador transcripcional que codifica un gen que regula la morfología de las esporas (COM1) en la forma de las esporas, la esporulación y la virulencia.	2-DE, MALDI-TOF MS	31	[56]
	Secretoma	Análisis del secretoma en respuesta a la privación de nitrógeno.	2-DE, MALDI-TOF MS, nESI-LC-MS/MS	59	[136]
	Micelio	Análisis del perfil de proteínas ubiquitiniladas bajo condiciones de estrés por falta de nitrógeno.	nLC-MS/MS	63	[137]
	Secretoma	Análisis del secretome en diferentes sistemas in vitro.	2-DE, nESI-LC-MS/MS	53	[138]
Phakopsora pachyrhyzi	Urediniosporas	Estudio de los eventos tempranos en el proceso de infección.	2-DE, MALDI-TOF/TOF MS/MS	117	[139]
	Apresorios, tubos germinados y urediniosporas germinadas	Estudio de los procesos que ocurren durante el desarrollo del apresorio.	2-DE, MALDI-TOF/TOF MS/MS	256	[140]
Phytophthora infestans	Secretoma	Estudio de las proteínas efectoras secretadas.	2-DE, MALDI-TOF-MS	9	[141]
	Quistes, quistes germinados y apresorios	Estudio de tres estados del desarrollo asexual.	2-DE, LC-MS/MS	13	[36]
Phytophthora palmivora	Micelios, esporangios, zoosporas, quistes y quistes germinados	Estudio del ciclo de desarrollo asexual.	2-DE, MALDI-TOF-MS	3	[58]
Phytophthora ramorum	Pared celular	Estudio de las proteínas secretadas y de las asociadas a la pared celular durante la infección de la planta.	LC-MS/MS	17	[142]
Puccinia triticina	Haustorios	Análisis del proteoma de los haustorios crecidos *in planta*.	2-DE, LC-MS/MS, GeLC-MS	260	[154]
Pyrenophora tritici-repentis	Micelio y secretoma	Evaluación de los proteomas de aislados avirulentos y virulentos del hongo.	2-DE, LC-MS/MS	63	[143]
Pleorotus ostreatus	Secretoma	Identificación de la proteína responsable de la actividad de la 17β-hidroxiesteroide deshidrogenasa.	1-DE, ^{18}O labelling, MALDI-TOF-MS, *de novo sequencing*-MS/MS	5	[144]
Rhizoctonia solani	Micelio	Optimización de los protocolos de extracción de proteínas para 2-DE.	2-DE, MALDI-TOF-MS	11	[145]
Sclerotinia sclerotiorum	Micelio y secretoma	Identificación de proteínas a gran escala.	2-DE, LC-MS/MS	113	[146]
	Esclerocios	Estudios de tres estadios del desarrollo de los esclerocios.	2-DE, LC-MS/MS	64	[147]
	Exudados de los esclerocios	Estudio de las proteínas secretadas en el exhudado de los esclerocios.	2-DE, LC-MS/MS	56	[148]
Serpula lacrymans	Secretoma	Análisis de genómica comparativa y funcional (transcriptoma y proteoma) para el estudio para el estudio de la maquinaria de descomposición de la madera, comparando varios Agaricomycetes.	LC-MS/MS	39	[149]
Stagnospora nodorum	Micelio	Identificación de dianas de las rutas de señalización analizando el proteoma de mutantes defectuosos en las proteínas G- alfa.	2-DE, LC-MS/MS	7	[150]
	Micelio	Análisis proteómico de mutantes defectuosos en las proteínas G-alfa.	iTRAQ, 2-D LC-MALDI-MS/MS	590	[151]
Thielaviopsis basicola	Micelio	Estudio del proteoma del hongo crecido in vitro en presencia de un extracto de raíces de plantas no hospedadoras (trigo, veza) y hospedadores susceptibles (algodón, lupinos).	2-DE, LC-MS/MS	41	[152]
	Micelio	Estudios genómicos, proteómicos y de virulencia de varios aislados silvestres.	2-DE, LC-MS/MS	43	[37]
Uromyces appendiculatus	Uredosporas	Estudio del proteoma de las uredosporas axesuales germinadas.	MudPIT	461	[74]
Ustilago maydis	Micelio	Estudio del proteoma del micelio.	2-DE, MALDI-TOF-MS, LC-MS/MS	250	[38]
Verticillium dahliae	Micelio	Comparación de los proteomas de aislados altamente (Vd1396-9) y ligeramente (Vs06-14) agresivos.	2-DE, LC-MS/MS	18	[153]

Existen excelentes revisiones que describen y analizan las características, protocolos y problemas inherentes a la 2-DE [45, 69] y en el capítulo 4 de este libro. Las limitaciones de la 2-DE han conducido al desarrollo de técnicas separación de proteínas o péptidos basadas en LC, como la cromatografía bidimensional en fase líquida (basada en un cromatoenfoque de alto

rendimiento en la primera dimensión seguida de una cromatografía en fase inversa de alta resolución en la segunda) [70], y la cromatografía nanocapilar en una dimensión llamada GeLC-MS/MS (que combina una separación de proteínas basada en el tamaño y una digestión en gel de las fracciones resultantes) [71].

A pesar de la existencia de diferentes métodos para la extracción y separación de proteínas, está claro que estos no son suficientes para permitir el análisis del proteoma completo. Algunos métodos son más sólidos y concluyentes que otros, en relación con el número de proteínas identificadas. Este es el caso de la Tecnología de Identificación de Proteínas Multidimensional (MudPIT, '*Multidimensional Protein Identification Technology*'), una estrategia basada en la separación mediante LC sin utilizar geles (llamado en inglés '*gel-free*' u '*off-gel strategy*'), lo que permite la identificación de un número mucho mayor de proteínas en comparación con los métodos basados en gel, aunque su desventaja es la falta de datos cuantitativos [24, 72, 73]. Esta tecnología es capaz de detectar e identificar proteínas rara vez vistas en el análisis basado en gel del proteoma, como proteínas de baja abundancia, factores de transcripción o proteín quinasas [24]. MudPIT se utilizó para analizar los mecanismos de germinación en *Uromyces appendiculatus* mediante la comparación de esporas asexuales germinadas y uredosporas inactivas [74]. En el capítulo 3 de este libro, se realiza una excelente revisión de la separación de proteías mediante métodos cromatográficos.

2.5. Análisis por espectrometría de masas e identificación de proteínas.

La MS es la técnica básica para el análisis proteómico debido a su precisión, resolución y sensibilidad (en el rango de concentración de femtomoles a attomoles), y a su alto rendimiento. La MS no sólo permite realizar un perfil proteico desde un punto de vista cualitativo y cuantitativo, sino también, y más importante, la identificación de las especies proteicas y la caracterización de modificaciones postraduccionales. Las proteínas pueden ser identificadas a partir de espectros de masas de proteínas intactas (en inglés '*top-down proteomics*') o de fragmentos peptídicos obtenidos tras una digestión enzimática (péptidos trípticos en su mayoría) o química (en inglés '*bottom-up proteomics*'). Las especies proteicas se identifican por comparación de los espectros experimentales y los teóricos obtenidos *in silico*, a través de la genómica, o bases de datos de secuencias de ESTs (en inglés '*Expressed Sequence Tags*') o de espectros de masas. A tal efecto, se encuentran disponibles diferentes instrumentos, algoritmos, bases de datos y repositorios [75, 76]. Para ver una revisión sobre la MS, existe un apartado en este libro sobre esta técnica.

Diferentes estrategias y algoritmos pueden ser utilizados para la identificación de proteínas: i) huella peptídica (PMF), que generalmente se utiliza si el genoma del organismo de estudio está secuenciado en su totalidad y está presente en las bases de datos (véase capítulo 17 de este libro), ii) espectrometría de masas en tándem (MS/MS), donde las secuencias de los péptidos se identifican por la correlación de espectros adquiridos de fragmentos de iones con los espectros teóricos previstos para cada péptido contenido en una base de datos de secuencias de proteínas, o mediante la correlación de espectros adquiridos de fragmentos de iones con las bibliotecas de espectros MS/MS construídas experimentalmente identificados en experimentos previos (véase capítulo 18 de este libro), iii) secuenciación *de novo*, donde las secuencias de péptidos son explícitamente leídos directamente de los espectros de iones fragmentados (véase capítulo 19 de este libro), y iv) enfoques híbridos, tales como los basados en la extracción de etiquetas de secuencias cortas de 3-5 residuos de longitud, seguido una búsqueda en la base de datos de "error-tolerancia" [76].

En una estrategia de proteómica basada en datos genómicos, el porcentaje de proteínas identificadas a partir de datos de MS depende de la disponibilidad de secuencias de ADN genómico o de ESTs. La construcción de repositorios de proteínas con una firma peptídica y derivados de espectros de MS abre nuevas posibilidades a la identificación de proteínas. La disponibilidad de se-

cuencias ESTs de organismos no secuenciados o "huérfanos", como es el caso de la mayoría de los hongos patógenos de plantas, incrementaría el porcentaje de las proteínas identificadas [77].

2.6. Cuantificación de las proteínas.

El uso de la proteómica, no sólo tiene como objetivo identificar las especies proteicas (principal objetivo de la proteómica descriptiva), sino también su cuantificación, al menos en términos relativos. La proteómica cuantitativa basada en MS es capaz de detectar la gran mayoría de las proteínas en las muestras, y de cuantificar una fracción significativa de las proteínas identificadas [78]. Además, se puede detectar cualquier cambio en la masa o en la secuencia de aminoácidos de una proteína (por ejemplo, sustituciones de aminoácidos o PTMs). La cuantificación absoluta de proteínas, más que la relativa, se mantiene como uno de los principales retos de la proteómica. Existen diferentes métodos para analizar el proteoma de una manera cuantitativa (para revisiones ver [79-81]): los métodos basados en gel y los basados en MS. Los métodos basados en gel usan principalmente geles 2-DE, seguida de una tinción, generalmente basada en CBB, en tinción fluorescente o en un pre-etiquetado, como es el caso de la electroforesis bidimensional diferencial cuantitativa con tinción fluorescente (2-DIGE) [29, 45, 69, 82]. La cuantificación mediante 2-DE presenta una alta sensibilidad, poder de resolución y capacidad de discriminar isoformas proteicas y proteínas con modificaciones post-traduccionales. La 2-DIGE aumenta el potencial de la 2-DE, ya que al ser etiquetadas las diferentes muestras con fluoróforos específicos éstas se pueden analizar en el mismo gel. Sin embargo, las técnicas de tinción son laboriosas, de reproducibilidad limitada y semi-cuantitativas. Por otra parte, la 2-DE presenta un rendimiento deficiente de cuantificación de proteínas con características particulares anteriormente comentadas, como los pesos moleculares extremos (altos o bajos), hidrofobicidad elevada, poca abundancia o p*I* extremos (ácidos o básicos).

Los métodos de cuantificación basados en MS pueden aplicarse tanto para estimar la abundancia absoluta o relativa de los niveles de proteína [78, 83, 84]. Además pueden clasificarse en aquellos métodos que usan marcaje de las proteínas o péptidos mediante el uso de isótopos estables (en inglés '*labelling-based methods*') y los que no (en inglés '*label-free methods'*). Así para el caso de los métodos de cuantificación basados en marcaje, la cuantificación absoluta utiliza un estándar externo para cuantificar un pequeño subconjunto de proteínas etiquetadas sin importar el número de muestras biológicas analizadas. Ejemplos de métodos de cuantificación absoluta son PSAQ ('*Protein Standard Absolute Quantification*'), SISCAPA ('*Stable Isotope Standards and Capture by Anti-Peptide Antibodies*'), QconCAT ('*Quantification used concatenated proteotypic peptides*') o AQUA ('*Absolute Quantification*') [83, 85]. En la cuantificación relativa, todas las proteínas están etiquetadas dentro de una única muestra biológica, y se compara con una muestra control con las proteínas sin etiquetar. Estas etiquetas se puede incorporar tanto con procedimientos *in vitro*, por el ejemplo los métodos ICAT ('*Isotope-Coded Affinity Tag*'), el etiquetado con ^{18}O, o iTRAQ ('*isobaric Tags for Relative and Absolute Quantification*'), como *in vivo*, por ejemplo los métodos SILAC ('Stable Isotope Labeling with Amino acids in cell Cultures'), o el etiquetado con ^{15}N [83]. Un enfoque alternativo es el llamado monitoreo seleccionado o de reacción múltiple (SRM o MRM, en inglés '*Selected or Multiple Reaction Monitoring*') que ha surgido precisamente para analizar cuantitativamente muestras biológicas complejas, como complemento de la proteómica basada en LC-MS/MS, debido a que no es óptima para una cuantificación sistemática de las proteínas identificadas por el carácter estocástico y la limitada sensibilidad del método, y sólo se realiza la cuantificación relativa de los componentes más abundantes [78, 86-88]. Para mayor información sobre la proteómica cuantitativa, se han incluído varios capítulos en este libro.

En la actualidad, los enfoques basados en MS sin marcaje ('*gel-free/label-free approaches*') para la cuantificación de proteínas son bastante populares debido a la mayor resolución de los instrumentos utilizados para la identificación (como el espectrómetro de masas Orbitrap) y el desarrollo de nuevos algoritmos para la cuantificación, apoyado por un riguroso diseño de investigación y validación estadística [84, 89, 90]. La cuantificación de proteínas por este método

se basa generalmente en dos categorías: en las mediciones de los cambios de intensidad de iones, tales como áreas de los picos de péptidos o alturas de los picos en la cromatografía, y en base al conteo de los espectros generados de las proteínas identificadas en el análisis MS/MS (en inglés '*spectral counting*'). La cuantificación basada en MS sin marcaje presenta una menor precisión (incertidumbre > 30%) en comparación con los métodos basados en marcaje. Sin embargo, la cuantificación tiene un rango dinámico mayor y, en general, es una metodología mucho más sencilla y más barata [83].

A pesar de que todas estas tecnologías tienen un gran potencial, su uso para el estudio de hongos fitopatógenos ha sido hasta ahora muy limitado. Taylor *et al.* utilizó iTRAQ para identificar las proteínas y cuantificar su abundancia relativa en *F. graminearum* cuando el micelio produce micotoxinas, imitando la infección de plantas en condiciones *in vitro* que inducen tricotecenos [91]. El conteo de espectros se ha utilizado en varios trabajos para lograr la cuantificación relativa de proteínas en hongos. Cooper *et al.* realizaron una semi-cuantificación de 461 proteínas a un nivel de confianza de más del 95% en *U. appendiculatus* comparando esporas germinadas e inactivas [74]. En nuestro grupo, se ha utilizado la técnica de separación de péptidos trípticos sin marcaje por LC para la identificación y cuantificación de proteínas utilizando un análisis LC-MS^E (basado en la fragmentación paralela de los péptidos trípticos y en la identificación y cuantificación simultánea de las proteínas) mediante las mediciones de los cambios de intensidad de los iones, en muestras de micelio y proteínas secretadas de *B. cinerea* [33, 42]. Esta metodología se basa en la medida de la intensidad de los iones precursores y toma las intensidades de iones de precursor de los tres péptidos más eficiente ionizados de una proteína con el fin de calcular una medida para su abundancia [92]. Estas abundancias se calculan por la suma de las intensidades de los tres péptidos más abundantes para cada proteína. En primer lugar, se realizó la evaluación y optimización de la cuantificación de las proteínas analizando diferentes diluciones de un mismo extracto de proteína de micelio (1500, 750, 300, 150 y 75 ng de proteína total inyectada en la columna y que corresponden a las diluciones 1 a 5 respectivamente), para observar el comportamiento de la cantidad relativa de las proteínas cuantificadas y ver si se mantiene la proporcionalidad entre las diluciones. La cuantificación de proteínas se realizó utilizando un estándar de enolasa (100 fmol por muestra/carrera) [92]. Un total de 51 proteínas se cuantificaron considerando que estuvieran presentes en las tres réplicas, en las cinco diluciones analizadas, con al menos tres péptidos detectados y con una CV < 25%. La misma tendencia de proporcionalidad se observó en los factores de dilución obtenidos a partir de ratios promedios de cantidad de proteína para cada proteína y cada dilución, aunque se observó una subestimación en la cuantificación causada por la sobrestimación de la dilución 1 [42]. Además, se evaluó la linearidad de la cuantificación examinando la suma de las intensidades de los tres péptidos más intensos de siete proteínas cuantificadas desde la más a menos abundante. Así, la respuesta de los péptidos fue lineal hasta la dilución 2 (correspondiente a 750 ng de proteína total inyectada en la columna) [42]. La inspección de los datos reveló que la respuesta de los péptidos no aumentó a la misma proporcionalidad a partir de 750 ng de cantidad total de proteínas, no teniendo entonces una respuesta lineal. La anchura y altura del pico cromatográfico no aumentó más a partir de cierta concentración de péptidos, muy probablemente debido a la saturación del detector del espectrómetro por la supresión iónica [42]. Por tanto, es necesario la evaluación de la concentración de proteína óptima inyectada en la columna para obtener una cuantificación más precisa. Además, se compararon dos cepas silvestres diferentes de *B. cinerea* (B05.10 y T4) con el fin de detectar variaciones en la abundancia de proteínas entre ellas [42]. En el caso de las muestras de micelio, se identificaron un total de 226 proteínas en B05.10 y 172 en la T4, de las cuales se cuantificaron 121 proteínas [42]. Los resultados obtenidos serán discutidos en el apartado 3.1. En el caso de las proteínas secretadas, se identificaron un total de 78 proteínas diferentes, de las cuales se cuantificaron 34 especies proteicas. Los resultados obtenidos serán discutidos en el apartado 3.2. [33]. Para más información sobre la técnica MS^E véase el capítulo 20 en este libro.

Otra forma de analizar las diferencias en la abundancia de proteínas es la combinación de 1-DE y LC-MS/MS basado en el conteo de espectros con el fin de semi-cuantificar las proteínas. Esta metodología se utilizó para estudiar los cambios en el secretoma de *B. cinerea* bajo diferentes condiciones de crecimiento [62, 93, 94].

2.7. Análisis de datos y validación estadística.

Las herramientas proteómicas generan una importante cantidad de datos, ya que un sólo experimento genera información de cientos o miles de proteínas. Por tanto, el análisis de los datos y las herramientas bioinformáticas son esenciales para este tipo de investigación, y en la mayoría de los casos lleva mucho tiempo su aprendizaje y requiere de especial manejo (ver revisiones [29, 30, 95]). Los programas de análisis para 2-DE permiten la comparación de múltiples geles de forma fiable y más o menos rápida, e incluyen el filtrado de las imágenes, la detección automáticas de las manchas proteicas, la normalización del volumen de cada mancha de proteína, y análisis diferenciales y estadísticos [96-98].

Otro reto en la proteómica es determinar si las identificaciones encontradas son correctas. Las herramientas estadísticas nos ayudan a validar la información. Aunque la validación estadística tras la identificación no se aplica de forma universal aún, su importancia ha sido reconocida por la mayoría de los investigadores e incluída en las políticas editoriales de algunas revistas líderes. En la actualidad existen un amplio número de paquetes comerciales con potentes algoritmos y herramientas estadísticas diseñados para ayudar a manejar la enorme cantidad de datos producidos en las investigaciones [99-101]. Ejemplos de test estadísticos son el análisis ANOVA, el FDR (en inglés '*False Discovery Rate*') o el PCA (en inglés '*Principle Component Analysis*') [29]. El análisis de datos y la validación estadística de describe ampliamente en el capítulo 5 de este libro.

3. *Botrytis cinerea:* el hongo fitopatógeno más estudiado mediante un enfoque proteómico.

3.1. Micelio.

B. cinerea Pers. Fr. (teleomorph *Botryotinia fuckeliana* (de Bary) Whetzel) es un importante patógeno necrotrófico de las plantas (tanto de vegetales, flores ornamentales de vivero, cultivos de campo y de huerta, productos agrícolas almacenados y transportados), que causa importantes pérdidas económicas en la agricultura [102]. En los últimos años, este hongo se ha convertido en un organismo modelo para estudiar los hongos fitopatógenos. El primer estudio proteómico de esta especie se publicó en 2006 y describía el proteoma parcial a partir de proteínas de micelio [53]. Un total de 400 manchas proteicas fueron detectadas en geles 2-DE, cubriendo un rango de p*I* de 5 a 8 y de peso molecular de 14 a 85 kDa. Se identificaron 22 especies proteicas de un total de 60 mediante el uso del análisis MALDI-TOF (en inglés '*Matrix-Assisted Laser Desorption/Ionization-Time-Of-Flight*') MS y ESI (en inglés '*electrospray ionization*') IT (en inglés '*ion trap*') MS/MS. Algunas de las proteínas identificadas fueron la malato deshidrogenasa (MDH), la gliceraldehído-3-fosfato deshidrogenasa (GADPH) y una ciclofilina, las cuales han sido relacionadas con la virulencia [53]. En un segundo estudio, el análisis proteómico comparativo del micelio de dos cepas silvestres de *B. cinerea*, que diferían en virulencia y en la producción de las toxinas, reveló la existencia de 73 diferencias cualitativas y cuantitativas en el mapa de proteínas 2-DE. Un total de 28 manchas proteicas fueron identificadas mediante el análisis MALDI-TOF MS y ESI IT MS/MS. Se identificaron las mismas proteínas mencionadas en el primer estudio, y aparecieron más abundantes o sólo en la cepa más virulenta [52]. Diecisiete manchas proteicas fueron identificadas como isoformas de la MDH y apareciendo más abundantes

en la cepa más virulenta. Esta enzima cataliza la conversión reversible entre el oxalacetato y el malato. El oxalacetato es un precursor de ácido oxálico que ha sido descrito como un factor de patogenicidad en *B. cinerea* [52]. Cuatro manchas proteicas fueron identificadas como isoformas de la GADPH. Además de su conocido papel en la glucolisis, la GAPDH participa en muchos otros sistemas y procesos celulares, y puede actuar como un factor de virulencia en diferentes organismos [52]. Las cuatro manchas proteicas identificadas en este estudio como GAPDH estuvieron presentes sólo en la cepa más virulenta, apoyando la hipótesis de que el metabolismo oxidativo en esta cepa pudiera ser más activo que el de la cepa menos virulenta, y sugiere un papel como potencial factor de virulencia de *B. cinerea* [52]. Una ciclofilina citosólica presente en la cepa más patogénica, no se detectó en el proteoma de la menos patogénica. Estas proteínas están involucradas en diferentes procesos celulares, incluyendo su papel como factor de virulencia en hongos patógenos de humanos y de plantas [52]. En estudios previos se propuso a la ciclofilina como factor de virulencia en las etapas posteriores de la infección, tales como la penetración y colonización de la planta en *B. cinerea* [103] y *Magnaporthe grisea* [104]. Un tercer trabajo más exhaustivo trató de establecer un mapa proteómico del micelio de *B. cinerea* durante la degradation de la celulosa [32]. Utilizando 2-DE combinada con el análisis MALDI-TOF/TOF MS/MS, se identificaron 306 proteínas. Los autores concluyeron que, dado que la celulosa es uno de los principales componentes de la pared celular vegetal, muchas de las proteínas identificadas pueden tener un papel crucial en el proceso de patogenicidad y estar involucradas en el ciclo de infección, por lo que pueden ser posibles candidatos para la elaboración de nuevos compuestos antifúngicos. Debido a la falta de genomas de hongos secuenciados en ese momento, el número de proteínas identificadas en estos estudios fue limitada, y principalmente se mostraban como proteínas hipotéticas o no anotadas [32].

Recientemente, nuestro grupo realizó un análisis del proteoma de micelio y del secretoma de seis cepas silvestres de *B. cinerea* aisladas de diferentes hospedadores (Tabla 1) utilizando 1- y 2-DE combinado con MALDI-TOF/TOF MS/MS [33], y la identificación y cuantificación de proteínas de micelio y del secretoma de las cepas B05.10 y T4 mediante un análisis LC-MS^E con péptidos trípticos no marcados [33, 42]. Mediante 1-DE, se comprobó la variabilidad del perfil proteico (variabilidad fenotípica) tanto del micelio como del secretoma entre las seis cepas (Figura 3). Como se describió anteriormente, el análisis 2-DE de las proteínas extraídas del micelio mostró 150 diferencias cualitativas y 19 diferencias cuantitativas estadísticamente significativas (*FDR* ≤ 0,05). Se identificaron 45 manchas proteicas (36 diferencias cualitativas y 9 cuantitativas estadísticamente significativas) que correspondieron a 40 especies proteicas únicas. Estas proteínas estuvieron relacionadas principalmente a procesos redox (18-35%), al metabolismo de carbohidratos y energía (11-19%) y al metabolismo de aminoácidos (8-13%). Además un número alto de manchas proteínas se identificaron como proteínas con función desconocida aunque con dominios conservados conocidos (13-21%).

Mediante el análisis LC-MS^E con péptidos trípticos no marcados de las muestras de micelio de las cepas B05.10 y T4, se identificaron un total de 226 proteínas en B05.10 y 172 en la T4, de las cuales 49 de ellas fueron detectados sólo en B05.10 y 7 en la T4 (Figura 4A). Por otra parte, un total de 121 proteínas fueron cuantificadas, de las cuales 49 tuvieron una relación (B05.10/T4) ≥ 2 y 7 una relación de (B05.10/T4) ≤ 0,5. Las dos principales categorías de las proteínas identificadas se relacionaron con la síntesis, ensamblaje y degradación de proteínas, y con el metabolismo de carbohidratos y de energía [42]. Una de las proteínas identificadas con diferencias cuantitativas entre las diferentes cepas fue la MDH identificada también en los estudios previos [52, 53]). Como se ha comentado anteriormente, el ácido oxálico se considera como un factor de patogenicidad en *B. cinerea*. Las funciones fisiológicas del ácido oxálico en la patogénesis son numerosas [18]: mejora la actividad de las poligalacturonasas para promo-

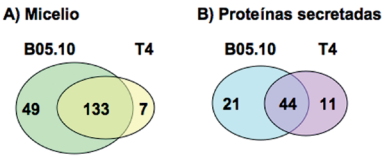


Figura 4. Diagramas de Venn para las proteínas identificadas en micelio (A) y en el secretoma (B) de las cepas B05.10 y T4 mediante el análisis LC-MS^E.

ver la degradación de la pared celular, inhibe las enzimas de protección de la planta, suprime el estallido oxidativo que generan las plantas, desregula el cierre de las células oclusivas de los estomas, media la señalización por pH, induce la muerte celular mediada por apoptosis y altera el estado redox celular en la planta. Cada una de estas actividades por separado puede facilitar la infección. Las funciones biológicas de ácido oxálico se basan en un equilibrio dinámico de la producción y degradación del mismo, y no simplemente en un sistema de síntesis máxima y de acumulación. La secreción de ácido oxálico crea un ambiente ácido adecuado para la expresión y secreción de algunos factores de virulencia tales como las enzimas que degradan la pared celular vegetal, de peptidasas, y de biosíntesis y secreción de fitotoxinas como el botrydial y el dihydrobotrydial que se secretan exclusivamente en un ambiente ácido [33]. Curiosamente, el nivel de la MDH fue mayor en las cepas 2100 y 2996, que son cepas aisladas de tejido verde (hojas de *Vicia faba* y plantas de *Phaseolus vulgaris*, respectivamente) cuyo pH > 6, y haciendo que estas cepas necesiten un flujo mayor de la producción de ácido oxálico para adquirir la adecuada medio ambiente para secretar las enzimas que degradan la pared celular, las peptidasas y las micotoxinas [33]. Otra de las enzimas identificadas fue la manitol deshidrogenasa (MTDH). Esta enzima fue identificada en dos manchas proteicas y detectada en todas las cepas estudiadas con diferencias cuantitativas estadísticamente significativas, excepto en la 2996 donde no se detectó uno de las manchas proteicas y en la otra mancha fue la menos abundante. En Ascomycetes, esta enzima puede producir manitol a través de la fructosa-6-P y la fructosa en el ciclo de manitol [105]. El manitol puede tener diferentes funciones en los hongos las cuales incluyen la osmorregulación, como mecanismo de almacenamiento de carbohidratos, como una fuente de poder reductor, como regulador del pH citoplasmático, como sumidero o fuente de protones y como quelatador de especies reactivas de oxígeno tanto *in vitro* como *in vivo* [106]. En estudios previos, se comprobó que el manitol se encuentra en grandes cantidades durante el desarrollo de micelio en *B. cinerea* [107]. Un posible papel fisiológico importante de la MTDH podría ser la conversión favorable de fructosa en manitol y, a la inversa, la degradación de manitol durante la respuesta a estrés osmótico [105]. Por otra parte, existe una posible interconexión del metabolismo del manitol con el metabolismo central de los carbohidratos. Por tanto, estos procesos podrían regular el flujo metabólico de carbono, así como estar involucrados en la respuesta al estrés osmótico. La degradación del manitol durante cuando el hongo se somete a estrés osmótico podría ser una respuesta para la osmoadaptación. En este caso, la MTDH se acumula, y se sugirió que podría estar dedicada preferentemente a la degradación manitol [105]. Por tanto, el manitol puede constituir un almacén de carbono fácilmente movilizable, no sólo utilizado para el crecimiento y diseminación del hongo, sino también para hacer frente al estrés osmótico y así mantener la presión de turgencia de las hifas. Otra especie proteica identificada con posible interés como factor de virulencia fue la cianida hidratasa que no se detectó en las cepas T4 y 2100 [33]. Este resultado fue previamente obtenido mediante LC-MS^E donde se detectó en la cepa B05.10 y no en la T4 [42]. La cianuro hidratasa cataliza la descomposición del cianuro de hidrógeno (HCN) en un compuesto menos tóxico: la formamida [108]. Esta enzima puede conferir tolerancia y ser un mecanismo de desintoxicación contra la generación de HCN por parte de la planta, ya que este compuesto es un mecanismo de defensa que generan las plantas cianogénicas para defenderse del ataque de los patógenos [109]. Estudios previos demostraron que *V. vinifera* y *P. vulgaris* son plantas cianogénicas [110, 111]. Sin embargo, no se conoce que se produzca cianogénesis en *Solanum lycopersum*, *V. fabae* o *Crocus sativus* [111, 112]. Existen varios genes de la cianuro hidratasa presentes en la secuencia del genoma de las cepas B05.10 y T4 con un 100% de homología. Este resultado confirmaría la versatilidad genética de *B. cinerea*, mediante el cual este hongo es capaz de adaptar su maquinaria de defensa a un hospedador particular, siendo en este caso las plantas cianogénicas [33]. Cabe destacar que entre las proteínas identificadas con función desconocida se encontró una esterasa en dos manchas proteicas muy abundantes excepto en las cepas 2100 y 2996. Esta proteína posee el péptido señal de secreción, y también fue detectada en el secretoma de las mismas cepas [33].

3.2. Proteínas secretadas: secretoma.

Una característica definitoria de los hongos fitopatógenos es la secreción de un gran número de enzimas de degradación y otras proteínas, que tienen diversas funciones en la adquisición de nutrientes, en la colonización del sustrato y en las interacciones ecológicas [113, 114]. El secretoma se ha definido como la combinación de las proteínas nativas secretadas y de la maquinaria celular involucrada en su secreción [115]. Varias enzimas fúngicas extracelulares, tales como la poligalacturonasa, la pectato liasa, la xilanasa y la lipasa, son necesarias para la virulencia en la interacción huésped-patógeno [116]. Mediante la proteómica, se puede estudiar la interacción entre las plantas y los patógenos mediada por moléculas secretadas e identificar proteínas efectoras fúngicas que facilitan la infección (factores de virulencia, micotoxinas) o respuestas de defensa de la planta desencadenadas por el ataque de los hongos (factores de avirulencia). A la luz de esto, y a diferencia de los animales, «los hongos digieren primero sus alimentos y después se la comen» [117], ilustrado por el gran número de enzimas hidrolíticas extracelulares necesarias para digerir una gran cantidad de sustratos potenciales. Por lo tanto, muchas de estas proteínas son de especial interés en el estudio de los hongos fitopatógenos y es un elemento clave en su estrategia de infección.

Hasta la fecha, se han publicado cinco estudios descriptivos del secretoma de *B. cinerea* para diseccionar su complejidad y versatilidad [33, 61, 62, 93, 94, 118, 119], en los que se identificaron 89, 126, 56, 105 y 76 proteínas, respectivamente. En el primer trabajo, se analizó el secretoma producido por el hongo en respuesta a una interacción simulada utilizando extractos de tomate y fresas maduras y hojas de *Arabidopsis* [93]. Se identificaron 89 proteínas de *B. cinerea* utilizando el análisis LC-MS/MS. Sesenta de estas proteínas contenían la secuencia péptido señal que indica la localización extracelular de las proteínas. Las proteínas identificadas correspondieron al metabolismo de los carbohidratos, a reacciones de oxidación/reducción, o fueron proteínas de transporte, peptidasas o factores de patogenicidad. *B. cinerea* secreta una amplia gama de enzimas que facilitan su penetración y colonización en el tejido de la planta, infectando una gran variedad de plantas hospedadoras. Se detectó cambios significativos en la abundancia relativa y la composición de las enzimas secretadas en función de la composición de los medios de cultivo empleados. En presencia de extractos de alimentos más favorables, tales como fresas y tomates maduros, el hongo aumentó las proteínas secretadas en comparación a cuando se usó extractos de hoja de *Arabidopsis*. Estos datos podrían proporcionar pistas importantes sobre cómo *B. cinerea* puede utilizar las proteínas secretadas para la infección y colonización de las plantas [93]. En el segundo trabajo, se estudió el impacto del grado de esterificación de la pectina sobre la secreción de enzimas en *B. cinerea*, debido a que en los cambios durante el proceso de maduración de las frutas parecen jugar un importante papel en la activación de la infección latente [94]. Mediante la combinación de 1-DE y LC-MS/MS, se identificaron 126 proteínas, de las cuales 87 proteínas se pronosticaron como secretadas, y algunas de las proteínas se identificaron como pectinasas [94]. Los resultados mostraron que el crecimiento de *B. cinerea* y la secreción de proteínas fueron similares en los cultivos con diferente esterificación de la pectina. Por lo tanto, es probable que la activación de este hongo desde el estado latente no es únicamente dependiente de cambios en el grado de esterificación de la pectina de la pared celular de la planta [94]. En un tercer trabajo, se indujo la secreción de las proteínas mediante la introducción controlada de varias fuentes de carbono y elicitores de origen vegetal (glucosa, almidón, pectina, carboximetilcelulosa (CMC) y extracto de pared celular de tomate (TCW)) en el medio de cultivo [61]. El uso de diferentes fuentes de carbono y de inductores de origen vegetal originó una respuesta del hongo con diferentes grados de complejidad, pasando de un estado constitutivo de crecimiento del hongo con un secretoma sencillo (mediante el uso de la glucosa como nutriente) hacia un comportamiento de secreción más complejo y posiblemente más patogénico (inducida por el extracto de TCW). Los correspondientes mapas 2-DE mostraron un aumento en el número y la variedad de manchas proteicas cuando se utilizaron la CMC o el extracto de TCW. La secreción de proteínas fue similar para los tratamientos con glucosa, almidón y pectina, y la mayoría de estas proteínas secretadas fueron enzimas de degradación de la pared celular. Sin embargo, se detectaron más proteínas con el tratamiento con CMC y aún más

con el extracto de TCW. Un total de 78 manchas proteicas fueron identificadas mediante el uso del análisis por MALDI-TOF/TOF MS/MS, que correspondieron a 56 proteínas únicas y a sus respectivas isoformas, y que representaron el 2,5 % de las proteínas secretadas predichas de *B. cinerea* [120]. Las secuencias de todas las proteínas identificadas contenían el péptido señal de secreción, tanto para las vías clásicas de secreción como para las no clásicas. Algunas de las proteínas identificadas están involucradas en el proceso de patogenicidad (por ejemplo, las pectin-metilesterasas, las xilanasas, o las proteasas) [19, 121].

Las proteínas secretadas en el comienzo de la infección, es decir, durante la germinación de las esporas en la superficie de las plantas o de heridas, son de especial interés, ya que juegan un papel esencial en el establecimiento de una infección exitosa. En el cuarto trabajo, esporas de *B. cinerea* se germinaron en condiciones que se asemejan al ambiente de la planta (esto es, en un medio sintético enriquecido con compuestos vegetales de bajo peso molecular, conteniendo un 50% de extractos de tomate, fresa, o kiwi, dentro de una bolsa de diálisis), y se colectaron las proteínas secretadas durante las primeras 16 h [62]. El mapa 2-DE del secretoma mostró un patrón de manchas similar tanto para todas las condiciones estudiadas como para el control sin extracto vegetal (con glucosa como fuente de carbono). De las 86 manchas proteicas cortadas, sólo 16 fueron identificadas por análisis MALDI-TOF MS correspondiendo a 11 proteínas únicas. Debido al bajo número de manchas proteicas identificadas, los extractos de proteínas se sometieron a 1-DE acoplado a LC-MS/MS y se identificaron 105 proteínas. Una conclusión interesante de estos datos fue que la composición del secretoma temprano no es tan variable como se esperaba, de modo que *B. cinerea* parece secretar durante su germinación un conjunto común de proteínas en todas las condiciones, que se completa con un menor número de proteínas específicas para ciertas condiciones [62].

Por último, nuestro grupo publicó recientemente la comparación de los secretomas de seis cepas silvestres de *B. cinerea* aisladas de diferentes plantas hospedadoras (Tabla 1), en condiciones de crecimiento mínimas (con sacarosa como fuente de carbono), mediante un análisis 1-DE combinado con MALDI-TOF/TOF MS/MS y de las cepas B05.10 y T4 por análisis LC-MS^E con péptidos trípticos sin marcaje [33]. El perfil proteico del secretoma de las diferentes cepas mostró las diferencias existentes entre las mismas y una complejidad relativamente baja (Figura 3). Mediante 1-DE, se identificaron 16 especies proteicas diferentes, conteniendo el péptido señal de secreción, y que pertenecían principalmente a las categorías de enzimas que degradan la pared celular, a proteolisis y a procesos redox. El perfil proteico 1-DE también mostró que en todas las cepas siempre se detectó un pequeño número de proteínas que son abundantemente secretadas. Así, para cada caso, de 2 a 6 proteínas constituyeron más del 50% de las proteínas totales extraídas. Debido a que el análisis por MALDI-TOF/TOF MS/MS sólo fue capaz de identificar las proteínas más abundantes, se utilizó una técnica más sensible como es el análisis por LC-MS^E sin marcaje para identificar y cuantificar las proteínas secretadas por las cepas B05.10 y T4 en las mismas condiciones descritas anteriormente [33]. En este caso, se identificaron 76 proteínas, conteniendo el péptido señal de secreción, 63 proteínas en la cepa B05.10 y 51 proteínas en la cepa T4, de las cuales 21 proteínas fueron únicas a B05.10 y 11 a T4 (Figura 4B). Las proteínas identificadas en ambas cepas pertenecieron principalmente a enzimas que degradan la pared celular (33% en B05.10 y 27% en T4), a proteolisis (19% en B05.10 y 20% en T4) y a reacciones redox (12% en B05.10 y 16% en T4). Cabe destacar que un notable porcentaje de las proteínas identificadas (27% en B05.10 y 33% en T4) tuvieron función desconocida aunque contaban con dominios conservados conocidos. Además, un total de 34 proteínas se cuantificaron. Se consideró el ratio entre B05.10/T4 (R) como $0,5 > R > 2$ con un t-Test con $p \leq 0,05$ y un CV $\leq$ 25%, y se encontraron 6 especies proteicas más abundantes y 12 menos abundantes en la cepa B05.10 con respecto a la cepa T4 [33]. Estas proteínas se relacionaron con la degradación de la pared celular vegetal como por ejemplo la glucoamilasa, la endo-poligalacturonasa, la pectinesterasa o peptidasas como la rhizopuspepsina-2.

Hay que destacar la gran cantidad de proteasas que se encuentran en el secretoma de *B. cinerea* [33, 62, 93]. Estas enzimas podría ser consideradas como potenciales factores fúngicos de virulencia, ya que, aparte de su evidente papel en la generación de aminoácidos para sostener el crecimiento de los hongos, las proteasas podrían tener también un papel importante en otros dos aspectos de su virulencia: podrían degradar (i) proteínas de la pared celular vegetal para promover la penetración de las hifas del hongo, y (ii) proteínas de defensa de las plantas para aumentar las posibilidades de una infección exitosa. La alta diversidad de las proteasas que se encuentran también puede ser reflejo de la diversidad de su sustrato. A diferencia de los polisacáridos, que tienen una estructura fija, las proteínas vegetales tienen una gran diversidad de estructuras por naturaleza, por lo que sería necesario un grupo diverso de proteasas para su degradación. Una de las proteínas secretadas más abundantes y que se detectó en todas las condiciones de cultivo en todos los estudios fue la aspártico proteasa BcAP8 o rhizopuspepsina-2 (BC1G_03070). El análisis proteómico del mutante *bc1g_03070* mostró algunas diferencias en el patrón de bandas/manchas proteicas observadas en los geles 1- y 2-DE [62]. La ausencia de esta proteasa en el secretoma causó la disminución de hasta un 90% en la actividad proteasa secretada. Por otra parte, causó un aumento en la intensidad de las bandas de proteínas de alto peso molecular en el perfil 1-DE y una disminución, o incluso la desaparición de bandas proteicas de bajo peso molecular [62]. Estos resultados sugieren que la abundancia de proteasas secretadas por *B. cinerea* causan la escisión significativa de las proteínas secretadas y, como consecuencia, provocan fragmentos más pequeños que, teniendo la secuencia de cobertura más pequeña, serían más difícil de identificar [62].

Aunque las cepas, y más importante, las condiciones de crecimiento del hongo, y la metodología utilizada para aislar las proteínas del secretoma fueron completamente diferentes, la disponibilidad de estos estudios nos da la oportunidad de comparar el secretoma de *B. cinerea* en condiciones muy diferentes. La comparación de estos datos parece dar a entender a primera vista que el secretoma de *B. cinerea* es altamente versátil y adaptativo, en el sentido de que se secretan diferentes conjuntos de proteínas dependiendo de las condiciones de crecimiento, la edad del micelio o las cepas utilizadas. Por lo tanto, *B. cinerea* puede alterar en gran medida la composición del grupo de proteínas secretadas para satisfacer los requisitos de las diferentes necesidades de crecimiento [21]. Así, por ejemplo, se han realizado unos trabajos recientes para estudiar el papel de las proteínas secretadas por *B. cinerea* con respecto a la adaptación al pH del ambiente [119] o a la toxicidad por metales pesados [118]. En el primer estudio, las proteínas relacionadas con la proteolisis se indujeron a pH 4 (similar al ambiente de la fruta madura), mientras que a pH 6 (similar al ambiente de la hoja) la mayoría de las proteínas más abundantes fueron las enzimas que degradan la pared celular [119]. Por lo tanto, el pH del ambiente podría regular la expresión de las proteínas secretadas en *B. cinerea*, mediante el cual el hongo activa la maquinaria necesaria para la invasión dependiendo el tejido y el hopedador. En el segundo trabajo, el hongo fue sometido a un estrés por cobre, zinc, níquel y cadmio [118]. De las 116 manchas proteicas variables se identificaron 55 proteínas únicas. El estudio reveló que la producción de oxidorreductasas y de enzimas que degradan la pared celular se modificó en respuesta a los metales [118]. Agrupando las proteínas identificadas en todos los estudios realizados en el secretoma, sólo se ha podido identificar aproximadamente el 10% de las proteínas secretadas predichas para *B. cinerea* [120] utilizando la metodología disponible en nuestros días.

En resumen, los resultados obtenidos por nuestro grupo en muestras de micelio y en proteínas secretadas de *B. cinerea* mediante el análisis por LC-MS^E concuerdan con los obtenidos mediante 1-DE y 2-DE, ya que la identificación y las cantidades de las proteínas se correlacionaron entre los dos enfoques [33]. Esto confirma que el uso de diferentes métodos proteómicos nos puede ayudar a validar los datos obtenidos y profundizar en la caracterización y comparación de los proteomas. Así, por ejemplo, un análisis proteómico que incluye técnicas basadas en gel (1- y 2-DE) y en LC sin marcaje se están utilizando por nuestro grupo para caracterizar y validar una colección de mutantes de *B. cinerea* cuyo ciclo de infección está afectado [122].

4. Perspectivas futuras.

En los últimos años se han realizado grandes avances en el campo de la proteómica de hongos, debido al incremento del número de genomas fúngicos disponibles y al desarrollo de protocolos de extracción para obtener extractos de proteínas de buena calidad. Sin embargo, la tremenda diversidad y flexibilidad en el genoma de los hongos hace que esta tarea sea difícil. Hasta la fecha, la mayoría de los estudios realizados en hongos fitopatógenos han estado limitados a los métodos basados en gel. La integración de los datos genómicos y proteómicos a gran escala proporcionaría la aclaración de los mecanismos de patogenicidad de los hongos patógenos de plantas y de la interacción con sus huéspedes. Ya que los hongos fitopatógenos causan importantes pérdidas económicas en numerosas cosechas, es necesario realizar estudios proteómicos a gran escala de estos organismos para identificar los factores involucrados en la patogenicidad. Así, el análisis proteómico se ha convertido en una excelente herramienta que está proporcionando gran cantidad de información sobre la patogenicidad de los hongos, permitiendo la identificación de nuevos factores de virulencia, la caracterización de rutas bioquímicas, de transducción de señales y de los mecanismos de infección, lo que nos permite comprender el ciclo y estilo de vida de estos organismos. A partir de esta información, se podrán descubrir nuevas dianas para la diagnosis de las enfermedades de las cosechas y diseñar nuevos fungicidas que sean respetuosos con el medio ambiente.

REFERENCIAS.

1. **Hawksworth DL, Kirk PM, Sutton BC, Pegler D.N. Ainsworth and Bisby's Dictionary of the Fungi.** Wallingford: PM Kirk, et al; 2001.

2. **Arora DK, Bridge PD and Bhatnagar D. Fungal biotechnology in agricultural, food and environmental applications.** New York: Marcel Dekker; 2004.

3. **Gonzalez-Fernandez R, Prats P and Jorrin-Novo JV. Proteomics of Plant Pathogenic Fungi.** J Biomed Biotechnol. 2010; vol. 2010: Article ID 932527, 36 pages. doi:10.1155/2010/.

4. **Kikot GE, Hours RA and Alconada TM. Contribution of cell wall degrading enzymes to pathogenesis of Fusarium graminearum: a review.** J Basic Microbiol. 2009; 49: 231-41.

5. **van Esse HP, Van't Klooster JW, Bolton MD, et al. The Cladosporium fulvum virulence protein Avr2 inhibits host proteases required for basal defense.** Plant Cell. 2008; 20: 1948-63.

6. **Friesen TL, Faris JD, Solomon PS, et al. Host-specific toxins: effectors of necrotrophic pathogenicity.** Cell Microbiol. 2008; 10: 1421-8.

7. **Gao X and Kolomiets MV. Host-derived lipids and oxylipins are crucial signals in modulating mycotoxin production by fungi.** Toxin Reviews. 2009; 28: 79-88.

8. **Kim KS, Min JY and Dickman MB. Oxalic acid is an elicitor of plant programmed cell death during Sclerotinia sclerotiorum disease development.** Mol Plant Microbe Interact. 2008; 21: 605-12.

9. **Lawrence CB, Mitchell TK, Craven KD, et al. At Death's Door: Alternaria Pathogenicity Mechanisms.** Plant Phatology Journal. 2008; 24: 101-11.

10. **McCouch S. Diversifying Selection in Plant Breeding.** PLoS Biol. 2004; 2: e347.

11. **Yang Y, Zhang H, Li G, et al. Ectopic expression of MgSM1, a Cerato-platanin family protein from Magnaporthe grisea, confers broad-spectrum disease resistance in Arabidopsis.** Plant Biotechnol J. 2009; 7: 763-77.

12. **Kim BS and Hwang BK. Microbial fungicides in the control of plant diseases Journal of Phytopathology.** 2007; 155: 641-53.

13. **Berg G. Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture.** Appl Microbiol Biotechnol. 2009; 84: 11-8.

14. **Van De Wouw AP and Howlett BJ. Fungal pathogenicity genes in the age of 'omics'.** Molecular Plant Pathology. 2011; 12: 507-14.

15. **Winnenburg R, Baldwin TK, Urban M, et al. PHI-base: a new database for pathogen host interactions.** Nucleic Acids Res. 2006; 34: D459-64.

16. **Winnenburg R, Urban M, Beacham A, et al. PHI-base update: additions to the pathogen host interaction database.** Nucleic Acids Res. 2008; 36: D572-6.

17. Afroz A, Ali GM, Mir A, et al. Application of proteomics to investigate stress-induced proteins for improvement in crop protection. Plant Cell Rep. 2011; 30: 745-63.

18. Amselem J, Cuomo CA, van Kan JA, et al. Genomic analysis of the necrotrophic fungal pathogens Sclerotinia sclerotiorum and Botrytis cinerea. PLoS Genet. 2011; 7: e1002230.

19. Choquer M, Fournier E, Kunz C, et al. Botrytis cinerea virulence factors: new insights into a necrotrophic and polyphageous pathogen. FEMS Microbiology Letters. 2007; 277: 1-10.

20. Egan MJ and Talbot NJ. Genomes, free radicals and plant cell invasion: recent developments in plant pathogenic fungi. Curr Opin Plant Biol. 2008; 11: 367-72.

21. Gonzalez-Fernandez R and Jorrin-Novo JV. Contribution of proteomics to the study of plant pathogenic fungi. J Proteome Res. 2012; 11: 3-16.

22. Hadwiger LA. Localization predictions for gene products involved in non-host resistance responses in a model plant/fungal pathogen interaction. Plant Science. 2009; 177: 257-65.

23. Staats M and van Kan JA. Genome Update of Botrytis cinerea Strains B05.10 and T4. Eukaryot Cell. 2012; 11: 1413-4.

24. Washburn MP, Wolters D and Yates JR, 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001; 19: 242-7.

25. Wolters DA, Washburn MP and Yates JR, 3rd. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001; 73: 5683-90.

26. Picotti P, Bodenmiller B, Mueller LN, et al. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. Cell. 2009; 138: 795-806.

27. Anderson NL, Anderson NG, Pearson TW, et al. A human proteome detection and quantitation project. Mol Cell Proteomics. 2009; 8: 883-6.

28. González-Fernández R and Jorrín-Novo JV. Proteomic protocols for the study of filamentous fungi. En: Gupta VK, Tuohy MG, Ayyachamy Met al (ed) Laboratory Protocols in Fungal Biology: Current Methods in Fungal Biology. New York, Heidelberg, Dordrecht, London: Springer, 2013, pp. 299-308.

29. Valledor L and Jorrin J. Back to the basics: Maximizing the information obtained by quantitative two dimensional gel electrophoresis analyses by an appropriate experimental design and statistical analyses. Journal of Proteomics. 2011; 74: 1-18.

30. Holzmuller P, Grebaut P, Brizard JP, et al. “Pathogeno-proteomics”. Ann N Y Acad Sci. 2008; 1149: 66-70.

31. Jorrin-Novo JV, Maldonado AM, Echevarria-Zomeno S, et al. Plant proteomics update (2007-2008): Second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge. J Proteomics. 2009; 72: 285-314.

32. Fernandez-Acero FJ, Colby T, Harzen A, et al. Proteomic analysis of the phytopathogenic fungus Botrytis cinerea during cellulose degradation. Proteomics. 2009; 9: 2892-902.

33. González-Fernández R, Aloria K, Valero-Galván J, et al. Proteomic analysis of mycelium and secretome of different Botrytis cinerea wild-type strains. Journal of Proteomics. 2014; 97:195-221.

34. Büttner P, Koch F, Voigt K, et al. Variations in ploidy among isolates of Botrytis cinerea: implications for genetic and molecular analyses. Current Genetics. 1994; 25: 445-50.

35. Ruiz-Herrera J. Fungal Cell Wall: Structure, Synthesis and Assembly. CRC Press; 1992.

36. Ebstrup T, Saalbach G and Egsgaard H. A proteomics study of in vitro cyst germination and appressoria formation in Phytophthora infestans. Proteomics. 2005; 5: 2839-48.

37. Coumans JV, Harvey J, Backhouse D, et al. Proteomic assessment of host-associated microevolution in the fungus Thielaviopsis basicola. Environ Microbiol. 2011; 13: 576-88.

38. Bohmer M, Colby T, Bohmer C, et al. Proteomic analysis of dimorphic transition in the phytopathogenic fungus Ustilago maydis. Proteomics. 2007; 7: 675-85.

39. Sulc M, Peslova K, Zabka M, et al. Biomarkers of Aspergillus spores: Strain typing and protein identification. International Journal of Mass Spectrometry. 2009; 280: 162-8.

40. Nandakumar MP and Marten MR. Comparison of lysis methods and preparation protocols for one- and two-dimensional electrophoresis of Aspergillus oryzae intracellular proteins. Electrophoresis. 2002; 23: 2216-22.

41. Shimizu M and Wariishi H. Development of a sample preparation method for fungal proteomics. FEMS Microbiol Lett. 2005; 247: 17-22.

42. Gonzalez-Fernandez R, Aloria K, Arizmendi JM, et al. Application of Label-Free Shotgun nUPLC-MSE and 2-DE Approaches in the Study of Botrytis cinerea Mycelium. Journal of Proteome Research. 2013; 12: 3042-56.

43. Gonzalez-Fernandez R, Redondo I and Jorrin-Novo JV. Making a Protein Extract from Plant Pathogenic Fungi for Gel- and LC-Based Proteomics. Methods Mol Biol. 2014; 1072: 93-109.

44. Nandakumar MP, Shen J, Raman B, et al. Solubilization of trichloroacetic acid (TCA) precipitated microbial proteins via NaOH for two-dimensional electrophoresis. J Proteome Res. 2003; 2: 89-93.

45. Gorg A, Weiss W and Dunn MJ. Current two-dimensional electrophoresis technology for proteomics. Proteomics. 2004; 4: 3665-85.

46. Rabilloud T. Use of thiourea to increase the solubility of membrane proteins in two-dimensional electrophoresis. Electrophoresis. 1998; 19: 758-60.

47. Kniemeyer O, Lessing F, Scheibner O, et al. Optimisation of a 2-D gel electrophoresis protocol for the human-pathogenic fungus Aspergillus fumigatus. Curr Genet. 2006; 49: 178-89.

48. Luche S, Santoni V and Rabilloud T. Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. Proteomics. 2003; 3: 249-53.

49. Rabilloud T. Solubilization of proteins for electrophoretic analyses. Electrophoresis. 1996; 17: 813-29.

50. Rabilloud T, Adessi C, Giraudel A, et al. Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis. 1997; 18: 307-16.

51. Herbert BR, Grinyer J, McCarthy JT, et al. Improved 2-DE of microorganisms after acidic extraction. Electrophoresis. 2006; 27: 1630-40.

52. Fernandez-Acero FJ, Jorge I, Calvo E, et al. Proteomic analysis of phytopathogenic fungus Botrytis cinerea as a potential tool for identifying pathogenicity factors, therapeutic targets and for basic research. Arch Microbiol. 2007; 187: 207-15.

53. Fernandez-Acero FJ, Jorge I, Calvo E, et al. Two-dimensional electrophoresis protein profile of the phytopathogenic fungus Botrytis cinerea. Proteomics. 2006; 6 Suppl 1: S88-96.

54. Grimplet J, Wheatley MD, Jouira HB, et al. Proteomic and selected metabolite analysis of grape berry tissues under well-watered and water-deficit stress conditions. Proteomics. 2009; 9: 2503-28.

55. Fragner D, Zomorrodi M, Kues U, et al. Optimized protocol for the 2-DE of extracellular proteins from higher basidiomycetes inhabiting lignocellulose. Electrophoresis. 2009; 30: 2431-41.

56. Bhadauria V, Wang LX and Peng YL. Proteomic changes associated with deletion of the Magnaporthe oryzae conidial morphology-regulating gene COM1. Biol Direct. 2010; 5: 61.

57. Noir S, Colby T, Harzen A, et al. A proteomic analysis of powdery mildew (Blumeria graminis f.sp. hordei) conidiospores. Mol Plant Pathol. 2009; 10: 223-36.

58. Shepherd SJ, van West P and Gow NA. Proteomic analysis of asexual development of Phytophthora palmivora. Mycol Res. 2003; 107: 395-400.

59. Kao SH, Wong HK, Chiang CY, et al. Evaluating the compatibility of three colorimetric protein assays for two-dimensional electrophoresis experiments. Proteomics. 2008; 8: 2178-84.

60. Medina ML and Francisco WA. Isolation and Enrichment of Secreted Proteins from Filamentous Fungi En: SpringerLink (Ed) 2D PAGE: Sample Preparation and Fractionation. Humana Press, 2008, pp. 275-85.

61. Fernandez-Acero FJ, Colby T, Harzen A, et al. 2-DE proteomic approach to the Botrytis cinerea secretome induced with different carbon sources and plant-based elicitors. Proteomics. 2010; 2010:

62. Espino JJ, Gutierrez-Sanchez G, Brito N, et al. The Botrytis cinerea early secretome. Proteomics. 2010; 10: 3020-34.

63. Supek F, Peharec P, Krsnik-Rasol M, et al. Enhanced analytical power of SDS-PAGE using machine learning algorithms. Proteomics. 2008; 8: 28-31.

64. Fryksdale BG, Jedrzejewski PT, Wong DL, et al. Impact of deglycosylation methods on two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-time of flight-mass spectrometry for proteomic analysis. Electrophoresis. 2002; 23: 2184-93.

65. Medina ML, Haynes PA, Breci L, et al. Analysis of secreted proteins from Aspergillus flavus. Proteomics. 2005; 5: 3153-61.

66. Medina ML, Kiernan UA and Francisco WA. Proteomic analysis of rutin-induced secreted proteins from Aspergillus flavus. Fungal Genet Biol. 2004; 41: 327-35.

67. Xu S, Chen J, Liu L, et al. Proteomics Associated with Virulence Differentiation of Curvularia lunata in Maize in China. Journal of Integrative Plant Biology 2007; 49: 487-196.

68. Phalip V, Delalande F, Carapito C, et al. Diversity of the exoproteome of Fusarium graminearum grown on plant cell wall. Curr Genet. 2005; 48: 366-79.

69. Rabilloud T, Vaezzadeh AR, Potier N, et al. Power and limitations of electrophoretic separations in proteomics strategies. Mass Spectrom Rev. 2009; 28: 816-43.

70. Pirondini A, Visioli G, Malcevschi A, et al. A 2-D liquid-phase chromatography for proteomic analysis in plant tissues. J Chromatogr B Analyt Technol Biomed Life Sci. 2006; 833: 91-100.

71. Recorbet G, Rogniaux H, Gianinazzi-Pearson V, et al. Fungal proteins in the extra-radical phase of arbuscular mycorrhiza: a shotgun proteomic picture. New Phytol. 2009; 181: 248-60.

72. McDonald T, Sheng S, Stanley B, et al. Expanding the subproteome of the inner mitochondria using protein separation technologies: one- and two-dimensional liquid chromatography and two-dimensional gel electrophoresis. Mol Cell Proteomics. 2006; 5: 2392-411.

73. Ye M, Jiang X, Feng S, et al. Advances in chromatographic techniques and methods in shotgun proteome analysis. TrAC Trends in Analytical Chemistry. 2007; 26: 80-4.

74. Cooper B, Neelam A, Campbell KB, et al. Protein accumulation in the germinating Uromyces appendiculatus uredospore. Mol Plant Microbe Interact. 2007; 20: 857-66.

75. Domon B and Aebersold R. Mass spectrometry and protein analysis. Science. 2006; 312: 212-7.

76. Nesvizhskii AI, Vitek O and Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat Methods. 2007; 4: 787-97.

77. Skinner W, Keon J and Hargreaves J. Gene information for fungal plant pathogens from expressed sequences. Current Opinion in Microbiology. 2001; 4: 381-6.

78. Malmström L, Malmström J and Aebersold R. Quantitative proteomics of microbes: Principles and applications to virulence. Proteomics. 2011; 11: 2947-56.

79. Elliott MH, Smith DS, Parker CE, et al. Current trends in quantitative proteomics. J Mass Spectrom. 2009; 44: 1637-60.

80. Han X, Aslanian A and Yates JR, 3rd. Mass spectrometry for proteomics. Curr Opin Chem Biol. 2008; 12: 483-90.

81. Zhu W, Smith JW and Huang CM. Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol. 2010; 2010: 840518.

82. Heinemeyer J, Scheibe B, Schmitz UK, et al. Blue native DIGE as a tool for comparative analyses of protein complexes. J Proteomics. 2009; 72: 539-44.

83. de Oliveira JM and de Graaff LH. Proteomics of industrial fungi: trends and insights for biotechnology. Appl Microbiol Biotechnol. 2011; 89: 225-37.

84. Porteus B, Kocharunchitt C, Nilsson RE, et al. Utility of gel-free, label-free shotgun proteomics approaches to investigate microorganisms. Appl Microbiol Biotechnol. 2011; 90: 407-16.

85. Brownridge P, Holman SW, Gaskell SJ, et al. Global absolute quantification of a proteome: Challenges in the deployment of a QconCAT strategy. Proteomics. 2011; 11: 2957-70.

86. Liebler DC and Zimmerman LJ. Targeted Quantitation of Proteins by Mass Spectrometry. Biochemistry. 2013; 52: 3797-806.

87. Holman SW, Sims PFG and Eyers CE. The use of selected reaction monitoring in quantitative proteomics. Bioanalysis. 2012; 4: 1763-86.

88. Gallien S, Duriez E and Domon B. Selected reaction monitoring applied to proteomics. J Mass Spectrom. 2011; 46: 298-312.

89. Nahnsen S, Bielow C, Reinert K, et al. Tools for label-free peptide quantification. Mol Cell Proteomics. 2013; 12: 549-56.

90. Hol apek M, Jirásko R and Lísa M. Recent developments in liquid chromatography-mass spectrometry and related techniques. Journal of Chromatography A. 2012; 1259: 3-15.

91. Taylor RD, Saparno A, Blackwell B, et al. Proteomic analyses of Fusarium graminearum grown under mycotoxin-inducing conditions. Proteomics. 2008; 8: 2256-65.

92. Silva JC, Gorenstein MV, Li GZ, et al. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics. 2006; 5: 144-56.

93. Shah P, Atwood JA, Orlando R, et al. Comparative proteomic analysis of Botrytis cinerea secretome. J Proteome Res. 2009; 8: 1123-30.

94. Shah P, Gutierrez-Sanchez G, Orlando R, et al. A proteomic study of pectin-degrading enzymes secreted by Botrytis cinerea grown in liquid culture. Proteomics. 2009; 9: 3126-35.

95. Deutsch EW, Lam H and Aebersold R. Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. Physiol Genomics. 2008; 33: 18-25.

96. Barrett J, Brophy PM and Hamilton JV. Analysing proteomic data. Int J Parasitol. 2005; 35: 543-53.

97. Marengo E, Robotti E, Antonucci F, et al. Numerical approaches for quantitative analysis of two-dimensional maps: a review of commercial software and home-made systems. Proteomics. 2005; 5: 654-66.

98. Wheelock AM and Goto S. Effects of post-electrophoretic analysis on variance in gel-based proteomics. Expert Rev Proteomics. 2006; 3: 129-42.

99. Bantscheff M, Schirle M, Sweetman G, et al. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem. 2007; 389: 1017-31.

100. Biron DG, Brun C, Lefevre T, et al. The pitfalls of proteomics experiments without the correct use of bioinformatics tools. Proteomics. 2006; 6: 5577-96.

101. Mueller LN, Brusniak MY, Mani DR, et al. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. J Proteome Res. 2008; 7: 51-61.

102. Elad Y, Williamson B, Tudzynski P, et al. Botrytis: biology, pathology and control. Dordrecht: Springer; 2007.

103. Viaud M, Brunet-Simon A, Brygoo Y, et al. Cyclophilin A and calcineurin functions investigated by gene inactivation, cyclosporin A inhibition and cDNA arrays approaches in the phytopathogenic fungus Botrytis cinerea. Mol Microbiol. 2003; 50: 1451-65.

104. Viaud MC, Balhadère PV and Talbot NJ. A Magnaporthe grisea Cyclophilin Acts as a Virulence Determinant during Plant Infection. The Plant Cell Online. 2002; 14: 917-30.

105. Dulermo T, Rascle C, Billon-Grand G, et al. Novel insights into mannitol metabolism in the fungal plant pathogen Botrytis cinerea. Biochem J. 2010; 427: 323-32.

106. Vélëz H, Glassbrook NJ and Daub ME. Mannitol metabolism in the phytopathogenic fungus Alternaria alternata. Fungal Genetics and Biology. 2007; 44: 258-68.

107. Dulermo T, Rascle C, Chinnici G, et al. Dynamic carbon transfer during pathogenesis of sunflower by the necrotrophic fungus Botrytis cinerea: from plant hexoses to mannitol. New Phytol. 2009; 183: 1149-62.

108. Schafer W. Molecular Mechanisms of Fungal Pathogenicity to Plants. Annual Review of Phytopathology. 1994; 32: 461-77.

109. Morrissey JP and Osbourn AE. Fungal resistance to plant antibiotics as a mechanism of pathogenesis. Microbiol Mol Biol Rev. 1999; 63: 708-24.

110. Franks TK, Hayasaka Y, Choimes S, et al. Cyanogenic glucosides in grapevine: polymorphism, identification and developmental patterns. Phytochemistry. 2005; 66: 165-73.

111. Jones DA. Why are so many food plants cyanogenic? Phytochemistry. 1998; 47: 155-62.

112. Martine M. Non-nutrient bioactive substances of pulses. British Journal of Nutrition. 2002; 88: S307-S19.

113. de Vries RP. Regulation of Aspergillus genes encoding plant cell wall polysaccharide-degrading enzymes; relevance for industrial production. Appl Microbiol Biotechnol. 2003; 61: 10-20.

114. Freimoser FM, Screen S, Hu G, et al. EST analysis of genes expressed by the zygomycete pathogen Conidiobolus coronatus during growth on insect cuticle. Microbiology. 2003; 149: 1893-900.

115. Tjalsma H, Bolhuis A, Jongbloed JD, et al. Signal peptide-dependent protein transport in Bacillus subtilis: a genome-based survey of the secretome. Microbiol Mol Biol Rev. 2000; 64: 515-47.

116. Deising HB and Kamoun S. The Secretome of Plant-Associated Fungi and Oomycetes. En: The Mycota. Springer Berlin Heidelberg, 2009, pp. 173-80.

117. Bennet JW. The molds of Katrina. Update (NY Acad Sci). 2006; 6-9.

118. Cherrad S, Girard V, Dieryckx C, et al. Proteomic analysis of proteins secreted by Botrytis cinerea in response to heavy metal toxicity. Metallomics. 2012; 4: 835-46.

119. Li B, Wang W, Zong Y, et al. Exploring pathogenic mechanisms of Botrytis cinerea secretome under different ambient pH based on comparative proteomic analysis. J Proteome Res. 2012; 11: 4249-60.

120. Choi J, Park J, Kim D, et al. Fungal secretome database: integrated platform for annotation of fungal secretomes. BMC Genomics. 2010; 11: 105.

121. Brito N, Espino JJ and Gonzalez C. The endo-beta-1,4-xylanase xyn11A is required for virulence in Botrytis cinerea. Mol Plant Microbe Interact. 2006; 19: 25-32.

122.

123. Maldonado AM, Echevarria-Zomeño S, Jean-Baptiste S, et al. Evaluation of three different protocols of protein extraction for Arabidopsis thaliana leaf proteome analysis by two-dimensional electrophoresis. Journal of Proteomics. 2008; 71: 461-72.

124. Wang W, Vignani R, Scali M, et al. A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. Electrophoresis. 2006; 27: 2782-6.

125. Quidde T, Osbourn AE and Tudzynski P. Detoxification of alpha-tomatine byBotrytis cinerea. Physiological and Molecular Plant Pathology. 1998; 52: 151-65.

126. Levis C, Giraud T, Dutertre M, et al. Telomeric DNA of Botrytis cinerea: a useful tool for strain identification. FEMS Microbiology Letters. 1997; 157: 267-72.

127. Godfrey D, Zhang Z, Saalbach G, et al. A proteomics study of barley powdery mildew haustoria. Proteomics. 2009; 9: 3222-32.

128. Bindschedler LV, McGuffin LJ, Burgis TA, et al. Proteogenomics and in silico structural and functional annotation of the barley powdery mildew Blumeria graminis f. sp. hordei. Methods. 2011; 54: 432-41.

129. Gao S, Liu T, Li Y, et al. Understanding resistant germplasm-induced virulence variation through analysis of proteomics and suppression subtractive hybridization in a maize pathogen Curvularia lunata. Proteomics. 2012; 12: 3524-35.

130. Paper JM, Scott-Craig JS, Adhikari ND, et al. Comparative proteomics of extracellular proteins in vitro and in planta from the pathogenic fungus Fusarium graminearum. Proteomics. 2007; 7: 3171-83.

131. Kwon SJ, Cho SY, Lee KM, et al. Proteomic analysis of fungal host factors differentially expressed by Fusarium graminearum infected with Fusarium graminearum virus-DK21. Virus Res. 2009; 144: 96-106.

132. Rampitsch C, Tinker NA, Subramaniam R, et al. Phosphoproteome profile of Fusarium graminearum grown in vitro under nonlimiting conditions. Proteomics. 2012; 12: 1002-5.

133. Yang FEN, Jensen JD, Svensson B, et al. Secretomics identifies Fusarium graminearum proteins involved in the interaction with barley and wheat. Molecular Plant Pathology. 2012; 13: 445-53.

134. Moretti M, Minerdi D, Gehrig P, et al. A Bacterial-Fungal Metaproteomic Analysis Enlightens an Intriguing Multicomponent Interaction in the Rhizosphere of Lactuca sativa. Journal of Proteome Research. 2012; 11: 2061-77.

135. Vincent D, Balesdent MH, Gibon J, et al. Hunting down fungal secretomes using liquid-phase IEF prior to high resolution 2-DE. Electrophoresis. 2009; 30: 4118-36.

136. Wang Y, Wu J, Park ZY, et al. Comparative Secretome Investigation of Magnaporthe oryzae Proteins Responsive to Nitrogen Starvation. Journal of Proteome Research. 2011; 10: 3136-48.

137. Oh Y, Franck WL, Han S-O, et al. Polyubiquitin Is Required for Growth, Development and Pathogenicity in the Rice Blast Fungus Magnaporthe oryzae. PLoS One. 2012; 7: e42868.

138. Jung Y-H, Jeong S-H, Kim SH, et al. Secretome analysis of Magnaporthe oryzae using in vitro systems. Proteomics. 2012; 12: 878-900.

139. Luster DG, McMahon MB, Carter ML, et al. Proteomic analysis of germinating urediniospores of Phakopsora pachyrhizi, causal agent of Asian soybean rust. Proteomics. 2010; 10: 3549-57.

140. Stone C, McMahon M, Fortis L, et al. Gene expression and proteomic analysis of the formation of Phakopsora pachyrhizi appressoria. BMC Genomics. 2012; 13: 269.

141. Torto TA, Li S, Styer A, et al. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen Phytophthora. Genome Res. 2003; 13: 1675-85.

142. Meijer HJ, van de Vondervoort PJ, Yin QY, et al. Identification of cell wall-associated proteins from Phytophthora ramorum. Mol Plant Microbe Interact. 2006; 19: 1348-58.

143. Cao T, Kim YM, Kav NN, et al. A proteomic evaluation of Pyrenophora tritici-repentis, causal agent of tan spot of wheat, reveals major differences between virulent and avirulent isolates. Proteomics. 2009; 9: 1177-96.

144. Matis M, Zakelj-Mavric M and Peter-Katalinic J. Mass spectrometry and database search in the analysis of proteins from the fungus Pleurotus ostreatus. Proteomics. 2005; 5: 67-75.

145. Lakshman DK, Natarajan SS, Lakshman S, et al. Optimized protein extraction methods for proteomic analysis of Rhizoctonia solani. Mycologia. 2008; 100: 867-75.

146. Yajima W and Kav NN. The proteome of the phytopathogenic fungus Sclerotinia sclerotiorum. Proteomics. 2006; 6: 5995-6007.

147. Liang Y, Rahman MH, Strelkov SE, et al. Developmentally induced changes in the sclerotial proteome of Sclerotinia sclerotiorum. Fungal Biol. 2010; 114: 619-27.

148. Liang Y, Strelkov SE and Kav NN. The proteome of liquid Sclerotial exudates from Sclerotinia sclerotiorum. J Proteome Res. 2010; 9: 3290-8.

149. Eastwood DC, Floudas D, Binder M, et al. The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. Science. 2011; 333: 762-5.

150. Tan KC, Heazlewood JL, Millar AH, et al. A signaling-regulated, short-chain dehydrogenase of Stagonospora nodorum regulates asexual development. Eukaryot Cell. 2008; 7: 1916-29.

151. Casey T, Solomon PS, Bringans S, et al. Quantitative proteomic analysis of G-protein signalling in Stagonospora nodorum using isobaric tags for relative and absolute quantification. Proteomics. 2010; 10: 38-47.

152. Coumans JV, Moens PD, Poljak A, et al. Plant-extract-induced changes in the proteome of the soil-borne pathogenic fungus Thielaviopsis basicola. Proteomics. 2010; 10: 1573-91.

153. El-Bebany AF, Rampitsch C and Daayf F. Proteomic analysis of the phytopathogenic soilborne fungus Verticillium dahliae reveals differential protein expression in isolates that differ in aggressiveness. Proteomics. 2010; 10: 289-303.

154. Song X, Rampitsch C, Soltani B, et al. Proteome analysis of wheat leaf rust fungus, Puccinia triticina, infection structures enriched for haustoria. Proteomics. 2011; 11:944-963.

Proteómica clínica. Biomarcadores y dianas terapeúticas

26

J. Ignacio Casal[1], Fernando Vivanco[2]

[1] Centro de Investigaciones Biológicas (CSIC). Madrid, Spain;

[2] Instituto de Investigación Sanitaria Fundación Jiménez Diaz, Universidad Complutense. Madrid, Spain.

1. Introducción.

Un biomarcador es una molécula, característica o propiedad que es objetivamente medible y evaluable como indicador de procesos biológicos normales, procesos patológicos o respuestas farmacológicas a una intervención terapéutica. En el ámbito de la práctica clínica se utiliza fundamentalmente para detectar una enfermedad (marcador diagnóstico), seguir su evolución (marcador pronóstico), o elegir un tratamiento (marcador predictivo) y monitorización. Sin embargo, la utilidad de los biomarcadores es múltiple, incluyendo no solo aspectos clínicos, sino también su aplicación en el desarrollo de fármacos, al poder ser empleados como dianas terapéuticas. Los biomarcadores pueden constituir una de las principales fuentes de información en el desarrollo de los procesos biológicos y ser de gran utilidad en el desarrollo futuro de nuevos medicamentos (1). En clínica humana la determinación de biomarcadores, que han sido validados en amplias cohortes de miles de muestras, influye directamente en las decisiones sobre el diagnóstico, pronóstico o tratamiento y seguimiento de los pacientes y de ahí procede su gran importancia (biomarcadores como indicadores de enfermedad). Desde una perspectiva proteómica los biomarcadores son proteínas o péptidos (individuales o grupos de ellos, denominados paneles o perfiles) (1,2). En las áreas oncológica y cardiovascular, que se utilizarán como ejemplos en este capítulo, la mayoría de los biomarcadores son de naturaleza bioquímica [moléculas (proteínas, péptidos) que se cuantifican en la sangre, orina] o de imagen (obtenidas mediante técnicas de imagen, como la tomografía computarizada multicorte, (TC), o la resonancia magnética nuclear (RMN), etc.) y que aparecen como alteraciones constituyentes de los tejidos o de los líquidos fisiológicos.

2. Selección de muestras biológicas en la búsqueda de biomarcadores.

La selección de las muestras biológicas para la búsqueda de nuevos biomarcadores es un paso crítico cuya finalidad es maximizar las diferencias entre las muestras a comparar minimizando el número de las mismas que es obligado analizar para obtener resultados estadísticamente significativos. Con frecuencia el problema fundamental es la selección de los controles adecuados, puesto que no es fácil, por ejemplo, disponer de muestras de corazón normal

(sano). Además, el concepto de “normal” o “sano”, desde un punto de vista proteómico es relativo, puesto que no conocemos el proteoma que constituye los tejidos humanos, máxime teniendo en cuenta la gran variabilidad intrínseca presente en las proteínas conocidas. Una opción para evitar esa variabilidad intrínseca es comparar en un mismo individuo, en una misma muestra, la zona patológica con la parte sana (región de arteria sana adyacente a la zona de arteria con placa de ateroma) (3,4). Cuando no es posible encontrar los controles sanos adecuados puede recurrirse al uso de modelos animales de la enfermedad que mimetice la patología en humanos. Igualmente, la selección de las muestras debería ser tal que se diferenciasen exclusivamente en lo que se desea estudiar, lo cual no es sencillo considerando la complejidad de las enfermedades humanas.

Entre la variedad de posibles muestras biológicas a estudiar las más utilizadas son los extractos de tejidos completos o subproteomas de los mismos (mitocondrias, membranas, etc.), células en cultivo o aisladas mediante microdisección con láser, el conjunto de proteínas secretadas por las dos muestras anteriores (denominado secretoma) y los distintos líquidos biológicos del organismo (plasma o suero, orina, saliva, líquido cefalorraquídeo o sinovial, etc.) y sus correspondientes subproteomas (glicoproteínas o fosfoproteínas plasmáticas, exosomas, micropartículas, etc.) (5,6).

El plasma (o el suero) es el material de partida preferido, juntamente con la orina, para la búsqueda de nuevos biomarcadores, porque además de su accesibilidad, es en plasma donde habitualmente se planea desarrollar algún ensayo para la determinación del posible nuevo biomarcador. Se asume que el plasma contiene proteínas circulantes procedentes de todos los tejidos del cuerpo con los que contacta, tanto en condiciones fisiológicas como en los procesos patológicos (7). La veracidad de tal asunción es difícil de establecer puesto que actualmente el proteoma completo del plasma humano no es conocido y los datos recientes más fiables establecen valores entre 2000 (8) y 3000 (9) proteínas. Igualmente se ha establecido la presencia de casi 2000 proteínas en la orina humana normal (10). Este número de proteínas plasmáticas tan elevado, junto con la presencia de un grupo muy reducido de ellas con gran concentración (las 20 proteínas mayoritarias constituyen el 99% del total) y su gran rango dinámico [diferencia de abundancia entre la proteína más concentrada (albúmina) y la de menor concentración (IL-6)] de 10 órdenes de magnitud], hace especialmente difícil la búsqueda de nuevos biomarcadores en plasma (7). Es por ello necesario deplecionar el plasma de las proteínas mayoritarias (11) para poder acceder a las minoritarias. Esto se lleva a cabo habitualmente mediante cromatografía de inmunoafinidad o métodos de ecualización de concentraciones de proteínas, entre otros. El análisis de las proteínas minoritarias suele llevarse a cabo mediante un fraccionamiento multidimensional, habitualmente una cromatografía de intercambio iónico seguida de otra de fase reversa de las distintas fracciones de la primera. En este caso, una muestra inicial de plasma se convierte en múltiples fracciones que requieren muchas horas de análisis instrumental (12, 13), lo que permite identificar un gran número de proteínas (el denominado proteoma profundo del plasma). Los potenciales biomarcadores suelen encontrarse a muy baja concentración en plasma (ng/ml) porque si su producción es local (en un tumor localizado, en un corazón isquémico) al alcanzar la circulación se diluyen en los aproximadamente 6 litros de sangre de nuestro organismo. Alternativamente se utilizan otros líquidos biológicos como el líquido cefalorraquídeo (LCR) (en la enfermedad de Alzheimer), el líquido intersticial en tumores, o la orina en las enfermedades renales, donde se supone una mayor concentración de la proteína de interés (14).

Actualmente existen unos 205 marcadores (proteínas) plasmáticos aprobados para la práctica clínica (15). Sin embargo, el número de nuevas proteínas aprobadas (por la FDA americana) como biomarcadores en los últimos años va decreciendo (15). Esta situación contrasta con el cada día mayor número de proteínas plasmáticas conocidas y la urgente necesidad de nuevos biomarcadores (sobre todo de fases tempranas) de las enfermedades de mayor prevalencia (cardiovasculares, oncológicas, neurodegenerativas, etc.)(15).

En los últimos años está cobrando una gran relevancia como material de partida para la identificación de posibles nuevos biomarcadores la orina, no solo su proteoma sino especialmente el estudio de su peptidoma (16). Ello es debido en gran medida al auge que está teniendo la aplicación de la electroforesis capilar acoplada on-line a los espectrómetros de masas (CE-MS) (17,18) en la determinación del peptidoma urinario, que consta actualmente de más de 5000 péptidos, recogidos en diversas bases de datos (19). El peptidoma urinario está empezando a ser utilizado como herramienta diagnóstica de diversas enfermedades cardiorenales (20). Igualmente hay un interés creciente en el estudio del proteoma de la saliva del que se han descrito más de 1000 proteínas (21).Existen diversos dispositivos que utilizando la microfluídica son capaces de determinar grupos de biomarcadores en saliva en un tiempo muy reducido (22).

Se acepta habitualmente, sobre todo para enfermedades multifactoriales como las cardiovasculares, oncológicas, neurodegenerativas u osteoarticulares, que un único marcador no es lo más adecuado para aplicarse a dichas enfermedades, sino la utilización de múltiples marcadores (paneles, perfiles) que caracterizan mejor la complejidad de tales patologías. La proteómica comparativa de expresión diferencial entre las proteínas de un mismo tipo de muestras en dos estados diferentes (sano/patológico; tratado/no tratado) es el método de elección para la identificación de potenciales perfiles proteicos asociados a una enfermedad. La identificación de estos perfiles en plasma (o en orina), característicos específicamente de la patología en estudio, se complica, porque además de la complejidad y el rango dinámico del plasma ya citados, muchos biomarcadores de enfermedad están en muy baja concentración (nM, pM), y existe una gran variabilidad en las proteínas humanas. No existe una única técnica que permita abordar el descubrimiento y caracterización de nuevos biomarcadores por lo que habitualmente es un proceso que se lleva a cabo en varias fases sucesivas utilizando plataformas y métodos diferentes.

3. Fases en el descubrimiento y validación de un nuevo biomarcador.

El proceso de descubrimiento y validación de un nuevo biomarcador, en general y mediante técnicas proteómicas, en particular, constituye una ardua y larga tarea, que suele desarrollarse siguiendo una serie de fases que se describen a continuación(23,24) (Figura 1):

Figura 1.- Representación esquemática de las fases que suelen llevarse a cabo en el descubrimiento y validación de un nuevo biomarcador. A la derecha, comparación entre el número inicial de muestras y analitos a lo largo de las distintas fases. Mientras que en la fase de descubrimiento puede usarse un número reducido de muestras, muy bien seleccionadas clínicamente, que contiene miles de proteínas o péptidos, en las fases finales de validación solo un pequeño número de proteínas candidatas son testadas en amplias cohortes de muestras (miles) para establecer su utilidad clínica.

Fase de descubrimiento. Esta fase inicial se lleva a cabo habitualmente mediante proteómica diferencial, comparando dos o más muestras, típicamente una célula, líquido biológico o tejido normal frente a la misma muestra de carácter patológico. Es de suma importancia definir clínicamente de forma muy precisa la patología objeto de estudio (seleccionar el grupo de pacientes y sus controles), evitando al máximo la presencia de otras patologías o condiciones que puedan complicar o confundir la interpretación de los resultados. Esta fase inicial es por definición sin sesgo alguno y puede llevarse a cabo mediante 2-DE, DIGE o preferentemente mediante LC-MS/MS (i.e., iTRAQ, SILAC). El resultado del estudio comparativo son listas de proteínas diferencialmente expresadas (presentes con distinta abundancia o concentración) entre el control y el estado patológico, en base al análisis semicuantitativo de los péptidos presentes en los datos MS o del marcaje isotópico. Estas listas de proteínas, a veces cientos de ellas, suelen contener abundantes falsos positivos (ver la descripción de estas técnicas en otros capítulos), por lo que las proteínas diferencialmente detectadas en esta fase inicial se consideran proteínas candidatas a biomarcadores (y no biomarcadores). Un aspecto especialmente relevante es intentar identificar nuevos biomarcadores que se correspondan con las fases tempranas de la enfermedad, porque sería altamente beneficioso para los pacientes. Esta es una labor complicada en cuanto a la obtención de muestras y porque en las fases iniciales la variación en la concentración de los potenciales biomarcadores es mínima. Mientras que la fase de descubrimiento es por esencia no-sesgada, todas las fases posteriores están dirigidas a verificar si las proteínas candidatas son verdaderos biomarcadores validables en grandes cohortes de muestras clínicas.

Fase de Confirmación (o Cualificación). Se utiliza para confirmar que la expresión de las proteínas candidatas está realmente alterada en las muestras patológicas, utilizando técnicas diferentes (ELISA, western blot, inmunohistoquímica) a las empleadas en la fase de descubrimiento y técnicas con una mayor exactitud en la cuantificación de las proteínas candidatas(s). Además se confirma la expresión diferencial de las proteínas de interés en plasma, si es el caso. La finalidad principal de esta fase es establecer una clara asociación entre la(s) proteína(s) candidata y la enfermedad estudiada, valorándose especialmente la sensibilidad (probabilidad de que las muestras patológicas sean realmente positivas en base a la detección de esa proteína), frente a la especificidad (probabilidad que una muestra sana resulte negativa). Lo que al final define el status de proteína candidata es la variación permanente en la abundancia de la proteína entre los dos estados comparados (sano/patológico). Actualmente la técnica de elección para las fases de cualificación y verificación es el ensayo de SRM (ver más adelante) que tiene mayor exactitud y sensibilidad que las técnicas utilizadas en la fase de descubrimiento y permite la evaluación simultánea de varios cientos de proteínas.

Fase de Verificación. Es una fase caracterizada por la realización de un amplio estudio, utilizando un gran número de muestras (cientos), con frecuencia (o adicionalmente) muestras de plasma, que incluya controles variados y muestras de distintas localizaciones (de distintas instituciones), para evaluar las posibles influencias ambientales y la variabilidad intrínseca de la población objeto de estudio. Se pretende confirmar la sensibilidad de las proteínas candidatas y se hace una primera aproximación a su especificidad. Es de resaltar que el número de muestras que se analizan va creciendo según se avanza en las sucesivas fases mientras que el número de proteínas candidatas que perviven va decreciendo. Las muestras utilizadas en cualquier fase deben ser independientes de las utilizadas en el resto de las fases.

Fase de Validación. Las escasas proteínas candidatas que llegan a esta fase son evaluadas mediante ensayos optimizados para cada una de ellas individualmente o como conjunto (perfil, panel), en miles de muestras, de forma que se establezca la variación en la población seleccionada. Estos ensayos deben tener un alto grado de sensibilidad y especificidad y un bajo coeficiente de variación (<10%) con un control muy estricto de un conjunto de indicadores (exactitud, reproducibilidad, precisión, linealidad, etc.) cuya descripción queda fuera de los límites de este capítulo. Con frecuencia hay que desarrollar nuevos ensayos específicos para las

proteínas nuevas (producción de nuevos anticuerpos, ELISAs), no conocidas previamente, y que han sido identificadas en la fase de descubrimiento. Para los biomarcadores así validados, se refinan y ajustan los ensayos de su cuantificación para que cumplan con los estrictos estándares exigidos para convertirse en un test o ensayo clínico. Los test clínicos requieren de la aprobación específica de las agencias nacionales o internacionales para ser utilizados con esa finalidad (diagnóstica, pronóstica). En caso de aprobación, los nuevos biomarcadores entrarían en la última fase de comercialización del ensayo o test. El uso de la técnica de SRM ("Selected Reaction Monitoring") como técnica para la validación y cuantificación de biomarcadores en fluidos biológicos y otro tipo de muestras biológicas se describe más adelante. Retrospectivamente puede comprobarse que, con frecuencia, la consecución de todas las fases, de los biomarcadores actualmente disponibles comercialmente, ha durado un periodo de más de 20 o 30 años desde la descripción inicial del biomarcador.

4. Necesidad de nuevos biomarcadores de la enfermedad cardiovascular.

Las enfermedades cardiovasculares, incluyendo el infarto agudo de miocardio, el fallo cardiaco, el ictus cerebral y la enfermedad arterial periférica, son la primera causa de muerte en los países industrializados y va aumentando progresivamente en los países en vías de desarrollo. El proceso patológico que subyace a la mayoría de ellas es el engrosamiento de la pared arterial de arterias de medio y gran calibre, debido a la formación de placas ateroscleróticas ricas en lípidos (colesterol y sus derivados) y recubiertas por una capa fibrosa. Las placas con alto contenido lipídico y abundantes células inflamatorias (macrófagos, linfocitos) se consideran inestables, vulnerables o de alto riesgo, por su propensión a fisurarse o romperse, exponiendo a la sangre diversos componentes altamente trombogénicos (factor tisular, colágeno), que activa el sistema de la coagulación, la formación de fibrina y la adhesión y agregación de plaquetas y genera un trombo, que bloquea total o parcialmente el vaso y precipita un evento agudo (25). Según dónde se localice la arteria con el trombo, esta oclusión puede dar lugar a un síndrome coronario agudo (infarto agudo de miocardio, angina de pecho), o a un accidente cerebrovascular (ictus o infarto cerebral) que puede ocasionar la muerte súbita o dejar graves secuelas a quienes lo sufren, entre otros. La manifestación clínica aterosclerótica se caracteriza, por tanto, por la formación de un trombo anclado en la placa rota o erosionada. Por ello, el término aterotrombosis (aterosclerosis y sus complicaciones trombóticas) define de forma más completa la naturaleza de la enfermedad. Aunque se han producido grandes avances en el tratamiento de estas enfermedades, la medicina actual no es capaz de predecir de manera adecuada quienes tienen riesgo de sufrir un evento agudo ni de detectar la presencia de placas ateroscleróticas vulnerables a la rotura. Por ello, uno de los mayores retos de la medicina cardiovascular es encontrar la manera de predecir el riesgo de un sujeto de sufrir un evento trombótico agudo e identificar biomarcadores que indiquen precozmente la situación de vulnerabilidad de las placas ateroscleróticas (26).

Ante la ausencia de biomarcadores, durante los últimos años se han venido utilizando los factores de riesgo cardiovascular para predecir el riesgo de sufrir un evento cardiovascular agudo (infarto de miocardio, ictus). Un factor de riesgo es una característica o circunstancia asociada a un incremento en la probabilidad de enfermar, que precede a su aparición y que se le atribuye causalidad o responsabilidad. Los factores de riesgo cardiovascular actualmente utilizados provienen de estudios epidemiológicos e incluyen la hipertensión, hipercolesterolemia, diabetes mellitus, tabaquismo, edad, etc., que permiten explicar aproximadamente un 50% de los casos. Sin embargo, sigue habiendo una gran incidencia de eventos isquémicos agudos inesperados, que no es posible predecir. Por ello, es necesario la búsqueda de nuevos biomarcadores diagnósticos y pronósticos de la enfermedad cardiovascular, que puedan ser identificados en

sangre, orina u otros líquidos biológicos (saliva) (27). Es importante resaltar que la pared vascular libera al torrente circulatorio moléculas que pueden reflejar los procesos patológicos que están ocurriendo en ella. Además, diversos componentes de la sangre participan directamente en la formación del trombo (factores de la coagulación, plaquetas). Por tanto hay un numeroso grupo de moléculas implicados en el proceso aterotrombótico que podían ser biomarcadores de la enfermedad. Sin embargo, no debe confundirse el concepto de biomarcador con el de proteínas asociadas o implicadas en la patología. Para ser considerado un biomarcador debe cumplir un conjunto de requisitos: 1) Ser específico de una enfermedad concreta. 2) Tener una alta sensibilidad: ser fácilmente cuantificable, incluso a bajos niveles. 3) Predictivo: variable con el curso de la enfermedad y el tratamiento. 4) Estable: iguales concentraciones a lo largo del día. Es deseable además, que sea no-invasivo, con facilidad para ser determinado en muestras de sangre, orina, saliva, etc.

5. Identificación de proteínas candidatas en patología cardiovascular mediante análisis proteómico.

Actualmente no hay ninguna proteína identificada mediante estrategias proteómicas que haya sido validada como biomarcador en alguna de las enfermedades cardiovasculares. Sin embargo muchas de las proteínas candidatas, identificadas mediante expresión diferencial (fase de descubrimiento), están siendo testadas en alguna de las fases descritas previamente. La figura 2 recoge de manera esquemática varias de las estrategias utilizadas en la fase de descubrimiento. Como puede apreciarse se han utilizado fundamentalmente extractos de arterias carótidas patológicas y controles sanos (obtenidas por endarterectomía) y en menor medida arterias coronarias, de muy difícil obtención. En la columna sombreada aparecen algunas de las proteínas candidatas (α1-antitripsina, hsp-27, fibrinógeno, etc.) identificadas y a su derecha el método utilizado (2-DE, HPLC, LC-MS, etc.). Mediante microdisección mediada por láser se han estudiado comparativamente, células musculares lisas (VSMC) o las tres capas fundamentales que componen los vasos (íntima, media y adventicia), obtenidas de arterias ateroscleróticas y de arterias sanas, o segmentos sanos y ateroscleróticos de una misma arteria. En los casos en los que el número de proteínas diferencialmente expresadas es elevado (137 proteínas en la capa media o 194 en la adventicia) se seleccionan solo un reducido grupo de ellas, mediante criterios adicionales, para proceder a las fases de verificación y validación. Una de las estrategias experimentales más ventajosas en la búsqueda de nuevos biomarcadores es el análisis de las proteínas secretadas (secretoma) por el tejido (o células) in vitro, como muestra la parte

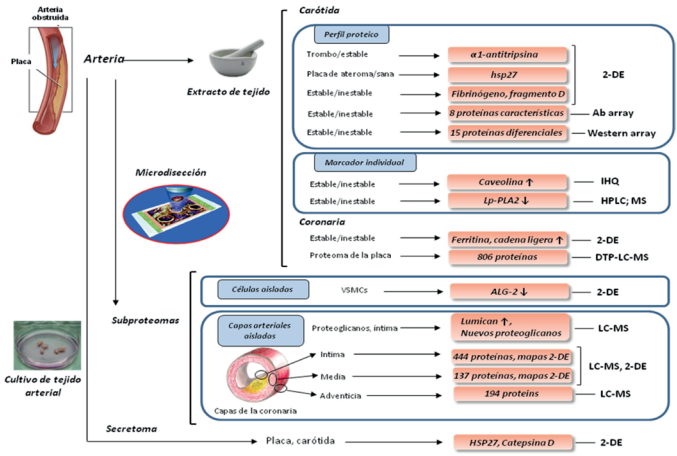


Figura 2.- Diversos tipos de muestras que han sido utilizados en la búsqueda de nuevos biomarcadores de aterosclerosis utilizando proteómica de expresión diferencial. El material de partida son segmentos de arterias sanas y de arterias con placas de ateroma que son analizados comparativamente, como tejido completo, en alguna de sus capas o de las células que están implicadas. La comparación entre placas estables e inestables proporciona, potencialmente, las proteínas asociadas a la vulnerabilidad de las placas de ateroma.

inferior de la figura 2. In vivo, las proteínas secretadas por las arterias (sanas o patológicas) se diluirían en el torrente circulatorio, mientras que in vitro, el tejido cultivado en un medio libre de proteínas, las proteínas secretadas son recogidas en un pequeño volumen de la placa de cultivo y fácilmente identificadas. Posteriormente puede detectarse su presencia en plasma.

6. Biomarcadores y descubrimiento de fármacos. Farmacoproteómica.

El descubrimiento de nuevos fármacos es un proceso inherentemente ineficiente y la explosión de datos disponibles en la actual era post-genómica aún no ha tenido gran impacto en la aprobación de nuevas entidades químicas terapéuticas (28,29). Considerando que las proteínas son las principales dianas de los fármacos y que existen cientos de miles de proteínas, incluyendo variantes de splicing y modificaciones postransduccionales, existe suficiente terreno para pasar de las 500 dianas proteicas actuales a por lo menos 3.000-10.000 nuevas dianas "drogables" (30). La proteómica puede contribuir en todas las fases del proceso de desarrollo de fármacos: identificación de dianas, validación, optimización de "leads" y desarrollo pre-clínico y clínico. Existen fundamentalmente tres tipos de biomarcadores en desarrollo de fármacos. **Biomarcadores de pronóstico,** que permiten conocer la evolución de la enfermedad en el individuo y deciden por tanto sobre a quién tratar o no tratar o como de agresivo ha de ser el tratamiento. Existen varios test genómicos de este tipo ya en uso (p.e Oncotype DX (Genomic Health) o el MammaPrint (Agendia). **Biomarcadores predictivos** se utilizan para discernir la probabilidad de que un paciente se beneficie de un tratamiento particular. Veremos varios ejemplos (HER2, EGFR) en el apartado de medicina personalizada. Por último, los **biomarcadores farmacodinámicos** miden los efectos a corto plazo del tratamiento y pueden en teoría ser utilizados para guiar la selección de dosis en fases tempranas de desarrollo clínico de un nuevo fármaco (31). Recientemente, la magnitud de la inhibición de la actividad quinasa BCR-ABL se demostró que correlacionaba con la evolución clínica, justificando la selección personalizada de dosis basada en ensayos previos (32). Los biomarcadores predictivos y farmacodinámicos se utilizan cada vez más en ensayos clínicos.

Una de las aproximaciones proteómicas al estudio de nuevas dianas terapéuticas consiste en el estudio y mapeo de rutas de señalización. Por ejemplo, el papel que juega la fosforilación quinasa-dependiente en el control de ciclo celular y la proliferación. Una forma de medir esta fosforilación es mediante anticuerpos anti-fosfoproteína dispuestos en formato de microarrays de fase reversa, lo que permite un estudio simultaneo de múltiples rutas de señalización (33-35) o bien utilizando métodos basados en MS para identificar y cuantificar selectivamente fosfopéptidos (36). (Ver capitulo 32).

En los últimos años se ha desarrollado la proteómica química para buscar y caracterizar alteraciones en niveles de proteínas que puedan ser dianas farmacológicas. En este tipo de proteómica, un compuesto químico (droga) se inmoviliza sobre un soporte solido tipo sefarosa o agarosa que posteriormente se utiliza para capturar por afinidad las dianas (proteínas) que interaccionan con ella a partir de un extracto celular o de tejido. Las proteínas retenidas son separadas electroforéticamente o por cromatografía líquida, seguido de su identificación por espectrometría de masas (37). Con la llegada de las nuevas técnicas de proteómica cuantitativa (ICAT, iTRAQ o SILAC) es ahora posible realizar estos experimentos de forma cuantitativa y a gran escala. Por ejemplo, estudios recientes han descrito el perfil de dianas para imatinib, nilotinib, dasatinib y bosutinib, todos ellos inhibidores de BCR-ABL utilizados en leucemia mieloide crónica (38,39). Estos estudios arrojaron datos interesantes sobre la selectividad de estos fármacos. Así, dasatinib y bosutinib fueron más promiscuos que imatinib o nilotinib. Estos

estudios también revelaron nuevas dianas para imatinib y nilotinib: la tirosin-quinasa DDR1 y la oxido-reductasa NQO2. El descubrimiento de otras dianas del sistema inmune predice serios efectos colaterales inmunosupresivos para dasatinib y bosutinib. En otro ejemplo, Bach et al. (40) caracterizaron el perfil de dianas de la roscotivina, un inhibidor de CDKs, describiendo un panel de 151 quinasas.

La proteómica química también se aplica a productos naturales. Por ejemplo, se han identificado los targets del macrolido antitumoral pladienolido. El factor de splicing SF3b fue identificado como diana y posteriormente validado por siRNA y experimentos de co-localización (41). Asimismo, mediante una aproximación similar se descubrió que la diana del antimicótico marino diazonamida A es la ornitina δ-amino transferasa, que juega un importante papel en el ensamblaje del huso mitótico y la división celular (42).

Otra interesante aplicación de la proteómica es al estudio del mecanismo de acción de un fármaco, en ocasiones en las propias células de interés. Mediante geles 2D y análisis MALDI-TOF se realizaron estudios de expresión de proteínas subsiguientes al tratamiento de células de adenocarcinoma ductal de páncreas con tricostatina A, un inhibidor de histona deacetilasa (HDAC) (43). La tricostatina A parece aumentar las proteínas que promueven muerte celular y reprimir aquellas implicadas en proliferación y crecimiento celular. Siguiendo una aproximación similar se estudió el mecanismo de acción de la aplidina, un antitumoral que induce apoptosis vía activación de RAC1-JNK (44). Un numero de proteínas implicadas en apoptosis estaban desreguladas, entre ellas Rab geranilgeraniltransferasa, proteina disulfuro isomerasa (PDI), cistationina γ-lyasa, ezrina, and ciclofilina A (CypA). El papel de PDI y CypA en la resistencia a Aplidina se confirmó funcionalmente. Por tanto, estas proteínas son candidatas a mediar, al menos parcialmente, la acción de Aplidina y podrían suministrar a las células una ruta de escape a la inducción de apoptosis por esta droga. En otras patologías, la aproximación farmacoproteómica fue especialmente útil en la identificación de alteraciones moleculares implicadas en diabetes tipo 2 y para la caracterización de los fármacos existentes (45).

Otro campo relevante es el del estudio de la resistencia a fármacos. Es bien sabido que la resistencia a la quimioterapia es un factor importante en el fracaso de muchos tratamientos del cáncer y otras patologías. En algunos casos, una fracción de tumores responde al tratamiento y otra no, mientras que en otros casos el tumor responde inicialmente pero eventualmente recrece hasta hacerse resistente. Esta resistencia generalmente presenta un componente multifactorial, ideal para ser analizado por estrategias de identificación masivas como la proteómica. Varios grupos han utilizado geles 2D en combinación con MALDI-TOF para monitorizar las alteraciones en niveles de proteínas tras la exposición a varios fármacos bien conocidos. Se estudió el efecto de la daunorubicina sobre las células de cáncer de páncreas (46), descubriéndose varias chaperonas que podrían contribuir a este efecto. En otro estudio se investigó la químioresistencia en células de melanoma (47). De nuevo varias chaperonas aparecieron implicadas en los mecanismos de resistencia. Sin embargo, estos datos nos obligan de nuevo a ser cautos con la sobredetección de chaperonas y otras proteínas "heat-shock" en estudios realizado con geles 2D, al tratarse de proteínas fácilmente alteradas en cualquier perturbación fisiológica. También se han realizado estudios del proteoma del plasma de pacientes con isquemia coronaria sensibles o resistentes a aspirina. Se observó un incremento en la expresión de un isotipo de la cadena γ del fibrinógeno y tres isotipos de haptoglobina, así como de otros tres isotipos de la proteína de unión de vitamina D en los pacientes resistentes a aspirina (48). Mediante arrays de fase reversa se demostró que la vía de señalización de TGF-β puede jugar un papel importante como marcador de químioresistencia en cáncer seroso de ovario avanzado (49).

7. Biomarcadores y medicina personalizada.

Las nuevas tecnologías –omicas de alto rendimiento están acercando a la realidad el concepto de medicina "personalizada" o "estratificada". El objetivo de la medicina personalizada es ofrecer "la medicina adecuada al paciente adecuado en el momento adecuado". Un requerimiento absoluto e inherente a este proceso es la necesidad de identificar no solo las dianas terapéuticas sino también los pacientes que las poseen y pueden por tanto responder a ellas. Por tanto, la identificación de marcadores diagnósticos de compañía o biomarcadores de fármacos es imperativa. El mantra "no biomarker, no drug" es repetido habitualmente por todas las compañías farmacéuticas. La proteómica tiene la ventaja de poder encontrar y aportar múltiples marcadores para un mismo problema. Dentro de este contexto tienen particular interés no solo los biomarcadores para la clasificación y/o estratificación de los pacientes sino los biomarcadores de pronóstico. Un biomarcador pronóstico permite la identificación de pacientes con diferentes riesgos de progresión o muerte.

Un marcador de enriquecimiento, estratificación o de compañía puede distinguir entre diferentes subgrupos poblacionales donde diferentes opciones de tratamiento son posibles, pero no puede guiar la elección de una particular terapia. Por ejemplo, la presencia de mutaciones de K-RAS en el uso de cetuximab frente a EGFR, en combinación con quimioterapia convencional en cáncer de pulmón (50), el uso de HER2 en cáncer de mama (51) mediados por anticuerpos monoclonales como Herceptin (trastuzumab) en pacientes de cáncer de mama con sobreexpresión de HER2 (52), el uso de inhibidores químicos de ALK en cáncer de pulmón conteniendo la fusión EML4-ALK o marcadores relacionados con angiogénesis en el tratamiento de varios tipos de cáncer con bevacizumab (53) han guiado decisiones clínicas sobre tratamiento dirigidos. Todos estos biomarcadores se descubrieron en base a estudios retrospectivos sobre si los pacientes respondían o no a la terapia.

El uso de biomarcadores para el descubrimiento de nuevos fármacos está asociado a la búsqueda de una terapia dirigida ("targeted therapy") contra un evento molecular específico. Sin embargo en muchos casos las drogas van dirigidas o afectan múltiples dianas moleculares. El descubrimiento de los inhibidores de PARP, como el olaparib, y la demostración de su efecto en pacientes con canceres asociados a BRCA 2 (54) han permitido el cierre del circulo terapéutico. En este sentido la inhibición de PARP se muestra como un biomarcador farmacodinámico. Tecnologías recientes que hacen uso de microarrays de proteínas de fase reversa han permitido la caracterización de los niveles de expresión de múltiples proteínas de forma simultánea y la caracterización por tanto de vías de señalización alteradas. Por ejemplo, se ha caracterizado la ruta PI3K como una mayor causa de activación en cáncer de ovario y ha permitido el descubrimiento de agentes contra AKT y mTOR (55).

Biomarcadores de respuesta a terapia. Algunos marcadores como CA125 o CA19-9, presentes en enfermedades benignas, pueden ser útiles para la monitorización de la progresión o la respuesta a terapia. Recientemente se han obtenido resultados esperanzadores que predicen la respuesta a terapias dirigidas. Perfiles de expresión de proteínas de pacientes de cáncer de pulmón tratados con gefitinib permitieron identificar nueve proteínas diferenciales entre respondedores y no respondedores (56). En otro estudio, el análisis por MALDI de sueros de pacientes pretratados se utilizó para identificar pacientes que se pueden beneficiar del tratamiento con inhibidores de la tirosina-quinasa EGFR. Se desarrolló un algoritmo capaz de predecir la respuesta al tratamiento en una cohorte de 139 pacientes (57). Aunque prometedor, al igual que en el caso del SELDI, no se pudo determinar la identidad de las proteínas implicadas. Un análisis proteómico y genómico de un gran panel de tumores de pacientes recibiendo tratamientos basados en taxol demostraron que las α-defensinas estaban asociadas estadísticamente con la respuesta a terapia en el momento de la cirugía (58).

A pesar de todo lo anterior, la aplicación de la proteómica al campo de la medicina personalizada aún está en su infancia y dista mucho de los avances conseguidos a nivel genómico. Recientes avances tecnológicos en la identificación y cuantificación de las proteínas, así como una mayor disponibilidad de biobancos y modelos animales deberían permitir un avance considerable en este campo. Se han publicado intentos para aunar estos esfuerzos, aunque sin resultados claros. En un reciente ejemplo, Marko-Varga y colaboradores proponen una metodología para desarrollar un análisis de sangre para comprobar el potencial desarrollo de enfermedad pulmonar intersticial en pacientes de cáncer de pulmón tratados con gefitinib (59), sin embargo no aportan datos concretos sobre el resultado final.

8. Biomarcadores en cáncer.

El cáncer continua siendo una de las mayores causas de mortalidad entre la población humana. Con el aumento de la esperanza de vida, la prevalencia de muchos canceres continuará aumentando en las próximas décadas y con ello el problema sanitario y social que el cáncer representa. Sin duda, la detección temprana, antes de la metástasis, es crítica para la recuperación y supervivencia de los pacientes. Actualmente, los métodos de detección se basan fundamentalmente en técnicas de imagen (mamografía para cáncer de mama, TC para cáncer de pulmón, colonoscopia para cáncer colorrectal, etc.), procedimientos en general invasivos y difícilmente aplicables al cribado masivo de poblaciones. Los biomarcadores pueden ser una de las mejores herramientas para el diagnóstico y pronóstico del cáncer, especialmente cuando se pueden realizar directamente en plasma o suero. La proteómica constituye una tecnología muy prometedora para la identificación de biomarcadores y dianas terapéuticas en cáncer. En este sentido han surgido iniciativas como la “Human Proteome Organization (HUPO” para coordinar la búsqueda de biomarcadores, principalmente en plasma (60).

Los biomarcadores en cáncer pueden ser DNA, mRNA, proteínas, metabolitos… y ser producidos bien por el mismo tumor o células adyacentes (estroma) en respuesta al tumor y/o condiciones asociadas, tales como inflamación. Los biomarcadores tumorales se pueden utilizar para el cribado de la población general, para el diagnóstico diferencial en pacientes sintomáticos y para el estadiaje clínico del cáncer. También se pueden utilizar para el estudio de la respuesta al tratamiento o como marcadores pronósticos de la progresión del cáncer. Los biomarcadores pueden ser detectados en tejidos, suero u otros fluidos biológicos, aunque desde un punto de vista práctico el suero y el plasma ofrecen las mejores condiciones y la menor invasividad (61). Sin embargo, los avances en este campo han sido escasos a pesar del intenso esfuerzo de los últimos años. Las razones son variadas: primero, una limitada disponibilidad de técnicas para el descubrimiento del proteoma profundo y una limitada comprensión de los rasgos moleculares que inducen la aparición de biomarcadores en cáncer; segundo, la limitación de recursos en el caso de investigadores individuales; tercero, la dificultad para establecer equipos integrados con los recursos apropiados y la correcta combinación de experiencia (básica, clínica, estadística, …) y, finalmente, la limitada disponibilidad de muestras biológicas bien caracterizadas que cumplan con los requisitos adecuados. A pesar de todo ello se van produciendo progresos, la FDA aprobó recientemente un test diagnóstico, basado en un inmunoensayo de 5 proteínas, para la identificación preoperativa de cáncer de ovario (OVA1™) (62).

Un marcador ideal debería poder ser fácilmente medible y reproducible, todo ello de forma coste-efectiva mediante un ensayo con alta sensibilidad y especificidad. Uno de los problemas tradicionalmente asociados con la identificación de biomarcadores tumorales es su baja sensibilidad y especificidad individual. Para aliviar este problema se utilizan combinaciones de varios marcadores, que suele resultar en una mejora significativa del poder diagnóstico. El

primer marcador en cáncer fue la cadena ligera de las inmunoglobulinas que se identificó en 1848 en un 75% de pacientes con mieloma (63). Ya recientemente en la década de los 60, el descubrimiento de la alfa-fetoproteina (AFP) y el antígeno carcino-embrionario (CEA), aplicados a cáncer hepático y cáncer de colon, respectivamente, abrió el camino a las nuevas técnicas e inmunoensayos. En 1980, se descubrió el antígeno específico de próstata (PSA) que se utiliza para la detección precoz de cáncer de próstata (64). Otros marcadores en uso son el CA125 para cáncer de ovario y el CA19-9 para cáncer de páncreas. Todos ellos presentan serias limitaciones debido a su baja sensibilidad y especificidad en estadios tempranos de la enfermedad. Algunos canceres, como mama o pulmón no poseen ni siquiera esta alternativa de biomarcadores. En la Tabla 1 se resumen algunos de las marcadores en uso clínico actualmente (adaptado de (64).

Tabla 1. Biomarcadores en uso clínico*.

Marcador tumoral	Tipo de cáncer	Año de descubrimiento	Aplicación	Origen
Alfa-fetoproteina	Hepatoma	1963	Diagnostico, estadio, recidiva, terapia	Suero
Calcitonina	Carcinoma medular de tiroides	1970s	Diagnostico, Monitorización de terapia	
CA125	Ovario	1981	Pronóstico, recidiva, monitorización de terapia	Suero
CA 15-3	Mama	1984-85	Monitorización de terapia	Suero
CA 19-9	Páncreas	1979	Monitorización de terapia	
Antígeno carcinoembrionario (CEA)	Colon	1965	Recidiva, Monitorización de terapia, metástasis	
ER y PgR	Mama	1970s	Clasificación de pacientes	Tumor
HER2	Mama	1985-86	Seleccionar pacientes para terapia	Tumor/suero
Gonadotrofina corionica humana- beta	Testicular	1938	Diagnostico, estadio, recidiva, terapia	Suero
Lactato deshidrogenasa	Célula germinal	1954	Diagnostico, estadio, recidiva, terapia	
Antígeno específico de próstata (PSA)	Próstata	1979	Cribado, Diagnóstico	Suero
Tiroglobulina	Tiroides	1956	Monitorización	Suero
KIT	Tumor gastrointestinal (GIST)		Diagnostico, selección de terapia	Tumor
EGFR	Colon		Selección de terapia	Tumor

* Adapted from Kulasingam and Diamandis. Nat Clin Pract Oncol. 2008;5:588-99

En cualquier caso se necesitan nuevos biomarcadores con la sensibilidad y especificidad adecuadas. En la actualidad la finalización de un buen número de genomas junto con avances recientes en genómica, proteómica y bioinformática han relanzado el interés por el descubrimiento de nuevos biomarcadores utilizando estas tecnologías de alto rendimiento. Estas tecnologías permiten detectar patrones y marcadores múltiples, más que marcadores únicos, creando un nuevo paradigma en la búsqueda de biomarcadores. La proteómica ha contribuido con un gran número de estrategias al descubrimiento de estos biomarcadores. Sin embargo, este esfuerzo aun no se ha traducido en resultados tangibles en cuanto a marcadores aprobados por las agencias reguladoras. Aunque es previsible que en un corto/medio plazo se materialicen en nuevas aprobaciones.

9. Bases moleculares para la identificación y detección de biomarcadores en cáncer.

Los mecanismos biológicos subyacentes a la aparición de proteínas diferencialmente expresadas entre una muestra biológica tumoral y otra normal pueden deberse a: 1, sobreexpresión génica, 2, aumento en la secreción de proteínas, 3, angiogénesis o mecanismos de invasión y 4, mecanismos de autoinmunidad.

- **1.** El incremento en la expresión de un gen puede deberse a un aumento en el número de copias (amplificación génica) o a un incremento en la actividad transcripcional. Un ejemplo puede ser el EGFR en cáncer colorrectal (65), que se ha utilizado como base para el desarrollo de anticuerpos monoclonales terapéuticos como Cetuximab.

- **2.** Respecto a la secreción de proteínas, alteraciones en el péptido señal de las proteínas pueden resultar en patrones atípicos de secreción, así como procesos de polarización y aumento de proteasas. Ejemplos de proteínas secretadas en la circulación tenemos la alfa-fetoproteína o la proteína HER2 (también conocida como ERBB2). La sobreexpresión de esta proteína está asociada con alto riesgo de recidiva y muerte en canceres de ovario y mama, respectivamente. HER2 es la diana del anticuerpo monoclonal terapéutico Trastuzumab (Herceptin®; Genentech). En el año 2000, la FDA aprobó el uso de un test para la detección de HER2 en suero, que constituye el primer test aprobado para el seguimiento de pacientes con cáncer de mama metastásico.

- **3.** La invasión de un tejido por el tumor podría permitir la liberación directa de moléculas en el fluido intersticial y la liberación subsiguiente por los linfáticos en la sangre. Para los tumores de tipo epitelial, las proteínas deben atravesar la membrana basal antes de aparecer en la sangre. Por ejemplo, el PSA se expresa abundantemente en las células epiteliales columnares prostáticas y se secreta en el lumen glandular, apareciendo en el plasma seminal. En pacientes con carcinoma de próstata, los niveles de PSA aumentan hasta 4-10 ng/ml debido a la rotura de las barreras anatómicas entre el lumen glandular y los capilares sanguíneos. En fases tardías los niveles pueden ir desde 10 ng/ml hasta 1000 ng/ml.

- **4.** Una estrategia que está ganando popularidad se basa en la capacidad que tienen los tumores para inducir una respuesta específica de anticuerpos (ver sección 11). Esta propiedad es particularmente cierta en aquellos tumores epiteliales que cursan con un elevado componente inflamatorio asociado, como por ejemplo cáncer de colon, próstata, ovario o mama. En estos casos los pacientes desarrollan autoanticuerpos que pueden ser utilizados para el diagnóstico del tumor, incluso en fases tempranas antes que se desarrolle el tumor. En el caso del cáncer de próstata, Wang y colaboradores describieron el uso de firmas de autoanticuerpos para mejorar la detección temprana, mediante el uso de una combinación de despliegue en fagos y microarrays de proteínas (66).

10. Estrategias de identificación y tipos de marcadores tumorales.

Existe una amplia variedad de tecnologías proteómicas disponibles para la identificación de proteínas diferencialmente expresadas como marcadores potenciales de cáncer, tanto en fluidos biológicos (suero, plasma, orina…) como en tejidos. Recientemente han aparecido una serie de técnicas de proteómica cuantitativa que han permitido avanzar considerablemente en esta área electroforesis bidimensional acoplada a DIGE (67), cromatografía liquida acoplada a espectrometría de masas utilizando técnicas como SILAC (68,69) o iTRAQ y microarrays de proteínas (33,70). Todas ellas han arrojado un gran número de potenciales marcadores que esperan validación. Existen múltiples revisiones que recogen los principales hallazgos para los diferentes tipos de cáncer. Por citar tan solo algunas, cáncer colorrectal (71), pulmón (72), próstata (73) o páncreas (74). Uno de los problemas recurrentes en la identificación de proteínas diferencialmente expresadas en cáncer y otras patologías mediante técnicas proteómicas, especialmente cuando no se han llevado a cabo exhaustivas labores de pre-fraccionamiento, es la presencia mayoritaria de proteínas celulares muy abundantes como las proteínas de choque térmico, cha-

peronas de varias clases, proteasas, etc. Desafortunadamente muchas de estas proteínas han sido reportadas en múltiples análisis proteómicos de diferentes patologías y constituyen lo que se ha venido en llamar "proteoma de stress celular". Debe llevarse a cabo un gran esfuerzo para separar los auténticos marcadores tumorales de otros marcadores irrelevantes o inespecíficos (75). Otro problema habitual es la identificación de proteínas presentes en varios tipos de tumor, pero no específicas de un tipo de tumor concreto.

A título de ejemplo, utilizando un ensayo multiplex basado en Luminex se analizó un panel de 40 proteínas plasmáticas previamente implicadas en cáncer de pulmón de células no pequeñas para distinguir entre estadio I y controles (76). Se detectaron varias citoquinas con capacidad para distinguir entre ambos casos y suficiente sensibilidad para detectar estadios tempranos. Otro biomarcador candidato (CXCL7) fue detectado utilizando un perfil proteómico diferencial entre sangre venosa pulmonar efluente bañando la vasculatura pulmonar y sangre arterial de los mismos individuos (77). Este marcador podría detectar cáncer de pulmón hasta 30 meses antes del diagnóstico clínico.

Otras estrategias como la peptidómica, basada en tecnologías SELDI-TOF o MALDI-TOF, cosecharon algunos éxitos iniciales (78,79), aunque han generado gran controversia por la baja reproducibilidad de la técnica y la dificultad para identificar las proteínas parentales. Otra técnica de reciente aparición es el uso de imagen basada en espectrometría de masas ("MALDI-Imaging"), utilizando cortes de tejidos frescos congelados que nos da una imagen virtual de la distribución de proteínas en un corte de tejido y a partir de ahí inferir posibles candidatos (80).

El uso de modificaciones post-transduccionales como biomarcadores de cáncer representa un área aun emergente. Niveles altos de fragmentos de calreticulina y la proteína disulfuro isomerasa A3 se han encontrado en el suero de pacientes con carcinoma de hígado que en los controles (81). Por otro lado, la alteración del proceso de glicosilación en cáncer es otra fuente importante de marcadores potenciales. Alteraciones en la fucosilación de la ribonucleasa de suero 1 y otras alteraciones en glicoproteínas han sido reportadas para carcinoma de páncreas (82). La mejora en las técnicas de espectrometría de masas como ETD, que permiten un mejor análisis de las modificaciones lábiles, nos permitirá aumentar nuestro conocimiento en este campo. Otros efectos como el "splicing" alternativo o aberrante pueden dar lugar a nuevos biomarcadores, como el "splicing" aberrante de SLC39A14 en cáncer de colon (83).

11. Uso de autoanticuerpos como biomarcadores.

Se sabe que los pacientes de cáncer desarrollan una respuesta inmune frente al tumor, incluidos autoanticuerpos que podría ser utilizada para el desarrollo de nuevos test diagnósticos (Figura 3). Estos autoanticuerpos han sido utilizados para el diagnóstico en múltiples tipos de cáncer (84,85). El uso de estrategias basadas en microarrays de proteínas ha permitido potenciar el descubrimiento de nuevos autoantígenos en diferentes tipos de cáncer como pulmón (86), colorrectal (70,87), próstata (66), ovario

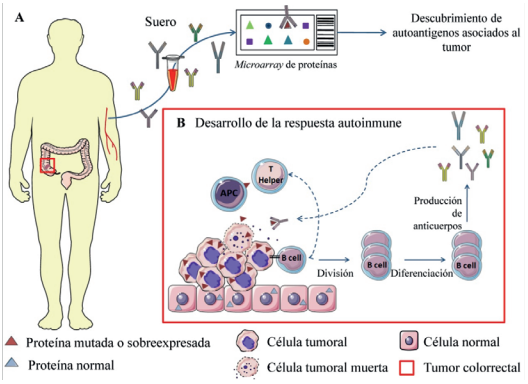


Figura 3. (A) Identificación de Antigenos Asociados a Tumores. (B) Esquema del desarrollo de la respuesta autoinmune frente a un tumor.

(88) y otros. Una de las principales ventajas de esta aproximación es que permite la detección del cáncer en estadios pre-sintomáticos y justifican la realización de estudios prospectivos (Ver también capitulo… Microarrays de Proteínas).

12. Herramientas estadísticas para la validación de biomarcadores candidatos.

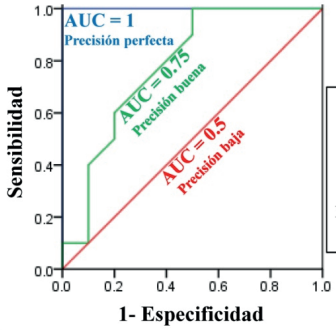
La aplicación de técnicas de alto rendimiento al descubrimiento de biomarcadores conduce al descubrimiento de grandes colecciones de biomarcadores, con el consiguiente porcentaje de falsos descubrimientos. Es por tanto imprescindible diseñar estudios de validación y caracterizar adecuadamente la sensibilidad (positivos verdaderos/ (positivos verdaderos + falsos negativos) y especificidad (verdaderos negativos/ (verdaderos negativos + falsos positivos)) de cada biomarcador candidato. Aquí se describen brevemente algunas de las técnicas estadísticas más utilizadas en estos estudios.

t-test (test de Student).

Se utiliza para evaluar si las medias de los grupos tumoral y control son estadísticamente diferentes entre sí, generalmente usando Microsoft Excel. Con el t-test, probamos la hipótesis de que la media de los valores de ELISA para el grupo tumoral y la media del grupo control sean iguales. Si el p-valor obtenido es superior a 0.05, la hipótesis es verdad con un 95% de probabilidad, si el p-valor es inferior a 0.05, hay 95% de probabilidad que las medias sean estadísticamente diferentes.

Curvas de ROC.

Los datos de ELISA para cada biomarcador individual se pueden evaluar para especificidad y sensibilidad mediante curvas ROC (Figura 4). En el caso de la técnica ELISA se introducen los datos de absorbancia a 450 nm de cada suero indicando a qué grupo pertenecía el suero (1: tumoral, 0: control). El programa calcula para un punto de corte dado los valores de sensibilidad y especificidad para cada marcador, calculando el porcentaje de verdaderos pacientes tumorales y verdaderos pacientes control. Efectúa esta operación utilizando cada valor de absorbancia como punto de corte. Así, se construye la curva ROC a partir del conjunto de datos obtenidos que corresponden a un valor de especificidad y sensibilidad para cada valor de absorbancia y marcador.



$$\textbf{Sensibilidad} = \frac{\text{Num de verdaderos positivos}}{(\text{Num de verdadero positivos} + \text{Num de falsos negativos})}$$

$$\textbf{Especificidad} = \frac{\text{Num de verdaderos negativos}}{(\text{Num de verdaderos negativos} + \text{Num de falsos positivos})}$$

Figura 4. Construcción de una curva ROC. La curva ROC se construye a partir de los valores de sensibilidad y especificidad calculadas a partir de las fórmulas del cuadro derecho. La curva azul representa la curva ROC de una prueba con una precisión perfecta, la curva verde representa la curva ROC de una prueba con una precisión buena y la curva roja indica que la prueba no consigue distinguir los diferentes grupos. Num, número.

Un análisis discriminante usando modelos lineales permite determinar el efecto de la combinación de los biomarcadores y calcular los AUC de dicha combinación (89). El análisis discriminante lineal consiste en analizar el efecto lineal de los parámetros (biomarcadores) sobre la separación entre las clases (1: tumoral, 0: normal). Brevemente, el programa analiza el poder discriminatorio entre las dos clases para cada biomarcador, conserva el mejor de todos y calcula el efecto de cada uno de los demás marcadores sobre su capacidad discriminatoria. El marcador que permite una mejor clasificación, en combinación con el primer seleccionado, se queda en el modelo y se repite el proceso hasta que no haya ningún marcador que mejore la clasificación. Finalmente, se obtiene un modelo final compuesto de los marcadores que permiten la mejor discriminación entre las dos clases. Los AUC y el análisis discriminante se pueden realizar con un programa adecuado (por ejemplo, JMP7 (SAS).

Bootstrapping.
El AUC de las curvas ROC calculado directamente con el modelo original y el conjunto completo de datos tiene un sesgo hacia valores altos, es decir, está sobreestimado. Para obtener un AUC corregido, no sesgado hacia valores altos, y que nos dé una estimación del AUC que podríamos obtener con una validación futura independiente, se puede utilizar bootstrapping con 1000 muestras replicadas (89). Los análisis de bootstrapping se ajustan mediante un modelo de regresión logística donde se modela la probabilidad de ser un paciente con cáncer o un individuo normal como una función de las variables.

13. Uso de SRM como técnica diagnóstica para la validación y cuantificación de biomarcadores en plasma u otros fluidos.

La técnica MRM (o SRM "selected reaction monitoring") ha sido utilizada durante décadas para la cuantificación en plasma de pequeñas moléculas o metabolitos de drogas en laboratorios de referencia. Recientemente se ha adaptado la técnica SRM a la cuantificación de proteínas en plasma o lisados celulares (90). Para ello las proteínas se digieren previamente con tripsina u otras enzimas a sus péptidos. Uno o más de estos péptidos cuya secuencia es única a esa proteína (llamados péptidos proteotípicos) se miden como subrogados de la concentración de proteína en combinación con los mismos péptidos sintéticos y marcados isotópicamente y de concentración conocida (91).

La proteómica dirigida basada en el uso de espectrometría de masas SRM tiene el potencial de una rápida verificación de los candidatos biomarcadores en plasma y, por tanto, cerrar el hueco entre descubrimiento y validación en el desarrollo de biomarcadores. La técnica SRM es extraordinariamente sensible y específica, basada en el uso de triples cuadrupolos, que cubre un rango dinámico de 4-5 órdenes de magnitud y alcanza un límite de cuantificación (LOQ) de 1 µg/ml. Tras la eliminación de las 12 proteínas más abundantes del plasma, el LOQ de SRM puede alcanzar los 25 ng/ml de plasma (92), pudiendo llegar a 2.5 ng/ml tras un fraccionamiento posterior por intercambio iónico. Este rango entraría dentro del rango mínimo aceptable de detección de marcadores en plasma (93). La SRM es una buena alternativa al ELISA como técnica de validación, evitando la necesidad de desarrollar anticuerpos específicos, que suele ser largo, costoso e impredecible.

Los datos obtenidos con este tipo de ensayos se almacenan en la página web del proyecto Peptide Atlas (www.srmatlas.org). Los ensayos SRM pueden ser aplicados directamente a la cuantificación de proteínas en múltiples muestras sin necesidad de desarrollos posteriores. El

repositorio de datos para péptidos proteotípicos en el SRMAtlas ofrece la posibilidad de ensayar clínicamente biomarcadores candidatos en grandes cohortes de pacientes.

14. Desafíos y perspectivas.

Aunque el descubrimiento de biomarcadores ha aumentado exponencialmente en los últimos años, la traslación de estos biomarcadores del laboratorio a la clínica y su aplicación al cribado masivo de poblaciones continúa siendo el gran desafío pendiente. Existen múltiples obstáculos para su aplicación clínica, incluyendo las propias limitaciones técnicas de la proteómica así como los grandes retos de validar combinaciones cada vez más amplias de marcadores sobre series necesariamente grandes de pacientes o el desarrollo de ensayos fácilmente aplicables en la rutina clínica.

Es previsible que en el futuro el uso de estos biomarcadores se realice a través de dispositivos miniaturizados automatizados aplicados a múltiples biomarcadores simultáneamente. Aunque se han realizado grandes avances en los últimos años, el camino a recorrer se prevé largo y difícil.

REFERENCIAS.

1. **Jain K. (Editor).** The Handbook of Biomarkers, Humana Press. 2010.

2. **Vivanco F, Padial LR, Darde VM, et al. Proteomic Biomarkers of Atherosclerosis.** Biomarker Insights, 2008; 12: 3:101-113.

3. **Vivanco F, de la Cuesta F, Barderas MG, Zubiri I, Alvarez-Llamas G. Cardiovascular Proteomics, 341-379, en: Platelet Proteomics, Garcia A, Senis Y, (eds.)** 2011, John Wiley.

4. **De la Cuesta F, Alvarez-Llamas G, Maroto A, et al. A Proteomic focus on the alterations occurring at the human atherosclerotic coronary intima.** Mol Cell Proteomics, 2011, 10:

5. **Alvarez-Llamas G, de la Cuesta F, Barderas ME, Darde V, Padial LR, Vivanco F. Recent advances in atherosclerosis-based proteomics: new biomarkers and a future perspective.** Expert Rev Proteomics. 2008; 5:679-691.

6. **De la Cuesta F, Alvarez-Llamas G, Gil-Dones F, Martin-Rojas T, Zubiri I, Pastor C, Barderas MG, Vivanco F. Tissue Proteomics in atherosclerosis: elucidating the molecular mechanism of cardiovascular diseases.** Expert Rev Proteomics. 2009; 6: 395-409.

7. **Anderson NL, Anderson NG. The human plasma proteome, history, character, and diagnostic prospects.** Mol Cell Proteomics. 2002; 1:845-867.

8. **Farrah T, Deutsch E, Omenn G, et al. A high confidence human plasma proteome reference set with estimated concentration in Peptide Atlas.** Mol Cell Proteomics. 2011; 10: M110.006353.

9. **Omenn G, States DJ, Adamski M et al. Overview of the HUPO plasma proteome project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database.** Proteomics. 2005; 5:3226-3245.

10. **Marimuthu A, O´Meally R, Chaerkady R. et al. A comprehensive map of the human urinary proteome.** J Proteom Res. 2011; 10:2734-2743.

11. **Darde VM, Barderas MG, Vivanco F. Depletion of high-abundance proteins in plasma by immunoaffinity subtraction for two-dimensional Difference gel Electrophoresis.** Methods Mol Biol. 2007; 357: 351-364.

12. **Hanash S, Pitteri SJ, Faca V. Mining the plasma proteome for cancer biomarkers.** Nature. 2008; 452:571-579.

13. **Faca V, Song K, Wang H. et al. A mouse to human search for plasma proteome changes associated with pancreatic tumor development.** PLOS Medicine. 2008;5:e123.

14. **Schutzer S, Liu T, Natelson B. et al. Establishing the proteome of normal human cerebrospinal fluid.** PLoS ONE. 2010; 5:e10980.

15. **Anderson LN. The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum.** Clin Chemistry. 2010; 56:177-185.

16. **Delles C, Schiffer E, Von zur Muhlen C. Urinary proteomic diagnosis of coronary artery disease: identification and clinical validation in 623 individuals.** J Hyperten. 2010; 28: 2316-2322.

17. Mischack H, Delles C, Klein J, Schanstra JP. Urinary proteomics based on capillary electrophoresis-coupled mass spectrometry in kidney disease: discovery and validation of biomarkers and clinical applications. Adv Chronic Kidney Dis 2010, 17.493-506.

18. Delles C, Diez J, Dominiczak AF. Urinary proteomics in cardiovascular disease: achievements, limits and hopes. Proteomics Clin Applic. 2011; 5:222-232.

19. Goog D, Zürbig P, Argilés A. et al. Naturally occrring human Urinary peptides for use in diagnosis of Chronic kidney disease. Mol Cell Proteomics. 2010; 9:2424-2437.

20. Von zur Muhlen C, Schiffer E, Zuerbig P, et al. Evaluation of urine proteome pattern analysis for its potential to reflect coronary artery atherosclerosis in symptomatic patients. J Proteom Res. 2009; 8:335–345.

21. Pfaffe T, Cooper-White J, Beyerlein P, Kostner K, Punyadeera C. Diagnostic potential of saliva: current state and future applications. Clin Chem. 2011; 57: 675-687.

22. Floriano P, Christodoulides N, Miller C. et al. use of saliva-based nano-biochip tests for acute myocardial infarction at the point of care: a feasibility study. Clin Chemistry. 2009, 55:1530-1538.

23. Rifaii N, Gillette M, Carr S. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotech. 2006; 24: 971-983.

24. Surinova S, Schiess R, Hüttenhain R, Cerciello F, Wollscheid B, Aebersold R. On the development of plasma protein biomarkers. J Proteom Res. 2011; 10:5-16.

25. Vivanco F, Martin-Ventura JL, Duran MC, et al. Quest for novel cardiovascular biomarkers by proteomic analysis. J Proteom Res 2005, 4, 1181-1191

26. Koening W, Khuseyinova N. Biomarkers of atherosclerotic plaque instability and rupture. Arterioscler Thromb Vasc Biol. 2007; 27:15-26.

27. Darde V, de la Cuesta F, Gil-Dones F, Alvarez-Llamas G, Barderas MG, Vivanco F. Analysis of the plasma proteome associated with acute coronary syndrome. Does a permanent protein signature exist in the plasma of ACS patients? J. Proteom. Res. 2010; 9:4420-4432.

28. Booth B, Zemmel R. Prospects for productivity. Nat Rev Drug Discov. 2004;3:451-6.

29. Brown D. Unfinished business: target-based drug discovery. Drug Discov Today. 2007;12:1007-12.

30. Rix U, Superti-Furga G. Target profiling of small molecules by chemical proteomics. Nat Chem Biol. 2009;5:616-24.

31. Sawyers CL. The cancer biomarker problem. Nature. 2008;452:548-52.

32. Shah NP, Lee FY, Luo R, Jiang Y, Donker M, Akin C. Dasatinib (BMS-354825) inhibits KITD816V, an imatinib-resistant activating mutation that triggers neoplastic growth in most patients with systemic mastocytosis. Blood. 2006;108:286-91.

33. Madoz-Gurpide J, Canamero M, Sanchez L, Solano J, Alfonso P, Casal JI. A proteomic analysis of cell signaling alterations in colorectal cancer. Mol Cell Proteomics. 2007;6:2150-64.

34. Gulmann C, Sheehan KM, Kay EW, Liotta LA, Petricoin EF, 3rd. Array-based proteomics: mapping of protein circuitries for diagnostics, prognostics, and therapy guidance in cancer. J Pathol. 2006;208:595-606.

35. Petricoin EF, 3rd, Espina V, Araujo RP, et al. Phosphoprotein pathway mapping: Akt/mammalian target of rapamycin activation is negatively associated with childhood rhabdomyosarcoma survival. Cancer Res. 2007;67:3431-40.

36. Ballif BA, Roux PP, Gerber SA, MacKeigan JP, Blenis J, Gygi SP. Quantitative phosphorylation profiling of the ERK/p90 ribosomal S6 kinase-signaling cassette and its targets, the tuberous sclerosis tumor suppressors. Proc Natl Acad Sci U S A. 2005;102:667-72.

37. Ong SE, Schenone M, Margolin AA, et al. Identifying the proteins to which small-molecule probes and drugs bind in cells. Proc Natl Acad Sci U S A. 2009;106:4617-22.

38. Remsing Rix LL, Rix U, Colinge J, et al. Global target profile of the kinase inhibitor bosutinib in primary chronic myeloid leukemia cells. Leukemia. 2009;23:477-85.

39. Rix U, Hantschel O, Durnberger G, et al. Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib reveal novel kinase and nonkinase targets. Blood. 2007;110:4055-63.

40. Bach S, Knockaert M, Reinhardt J, et al. Roscovitine targets, protein kinases and pyridoxal kinase. J Biol Chem. 2005;280:31208-19.

41. Kotake Y, Sagane K, Owa T, et al. Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat Chem Biol. 2007;3:570-5.

42. Wang G, Shang L, Burgett AW, Harran PG, Wang X. Diazonamide toxins reveal an unexpected function for ornithine delta-amino transferase in mitotic cell division. Proc Natl Acad Sci U S A. 2007;104:2068-73.

43. Cecconi D, Donadelli M, Rinalducci S, et al. Proteomic analysis of pancreatic endocrine tumor cell lines treated with the histone deacetylase inhibitor trichostatin A. Proteomics. 2007;7:1644-53.

44. Gonzalez-Santiago L, Alfonso P, Suarez Y, et al. Proteomic analysis of the resistance to aplidin in human cancer cells. J Proteome Res. 2007;6:1286-94.

45. Chapal N, Molina L, Molina F, Laplanche M, Pau B, Petit P. Pharmacoproteomic approach to the study of drug mode of action, toxicity, and resistance: applications in diabetes and cancer. Fundam Clin Pharmacol. 2004;18:413-22.

46. Moller A, Malerczyk C, Volker U, Stoppler H, Maser E. Monitoring daunorubicin-induced alterations in protein expression in pancreas carcinoma cells by two-dimensional gel electrophoresis. Proteomics. 2002;2:697-705.

47. Sinha P, Poland J, Kohl S, et al. Study of the development of chemoresistance in melanoma cell lines using proteome analysis. Electrophoresis. 2003;24:2386-404.

48. Lopez-Farre AJ, Mateos-Caceres PJ, Sacristan D, et al. Relationship between vitamin D binding protein and aspirin resistance in coronary ischemic patients: a proteomic study. J Proteome Res. 2007;6:2481-7.

49. Carey MS, Agarwal R, Gilks B, et al. Functional proteomic analysis of advanced serous ovarian cancer using reverse phase protein array: TGF-beta pathway signaling indicates response to primary chemotherapy. Clin Cancer Res. 2010;16:2852-60.

50. Karapetis CS, Khambata-Ford S, Jonker DJ, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N Engl J Med. 2008;359:1757-65.

51. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med. 2001;344:783-92.

52. Hudis CA. Trastuzumab--mechanism of action and use in clinical practice. N Engl J Med. 2007;357:39-51.

53. Karihtala P, Maenpaa J, Turpeenniemi-Hujanen T, Puistola U. Front-line bevacizumab in serous epithelial ovarian cancer: biomarker analysis of the FINAVAST trial. Anticancer Res. 2010;30:1001-6.

54. Fong PC, Boss DS, Yap TA, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med. 2009;361:123-34.

55. Shayesteh L, Lu Y, Kuo WL, et al. PIK3CA is implicated as an oncogene in ovarian cancer. Nat Genet. 1999;21:99-102.

56. Okano T, Kondo T, Fujii K, et al. Proteomic signature corresponding to the response to gefitinib (Iressa, ZD1839), an epidermal growth factor receptor tyrosine kinase inhibitor in lung adenocarcinoma. Clin Cancer Res. 2007;13:799-805.

57. Taguchi F, Solomon B, Gregorc V, et al. Mass spectrometry to classify non-small-cell lung cancer patients for clinical outcome after treatment with epidermal growth factor receptor tyrosine kinase inhibitors: a multicohort cross-institutional study. J Natl Cancer Inst. 2007;99:838-46.

58. Bauer JA, Chakravarthy AB, Rosenbluth JM, et al. Identification of markers of taxane sensitivity using proteomic and genomic analyses of breast tumors from patients receiving neoadjuvant paclitaxel and radiation. Clin Cancer Res. 2010;16:681-90.

59. Marko-Varga G, Ogiwara A, Nishimura T, et al. Personalized medicine and proteomics: lessons from non-small cell lung cancer. J Proteome Res. 2007;6:2925-35.

60. Hanash S. Building a foundation for the human proteome: the role of the Human Proteome Organization. J Proteome Res. 2004;3:197-9.

61. Hanash SM, Baik CS, Kallioniemi O. Emerging molecular biomarkers-blood-based strategies to detect and monitor cancer. Nat Rev Clin Oncol. 2011;8:142-50.

62. Fung ET. A recipe for proteomics diagnostic test development: the OVA1 test, from biomarker discovery to FDA clearance. Clin Chem. 2010;56:327-9.

63. Jones HB. On a New Substance Occurring in the Urine of a Patient with Mollities Ossium. Phil. Trans. R. Soc. 1848;138:55-62.

64. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat Clin Pract Oncol. 2008;5:588-99.

65. Moroni M, Veronese S, Benvenuti S, et al. Gene copy number for epidermal growth factor receptor (EGFR) and clinical response to antiEGFR treatment in colorectal cancer: a cohort study. Lancet Oncol. 2005;6:279-86.

66. Wang X, Yu J, Sreekumar A, et al. Autoantibody signatures in prostate cancer. N Engl J Med. 2005;353:1224-35.

67. Alfonso P, Nunez A, Madoz-Gurpide J, Lombardia L, Sanchez L, Casal JI. Proteomic expression analysis of colorectal cancer by two-dimensional differential gel electrophoresis. Proteomics. 2005;5:2602-11.

68. Ong SE, Blagoev B, Kratchmarova I, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002;1:376-86.

69. Luque-Garcia JL, Martinez-Torrecuadrada JL, Epifano C, Canamero M, Babel I, Casal JI. Differential protein expression on the cell surface of colorectal cancer cells associated to tumor metastasis. Proteomics. 2010;10:940-52.

70. Babel I, Barderas R, Diaz-Uriarte R, Martinez-Torrecuadrada JL, Sanchez-Carbayo M, Casal JI. Identification of tumor-associated autoantigens for the diagnosis of colorectal cancer in serum using high density protein microarrays. Mol Cell Proteomics. 2009;8:2382-95.

71. Barderas R, Babel, I. and J I Casal. Colorectal cancer proteomics, molecular characterization and biomarker discovery. Proteomics Clin Appl. 2010;4:159-78.

72. Cho JY, Sung HJ. Proteomic approaches in lung cancer biomarker development. Expert Rev Proteomics. 2009;6:27-42.

73. Larkin SE, Zeidan B, Taylor MG, et al. Proteomics in prostate cancer biomarker discovery. Expert Rev Proteomics. 2010;7:93-102.

74. Cecconi D, Palmieri M, Donadelli M. Proteomics in pancreatic cancer research. Proteomics. 2011;11:816-28.

75. Wang P, Bouwman FG, Mariman EC. Generally detected proteins in comparative proteomics-a matter of cellular stress response? Proteomics. 2009;9:2955-66.

76. Kellar KL, Douglass JP. Multiplexed microsphere-based flow cytometric immunoassays for human cytokines. J Immunol Methods. 2003;279:277-85.

77. Yee J, Sadar MD, Sin DD, et al. Connective tissue-activating peptide III: a novel blood biomarker for early lung cancer detection. J Clin Oncol. 2009;27:2787-92.

78. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet. 2002;359:572-7.

79. Villanueva J, Shaffer DR, Philip J, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest. 2006;116:271-84.

80. Caprioli RM. Deciphering protein molecular signatures in cancer tissues to aid in diagnosis, prognosis, and therapy. Cancer Res. 2005;65:10642-5.

81. Chignard N, Shang S, Wang H, et al. Cleavage of endoplasmic reticulum proteins in hepatocellular carcinoma: Detection of generated fragments in patient sera. Gastroenterology. 2006;130:2010-22.

82. Barrabes S, Pages-Pons L, Radcliffe CM, et al. Glycosylation of serum ribonuclease 1 indicates a major endothelial origin and reveals an increase in core fucosylation in pancreatic cancer. Glycobiology. 2007;17:388-400.

83. Thorsen K, Mansilla F, Schepeler T, et al. Alternative splicing of SLC39A14 in colorectal cancer is regulated by the Wnt pathway. Mol Cell Proteomics. 2011;10:M110 002998.

84. Tan HT, Low J, Lim SG, Chung MC. Serum autoantibodies as biomarkers for early cancer detection. FEBS J. 2009;276:6880-904.

85. Casal JI, Barderas R. Identification of cancer autoantigens in serum: toward diagnostic/prognostic testing? Mol Diagn Ther. 2010;14:149-54.

86. Qiu J, Choi G, Li L, et al. Occurrence of autoantibodies to annexin I, 14-3-3 theta and LAMR1 in prediagnostic lung cancer sera. J Clin Oncol. 2008;26:5060-6.

87. Babel I, Barderas R, Diaz-Uriarte R, et al. Identification of MST1/STK4 and SULF1 proteins as autoantibody targets for the diagnosis of colorectal cancer by using phage microarrays. Mol Cell Proteomics. 2011;10:M110 001784.

88. Chatterjee M, Mohapatra S, Ionan A, et al. Diagnostic markers of ovarian cancer by high-throughput antigen cloning and detection on arrays. Cancer Res. 2006; 66: 1181-90.

89. Visintin I, Feng Z, Longton G, et al. Diagnostic markers for early detection of ovarian cancer. Clin Cancer Res. 2008;14: 1065-72.

90. Barr JR, Maggio VL, Patterson DG, Jr., et al. Isotope dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. Clin Chem. 1996;42:1676-82.

91. Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol. 2008;4:222.

92. Keshishian H, Addona T, Burgess M, Kuhn E, Carr SA. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. Mol Cell Proteomics. 2007;6: 2212-29.

93. Schiess R, Wollscheid B, Aebersold R. Targeted proteomic strategy for clinical biomarker discovery. Mol Oncol. 2009;3: 33-44.

MS Imaging

Francesc J. Márquez
Bruker Española S.A.

1. Introducción.

El objetivo de la Proteómica es identificar las proteínas que se producen en las células o tejidos y conocer su función biológica. Por esa razón es necesaria la utilización de herramientas analíticas de una gran potencialidad, como es la espectrometría de masas. En este sentido, la técnica MS Imaging permite la investigación de la distribución espacial de péptidos o proteínas, u otro tipo de moléculas biológicas directamente sobre un tejido. Para ello normalmente se utilizan espectrómetros de masas de iones secundarios (SIMS) o bien espectrómetros de masas con fuente MALDI (Desorción e Ionización por Láser Asistida por Matrices) y en menor medida espectrómetros de masas con fuente DESI (Desorción e Ionización mediante Electroespray).

Estas técnicas se han desarrollado para proporcionar información química y espacial principalmente para diferentes clases de biomoléculas en tejidos (péptidos, proteínas, metabolitos, lípidos, etc.). El objetivo es obtener esta información con la mayor resolución espacial sin comprometer la distribución espacial original de las moléculas de interés, con una sensibilidad que permita obtener información biológica relevante y sin la necesidad de usar reactivos de fluorescencia, radioactividad o inmunohistoquímica [1].

La mayoría de las técnicas utilizadas para identificar los compuestos biológicos directamente sobre tejido como la fluorescencia o la inmunohistoquímica suelen estar limitadas a un número de compuestos. Además frecuentemente se necesita un conocimiento previo de las moléculas de interés y por lo tanto son menos efectivas como herramienta para descubrir nuevos compuestos.

El MS Imaging tiene la ventaja de poder realizar análisis complejos de múltiples componentes, desde metabolitos o fármacos a proteínas de tamaño medio, con una resolución espacial de 50 mm o incluso mejor [2].

En esencia la técnica de MS-Imaging se basa en la adquisición de un espectro de masas en cada punto dentro de un área definida sobre un tejido, obteniendo así una colección de espectros de masas de toda el área. Los espectros obtenidos se promedian y a partir del espectro promedio es posible trazar la intensidad relativa de un determinado a través del tejido y visualizar la distribución de cada uno de los iones moleculares presentes en los espectros (Figura 1). Este tipo de análisis se realiza mediante un mapa de pixels de intensidades moleculares en el cual se puede visualizar la intensidad de cada analito mediante un código de colores a semejanza de un mapa de calor y compararlo con la imagen histológica (Figura 2).

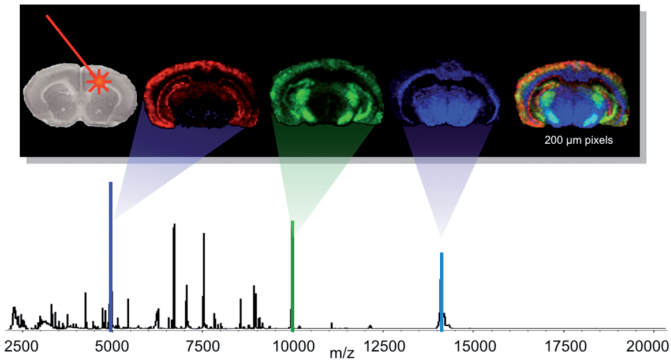


Figura 1: Adquisición de los espectros de masas en cada punto dentro de un área definida del tejido. A partir del espectro promedio se localizan los analitos en el tejido.
Utilizadas con el permiso de Bruker.

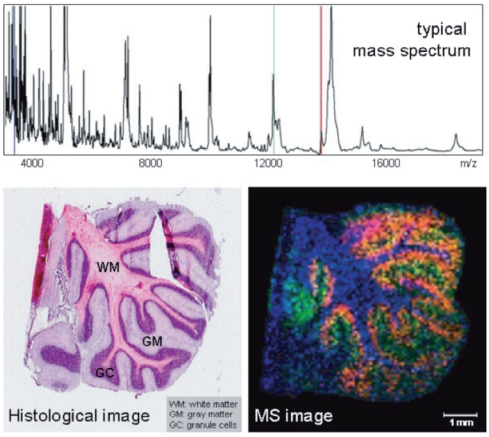


Figura 2: Visualización de las intensidades de cada analito mediante un código de colores y comparararación con la imagen histológica.
Utilizadas con el permiso de Bruker.

Los primeros estudios de MS Imaging se realizaron mediante espectrometría de masas con microsondas láser (LMMS) y análisis de masas con microsondas láser (LAMMA), disponibles de manera comercial en los últimos años de la década de los 70 [3]. La LMMS se utilizaba tanto para el análisis de muestras biológicas como para muestras inorgánicas. Un pulso de láser ultravioleta (UV) focalizado se usaba para dersorber e ionizar muestras sólidas sin matriz y se implementaron diferentes tipos de analizadores como por ejemplo el Tiempo de Vuelo (TOF) o la Resonancia Ciclotrónica de Iones por Transformada de Fourier (FT-ICR) [4-5]. En la técnica LAMMA el láser se utilizaba para generar pulsos de luz e ionizar los analitos que eran acelerados hacia el interior del espectrómetro de masas. Esta técnica se usaba para el análisis de muestras en polvo, en las que las moléculas se excitan a un estado de ionización mediante el haz de láser focalizado [6,7]. El desarrollo de la técnica SIMS a finales de la década de los 60, combinada con la alta sensibilidad de los instrumentos TOF, hizo del TOF-SIMS una poderosa herramienta para la técnica MS Imaging [8]. Fue posteriormente cuando los MALDI-TOF se utilizaron para experimentos de MS Imaging permitiendo un rango de masas mayor que el TOF-SIMS y ampliar la técnica al análisis de péptidos y proteínas [9].

En analizador de tiempo de vuelo (TOF) es el de elección tanto para MALDI como SIMS para realizar MS Imaging. La velocidad, sensibilidad y amplio rango de masas (m/z de aprox. 1-100.000) lo hacen especialmente atractivo para esta aplicación. Últimamente ha habido diferentes desarrollos instrumentales para incrementar la velocidad de análisis, la resolución de los espectros, la exactitud de masa, capacidad de realizar ensayos de MS/MS y resolución espacial. La incorporación de la movilidad iónica a los espectrómetros de masas ha añadido a la técnica la posibilidad de un paso de separación de las moléculas en muestras complejas en fase gas

previo al análisis y permite una mejor resolución de las mismas. El desarrollo de fuentes MALDI con mayor velocidad de repetición del láser y menos energéticas permiten no sólo reducir el tiempo de barrido del tejido en el análisis, sino que además permiten la tinción con hematoxilina y eosina del mismo corte de tejido que se ha analizado en la fuente MALDI. Esto permite correlacionar sin ambigüedad los resultados obtenidos mediante MS-Imaging y la tinción con hematoxilina y eosina. Así mismo, el desarrollo de dispositivos que permiten focalizar el haz del láser y modular su energía hace aumentar la resolución espacial de la técnica y maximizar la ionización. El desarrollo de las fuentes MALDI también permite realizar MS-Imaging en analizadores de alta resolución y exactitud de masa como los FT-ICR o los FT-Orbitrap capaces de separar en un espectro de masas los iones con la misma masa nominal y establecer una asignación de las fórmulas empíricas de pequeñas moléculas y lípidos de manera fiable. Además este tipo de equipos permiten realizar diferentes técnicas de MS/MS.

Aún con el desarrollo de la instrumentación, la preparación de la muestra es un factor extremadamente importante y determinante para conseguir la mayor información química y mayor resolución espacial posible [10,11].

2. Métodos de ionización para MS Imaging.

2.1. Secondary Ion mass spectrometry (SIMS).

La espectrometría de masas SIMS utiliza la alta energía (5-25 KeV) de partículas cargadas de un haz de iones primarios, como por ejemplo Ar^+, Ga^+, In^+, Cs^+, que se aceleran y se hacen impactar sobre la superficie de la muestra. La energía cinética asociada de los iones primarios se transfiere a la muestra tras el impacto. El proceso de transferencia de energía afecta a sólo a las capas externas de la muestra, típicamente 10 Å. Esta transferencia de energía induce una cascada de colisiones entre átomos y moléculas de la superficie que se liberan de la superficie cuando su energía cinética es mayor que la energía de enlace del sustrato formando iones (iones secundarios) y estos iones son los que proporcionan la información cualitativa y cuantitativa sobre la composición molecular de la muestra [8, 12-14]

La técnica SIMS principalmente desorbe e ioniza elementos y pequeñas moléculas. El rango de de masas utilizado está limitado a m/z 1.000 aproximadamente como resultado de una amplia fragmentación de la superficie [8], aunque el uso de una matriz orgánica como las utilizadas para MALDI puede superar este límite y mejora la detección para masas más elevadas. El punto fuerte de esta técnica es la posibilidad de proporcionar una elevada resolución espacial.

En SIMS se hace una distinción entre el modo estático y modo dinámico. Estas dos técnicas difieren en el grado de daño sobre la superficie y se usan para propósitos diferentes. El SIMS estático usa una dosis de iones primarios más baja y minimiza la interacción de los iones a la mono-capa de moléculas en la parte superior. Cada ion primario golpea sobre una región inalterada y la interacción entre el haz de iones primarios y los átomos y moléculas de la muestra se limita a menos de un 1% [15].

El SIMS dinámico es más destructivo ya que se usa una dosis de iones primarios mayor que resulta en interacciones en las capas más profundas de la superficie a estudiar. Las aplicaciones para las que se utilizan las técnicas son diferentes: El SIMS estático se utiliza para imaging cualitativo [16] mientras que el SIMS dinámico se utiliza para estudios de imagen cuantitativos elementales [17,18].

2.2. MALDI.

La técnica MALDI se ha utilizado de manera habitual para el estudio de moléculas grandes como proteínas, péptidos, polímeros y lípidos debido a la propiedad de la técnica para generar iones macromoleculares intactos. El uso de la ionización MALDI en estudios de imagen sobre muestras biológicas para el análisis de péptidos y proteínas directamente desde secciones de tejido fue introducida en 1997 por el Profesor Caprioli y sus colaboradores [9, 19, 20]. El mecanismo de formación de iones en MALDI es un fenómeno complejo y los procesos de desorbción y formación de iones han sido ampliamente estudiados [21, 22].

Un experimento de MS Imaging con MALDI requiere de la muestra colocada en una superficie conductiva a la cual se le añade matriz para favorecer la ionización. La adición de una matriz a la superficie de una muestra cumple con diferentes propósitos, la extracción de las moléculas de analito del tejido, la formación de los cristales de la misma juntamente con el analito y la absorción de la energía del láser para una ionización suave de las moléculas de la muestra. Una vez depositada la matriz, la muestra se irradia con un láser pulsado, que induce la desorbción y la ionización de los cristales, realizando un barrido de la superficie a analizar. Los iones formados en estado gaseoso se separan según su relación m/z por el analizador de masa. De esta manera, la técnica MALDI permite analizar áreas de tejido con una resolución espacial de hasta 20 mm y con un rango de masas de 500 a más de 100.000 m/z [2].

2.3. DESI.

La técnica de ionización electrospray por desorbción (DESI) se produce mediante la interacción de gotas cargadas generadas por electrospray con la superficie de la muestra. El impacto de las gotas con la superficie produce una segunda generación de gotas cargadas con moléculas de la superficie a analizar disueltas en el interior. Estas gotas evolucionan en el proceso de electrospray para generar los iones en forma gaseosa que son dirigidos hacia el interior del espectrómetro de masas [23, 24]. El ángulo entre la pluma del electrospray que contiene los analitos y la entrada al espectrómetro de masas debe ser optimizado para permitir la entrada del mayor número posible de iones al analizador. Los iones medidos desde la superficie de los tejidos por DESI son principalmente lípidos monocargados, por lo que su uso se limita principalmente a estudios de lipidómica [24, 25, 26] Esta técnica, en general proporciona una resolución espacial menor que las técnicas SIMS o MALDI aunque no necesita un tratamiento previo de la muestra.

3. Instrumentación.

Durante los últimos años ha habido diferentes desarrollos técnicos con el objetivo de incrementar la velocidad de análisis, la resolución de masa, la exactitud de masa, la capacidad de realizar experimentos de MS/MS y la resolución espacial en el campo del MS-Imaging.

Mientras los espectrómetros de masas con analizadores de tiempo de vuelo (TOF) proporcionan una gran velocidad de análisis para MALDI Imaging, el poder de resolución de masas es menor que con otras técnicas, lo que puede ocasionar pérdida de información química y espacial. Así, hay otras técnicas establecidas o en desarrollo para contrarrestar esta limitación como pueden ser la movilidad iónica, el MALDI FT-ICR o el MALDI-Orbitrap.

La separación por movilidad iónica permite una separación de los iones en estado gas con la misma masa nominal que provienen de la superficie del tejido según su movilidad y por tanto en la diferencia de velocidades bajo la acción de un campo eléctrico y con la acción de un gas tampón a contracorriente. El tiempo de migración de cada ión depende de la masa, la carga, la forma y el tamaño de cada ión proporcionando la capacidad de distinguir entre los diferentes

analitos. Cuanto mayor es la superficie del ión, más área tiene el gas tampón para colisionar y por tanto el ión necesita de más tiempo para llegar al final del tubo de movilidad iónica [27,28]. Esta técnica permitiría separar los diferentes grupos de moléculas según su estructura tridimensional y mejorar su análisis sobre el tejido.

Los espectrómetros de masas de tipo FT-ICR u Orbitrap con fuente MALDI son otros de los instrumentos utilizados debido a que presentan un elevado poder de resolución de masas. Así el FT-ICR puede resolver e identificar iones con la misma masa nominal que se están presentes de manera común en espectros de MALDI-Imaging [26, 29, 30]. Además, la elevada exactitud de masa de estos equipos (mejor de 1 ppm) y su capacidad para realizar análisis de MS/MS se ha utilizado para identificar con un elevado grado de confianza pequeñas moléculas, fármacos y sus metabolitos sobre tejidos biológicos [30, 31, 32].

Otro de los objetos de mejora de la técnica de MS Imaging ha sido la resolución espacial. En el caso del SIMS, como se ha comentado anteriormente, la resolución espacial es elevada por el pequeño tamaño de haz de iones. Sin embargo, en MALDI la resolución espacial no sólo está limitada por el tamaño del foco del láser sino también por el tamaño de los cristales de matriz depositados en el tejido.

En los últimos años se han experimentado con nuevos dispositivos ópticos para evitar la limitación causada por el tamaño del foco de láser como por ejemplo el microscopio de masas estigmático [1,33]. Con este sistema la resolución espacial que se puede obtener es independiente del tamaño del área de desorbción/ionización. O bien emplear un objetivo coaxial para disminuir el foco de laser a un micrómetro o mejor [34]. Sin embargo, los nuevos modelos comerciales de MALDI-TOF y TOF/TOF ya permiten modular el foco del láser llegando por debajo de cinco micrómetros. También se han desarrollado tecnologías que proporcionan una modulación del haz del laser y obtener un pulso de láser que funciona perfectamente bien con todas las matrices y maximiza la ionización de la muestra [35].

Otro aspecto importante es la velocidad de disparo del láser. Si tenemos en cuenta una sección de tejido de 1 cm^2 y la medimos con una resolución espacial de 50 micrómetros, se requieren 40.000 puntos de disparo y de entre 100 a 1000 disparos de laser en cada punto. Con un láser de nitrógeno estándar a 20 Hz el tiempo para recolectar todos los datos estaría por encima de 55 horas, lo que es poco práctico. La introducción de los láseres de Nb:YAG con velocidades de repetición de entre 200 1000 Hz en instrumentos comerciales ha hecho la técnica más práctica aumentando la velocidad de medida.

Respecto al tamaño de los cristales de la matriz, la preparación de la muestra es crítica a la hora de obtener buena calidad de resultados en cuanto a reproducibilidad y resolución.

4. Preparación de la muestra.

El flujo de trabajo para la preparación de muestra consta de diferentes fases: recolección del tejido, el corte de tejido y la deposición de la matriz en el caso de utilizar las técnicas MALDI o SIMS (Figura 3).

4.1 Recolección del tejido.

El primer paso crítico para la preparación de la muestra es la recolección y almacenamiento de las mismas, puesto que se necesita un procedimiento que mantenga la integridad y la

localización espacial de las moléculas en el tejido original. Como en todos los tipos de muestras biológicas, una vez cortado el tejido se produce la degradación y la modificación de las moléculas en un proceso que continuará en el tiempo. Esta evolución a nivel proteómico puede ser que sea más lenta que a nivel metabólico, pero aún así es rápida y varía lo suficiente como para que la preservación y procesado del tejido deba ser reproducible y bajo unas condiciones estándar. En cada paso, pequeñas variaciones experimentales en la manipulación del tejido pueden producir cambios en los perfiles de biomarcadores que pueden afectar al análisis por MS-Imaging en el transcurso del experimento.

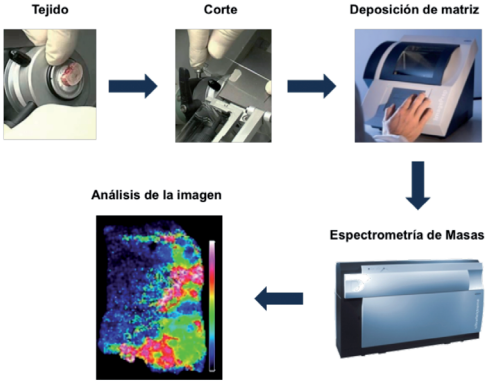


Figura 3. Preparación de la muestra para MS Imaging.
Utilizadas con el permiso de Bruker.

Muchos de los procedimientos básicos se han desarrollado de acuerdo con los métodos establecidos y usados comúnmente en histología, desde el corte del tejido hasta el punto en que el éste se fija en el porta-objetos.

El protocolo más comúnmente empleado para el almacenamiento del tejido es congelarlo inmediatamente después de su extracción y mantenerlo guardado a -80ºC en isopentano líquido o nitrógeno líquido hasta su análisis. De esta manera se minimiza la degradación de las proteínas. Otra posibilidad es usar metanol o isopropanol a temperaturas de -70ºC o menores. Hay que tener en cuenta que es importante mantener la forma del tejido en este paso para facilitar la obtención de réplicas biológicas del mismo tejido. Sin embargo, incluso el proceso de congelado se pueden producir roturas y fragmentaciones de partes del tejido, lo que dificultaría el corte y la comparación de estas réplicas. Para mantener la forma del tejido suele utilizarse papel de aluminio o film de plástico para envolver la pieza extraída. Sin embargo, si se requieren múltiples secciones del mismo tejido para realizar el análisis o para la validación de los experimentos es recomendable realizar todas las secciones del tejido que se necesite en el mismo momento para evitar la manipulación sucesiva del tejido y que puedan producirse procesos de oxidación o de deterioro del mismo. Para estudiar de manera lo más exacta posible el proteoma durante varios estadios de una enfermedad, respuesta a fármacos o desarrollo de un organismo, puede ser necesaria la desnaturalización de todos los encimas proteolíticos para prevenir una rápida degradación de la muestra.

El método de desnaturalización más comúnmente usado en técnicas inmunohistoquímicas para prevenir la degradación post-mortem de proteínas y péptidos en el tejido es la fijación del mismo en paraformaldehído y embeberlo en parafina. Sin embargo esta manera de conservar el tejido no es la más adecuada para la técnica de MS Imaging puesto que se forman interacciones entre proteínas por la formación de puentes de metileno [36, 37] entre ellas. Aunque se han realizado estudios de LC-MS y LC-MS/MS desde tejido embebido en parafina mediante un protocolo de desparafinización y posterior digestión enzimática [38, 39]. También se están desarrollando métodos para realizar análisis con MALDI directamente sobre tejidos conservados en bloques de parafina [40]. La desnaturalización de las proteínas sin perder la morfología del tejido se puede realizar también por calor. El método se basa en un calentamiento rápido de la muestra de tejido que evita la actividad enzimática.

4.2 Cortes de tejido.

Para realizar el corte de los tejidos es importante tener en cuenta el grosor de la sección, la

temperatura y el uso de medios para embeber el tejido. Los tejidos congelados se cortan en secciones finas en un microtomo refrigerado para montarlas sobre un portaobjetos. El grosor óptimo de las muestras es entre 10 y 20 micrómetros para asegurar que suficientes moléculas de analito estarán disponibles para su extracción. El corte suele realizarse a una temperatura de -5 a -20ºC dependiendo del tipo de tejido, los tejidos más grasos necesitan menores temperaturas que los tejidos de alto contenido acuoso.

Las técnicas de patología tradicional requieren de manera usual embeber el tejido en un material de soporte para realizar los cortes de tejido como la parafina. El caso ideal es disponer de un bloque de tejido congelado lo suficientemente grande para no necesitar ningún soporte que pueda interferir en la ionización, sin embargo cuando las piezas de tejido son pequeñas, puede ser necesario un material de soporte que ayude en el corte. Para los tejidos no fijados que necesitan ser cortados en un micrótomo refrigerado, el soporte utilizado para facilitar el corte suele ser el medio OCT. Sin embargo, al cortar el tejido se forma una capa de OCT sobre el mismo y actúa como un supresor de iones. Se ha comprobado que el medio OCT ioniza muy fácilmente en MALDI proporcionando espectros con un gran número de picos de contaminación [11,41] enmascarando las moléculas de interés, por lo que se recomienda no utilizar este medio. Incluso restos de OCT en el microtomo puede conducir a espectros de mala calidad. Cuando es inevitable embeber el tejido en un medio de soporte, se puede utilizar algún material blando no polimérico como por ejemplo la gelatina, que es compatible para el análisis por espectrometría de masas. El tejido se embebe en 10% de gelatina a 30ºC inmediatamente después de la disección y se congela a -80ºC previniendo el daño sobre el tejido.

El tejido seccionado puede colocarse directamente sobre una superficie conductiva, suele utilizarse una placa de MALDI de metal o bien porta-objetos de vidrio de superficie recubierta con oxido de titanio e indio (ITO).

Una vez montado el tejido sobre la superficie puede ser necesario realizar algún tratamiento de la muestra como por ejemplo, pasos de lavado, digestiones enzimáticas o la deposición de matriz.

Para el análisis de péptidos o proteínas se recomienda realizar pasos de lavado para eliminar de la superficie del tejido lípidos y sales. Un método común es el lavado con etanol 70-80% a 4ºC, seguido de un lavado con etanol al 90-100% para ayudar a completar la deshidratación del tejido, fijarlo y prevenir la degradación del proteoma [11,42.43]. Es importante retirar las sales para que la cristalización de la matriz produzca cristales homogéneos, incrementando la eficiencia de la ionización de péptidos y proteínas. Además las sales y los lípidos ionizan fácilmente pudiendo causar la supresión de iones y la saturación del detector. Para aquellos tejidos de alto contenido en lípidos un lavado con cloroformo puede ser recomendable para eliminarlos [44].

Hay que tener en cuenta que protocolos de lavado más drásticos o aumentar el número de lavados pueden dañar el tejido y eliminar los péptidos de interés.

También pueden realizarse digestiones enzimáticas sobre el tejido (por ejemplo una digestión tríptica) para mejorar la detección e identificación de las proteínas. Este tipo de digestiones suele realizarse porque generalmente no es posible detectar las proteínas más grandes mediante MS-Imaging. El límite de masa para una detección razonablemente sensible está en torno a los 25-30 kDa, dependiendo de la instrumentación usada. La digestión ***in situ*** de estas proteínas sobre el tejido permite localizar las proteínas con mayor masa. Los péptidos formados están en relación directa con la abundancia y la localización de las proteínas de las que provienen en el tejido y por lo tanto la información sobre la localización y la abundancia de las proteínas vendrá determinada por los péptidos detectados. Además la fragmentación de los péptidos en el espectrómetro de masas permite también tener una mayor posibilidad de identificación de proteínas desconocidas directamente desde el tejido [45].

La digestión *in situ* sobre el tejido requiere que éste mantenga cierto grado de humedad para mantener la actividad enzimática, pero hay que tener en cuenta que esto aumenta la posibilidad de la difusión de los péptidos durante la incubación. La enzima puede aplicarse sobre el tejido mediante un espray que cubra todo el tejido o bien mediante una gota sobre las zonas del tejido de interés cuando no es necesaria una resolución espacial elevada. La ventaja en este último caso es un mayor control sobre las condiciones de incubación y la difusión de los péptidos.

4.3 Selección y deposición de la matriz sobre el tejido para MALDI y SIMS.

Así como para la ionización DESI no es necesario el uso de una matriz para el análisis de la muestra mediante MS-Imaging, ésta sí que es necesaria para la ionización MALDI y para SIMS, en el caso de utilizar esta técnica aumentando su rango de masas hasta m/z 2000 y poder analizar péptidos.

La deposición de la matriz para MALDI y ME-SIMS (Matrix-enhanced SIMS) es un paso crucial en el protocolo de preparación de la muestra para MS Imaging. La deposición de la matriz debe ser homogénea, reproducible, proporcionar suficiente sensibilidad y de fácil uso. Además, el tamaño de los cristales de la matriz determina el máximo de resolución espacial del análisis.

Así mismo, es importante elegir la matriz adecuada al tipo de moléculas de interés en el análisis para conseguir buenos resultados. Las matrices más utilizadas en MS Imaging y su aplicación se resumen en la siguiente tabla:

Matriz	Abreviatura	Aplicación
Ácido 2,5-Dihidroxibenzoico [46-48]	DHB	Péptidos, proteínas pequeñas, glicoproteínas, glicopéptidos, azúcares.
Acido a-Ciano-4-hidroxicinámico [48,49]	HCCA	Péptidos, proteínas pequeñas, glicopéptidos
Acido 3,5-Dimetoxi-4-hidroxicinamico (Acido sinapínico) [48, 50, 51]	SA	Péptidos, proteínas grandes, lípidos
Acido 3-Hidroxipicolínico [48, 52]	3-HPA	Péptidos, glicoproteínas, oligonucleótidos
2,4-Dinitrofenilhidrazina [53]	2,4-DNPH	Péptidos, proteínas
HCCA/anilina [54]		Péptidos
HCCA/N,N-dimetilanilina [54]		Péptidos
HCCA/2-amino-4-metil-5-nitropiridina [54]		Péptidos

La deposición de la matriz debe ser considerada con atención. Las condiciones ambientales (humedad, presencia de oxigeno, etc.) durante la deposición pueden influir en la cristalización y la interacción entre la matriz y los analitos. Por ejemplo, si la matriz llega a una superficie deshidratada o casi deshidratada la extracción de los analitos por parte de la matriz es más difícil. A continuación se detallarán los métodos de aplicación de matriz más utilizados:

- **Deposición de gotas manualmente:** La disolución de matriz sobre la muestra se aplica manualmente mediante una pipeta sobre las regiones del tejido que se desea. Es un método simple y rápido ideal para el análisis de un perfil de péptidos o proteínas, pero no es un método recomendado para obtener imágenes de calidad. Puede darse un efecto de difusión dentro del punto de matriz, el tamaño del punto es grande y la distribución de los cristales de matriz es irregular.

- **Nebulización neumática:** La disolución de matriz se aplica sobre la muestra mediante un pulverizador o un aerógrafo de los utilizados habitualmente en aplicaciones de cromatografía de capa fina [11, 55]. Un espray suave permite la formación de una capa de cristales homogénea en la superficie del tejido. Sin embargo, el espray tiene que ser optimizado previamente a cada aplicación y la reproducibilidad puede ser un problema. En este sentido, una buena reproducibilidad es muy importante en las aplicaciones biomédicas. Este requisito ha sido un factor determinante en el desarrollo de sistemas de deposición de matriz automáticos. Stoeckli y sus colaboradores, desarrollaron un sistema de nebulización neumática controlando las condiciones ambientales de temperatura y humedad obteniendo buena homogeneidad [56].

- **Sistema piezoeléctrico:** Pequeñas gotas de la disolución de matriz se aplican con un dispositivo piezoeléctrico a modo de una impresora de inyección de tinta. Esta técnica permite la deposición de pequeños volúmenes por punto (100 pL) por ciclo. Cada punto de matriz tiene un diámetro de unos 150 micrómetros o incluso menores dependiendo de la cantidad total de matriz aplicada por cada ciclo. La deposición es precisa, altamente reproducible y las gotas son uniformes [43, 57, 58]. Debido al tamaño de la gota, su uso es más adecuado para cuando no se necesita una gran resolución espacial . Para realizar MS Imaging con mayor resolución el tamaño de los cristales de matriz debe ser igual o menor que el tamaño de un solo pixel.

- **Sistema de vaporización vibracional (ImagePrep™ de Bruker Daltonics):** El sistema produce un aerosol de matriz creado por vaporización vibracional bajo condiciones controladas en un sistema cerrado. El sistema incorpora un sensor que mide la luz dispersada por la capa de matriz depositada en la muestra de tejido lo que proporciona un control sobre el tamaño de la gota. Esto permite limitar el efecto de difusión en la superficie y el tiempo de interacción entre la matriz y los analitos. Els sitema puede optimizarse para una máxima co-cristalización. El equipo es ideal para la aplicación de pequeñas cantidades de matriz (10 mL como mucho) y proporciona una capa de matriz homogénea con cristales de diámetro entre 20 y 50 micrómetros. Además, el sistema puede utilizarse también fácilmente para aplicar enzimas para realizar experimentos de digestión sobre el tejido [2].

Existen otros métodos de aplicación de matriz, pero únicamente se han descrito para el análisis de lípidos debido a que el tiempo de interacción entre el analito y la matriz es limitado. Estas metodologías son:

- **Sublimación de matriz:** La sublimación de la matriz se realiza para evitar el uso de disolventes. El objetivo es eliminar la posibilidad de difusión de las moléculas de analito en el tejido durante la deposición del la matriz. La sublimación ocurre a bajas presiones y altas temperaturas (aproximadamente a 125ºC) lo que permite a la matriz condensar en la superficie más fría de la muestra. Este método asegura incrementar la pureza de la matriz y la reducción del tamaño de los cristales de la misma y aumentar la resolución de la imagen. Esta técnica es muy recomendable para SIMS debido al pequeño tamaño de los cristales [59].

- **Matriz molida:** En este método, la matriz finamente molida se aplica sobre el tejido sin utilizar disolventes y utilizando un cedazo de acero inoxidable libre de estaño de 20 micrómetros. Como en el caso de la sublimación, el efecto de la difusión es limitado debido a la ausencia de disolventes [60].

La elección del método de deposición de la matriz es pues determinante para el tipo de aplicación que se quiera perseguir. Los métodos automáticos son mejores para cuando se necesita analizar un gran número de muestras y se necesita un alto rendimiento. Los métodos manuales son más adecuados cuando se quieren desarrollar nuevos protocolos.

5. Análisis de datos.

Una vez se ha adquirido la imagen, el siguiente paso es siempre identificar las especies que muestran una localización específica en el tejido o la distribución de las mismas. Conjuntamente con el desarrollo del trabajo experimental de MS Imaging, se ha desarrollado en los últimos años software para generar imágenes que permitan correlacionar espectros de masas con una posición (Figura 2). Entre ellos están el BioMap de Novartis, FlexImaging de Bruker Daltonics y AMOLF's "datacube". También grupos de investigación han desarrollado su propio software de análisis [61]. Hay que tener en cuenta que los espectros de masas generados en un experimento de MS Imaging muestran en general alrededor de 400 o 500 picos y el tamaño de los ficheros obtenidos pueden variar entre 100 Mbytes hasta pocos Gbytes por muestra, dependiendo de la resolución espacial y del analizador de masas utilizado. Así pues, la interpretación manual de los datos se convierte en una tarea ardua y dificultosa además de necesitar mucho tiempo. Para facilitar este trabajo se ha desarrollado software basado en herramientas estadísticas para ayudar a la interpretación de los datos.

El análisis de componentes principales (PCA) [62, 63] es un tipo de análisis estadístico que se utiliza ampliamente cuando existen múltiples variables. Por lo que este tipo de análisis es fácilmente aplicable a los datos obtenidos en los análisis de Imaging. El análisis de componentes principales es un método estadístico no supervisado que se usa para identificar grupos que presenten variables correlacionadas muy próximas entre sí, en el caso del MS Imaging se utilizan las coordinadas espaciales y la masa e intensidad de las señales en el espectro de masas.

Otro tipo de análisis utilizado es el análisis semántico latente probabilístico (pLSA) que proporciona conjuntos de datos que combinan la reducción de ruido y la clasificación automática del tejido [64, 65].

También se ha utilizado el "clustering" jerárquico [66] en el que los conjuntos de datos se agrupan por similitud, obteniendo un gráfico en forma de árbol que separa clases para cada espectro. Las ramas del árbol que están más alejadas presentan mayores diferéncias que las que están más cercanas.

El MS Imaging se ha aplicado recientemente a un amplio abanico de tejidos y análisis. Mientras que al principio de la técnica era suficiente el poder ver una distribución de un marcador en una sección de tejido, en la actualidad resulta imprescindible mostrar una relevancia bioquímica revelando la identidad de alguno de los marcadores y su validación como se ha mostrado en diversos trabajos publicados.

REFERENCIAS.

1. M. A. F. Altelaar, I. M. Taban, L. A. McDonnell, P. D. E. M. Verhaert, et al., Int. J. Mass Spectrom. 2007, 260, 203–211.
2. M. Schuerenberg, C. Luebbert, S.O. Deininger, R. Ketterlinus, D. Suckau, Nat. Methods 4 (2007) iii
3. F. Hillenkamp, E. Unsöld, R. Kaufmann, R. Nitsche, Appl. Phys. A: Mater. Sci. Process. 8. 1975, 341.
4. H. Struyf, W. van Roy, L. van Vaeck, R. van Grieken, R. Gijbels, Anal. Chim. Acta. 1993, 283, 139.
5. L. van Vaeck, H. Struyf, W. van Roy, F. Adams, Mass Spectrom. Rev. 1994, 13, 189.
6. H. Vogt, H.J. Heinen, S. Meier, R. Wechsung, Fresen. J. Anal. Chem. 1981, 308, 195.
7. H.J. Heinen, S. Meier, H. Vogt, R. Wechsung, Int. J. Mass Spectrom. Ion Phys. 1983, 47, 19.
8. ToF-SIMS: Surface Analysis by Mass Spectrometry, IM Publications and Surface Spectra Limited, Chichester, Manchester, 2001.

9. R.M. Caprioli, T.B. Farmer, J. Gile, Anal. Chem. 1997, 69, 4751.

10. B.K. Kükrer, I.M. van der Wiel, J. Stauber, C. Güzel, J.M. Kros, T.M. Luider, R.M.A. Heeren, Proteomics 2009, 9, 2622.

11. S.A. Schwartz, M.L. Reyzer, R.M. Caprioli, J. Mass Spectrom. 2003 38 699.

12. A. Adriaens, R. van Ham, L. van Vaeck, in: J.C. Vickerman, D. Briggs (Eds.), ToFSIMS: Surface Analysis by Mass Spectrometry, IM Publications and Surface Spectra Limited, Chichester, Manchester. 2001, p. 195.

13. A. Delcorte, in: J.C. Vickerman, D. Briggs (Eds.), ToF-SIMS: Surface Analysis by Mass Spectrometry, IM Publications and Surface Spectra Limited, Chichester, Manchester. 2001, p. 161.

14. H.M. Urbassek, in: J.C. Vickerman, D. Briggs (Eds.), ToF-SIMS: Surface Analysis by Mass Spectrometry, IM Publications, Chichester. 2001, p. 139.

15. T.L. Colliver, C.L. Brummel, M.L. Pacholski, F.D. Swanek, A.G. Ewing, N. Winograd, Anal. Chem. 1997, 69, 2225.

16. G.B. Eijkel, B.K. Kükrer, I.M. van der Wiel, J.M. Kros, T.M. Luider, R.M.A. Heeren, Surf. Interface Anal. 2009.

17. S. Chandra, G.H. Morrison, Int. J. Mass Spectrom. Ion Process. 1995, 143, 161.

18. S. Chandra, D.R. Smith, G.H. Morrison, Anal. Chem. 2000, 72, 104A.

19. P. Chaurand, M. Stoeckli, R. M. Caprioli. Anal. Chem. 1999, 71, 5263.

20. M. Stoeckli, T. B.Farmer, R. M. Caprioli, J. Am. Soc. Mass Spectrom. 1999, 10, 67.

21. R. Knochenmuss. Analyst. 2006, 131, 966.

22. K. Dreisewerd, Chem. Rev. 2003, 103, 395.

23. Z. Takats, J.M.Wiseman, B. Gologan, R.G. Cooks, Science. 2004, 306, 471.

24. R.G. Cooks, Z. Ouyang, Z. Takats, J.M. Wiseman, Science. 2006, 311, 1566.

25. J.M.Wiseman, D.R. Ifa, A. Venter, R.G. Cooks, Nat. Protoc. 2008, 3, 517.

26. J. Pol, V. Vidova, G. Kruppa, V. Kobliha, P. Novak, K. Lemr, T. Kotiaho, R. Kostiainen, V. Havlicek, M. Volny, Anal. Chem. 2009, 81, 8479.

27. J.A. McLean, W.B. Ridenour, R.M. Caprioli, J. Mass Spectrom. 2007, 42, 1099.

28. S.N. Jackson, M. Ugarov, T. Egan, J.D. Post, D. Langlais, J.A. Schultz, A.S. Woods, J. Mass Spectrom. 2007, 42, 1093.

29. D.S. Cornett, S.L. Frappier, R.M. Caprioli, Anal. Chem. 2008, 80, 5648.

30. I.M. Taban, A.F.M. Altelaar, Y.E.M. Van der Burgt, L.A. McDonnell, R.M.A. Heeren, J. Fuchser, G. Baykut, J. Am. Soc. Mass Spectrom. 2007, 18, 145.

31. T.K. Sinha, S. Khatib-Shahidi, T.E. Yankeelov, K. Mapara, M. Ehtesham, D.S. Cornett, B.M. Dawant, R.M. Caprioli, J.C. Gore, Nat. Methods 5 (2008) 57.

32. E.A. Stemmler, C.R. Cashman, D.I. Messinger, N.P. Gardner, P.S. Dickinson, A.E. Christie, Peptides 2007, 28. 2104.

33. S. L. Luxembourg, T. M. Mize, L. A. McDonnell, R. M. A. Heeren, Anal. Chem. 2004, 76, 5339.

34. M. Koestler, D. Kirsch, A. Hester, A. Leisner, S. Guenther, B. Spengler, Rapid Commun. Mass Spectrom. 2008, 22, 3275.

35. A. Holle, A. Haase, M. Kayser, J. Hohndorf, J.Mass Spectrom. 2006, 41, 705.

36. S. S. Rubakhin, W. T. Greenough, J. V. Sweedler, Anal. Chem. 2003, 75, 5374.

37. J. Stauber, R. Lemaire, J. Franck, D. Bonnel, et al., J. Proteome Res. 2008, 7, 969.

38. D. K. Crockett, Z. Lin, C. P. Vaughn, M. S. Lim, K. S. J. Elenitoba-Johnson, Lab. Invest. 2005, 85, 1405.

39. D. E. Palmer-Toy, B. Krastins, D. A. Sarracino, J. B. J. Nadol, S. N. Merchant, J. Proteome Res. 2005, 4, 2404–2411.

40. R. Lemaire, A. Desmons, J. C. Tabet, R. Day, R. et al., J. Proteome Res. 2007, 6, 1295.

41. P. Chaurand, J. L. Norris, D. S. Cornett, J. A.Mobley, R. M. Caprioli, J. Proteome Res. 2006, 5, 2889.

42. P. Chaurand, S. A. Schwartz, D. Billheimer, B.J. Xu, et al. Anal. Chem. 2004, 76, 1145.

43. H.R. Aerni, D. S. Cornett, R. M. Caprioli. Anal. Chem. 2006, 78, 827.

44. R. Lemaire, M. Wisztorski, A. Desmons, J. C. Tabet, et al. Anal. Chem. 2006, 78, 7145.

45. M. R. Groseclose, M. Andersson, W. M. Hardesty, R. M. Caprioli, J. Mass Spectrom. 2007, 42, 254.

46. I. Fournier, C. Marinach, J.C. Tabet, G. Bolbach, J. Am. Soc. Mass Spectrom. 2003, 14, 893.

47. K. Strupat, M. Karas, F. Hillenkamp, Int. J. Mass Spectrom. Ion Process. 1991, 111, 89.

48. A. Tholey, E. Heinzle, Anal. Bioanal. Chem. 2006, 386, 24.

49. R.C. Beavis, T. Chaudhary, B.T. Chait, Org. Mass Spectrom. 1992, 27, 156.

50. R.C. Beavis, B.T. Chait, K.G. Standing, Rapid Commun. Mass Spectrom. 1989, 3, 436.

51. I. Fournier, J.C. Tabet, G. Bolbach, Int. J. Mass Spectrom. 2002, 219, 515.

52. C. Pan, S. Xu, H. Zhou, Y. Fu, M. Ye, H. Zou, Anal. Bioanal. Chem. 2007, 387, 193.

53. R. Lemaire, A. Desmons, J.C. Tabet, R. Day, M. Salzet, I. Fournier, J. Proteome Res. 2007, 6, 1295.

54. R. Lemaire, J.C. Tabet, P. Ducoroy, J.B. Hendra, M. Salzet, I. Fournier, Anal. Chem. 2006, 78, 809.

55. A.F.M. Altelaar, I. Klinkert, K. Jalink, R.P.J. de Lange, R.A.H. Adan, R.M.A. Heeren, S.R. Piersma, Anal. Chem. 2006 78, 734.

56. M. Stoeckli, D. Staab, S. Capretta, 54th American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics, Seattle, WA, USA. 2006.

57. R.L. Adams, J. Roy, J. Appl. Mech. 1986, 53, 193.

58. T. Nakanishi, I. Ohtsu, M. Furuta, E. Ando, O. Nishimura, J. Proteome Res. 2005, 4, 743.

59. J.A. Hankin, R.M. Barkley, R.C. Murphy, J. Am. Soc. Mass Spectrom. 2007, 18, 1646.

60. S.M. Puolitaival, K.E. Burnum, D.S. Cornett, R.M. Caprioli, J. Am. Soc. Mass Spectrom. 2008, 19, 882.

61. E. Astigarraga, G. Barreda-Gómez, L. Lombardero, O. Fresnedo, F. Castaño, M. T. Giralt, B. Ochoa, R. Rodríguez-Puertas, J. A. Fernández. Anal. Chem. 2008, 80, 9105.

62. Y.-L. Chou, Statistical analysis, with business and economic applications, Holt, Rinehart and Winston, New York. 1975, p. 17.9.

63. M.E. Wall, A. Rechtsteiner, L.M. Rocha, in: D.P. Berrar, W. Dubitzky, M. Granzow (Eds.), A Practical Approach to Microarray Data Analysis, Kluwer Academic Publishers. 2003, p. 91.´

64. M. Hanselmann, M. Kirchner, B.Y. Renard, E.R. Amstalden, K. Glunde, R.M.A. Heeren, F.A. Hamprecht, Anal. Chem. 2008, 80, 9649.

65. M. Hanselmann, U. Köthe, M. Kirchner, B.Y. Renard, E.R. Amstalden, K. Glunde, R.M.A. Heeren, F.A. Hamprecht, J. Proteome Res. 2009, 8, 3558.

66. S.O Deninger, M.P. Ebert, A Fütterer,† M. Gerhard, C. Röcken. J. Proteome Res. 2008, 7, 5230.

Estrategias en proteómica de alto rendimiento (high throughput). Caracterización de proteomas. LC-MS/MS. FDR

Joaquin Abian[1], Marina Gay[1], Montserrat Carrascal[1]

[1] Laboratorio de Proteómica CSIC/UAB, Instituto de Investigaciones Biomédicas de Barcelona. Consejo Superior de Investigaciones Científicas.

1. Introducción.

La rápida evolución de las herramientas analíticas utilizadas en proteómica en la última década ha permitido la identificación de un número cada vez mayor de moléculas en un único análisis. Paralelamente, el significado del término 'proteómica a gran escala' se ha ido adaptando a este aumento del potencial analítico.

A finales de los años 90, las estrategias de análisis proteómico a gran escala se basaban en la utilización de geles bidimensionales (2D) y métodos de mapeo peptídico utilizando técnicas de MALDI-TOF. Inicialmente, el número de proteínas identificadas en estos trabajos pioneros no superaba las pocas decenas. Así, en uno de los estudios más relevantes de la época, dirigido al análisis a gran escala del proteoma de la levadura, se reportaba la identificación de tan sólo 135 manchas de proteínas [1]. En años posteriores, la aparición de robots para el picado y digestión automatizada de las manchas en el gel permitió aumentar de forma importante la eficiencia de estos procedimientos. Sin embargo, la utilización de geles 2D para la separación de proteínas determinaba una buena parte de las limitaciones de la técnica (p.e. rangos de tamaño molecular y de punto isoeléctrico analizables) haciendo evidente la necesidad de la implementación de estrategias alternativas [2]. El desarrollo de las técnicas de *shotgun proteomics* basadas en la digestión de proteomas y su análisis directo mediante cromatografía capilar multidimensional acoplada a la espectrometría de masas en tándem permitió solventar muchos de los problemas derivados del análisis en geles bidimensionales. Uno de los primeros trabajos realizados mediante esta aproximación, describió la identificación de cerca de 1400 proteínas del proteoma de la levadura en un único análisis, una cantidad que implicaba un aumento de un orden de magnitud respecto a las técnicas basada en geles 2D. El conjunto de proteínas identificadas incluía además proteínas que por sus características fisicoquímicas o por su abundancia relativa eran raramente identificadas en estrategias basadas en la utilización de geles 2D [3]. La eficiencia de estos procedimientos *shotgun* ha ido aumentando con la disponibilidad de nuevos sistemas de cromatografía capilar comerciales y de espectrómetros de masa cada vez más rápidos y de mayor resolución. Actualmente esta estrategia constituye el método estándar para la descripción de proteomas en la mayoría de laboratorios y ha sido la base para la identificación y cuantificación del primer proteoma completo de una célula modelo, la levadura [4].

Estas nuevas tecnologías conllevan, sin embargo, sus propios problemas y limitaciones. La información obtenida en un análisis *shotgun* se refiere a pequeñas secuencias peptídicas cuya asignación a una proteína específica no siempre es posible debido a que la secuencia puede ser compartida por proteínas diferentes. Es por tanto necesario el desarrollo de procedimientos que permitan inferir la colección de proteínas presentes en la muestra a partir de la información péptido-céntrica disponible. Esta limitación es todavía más grave cuando se requiere un análisis cuantitativo, ya que la cantidad total de un péptido determinado puede ser fruto de la contribución de proteínas o fragmentos proteicos diferentes que se encuentran en proporciones diferentes en las muestras. Por otro lado, la interpretación de las decenas de miles de espectros generados en un estudio de *shotgun proteomics* constituye otro de los retos más importantes en su aplicación. Los resultados obtenidos mediante cualquiera de los algoritmos disponibles para la identificación *de novo* o en bases de datos de las colecciones de espectros experimentales debe estudiarse para eliminar falsas identificaciones. Dada la imposibilidad práctica de un análisis manual de estas colecciones, se han desarrollado procedimientos para su depurado automático y para la determinación del grado de calidad de la colección de datos resultante.

2.- *Shotgun proteomics*.

En 1995 el grupo de John Yates III describía una estrategia para identificar proteínas a gran escala a partir de los espectros de masas en tándem derivados del ión correspondiente a uno de los péptidos de digestión de la proteína y de su comparación con los espectros teóricos derivados de las secuencias anotadas en bases de datos [5]. El procedimiento podía aplicarse tanto a la identificación de una única proteína aislada mediante cromatografía o electroforesis en gel, como a la de una mezcla compleja de proteínas mediante el análisis exhaustivo de la colección de péptidos obtenidos después de su digestión. Esta última estrategia es la base de las técnicas de análisis de proteomas que se denomina *shotgun proteomics* (traducido como proteómica de perdigonada) (Figura 1). La capacidad de esta tecnología para el análisis a gran escala de proteomas fue demostrada posteriormente por el mismo grupo en varios trabajos en los que se estudiaron el proteoma de la levadura y del arroz y donde los autores desarrollaron un sistema analítico que bautizaron como MudPIT (*Multidimensional Protein Identification Technology)* [3, 6].

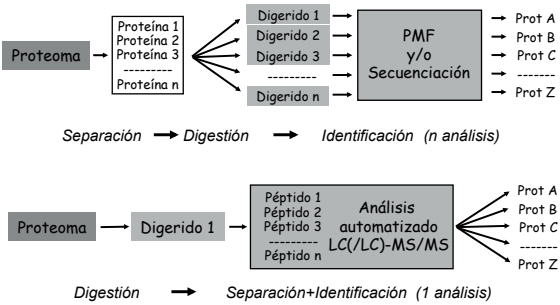


Figura 1.- Análisis *shotgun*. A diferencia de los métodos clásicos donde se identifica independientemente cada una de las proteínas presentes en la muestra una vez separadas, los métodos *shotgun* analizan la muestra directamente después de la digestión enzimática de sus componentes. El análisis shotgun normalmente requiere una o más etapas de separación acopladas a la espectrometría de masas.

El método de separación por excelencia en *shotgun proteomics* es la cromatografía líquida. A diferencia de las técnicas de separación en geles bidimensionales, los sistemas basados en la cromatografía líquida son reproducibles y fácilmente automatizables y pueden acoplarse directamente al análisis espectrométrico.

En la configuración actualmente más estándar, los péptidos de digestión separados en un sistema cromatográfico provisto de una columna nano-capilar (50-100 µm di) de fase reversa son

analizados en línea mediante instrumentos para espectrometría de masas en tándem con fuentes de microelectrospray (también denominadas como nanospray de flujo continuo o simplemente nanospray) [7]. Aunque la aplicación de la cromatografía capilar o nanocromatografía no es común en otros ámbitos de la química analítica, esta técnica constituye una herramienta imprescindible en el análisis proteómico. La principal ventaja de estos sistemas para cromatografía líquida-espectrometría de masas (LC/MS) deriva de la alta sensibilidad obtenida mediante el uso de flujos cromatográficos inferiores al microlitro por minuto utilizando fuentes de microspray [8]. La disponibilidad de columnas y sistemas comerciales para cromatografía capilar fue durante mucho tiempo una de las principales limitaciones para su uso generalizado ya que requería la adaptación de los sistemas convencionales disponibles y la construcción casera de columnas e interfases [8, 9]. Actualmente esta disponibilidad ha dejado de ser un problema y es posible encontrar en el mercado una variedad de instrumentación y de columnas de diferentes tamaños y con diferentes soportes. Aun así, las técnicas de cromatografía capilar no han adquirido todavía la reproducibilidad y robustez de los sistemas analíticos convencionales principalmente debido a los problemas derivados del trabajo a nanoflujos. Así, el efecto de los volúmenes muertos en estos sistemas y la dificultad para determinar y controlar el flujo de disolvente son dos de los problemas más comunes en el trabajo diario con estas técnicas. Por otro lado, el volumen de inyección óptimo en estos sistemas es muy pequeño, situándose en el orden de los nanolitros. Con objeto de poder trabajar con volúmenes mayores, que permitan además inyecciones más reproducibles, suelen emplearse cartuchos con rellenos de fase reversa o intercambio iónico que, acoplados en línea a la columna analítica mediante un sistema de válvulas, permiten desalar y preconcentrar la muestra antes de su desorción y separación en la columna cromatográfica. Estos cartuchos suelen tener un diámetro interno superior al de la columna analítica lo que permite además utilizar flujos mayores durante la inyección y preconcentración de la muestra y por tanto reducir de forma importante el tiempo total de análisis.

El número de péptidos analizable mediante un sistema LC/MS depende de parámetros cromatográficos (número de platos, capacidad de carga) y espectrométricos (principalmente, velocidad de barrido, resolución de masa y sensibilidad).

El término capacidad de separación de picos (*peak capacity*) es un parámetro cromatográfico que se utiliza para indicar el número de compuestos que pueden separarse con resolución unidad en un tiempo determinado [10]. Los sistemas cromatográficos basados en columnas capilares convencionales (50-100 µm di, 10-20 cm de largo, partículas de 3-5 µm) proveen de una capacidad de separación de aproximadamente 100 a 300 en tiempos de cromatografía de 60-90 minutos. El detector espectrométrico por su parte permite resolver compuestos que coeluyen en la cromatografía aportando una dimensión de separación adicional. Una trampa iónica lineal moderna es capaz de realizar alrededor de 50 espectros de secuenciación en el tiempo de elución de un pico cromatográfico (aprox. 20-30 segundos) de forma que la capacidad de separación de picos combinada se encontraría en el orden de 10^4. Considerando los aproximadamente 21,000 genes del genoma humano y que cada secuencia proteica genera una media de 50 péptidos de digestión tríptica, el número de componentes a separar se encontraría en el orden de 10^6. Aunque no todos los genes están expresados en una célula en un momento determinado, dado que el cálculo anterior no tiene en cuenta la complejidad adicional derivada del procesamiento alternativo de mRNAs, de las modificaciones prostraduccionales de las proteínas, de la ocurrencia de digestiones parciales o no canónicas, interferencias no peptídicas, etc, se hace evidente que el análisis de un proteoma completo o sub-proteoma supone una dura prueba para estas tecnologías.

La capacidad de separación cromatográfica puede mejorarse reduciendo el tamaño del soporte cromatográfico y aumentando la longitud de la columna y el tiempo de separación. Por ejemplo, utilizando columnas nanocapilares de 50 µm de diámetro interno y de 20 cm de longitud rellenas con partículas de 0.8 µm de diámetro, es posible alcanzar capacidades de separación superiores a 400 en cromatogramas de 50 minutos [11]. Acopladas a un analizador de trampa

iónica lineal, este tipo de columnas permitió identificar, a partir de digeridos del proteoma del microorganismo *Shewanella oneidensis* y en un único análisis de 50 minutos, unos 4000 péptidos trípticos correspondientes a alrededor de 1000 proteínas diferentes [11]. En las condiciones empleadas para este análisis en concreto, el factor limitante fue la velocidad de barrido del espectrómetro de masas. Dado que las nuevas generaciones de espectrómetros proveen de velocidades entre 2-3 veces superiores a las utilizadas en los estudios anteriores, es probable que este mismo análisis produjera actualmente un número todavía mayor de identificaciones.

Otra posibilidad para mejorar la capacidad de separación del sistema cromatográfico es la reducción del diámetro de la columna capilar. Con columnas de 15 µm de diámetro interno rellenas con partículas de 3 µm y trabajando a flujos de disolvente de 20 nl/min ha sido posible conseguir capacidades de separación del orden de 10^3. El bajo flujo utilizado con este tipo de columnas produce además un incremento importante de la sensibilidad de la fuente de nanospray situando el límite de detección por debajo de 75 zeptomols de proteína individual mediante monitorización por FT-ICR y de 7 attomols mediante análisis MS/MS en una trampa lineal [12]. Así, la monitorización de masas peptídicas mediante un analizador FT-ICR acoplado a un sistema de este tipo permitió un recubrimiento del 14% del proteoma teórico del microorganismo *Deinococcus radiodurans* en un único análisis y a partir de únicamente 2.5 ng de muestra [12].

Tanto la utilización de partículas submicrométricas como la de capilares de diámetro inferior a 20 µm requieren el uso de sistemas especiales de nanocromatografía capaces de trabajar a presiones muy altas (10-20 kpsi). Si bien recientemente han aparecido sistemas comerciales para ultra-HPLC (uHPLC) capaces de producir capacidades de pico de alrededor de 450 con soportes de 1.7 µm y columnas de 75 µm x 25 cm trabajando a 9500 psi, la no disponibilidad comercial durante mucho tiempo de esta instrumentación ha hecho que hasta ahora su aplicación quedara restringida a un número limitado de laboratorios.

2.1 Cromatografía multidimensional, MDLC.

Para aumentar la capacidad de separación de los sistemas LC/MS se recurre a la utilización de métodos de separación cromatográfica multidimensional (MDLC) donde la muestra se somete a dos o más procesos cromatográficos consecutivos que utilizan mecanismos de separación de selectividad diferente (por lo que se utiliza el término separación ortogonal).

En MDLC los diferentes procesos de separación pueden aplicarse acoplados en línea o en etapas separadas. El término 'sistema integrado' aplicado a los sistemas MDLC es la traducción del término ingles '*comprehensive MDLC*', que fue acuñado por Bushey y Jorgenson para referirse a sistemas LC/LC en línea donde todo el eluído de una columna es transferido a una segunda unidad, diferenciándose así de aquellos sistemas en los que solo una o unas pocas fracciones seleccionadas son analizadas en una segunda etapa [13]. Los sistemas bidimensionales integrados utilizan un sistema de válvulas para el lavado y derivación automática de los componentes separados en la primera dimensión hacia la segunda etapa cromatográfica. En los sistemas *off-line*, por su parte, el material eluido de la primera etapa cromatográfica se recoge manualmente o mediante un colector de fracciones para posteriormente, cada una de estas fracciones, ser reinyectada en el sistema LC/MS. Cada una de estas dos aproximaciones tiene sus ventajas e inconvenientes.

Cuando se cumple la condición de ortogonalidad entre las diferentes etapas, la capacidad de separación de picos del sistema multidimensional corresponde al producto de las capacidades de separación de cada una de las etapas independientes [14, 15]. La capacidad de separación obtenida en un sistema LC/LC es muy inferior a este valor teórico dado que, a diferencia del acoplamiento LC/MS donde el muestreo por el espectrómetro de masas del material transferido desde la primera dimensión es continuo y el análisis muy rápido, en el acoplamiento LC/LC se trabaja típicamente con un número limitado de fracciones que acumulan (vuelven a mezclar)

el material eluído durante tiempos relativamente largos. De esta forma, en estos sistemas, la capacidad de separación corresponde al producto de la capacidad de separación de la última columna y del número de fracciones o ciclos de separación que se llevan a cabo [15]. A pesar de ello, incluso los sistemas LC/LC más primitivos eran capaces de producir capacidades de separación superiores a 10^3 [16, 17].

Los procedimientos de análisis multidimensional generan muestras menos complejas y más adecuadas al poder de resolución de la columna y a la velocidad de barrido del espectrómetro. Por otro lado, es posible partir de cantidades mayores de material a fraccionar, aumentándose por tanto la capacidad de carga del sistema y, en consecuencia, la posibilidad de identificación de proteínas minoritarias.

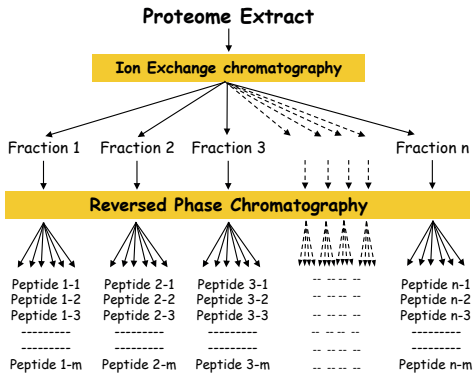


Figura 2.- Cromatografía Líquida Multidimensional. La adición de etapas cromatográficas en el análisis de los componentes de una mezcla compleja provee de un mayor poder de separación. En condiciones óptimas, la capacidad de separación de picos del sistema multidimensional corresponde al producto de las capacidades de separación de cada una de las etapas independientes (en el ejemplo de la figura, m x n)

Actualmente, en proteómica, la combinación utilizada generalmente es la que acopla una primera etapa basada en una columna de intercambio catiónico fuerte (SCX) seguida de otra de fase reversa. En esta configuración, la utilización del intercambio iónico provee de una primera etapa con una alta capacidad de carga mientras que la fase reversa permite el lavado efectivo de sales y otros contaminantes polares, aporta una buena resolución en la separación de péptidos y se adapta perfectamente al análisis por electrospray. Estos sistemas incluyen una precolumna de fase reversa que se utiliza para la preconcentración de los componentes antes de ser eluidos hacia la segunda columna.

Los primeros sistemas de cromatografía bidimensional integral (en línea) acoplados a la espectrometría de masas (2DLC/MS) aplicados al análisis de mezclas de proteínas y de digeridos proteicos fueron descritos por el grupo de Moseley. Estos sistemas utilizaban una columna de intercambio catiónico [18] o de exclusión molecular [19] acoplada a una columna de fase reversa mediante un sistema de válvulas. En el primer caso se utilizaron columnas capilares tanto en la etapa de intercambio iónico (750 µm id x 12 cm) como en la de fase reversa (500 µm d.i. x10 cm). Excepto por la falta de una columna de preconcentración intermedia y la utilización de un divisor de flujo (1/10) a la salida de la columna de fase reversa, que trabajaba a 50 µl/min, para adaptarla al flujo de entrada de la fuente de electrospray, la configuración de estos sistemas era muy similar a la utilizada actualmente.

La capacidad de separación calculada en estos primeros trabajos era relativamente baja. El sistema LC/LC proveía de capacidades de separación de solo 512 y en conjunto, considerando el espectrómetro de masas, de 2500. La gran potencialidad de los sistemas acoplados MDLC/MS, sin embargo, fue demostrada poco después con la descripción del sistema MudPIT [20] que probablemente marcó el inicio de la generalización de esta estrategia en el análisis proteómico.

2.2 MudPIT.

El sistema MudPIT integraba un sistema de cromatografía multidimensional acoplada a la espectrometría de masas con un sistema automatizado para la búsqueda de los espectros de fragmentación en una base de datos de secuencias proteicas.

Este sistema se describió inicialmente como parte de la estrategia denominada DALP (*Direct Analysis of Large Protein Complexes*) que estos autores aplicaron al estudio de los complejos ribosomales 80S y 40S de *Saccharomyces* [20]. El sistema cromatográfico DALP se construyó inicialmente utilizando columnas microbore (1 mm d.i.) de intercambio iónico y de fase reversa. Posteriormente, con objeto de aumentar la sensibilidad del análisis, el sistema se miniaturizó utilizando una columna capilar bifásica (100 µm d.i.) que actuaba además como aguja de electrospray.

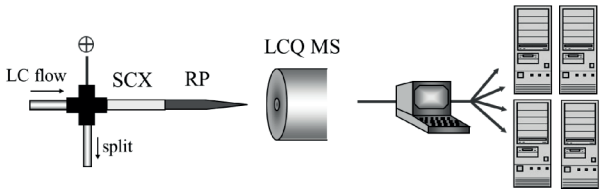


Figura 3 La tecnología MudPIT. En el sistema MudPIT original, la columna cromatográfica, una columna bifásica conteniendo soportes cromatográficos de intercambio iónico y fase reversa, era a la vez la aguja de electrospray. El material a separar se depositaba en la cabeza de esta columna. El material absorbido en la fase de intercambio iónico se eluía hacia el soporte de fase reversa, mediante inyecciones de una solución acuosa de sal. Los componentes retenidos en el soporte de fase reversa se separaban y eluían hacia el espectrómetro, donde eran analizados, mediante un gradiente de disolvente orgánico. La utilización de varios de estos ciclos de elución con concentraciones crecientes de sal permitía analizar de forma fraccionada el total de los componentes de la muestra.

En la columna capilar bifásica, los soportes cromatográficos de intercambio catiónico fuerte (SCX) y de fase reversa estaban dispuestos en línea (figura 3). La mezcla de péptidos de digestión acidificada era introducida en la columna de forma que los péptidos quedaban retenidos en la fase SCX. Mediante un aumento de la concentración de sal en el eluyente, una fracción de este material era desorbida hacia la fase reversa. Los péptidos quedaban retenidos en la fase reversa de la que eran eluídos directamente hacia el espectrómetro de masas mediante un gradiente de acetonitrilo. Este procedimiento se repetía varias veces utilizando concentraciones crecientes de sal hasta que todo el material inicialmente retenido en la fase de intercambio iónico era desorbido y analizado.

La aplicación de una versión mejorada de este sistema, bautizada ya como MudPIT, al análisis de la levadura permitió identificar más de 5000 péptidos correspondientes a 1484 proteínas, la mayor colección obtenida hasta ese momento con cualquier otra técnica [3, 6]. La capacidad de separación de picos del sistema MDLC-electrospray-MS/MS se estimó en 25,000, valor que se incrementaba hasta cerca de 10^8 si se utilizaba el espectrómetro únicamente como detector de masas. Los datos obtenidos en este estudio demostraron además una de las características propias del análisis *shotgun*: la eliminación de los sesgos en la identificación de proteínas de baja expresión o de propiedades fisicoquímicas (Mr, pI, solubilidad) extremas propios de las técnicas de análisis en gel. El procedimiento mostraba una buena reproducibilidad cromatográfica (0.5%) y rangos dinámicos del orden de 10^4 que indicaban su capacidad para detectar proteínas que se encontrasen a niveles de 100 copias/célula.

En paralelo con estos trabajos, el sistema MudPIT fue empleado para la caracterización inicial del proteoma de *Plasmodium falciparum,* [21] la secuenciación de cuyo genoma se había descrito ese mismo año. En este estudio, que abordó el proteoma en las cuatro etapas del ciclo vital del plasmodio, se pudieron describir 2415 proteínas de este parásito que debieron caracterizarse en presencia de miles de proteínas del huésped. Asimismo, Schirmer et al. aplicaron este sistema a la caracterización de proteínas integrales de la membrana de la envuelta nuclear [22]. En el conjunto del trabajo se identificaron 2931 proteínas.

Las principales ventajas de los sistemas integrados residen en su automatismo y en la eficiencia conseguida con la utilización del total de la muestra disponible y la eliminación de las pérdidas implícitas a cualquier proceso de colección y reinyección de fracciones [23]. En los sistemas *off-line*, sin embargo, dado que las dos etapas de separación se llevan a cabo de forma independiente, es posible optimizar cada una por separado lo que puede redundar en un procedimiento más eficiente en su conjunto, especialmente cuando la cantidad de muestra no es una limitación [24]. Así, en un trabajo en el que se comparaba la eficiencia de un sistema MDLC *off-line* con los datos previamente reportados mediante MudPIT en el análisis del proteoma de la levadura, el grupo de Gygi pudo caracterizar más de 7500 péptidos correspondientes a más de 1500 proteínas en un único análisis [24]. Estos valores superaban a los reportados por el grupo de Yates [3], cuya colección de datos era la combinación de resultados del análisis de tres fracciones distintas del extracto de levadura y que en conjunto utilizaban un 35% más de muestra (1.35 mg vs 1 mg). El trabajo *off-line* permitía un tratamiento de la muestra más efectivo, con la posibilidad de usar detergentes que posteriormente eran eliminados durante la separación por intercambio iónico y que producían un aumento en el número de proteínas de membrana detectadas en el estudio.

El fraccionamiento *off-line* permite además el tratamiento de las fracciones para su enriquecimiento en componentes específicos antes de su análisis por fase reversa. Este es un procedimiento general en el análisis del fosfoproteoma donde los fosfopéptidos en las fracciones de intercambio iónico son purificados mediante columnas IMAC o de óxido de titanio [25]. Por otro lado, el acoplamiento *off-line* es menos sensible a los problemas de compatibilidad entre sistemas cromatográficos diferentes, lo que permite combinar sistemas que, por el tipo de disolventes utilizados, por el volumen de eluído generado o por la diferencia de flujos de solvente que utilizan, resultan difíciles de acoplar en línea. Los sistemas *off-line* permiten además evitar los cuellos de botella originados por etapas de diferente duración (p.e., varios espectrómetros pueden utilizarse en paralelo para el análisis de fracciones diferentes obtenidas en la primera dimensión).

2.3. AMT *(Accurate mass and time tags).*

Si bien una gran parte de los trabajos en *shotgun proteomics* emplean técnicas de espectrometría de masas en tándem para la identificación de las secuencias peptídicas, esta identificación puede hacerse también en base a otras características como las masas exactas y las propiedades cromatográficas de estos péptidos.

El grupo de Smith ha desarrollado tecnologías dirigidas al análisis ultrasensible a gran escala de proteomas basadas en el cálculo de la masa exacta de los péptidos mediante sistemas FT-ICR, TOF y Orbitrap y en los datos de tiempo de retención obtenidos en sistemas cromatográficos de alta eficiencia [26]. Esta estrategia, denominada AMT (*Accurate Mass and Time tags*), implica un análisis previo de las muestras mediante espectrometría de masas en tándem. Los péptidos identificados con alta confianza en este análisis entran a formar parte de una base de datos conteniendo sus masas exactas y los valores de su tiempo de elución normalizado frente a una separación de referencia. Esta información se utiliza posteriormente para la identificación de los péptidos detectados mediante espectrómetros de masas de alta resolución de una única etapa, teniendo en cuenta tanto la coincidencia de tiempos y masas como consideraciones estadísticas acerca de la complejidad de la muestra.

La sustitución del procedimiento convencional basado en la determinación del espectro de secuenciación de cada péptido individual por el de un análisis de masa en donde cada espectro aporta información sobre múltiples péptidos permite incrementar de forma significativa tanto la velocidad de análisis como el recubrimiento del proteoma en cuestión. Así, mediante la utilización de la tecnología de marcas AMT con columnas ultrarrápidas, se ha podido conseguir la identificación de 1656 péptidos en un análisis de 120 segundos, lo que corresponde a una velocidad de caracterización de 246 proteínas por minuto, la mayor reportada hasta ese momento [11, 26].

2.4. Espectrómetros de masas.

Los analizadores de masas más frecuentemente utilizados para la identificación de proteínas mediante secuenciación han sido las trampas de iones y los sistemas híbridos Q-TOF y Q-LIT. Más recientemente, los sistemas híbridos LIT-Orbitrap se han convertido probablemente en la instrumentación de referencia para análisis *shotgun* a gran escala. Cabe destacar que los trabajos pioneros del grupo de Hunt a principios de los años 90 se llevaron a cabo mediante un espectrómetro de triple cuadrupolo [27]. Aunque los espectrómetros de triple cuadrupolo no han jugado ningún papel relevante en el análisis de proteomas a gran escala, recientemente estos sistemas han empezado a ocupar un lugar prominente en proteómica cuantitativa dada su alta sensibilidad y linealidad en el análisis cuantitativo de moléculas diana seleccionadas.

Las trampas de iones han sido uno de los sistemas más utilizados para la identificación de proteínas tanto en formato de acoplamiento LC/MS como mediante nanospray. Muchos de los trabajos pioneros de *shotgun proteomics* que hemos comentado anteriormente fueron realizados mediante este tipo de sistemas [6]. Las trampas de iones son analizadores estables, con una resolución moderada pero que ofrecen un alto rendimiento de fragmentación. Su capacidad para acumular iones a partir de muestras con concentraciones muy bajas de material hasta conseguir acumular suficiente señal para obtener un espectro utilizable, hacían de estos detectores un sistema único.

Los sistemas Q-TOF permiten, por su parte, la obtención de espectros de fragmentación a resolución alta, con errores entre 2 y 5 ppm. Como en otros sistemas que utilizan un cuadrupolo o una trampa de iones como primer analizador, el ión precursor puede asimismo seleccionarse en una ventana relativamente estrecha, en algunos casos inferior a 1 Da. Ambos factores determinan la obtención de espectros que mejoran la eficiencia de la búsqueda posterior en bases de datos, una de las etapas más críticas en el análisis de proteomas a gran escala.

El método de fragmentación por excelencia en espectrometría de masas es la disociación inducida por colisión (CID), en el que se producen fragmentos de las moléculas a analizar mediante colisión con átomos o moléculas de gases nobles como el argón o el xenón (triples cuadrupolos) o el helio (trampas iónicas). Remitimos al lector al capítulo 16 de este mismo libro y a otros trabajos para una descripción más detallada de la fragmentación peptídica mediante CID [28]. Más recientemente se han introducido métodos de fragmentación basados en la irradiación de la muestra con luz láser, como en el caso de la disociación por absorción de radiación infrarroja (*InfraRed MultiPhoton Dissociation*, IRMPD), o con electrones libres o radicales químicos, como en las denominadas disociación por captura electrónica (ECD) y disociación por transferencia electrónica (ETD), respectivamente, que ofrecen una información distinta a la obtenida por CID. La ECD y ETD producen principalmente fragmentaciones en la cadena lineal del péptido, respetando en gran medida las modificaciones de las cadenas laterales, por lo que son procedimientos de interés en el análisis de modificaciones postraduccionales (PTMs) y son especialmente eficientes en el caso de polipéptidos con alto número de cargas y de tamaño grande.

Como ya se indicó en el apartado anterior, los espectrómetros de alta resolución de una sola etapa (TOF, FT-ICR, Orbitrap) pueden también emplearse para la caracterización de proteomas utilizando la técnica de marcas AMT. Esta estrategia requiere un análisis previo del proteoma en cuestión, mediante un análisis convencional por espectrometría de masas en tándem, para generar una base de datos de tiempos de retención y masas exactas que se utilizarán posteriormente para la identificación. A pesar de los condicionantes técnicos implicados en esta aproximación, la técnica de marcas AMT se ha aplicado con éxito al análisis diferencial y ultrasensible de proteomas en estados diferentes. En los estudios realizados, el analizador TOF demostró ser la mejor elección en el caso de análisis cromatográficos ultrarrápidos (< 2 min) debido a que este analizador permitía velocidades de barrido relativamente altas (0.2 s por espectro). Para separaciones más lentas, el analizador FT-ICR produjo mejores resultados debido a la exactitud de masa alcanzable con este instrumento [29].

3 Prefraccionamiento de muestras.

La cobertura conseguida en el análisis de un proteoma puede aumentarse mediante diferentes estrategias para el fraccionamiento de la muestra previo al análisis *shotgun*. Este prefraccionamiento puede realizarse a nivel del material biológico de partida (subfraccionamiento celular), de sus componentes proteicos en base a sus propiedades fisicoquímicas (solubilidad, filtración, ultracentrifugación), electroforéticas (IEF, SDS-PAGE) o cromatográficas (exclusión molecular, cromatografía de afinidad, cromatografía líquida). El objetivo de estos procedimientos es aportar un nivel adicional de separación al análisis por LC/MS o MDLC/MS.

Algunos métodos como el fraccionamiento de las proteínas mediante cromatografía líquida previo a su análisis por LC/MS, o la separación de proteínas en un gel monodimensional, que posteriormente se corta en bandas que se digieren y analizan separadamente, han mostrado un rendimiento del mismo orden o superior al obtenido mediante sistemas multidimensionales equivalentes basados en una primera etapa de intercambio iónico [23]. En un estudio en el que se evaluaron diversos métodos de fraccionamiento y análisis, la SDS-PAGE superó al fraccionamiento por IEF con cerca del doble de identificaciones en el análisis *shotgun* [30]. De hecho, la combinación del fraccionamiento en gel y el análisis LC-MS/MS fue la estrategia que produjo mejores resultados superando a los de un análisis MudPIT [30]

La separación de las proteínas en un gel monodimensional constituye una técnica preparativa simple y rápida que provee de una alta capacidad de carga. En su trabajo clásico de la caracterización del fosfoproteoma nuclear de la células Hela, el grupo de Gygi utilizó un gel de 1.5 mm (15 x 15 cm) para el fraccionamiento de 5 mg de extracto inicial [31]. El análisis de las 10 secciones en las que se dividió el gel mediante un sistema *off-line* SCX – fase reversa, permitió la caracterización de 2002 puntos de fosforilación en 1715 fosfopéptidos correspondientes a 967 proteínas diferentes.

A diferencia de las técnicas basadas en gel, los procedimientos de prefraccionamiento basados en la cromatografía tienen la ventaja de ser fácilmente automatizables. Wagner et al. describió un sistema integrado para el fraccionamiento de proteínas utilizando 4 columnas de fase reversa encargadas de separar las fracciones provenientes de una columna de SCX [32]. El sistema estaba destinado a la determinación de proteínas de tamaño inferior a 14 KDa para lo que la columna de SCX venía a su vez precedida por una columna RAM (*Restricted Acces Media*). Utilizando extractos celulares y hemodializados como muestras de partida, el sistema era capaz de resolver, en un tiempo de 96 minutos, hasta 1000 picos de péptidos y proteínas que podían ser luego sometidos a análisis por mapeo peptídico o por secuenciación.

En el análisis LC/MS de muestras complejas, la velocidad de barrido del espectrómetro de masas es frecuentemente el factor limitante que determina la capacidad de separación de picos del sistema. Durante un análisis *shotgun*, el espectrómetro de masas trabaja en ciclos donde un espectro de masas viene seguido de los espectros de secuenciación de un determinado número de las señales mayoritarias detectadas en aquél. Este modo de trabajo implica que el número de compuestos coeluyentes detectables en el espacio de tiempo definido por el ancho de pico cromatográfico viene determinado por el tiempo necesario para realizar cada ciclo de análisis de masa. Por otro lado, la selección automatizada de las señales más intensas para su fragmentación lleva a la discriminación de señales minoritarias que escapan a la detección. En un trabajo reciente se calculaba que, utilizando uno de los sistemas mas veloces disponibles actualmente (Orbitrap Velos), era posible detectar alrededor de 100,000 señales peptídicas en un análisis LC/MS del proteoma de células HeLa. Sin embargo, debido a limitaciones relacionadas con la sensibilidad y velocidad de barrido del espectrómetro, solo un 16% de estas señales eran finalmente sometidas a análisis de secuencia [33].

Una estrategia para aumentar la fracción de señales peptídicas sometidas a análisis de secuencia en un análisis *shotgun* es la probablemente mal denominada de fraccionamiento en fase gaseosa (GPF, *Gas Phase Fractionation,* figura 4). En este procedimiento, la muestra se somete a varios análisis en cada uno de los cuales la búsqueda de precursores se restringe a un rango de masas [34] o de intensidades relativas [35] diferente. La eficiencia de este procedimiento puede optimizarse mediante la selección de rangos que se adaptan al número total de señales presentes en distintas zonas del cromatograma. La utilización de este procedimiento implica la reinyección de la misma muestra repetidas veces por lo que queda restringido a situaciones donde la velocidad de análisis y la cantidad de muestra no son factores limitantes.

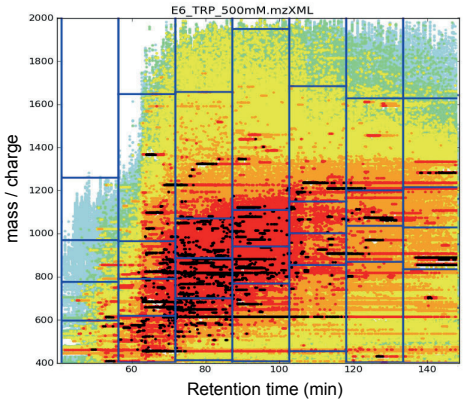


Figura 4.- Fraccionamiento en fase gaseosa (GPF). Mediante el análisis del número de señales por unidad de masa/carga a lo largo del tiempo de separación cromatográfica se determinan ventanas de rangos de masa que contienen un número similar de señales a identificar. El análisis de la muestra en análisis independientes, cada uno de ellos focalizado en una de estas ventanas de m/z, mejora la eficiencia del espectrómetro, limitada por su velocidad de barrido máxima, aumentando el número de identificaciones.

Diversos parámetros de interés para la optimización de este procedimiento han sido discutidos en la literatura [30, 36]. En el análisis del proteoma de *Pseudomonas aeruginosa* la utilización de 4 sectores de rango de masa aumentó alrededor de un 70-100% el número de identificaciones en un sistema LTQ y de un 70% en un Orbitrap en relación al mismo análisis utilizando un único barrido convencional [36]. El aumento del número de sectores produce mejoras menores a medida que el número de péptidos en cada sector se va ajustando a la eficiencia del ciclo de barrido del espectrómetro de masas. En otro estudio comparativo en el que se utilizaron cantidades idénticas de material de partida (65 µg de un digerido tríptico de organelas de levadura), los métodos GPF basados en rangos de masa (11 sectores), o de intensidad (9 rangos con tres iones cada uno) produjeron un número de proteínas identificadas muy similares entre sí, si bien estos valores eran aproximadamente un 25% inferiores a los conseguidos en un sistema MudPIT que utilizaba 13 etapas de elución [30].

4.- Análisis cuantitativo.

El análisis cuantitativo a gran escala de proteomas constituye uno de los retos más importantes del próximo futuro. La determinación de los niveles de las proteínas en diferentes tipos celulares y condiciones fisiológicas es un paso necesario para la comprensión de los mecanismos implicados en la función celular y es fundamental en el desarrollo de campos como la biología de sistemas. Los procedimientos para el análisis cuantitativo de proteomas pueden dividirse en dos grandes grupos en base al uso o no de isótopos estables.

Los métodos que utilizan isótopos estables utilizan algún tipo de marcaje isotópico diferencial de los componentes presentes en muestras diferentes que permite posteriormente su cuantificación relativa en un único análisis *shotgun*.

El marcaje isotópico de las proteínas se realiza utilizando reactivos de derivatización especiales como el ICAT [37] o el cICAT [38], o mediante marcaje metabólico *in vivo* a partir de moléculas

precursoras marcadas como es el caso del SILAC [39]. De la misma forma, para el marcaje isotópico de los péptidos se utilizan reactivos como el iTRAQ [40] o el TMT [41], el marcaje enzimático con ^{18}O utilizando tripsina en $H_2{}^{18}O$ [42] así como otros procedimientos utilizados clásicamente para la derivatización química de péptidos como la acetilación [43].

Todos estos métodos proporcionan información cuantitativa relativa. Para un análisis cuantitativo absoluto se utilizan las técnicas clásicas de dilución isotópica utilizando péptidos o proteínas marcados con isótopos estables como estándares internos. El método AQUA utiliza péptidos sintéticos que se añaden a la muestra durante o después de la digestión enzimática [44]. El método PSAQ (*Protein Standard for Absolute Quantification*) utiliza proteínas completas marcadas isotópicamente como estándares internos que se añaden a la muestra inicial [45]. El método QCAT hace uso de proteínas quiméricas constituidas por una serie de péptidos proteotípicos [46].

Por su parte, los métodos denominados *label-free* utilizan directamente la intensidad de las señales iónicas de los compuestos detectados o variables estadísticas experimentales, como el número de espectros adquiridos o el contaje de péptidos, para determinar la abundancia de los componentes presentes en muestras diferentes y que son analizadas separadamente [47].

Los detalles y compromisos de estas y otras técnicas se han analizado extensivamente en la literatura [48] y en un capítulo de este libro (Capítulo 17). Por otro lado, su aplicación al análisis cuantitativo a gran escala de proteomas ha sido hasta el momento relativamente limitada. El método SILAC ha sido aplicado con éxito por el grupo de Mann para la caracterización de los cambios *in vivo* en el fosfoproteoma de células Hela estimuladas con EGF [25]. En este estudio, los autores reportan la caracterización de 10000 fosfopéptidos en un total de 116 análisis LC/MS utilizando dos sistemas híbridos LIT-FT-ICR y LIT-Orbitrap. La cuantificación de los sitios de fosforilación se lleva a cabo utilizando el software MSQuant desarrollado por el mismo grupo (http://msquant.sourceforge.net/). Del conjunto de fosfopéptidos identificados, 1046 mostraron cambios superiores a dos veces por acción de la EGF, habiéndose calculado una desviación estándar de la cuantificación inferior al 20%. Este mismo grupo ha desarrollado una estrategia basada en el marcaje SILAC, espectrometría de masas de alta resolución (Orbitrap) y un conjunto de algoritmos integrados dentro del paquete MaxQuant [49] que ha demostrado ser capaz de identificar y cuantificar más de 5000 proteínas en células eucariotas [49, 4, 50]. La base de este alto recubrimiento, que esencialmente corresponde al proteoma completo de las células estudiadas, reside en el cálculo de alta precisión de las masas de las señales de los péptidos detectados. Mediante la aplicación de procedimientos de recalibración, los algoritmos del paquete MaxQuant proveen de medidas con errores medios del orden de 300 ppb (desviación.estándar, 409 ppb), seis veces inferiores a los obtenidos normalmente en estos instrumentos (error medio de 1.8 ppm con desviaciones estándar de 2.5 ppm). Esta precisión en el cálculo de la masa incrementa la eficiencia del motor de búsqueda utilizado (Mascot) de forma que el sistema es capaz de identificar el 73% de todos los pares de señales SILAC detectadas. Por su parte, Shen et al, utilizando un marcaje con ^{15}N, demostraron la posibilidad de realizar cuantificaciones con cantidades mínimas de material utilizando un sistema nanocapilar de alta eficiencia acoplado a un instrumento FT-ICR [12]. Utilizando únicamente 2.5 ng de material fue posible la cuantificación de 436 proteínas en un análisis de 3 horas. Dado que el análisis FT-ICR era de una sola etapa (no MS/MS), la identificación de los correspondientes péptidos en los cromatogramas se realizaba mediante la estrategia de AMT mencionada anteriormente [26].

5.- Automatismos para análisis de alta velocidad.

El número de muestras analizadas por unidad de tiempo en un análisis *shotgun* puede aumentarse utilizando baterías instrumentales compuestas por sistemas LC/MS completos, cada uno de ellos trabajando de forma independiente. Este tipo de aproximación se ha utilizado en alguno de los trabajos clásicos en proteómica, como el de la caracterización sistemática de los complejos proteicos de *Saccharomyces* [51], pero se trata de una estrategia accesible a unos pocos laboratorios y de alto coste económico.

En un análisis LC/MS la mayor pérdida de eficiencia corresponde a los tiempos muertos durante los pasos de inyección de muestra y de lavado y estabilización de las columnas en los que el espectrómetro de masas, generalmente el equipo de mayor coste del sistema, no está adquiriendo ninguna información. Para mejorar el ciclo de trabajo sin necesidad de recurrir a baterías de espectrómetros de masas, se han desarrollado sistemas que hacen uso de varias columnas analíticas trabajando en paralelo y acopladas a un único analizador. Estos sistemas se clasifican según su modo de trabajo en sistemas multiplexados y sistemas escalonados (*staggered*). Los sistemas multiplexados alternan muy rápidamente la fuente de entrada al espectrómetro de forma que se muestrean alternativamente los diferentes eluídos de un conjunto de columnas que trabajan en paralelo. En los sistemas escalonados las diferentes columnas trabajan asincrónicamente, de forma que el eluído de cada columna se introduce durante los periodos en los que las otras columnas se encuentran en ciclos de carga, lavado o estabilización. Estos procedimientos son bien conocidos en la industria farmacéutica en áreas como el análisis ADME o el cribado de librerías de fármacos donde es precisa una alta velocidad de análisis pero están poco extendidos todavía en Proteómica.

El grupo de Smith describió un sistema microcapilar basado en la utilización de dos columnas acopladas a la fuente de electrospray de un único analizador FT-ICR. Mediante un sistema de válvulas era posible monitorizar la separación en una columna mientras la otra se encontraba en periodo de carga y estabilización [52]. Más recientemente, dos unidades de este tipo se acoplaron para generar un sistema integrado para LC/MS constituido por 4 columnas trabajando en paralelo en modo escalonado, cada una de ellas acopladas a una aguja de electrospray que se posiciona frente a la entrada del espectrómetro en el momento adecuado mediante un robot. Con esta configuración es posible conseguir ciclos de trabajo del 100% [53]. A pesar de la aparente complejidad de un sistema capilar de este tipo, los autores reportan una alta reproducibilidad y robustez a largo plazo y afirman que el sistema se usa de forma rutinaria en su laboratorio para el análisis de proteomas [53].

6.-Análisis de los datos de secuenciación.

Si bien el rápido desarrollo de las técnicas instrumentales para el análisis de proteomas ha permitido obtener colecciones de decenas de miles de espectros en cuestión de horas. Las herramientas informáticas y bioinformáticas necesarias para su interpretación y validación, así como para su análisis en el contexto de la biología del objeto de estudio, han quedado típicamente varios pasos por detrás de estos avances.

En los trabajos pioneros de Yates y cols., estos autores desarrollaron un algoritmo de correlación entre los espectros de secuenciación y las secuencias alojadas en una base de datos que fue el origen del conocido programa Sequest. Diversos motores de búsqueda como Sequest, Mascot, Phenyx u OMMSA son actualmente aplicados de forma rutinaria en análisis *shotgun*

para la identificación de secuencias en bases de datos (Tabla 1). Las diferencias entre estos algoritmos y sus ventajas relativas han sido descritas detalladamente en la literatura [54-57]. Existen además diversos paquetes informáticos libres (p.e. Trans Proteomics Pipeline, TPP, obtenible en http://sourceforge.net/projects/sashimi/, OpenMS [58] obtenible de http://open-ms.sourceforge.net/) y comerciales (p.e ProteinScape, Scaffold) dirigidos a la evaluación integrada de los datos y a su clasificación.

Tabla 1. Herramientas para la identificación de proteínas mediante mapeo peptídico o secuenciación de péptidos en bases de datos o de novo (ver también ref 55-56).

Mapeo Peptídico	
Aldente	http://www.genebio.com/products/phenyx/aldente/
Mascot	www.matrixscience.com
MS-Fit	prospector.ucsf.edu
ProFound	http://prowl.rockefeller.edu/prowl-cgi/profound.exe
Secuenciación de péptidos	
Andrómeda	www.andromeda-search.org
GPM X! Hunter	http://h201.thegpm.org/tandem/thegpm_hunter.html
EigenMS	[91]
Mascot	www.matrixscience.com
MS-Tag	prospector.ucsf.edu
OMSSA	pubchem.ncbi.nlm.nih.gov/omssa/
Phenyx	http://www.genebio.com/products/phenyx/
Popitam	https://code.google.com/p/popitam/
SEQUEST	fields.scripps.edu/sequest
SHERPA	www.hairyfatguy.com/Sherpa/
Spectrum Mill	http://www.chem.agilent.com/en-US/products-services/Software-Informatics/Spectrum-Mill/Pages/default.aspx
X!TANDEM	www.thegpm.org/TANDEM/
De Novo	
DeNovoID	http://proteomics.mcw.edu/denovoid.html
DenovoX	www.thermo.com/
Lutefisk	www.hairyfatguy.com/lutefisk/
MSNovo	msms.usc.edu/supplementary/msnovo
NovoHMM	http://www-huber.embl.de/users/befische/software.html
Peaks	www.bioinfor.com/peaks/products/studio.html
PepNovo	proteomics.ucsd.edu/software-tools/531-2/
SHERENGA	[92]
Basados en Tags de Secuencia	
GutenTag	fields.scripps.edu/downloads.php
PEAKS-SPIDER	www.bioinfor.com/peaks/features/spider.html
MS BLAST	genetics.bwh.harvard.edu/msblast/
FASTA	www.ebi.ac.uk/fasta33
Inferencia	
DBParser	[93]
DTASelect	fields.scripps.edu/downloads.php
EBP	bioinf.itmat.upenn.edu/ebp/
EPIR	[94]
IDPicker	fenchurch.mc.vanderbilt.edu/software.php
Isoform Resolver	[95]
Mass Sieve	[96]
MAYU	prottools.ethz.ch/muellelu/web/LukasReiter/Mayu/
MFPaQ	mfpaq.sourceforge.net/
PRISM	[97]
Protein Prophet	proteinprophet.sourceforge.net/prot-software.html
Pro_Probe	[98]
Scaffold	www.proteomesoftware.com/Proteome_software_prod_Scaffold.html
VEMS	portugene.com/vems.html
PANORAMICS	[99]
ProteinScape	http://www.bruker.com/products/mass-spectrometry-and-separations/software.html
PROVALT	[85]
Qscore	[100]

La identificación de proteínas o péptidos en el caso de especies con genomas poco o nada secuenciados debe llevarse a cabo mediante un análisis *de novo* (es decir, sin recurrir a la información de las bases de datos, al menos en primera instancia). En el pasado reciente el análisis de novo implicaba el análisis manual de los espectros, procedimiento obviamente inviable en análisis a gran escala. Actualmente se dispone de diversas aplicaciones software que utilizan diferentes algoritmos matemáticos para deducir la secuencia o trozos (*tags*) de la secuencia del péptido que ha podido producir el espectro analizado (Tabla 1). Generalmente, la proteína o el posible gen de origen debe identificarse en un segundo paso a partir de los tags obtenidos mediante análisis de similitud utilizando aplicaciones como BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) o FASTA (http://www.ebi.ac.uk/Tools/sss/fasta/). Cuando se utilizan estos programas de búsqueda por homología debe seleccionarse una matriz adecuada para el análisis que tenga en cuenta el pequeño tamaño de las secuencias problema. Asimismo, debe tenerse en cuenta que la puntuación de estas herramientas valoran la semejanza funcional de las secuencias y no su semejanza química. Estos condicionantes han sido tenidos en cuenta en aplicaciones como MS-BLAST [59] que utiliza matrices optimizadas para la búsqueda de secuencia derivadas del análisis espectrométrico y que es accesible vía web (Tabla 1). Junqueira et al. describieron un método automatizado para el análisis de novo de proteomas que utilizaba diversas herramientas para el filtrado de los datos (EagleEye), la búsqueda en bases de datos (Mascot, PepNovo) y el análisis de homología (MS-BLAST) aplicándolo al análisis del insecto *Triatoma infestans* y de *Araucaria angustifolia* [60].

Un problema frecuentemente obviado en la interpretación de espectros de secuenciación en el análisis de muestras complejas es el derivado de la coelución de péptidos de masas similares. En general, los motores de búsqueda convencionales asignan cada espectro a una secuencia, soslayando el hecho de que, debido a la relativa baja resolución con la que se aíslan los iones precursores, frecuentemente los espectros de secuenciación corresponden a mezclas de espectros de péptidos diferentes [61-65]. Recientemente, el grupo de Mann ha desarrollado un motor de búsqueda, denominado Andrómeda, que incluye un algoritmo capaz de producir identificaciones secundarias en el análisis de este tipo de espectros [65]. Estos autores han mostrado un incremento de entre el 5.7% y el 10.7% en el número de identificaciones en función de la resolución utilizada para el aislamiento del precursor (4 y 2 Th respectivamente). Andrómeda ha sido integrado en el paquete MaxQuant [49] aunque puede ser utilizado también de forma independiente [65] (Tabla 1).

Una de las principales limitaciones de estas herramientas es el tiempo necesario para el análisis del enorme volumen de datos típicamente producidos en los análisis *shotgun*. Las modificaciones químicas producidas en los péptidos para su análisis o cuantificación, así como las modificaciones postraduccionales, extienden el número de elementos a combinar en la búsqueda más allá de los 20 aminoácidos típicos. Cuando deben considerarse varias de estas posibilidades los tiempos de búsqueda pueden aumentar exponencialmente hasta hacer la operación ineficiente. Por ejemplo, en nuestro laboratorio, utilizando un ordenador personal y utilizando datos de baja resolución (tolerancia de 2 Th para el error de masa del ion precursor), el programa OMMSA emplea aproximadamente 20 ms de cálculo por espectro de secuenciación. Cuando esta búsqueda requiere tener en cuenta la presencia de fosforilaciones (modificación en S, T, Y y deshidratación de S y T), el tiempo de cálculo por espectro pasa a 260 ms. El problema es todavía más grave con los algoritmos de secuenciación de novo que requieren una capacidad de cálculo mucho mayor que los buscadores en bases de datos y por lo que raramente han sido utilizados en análisis a gran escala.

La disponibilidad de ordenadores basados en multiprocesadores con alta capacidad de cálculo, de sistemas distribuidos o de clusters de ordenadores así como de software capaz de utilizar la potencia computacional ofrecida por estos sistemas son por tanto requisitos necesarios en el análisis a gran escala de proteomas. Afortunadamente, la disponibilidad de la nueva generación de espectrómetros de alta resolución (FT-ICR, LIT-Orbitrap) capaces de determinar

la masa del ión precursor con alta precisión ha permitido aumentar de forma relevante la velocidad de cálculo. De esta forma, la misma búsqueda del ejemplo anterior, realizada con espectros obtenidos en un sistema Orbitrap (tolerancia de 0.02 Th para el ión precursor) requiere 10 ms por espectro cuando no se contemplan modificaciones, y solo aumenta hasta 17 ms en la búsqueda de péptidos fosforilados.

6.1 Filtrado de datos.

Todos los motores de búsqueda indicados anteriormente calculan uno o varios parámetros o puntuaciones (*scores*) indicadores de la calidad de la identificación y que pueden relacionarse con la probabilidad de que la identificación no se haya producido al azar. La utilización de datos de muy alta puntuación minimiza el número de falsos positivos en la colección a expensas de proveer de un número menor de identificaciones totales y por tanto un recubrimiento menor del proteoma analizado. De esta forma, en el proceso de filtrado debe conseguirse un equilibrio óptimo entre sensibilidad y especificidad. Sin embargo, los algoritmos de búsqueda pueden producir identificaciones de alta puntuación incluso con espectros de ruido, por lo que resulta imposible separar de forma absolutamente fiable las asignaciones correctas de las incorrectas. En este contexto, la valoración manual por un experto de las secuencias asignadas a un espectro probablemente sigue siendo la validación más robusta de esta asignación. Este procedimiento es, sin embargo, inaplicable en un análisis *shotgun* en el que deben analizarse de forma intensiva grandes colecciones de espectros. El filtrado de los falsos positivos en el proceso de identificación automatizada es por tanto uno de los principales problemas a considerar en este tipo de tareas.

La proporción estimada de falsos positivos en una población de péptidos suele expresarse como la tasa de asignaciones incorrectas o FDR (*False Discovery Rate*). La FDR puede determinarse de forma eficaz determinando el número de asignaciones que se obtendría, a partir de esa colección de espectros y teniendo en cuenta una determinada puntuación de corte, si la búsqueda se realizara contra una base de datos señuelo compuesta por secuencias proteicas aleatorizadas o invertidas [66] (Figura 5). Aunque la búsqueda en la base de datos señuelo, comúnmente denominada *mock* o *decoy*, puede hacerse de forma independiente a la de la base de datos natural, también denominada diana o *target*, normalmente el proceso se realiza en un solo paso utilizando bases de datos combinadas conteniendo los dos conjuntos de secuencias proteicas.

El valor de FDR viene definido por:

$$FDR = \frac{FP}{FP + TP}$$

Donde FP define el número de falsos positivos en la colección y TP el de los positivos correctos. En base a esta fórmula, el FDR de una colección asignada en una base de datos combinada y filtrada a un determinado valor de puntuación puede calcularse como:

$$FDR = \frac{2 \times Decoy}{Decoy + T\arg et}$$

Donde *Decoy* son las asignaciones a secuencias en la base de datos señuelo y *Target* las asignaciones en la base de datos diana. El factor 2 resulta de considerar que un espectro que no corresponde a una secuencia en la base de datos tiene las mismas posibilidades de asignarse erróneamente a una secuencia de la base de datos natural que de la señuelo, por lo que el número de falsos positivos será el doble que el número de identificaciones en la base de datos señuelo [66]. Otros autores, sin embargo, consideran que esta fórmula no refleja correctamente la proporción de falsos positivos en las colecciones de péptidos identificados en la

base de datos natural que es con las que finalmente se trabaja. En base a este razonamiento, el valor de FDR se calcula como la razón de asignaciones en la base de datos *decoy* frente a las asignaciones en la natural. Este cálculo alternativo genera valores que, para los niveles de filtrado utilizados comúnmente, son aproximadamente la mitad de los calculados con la fórmula indicada arriba [67, 68].

Generalmente, las colecciones descritas en la literatura se seleccionan con valores de FDR inferiores al 5%. Este valor depende en gran medida de la instrumentación empleada y con aparatos de alta resolución es común reportar valores de FDR del orden del 0.1%.

Por su parte, los algoritmos disponibles actualmente para el análisis *de novo* producen resultados cuya bondad resulta a veces difícil de evaluar. Algunos de los problemas asociados a este tipo de búsqueda ya han sido comentados más arriba en relación a los algoritmos de búsqueda en bases de datos. Todos los programas de búsqueda *de novo* que hemos podido utilizar tienden a producir una gran cantidad de secuencias no correctas y son frecuentes las secuencias puntuadas con alta fiabilidad a partir de espectros de ruido que no corresponden a péptidos. Los espectros de fragmentación por CID no suelen contener una información completa de la secuencia lo que produce zonas ambiguas donde más de una combinación de aminoácidos o un distinto orden de éstos es posible. Algunas herramientas de búsqueda como Lutefisk marcan estas zonas de secuencia con el valor de su peso molecular mientras que otras, más voluntariosas, rellenan estos huecos siempre. La problemática de esta última estrategia se evidencia por una frecuencia anormal de asignaciones de alta probabilidad que corresponden a secuencias donde las zonas ambiguas se han rellenado con cadenas de residuos de triptófano o que contienen un alto número de otros aminoácidos de alto peso molecular como tirosina o cisteína. Aunque el triptófano es un aminoácido poco frecuente, su alto peso molecular permite rellenar los espacios vacíos de forma más efectiva, objetivo final del algoritmo de cálculo.

La utilización de filtros de calidad de los espectros como paso previo a la búsqueda es un factor vital para evitar la producción de asignaciones casuales [69, 70]. Asimismo, la eliminación de espectros de interferencias conocidas antes de los análisis reduce los tiempos de búsqueda, disminuye el número de proteínas identificadas con asignaciones únicas y en general mejora el FDR de las colecciones obtenidas [60]. Por otro lado, la utilización combinada de puntuaciones diferentes como la correlación cruzada (Xcorr) o el dCn del algoritmo Sequest o el *D-value* derivado de Peptide Prophet puede aumentar el grado de resolución a la hora de separar las asignaciones en la bases diana y señuelo [71](figura 5).

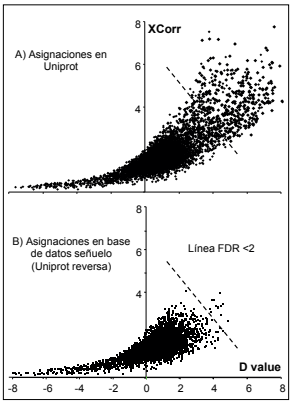


Figura 5.- Selección de identificaciones correctas mediante dos puntuaciones diferentes. La utilización de puntuaciones diferentes, en este caso, Xcorr (Sequest) y *D value* (Peptide Prophet), que evalúan la bondad de la asignaciones en base a parámetros y algoritmos diferentes, permite resolver con mayor eficacia el conjunto de identificaciones en las bases de datos *target* y *decoy*.

Probablemente el método más eficaz para la producción de colecciones de alta confianza consiste en el uso de dos o más motores de búsquedas diferentes. Los motores de búsqueda basados en algoritmos diferentes suelen tener un grado de coincidencia alto en la asignación de secuencias correctas, pero es más difícil que asignen de la misma forma un falso positivo. Las identificaciones asignadas por más de un motor de búsqueda tienen por tanto una mayor probabilidad de ser correctas por lo que es posible aumentar la calidad de una colección seleccionando únicamente aquellas secuencias cuya asignación coincide en dos o más buscadores. Diversos trabajos han utilizado esta estrategia para aumentar la calidad de los conjuntos de datos reportados [72, 73]. PEAKS (PEAKS, Bioinformatic Solutions Inc. Canadá) fue uno de los primeros programas comerciales que implementó una aplicación (InChorus, http://www.bioinfor.com/peaks/features/inchorus.html) dirigida a combinar los resul-

tados procedentes de diversos buscadores en bases de datos. Los aspectos cuantitativos de esta aproximación se han analizado en diversos trabajos [67, 74-76]. Jones et al. describieron un método para el cálculo de un FDR combinado en conjuntos de identificaciones coincidentes en dos o más motores de búsqueda, consiguiendo, para un FDR dado, un aumento en el número de identificaciones del 35% respecto al conseguido con cualquiera de los motores de búsqueda individuales [67]. Se han utilizado además diversas aproximaciones estadísticas para combinar las probabilidades de la identificación, derivadas de las puntuaciones de diferentes motores de búsqueda, en un único valor de probabilidad con el objeto de aumentar la sensibilidad del proceso de filtrado y un cálculo más preciso del valor de FDR [74, 75]. Un paquete informático que utiliza estas aproximaciones, MSblender, está disponible en la Red (http://www.marcottelab.org) [75]

El uso de instrumentos de alta resolución como Orbitraps y FT-ICRs, que permiten determinar con alta precisión la masa de iones precursores y producto, supone una importante ventaja para la asignación de las secuencias correctas [77, 49, 65]. Así, la determinación de la masa exacta del precursor disminuye enormemente el universo de búsqueda en programas como Sequest que seleccionan para su comparación con el perfil del espectro aquellas secuencias de la base de datos cuya masa coincide con la medida experimentalmente, reduciendo de esta forma la probabilidad de falsas identificaciones y aumentando la velocidad de análisis. Como ya se ha comentado, la determinación de masa exacta en instrumentos de alta resolución puede mejorarse mediante procedimientos de recalibración basados en el análisis de la distribución de las masas de las señales en la muestra. El paquete MaxQuant, citado más arriba, utiliza esta estrategia para disminuir el error de medida al nivel de 200 ppb y consiguiendo de esta forma ratios de identificación del 50% [49].

La eficiencia del análisis de secuencia puede aumentarse utilizando datos espectrométricos complementarios como la composición de aminoácidos obtenida a partir del análisis de los iones imonio o de medidas de masa exacta, o mediante el uso de marcaje isotópico. Uno de los métodos más extendidos y prácticos es la combinación de datos MS^2 y MS^3 del péptido problema que pueden obtenerse en sistemas QIT o LIT [78]. Asimismo, la utilización de métodos de fragmentación alternativos como ECD o ETD provee de espectros con información de secuencia que es complementaria a la obtenida mediante CID [79] y que puede ser utilizada conjuntamente para la identificación [80, 81].

6.2. Gene/Protein inference.

El análisis *shotgun* provee de una información péptido-céntrica, es decir que se refiere estrictamente a la secuencia de los péptidos presentes en las mezclas de digestión enzimática propias del análisis *shotgun*. Estos péptidos han perdido en el proceso su relación de identidad con las proteínas que los contenían. Una vez eliminada esta relación, la asignación de una determinada secuencia peptídica a su proteína de origen no es trivial. Los péptidos identificados pueden corresponder a zonas homólogas de proteínas diferentes o provenir de fragmentos proteicos con entidad propia en cuanto a su actividad biológica. Esta ambigüedad conlleva un problema de difícil solución en el análisis cuantitativo donde la cantidad total de un determinado péptido corresponde a la suma de todas estas posibles contribuciones y donde la dinámica de las moléculas de origen puede ser diferente.

El denominado problema de la inferencia de proteínas (*The Protein Inference Problem*) fue bautizado así y analizado en detalle por Nesvizhskii y Aebersold en 2005 [82]. En el pasado este problema no fue considerado de forma adecuada lo que ha generado la presencia en la literatura de colecciones de proteínas con datos redundantes o con identificaciones erróneas que dificultan o impiden la valoración biológica de los resultados obtenidos. Ante esta evidencia se ha realizado un importante esfuerzo con objeto de normalizar el uso de procedimientos adecuados para el manejo de la información resultante del análisis proteómico. En 2006 se publicó

el denominado Informe de París [83], un conjunto de recomendaciones dirigidas a asegurar la calidad de los datos publicados en la literatura, recogidas inicialmente en las instrucciones para publicación de la revista Molecular and Cellular Proteomics y adoptadas posteriormente por otras publicaciones científicas. Tal como se describe en esas instrucciones (http://www.mcponline.org/site/misc/ParisReport_Final.xhtml), los criterios mínimos para anotar el conjunto de proteínas asignables a partir de una colección de péptidos requieren ajustarse a la *lex parsimoniae*. Según este principio, la colección de proteínas derivada de un análisis proteómico debe estar constituida por el conjunto mínimo de éstas que describa adecuadamente la colección de péptidos identificados.

Actualmente, diversos algoritmos como Protein Prophet, integrado recientemente en el paquete TPP, Scaffold o el Pro Group Algorithm incluido en Protein Pilot (Tabla 1) son usados comúnmente para la inferencia de proteínas. Pueden encontrarse además en la literatura diversos procedimientos para evaluar la bondad de las proteínas asignadas principalmente basados a su vez en la bondad de las identificaciones de las secuencias peptídicas que apuntan a la proteína y del número de éstas [84-87]. Programas como MAYU [88] o PSEP [89] permiten el cálculo del FDR de estas colecciones. Mientras que PSEP calcula el FDR de las colecciones de proteínas identificadas, MAYU provee de estimaciones del FDR de las identificaciones individuales (Tabla 1).

7.- Conclusiones y Perspectiva.

En los últimos años, la eficiencia de las herramientas disponibles para el análisis de proteomas a gran escala ha aumentado enormemente. Si bien el análisis exhaustivo de los componentes de un proteoma implica todavía grandes dificultades y no puede considerarse resuelto, los avances en sensibilidad, rapidez y resolución de los instrumentos actuales, así como de las herramientas para el análisis de datos, son continuos. Desde las 196 proteínas de la levadura descritas por Shevchenko et al en 1996 [1] hasta las más de 4000 identificadas y cuantificadas por el grupo de Mann en 2008 [4] se ha recorrido un largo camino. Esta dinámica y el potencial y madurez de la instrumentación actualmente disponible han llevado a considerar la factibilidad del estudio a gran escala del proteoma humano y ya se ha iniciado un proyecto internacional con este objetivo, el Proyecto del Proteoma Humano (HPP, *Human Proteome Project*) (http://www.hupo.org/research/hpp/). El HPP es un proyecto multicéntrico internacional y auspiciado por HUPO y que, en una primera etapa, pretende identificar y monitorizar cada uno de los productos de los aproximadamente 21000 genes del genoma humano. El HPP por tanto puede constituir el mayor proyecto de proteómica a gran escala de esta década.

Una caracterización verdaderamente exhaustiva de los componentes de un proteoma tiene que lidiar sin embargo con la enorme complejidad derivada de las modificaciones postraduccionales de las proteínas, y de otros procesos relacionados con la expresión, maduración y metabolismo proteico. Un análisis de este tipo se enfrenta a un universo analítico de cientos de miles, quizás millones de componentes proteicos diferentes cuyo abordaje global mediante métodos *shotgun* y secuenciación por espectrometría de masas en tándem no es trivial. Incluso el estudio de subconjuntos aislados del proteoma como el glicoma, el nitroproteoma o el fosfoproteoma, constituye un test de prueba para las estrategias actuales. Una aproximación a este problema requerirá no solo una instrumentación analítica más eficiente y herramientas bioinformáticas capaces de deconvolucionar la compleja información obtenible de la muestra biológica, sino también el desarrollo de nuevas estrategias de análisis a gran escala que como las denominadas de proteómica *Top-Down* permitan ligar de forma más sencilla y específica, la información de secuencia derivada del análisis instrumental con la molécula original presente en la muestra [90].

REFERENCIAS.

1. Shevchenko A., Jensen O. N., Podtelejnikov A. V., et. al. Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. Proc. Natl. Acad. Sci. USA 1996, 93, 14440-14445.

2. Gygi S. P., Corthals G. L., Zhang Y., Rochon Y., Aebersold R.. Evaluation of two-dimensional gel electrophoresis based proteome analysis technology. Proc. Natl. Acad Sci. USA 2000, 97, 9390–9395.

3. Washburn M.P., Wolters D., Yates J.R. III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol., 2001, 19, 242-247.

4. de Godoy L. M. F., Olsen J. V., Cox J., et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 2008, 455, 1251-1255.

5. Yates J. R. 3rd, Eng J. K., McCormack A. L., Schieltz D. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem. 1995, 67, 1426-36.

6. Wolters D. A., Washburn M. P., Yates J. R., III. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. Anal. Chem. 2001, 73, 5683-5690

7. Emmett M. R., Caprioli R. M. Micro-electrospray mass spectrometry: ultra-high-sensitivity analysis of peptides and proteins. J. Am. Soc. Mass Spectrom. 1994, 5, 605-613.

8. Abian J., Oosterkamp A.J., Gelpí E. Comparison of conventional, narrow-bore and capillary liquid chromatography/mass spectrometry for electrospray ionization mass spectrometry: practical considerations. J. Mass Spectrom. 1999, 34, 244-254.

9. Carrascal M., Abian J. Capillary Separations. En Aguilar M.-I., editor. HPLC of Peptides and Proteins: Methods and Protocols Methods in Molecular Biology vol 251, Humana Press Inc. NJ, USA 2004, pp 143-164.

10. Shen Y., Lee M.L. General Equation for Peak Capacity in Column Chromatography. Anal. Chem. 1998, 70, 3853-3856.

11. Shen Y., Smith R. D., Unger K.K., Kumar D., Lubda D., Ultrahigh-Throughput Proteomics Using Fast RPLC Separations with ESI-MS/MS. Anal. Chem. 2005, 77, 6692-6701.

12. Shen Y., Tolic N., Masselon Ch., et al. Ultrasensitive Proteomics Using High-Efficiency On-Line Micro-SPE-NanoLC-NanoESI MS and MS/MS. Anal. Chem. 2004, 76, 144-154.

13. Bushey M. M., Jorgenson J.W. Automated Instrumentation for Comprehensive Two-Dimensional High-Performance Liquid Chromatography of Proteins. Anal. Chem. 1990, 62, 161-167.

14. Giddings, J. C. Concepts and comparisons in multidimensional separation. J. High Resolut. Chrom. 1987, 10, 319-323.

15. Giddings, J. C. Use of multiple dimensions in analytical separations in multidimensional chromatography. En: Cortes, H. J., Editor, Multidimensional Chromatography: techniques and applications; Marcel Dekker: New York, 1990; pp. 1-27.

16. Holland L.A., Jorgenson J.W., Separation of Nanoliter Samples of Biological Amines by a Comprehensive Two-Dimensional Microcolumn Liquid Chromatography System. Anal. Chem. 1995, 67, 3275-3283.

17. Opiteck G. J., Ramirez S. M., Jorgenson J. W., Moseley M.A. III. Comprehensive Two-Dimensional High-Performance Liquid Chromatography for the Isolation of Overexpressed Proteins and Proteome Mapping. Anal. Biochem 1998, 258, 349-361.

18. Opiteck G. J., Jorgenson J. W., Comprehensive On-Line LC/LC/MS of Proteins. Anal. Chem. 1997, 69, 1518-1524.

19. Opiteck G. J., Jorgenson J. W., Two-Dimensional SEC/RPLC Coupled to Mass Spectrometry for the Analysis of Peptides. Anal. Chem. 1997, 69, 2283-2291.

20. Link A. J., Eng J., Schieltz D.M., et al. Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. 1999, 17, 676-682.

21. Florens L., Washburn M.P., Raine J.D., et al., A proteomic view of the Plasmodium falciparum life cycle. Nature 2002, 419, 520-526.

22. Schirmer E. C., Florens L., Guan T., Yates J. R. III, Gerace L. Nuclear Membrane Proteins with Potential Disease Links Found by Subtractive Proteomics. Science 2003, 301, 1380-1382.

23. Dowell J. A., Frost D. C., Zhang J., Li L. Comparison of Two-Dimensional Fractionation Techniques for Shotgun Proteomics. Anal. Chem. 2008, 80, 6715-6723.

24. Peng J., Elias J. E., Thoreen C. C., Licklider L. J., Gygi S. P. Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC-MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome. J. Proteome Res. 2003, 2, 43-50.

25. Olsen J.V., Blagoev B., Gnad F., et al. Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. Cell 2006, 127, 635-648.

26. Smith R.D., Tang K., Shen Y. Ultra-sensitive and quantitative characterization of proteomes. Mol. BioSyst., 2006, 2, 221-230.

27. Hunt D.F., Henderson R.A., Shabanowitz J., et al. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science 1992, 255, 1261-1263.

28. Burlingame A. L., Boyd R. K., Gaskell S. J. Mass Spectrometry. Anal. Chem. 1994, 66, 634R-683R.

29.- Shen Y., Strittmatter E. F., Zhang R., et al. Making Broad Proteome Protein Measurements in 1-5 min Using High-Speed RPLC Separations and High-Accuracy Mass Measurements. Anal. Chem. 2005, 77, 7763-7773.

30. Breci L., Hattrup E., Keeler M., Letarte J., Johnson R., Haynes P.A. Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing. Proteomics 2005, 5, 2018-2028.

31. Beasusoleil S. A., Jedrychowski M., Schwartz D., et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc. Nat. Acad. Sci. 2004, 101, 12130-5

32. Wagner K., Miliotis T., Marko-Varga G., Bischoff R., Unger K.K. An Automated On-Line Multidimensional HPLC System for Protein and Peptide Mapping with Integrated Sample Preparation. Anal. Chem. 2002, 74, 809-820

33. Michalski A., Cox J., Mann M. More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC-MS/MS. J. Proteome Res. 2011, 10, 1785–1793

34. Spahr, C. S., Susin, S. A., Bures, E. J., et. Al., Simplification of complex peptide mixtures for proteomic analysis: Reversible biotinylation of cysteinyl peptides. Electrophoresis 2000, 21, 1635– 1650.

35. Yi E. C., Marelli M., Lee H., et al. Approaching complete peroxisome characterization by gas-phase fractionation. Electrophoresis. 2002, 23, 3205-16.

36. Scherl A., Shaffer S.A., Taylor G.K., Kulasekara H.D., Miller S.I., Goodlett D.R., Genome-Specific Gas-Phase Fractionation Strategy for Improved Shotgun Proteomic Profiling of Proteotypic Peptides. Anal. Chem. 2008, 80, 1182-1191.

37. Gygi S. P., Rist B., Gerber S. A., Turecek F., Gelb M. H., Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 1999, 17, 994-999

38. Li J., Steen H., Gygi S.P. Protein Profiling with Cleavable Isotope-coded Affinity Tag (cICAT) Reagents.The Yeast Salinity Stress Response. Mol. Cell. Proteomics 2003, 2, 1198-1204.

39. Ong S.E., Blagoev B., Kratchmarova I., et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics 2002, 1, 376–386.

40. Ross P. L., Huang Y. N., Marchese J. N., et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics 2004, 3, 1154-69.

41. Dayon L., Hainard A., Licker V., et al. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal. Chem. 2008, 80, 2921-31.

42. Schnolzer M., Jedrzejewski P., Lehmann W.D., Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. Electrophoresis 1996, 17, 945-953.

43. Chakraborty A., Regnier F.E. Gist for comparative proteomics. J. Chromatogr. A. 2002, 949, 173-178.

44. Gerber S.A., Rush J., Stemman O., Kirschner M.W., Gygi S.P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad. Sci. USA 2003, 100, 6940–6945.

45. Brun V., Dupuis A., Adrait A., Marcellin M., Thomas D., Court M., et al. Isotope-labeled protein standards: toward absolute quantitative proteomics. Mol. Cell. Proteomics 2007, 6, 2139–49.

46. Beynon R.J., Doherty M.K., Pratt J.M., Gaskell S.J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. Nat. Methods 2005, 2, 587–9.

47. Old W.M., Meyer-Arendt K., Aveline-Wolf L., et al. Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics. Mol. Cell. Proteomics 2005, 4, 1487-1502.

48. Panchaud A., Affolter M., Moreillon P., Kussmann M. Experimental and computational approaches to quantitative proteomics: Status quo and outlook. Journal of Proteomics 2008, 71, 19-33.

49. Cox J., Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 2008, 26, 1367-1372.

50. Graumann, J., Hubner N.C., Kimet J.B. al. SILAC-labeling and proteome quantitation of mouse embryonic stem cells to a depth of 5111 proteins. Mol. Cell Proteomics 2008, 7, 672–683.

51. Ho Y, Gruhler A., Heilbut A, et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 2002, 415, 180-183.

52. Shen Y, Tolic N., Zhao R., et al. High-Throughput Proteomics Using High-Efficiency Multiple-Capillary Liquid Chromatography with On-Line High-Performance ESI FTICR Mass Spectrometry. Anal. Chem. 2001, 73, 3011-3021.

53. Livesay E. A., Tang K., Taylor B.K., et al. Fully automated, four-column capillary LC-MS system for maximizing throughput in proteomic analyses. Anal. Chem. 2008, 80, 294-302.

54. Elias J. E., Haas W., Faherty B. K., Gygi S. P. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat. Methods, 2005, 2, 667-675.

55. Nesvizhskii A.I., Vitek O., Aebersold R., Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat. Methods, 2007, 4, 787-797

56. Matthiesen R. Methods, algorithms and tools in computational proteomics: A practical point of view. Proteomics 2007, 7, 2815-2832.

57. McHugh L., Arthur J. W. Computational Methods for Protein Identification from Mass Spectrometry Data. PLoS Computational Biology 2008, 4 , 1-12.

58. Kohlbacher O., Reinert K., Gropl C., et al., TOPP – the OpenMS proteomics pipeline. Bioinformatics 2007, 23, e191–e197.

59. Shevchenko A., Sunyaev S., Loboda A., et al. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. Anal. Chem. 2001, 73, 1917-1926.

60. Junqueira M., Spirin V., Balbuena T.S., et al. Protein Identification Pipeline for the Homology Driven Proteomics. J. Proteomics 2008, 71, 346-356

61. Zhang N., Li X. J., Ye M., Pan S., Schwikowski B., Aebersold R. ProbIDtree: an automated software program capable of identifying multiple peptides from a single collision-induced dissociation spectrum collected by a tandem mass spectrometer. Proteomics 2005, 5, 4096-4106.

62. Bern M., Finney G., Hoopmann M. R., Merrihew G., Toth M. J., MacCoss M. J. Deconvolution of mixture spectra from ion-trap data-independent-acquisition tandem mass spectrometry. Anal. Chem. 2010, 82, 833-841.

63. Wang J., Perez-Santiago J., Katz J. E., Mallick P., Bandeira N. Peptide identification from mixture tandem mass spectra. Mol. Cell. Proteomics 2010, 9, 1476-1485.

64. Houel S., Abernathy R., Renganathan K., Meyer-Arendt K., Ahn N. G., Old W. M. Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies. J. Proteome Res. 2010, 9, 4152-4160.

65. Cox J., Neuhauser N., Michalski A., et al. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. J. Proteome Res. 2011, 10, 1794-1805.

66. Elias J. E., Gygi S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods 2007, 4, 207-214.

67. Jones A. R., Siepen J. A., Hubbard S. J., Paton N. W. Improving sensitivity in proteome studies by analysis of false discovery rates for multiple search engines. Proteomics 2009, 9, 1220-1229.

68. Käll L., Storey J. D., MacCoss M. J., Noble W. S. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. J. Proteome Res. 2008, 7, 29-34.

69. Sadygov R.G., Eng J., Durr E., et al. Code developments to improve the efficiency of automated MS/MS spectra interpretation. J. Proteome Res. 2002, 1, 211-215.

70. Bern M., Goldberg D., McDonald W.H., Yates J.R. 3rd. Automatic quality assessment of peptide tandem mass spectra. Bioinformatics 2004, 20 (Suppl. 1), I49-I54.

71. Carrascal M., Ovelleiro D., Casas V., Gay M, Abian J.. Phosphorylation analysis of primary human T lymphocytes using sequential IMAC and titanium oxide enrichment. J Proteome Res. 2008, 7, 5167-76.

72. Balgley B. M., Laudeman T., Yang L., Song T., Le C. S. Comparative Evaluation of Tandem MS Search Algorithms Using a Target-Decoy Search Strategy. Mol. Cell. Proteomics 2007, 6, 1599-1608.

73. Carrascal M., Gay M., Ovelleiro D., Casas V., Gelpí E., Abian J. Characterization of the Human Plasma Phosphoproteome Using Linear Ion Trap Mass Spectrometry and Multiple Search Engines. J. Proteome Res. 2010, 9, 876-884.

74. Searle B. C., Turner M., Nesvizhskii A. I. Improving Sensitivity by Probabilistically Combining Results from Multiple MS/MS Search Methodologies. J. Proteome Res. 2008, 7, 245-253.

75. Kwon T., Choi H., Vogel Ch., Nesvizhskii A. I., Marcotte E. M. MSblender: A probabilistic Approach for Integrating Peptide Identifications from Multiple Database Search Engines. J. Proteome Res. 2011, In Press.

76. Alves G., Wu W. W., Wang G., Shen R. F., Yu Y. K. Enhancing peptide identification confidence by combining search methods. J. Proteome Res. 2008, 7, 3102-3113.

77. Frank AM, Savitski MM, Nielsen ML, Zubarev RA. et al. De Novo Peptide Sequencing and Identification with Precision Mass Spectrometry. J. Proteome Res. 2007, 6, 114-123.

78. Olsen J. V., Mann M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. Proc. Nat. Acad. Sci. USA 2004, 101, 13417-13422.

79. Creese A J, Cooper H J. Liquid Chromatography Electron Capture Dissociation Tandem Mass Spectrometry (LC-ECD-MS/MS) versus Liquid Chromatography Collision-induced Dissociation Tandem Mass Spectrometry (LC-CID-MS/MS) for the Identification of Proteins. J. Am. Soc. Mass Spectrom. 2007, 18,891-897.

80. Horn D. M., Zubarev R. A., McLafferty F. W. Automated De Novo Sequencing of Proteins by Tandem High-resolution Mass Spectrometry. Proc. Natl. Acad. Sci. USA 2000, 97, 10313-10317.

81. Savitski M. M., Nielsen M. L., Kjeldsen F., Zubarev R. A. Proteomics-grade De Novo Sequencing Approach. J. Proteome Res. 2005, 4, 2348-2354.

82. Nesvizhskii A. I., Aebersold R., Interpretation of Shotgun Proteomic Data. The Protein Inference Problem. Mol. Cell. Proteomics 2005, 4,1419-1440.

83. Bradshaw R.A., Burlingame A.L., Carr S., Aebersold R. Reporting Protein Identification Data. The Next Generation Of Guidelines. Mol. Cell. Proteomics 2006, 5, 787-788.

84. MacCoss M. J., Wu C. C., Yates J. R., 3rd. Probability-based validation of protein identifications using a modified SEQUEST algorithm. Anal. Chem. 2002, 74, 5593–5599.

85. Nesvizhskii A. I., Keller A., Kolker E., Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. Anal.Chem. 2003, 75, 4646-4658.

86. Weatherly D. B., Atwood J. A., 3rd, Minning T. A., Cavola C., Tarleton R. L., Orlando R. A heuristic method for assigning a falsediscoveryrate for protein identifications from Mascot database search results. Mol. Cell. Proteomics 2005, 4, 762-772.

87. Price T. S., Lucitt M. B., Wu W., et al. EBP, a program for protein identification using multiple tandem mass spectrometry datasets. Mol. Cell. Proteomics 2007, 6, 527-536.

88. Reiter L., Claassen M., Schrimpf S.P., et al. Protein Identification False Discovery Rates for Very Large Proteomics Data Sets Generated by Tandem Mass Spectrometry. Mol. Cell. Proteomics 2009, 8, 2405-2417.

89. Tang W. H., Shilov I. V., Seymour S. L. Nonlinear Fitting Method for Determining Local False Discovery Rates from Decoy Database Searches. J. Proteome Res. 2008, 7, 3661-3667.

90. Durbin K.R., Tran J.C., Zamdborg L., et al. Intact mass detection, interpretation, and visualization to automate Top-Down proteomics on a large scale. Proteomics 2010, 10, 3589-3597.

91. Bern M W, Goldberg D. De novo analysis of peptide tandem mass spectra by spectral graph partitioning. Journal of Computational Biology 2006, 13, 364-78.

92. Dancík V, Addona T A, Clauser K R, Vath J E. et al.. De novo peptide sequencing via tandem mass spectrometry. Journal of Computational Biology 1999, 6, 327-42.

93. Yang, X., Dondeti, V., Dezube, R., et al. DBParser: web-based software for shotgun proteomic data analyses. J. Proteome Res. 2004, 3, 1002-8.

94. Kristensen D.B., Brønd J.Ch., Nielsen P.A., et al. Experimental Peptide Identification Repository (EPIR). An Integrated Peptide-Centric Platform For Validation And Mining Of Tandem Mass Spectrometry Data. Mol Cell Proteomics, 2004 3, 1023-38.

95. Resing K. A., Meyer-Arendt K., Mendoza A. M., et al. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. Anal. Chem. 2004, 76, 3556-3568

96. Slotta D. J., McFarland M.A., Makusky A.J., Markey S.P. Mass Sieve: A new tool for mass spectrometry-based proteomics. J. Biomol. Tech. 2007;18:7.

97. Kiebel G. R., Auberry K. J., Jaitly N., et al. PRISM: a data management system for high-throughput proteomics. Proteomics. 2006 Mar;6(6):1783-90.

98. Sadygov, R. G., Liu, H. B. and Yates, J. R. Statistical models for protein validation using tandem mass spectral data and protein amino acid sequence databases. Anal. Chem. 2004, 76, 1664-1671.

99. Feng J., Garrett W. M., Naiman D.Q., Cooper B. Correlation of Multiple Peptide Mass Spectra for Phosphoprotein Identification *J. Proteome Res.*, 2009, 8, 5396-5405.

100. Moore R. E., Young M. K., Lee T. D. Qscore: an algorithm for evaluating SEQUEST database search results. J. Am. Soc. Mass Spectrom. 2002, 13, 378-386.

Aplicaciones de las herramientas proteómicas en Tecnología de la Alimentación

Miguel Ángel Sentandreu[1], Carlos B Amat[2] y Enrique Sentandreu[1]

[1] Instituto de Agroquímica y Tecnología de Alimentos (CSIC). Valencia, Spain;

[2] INGENIO [CSIC-UPV], Universidad Politécnica de Valencia, Spain.

1. Introducción.

Las técnicas proteómicas se usan cada vez más en la valoración de alimentos tanto procesados como en crudo, así como en el control, la optimización y el desarrollo de nuevos productos en tecnología y biotecnología de alimentos (1). Y ello por razones de peso.

En primer lugar, las proteínas y los péptidos son componentes principales de los alimentos y juegan un papel decisivo en la determinación de sus propiedades nutricionales y funcionales. Como parte de sistemas bioquímicos extremadamente complejos, las proteínas alimentarias son mezclas en las que se encuentran simultáneamente componentes derivados del procesamiento postraduccional (fosforilación, glicosilación, puentes disulfuro) y del corte y empalme alternativo del ARN pre-mensajero.

En segundo lugar, los alimentos que llegan al consumidor se han expuesto a toda una serie de diferentes procesos. Hasta el producto más simple, por ejemplo una pieza de fruta, ha sido expuesta, antes y después de su recolección, a una serie de condiciones de estrés biótico, abiótico y mecánico. El procesamiento de productos más elaborados también conlleva una serie de operaciones estresantes. El estrés altera la homeostasis del producto fresco y se manifiesta en cambios, tanto en su composición proteica como en la estructura y conformación de sus proteínas y en modificaciones postraduccionales. También en los productos procesados se pueden producir cambios en la conformación de las proteínas y hasta proteolisis. Es necesario comprender estos cambios para ahondar en nuestro conocimiento de las respuestas a procesos bioquímicos específicos, para mejorar más las estrategias de procesamiento de los alimentos y para ofrecer productos de alta calidad que satisfagan los requisitos de los consumidores (2).

En tercer lugar, las proteínas son en gran medida responsables de las características de muchos productos alimenticios durante su preparación. No sólo de las propiedades fisicoquímicas como la viscosidad o la conductividad térmica, sino también de las propiedades nutricionales y sensoriales de los productos derivados de la leche, la carne o los cereales, que dependen de su contenido en proteínas. Por otra parte, algunos procesos de degradación y de contaminación biológica afectan a la composición proteica de los alimentos. Los procesos técnicos que se emplean en la preparación de productos alimenticios contribuyen aún más a aumentar

la heterogeneidad de los sistemas proteicos, induciendo una serie de reacciones químicas como la proteolisis, la glicosilación no enzimática y otras modificaciones covalentes. Una fuente adicional de complejidad que dificulta la completa caracterización del proteoma de los alimentos es el amplio intervalo dinámico asociado a la expresión proteica. De forma similar a los sistemas biológicos de los que procede, el proteoma de los alimentos es un ente dinámico que refleja el medio ambiente y el estado fisiológico de los organismos originarios y evoluciona con el tiempo (3).

Por último, el metabolismo humano es principalmente la expresión de un estado de equilibrio entre la biosíntesis y la degradación de las proteínas, que funcionan como enzimas, transportadores, canales, hormonas y otras moléculas señalizadoras o que proporcionan elementos estructurales de las células, los órganos y el esqueleto (4). Los estudios de los efectos de los nutrientes en el proteoma tienen el potencial de explicar, a nivel molecular, muchos de los cambios fisiológicos asociados con los estímulos nutricionales (5).

Al igual que en otras áreas, la composición proteica de los alimentos ha sido estudiada desde hace tiempo por los métodos de la bioquímica tradicional, principalmente la separación de las proteínas por técnicas de electroforesis en gel de poliacrilamida y posterior identificación por inmunodetección o secuenciación N-terminal mediante la degradación de Edman. A pesar de su gran utilidad, estas técnicas presentan limitaciones importantes para matrices tan complejas como los alimentos. Aparte de ser métodos lentos y poco sensibles, con el solo uso de estas técnicas no era posible el análisis de proteínas a gran escala. El análisis proteómico basado en la identificación de secuencias peptídicas por espectrometría de masas ha supuesto un avance muy importante en la identificación masiva de proteínas. De hecho, desde los años 80 del pasado siglo se han dedicado considerables esfuerzos a la caracterización estructural de proteínas y péptidos de los alimentos. Aunque el uso de diversos métodos espectroscópicos ha rendido mucha información útil, no han proporcionado información sobre los cambios estructurales o dinámicos experimentados por regiones específicas y además, estos métodos requieren concentraciones relativamente elevadas, grandes cantidades de proteínas purificadas y un tiempo de análisis prolongado. En contraste, la espectrometría de masas (MS) puede revelar de forma apropiada y precisa los cambios composicionales de una proteína grande en un plazo de tiempo relativamente corto. Hasta el año 2000, se habían determinado las masas moleculares de un total de 20 proteínas animales y 25 de origen vegetal mediante MS (6). A partir de esa fecha se ha producido una escalada en la obtención de resultados. Esta revolución ha sido posible gracias a la confluencia de diferentes avances científicos entre los que hay que destacar el desarrollo en los años noventa de las técnicas de ionización "suave" acopladas a la espectrometría de masas como son la ionizacion por electrospray (ESI del inglés "Electrospay Ionization") y la ionización por desorción láser asistida por matriz (MALDI del inglés "Matrix-Assisted Laser Desorption Ionization"). Igualmente importante ha sido la secuenciación de genomas a gran escala junto con el desarrollo de las bases de datos de proteínas tal como las conocemos actualmente. En toda esta revolución proteómica también ha jugado un papel muy importante el desarrollo de las potentes herramientas bioinformáticas capaces de procesar la gran cantidad de información generada por los espectrómetros de masas. La Figura 1, que muestra el número de trabajos de investigación en ciencia y tecnología de alimentos que emplean técnicas proteómicas (7), ilustra sobradamente esta evolución.

Las proteínas son responsables en gran medida de las características globales de muchas formulaciones alimenticias. Las propiedades nutritivas, reológicas y sensoriales de los productos derivados de la leche, los cereales o la carne dependen de la composición de su fracción proteica. El caso más llamativo acaso sea el de los productos derivados de la harina de trigo, cuyas características óptimas están determinadas por las proteínas del gluten. Precisamente el cuerpo de este capítulo se inicia con un repaso a las aplicaciones de la proteómica a la determinación de la calidad de los alimentos.

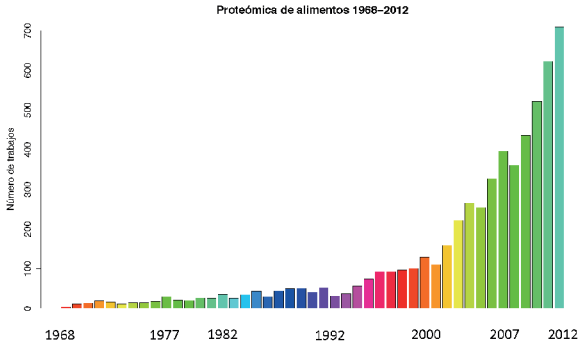


Figura 1: Distribución cronológica de los trabajos en proteómica de alimentos que aparecen en la base de datos Food Science and Technology Abstracts (FSTA) cuyos títulos incluyen términos del glosario de Patterson y Aebersold (7). De los 4 trabajos en 1968 se ha pasado a 709 trabajos en 2012.

Las proteínas de los alimentos son fuentes de péptidos biológicamente activos. La secuencia de un péptido activo permanece inactiva en el seno de la proteína hasta que es liberado durante la digestión gastrointestinal o el procesamiento del alimento. En ese momento pueden actuar como compuestos reguladores con actividad similar a la de una hormona, y exhibir un amplio abanico de funciones biológicas. Por eso la segunda parte de este capítulo se centra en la llamada proteómica nutricional.

Otra parte notable de este capítulo se dedica a la seguridad de los componentes de los alimentos y a sus interacciones durante el procesamiento. El ejemplo más notable de modificaciones inducidas por los tratamientos que pueden afectar la calidad global del alimento se refiere a los productos del tratamiento térmico sobre materiales como la leche, la carne o los cereales. Las principales modificaciones inducidas por el tratamiento térmico en las proteínas lácteas son la desnaturalización de las proteínas del suero y la compleja serie de interacciones covalentes conocidas como reacción de Maillard, que producen una disminución en la calidad nutricional y la eventual formación de compuestos tóxicos. La proteómica también se emplea en tecnología de alimentos en la validación y optimización de procesos, el control de calidad y la reducción de las variaciones entre lotes de fabricación. También se aplica a la detección de las alteraciones causadas por el empleo de organismos modificados genéticamente de origen vegetal y a la seguridad alimentaria, especialmente la relacionada con la identificación de alérgenos y microorganismos.

Finalmente, otro de los aspectos principales de la investigación en tecnología de alimentos es la caracterización de los compuestos que dotan a los alimentos de su peculiar sabor, aroma y en general las características sensoriales tan apreciadas por los consumidores. En razón de sus propiedades estructurales, las proteínas pueden establecer interacciones covalentes o no covalentes con diferentes componentes del alimento, entre ellos el agua, lo que resulta en cambios en la hidratación y la solubilidad de la proteína, en la viscosidad de las disoluciones, en la formación de películas, la gelificación y la adsorción a la interfase entre las fases acuosa y lipídica.

2. Caracterización de la materia prima.

2.1. Estudio de la composición de los alimentos.

Las herramientas de proteómica combinadas de manera adecuada con distintas estrategias de fraccionamiento y separación han hecho posible un aumento espectacular de nuestro conocimiento sobre la composición proteica de los alimentos, algo que es fundamental para poder conocer de manera precisa las propiedades nutritivas, sensoriales y funcionales de los mismos. Uno de los principales avances aportados por el análisis proteómico ha sido el poder identificar las proteínas minoritarias de los alimentos pero que son relevantes desde un punto de vista

fisiológico, nutricional o tecnológico. Esto ha sido posible gracias al desarrollo de sistemas de fraccionamiento previo capaces de eliminar las proteínas más abundantes, las cuales con frecuencia representan un número reducido de todo el conjunto total de proteínas pero que interfieren en gran medida en el análisis de las demás. Un caso ilustrativo de esta situación es el estudio del proteoma de la leche realizado por Yamada y col. (8) donde se llevó a cabo un estudio de las proteínas minoritarias de leche y calostro gracias a la eliminación previa de las proteínas mayoritarias por inmunoadsorción, seguido de una separación por electroforesis bidimensional. En otros estudios, el proceso de extracción proteica está dirigido al estudio del proteoma de un orgánulo concreto. Por ejemplo, en el estudio de Bianchi y col. (9) el proceso de fraccionamiento previo que desarrollaron tenía como finalidad la extracción selectiva de los glóbulos de grasa de la leche de vaca para poder analizar su proteoma y estudiar su relación con el estado fisiopatológico de la glándula mamaria. Aparte del proteoma de la leche, también se ha estudiado la composición proteica de otros alimentos como el huevo (10), el trigo (11-13), el arroz (14), o el guisante (15). En todos ellos se realiza un estudio general del alimento, tratando de estudiar y caracterizar el máximo número de proteínas posible.

2.2. Evaluación de la autenticidad de de los alimentos.

Este tipo de estudios están en relación con la legislación vigente en materia de etiquetado. En la actualidad, los consumidores exigen que se les informe de manera clara y precisa sobre los alimentos que consumen, siendo dicha información un aspecto que influye de manera relevante en el criterio de compra. Esto es especialmente importante en el caso de los alimentos procesados donde la diferenciación visual de los distintos ingredientes no es evidente. Factores como el precio, el estilo de vida, la salud o la religión pueden influir en la decisión individual a la hora de comprar un alimento. Es por ello que existe una obligación de declarar los ingredientes de los alimentos en el etiquetado. La legislación en materia alimentaria debe proteger al consumidor frente a etiquetados falsos o incorrectos, algo que normalmente se lleva a cabo con objeto de aumentar el beneficio económico. Para conseguir esto es necesario disponer de métodos robustos, precisos y sensibles capaces de controlar esta forma de proceder (16). Los métodos tradicionalmente empleados para este propósito han sido principalmente los inmunoensayos y los análisis de ADN. Los inmunoensayos han sido comúnmente empleados en el control del fraude de los alimentos debido a su facilidad de manejo, su elevada sensibilidad y la capacidad de analizar un gran número de muestras al mismo tiempo. Su principal inconveniente es que es necesario disponer de anticuerpos altamente específicos, algo que no siempre es fácil de conseguir, lo que genera un riesgo de tener reacciones cruzadas. Otro problema importante es que el procesado de ciertos alimentos puede afectar al imnunoensayo debido a la desnaturalización de las proteínas, influyendo negativamente en el reconocimiento de éstas por parte del anticuerpo. Otra limitación de los inmunoensayos es que, al no estar basados en diferencias a nivel de secuencia, su poder discriminante es bajo a la hora de diferenciar ingredientes estrechamente relacionados, como la diferenciación de carnes de pollo y pavo, por ejemplo (17). Los métodos basados en el análisis de ADN, por el contrario, sí que presentan un elevado poder discriminante y sensibilidad puesto que se basan en la detección de fragmentos de ADN específicos de los componentes del alimento. El problema es que durante el procesado de los alimentos se puede producir la degradación del ADN, lo que afecta negativamente al tamaño de los fragmentos a amplificar, pudiendo por tanto reducir la especificidad del ensayo. Existe también el problema de que, al ser matrices complejas, la extracción del ADN resulta complicada, lo que dificulta la estandarización de los protocolos, teniendo que optimizarlos en cada caso particular para asegurar una correcta recuperación. Esto resulta especialmente problemático en el caso de los análisis cuantitativos (18;19).

Los avances de los últimos años en espectrometría de masas aplicada al análisis proteómico representan una alternativa seria e interesante a los métodos mencionados. Su poder discriminante es comparable a los análisis de ADN ya que se basan en la detección de péptidos/proteínas biomarcadores específicos del alimento a detectar. Sin embargo, la aproximación proteómica

presenta una mayor fiabilidad y robustez que los inmunoensayos y los análisis de ADN en el análisis de alimentos procesados. La secuencia primaria de aminoácidos de las proteínas es considerablemente más resistente que el ADN al procesado de los alimentos, lo que unido a una mayor facilidad de extracción, permitiría el desarrollo de métodos cuantificación más fiables (20).

Las aplicaciones de la proteómica en el control del fraude en los alimentos se han centrado principalmente en el análisis de la composición de los mismos. Así, Mazzeo y col. (21) desarrollaron una estrategia basada en el análisis por espectrometría de masas MALDI-TOF para identificar los perfiles proteicos característicos de distintas especies de pescado. El estudio de las parvalbúminas les permitió la identificación de proteínas biomarcadoras específicas de cada especie. En otro estudio, Carrera y col. (22) a través de una aproximación proteómica clásica consistente en la extracción de proteínas, electroforesis bidimensional, digestión en gel de las proteínas diana y análisis de *novo* de los péptidos generados mediante MALDI-TOF y cromatografía líquida acoplada a espectrometría de masas en tándem (LC-ESI-MS/MS) consiguieron caracterizar péptidos biomarcadores de las distintas especies de pescado de la familia de los Merlúcidos. De modo similar, también se ha llevado a cabo la caracterización de péptidos específicos de distintas especies de langostinos comerciales (23).

Aunque en menor medida, también se han descrito aplicaciones de la proteómica en la identificación de distintas especies de carne para el control del fraude que supone el empleo de carnes de menor valor económico en la fabricación de productos cárnicos. En un estudio pionero, Taylor y Linforth (24) ya vislumbraron en 1993 el potencial de la espectrometría de masas para identificar los distintos tipos de carne que pueden estar presentes en un alimento tomando como criterio discriminante las diferencias en los valores de masa de la hemoglobina y mioglobina de las diferentes especies animales. De esta manera, estos autores fueron capaces de detectar la presencia de un 10 % de hemoglobina de caballo en una mezcla con hemoglobina de vacuno, destacando no obstante la dificultad que supone el trabajar con mezclas reales de carne en vez de con proteínas puras. Recientemente, Sentandreu y col. (20) han desarrollado una estrategia distinta consistente en la extracción de proteínas musculares, enriquecimiento de proteínas diana mediante isoelectroenfoque en medio líquido, digestión con tripsina y posterior caracterización de péptidos biomarcadores específicos de especie mediante MALDI-TOF y LC-ESI-MS/MS. Un esquema del protocolo desarrollado por estos autores se ilustra en la Figura 2. Mediante esta aproximación ha sido posible detectar la presencia de un 0,5 % de carne de pollo en una mezcla con carne de cerdo. Aparte de su bajo límite de detección, el método tiene la ventaja de poder emplearse para la cuantificación tanto de carne fresca como cocinada mediante el empleo de péptidos marcados con isótopos estables (^{13}C, ^{15}N).

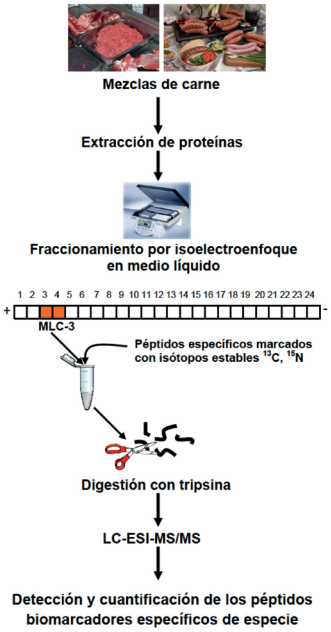


Figura 2: Aproximación proteómica desarrollada por Sentandreu y col.(20) para detectar cuantitativamente la presencia de carne de pollo en mezclas de carne y productos cárnicos. El método se basa en la identificación de péptidos específicos de la especie aviar (Gallus gallus) generados a partir de la digestión en solución con tripsina de la cadena ligera de miosina 3 (MLC-3), enriquecida previamente por isoelectroenfoque en medio líquido en el intervalo de pH 4-7. En estas condiciones, MLC-3 se concentra en las fracciones 3 y 4. El empleo de péptidos marcados con isótopos estables permite la cuantificacion de los péptidos biomarcadores, determinando así el porcentaje de carne de pollo en la mezcla inicial.

Se han desarrollado diferentes aproximaciones para poder identificar tanto las distintas especies de leche que se pueden emplear en la fabricación de quesos (25;26) como para detectar la posible presencia de adulterantes en la leche desnatada en polvo, como proteínas de soja y de guisante (27). En otro interesante estudio llevado a

cabo por Chambery y col. (28) se pone de manifiesto la utilidad de las herramientas de proteómica en la identificación de vinos de alta calidad mediante la digestión tríptica de las proteínas del vino y el posterior análisis por espectrometría de masas MALDI-TOF. De este modo, estos autores obtienen lo que ellos denominan los "códigos de masa" característicos de cada tipo de vino.

Además del control del fraude en la composición de alimentos, se han descrito algunas aplicaciones de la proteómica en la determinación del origen geográfico de los alimentos. Wang y col. (29) desarrollaron un método para determinar el origen geográfico de la miel proveniente de diferentes regiones del mundo. Para ello, estos autores realizaron una extracción de las proteínas de los distintos tipos de miel seguido de un análisis MALDI-TOF de los valores de masa de las proteínas intactas. El espectro de masas de proteínas obtenido para cada tipo de miel se empleó en obtener un código de barras de masas proteicas característico que permite discriminar entre las muestras de miel de distinto origen geográfico.

3. Proteómica nutricional.

El gran impacto que la dieta tiene en la salud humana es algo que se conoce desde la antigüedad. Así como la nutrición tradicional se ha ocupado de estudiar el aporte de nutrientes necesarios para alimentar adecuadamente a la población en general, hoy en día la nutrición particulariza en aspectos más concretos relacionados con promover la salud y/o evitar o retrasar la aparición de enfermedades. También considera la variabilidad humana en cuestión de gustos y necesidades nutricionales a la hora de comer, asumiendo la idea de que las personas responden de modo diferente a una misma dieta en función de sus características genéticas, su entorno o el estilo de vida que lleven. En este nuevo escenario el estudio proteómico aplicado a la investigación nutricional se está convirtiendo en una herramienta fundamental para estudiar la expresión de nuestro genoma en respuesta a una dieta concreta. Aldred y col. (30) llevaron a cabo un estudio de intervención humana para conocer el efecto de la suplementación en la dieta con alfa-tocoferol sobre la expresión del proteoma plasmático de individuos sanos. Las muestras de plasma de los individuos sometidos a suplementación con alfa-tocoferol se analizaron a los 14 y 28 días mediante electroforesis bidimensional y posterior identificación de las proteínas de interés mediante digestión tríptica y análisis de la huella peptídica por espectrometría de masas MALDI-TOF. De este modo se consiguió demostrar que la suplementación en la dieta con alfa-tocoferol durante 28 días produce un claro aumento de distintas isoformas de la apolipoproteína A1, lo que podría estar en relación con el efecto protector que se atribuye al alfa-tocoferol frente a las enfermedades cardiovasculares.

A través de un análisis proteómico similar, Arbones-Mainar y col. (31) llevaron a cabo un interesante estudio sobre el efecto que tiene en el proteoma hepático el suplementar con aceite de oliva virgen la dieta de ratones. Estos autores mostraron que la incorporación de aceite de oliva virgen a la dieta de los ratones influía en el metabolismo hepático de los lípidos aumentando los niveles de glutamato deshidrogenasa, glutatión sintasa e hidroxiacil glutatión hidrolasa, enzimas que pueden contribuir a disminuir el estrés oxidativo promovido por la esteatosis hepática y ejercer así un efecto protector frente al desarrollo de enfermedades lipotóxicas como la ateroesclerosis.

Aunque los estudios realizados no son numerosos por el momento, la proteómica ha demostrado ser una herramienta eficaz en la identificación de biomarcadores nutricionales en aspectos de relevancia para nuestra salud como la modulación de la inmunidad en función de la dieta (32) o el efecto protector de algunos alimentos frente a enfermedades como el cáncer (33), lo que representa una aproximación de futuro para conocer mejor la relación entre los alimentos y la salud.

En lo referente a la biodisponibilidad de los alimentos, la cromatografía líquida acoplada a la espectrometría de masas en tándem se ha empleado con éxito en estudios de digestibilidad de la carne de cerdo para poder identificar los péptidos generados por acción de las enzimas gastrointestinales sobre las proteínas del músculo. Después de una hidrólisis secuencial con pepsina y pancreatina, Escudero y col. (34) obtuvieron un hidrolizado de proteínas de carne de cerdo que, después de desproteinizarlo, fue analizado directamente por LC-ESI-MS/MS. Mediante esta estrategia se pudieron identificar 51 péptidos con un tamaño comprendido entre 6 y 16 aminoácidos provenientes de las principales proteínas del músculo. El hecho de ser ricos en prolina hace pensar que estos péptidos son probablemente resistentes a ser hidrolizados en fragmentos más pequeños durante la digestión. Una aproximación peptidómica similar fue utilizada por Boutrou y col. (35) para evaluar la resistencia de fosfopéptidos provenientes de la beta-caseína a la acción de las enzimas proteolíticas de la membrana intestinal. De este modo concluyeron que tanto las fosforilaciones como las uniones de coordinación con cationes como el zinc o el cobre confieren a los péptidos derivados de la caseína una mayor resistencia a la acción de las enzimas digestivas.

3.1. Péptidos con actividad biológica o funcional.

Como ya se ha comentado, aparte de aspectos estrictamente relacionados con la nutrición hay un interés creciente por investigar la bioquímica de alimentos y/o componentes alimentarios que permitan asegurar el estado de salud y reducir o retrasar el riesgo de padecer ciertas enfermedades, especialmente las denominadas degenerativas tales como diabetes, obesidad, hipertensión, infecciones gastrointestinales o algunos tipos de cáncer, que constituyen las principales causas de muerte en los países desarrollados. Esta propiedad de los alimentos de ejercer un beneficio directo sobre la salud mediante la regulación de diferentes procesos bioquímicos y/o biológicos es lo que se conoce como función fisiológica o terciaria.

La función fisiológica o terciaria de un número importante de componentes alimentarios ha sido objeto de estudio durante los últimos años (36-38). Dentro de éstos, los péptidos ocupan un lugar destacado ya que son constituyentes capaces de regular numerosas funciones biológicas en el organismo. Los péptidos que poseen estas características se les conoce como péptidos funcionales. Una de las propiedades que ha despertado mayor interés es la capacidad de algunos péptidos de regular la presión arterial en el organismo. Aunque numerosos trabajos de la literatura describen la identificación de péptidos antihipertensivos mediante el esquema clásico de extracción a partir del alimento, purificación mediante distintas etapas cromatográficas y elucidación de la secuencia mediante la degradación de Edman (39), los trabajos más recientes muestran un empleo cada vez mayor de la espectrometría de masas para obtener la secuencia de los mismos. La identificación de péptidos funcionales mediante espectrometría de masas no está exenta de dificultades ya que en muchos casos se trata de péptidos de pequeño tamaño (2-5 aminoácidos), de los que se necesita espectros de fragmentación MS/MS de gran calidad que aporten información suficiente para cubrir la totalidad de la secuencia y poder elucidarla sin recurrir necesariamente a la comparación de secuencias en bases de datos tal como hacen algunos motores de búsqueda. Por esta razón, no todos los espectrómetros de masas son adecuados para su empleo en este tipo de estudios.

Se han caracterizado péptidos antihipertensivos provenientes de hidrolizados de proteínas de alimentos empleando diferentes enzimas gastrointestinales. Bougatef y col. (40) han descrito recientemente la identificación de un péptido con actividad antihipertensiva proveniente de un hidrolizado de proteínas de pescado. Después de fraccionar el hidrolizado de proteínas mediante cromatografía de exclusión molecular y fase reversa, estos autores identificaron la secuencia Ala-Gly-Ser mediante la infusión directa de la fracción con mayor actividad antihipertensiva en un triple cuadrupolo. En el trabajo llevado a cabo por Escudero y col. (41), esquematizado en la Figura 3, se desarrollaron dos estrategias distintas y complementarias para la identificación de péptidos con actividad antihipertensiva a partir de un hidrolizado de carne de cerdo empleando

pepsina y pancreatina. En una primera aproximación, el hidrolizado desproteinizado se fraccionó mediante cromatografía liquida de fase reversa, recogiendo las fracciones eluídas de la columna y determinando su actividad antihipertensiva *in Vitro*. Posteriormente, aquellas fracciones que mostraron una actividad antihipertensiva destacada fueron analizadas por espectrometría de masas MALDI-TOF/TOF con objeto de poder identificar los péptidos contenidos en ellas. En una segunda aproximación, complementaria a la primera, el mismo hidrolizado de carne desproteinizada se analizó directamente por cromatografía de fase reversa acoplada a una fuente de nanoelectrospray y un espectrómetro de masas cuadrupolo-tiempo de vuelo. De los péptidos identificados por esta vía, aquellos que más se adecuaban a los criterios de tamaño y estructura de péptidos con conocida actividad antihipertensiva *in vitro* fueron sintetizados para la posterior evaluación de su capacidad de inhibir a la enzima convertidora de angiotensina (ECA).

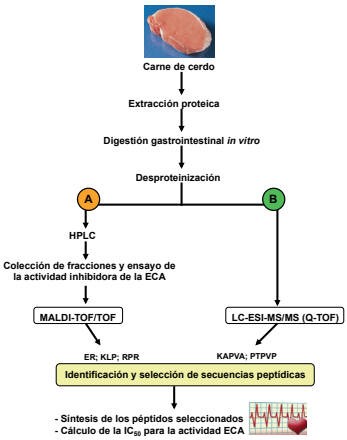


Figura 3:Estrategia desarrollada por Escudero y col.(41) para identificar péptidos con potencial actividad antihipertensiva generados a partir de la digestión gastrointestinal *in vitro* de la carne de cerdo. La identificación y selección de los péptidos se realizó en función de la capacidad de éstos de inhibir a la enzima convertidora de angiotensina (ECA) que forma parte del mecanismo de regulación de la presión arterial en el organismo.

La digestión de las proteínas de los alimentos por acción de las enzimas gastrointestinales no es la única vía de generación de péptidos bioactivos. En el trabajo realizado por Nielsen y col. (42) se ilustra la generación de péptidos con efecto antihipertensivo a partir de la fermentación de las proteínas de la leche por acción de distintas cepas de bacterias ácido lácticas normalmente empleadas en la fabricación del yogur y otros productos lácteos. En el trabajo de Meyer y col. (43) se estudió la generación de los tripéptidos antihipertensivos Val-Pro-Pro e Ile-Pro-Pro durante la maduración de distintos tipos de queso suizo mediante el empleo de un sistema HPLC-ESI-MS^3 integrado por una trampa de iones lineal.

Aparte de la actividad antihipertensiva, la espectrometría de masas también se ha empleado en la caracterización de péptidos provenientes de los alimentos con otras propiedades funcionales como la actividad antioxidante (44;45) o anticolesterolémica (46). Otra aplicación relevante de la proteómica es la identificación de péptidos antimicrobianos producidos por las bacterias ácido lácticas que se emplean normalmente durante los procesos de fermentación de alimentos, los conocidos comúnmente como bacteriocinas. El trabajo de Borrero y col. (47) es ilustrativo de las limitaciones de la degradación de Edman en la secuenciación de la garvimicina, una bacteriocina circular, frente a la utilidad de la secuenciación *de novo* empleando la espectrometría de masas en tándem.

4. Seguridad Alimentaria.

La aplicación de las herramientas proteómicas en seguridad alimentaria se ha centrado principalmente en dos grandes problemáticas: Detectar la posible contaminación microbiana en los alimentos y evaluar la seguridad de sus ingredientes para la salud humana.

Las aplicaciones encaminadas a detectar la contaminación microbiana de los alimentos se han centrado en desarrollar estrategias para la identificación del perfil proteico característico de cada tipo de bacteria. El objetivo principal es desarrollar métodos rápidos, sensibles y fiables para poder detectar los microorganismos que pueden afectar a la calidad y la seguridad de los

alimentos tanto en el procesado como durante su almacenamiento. El trabajo de Ayyadurai y col. (48) ilustra muy bien este tipo de aplicación. Estos autores emplearon la espectrometría de masas MALDI-TOF como un método rápido para obtener perfiles proteicos característicos de las distintas especies del género *Yersinia,* pudiendo así identificar de manera precisa las especies patógenas causantes de infecciones alimentarias y diferenciándolas de las especies inofensivas de *Yersinia* que pueden encontrarse en el ambiente. De un modo parecido, Fernández-No y col. (49) utilizaron esta estrategia para obtener perfiles proteicos capaces de identificar sin ambigüedad las especies bacterianas productoras de aminas biogénicas, responsables de la contaminación en pescado y otras especies marinas. Otra aplicación interesante de la proteómica en seguridad alimentaria se ilustra en el trabajo desarrollado por Bossi y col. (50) donde, mediante un estudio cuantitativo por electroforesis bidimensional y análisis de las proteínas significativas por LC-ESI-MS/MS, se evalúa la capacidad de los taninos presentes en el vino y otras bebidas de inhibir el crecimiento de bacterias indeseables que pueden afectar a la calidad de las mismas. Más concretamente, este trabajo representa la primera aproximación proteómica dirigida a evaluar la capacidad del ácido tánico en alterar la expresión de proteínas de *Lactobacillus hilgardii.*

Las aplicaciones encaminadas a evaluar la seguridad de los ingredientes alimentarios normalmente se han centrado en la caracterización y el control de ciertas sustancias tóxicas naturalmente presentes en los alimentos como las lecitinas presentes en los guisantes crudos o procesados incorrectamente, las cuales son responsables de la aparición de alteraciones gastrointestinales (51). El análisis proteómico también se ha empleado en el desarrollo de métodos alternativos y complementarios a los inmunoensayos para la detección específica de las proteínas del gluten de trigo. La intolerancia que padecen ciertas personas a dichas proteínas es lo que se conoce como enfermedad celíaca, la cual supone un grave problema para este grupo de población que se ve obligado a llevar una dieta libre de gluten para evitar problemas de inflamación intestinal y enteropatías. El trabajo realizado por Sealey-Voyksner y col. (52) demuestra la utilidad de la cromatografía líquida acoplada a la espectrometría de masas en tándem a la hora de disponer de un método preciso y sensible capaz de detectar y cuantificar péptidos diana específicos del gluten de trigo en un intervalo entre 0,01 – 100 ppm, adecuado para certificar que los productos etiquetados como libres de gluten cumplen con este requisito.

Otra aplicación interesante que refleja el potencial del análisis proteómico en garantizar la seguridad de alimentos es el método desarrollado por Ocaña y col. (53) para poder detectar, por espectrometría de masas, la presencia de proteínas de origen animal en alimentos que deberían de estar libres de ellas. A raíz del problema ocurrido con la aparición de la encefalopatía espongiforme bovina (conocida coloquialmente como “enfermedad de las vacas locas”), se prohibió el uso de esta fuente de proteínas para la alimentación de animales destinados al consumo humano. La estrategia que desarrollaron estos autores se basa en la detección de péptidos específicos de la gelatina de cerdo, obtenidos por digestión con HCl e identificados por MALDI-TOF y LC-ESI-MS/MS.

4.1. Alergias debidas a la ingestión de alimentos.

Cada vez se dan más casos de personas alérgicas a ciertos alimentos. A pesar de que no se puede hablar de un fenómeno masivo, lo cierto es que el número de individuos afectados por este problema ha aumentado de forma progresiva en los últimos años. A diferencia de otras formas de intolerancia alimentaria, las alergias se deben a una reacción inmunológica causada por ciertas proteínas de los alimentos que son reconocidas específicamente por las inmunoglobulinas E (IgE) producidas por los individuos susceptibles. El identificar correctamente las proteínas causantes de alergias en los alimentos resulta de vital importancia tanto para el diagnóstico clínico de la persona como para el desarrollo de nuevos alimentos o tecnologías de fabricación en un mercado determinado. En este sentido, la proteómica se ha aplicado con éxito en la caracterización de proteínas alergénicas provenientes de distintos alimentos. En este tipo de

estudios, más que el análisis masivo del proteoma de un alimento, se busca una aproximación dirigida que permita identificar específicamente las proteínas que causan hipersensibilidad. En este contexto, el combinar las aproximaciones proteómica e inmunológica puede ser de gran ayuda para la consecución de este objetivo, algo que se aprecia bien en el trabajo de Sotkovsky y col. (54). En él se llevó a cabo la identificación de las proteínas del trigo que son causantes de alergias; para ello, las proteínas de un extracto de trigo se separaron en una primera etapa por electroforesis mono- y bi-dimensional. En una segunda etapa, las proteínas separadas en los geles fueron transferidas a membranas para realizar un análisis por western-blot incubando dichas membranas con el suero de pacientes alérgicos al trigo. La unión específica de las IgE de los pacientes a proteínas específicas se visualizó posteriormente por revelado con un anticuerpo secundario, reconociendo así las proteínas inmunorreactivas. La identificación precisa de cada mancha inmunorractiva se llevó a cabo por escisión del gel, digestión con tripsina y análisis de los fragmentos empleando tanto la espectrometría de masas MALDI-TOF como un sistema LC-ESI-Q/TOF.

Además de las proteínas del trigo, la proteómica también se ha aplicado con éxito en la caracterización de las proteínas causantes de alergias en otros alimentos y bebidas como los cacahuetes (55), la leche (56), el vino (57), el apio (58) o el arroz (59), por poner algunos ejemplos.

4.2. Seguridad de los alimentos modificados genéticamente.

La capacidad para el análisis masivo de proteínas que posee la espectrometría de masas puede ser de gran utilidad a la hora de estudiar los cambios que se producen en el proteoma de un alimento modificado genéticamente con respecto al alimento convencional. Los alimentos modificados genéticamente, abreviadamente conocidos como GMO (del inglés "Genetic Modified Organism"), plantean problemas de seguridad alimentaria puesto que los cambios introducidos en el genoma pueden tener consecuencias difíciles de evaluar. De esta problemática surge el concepto de "equivalencia sustancial", que hace referencia al hecho de que un alimento modificado genéticamente tenga la misma composición que el alimento tradicional. Este concepto fue introducido por la OCDE en 1993, antes de que se aprobara la comercialización del primer alimento modificado genéticamente. Si se cumple el criterio de equivalencia sustancial, el alimento transgénico normalmente es considerado como seguro. Por el contrario, en caso de identificarse alguna diferencia se deberán realizar ensayos adicionales tanto de composición como inmunológicos y toxicológicos. El planteamiento que justifica todo esto es que el proceso de modificación genética puede dar lugar a alteraciones aleatorias de ciertos genes o de sus mecanismos de regulación que pueden generar alteraciones importantes no detectables en un análisis rutinario de composición nutricional. Para ello se necesitan estrategias analíticas más resolutivas y precisas como las herramientas proteómicas. Un buen ejemplo es el trabajo realizado por Corpillo y col. (60), donde se lleva a cabo un análisis proteómico para comprobar la equivalencia sustancial de tomates que han sido manipulados genéticamente para hacerlos resistentes a la acción de virus. Estos autores emplearon la electroforesis bidimensional y la espectrometría de masas MALDI-TOF para comparar el proteoma de la variedad transgénica con el proteoma de los tomates sin manipular. Al no encontrar diferencias ni en las proteínas identificadas ni en los niveles de expresión de las mismas, concluyeron que la variedad transgénica es sustancialmente equivalente al tomate natural y por tanto apta para el consumo. Por el contrario, Zolla y col. (61) realizaron un estudio proteómico comparando variedades de maíz transgénico con las variedades naturales, concluyendo que la manipulación genética iba acompañada de notables efectos colaterales puesto que modificaba la expresión de un buen número de proteínas en la adaptación de la planta a las condiciones ambientales.

Otros estudios han centrado el análisis proteómico en evaluar algún aspecto concreto de la seguridad de los alimentos modificados genéticamente. Batista y col. (62), por ejemplo, estudiaron las proteínas alergénicas de la soja transgénica, comparándolas con el perfil de proteínas

inmunorreactivas de la soja normal. El estudio concluyó que la modificación genética de la soja no alteraba la expresión de las proteínas alergénicas y por tanto no existía un riesgo añadido de alergia en la variedad transgénica.

Otra aplicación importante relacionada con la seguridad de los alimentos modificados genéticamente es el desarrollo de estrategias para diferenciar la variedad modificada de su homóloga natural en un mismo alimento. Así, Ocaña y col. (63) desarrollaron una estrategia dirigida a poder diferenciar la soja y el maíz transgénico mediante la identificación de la proteína expresada específicamente por éstas a consecuencia de la manipulación genética. Estos autores destacan la necesidad de desarrollar diferentes etapas de fraccionamiento/enriquecimiento para poder detectar la proteína transgénica en una mezcla conteniendo el 0,9 % de semillas de soja modificada mediante digestión con tripsina e identificación de los péptidos por MALDI-TOF y LC-ESI-QTOF.

5. Efectos debidos al procesado y almacenamiento de los alimentos.

Si el proteoma de los alimentos ya es de por sí suficientemente complejo, el procesado y almacenamiento de éstos pueden dar lugar a modificaciones postraduccionales que añaden aún más complejidad al estudio del mismo. Las principales modificaciones de las proteínas en estas etapas se relacionan con procesos de oxidación, con reacciones de Maillard o con hidrólisis proteica por rotura de enlaces peptídicos. Todo esto puede tener importantes consecuencias en las propiedades nutritivas, tecnológicas o toxicológicas de los alimentos. La utilidad de las herramientas proteómicas en el estudio y control de estos procesos se pude apreciar en una serie de casos ilustrativos, como el trabajo llevado a cabo por Meltretter y col. (64), donde la espectrometría de masas MALDI-TOF se empleó con éxito en estudiar los cambios que se producen en las proteínas de la leche después de someterla a distintos tratamientos térmicos. Concretamente, estos autores se centraron en estudiar los productos derivados de la reacción de Maillard entre las proteínas de la leche y la lactosa. La conclusión es que la aproximación proteómica resulta fiable, más sencilla y rápida que otros métodos empleados en el estudio de la formación de estos productos. En el trabajo de Álvarez y col. (65) se pone de manifiesto la utilidad de la espectrometría de masas como herramienta para estudiar los cambios que pueden sufrir las proteínas de los alimentos sometidos a tratamientos por altas presiones. Estos autores constatan cómo las altas presiones pueden afectar de modo distinto la estructura terciaria de las proteínas del suero lácteo según sea la procedencia de dichas proteínas.

Bobalova y Chmelik (66) llevaron a cabo una interesante aproximación proteómica para poder identificar la glicosilación que sufren las proteínas de la cebada durante el proceso de malteado. El trabajo combina adecuadamente el desarrollo de un método rápido de fraccionamiento del extracto de malta empleando una columna monolítica de intercambio aniónico con la identificación, mediante espectrometría de masas MALDI-TOF/TOF, de las proteínas modificadas al glicosilarse.

Uno de los principales cambios bioquímicos que tienen lugar durante el procesado de alimentos como el queso o los productos cárnicos es la hidrólisis de las proteínas por acción de diferentes grupos de enzimas proteolíticas. En la fabricación del queso, por ejemplo, las caseínas de la leche son hidrolizadas por acción de distintas endo- y exo-peptidasas liberadas por las bacterias lácticas empleadas en el proceso. Sforza y col. (67) ya evidenciaron en al año 2003 la utilidad de la espectrometría de masas acoplada a la separación por HPLC en la caracterización de los péptidos que se generan durante la maduración del queso Grana Padano debido a esta proteolisis. Según estos autores, las secuencias identificadas en este trabajo podrían tener una

potencial actividad biológica, lo que confiere un interés añadido a la caracterización de estos productos de hidrólisis proteica de los alimentos.

Durante el proceso de maduración de la carne también se produce una importante degradación de las principales proteínas constituyentes del músculo. Después del sacrificio del animal, las propias enzimas proteolíticas del músculo continúan siendo activas, hidrolizando las proteínas clave que conforman las miofibrillas. Esto da lugar a una disminución de la resistencia mecánica del músculo, lo que está directamente relacionado con el aumento de la terneza de la carne que representa el atributo de calidad más apreciado por los consumidores. Todo este proceso es enormemente complejo y depende de numerosos factores que por el momento no se conocen en profundidad, lo que conlleva que la industria cárnica se enfrente a un importante problema de elevada variabilidad en la terneza final de las canales y de imposibilidad de predecir y estandarizar tal variabilidad. En los últimos años la aplicación de las herramientas proteómicas ha contribuido de manera significativa a avanzar en la comprensión de estos complejos mecanismos bioquímicos relacionados con la conversión del músculo en carne. Esto se aprecia bien en el trabajo realizado por Lametsch y col. (68), donde se emplea una aproximación proteómica clásica (extracción de proteínas seguida de electroforesis bidimensional e identificación de manchas relevantes por MALDI-TOF) para estudiar los cambios en el proteoma del músculo porcino durante la maduración de la carne y su relación con el desarrollo de la terneza a las 72 h postmortem. De este modo concluyeron que los principales cambios durante la maduración se relacionaban con la generación de fragmentos de actina y de otras proteínas miofibrilares como la miosina o la titina. En un estudio sobre la maduración de la carne de vacuno, Sawdy y col. (69) también emplearon la espectrometría de masas para identificar dos fragmentos de la cadena pesada de miosina que se correlacionaban con el desarrollo de la terneza de la carne a los 7 días.

Al igual que con la maduración de la carne, la proteómica ha jugado un papel relevante en la caracterización de la proteolisis postmortem que tiene lugar durante el curado de los productos cárnicos como el jamón (70). A diferencia de la maduración de la carne, periodo que se suele extender hasta dos semanas después del sacrificio, el curado del jamón se prolonga durante periodos de 12 meses o incluso más. Esto tiene como consecuencia una degradación de las proteínas del músculo mucho más intensa, dando lugar a la generación de péptidos de pequeño tamaño que son responsables, junto con otros compuestos, de la generación del aroma y sabor típico de estos productos. Las estrategias desarrolladas para la identificación de estos oligopéptidos se han basado principalmente en etapas sucesivas de fraccionamiento por técnicas cromatográficas y posterior identificación directa mediante LC-ESI-MS/MS (71).

6. Conclusiones y perspectivas futuras.

A pesar de que actualmente el uso de la tecnología proteómica en ciencia de alimentos todavía es limitado, la tendencia de los últimos años indica que ésta es una aplicación con gran potencial de futuro. Como se puede apreciar en la Figura 1, las publicaciones donde se hace uso de las distintas herramientas proteómicas disponibles para el estudio y caracterización de alimentos han aumentado de forma significativa en los últimos diez años.

Las aplicaciones que se describen en los diferentes apartados del presente capítulo permiten tener una idea bastante precisa de cómo el análisis proteómico puede jugar un papel decisivo a la hora de tener un conocimiento más completo y preciso de la composición de los alimentos, tanto desde el punto de vista nutricional como de la actividad fisiológica que sus componentes pueden ejercer en nuestro organismo. La proteómica alimentaria puede convertirse en un vehí-

culo eficaz para entender mejor las complejas interacciones que existen entre la composición proteica de los alimentos y los procesos de producción y almacenamiento de los mismos. Como se puede comprobar, las aplicaciones proteómicas en alimentación incluyen tanto visiones generales del proteoma como estudios dirigidos hacia un grupo particular de proteínas o péptidos. El estudio general nos permitirá tener una visión de conjunto del proteoma de los alimentos y de los cambios que éste puede sufrir como consecuencia de los diferentes procesos de producción. Los estudios dirigidos hacia la identificación de proteínas o péptidos específicos demuestran, por otro lado, ser una herramienta con gran potencial en aspectos como la caracterización de componentes de los alimentos con actividad funcional, beneficiosa para nuestro organismo, o en la identificación de componentes proteicos causantes de efectos negativos para la salud como alergias e intolerancias. La proteómica alimentaria tiene igualmente un gran potencial en garantizar la calidad de los alimentos mediante la certificación de un tipo determinado de composición o una tecnología concreta de producción, lo que sin duda puede contribuir al diseño de nuevos alimentos y tecnologías de fabricación.

REFERENCIAS.

1. Gaso-Sokac D, Kovac S, Josic D. Application of Proteomics in Food Technology and Food Biotechnology: Process Development, Quality Control and Product Safety. Food Technology and Biotechnology 2010 Jul;48(3):284-95.

2. Pedreschi R, Hertog M, Lilley KS, Nicolai B. Proteomics for the Food Industry: Opportunities and Challenges. Critical Reviews in Food Science and Nutrition 2010;50(7):680-92.

3. Mamone G, Picariello G, Caira S, Addeo F, Ferranti P. Analysis of food proteins and peptides by mass spectrometry-based techniques. Journal of Chromatography A 2009 Oct 23;1216(43):7130-42.

4. Daniel H. Genomics and proteomics: importance for the future of nutrition research. British Journal of Nutrition 2002 May;87:S305-S311.

5. Velazquez-Arellano A, Encarnacion S. Role of proteomics in the study of nutrition. J Nutrigenet Nutrigenomics 2008;1(1-2):55-8.

6. Alomirah HF, Alli I, Konishi Y. Applications of mass spectrometry to food proteins and peptides. Journal of Chromatography A 2000 Sep 29;893(1):1-21.

7. Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. Nature Genetics 2003 Mar;33:311-23.

8. Yamada M, Murakami K, Wallingford JC, Yuki Y. Identification of low-abundance proteins of bovine colostral and mature milk using two-dimensional electrophoresis followed by microsequencing and mass spectrometry. Electrophoresis 2002 Apr;23(7-8):1153-60.

9. Bianchi L, Puglia M, Landi C, Matteoni S, Perini D, Armini A, et al. Solubilization methods and reference 2-DE map of cow milk fat globules. J Proteomics 2009 Jul 21;72(5):853-64.

10. Raikos V, Hansen R, Campbell L, Euston SR. Separation and identification of hen egg protein isoforms using SDS-PAGE and 2D gel electrophoresis with MALDI-TOF mass spectrometry. Food Chemistry 2006;99(4):702-10.

11. Gao LY, Ma WJ, Chen J, Wang K, Li J, Wang SL, et al. Characterization and Comparative Analysis of Wheat High Molecular Weight Glutenin Subunits by SDS-PAGE, RP-HPLC, HPCE, and MALDI-TOF-MS. Journal of Agricultural and Food Chemistry 2010;58(5):2777-86.

12. Liu L, Wang AL, Appels R, Ma JH, Xia XC, Lan P, et al. A MALDI-TOF based analysis of high molecular weight glutenin subunits for wheat breeding. Journal of Cereal Science 2009;50(2):295-301.

13. Zhang Q, Dong YM, An XL, Wang AL, Zhang YZ, Li XH, et al. Characterization of HMW glutenin subunits in common wheat and related species by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). Journal of Cereal Science 2008;47(2):252-61.

14. Aki T, Yanagisawa S. Application of rice nuclear proteome analysis to the identification of evolutionarily conserved and glucose-responsive nuclear proteins. J Proteome Res 2009 Aug; 8(8):3912-24.

15. Bourgeois M, Jacquin F, Savois V, Sommerer N, Labas V, Henry C, et al. Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. Proteomics 2009 Jan;9(2):254-71.

16. Sentandreu MA, Sentandreu E. Peptide biomarkers as a way to determine meat authenticity. Meat Science 2011 Nov;89(3):280-5.

17. Hsieh YHP, Sheu SC, Bridgman RC. Development of a monoclonal antibody specific to cooked mammalian meats. Journal of Food Protection 1998 Apr;61(4):476-81.

18. Woolfe M, Primrose S. Food forensics: using DNA technology to combat misdescription and fraud. Trends in Biotechnology 2004 May;22(5):222-6.

19. Paredi G, Sentandreu MA, Mozzarelli A, Fadda S, Hollung K, de Almeida AM. Muscle and meat: New horizons and applications for proteomics on a farm to fork perspective. Journal of Proteomics 2013 Aug 2;88:58-82.

20. Sentandreu MA, Fraser PD, Halket J, Patel R, Bramley PM. A Proteomic-Based Approach for Detection of Chicken in Meat Mixes. Journal of Proteome Research 2010 Jul;9(7):3374-83.

21. Mazzeo MF, De Giulio B, Guerriero G, Ciarcia G, Malorni A, Russo GL, et al. Fish Authentication by MALDI-TOF Mass Spectrometry. Journal of Agricultural and Food Chemistry 2008;56(23):11071-6.

22. Carrera M, Cañas B, Piñeiro C, Vázquez J, Gallardo JM. De novo mass spectrometry sequencing and characterization of species-specific peptides from nucleoside diphosphate kinase B for the classification of commercial fish species belonging to the family Merlucciidae. Journal of Proteome Research 2007 Aug;6(8):3070-80.

23. Ortea I, Cañas B, Calo-Mata P, Barros-Velazquez J, Gallardo JM. Arginine Kinase Peptide Mass Fingerprinting as a Proteomic Approach for Species Identification and Taxonomic Analysis of Commercially Relevant Shrimp Species. Journal of Agricultural and Food Chemistry 2009;57(13):5665-72.

24. Taylor AJ, Linforth R, Weir O, Hutton T, Green B. Potential of Electrospray Mass-Spectrometry for Meat Pigment Identification. Meat Science 1993;33(1):75-83.

25. Cuollo M, Caira S, Fierro O, Pinto G, Picariello G, Addeo F. Toward milk speciation through the monitoring of casein proteotypic peptides. Rapid Communications in Mass Spectrometry 2010 Jun;24(11):1687-96.

26. Pappa EC, Robertson JA, Rigby NM, Mellon F, Kandarakis I, Mills ENC. Application of proteomic techniques to protein and peptide profiling of Teleme cheese made from different types of milk. International Dairy Journal 2008;18(6):605-14.

27. Luykx DMAM, Cordewener JHG, Ferranti P, Frankhuizen R, Bremer MGEG, Hooijerink H, et al. Identification of plant proteins in adulterated skimmed milk powder by high-performance liquid chromatography-mass spectrometry. Journal of Chromatography A 2007 Sep 14;1164(1-2):189-97.

28. Chambery A, del Monaco G, Di Maro A, Parente A. Peptide fingerprint of high quality Campania white wines by MALDI-TOF mass spectrometry. Food Chemistry 2009;113(4):1283-9.

29. Wang J, Kliks MM, Qu WY, Jun SJ, Shi GY, Li QX. Rapid Determination of the Geographical Origin of Honey Based on Protein Fingerprinting and Barcoding Using MALDI TOF MS. Journal of Agricultural and Food Chemistry 2009;57(21):10081-8.

30. Aldred S, Sozzi T, Mudway I, Grant MM, Neubert H, Kelly FJ, et al. Alpha tocopherol supplementation elevates plasma apolipoprotein A1 isoforms in normal healthy subjects. Proteomics 2006 Mar;6(5):1695-703.

31. Arbones-Mainar JM, Ross K, Rucklidge GJ, Reid M, Duncan G, Arthur JR, et al. Extra virgin olive oils increase hepatic fat accumulation and hepatic antioxidant protein levels in APOE-/- mice. J Proteome Res 2007 Oct;6(10):4041-54.

32. Admyre C, Johansson SM, Qazi KR, Filen JJ, Lahesmaa R, Norman M, et al. Exosomes with immune modulatory features are present in human breast milk. J Immunol 2007 Aug 1;179(3):1969-78.

33. Breikers G, van Breda SG, Bouwman FG, van Herwijnen MH, Renes J, Mariman EC, et al. Potential protein markers for nutritional health effects on colorectal cancer in the mouse as revealed by proteomics analysis. Proteomics 2006 May;6(9):2844-52.

34. Escudero E, Sentandreu MA, Toldrá F. Characterization of Peptides Released by in Vitro Digestion of Pork Meat. Journal of Agricultural and Food Chemistry 2010 Apr 28;58(8):5160-5.

35. Boutrou R, Coirre E, Jardin J, Leonil J. Phosphorylation and Coordination Bond of Mineral Inhibit the Hydrolysis of the beta-Casein (1-25) Peptide by Intestinal Brush-Border Membrane Enzymes. Journal of Agricultural and Food Chemistry 2010;58(13):7955-61.

36. Shahidi F. Nutraceuticals and functional foods: whole versus processed foods. Trends in Food Science & Technology: 20 (9, Natural and Safe Foods) 376-387 2009;20(9, Natural and Safe Foods):376-87.

37. Kussmann M, Affolter M. Proteomics at the center of nutrigenomics: Comprehensive molecular understanding of dietary health effects. Nutrition: 25 (11-12) 1085-1093 2009;25(11-12):1085-93.

38. Iwaniak A, Minkiewicz P. Biologically active peptides derived from proteins - a review. Polish Journal of Food and Nutrition Sciences: 58 (3) 289-294 2008;58(3):289-94.

39. Arihara K, Nakashima Y, Mukai T, Ishikawa S, Itoh M. Peptide inhibitors for angiotensin I-converting enzyme from enzymatic hydrolysates of porcine skeletal muscle proteins. Meat Science 2001;57(3):319-24.

40. Bougatef A, Balti R, Nedjar-Arroume N, Ravallec R, Adje EY, Souissi N, et al. Evaluation of angiotensin I-converting enzyme (ACE) inhibitory activities of smooth hound (Mustelus mustelus) muscle protein hydrolysates generated by gastrointestinal proteases: identification of the most potent active peptide. European Food Research and Technology 2010;231(1):127-35.

41. Escudero E, Sentandreu MA, Arihara K, Toldrá F. Angiotensin I-Converting Enzyme Inhibitory Peptides Generated from in Vitro Gastrointestinal Digestion of Pork Meat. Journal of Agricultural and Food Chemistry 2010 Mar 10;58(5):2895-901.

42. Nielsen MS, Martinussen T, Flambard B, Sorensen KI, Otte J. Peptide profiles and angiotensin-I-converting enzyme inhibitory activity of fermented milk products: Effect of bacterial strain, fermentation pH, and storage time. International Dairy Journal 2009;19(3):155-65.

43. Meyer J, Butikofer U, Walther B, Wechsler D, Sieber R. Hot topic: Changes in angiotensin-converting enzyme inhibition and concentrations of the tripeptides Val-Pro-Pro and Ile-Pro-Pro during ripening of different Swiss cheese varieties. Journal of Dairy Science 2009 Mar;92(3):826-36.

44. Apaporn B, Chantragan S, Aphichart K, Polkit S. An antioxidant protein in Curcuma comosa Roxb. Rhizomes. Food Chemistry: 124 (2) 476-480 2011;124(2):476-80.

45. Gomez-Ruiz JA, Lopez-Exposito I, Pihlanto A, Ramos M, Recio I. Antioxidant activity of ovine casein hydrolysates: identification of active peptides by HPLC-MS/MS. European Food Research and Technology 2008;227(4):1061-7.

46. Zhong F, Zhang XM, Ma JG, Shoemaker CF. Fractionation and identification of a novel hypocholesterolemic peptide derived from soy protein Alcalase hydrolysates. Food Research International 2007;40(6):756-62.

47. Borrero J, Brede DA, Skaugen M, Diep DB, Herranz C, Nes IF, et al. Characterization of Garvicin ML, a Novel Circular Bacteriocin Produced by Lactococcus garvieae DCC43, Isolated from Mallard Ducks (Anas platyrhynchos). Applied and Environmental Microbiology 2011 Jan;77(1):369-73.

48. Ayyadurai S, Flaudrops C, Raoult D, Drancourt M. Rapid identification and typing of Yersinia pestis and other Yersinia species by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. Bmc Microbiology 2010;10.

49. Fernandez-No IC, Bohme K, Gallardo JM, Barros-Velazquez J, Cañas B, Calo-Mata P. Differential characterization of biogenic amine-producing bacteria involved in food poisoning using MALDI-TOF mass fingerprinting. Electrophoresis 2010 Mar;31(6):1116-27.

50. Bossi A, Rinalducci S, Zolla L, Antonioli P, Righetti PG, Zapparoli G. Effect of tannic acid on Lactobacillus hilgardii analysed by a proteomic approach. J Appl Microbiol 2007 Mar;102(3):787-95.

51. Nasi A, Picariello G, Ferranti P. Proteomic approaches to study structure, functions and toxicity of legume seeds lectins. Perspectives for the assessment of food quality and safety. Journal of Proteomics 2009 Apr 13;72(3):527-38.

52. Sealey-Voyksner JA, Khosla C, Voyksner RD, Jorgenson JW. Novel aspects of quantitation of immunogenic wheat gluten peptides by liquid chromatography-mass spectrometry/mass spectrometry. Journal of Chromatography A 2010 Jun 18;1217(25):4167-83.

53. Ocaña MF, Neubert H, Przyborowska A, Parker R, Bramley P, Halket J, et al. BSE Control: Detection of gelatine-derived peptides in animal feed by mass spectrometry. Analyst 2004 Feb;129(2):111-5.

54. Sotkovsky P, Hubalek M, Hernychova L, Novak P, Havranova M, Setinova I, et al. Proteomic analysis of wheat proteins recognized by IgE antibodies of allergic patients. Proteomics 2008 Apr;8(8):1677-91.

55. Chassaigne H, Norgaard JV, van Hengel AJ. Proteomics-based approach to detect and identify major allergens in processed peanuts by capillary LC-Q-TOF (MS/MS). Journal of Agricultural and Food Chemistry 2007;55(11):4461-73.

56. Bertino E, Gastaldi D, Monti G, Baro C, Fortunato D, Perono GL, et al. Detailed proteomic analysis on DM: insight into its hypoallergenicity. Front Biosci (Elite Ed) 2010;2:526-36.

57. Wigand P, Tenzer S, Schild H, Decker H. Analysis of Protein Composition of Red Wine in Comparison with Rose and White Wines by Electrophoresis and High-Pressure Liquid Chromatography-Mass Spectrometry (HPLC-MS). Journal of Agricultural and Food Chemistry 2009;57(10):4328-33.

58. Faeste CK, Jonscher KR, Sit L, Klawitter J, Lovberg KE, Moen LH. Differentiating cross-reacting allergens in the immunological analysis of celery (Apium graveolens) by mass spectrometry. Journal of AOAC International: 93 (2) 451-461 2010;93(2):451-61.

59. Koller A, Washburn MP, Lange BM, Andon NL, Deciu C, Haynes PA, et al. Proteomic survey of metabolic pathways in rice. Proceedings of the National Academy of Sciences of the United States of America 2002 Sep 3;99(18):11969-74.

60. Corpillo D, Gardini G, Vaira AM, Basso M, Aime S, Accotto GR, et al. Proteomics as a tool to improve investigation of substantial equivalence in genetically modified organisms: The case of a virus-resistant tomato. Proteomics 2004 Jan;4(1):193-200.

61. Zolla L, Rinalducci S, Antonioli P, Righetti PG. Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a result of genetic modifications. Journal of Proteome Research 2008 May;7(5):1850-61.

62. Batista R, Martins I, Jeno P, Ricardo CP, Oliveira MM. A proteomic study to identify soya allergens--the human response to transgenic versus non-transgenic soya samples. Int Arch Allergy Immunol 2007;144(1):29-38.

63. Ocaña MF, Fraser PD, Patel RKP, Halket JM, Bramley PM. Mass spectrometric detection of CP4 EPSPS in genetically modified soya and maize. Rapid Communications in Mass Spectrometry 2007;21(3):319-28.

64. Meltretter J, Birlouez-Aragon I, Becker CM, Pischetsrieder M. Assessment of heat treatment of dairy products by MALDI-TOF-MS. Molecular Nutrition & Food Research 2009;53(12):1487-95.

65. Alvarez PA, Ramaswamy HS, Ismail AA. Effect of high-pressure treatment on the electrospray ionization mass spectrometry (ESI-MS) profiles of whey proteins. International Dairy Journal 2007;17(8):881-8.

66. Bobalova J, Chmelik J. Proteomic identification of technologically modified proteins in malt by combination of protein fractionation using convective interaction media and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Journal of Chromatography A 2007 Sep 7;1163(1-2):80-5.

67. Sforza S, Ferroni L, Galaverna G, Dossena A, Marchelli R. Extraction, semi-quantifcation, and fast on-line identification of oligopeptides in Grana Padano cheese by HPLC-MS. Journal of Agricultural and Food Chemistry 2003;51(8):2130-5.

68. Lametsch R, Karlsson A, Rosenvold K, Andersen HJ, Roepstorff P, Bendixen E. Postmortem proteome changes of porcine muscle related to tenderness. Journal of Agricultural and Food Chemistry 2003 Nov 19;51(24):6992-7.

69. Sawdy JC, Kaiser SA, St Pierre NR, Wick MP. Myofibrillar 1-D fingerprints and myosin heavy chain MS analyses of beef loin at 36 h postmortem correlate with tenderness at 7 days. Meat Science 2004;67(3):421-6.

70. Di Luccia A, Picariello G, Cacace G, Scaloni A, Faccia M, Liuzzi V, et al. Proteomic analysis of water soluble and myofibrillar protein changes occurring in dry-cured hams. Meat Science 2005 Mar;69(3):479-91.

71. Mora L, Valero ML, del Pino MMS, Sentandreu MA, Toldra F. Small peptides released from muscle glycolytic enzymes during dry-cured ham processing. Journal of Proteomics 2011 Apr 1;74(4):442-50.

METALÓMICA. Aspectos medioambientales

30

José Luis Gómez Ariza y Tamara García Barrera[1]

[1]Departamento de Química y Ciencia de los Materiales. Facultad de Ciencias Experimentales. Campus de El Carmen, Huelva, Spain.

1. Los metales en los seres vivos. METALÓMICA.

Los metales desempeñan un papel esencial en el metabolismo de los organismos vivos, y a diferencia de los metabolitos convencionales, estos elementos no se producen ni se consumen en las reacciones. Por ello, el estudio del metabolismo de los metales de transición constituye una interfase única entre la química y la biología. La disponibilidad de estos elementos y sus químicas singulares constituyen fuerzas impulsoras de la evolución de la vida en la Tierra (1). Este hecho explica la existencia de organismos extremófilos y la variedad de organismos que surgen bajo condiciones geoquímicas diferentes (2).

Es interesante considerar la biología de sistemas relacionada con el metabolismo de los metales de transición, en la que el metaloma se define como el contenido de metales en la célula, su especiación y la localización en cada célula, tejido u organismo (3). Un estudio sistemático del metabolismo de los metales de transición requiere el conocimiento de cómo los organismos detectan, incorporan y usan estos metales (1).

Los iones metálicos generan uniones fuertes entre proteínas que superan los problemas estéricos, las repulsiones electrostáticas y otras interacciones no covalentes que impiden la formación de asociaciones entre proteínas monoméricas (4). La importancia de los iones metálicos en los sistemas biológicos se pone de manifiesto por su influencia en el comportamiento de más de un 50 % de las proteínas (4), y en el hecho de que las metaloproteínas representan, aproximadamente, el 30 % de las proteínas conocidas (5, 6).

Las metaloproteínas utilizan las propiedades únicas de los metales, en conjunción con el conjunto de macromoléculas presentes en los seres vivos, para desarrollar los importantes procesos que hacen posible la vida (6). Ello refleja la capacidad de las células eucarióticas para detectar e interaccionar con los metales bajo tres conceptos: afinidad, alosterismo y accesibilidad (7). Además, la compresión del papel biológico de las metaloproteínas en sus entornos nativos requiere el estudio de sus complejas interacciones en la célula, en cuyo interior existe un medio viscoso que contiene elevadas concentraciones otras muchas moléculas, las cuales originan numerosas reacciones competitivas (8).

Se ha comprobado que microorganismos como la *Escherichia coli*, adquieren elementos del medio de cultivo y los concentran alcanzando varios órdenes de magnitud hasta valores del orden de 2×10^5 átomos por célula, así el Zn y el Fe presentan una concentración total aproximada 0,1

mM en este tipo de células (9). Otros metales también se concentran en la *E. coli* hasta alcanzar un nivel fijo de concentración, no en forma libre en el citosol, ya que habitualmente los metales se incorporan a proteínas específicas mediante mecanismos moleculares que evitan los problemas de toxicidad asociados a la presencia de esos metales a las concentraciones encontradas (10). Por tanto, la regulación de los metales en las células difiere considerablemente respecto a la observada *in vitro*, al estar las uniones metal-biomolécula completamente moduladas por los filtros moleculares biológicos (11). En este escenario, es fundamental el papel desempeñado por las proteínas "acompañantes" (metalochaperonas) que captan los metales, los conducen, y los distribuyen entre las enzimas y biomoléculas que las necesitan, en el entorno celular adecuado y en el momento preciso para que ejerzan su función, preparándolas, asimismo, para que se produzcan las interacciones necesarias (12). Asimismo, pueden destacarse las proteínas transportadoras, que "conducen" dichos metales, especialmente en los procesos de translocación celular e interorgánica, contribuyendo de manera decisiva en el tráfico de los metales, su homeostais, señalización celular, detoxificación y metabolismo (Figura 1).

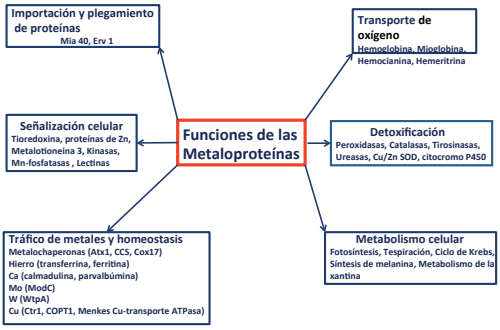


Figura 1. Papel de las protreínas en el metabolismo de las células.

En este contexto se sitúan los conceptos de metaloma y metalómica. El metaloma hace referencia al conjunto de especies de metales y metaloides presentes en una célula o tejido, considerando su identidad, cantidad y localización (5). El estudio de estos sistemas puede abordarse considerando la presencia de los diversos elementos, es decir, la relación de concentraciones de todos los elementos (heteroelementos-elementos diferentes de C, H, O, N), en un sistema biológico, célula, orgánulo, fluido o tejido, que se denomina también ionoma (13). También puede considerarse la presencia de complejos metálicos asociados a un determinado tipo de biomolécula, por ejemplo, el metaloproteoma (conjunto de metaloproteínas) o metalometaboloma (conjunto de metalometabolitos). El metaloma debe considerarse como un concepto dinámico, ya que las entidades vivas están en constante cambio y sufren la interacción del medio en que se encuentran, por lo que no puede hablarse de un metaloma concreto asociado a un organismo vivo concreto, y sólo es posible hacerlo en unas determinadas condiciones o situación.

La metalómica es la forma de estudiar y caracterizar el metaloma (la metodología), así como las interacciones y relaciones funcionales de los metales y sus especies con las biomoleculas, y los diversos componentes celulares, como genes, orgánulos, proteínas, metabolitos, y otras moléculas. Sin embargo, un estudio metalómico no puede limitarse a la identificación y análisis de metales o especies metálicas en los sistemas biológicos, sino que debe además establecer su implicación y función dentro de dichos sistemas. Ello requiere el estudio de moléculas de muy diverso tipo, pequeños metalometabolitos como la selenometionina o selenoetionina, metaloproteínas, ácidos nucleicos que contienen fósforo, y otras, lo que nos da una idea de su amplio campo de aplicación.

2. Los metales en moléculas de interés biológico.

De los veintiséis elementos reconocidos en la actualidad como esenciales, once son esenciales para toda forma de vida (C, O, H, N, Na, K, Ca, Mg, P, S y Cl) y quince son esenciales para la mayoría de sistemas biológicos, y se clasifican en elementos traza (requeridos en la dieta en

mg/día; Fe, Cu y Zn) y ultratraza (requeridos en la dieta en µg/día): V, Cr, Mn, Co, Ni, Mo, W, B, Si, Se, F, I). Algunos elementos tienen también importancia biológica ya que pueden utilizarse en drogas o sondas que los contienen como el Y, Cr, Tc, Co, Pt, Ag, Au, Cd y Hg. Un ejemplo es el cisplatino empleado como tratamiento contra el cáncer (14-17).

Shoeder *et al.* considera que los elementos esenciales pueden clasificarse en dos grupos: (i) elementos que participan en reacciones bioquímicas (elementos traza esenciales) y (ii) elementos con otras funciones. Este último grupo puede dividirse a su vez en aquellos elementos que no causan daño a los seres vivos a las concentraciones en las que habitualmente se encuentran en el medio ambiente, y elementos perjudiciales. De esta forma, existen elementos con funciones esenciales probadas (F, Si, Cr, Mn, Fe, Co, Ni, Cu, Zn, Se, Mo y I) que producen deficiencias nutricionales en humanos (F, Cr, Fe, Cu, Zn, Se y I) o animales (Si, Mn, Co, Ni y Mo), y otros elementos con función esencial sospechada pero mecanismo de acción desconocido (V, As, B, Br, Cd, Li, Pb y Sn). Finalmente, existen elementos como el Hg, que no tienen ninguna función biológica conocida. Sin embargo, la clasificación del carácter esencial de los elementos no es absoluta ya que muchos elementos considerados históricamente como tóxicos se consideran en la actualidad esenciales como ocurre con el Cr, Se (16, 18) y W, el cual se ha incluido recientemente a la lista de elementos que pueden encontrarse en los seres vivos (14).

La Tabla 1 muestra una clasificación de los elementos en relación a su función biológica (16, 19). Como puede observarse, existen algunos elementos que presentan un doble carácter esencial-tóxico como el As, Se y Cr, dependiendo de las diferentes especies químicas en las que puede encontrarse (p.ej. el As inorgánico es tóxico pero la arsenobetaína es inocua). Además, en la actualidad, numerosos trabajos se basan en el estudio de las interacciones (antagónicas, sinérgicas o competitivas) entre elementos y así, se sabe que el Se (20) y Zn (21) tienen efecto protector frente a la acción tóxica del Hg o Cd (22).

Tabla 1. Clasificación de los elementos en relación a su carácter esencial.

	As	S	Bi	B	Br	Cd	Ca	C	Zn	Cl	Co	Cu	Cr	Sn	F	P	H	Fe	I	Li	Mg	Mn	Hg	Mo	Ni	N	O	Pb	K	Se	Si	Na	Tl	V	W
Esenciales para toda forma de vida		X					X	X		X						X	X				X					X	X		X			X			
Esenciales traza (mg/día)									X			X						X																	
Esenciales ultratraza (µg/día)				X							X		X		X				X			X		X	X					X	X			X	X
Conocidos problemas de nutrición y deficiencias en humanos									X			X	X		X			X	X											X					
Deficiencias solamente conocidas en animales											X											X		X	X						X				
Con función esencial sospechada pero con un mecanismo de acción desconocida	X			X	X	X								X						X								X						X	
Tóxicos	X		X			X							X							X			X					X		X			X		

La función biológica y toxicidad de un elemento depende de la especie química, y a su vez, la forma en la que se encuentra el elemento depende de sus propiedades químicas. Los ligandos en química bioinorgánica son generalmente cadenas laterales de aminoácidos o constituyentes de ácidos nucleicos. La coordinación depende críticamente de la estructura tridimensional de las proteínas y de la estructura terciaria de los ácidos nucleicos (14). Sin embargo, los metales pueden unirse también a grupos prostéticos de metaloproteínas (p.e. hierro-porfirina IX, magnesio-clorofila), bleomicina, sideróforos, coenzimas (p.e. cobalamina-Co), y metilcobalamina, que puede transferir un metilcarbanión a sales de Hg, Pb, y Sn en disolución acuosa, a través de una reacción de biometilación que probablemente contribuye a la toxicidad de estos elementos. Finalmente, los metales pueden unirse a entidades complejas como las membranas celulares, virus y compartimentos intracelulares (p.e. ribosoma, mitocondria y retículo endoplasmático) (14). En este sentido, algunos elementos como el Cu, Zn, Cd, Hg y Ag, se coordinan a proteínas a través de un átomo de azufre, debido a su afinidad por este elemento (p.ej. metionina,

cisteína, fitoquelatinas, metalotioneínas y albúmina), y otros, a través de un átomo de nitrógeno u oxígeno como Mo, Mn, Fe, Co, Ni, Cu y Zn (p.ej. histidina, tirosina, ácido glutámico, asparagina y la malloría de metaloenzimas). Los metaloides como el As, Se y I, se unen a metabolitos mediante unión covalente (ácido metilarsónico, arsenobetaina, arsenoazúcares, selenoaminoácidos, selenopéptidos y selenoproteínas). Otros elementos como Al, Ni y Fe se coordinan con pequeños ligandos orgánicos como el ácido cítrico o la nicotina. El Mg, V, Fe, Co y Ni se coordinan con ligandos de tetrapirrol (p.ej. porfirinas, cobalaminas y el F430); Ca, Sr, Ba, La y Pb forman complejos con polisacáridos como el ramnogalacturonan-II, y finalmente, Pt, Ru, Cr y Ni se coordinan a ácidos nucleicos y sus constituyentes (23). En las selenoproteinas (p.e. glutatión peroxidasa, selenoproteína P), el selenio presenta una unión fuerte ya que la selenocisteína es codificada por el codón UGA y por tanto es una parte integral constituyente de la proteína (24).

Una de las principales funciones de los metales en los sistemas biológicos está ligada a las metaloenzimas, las cuales requieren la participación de un ión metálico (o iones) en un sitio activo para llevar a cabo su función. Las metaloenzimas constituyen una subclase de las metaloproteínas, las cuales incorporan uno o más átomos como parte normal de sus estructuras. Como se ha comentado previamente, se estima que los heteroelementos están presentes en un tercio de las biomoléculas existentes en una célula (25). La Tabla 2 muestra para la mayoría de elementos las biomoléculas con las que se unen o tienen relación, así como su función biológica (26).

3. Procedimientos experimentales en METALÓMICA.

Los análisis de especiación clásicos se centran en el estudio de especies bien conocidas, generalmente compuestos organometálicos de origen antropogénico, así como de sus productos de degradación medioambiental. Algunos ejemplos son el tributilestaño, metilmercurio, arsénico inorgánico, cromo (VI), cromo (III), etc. En la mayoría de los casos, estos analitos son estables debido a que existe una unión covalente entre el metal o metaloide y un átomo de carbono, lo cual hace que el procedimiento de preparación de la muestra sea sencillo. Además, en la mayoría de los casos se dispone de patrones comerciales o bien se pueden sintetizar con facilidad.

En metalómica, los analitos de interés son especies metálicas endógenas de los sistemas biológicos. Durante millones de años, la evolución ha dado lugar a una gran variedad de ligandos biológicos con gran capacidad para coordinarse con elementos traza. La función de numerosas proteínas no podría nunca explicarse sin la intervención de algún elemento traza que actúa como cofactor. Existe por tanto, una gran variedad de analitos entre los que se incluyen ligandos de baja masa molecular (citrato, oxalato, tartrato, aminoácidos u oligopéptidos), moléculas macrocíclicas quelantes, y macromoléculas como las proteínas, polisacáridos o fragmentos de ADN. A todo ello hay que sumarle la complejidad de los sistemas biológicos que añade dificultad a su análisis.

En contraste con los procedimientos proteómicos, la preparación de la muestra en metalómica es crítica, siendo los principales problemas la contaminación de la misma con elementos traza, la aparición de artefactos como consecuencia de la manipulación de la muestra, y la labilidad del enlace del metal o metaloide con la proteína. No obstante, en metalómica, la presencia del metal ligado a la biomolécula representa una ventaja indudable en el seguimiento o monitorización de los análisis, ya que éste puede servir de “etiqueta” de la biomolécula cuando en la plataforma instrumental utilizada se incluye un detector atómico específico de elevada sensibilidad. Generalmente el plasma de acoplamiento inductivo con detector de masas (ICP-MS).

3.1. Precauciones en la preparación de la muestra.

Siendo la metalómica heredera de la especiación y muy próxima a la proteómica, lo más fácil durante estos años ha sido la aplicación o adaptación de los procedimientos de tratamiento de muestra existentes. Sin embargo, este hecho ha generado diversos problemas, que obligan a adoptar precauciones específicas para este tipo de análisis (27):

- 1. Evitar el uso de piezas metálicas para la escisión de órganos o durante el procedimiento analítico ya que podrían contaminar las muestras. Este tipo de material debe sustituirse por utensilios de teflón o porcelana.

- 2. En determinados casos la extracción mediante sonda de ultrasonidos puede afectar la unión del metal a la proteína.

- 3. Algunas metaloproteínas no son estables en el tiempo, aunque las muestras biológicas se almacenen congeladas, pudiendo producirse pérdidas de determinadas especies.

- 4. Todos los reactivos utilizados durante el procedimiento analítico deben ser de la máxima calidad y el contenido de elementos traza el mínimo posible con el objetivo de evitar la aparición de artefactos.

- 5. El pH y la fuerza iónica deben controlarse de tal forma que el procedimiento analítico no suponga la pérdida del metal o metaloide. El pH debe mantenerse próximo al fisiológico para evitar la desnaturalización de las proteínas y la mencionada pérdida del metal. Esta consideración no es tan crítica en el caso de las selenoproteínas (ej. glutatión peroxidasa y selenoproteina P) en las que el enlace con el selenio es covalente y la selenocisteína está genéticamente codificada en estas proteínas.

- 6. La homogeneización de tejidos debe llevarse a cabo con una disolución reguladora adecuada que contenga un agente reductor apropiado que evite la pérdida del metal bajo condiciones oxidantes.

- 7. Debe evitarse el uso de AEDT (ácido etilendiaminotetraacético), utilizado como anticoagulante o inhibidor de las metaloproteasas. En lugar de AEDT puede utilizarse heparina como agente anticoagulante. Otros agentes complejantes que deben evitarse son el tris(hidroximetil)aminometano (Tris) y el ácido N-2-hidroxietilpiperacin-N'-2-etanosulfónico (Hepes) ya que forman complejos con el Zn^{2+} y el Cu^{2+}, respectivamente. El DTT (ditiotreitol), también muy utilizado en procedimientos proteómicos como agente reductor, para evitar la formación de puentes disulfuro entre las proteínas, debe sustituirse por glutatión reducido (GSH) a una concentración próxima a la que existe en el citosol. Más recientemente se ha propuesto el uso de cloruro de tris(2-carboxietil)fosfina (TCEP), ya que carece de propiedades complejantes (28, 29).

Por otro lado, es necesario detener los procesos bioquímicos y enzimáticos que pueden producirse en los tejidos biológicos durante su almacenamiento, especialmente en tejidos animales. Por lo que es importante conservar la muestra a bajas temperaturas (-80 ºC) o aplicar procedimientos de liofilización.

3.1.1. Selección de tejidos biológicos y fluidos para el análisis.

En principio, cualquier tejido biológico, célula (por ejemplo eritrocitos) o fluido biológico (plasma sanguíneo, fluido cerebroespinal, etc), puede extraerse de un mamífero sometido a experiencias de exposición o de vida libre, para llevar a cabo estudios metalómicos. La mayoría de estudios con órganos, suelen llevarse a cabo en hígado o riñón, por ser metabólicamente muy activos,

y más fácilmente accesibles, aunque en muchos casos la facilidad depende del tamaño de la especie. A veces resulta interesante utilizar el cerebro como órgano diana, aunque está normalmente protegido de la acción de los contaminantes por la barrera hematoencefálica (BBB) (30), y su extracción es más difícil ya que hay que abrir la cavidad craneal. No obstante, aunque el cerebro está protegido del influjo de xenobióticos por la barrera hematoencefálica, se incluye en numerosos estudios, ya que en determinados casos la presencia de un xenobiótico en el organismo, aunque no atraviese la BBB, puede dificultar el acceso al cerebro de un elemento traza.

3.1.2. Preparación de extractos celulares a partir de órganos o plasma para análisis metalómico.

Cuando se estudian organismos de pequeño tamaño como por ejemplo el ratón, suele ser necesario utilizar muestras compuestas por órganos de diferentes individuos para disponer de una cantidad de tejido suficiente que permita hacer el análisis. El primer paso consiste en la disgregación del tejido y la rotura de las células, que suele llevarse a cabo mediante métodos mecánicos o químicos. Los métodos mecánicos se basan generalmente en el uso de un homogeneizador vidrio-teflón (tejidos animales), vidrio-vidrio (tejidos vegetales), un crio-homogeneizador o un dispositivo de ultrasonidos. También se han utilizado agitadores de alta velocidad (sistema Turrax) o bolas de vidrio. Habitualmente, se utiliza nitrógeno líquido para facilitar la rotura celular. Los métodos químicos consisten en la adición de detergentes en los que las membranas celulares son solubles. En otras ocasiones se aplica una lisis enzimática con enzimas específicas que eliminan las paredes celulares. En determinadas ocasiones el uso de nitrógeno líquido resulta suficiente para la lisis celular y extracción del analito. En este sentido, las muestras de plantas deben simplemente cortarse en trozos y tratarse con nitrógeno líquido. No obstante, el método de lisis depende del tipo de tejido, así para el músculo debe utilizarse un homogeneizador de alta velocidad o bien enzimas hidrolíticas (28, 31). La figura 2 muestra los diferentes métodos de disgregación de tejidos y rotura celular que se pueden aplicar a muestras órganos (hígado, músculo y riñón), glóbulos rojos, levadura, bacterias y plantas. Tras la rotura celular, la extracción se lleva a cabo a pH adecuado utilizando un tampón.

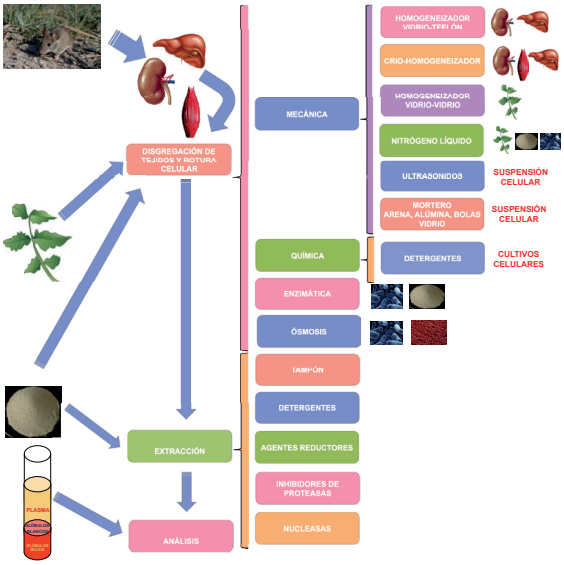


Figura 2.- Métodos de disgregación de tejidos y rotura celular para muestras de órganos (hígado, músculo y riñón), glóbulos rojos, levadura, bacterias y plantas.

El uso de detergentes como el dodecilsulfato sódico (SDS) o el 3-((3-Colamidopropil)- dimetilamonio)-propano sulfonato (CHAPS) es recomendable cuando el elemento forma parte de la estructura de la proteína o se encuentra fuertemente complejado (32). Estos detergentes eliminan las interacciones de los lípidos con las proteínas y disuelven las metaloproteínas previniendo de esta forma interacciones hidrofóbicas (33). Los detergentes no iónicos y los que forman zwitteriones son más apropiados para evitar la desnaturalización de las proteínas (34). El SDS se ha utilizado para la extracción de selenoproteínas de levadura y tejidos de mamíferos (33, 35, 36). En determinados casos se ha observado que el uso del SDS puede interferir en la separación de proteínas en la primera dimensión (34). No obstante, algunos autores afirman que si el SDS se combina con un exceso de detergentes no iónicos o zwitteriónicos (por ejemplo Triton X-100

y CHAPS) se elimina esta interferencia (37). El uso del SDS interfiere también en el análisis posterior mediante espectrometría de masas ya que suprime la señal cuando se utilizan fuentes de electrospray (ESI) o ionización-desorción láser asistida por matriz (MALDI), así como cuando se emplea la cromatografía de líquidos acoplada a plasma de acoplamiento inductivo con detector de masas (LC-ICP-MS) (38). El uso del SDS como agente solubilizante/desnaturalizante no es recomendable para todos los tipos de proteínas ya que las proteínas muy ácidas no se unen al mismo. En este caso se recomienda utilizar un agente catiónico, el cual no es habitual para la electroforesis en gel bidimensional (2D-GE) (37). En la primera dimensión de la 2D-GE suelen utilizarse detergentes no-iónicos como Triton X y NP-40. En este caso, el CHAPS es más efectivo para la solubilización y más compatible con el isoelectroenfoque (33, 39). En la actualidad, también es posible utilizar mezclas comerciales de detergentes e inhibidores de proteasas.

Los agentes reductores se utilizan para evitar la oxidación de las proteínas y la formación de los grupos disulfuro. En estudios metalómicos se han utilizado DTT, ditioeritiol (DTE), β-mercaptoetanol, tributilfosfina (TBP) y tris-carboxietilfosfina (TCEP). No obstante, a pesar de las ventajas que aporta el uso de estos agentes reductores, en algunos casos puede producirse la reoxidación de los puentes disulfuro, precipitación de las proteínas o reducción del propio agente reductor (37, 40). Además, como se ha comentado previamente, muchos de estos agentes reductores presentan propiedades quelantes por lo que su uso puede provocar la pérdida de los metales débilmente unidos a las proteínas o su redistribución en la molécula. En este sentido, el glutatión reducido (GSH) ha resultado ser adecuado debido a que de forma natural se encuentra en el citosol de la célula (27), o bien el tris-carboxietilfosfina (TCEP), que carece de propiedades complejantes (28, 29).

Para la extracción de proteínas, es imprescindible utilizar inhibidores de proteasas como el fluoruro de fenilmetilsulfonilo (PMSF), fluoruro de aminoetil-bencilsulfonilo (AEBSF), AEDT, pepstatina, benzamidina, leupeptina y aprotinina. Sin embargo en el caso de las metaloproteínas es importante evitar el uso de agentes quelantes como ya se ha comentado.

Cuando tras la extracción de metaloproteínas se realiza una separación mediante 2D-GE, deben añadirse nucleasas para romper las cadenas de ADN y ARN, evitando así posibles interferencias en la separación electroforética. Asimismo, durante la ultracentrifugación, un porcentaje de proteínas puede perderse por precipitación con el ADN y ARN si no se añaden nucleasas, siendo la benzonasa (41).

El procedimiento general para la preparación de extractos citosólicos de metaloproteínas se ilustra en la figura 2.

3.2. Acoplamientos instrumentales y estrategias analíticas para el análisis de metaloproteínas.

3.2.1. Acoplamientos instrumentales para la determinación de metaloproteínas.

La figura 3 muestra las técnicas analíticas acopladas más utilizadas en metalómica. El procedimiento general para el aislamiento e identificación de metaloproteínas contempla la combinación de tres tipos de técnicas: 1. técnicas analíticas de separación, 2. técnicas espectroscópicas y 3. espectrometría de masas.

Las técnicas analíticas de separación constituyen habitualmente la primera etapa del análisis metalómico y su objetivo es conocer la masa molecular de la metaloproteína (cromatografía de exclusión de tamaño-SEC) o purificarla (SEC, IEC-cromatografía de intercambio iónico, RP-fase reversa, IP-pares iónicos, AC-cromatografía de afinidad, HILIC-cromatografía de interacción hidrofílica). Generalmente suelen utilizarse columnas a escala preparativa que permiten el aislamiento de fracciones tras la monitorización del tiempo de retención con un detector no destruc-

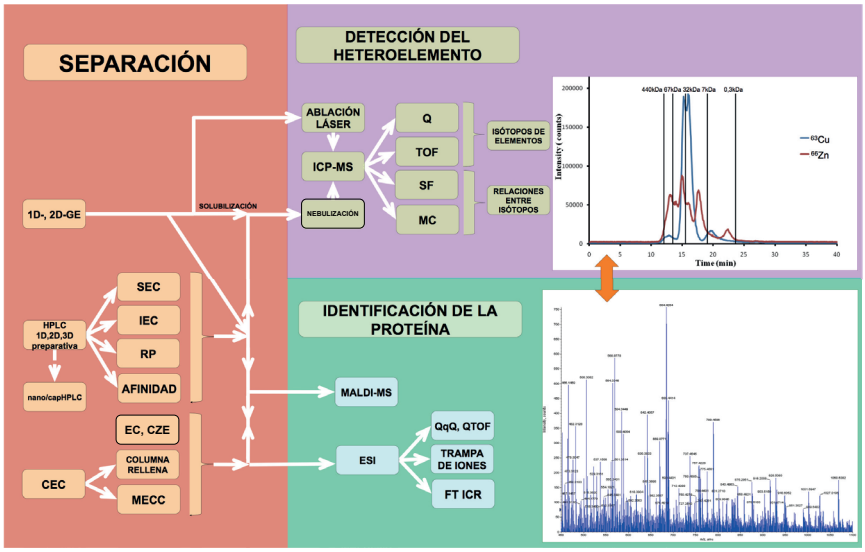


Figura 3.- Técnicas analíticas acopladas de uso frecuente en metalómica.

tivo, como puede ser el UV. El uso de la cromatografía de líquidos capilar y nanocromatografía (nHPLC) es imprescindible para asegurar un bajo consumo de muestra. La cromatografía de líquidos bi- o tridimensional permite reducir la complejidad de la muestra aumentando capacidad de pico. De esta forma pueden utilizarse los sistemas bidimensionales: IEC-RP, HILIC-RP y RP-RP, éste último mediante el empleo de diferentes valores de pH en cada dimensión (42).

La separación también puede llevarse a cabo mediante electroforesis en gel de poliacrilamida (PAGE), capilar (EC) o electroforesis capilar de zona (CZE), y electrocromatografía capilar (CEC) (en columna rellena o cromatografía capilar electrocinética micelar (MECC)). En determinadas ocasiones, la labilidad de la unión metal(-oide)-proteína no permite su separación mediante RP o IEC. En estos casos, la electroforesis capilar constituye una alternativa muy valiosa (23, 43, 44). La electroforesis en gel de poliagrilamida (PAGE) de una dimensión o bidimensional (2D) es una técnica de alta resolución para la separación de proteínas. Las proteínas que contienen algún heterolemento unido covalentemente pueden separarse fácilmente, pero si los metales forman complejos con las mismas es necesario trabajar en condiciones no desnaturalizantes y por tanto evitar la presencia de SDS.

El acoplamiento en línea de estas técnicas analíticas de separación con un detector atómico multielemental, de elevada sensibilidad y selectividad, permite la monitorización de los metales (metaloides) en las proteínas separadas. Para tal fin, el sistema más utilizado es el plasma de acoplamiento inductivo acoplado a detector de masas (ICP-MS), cuya interfase consiste en un capilar que se introduce directamente en el sistema de nebulización del ICP (técnicas cromatográficas y electroforéticas capilares) o ablación láser (electroforesis en gel). Indudablemente, el ICP-MS es una herramienta muy valiosa en este campo, debido fundamentalmente a: (i) capacidad de realizar un análisis multiisotópico, incluyendo no-metales como S, Se y P; (ii) alta sensibilidad; (iii) elevada tolerancia a la matriz de la muestra y (iv) amplio rango lineal. La ablación láser (LA) acoplada a ICP-MS permite la monitorización directa de la presencia de metales(-oides) en el gel de electroforesis (1D ó 2D). El resultado es un electroferograma en el cual la cantidad de un elemento es función de su posición en el gel. Mediante la ablación láser la escisión de las manchas del gel y su digestión no son necesarias, lo cual representa una gran ventaja ya que constituyen habitualmente importantes fuentes de contaminación. El análisis cuantitativo mediante LA-ICP-MS es rápido y la señal es directamente proporcional a la cantidad de elemento en el gel, por lo que se elimina el problema de la recuperación de las proteínas presentes en el mismo. Sin embargo, sólo una pequeña parte de la banda o mancha de proteínas (10-50 µm)

llegan al ICP-MS tras la ablación, lo cual dificulta la aplicación de esta técnica a proteínas poco abundantes (45).

La señal que proporciona la detección mediante ICP-MS no depende de las especies de carácter orgánico que acompañan al elemento en la elución cromatográfica. Ello hace posible la cuantificación de la metaloproteína, siempre que su estructura primaria y estequiometría sean conocidas. Sin embargo, algunos elementos importantes desde el punto de vista biológico, como el azufre o el fósforo, no se detectan con suficiente sensibilidad debido a su elevado potencial de ionización. Tampoco son ionizados con eficiencia los halógenos y el flúor (éste último tiene interés en el estudio del metabolismo de las drogas). Otro problema adicional es la presencia de interferencias poliatómicas que desvirtúan la señal para determinadas relaciones *m/z*. Sin embargo, en la actualidad es posible eliminar tales interferencias mediante el uso de un sistema de alta resolución (sector magnético-ICP-SF-MS) o con una celda de colisión/reacción. El detector de sector magnético, SF, proporciona una señal de fondo muy baja y buenos límites de detección para casi todos los elementos. Sin embargo, para medidas a alta resolución, la sensibilidad disminuye drásticamente. La celda de colisión/reacción permite la eliminación de interferencias mediante la colisión de iones poliatómicos con He o Xe o bien, mediante la reacción de transferencia de carga con H_2 o NH_3. También puede utilizarse oxígeno, en cuyo caso la especie detectada será un nuevo ión producto que se mide a una *m/z* diferente a la del ión original y que no resulta afectada por las interferencias.

La medida de las diferentes relaciones *m/z* a lo largo del cromatograma requiere el uso de sistemas con una alta velocidad de adquisición ya que de no ser así, la relación entre los metales o metaloides puede variar durante un tiempo más corto que el requerido por el sistema de detección para el barrido de las *m/z*. El sistema cuadrupolar (Q) es un sistema más lento que el de tiempo de vuelo (TOF) el cual produce un espectro de masas atómico en un tiempo inferior a 50 µs (46). Sin embargo, el ICP-Q-MS es del orden de diez veces más sensible que el ICP-TOF-MS por lo que éste último ha sido menos utilizado. El uso de un sistema multicolector (MC) para la detección cromatográfica soluciona todos estos problemas aunque su coste es muy elevado (47). El ICP-MS permite además de la medida de isótopos de los elementos de interés (cuadrupolo-Q, tiempo de vuelo-TOF), la determinación de relaciones isotópicas utilizando sistemas de alta resolución (sector magnético-SF, multicolector-MC).

Otro componente instrumental importante en el análisis de metaloproteínas es la espectrometría de masas orgánicas. Por ello se usa de forma complementaria el ICP-MS (información atómica, metales y heteroelementos), ESI-MS (información molecular) y MALDI-MS (información molecular). La sensibilidad del ICP-MS es excelente cuando se utiliza directamente en extractos biológicos, pero la pérdida de información relativa a la unión de estos elementos a biomoléculas hace imprescindible el uso del acoplamiento con técnicas analíticas de separación de alta resolución. La información relativa al complejo metal-biomolécula se pierde habitualmente en MALDI-MS debido al uso de matrices ácidas y al propio mecanismo de ionización. Sin embargo, el MALDI-MS es muy valioso para el análisis de selenopéptidos (unión covalente), y en la actualidad existen matrices para MALDI que permiten conservar el contenido metálico de las proteínas. El sistema ESI-MS proporciona teóricamente información valiosa sobre los complejos metálicos, pero la experiencia demuestra que esta técnica no siempre ofrece esta información y requiere un alto grado de purificación del analito (45). El sistema ESI-MS se ha aplicado en estudios de metalotioneínas obteniéndose espectros de MS de la proteína sin metales (condiciones ácidas) y de la proteína intacta que permiten conocer la estequiometría metal:biomolécula (48). Por otra parte, la cuantificación es difícil con las técnicas de espectrometría de masas moleculares ya que los resultados dependen de la ionización de la molécula. El efecto matriz es muy considerable en ESI-MS, importante en MALDI-MS y poco habitual en ICP-MS. El sistema ESI-MS es mejor que el MALDI-MS para la espectrometría de masas en tándem y acoplamientos en línea a técnicas analíticas de separación (HPLC, CE). La principal diferencia entre estas técnicas de espectrometría de masas molecular es que el sistema ESI-MS es sensible a la concentración

mientras que el MALDI-MS es sensible a la masa y sólo las uniones covalentes se conservan. Los analizadores más utilizados en espectrometría de masas con ionización mediante ESI son el triple cuadrupolo (QqQ), el cuadrupolo-tiempo de vuelo (QTOF), la trampa de iones y el de resonancia ciclotrón con transformada de Fourier (FT-ICR-MS).

3.2.2. Estrategias analíticas para la determinación de metaloproteínas.

Los estudios proteómicos se han catalogado como "de arriba abajo" (top-down) y "de abajo arriba" (bottom-up) y proteómica de escopetazo (shotgun), de acuerdo con la estrategia empleada (5, 49):

- **1.** En los planteamientos de "arriba abajo" se separan las proteínas mediante LC y se fragmentan mediante disociación multifotónica en el infrarrojo (IRMPD) o disociación por captura de electrones (ECD). La composición de los aminoácidos de los fragmentos se analiza por espectrometría de masas. La identificación se basa en las correlaciones entre la composición de aminoácidos de los iones formados y las proteínas que tienen posibilidad de estar presentes.

- **2.** Los estudios de "abajo arriba" se basan en la electroforesis en gel 2D seguida de una digestión con tripsina de las manchas y el análisis mediante MS. La identificación de algunos péptidos es habitualmente suficiente para identificar la proteína.

- **3.** Por otro lado, la proteómica de escopetazo (shotgun), separación multidimensional MS/MS o tecnología multidimensional de identificación de proteínas (multidimensional protein identification technology o MudPIT), comienza con el tratamiento de la muestra con tripsina para hidrolizar todas las proteínas. Los péptidos resultantes se separan por cromatografía multidimensional (2D-nHPLC) y se secuencian por espectrometría de masas tándem (MS/MS). Los péptidos secuenciados se emplean para realizar búsquedas en bases de datos e identificar las proteínas de la muestra original.

Para el estudio de metaloproteínas es necesario además, describir los sitios de unión del metal(-oide), la estequiometría metal-proteína y la estructura dependiente del metal o los cambios conformacionales que se produzcan. Este tipo de estudios es incompatible con procesos de desnaturalización de las proteínas que provocarían la pérdida de los elementos que no estén unidos covalentemente. Una estrategia analítica válida para el estudio de metaloproteínas requiere (5): (i) la conservación de la unión metal(-oide)-proteína, (ii) la detección específica de metaloproteínas y (iii) la purificación y/o preconcentración de metaloproteínas.

Estos requerimientos dan lugar a dos estrategias analíticas generales para la determinación de proteínas que contienen metales como cofactores no unidos covalentemente:

- **1.** Extracción del proteoma, sin tener en cuenta las uniones metal-proteína, seguida del aislamiento de algunas metaloproteínas con afinidad por un determinado metal(-oide) particular para su identificación posterior mediante cualquiera de los procedimientos proteómicos citados anteriormente. Este procedimiento permite la detección de metaloproteínas con afinidad por el metal en la muestra, pero no sobre si los complejos metálicos están actualmente presentes ni con qué proteínas.

- **2.** Extracción de los complejos metal(-oide)-proteína intactos, su separación e identificación. El énfasis se pone ahora en preservar el complejo. La unión sólo puede eliminarse cuando se alcanza un alto grado de purificación que permite tener certeza de que tal unión existía antes de la desnaturalización.

4. Aplicación de las técnicas metalómicas en estudios medioambientales.

4.1. Experiencias de exposición en seres vivos.

Una forma de evaluar la repercusión de contaminantes y otros factores exógenos en los bioindicadores que viven en un determinado hábitat es mediante experiencias de exposición, pudiendo evaluarse la respuesta bioquímica de mamíferos modelo a la presencia de niveles bajos (crónica) o elevados (aguda) de xenobióticos. De esta forma es posible extrapolar los resultados al hombre, si se elige adecuadamente el modelo de mamífero para el estudio. En relación a ello, deben considerarse diversas cuestiones que permitan un adecuado diseño de la experiencia.

4.1.1. Selección de animales modelo.

La selección de un modelo animal adecuado que permita extrapolar los resultados al hombre es fundamental. La FDA (US Food and Drug Asministration) ha constatado que la mayoría de errores cometidos en la investigación de las drogas se deben al uso de modelos inadecuados (50). En principio, las ratas, ratones y conejos son los modelos más utilizados, aunque los resultados que se obtienen con ellos no son siempre coincidentes. Las ratas tienen muchas características fisiológicas comunes con el hombre, por ejemplo la frecuencia cardíaca (51), o similitud genética, ya que aproximadamente el 90 % de los genes de la rata coinciden con los humanos (52). Por ello, la rata constituye un buen modelo para el estudio de enfermedades en el hombre como hipertensión, envejecimiento, enfermedades infecciosas, afecciones neurológicas (50, 51). Sin embargo, las ratas metabolizan el arsénico de forma diferente a los humanos ya que acumulan el As (III) en los eritrocitos uniéndose a la hemoglobina (53, 54). A pesar de ello, pueden utilizarse modelos de ratas genéticamente modificadas que contengan genes humanos (50).

Otras cuestiones importantes a tener en cuenta en la selección del organismo modelo son el sexo, la edad y el tipo de muestra. En relación al sexo, es preferible trabajar con machos para evitar influencias hormonales que puedan repercutir en los resultados obtenidos. Además los testículos son órganos que permiten importantes informaciones toxicológicas (55). La edad es también uno de los factores importantes a tener en cuenta, así, algunos mecanismos de detoxificación, como la excreción biliar de GS-$HgCH_3$, se produce en adultos pero no en jóvenes, porque aún no han desarrollado este mecanismo metabólico (56). Asimismo, hay que considerar el tipo y cantidad de muestra que se precisa para el análisis. Si la muestra objeto de estudio es sangre, puede ser más apropiado como organismo modelo la rata que el ratón, ya que a partir del ratón de laboratorio convencional (*Mus musculus*) sólo pueden extraerse 800-900 µl como máximo mediante punción cardíaca, lo que representa una cantidad de muestra muy pequeña para la mayoría de los análisis. Si se trata de muestras de bilis, es más adecuado utilizar el conejo, no sólo por la cantidad de fluido que puede obtenerse, sino por la mayor facilidad de llevar a cabo la extracción desde un punto de vista quirúrgico (57).

Los animales pueden exponerse a uno o más metales/metaloides tóxicos mediante administración oral, intravenosa, interperitoneal o subcutánea (30). Es imprescindible controlar todas las condiciones de hábitat durante el experimento de forma que no aumente el estrés del animal.

4.1.2. Selección de la dosis apropiada de contaminante.

Es importante seleccionar adecuadamente la dosis de contaminante o tóxico para la experiencia de exposición, según se quiera obtener información sobre toxicidad aguda (lo que implica la administración de grandes dosis) o crónica, que precisa la administración de dosis pequeñas durante periodos de tiempo largos. Debe considerarse que los mecanismos bioquímicos impli-

cados en la toxicidad aguda o crónica de los contaminantes pueden diferir entre sí, así la toxicidad aguda del As(III) origina la inhibición de multitud de enzimas (17), mientras que la toxicidad crónica provoca la inhibición de la biosíntesis de las selenoproteínas (58).

4.1.3. Ejemplos de ensayos de exposición del ratón de laboratorio (*Mus musculus*) a elementos tóxicos.

Es bien conocido que la exposición crónica a elementos tóxicos como cadmio, arsénico, mercurio o cromo incrementa la incidencia del cáncer y los desórdenes neurológicos, sin embargo, los mecanismos que determinan la capacidad de estos elementos para transformar las células permanecen desconocidos (59). Debido a ello, las experiencias de exposición con animales modelo pueden proporcionar una información muy útil sobre la respuesta biológica de dichos organismos a los contaminantes, así como los procesos de homeostasis que experimentan los metales y su tráfico entre los diversos órganos. De esta forma se ha establecido el carácter carcinogénico de Cd y As que inducen daños en el DNA de eritrocitos de la médula ósea de ratón (60), y se ha comprobado que el tratamiento combinado de animales expuestos a Cd con Se y Zn provoca una disminución de las concentraciones de Cd en plasma y testículo y una protección más eficiente frente a daños orgánicos (61). También se han estudiado otros efectos de metales como Cd y Zn sobre la maquinaria del tráfico de estos elementos en células hepáticas, mediante inyección intraperitoneal de estos elementos (62). Pueden citarse otros ejemplos que demuestran el interés de los ensayos de exposición con animales modelo para conocer los mecanismos de toxicidad, tráfico e interacciones, a nivel celular y orgánico, de los metales a los que potencialmente estamos expuestos en el medio ambiente (63, 64).

4.1.3.1. Exposición del ratón de laboratorio (*Mus musculus*) a Cd.

Se ha descrito la exposición del ratón *M. musculus* a Cd (30, 65) mediante inyección subcutánea (100 µL) de dosis crecientes de disolución acuosa de $CdCl_2$ durante 10 días. Durante la exposición se separaron tres subgrupos a los 2, 6 y 10 días (grupos A, B y C, respectivamente). Asimismo, se utilizaron grupos control a los que se inyectaron diariamente 100 µL de suero salino fisiológico.

Para el correspondiente estudio metalómico se extrajeron hígado, riñón y cerebro, que se trataron según el esquema de la figura 4 para la obtención de los extractos citosólicos. Los extractos se analizaron mediante un acoplamiento SEC-ICP-MS, empleando una columna cromatográfica de exclusión de tamaño Superdex™-75, que permite un rango de separación efectivo para mo-

Figura 4.- Esquema de obtención del extracto citosólico de células y tejidos.

léculas de masas comprendidas entre 3 y 70 kDa, que se acopla al nebulizador del dispositivo ICP-MS mediante un tubo de PEEK de 30 cm de 0.17 mm de diámetro interno. La calibración de sistema se realiza mediante metalobiomoléculas con masas en el rango de trabajo del sistema, como ferritina, Mr 900 kDa; albúmina del suero bovino (BSA), Mr 66,3 kDa; superóxido dismutasa (SOD, que contiene Cu y Zn), Mr 32.5 kDa; metalotioneina I (unida a Cd, Cu y Zn), Mr 6 kDa; y arsenobetaína (AsB, con As), Mr 178 Da.

Los resultados demuestran la acción protectora del hígado frente al cadmio, que es atrapado como Cd-metalotioneina (Fig. 5a). Puede observarse el aumento progresivo de la intensidad del pico a 6-7 kDa. Además, en las figuras 5b y c se observan otras moléculas unidas a Cu y Zn, con máximos a 7 kDa que aumentan ligeramente con la exposición aunque con menor intensidad que el pico de Cd. En estas figuras pueden también observarse picos de Cu y Zn con tiempos de retención coincidentes con los patrones de superóxido dismutasa (SOD, 32 kDa). Posiblemente, el estrés oxidativo provocado por la exposición a Cd puede ser el origen de esta sobrexpresión de Cu,Zn-SOD que alcanza mayor intensidad en los animales más expuestos (grupo C).

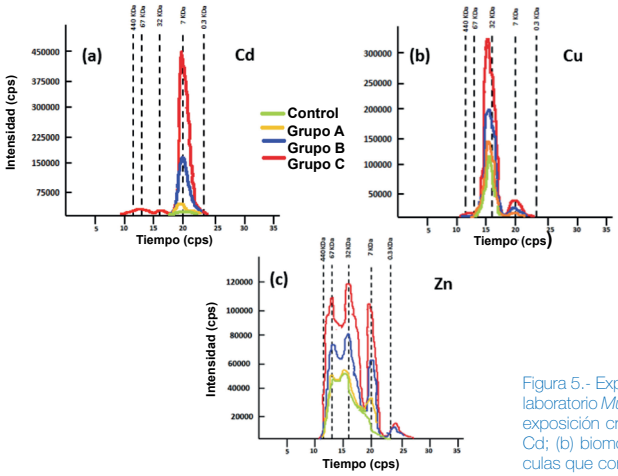


Figura 5.- Experiencia de exposición a Cd del ratón de laboratorio *Mus musculus*. Los grupos A, B y C indican exposición creciente. (a) Biomoléculas que contienen Cd; (b) biomoléculas que contienen Cu; (c) biomoleculas que contienen Zn

4.1.3.2. Exposición del ratón de laboratorio (*Mus musculus*) a Hg.

Cuando la experiencia de exposición se repite con los mismos animales modelo (*Mus musculus*) mediante inyección subcutánea de mercurio (120 µL de disolución acuosa de $HgCl_2$ hasta alcanzar una dosis de 3 mg de Hg por Kg de peso y por día) durante 14 días (65), se observa la aparición de un pico en extracto citosólico de hígado que se relaciona con una biomolécula de Cu con tiempo de retención coincidente con el patrón de Cu,Zn-SOD (32 kDa) que disminuye en intensidad un 60 % con la exposición (Fig. 6a, grupo C), mientras que el pico equivalente de la molécula de mercurio (el posible Hg-SOD) aumenta en intensidad un 30 % (Fig. 6b; grupo C), aunque la intensidad de este pico es muy baja. Estos hechos están de acuerdo con la concentración total de estos elementos en el extracto de hígado, en los cuales los niveles de Cu disminuyen de 1,5 a 0,75 µg g^{-1} (grupos A a C) y los de mercurio aumentan de 0.075 a 0,125 µg g^{-1} (grupos A a C). Es destacable la baja concentración de mercurio en el extracto de hígado a pesar de la elevada exposición de los ratones a este elemento en el grupo C. Para explicar estos hechos debe considerarse la concentración de estos metales en el riñón, lo cuál muestra que la concentración de Cu aumenta de 0.5 a 1,5 µg g^{-1} (de los grupos A a C) y la de Hg de 2 a 14 µg g^{-1}; además los perfiles SEC-ICP-MS en riñón indican quela intensidad de los picos relacionados con la metalotioneina de Cu (Cu-MT, Fig. 6c) y metalotioneina de Hg (Hg-MT, Fig. 6d) aumentan notablemente, especialmente el de Hg-MT, lo que sugiere que el Hg compite con el Cu por los sitios activos de la SOD en el hígado, lo que reduce el tiempo de residencia del

mercurio en este órgano (66), movilizándose a los riñones (67) para su eliminación por la orina. Esta hipótesis puede explicar la presencia elevada de Hg en los riñones, principalmente como Hg-MT, que es una molécula asociada a la detoxificación de mercurio (68).

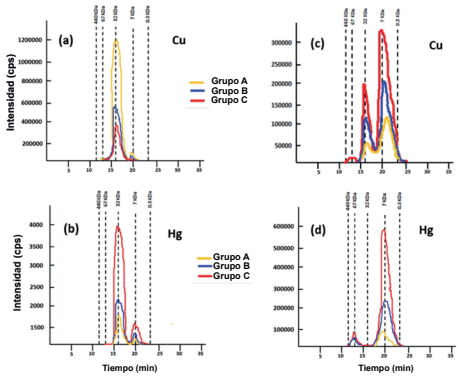


Figura 6.- Experiencia de exposición a Hg del ratón de laboratorio *Mus musculus*. Los grupos A, B y C indican exposición creciente. (a) Biomoléculas que contienen Cu en el hígado; (b) biomoléculas que contienen Hg en el hígado; (c) biomoleculas que contienen Cu en el riñón; (d) biomoléculas que contienen Hg en el riñón.

4.1.3.3. Exposición del ratón de laboratorio (*Mus musculus*) a As.

Cuando los ratones de laboratorio se someten a exposición oral de arsénico en forma de arsenito sódico disuelto en agua, se observa que el elemento se encuentra fundamentalmente acumulado en el riñón (Fig. 7) en la que se aprecian dos picos vinculados a la presencia de As de baja masa molecular, en el rango de 200 a 130 Da, que pueden relacionarse con monometilarsénico y dimetilarsénico (MMA y DMA), cuya intensidad aumenta con la exposición. En hígado el comportamiento del arsénico es el inverso, y estos dos picos decrecen con la dosis. Posiblemente, este comportamiento de las especies de arsénico puede relacionarse con el tráfico progresivo de este elemento hacia los riñones para su correspondiente excreción.

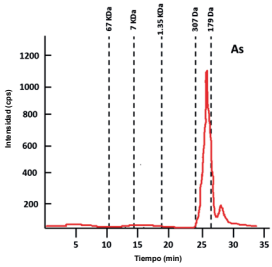


Figura 7.- Experiencia de exposición a As del ratón de laboratorio *Mus musculus*. Extracto citosólico de riñón.

4.2. Aplicación de procedimientos metalómicos en estudios medioambientales.

Estos procedimientos metalómicos pueden aplicarse para evaluar la respuesta biológica de los organismos de vida libre en sistemas medioambientales (30, 41). El estudio en este caso es mucho más complejo, pero presenta la ventaja de que la exposición puede producirse por inhalación, ingestión o cualquier otra vía de forma conjunta, mientras que la exposición programada en el laboratorio suele producirse sobre una base experimental mucho más simple.

El ejemplo que se comentará a continuación se centra en el Parque Nacional de Doñana, una de las más importantes reservas biológicas en Europa, donde millones de aves migradoras anidan cada año procedentes de África. No obstante, el Parque esta sometido a la acción de contaminantes debido a las actividades agrícolas, mineras e industriales de su entorno. Por ello, la aplicación de los procedimientos metalómicos previamente descritos a animales de vida libre de este ecosistema puede proporcionar una información muy interesante sobre la respuesta biológica a la contaminación. En este sentido se ha seleccionado un bioindicador usado frecuentemente en esta zona, el ratón *Mus spretus*, que es una especie no protegida, que campea en el entorno de su madriguera y por ello refleja bien las características del entorno en el que se le captura.

El estudio que se describe considera áreas contaminadas y no contaminadas de Doñana, Fig. 8, pudiendo citarse el “Lucio del Palacio” (LDP), zona situada en el centro del Parque que

puede considerarse como área control, con baja concentración de contaminantes; "El Arroyo de la Rocina" (ROC), con abundantes cultivos de fresas, cítricos y vid, que recibe la contaminación difusa de plaguicidas y agroquímicos, así como metales procedentes de las aguas ácidas procedentes de las actividades mineras de Riotinto; "El Partido" (PAR) y "El Arroyo del Ajolí", con influencias agrícolas; y, finalmente, "El Matochal" (MAT), próximo al río Guadiamar, también con influencia agrícola y minera, especialmente tras el accidente minero de Aznalcollar.

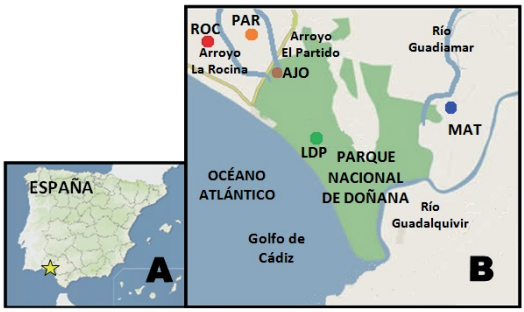


Figura 8.- Puntos de muestreo en el Parque Nacional de Doñana y alrededores (suroeste de España). A, Localización; B, puntos de muestreo: ROC, Arroyo de la Rocina; MAT, El Matochal; PAR, Arroyo del Partido; AJO, Arroyo del Ajolí; LDP, Lucio del Palacio (parte central del Parque).

Los resultados se recogen en la Fig. 9, donde se muestran los perfiles metalómicos, correspondientes a extractos citosólicos de órganos de elevada actividad metabólica como el hígado y el riñón. La sobrexpresión o inhibición de metalobiomoléculas causada por contaminantes se siguió mediante SEC-ICP-MS, los resultados más interesantes se observan con elementos como Cu, Zn y Cd (Fig. 9). Pueden apreciarse similitudes en la expresión de metalotioneinas, pudiendo destacarse la sobrexpresión de los picos marcados por Cu, Zn y Cd de 7 kDa en el extracto de hígado (Fig. 9), los cuales se relacionan bien por su tiempo de retención con el estándar de metalotioneina I (Fig. 9-A, 9-C y 9-E). La intensidad del pico de Cu de la zona contaminada MAT (Fig 9-A) es claramente superior a la de otros puntos de muestreo, también afectados por la contaminación, como ROC y AJO. El pico equivalente marcado por Cd (Fig. 9-E) presenta una menor intensidad, siendo más destacadas las señales correspondientes a MAT y LDP.

Otro pico trazado por Cu que se relaciona con moléculas de 32 kDa presenta mayor sensibilidad en muestras de LDP, MAT, ROC and PAR, aunque la respuesta es inferior en la zona de menor contaminación (AJO) (Fig. 9-A). En la Fig. 3-C pueden verse dos señales trazadas por la presencia de Zn que coinciden con los estándares de 67 y 32 kDa, respectivamente. En este último caso, la mayor respuesta corresponde de nuevo a LDP (Fig. 9-C), con menor intensidad

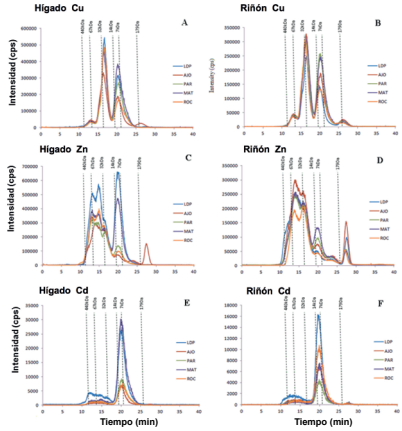


Figura 9.- Cambios de expression de metal-biomoléculas del ratón de campo Mus *spretus* en distintas áreas del Parque Nacional de Doñana. Técnica metalómica utilizada en la obtención de los perfiles SEC-ICP-MS. A,C, E (extractos citosólicos de hígado); B,D,F (extractos citosólicos de riñón).

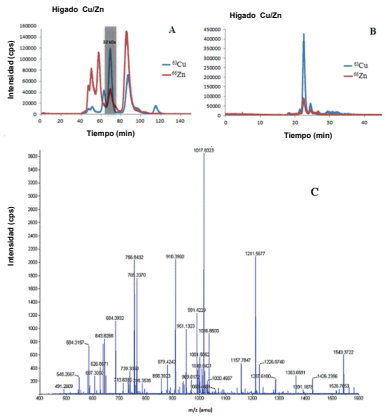


Figura 10.- Identificación de la biomolécula de Cu/Zn de 32 kDa. (A) Extracto citosólico de hígado (*Mus spretus*) del Lucio del Palacio (LDP) caracterizado por SEC-ICP-MS; (B) Cromatograma obtenido por cromatografía de intercambio aniónico de la fracción (A); (C) Espectro de masas de la fracción obtenida en (B) y sometida a digestión con tripsina.

de respuesta en MAT y ROC, a pesar de ser áreas teóricamente más contaminadas. Para confirmar que el pico de Cu y Zn de masa molecular 32 kDa corresponde a la superóxido dismutasa (SOD), se aisló el extracto correspondiente a partir del hígado, usando la muestra más concentrada (LDP) y una columna de exclusión de tamaño preparativa acoplada al ICP-MS (Fig. 10-A). A esta fracción se le eliminan las sales, se liofiliza y se analiza por cromatografía analítica de intercambio aniónico (Fig. 10-B), la cuál demuestra la pureza del pico asociado a la presencia de Cu y Zn que se obtiene con la columna AEC mencionada. Utilizando una columna AEC preparativa análoga se obtiene una fracción, que liberada de sales y sometida a digestión con tripsina, permite la identificación de la proteína por nESI-QqQ-TOF. El espectro de masas resultante, que se muestra en la Fig. 10C permite la secuenciación de la proteína usando péptidos de doble carga con valores de m/z 584.31, 684.39 y 756.84, los cuales sometidos a análisis por MS/MS permite la identificación de la proteína en bases de datos como la MASCOT. De esta forma se confirma la presencia de la molécula de Cu,Zn-superóxido dismutasa asociada al pico de 32 kDa en la Fig. 9-C.

Por otro lado, es interesante señalar la presencia en los cromatogramas de los extractos citosólicos de riñón, Fig. 9, con picos de Cu, Zn y Cd (Fig. 9-B, 9-D y 9-F) relacionados con moléculas de 7 kDa, de diferente intensidad. En este sentido, la mayor intensidad de respuesta se observa en los picos de Cu y Zn de MAT, PAR y AJO, lo que se correlaciona bien con la mayor concentración de estos elementos en el extracto citosólico. Sin embargo, la mayor intensidad del pico de Cd se observa en LDP (Fig. 9-F), coincidiendo también con una mayor presencia de este elemento en el correspondiente extracto. Otro elemento que ha proporcionado resultados interesantes al aplicar el estudio metalómico en la zona considerada es el arsénico. Es bien sabido que las especies organoarsenicales, como la arsenobetaina, la arsenocolina, los arsenorobósidos y arsenolípidos están presentes en los organismos acuáticos y animales de origen marino (69, 70); sin embargo, los mamíferos metabolizan el arsénico inorgánico en el hígado (71). Después de su absorción en el tracto gastrointestinal, se producen especies metiladas de arsénico que son excretadas por la orina como formas metiladas de arsénico pentavalente (72-75). Además, se ha demostrado, que la unión del arsénico a ciertas proteínas constituye un mecanismo de detoxificación (76). Además, si se considera que el arsénico trivalente es químicamente más reactivo que las especies pentavalentes y se une a numerosas proteínas, como tubulina y actina (77), hemoglobina (53, 78), metalotioneina (79) y otras.

Puede verse en la figura 11 la presencia de especies de As de baja masa molecular en el extracto citosólico de riñón analizado mediante SEC-ICP-MS (usando una columna Superdex™-75). Las intensidades de los picos cromatográficos asociados a la presencia de As es más elevada en ROC, donde también se observan concentraciones superiores de As total en los extractos. Generalmente, la intensidad de las señales trazadas por arsénico son superiores en riñón que en hígado, por esta razón los estudios posteriores se han realizado en extractos de riñón de ratones capturados en ROC. Para aumentar la resolución de los cromatogramas de estas especies de As de baja masa molecular se utiliza una columna de exclusión de tamaños con rango efectivo de 0,1 a 7 kDa (Fig. 12-A). Se pueden apreciar dos picos de metabolitos de baja masa molecular. Para identificar estas especies de As, los extractos se fraccionaron con una columna SEC preparativa (Hiload 26/60, superdex 30 Prep), recogiéndose el pico asociado a una masa molecular de 100 a 300 Da (Fig. 12-B). Sin embargo, para la confirmación de los ana-

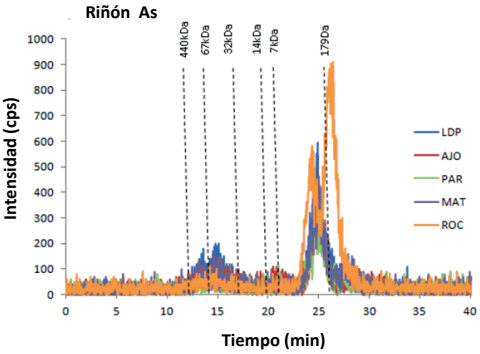


Figura 11.- Metabolitos de arsénico en extractos citosólicos de riñón de *Mus spretus* capturado en ROC. Croamtograma obtenido mediante SEC-ICP-MS con columna preparativa.

litos se utiliza sobre los extractos una separación cromatográfica de intercambio iónico para su purificación (Figs. 12-C y D). El cromatograma que se obtiene por cromatografía de intercambio aniónico muestran la presencia predominante de formas metiladas de arsénico (MMA y DMA) en el extracto de riñón (Fig. 12-D). Cuando se emplea cromatografía de intercambio catiónico (Fig. 12-C), se observa un pico a 4 min con tiempo de retención idéntico al patrón de arsenobdetaína. No obstante, debido a que la arsenobetaína no está presente en los organismos terrestres la especie detectada precisa de identificación por espectrometría de masas en tándem.

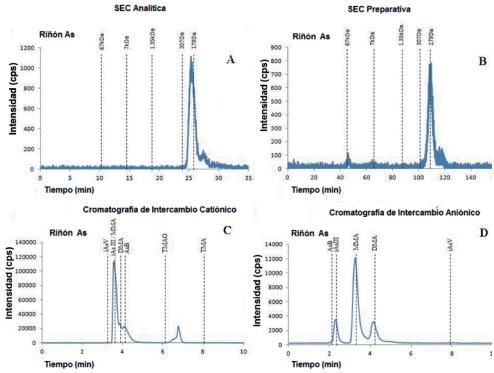


Figura 12.- Metabolitos de As caracterizados en extractos de riñón de *Mus spretus* capturado en ROC. (A) Perfil obtenido mediante SEC-ICP-MS con columna SuperdexTM-Peptide para el rango 0,3 a 7 kDa; (B) Perfil cormatográfico obtenido con columna preparativa naload 26/60 Superdex 30 Prep; (C) Metabolitos de As caracterizados mediante cromatografía de intercambio catiónico (columna Supelcosil SCX); (D) (C) Metabolitos de As caracterizados con cromatografía de inercambio aniónico (columna Hamilton PRP X-100).

REFERENCIAS.

1. **Thiele DJ, Gitlin JD. Assembling the pieces.** Nature Chem Biol. 2008; 4: 145-7

2. **Raymond J, Segre D. The effect of oxygen on biochemical networks and the evolution of complex life.** Science. 2006; 311: 1764-7

3. **Williams RJP. Chemical selection of elements by cells.** Coord Chem Rev. 2001; 216: 583-95

4. **Yi L. Metal ions as matchmakers for proteins.** PNAS. 2010; 107: 1811-2

5. **Mounicou S, Szpunar J, Lobinski R. Metallomics: the concept and methodology.** Chem Soc Rev. 2009; 38: 1119-38

6. **Martínez-Fábregas J, Rubio S, Díaz-Quintana A, Díaz-Moreno I, de la Rosa MA. Proteomic tools for the analysis of transient interactions between metalloproteins.** FEBS J. 2011; 278: 1401-10

7. **Waldron KJ, Robinson N J. How do bacterial cells ensure that metalloproteins get the correct metal?.** Nat Rev Microbiol. 2009; 7: 25-35

8. **Gierasch L, Gershenson A. Post-reductionist protein science, or putting Humpty Dumpty back together again.** Nat Chem Biol. 2009; 5: 774-7

9. **Lydia AF, O'Halloran, TV. Transition Metal Speciation in the Cell: Insights from the Chemistry of Metal Ion Receptors.** Science. 2003; 300: 931-6

10. **Caryn EO, O´Halloran, TV. Femtomolar Sensitivity of Metalloregulatory Proteins Controlling Zinc Homeostasis.** Science. 2001; 292: 2488-92

11. **Maret W. Metalloproteomics, metalloproteomes, and the annotation of metalloproteins.** Metallomics. 2010; 2: 117-25

12. **O'Halloran TV, Culotta VC. Metallochaperones, an Intracellular Shuttle Service for Metal Ions.** J Biol Chem. 2000; 275: 25057-60

13. **Lahner B, Gong J, Mahmoudian M, Smith EL, Abid JB, Rogers EE, Guerinot ML, Harper JF, Ward JM, McIntyre L, Schroederk JI, Salt DE. Genomic scale profiling of nutrient and trace elements in Arabidopsis thaliana.** Nat Biotechnol. 2003; 21: 1215-21

14. **Lippard SJ, Berg JM. Principles of Bioinorganic Chemistry.** California: Mill Valley; 1994

15. **Cowan, JA. Inorganic Biochemistry.** An Introduction. 2 ed. Nueva York: Willey-VCH; 1997

16. **Escanero JF. Minerales: Elementos Traza.** En: Cocho JA, Escanero JF, González de Buitrago JM. Elementos Traza: Aspectos Bioquímicos, Analíticos y Clínicos. Barcelona: Sociedad Española de Bioquímica Clínica y Patología Molecular; 1998. p. 11-29

17. Gailer J. Arsenic-selenium and mercury-selenium binds in biology. Coord Chem Rev. 2007; 251: 234-54

18. Shroeder HA, Nason AP. Trace-element analysis in clinical chemistry. Clin Chem. 1971; 17: 461-74

19. Gutiérrez Ríos, E. Química Inorgánica. Barcelona: Reverté; 1993

20. Frisk P, Yaqob A, Lindh U. Indications of selenium protection against cadmium toxicity in cultured K-562 cells. Sci Total Envirom. 2002; 296:189-97

21. Messaoudi I, El Heni J, Hammouda F, Said K, Kerkeni A. Protective effects of selenium, zinc, or their combination on cadmium-induced oxidative stress in rat kidney. Biol Trace Elem Res. 2009; 130:152-61

22. García-Barrera T, Gómez-Ariza JL, González-Fernández M, Moreno F, García-Sevillano MA, Gómez-Jacinto V. Biological responses related to agonistic, antagonistic and synergistic interactions of chemical species. Anal Bioanal Chem. 2012; 403: 2237-53.

23. Schaumlöffel D, Lobinski R. Isotope dilution technique for quantitative analysis of endogenous trace element species in biological systems. Int J Mass Spectrom. 2005; 242: 217-23

24. Berry MJ, Banu L, Chen YY, Mandel SJ, Kieffer JD, Harney JW, Larsen PR. Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region. Nature. 1991; 353: 273-6

25. Bettmer J, Montes Bayón M, Ruiz Encinar J, Fernández Sánchez ML, Fernández de la Campa MR, Sanz Medel A. The emerging role of ICP-MS in proteomic analysis. J Proteomics. 2009; 72: 989-1005

26. Cocho JA, Escanero JF, González de Buitrago JM. Elementos Traza: Aspectos Bioquímicos, Analíticos y Clínicos. Barcelona: Sociedad Española de Bioquimica Clínica y Patología Molecular; 1998

27. Gómez-Ariza JL, Zeini-Jahromi E., González-Fernández M, García-Barrera T, Gailer J. Liquid chromatography-inductively coupled plasma-based metallomic approaches to probe health-relevant interactions between xenobiotics and mammalian organisms. Metallomics. 2011; 3: 566-77

28. Mesko MF, Hartwig CA, Bizzi CA, Pereira JSF, Mello PA, Flores EMM. Sample preparation strategies for bioinorganic analysis by inductively coupled plasma mass spectrometry. Int J Mass Spectrom. 2011; 307: 123-36

29. García-Sevillano, González-Fernández M, Jara-Biedma R, García-Barrera T, López-Barea J, Pueyo C, Gómez-Ariza JL. Biological response of free-living mouse *Mus spretus* under environmental stress based on assessment of metal-binding biomolecules by SEC-ICP-MS. DOI: 10.1007/s00216-012-6274-2

30. Gonzalez-Fernandez M, Garcia-Sevilliano MA, Jara-Biedma R, Garcia-Barrera T, Vioque A, Lopez-Barea J, Pueyo C, Gomez-Ariza JL. Size characterization of metal species in liver and brain from free-living *(Mus spretus)* and laboratory *(Mus Musculus)* mice by SEC-ICP-MS: Application to environmental contamination assessment. J Anal At Spectrom. 2011; 26: 141-9

31. Simpson RJ, Preparation of Cellular and Subcellular Extracts. En: John Inglis. Proteins and Proteomics, a Laboratory Manual. Nueva York: Cold Spring Harbor Laboratory Press; 2003. p. 91-143

32. Szpunar J, Bouyssiere B, Lobinski R. Sample Preparation for Speciation Analysis for Metallobiomolecules. En: Mester Z, Sturgeon R. Sample Preparation for Trace Element Analysis. Amsterdam: Elsevier; 2003. p. 1185-210

33. Chassaigne H, Chéry CC, Bordin G, Vanhaecke F, Rodríguez AR. 2-Dimensional gel electrophoresis technique for yeast selenium containing proteins-sample preparation and MS for processing 2D gel proteins spots. J Anal At Spectrom. 2004; 19: 85-95

34. Cañas B, Piñeiro C, Calvo E, López-Ferrer D, Gallardo JM. Trends in sample preparation for classical and second generation proteomics. J Chromatogr A. 2007; 1153: 235-58

35. Leopold I, Fricke B. Inhibition, reactivation and determination of metal ions in membrane metalloproteases of bacterial origin using high-performance liquid chromatography coupled on-line with inductively coupled plasma mass spectrometry. Anal Biochem. 1997; 252: 277-85

36. Casiot C, Szpunar J, Lobinski R, Potin-Gautier M. Sample preparation and HPLC separation approaches to speciation analysis of selenium in yeast by ICP-MS. 1999; 14: 645-50

37. Shaw M, Riederer BM. Sample preparation for two-dimensional gel electrophoresis. Proteomics. 2003; 3: 1408-17

38. Botelho D, Wall MJ, Vieira DB, Fitzsimmons S, Liu F, Doucette A. Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separations. J Proteome Res. 2010; 9: 2863-70

39. Tastet L, Schaumloffel, Lobinski R. ICP-MS-assisted proteomics approach to the identification of selenium-containing proteins in selenium-rich yeast. J Anal At Spectrom. 2008; 23: 309-17

40. Hoving S, Gerrits B, Voshol H, Müller D, Roberts RC, Oostrum J. Preparative two-dimensional gel electrophoresis at alkaline pH using narrow range immobilized pH gradients. 2002; 2: 127-34

41. González-Fernández M, García-Barrera T, Arias-Borrego A, Bonilla-Valverde D, López-Barea J, Pueyo C, Gómez-Ariza JL. Metal-binding molecules in the organs of *Mus musculus* by size-exclusion chromatography coupled with UV spectroscopy and ICP-MS. Anal Bioanal Chem. 2008; 390: 17-28

42. Gilar M, Olivova P, Daly AE, Gebler JC. Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem. 2005; 77: 6426-34

43. Michalke B. Capillary electrophoresis in speciation analysis. En: Cornelis R, Caruso J, Crews H, Heumann KG. Handbook of elemental speciation: techniques and methodology. Chichester: John Wiley & Sons; 2003. p. 201-23

44. Schaumlöffel D, Ouerdane L, Bouyssiere B, Lobinski R. Speciation analysis of nickel in the latex of a hyperaccumulating tree Sebertia acuminata by HPLC and CZE with ICP MS and electrospray MS-MS detection. J Anal At Spectrom. 2003; 18: 120-7

45. Lobinski R, Schaumlöffel D, Szpunar J. Mass spectrometry in bioinorganic analytical chemistry. Mass Spectrom Rev. 2006; 25: 255-89

46. Heisterkamp M, Adams FC. Gas chromatography-Inductively coupled plasma-Time-of-flight mass spectrometry for the speciation analysis of organoleadcompounds in environmental water samples. Fresenius J Anal Chem. 2001; 370: 597–605

47. Krupp EM, Donard OFX. Isotope ratios of transient signals with GCMC-ICP-MS. Int J Mass Spectrom. 2005; 242: 233-42

48. Mounicou S, Ouerdane L, LÁzou B, Passagne I, Ohayon-Courtès C, Szpunar J, Lobinski R. Identification of Metallothionein Subisoforms in HPLC using accurate mass and online sequencing by electrospray hybrid linear ion trap-orbital ion trap mass spectrometry. Anal Chem. 2010; 82: 6947-57

49. Nagele E, Vollmer M, Horth P, Vad C. 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. Expert Rev Proteomics. 2004; 1: 37-46

50. Cozzi J, Fraichard A, Thiam K. Use of genetically modified rat models for translational medicine. Drug Discov Today. 2008; 13: 488-94.

51. Abbott A. Laboratory animals: the Renaissance rat. Nature. 2004; 428: 464-6

52. Lindblad-Toh K. Genome sequencing: Three's company. Nature. 2004; 428: 475-6

53. Lu M, Wang H, Li X-F, Lu X, Cullen WR, Arnold LL, Cohen SM, Le XC. Evidence of Hemoglobin Binding to Arsenic as a Basis for the Accumulation of Arsenic in Rat Blood. Chem Res Toxicol. 2004; 17: 1733-42

54. Cohen SM, Arnold LL, Uzvolgyi E, Cano M, John MS, Yamamoto S, Lu X, Le XC. Possible role of dimethylarsinous acid in dimethylarsinic acid-induced urothelial toxicity and regeneration in the rat. Chem Res Toxicol. 2002; 15: 1150-7

55. Pafizek J, Zahor Z. Effect of cadmium salts on testicular tissue. Nature. 1956; 177: 1036-37

56. Ballatori N, Clarkson TW. Developmental in the biliary excretion of methylmercury and gluthation. Science. 1982; 216: 61-3

57. Gailer J, Madden S, Buttigieg GA, Denton MB, Younis HS. Appl Organomet Chem. 2002; 16: 72-5

58. Ganyc D, Talbot S, Konate F, Jackson S, Schanen B, Cullen W, Self WT. Environ Health Perspect. 2007; 115: 346-53

59. Martinez-Zamudio R, Ha HC. Environmental epigenetics in metal exposure. Epigenetics. 2011; 820-7

60. Lewinska D, Arkusz J, Stanczyk M, Palus J, Dziubałtowska E, Stepnik M. Comparison of the effects of arsenic and cadmium on benzo(a)pyrene-induced micronuclei in mouse bone-marrow. Mutat. Res. 2007; 632: 37-43

61. Saïd L, Banni M, Kerkeni A, Saïd K, Messaoudi I. Influence of combined treatment with zinc and selenium on cadmium induced testicular pathophysiology in rat. Food Chem. Toxicol. 2010; 48: 2759-65

62. Staneviciene I, Sadauskiene I, Lesauskaite V, Ivanoviene L, Kasauskas A, Ivanov L. Subacute effects of cadmium and zinc ions on protein synthesis and cell death in mouse liver. Medicina (Kaunas). 2008; 44: 131-8

63. Kenyon EM, Hughes MF, Adair BM, Highfill JH, Crecelius EA, Clewell HJ, Yager JW. Tissue distribution and urinary excretion of inorganic arsenic and its methylated metabolites in C57BL6 mice following subchronic exposure to arsenate in drinking water. Toxicol Appl Pharm. 2008; 232: 448-55

64. Brandao R, Santos FW, Farina M, Zeni G, Bohrer D, Rocha JBT, Nogueira CW. Antioxidants and metallothionein levels in mercury-treated mice. Cell Biol Toxicol. 2006; 22: 429-38

65. González-Fernandez M, García-Sevillano MA, Jara-Biedma R, Navarro-Roldán F, García-Barrera T, López-Barea J, Pueyo C, Gómez-Ariza JL. Use of metallomics in environmental pollution assessment using mice *Mus musculus*/*Mus spretus* as bioindicators. Current Analytical Chemistry, 2012; en prensa

66. Zalups RK. Progressive losses of renal mass and the renal and hepatic disposition of administered inorganic mercury. Toxicol Appl Pharmacol. 1995; 130; 121-31

67. Zalups RK. Molecular interactions with mercury in the kidney. Pharmacol Rev. 2000; 52: 113-43

68. Zalups RK, Koropatnick J. Temporal changes in metallothionein gene transcription in rat kidney and liver: Relationship to content of mercury and metallothionein protein. J Pharmcol Exp Ther. 2000; 295: 74-82

69. Edmonds JS, Francesconi K A. Arseno-sugars from brown kelp (Ecklonia radiata) as intermediates in cycling of arsenic in a marine ecosystem. Nature. 1981; 289: 602-4

70. Edmonds J S, Francesconi K A. Transformations of arsenic in the marine environment. Experientia. 1987; 43: 553-7

71. Naranmandura H, Suzuki KT. Formation of dimethylthioarsenicals in red blood cells. Toxicol Appl Pharmacol. 2008; 227: 390-9

72. Mandal BK, Ogra Y, Suzuki KT. Identification of dimethylarsenious and monomethylarsenious acids in human urine of the arsenic affected areas in West Bengal, India. Chem Res Toxicol. 2001; 14: 371-8

73. Suzuki KT, Mandal BK, Ogra Y. Speciation of arsenic in body fluids. Talanta. 2002; 58: 111-9

74. Del Razo LM, García-Vargas GG, Albores A, Vargas H, Gonsebatt ME, Montero R, Ostrosky-Wegman P, Kelsh M, Cebrián ME. Altered profile of urinary arsenic metabolites in adults with chronic arsenicism. A pilot study. Arch Toxicol. 1997; 71: 211-7

75. Hughes MF, Kenyon EM, Edwards BC, Mitchell CT, Del Razo LM, Thomas DJ. Accumulation and metabolism of arsenic in mice after repeated oral administration of arsenate. Toxicol Appl Pharmacol. 2003; 191: 202-10

76. Aposhian HV, Aposhian MM. Arsenic toxicology: Five questions. Chem Res Toxicol. 2006; 19: 1-15

77. Menzel DB, Hamadeh HK, Lee E, Meacher DM, Said V, Rasmussen RE, Greene H, Roth RN. Arsenic binding proteins from human lymphoblastoid cells. Toxicol Lett. 1999; 105: 89-101

78. Winski SL, Carter DE. Interactions of rat red blood cell sulfhydryls with arsenate and arsenite. J Toxicol Environ Health. 1995; 46: 379-97

79. Ngu TT, Stillman MJ. Arsenic binding to human metallothionein. J Am Chem Soc. 2006; 128: 12473-83

Intensity Fading MALDI-TOF MS en Proteómica Funcional e Interactómica

Sílvia Bronsoms[1], Giovanni Covaleda[2] y Francesc Xavier Aviles[3]

[1] Responsable Técnico del Servicio de Proteómica, Universitat Autònoma de Barcelona, Spain;

[2] Post-doc, Grupo de Ingeniería de Proteínas y Proteómica, Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Spain;

[3] Profesor y director del grupo de Ingeniería de Proteínas y Proteómica; Institut de Biotecnologia i de Biomedicina y Dept. de Bioquímica, Universitat Autònoma de Barcelona, Spain.

1. Introducción.

La caracterización de las interacciones no covalentes entre las macromoléculas biológicas es un elemento clave para comprender y predecir el funcionamiento de los sistemas biológicos (1). Las interacciones proteína-proteína y proteína-ligando constituyen la base para el estudio, entre otros, de las cascadas de señalización celular, la regulación bioquímica y, a un nivel superior, nos permiten determinar la conectividad de un proteoma en un contexto celular concreto. Se trata de interacciones dinámicas, que cambian en el tiempo y en el espacio, adaptándose a las diferentes condiciones fisiológicas. El análisis de estos cambios nos permite estudiar los mecanismos de adaptación celular, de defensa, de patogenicidad o de progresión de enfermedades (2).

Tradicionalmente el estudio de las interacciones proteína-proteína se ha desarrollado de forma focalizada, concentrándose en proteínas concretas y utilizando metodologías espectroscópicas (absorción en el UV, fluorescencia, dicroísmo circular), resonancia magnética nuclear, ultracentrifugación, cromatografía o biacore entre otras. Los avances que se han producido estos últimos años con la era de la proteómica y la llegada de la biología de sistemas ha requerido la puesta a punto de metodologías de alto rendimiento, que permitan estudiar interacciones a gran escala de forma directa y rápida (3). Este cambio de paradigma en el campo de la biología molecular/celular se ha producido paralelamente a la explosión de la espectrometría de masas (MS, como abreviatura inglesa) como herramienta de trabajo para la identificación y caracterización de proteínas y complejos proteicos (4). Las dos ventajas principales son su velocidad y sensibilidad. La elevada velocidad de análisis, junto con la baja cantidad de muestra requerida, está permitiendo utilizar esta técnica para la identificación de las proteínas presentes en multitud de muestras y condiciones biológicas distintas, determinar su nivel de expresión y caracterizar sus modificaciones post-traduccionales. En estos últimos años también se han desarrollado estrategias a gran escala para la caracterización de las interacciones proteína-proteína, que combinan estrechamente la MS con otras técnicas, por ejemplo la purificación por afinidad (5). Se trata de estrategias indirectas de detección de interacciones que constan de una primera etapa de aislamiento/purificación del complejo y una segunda etapa de identificación de los componentes por MS. Finalmente, también se realiza la caracterización directa de las interacciones no-covalentes por MS en la cual el complejo proteína-proteína es directamente ionizado y analizado por MS. Este análisis directo de los complejos biomoleculares

nos puede facilitar información sobre la identidad de sus componentes, de su estequiometría e incluso de su afinidad (6).

2. Caracterización de Interacciones no Covalentes Utilizando MALDI-TOF MS.

La MALDI-TOF MS (Matrix-Assisted Laser Desorption Ionization Time Of Flight MS), juntamente con la ionización por electroespray, proporcionan un método de ionización suave por lo cual acostumbran a ser las técnicas de elección para el análisis de interacciones no covalentes por MS. La ionización por electroespray es la técnica de ionización más suave y también la primera dónde se observó que algunos componentes unidos de forma no covalente podían ser estudiados por MS (7). Aunque la capacidad de la MALDI MS para la detección de interacciones no covalentes ha sido demostrada desde hace años (8, 9) todavía no ha encontrado una aplicabilidad general, probablemente debido a las precauciones experimentales que deben tomarse para evitar la disrupción del complejo o a la presencia de oligómeros no específicos.

Existen una serie de parámetros que afectan de forma directa a la formación del complejo no covalente en la MALDI MS: la matriz, el pH, los disolventes, la fuerza iónica, la concentración de la muestra o los disparos de laser (first-shot phenomenon) entre otros (10, 11). Pero teniendo en cuenta estos parámetros, y evitando condiciones demasiado severas durante la preparación de la muestra, la MALDI MS puede utilizarse para la detección de interacciones no covalentes (12-15).

3. La Estrategia de Intensity Fading MALDI-TOF MS.

La Intensity Fading (IF) MALDI-TOF MS es una aproximación que permite detectar la formación de complejos no covalentes por MALDI-TOF MS a través del seguimiento de las intensidades de las señales moleculares (m/z) que se ven disminuidas (fading) al ser adicionada a la muestra de análisis una molécula diana o una molécula interactora (15).

En el caso de proteínas, esta estrategia se basa en la comparación de los espectros totales de masas MALDI-TOF MS de una mezcla o extracto biológico antes y después de la adición de una proteína diana (también llamada receptora) o de un ligando proteico (16). No se suele buscar ni detectar en el espectro inicial la(s) señal(es) correspondiente(s) a complejos no covalentes formados en la mezcla o extracto, sino la disminución de señal(es) de sus componentes "interactores" al ser desplazado su equilibrio hacia la formación de complejo(s) por la adición de la molécula diana (Figura 1a). Además, habitualmente no se observan las señales correspondientes a complejos no covalentes o a la molécula diana, debido al uso de una ventana analítica de masas restringida, a la baja concentración que alcanzan en la muestra o a la disminución de la eficiencia de detección al aumentar la relación masa/carga de dichas moléculas (especialmente si se usan detectores convencionales, p.e. tipo placas microcanal o MCPs) (17).

Esta aproximación experimental puede llevarse a cabo siguiendo dos protocolos: el protocolo directo, dónde se compara el espectro de masas MALDI de la muestra biológica antes y después de la incubación con la molécula diana (Figura 1a); y el protocolo indirecto, en dónde se incorpora un paso de captura por afinidad, utilizando moléculas diana inmovilizadas en micro-

partículas magnéticas o microcolumnas (Figura 1b) (18).

Con tal de ilustrar el protocolo experimental y mostrar la aplicabilidad de esta metodología se describen a continuación los detalles experimentales de esta aproximación y se complementa con ejemplos obtenidos con interacciones modelo, utilizando proteasas como molécula diana e inhibidores proteicos como moléculas ligando (con rangos de masas entre 20-40 kDa y 4-20 kDa, respectivamente).

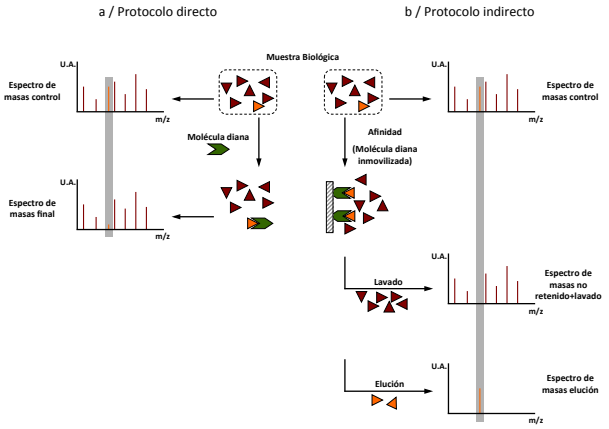


Figura 1. Esquema de la aproximación de Intensity Fading MALDI-TOF MS. A: Protocolo directo. Se detecta la desaparición de la molécula ligando por MS después de la incubación con la molécula diana. B: Protocolo indirecto. Inmovilización de la molécula diana y posterior unión del extracto, lavado y elución de la molécula ligando.

3.1. Equipamiento.

Los datos presentados en este capítulo fueron obtenidos utilizando espectrómetros de masas MALDI-TOF Ultraflex y UltrafleXtreme (Bruker Daltonics, Bremen, Alemania), pero pueden utilizarse equipos de otras casas comerciales que posean una fuente de ionización MALDI.

3.2. Preparación de las muestras.

3.2.1. Muestras simples para la puesta a punto de la aproximación.

La puesta a punto de la aproximación experimental se realizó utilizando inhibidores y proteasas como modelo de interacción proteína-proteína. En todos los casos se eligieron parejas inhibidor-proteína con una interacción ya conocida. La mayoría de inhibidores y proteasas ensayados se adquirieron de casas comerciales (Sigma y Bachem), pero algunos de ellos fueron obtenidos de forma recombinante en el laboratorio (p.e. carboxipeptidasas, CPs, tales como CPA4, así como de la latexina o inhibidor humano de CPs, y el inhibidor de carboxipeptidasa de patata, PCI).

Una vez elegida la proteasa diana a ensayar, se prepara una disolución madre con una mezcla de moléculas que no se unan a la proteasa y con una que sí lo haga. Se prepara la disolución con un tampón adecuado para que se pueda producir la interacción, normalmente se elige un tampón similar al que se usa para el ensayo de actividad de la enzima (p.e. TrisHCl 10 mM pH 7.5). La concentración de la muestra de ensayo debe idealmente mantenerse en el rango submicromolar, ya que se ha observado que (para el tipo de proteínas aquí ensayado) el fenómeno de IF se observa con más dificultad a partir de concentraciones de analito de 15 a 20 μM.

3.2.2. Muestras complejas y extractos biológicos.

En el caso de extractos biológicos habitualmente se trabaja con prefraccionamiento de la muestra o disminución de su complejidad, para minimizar los fenómenos de supresión de señal y aumentar así el número total de moléculas ionizadas y por lo tanto mejorar la representatividad de los péptidos y pequeñas proteínas presentes en el extracto original (19-20). En caso de extractos poco complejos los análisis pueden resultar más limpios y sensibles si se realiza un paso de desalado mediante micro-columnas de C4 o C18 o de diálisis (por flotación) usando filtros de 0.025 μm.

En el caso de muestras más complejas, donde es más probable que los ligando no ionizen bien debido a fenómenos como la supresión de señal, el extracto, normalmente un liofilizado,

se disuelve en un tampón de baja fuerza iónica (agua milliQ o TrisH-Cl 10 mM pH 7.5), se centrifuga y el sobrenadante se prefracciona mediante una cromatografía (19). Esta puede ser de fase reversa, de intercambio iónico o de gel-filtración, entre otras. Se colecta el resultado de la cromatografía en fracciones (de 10 a 50 en función de la complejidad de la muestra) y se liofilizan (Figura 2).

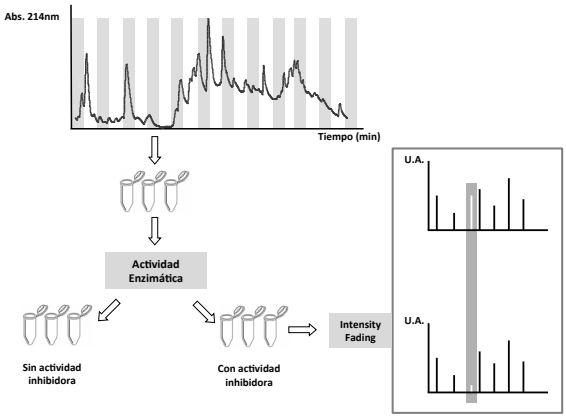


Figura 2. Esquema de la aproximación de Intensity Fading MALDI-TOF MS con pre-fraccionamiento. El extracto se somete a una separación cromatográfica y se ensaya la actividad enzimática de las alícuotas obtenidas. Las fracciones que presentan actividad inhibidora son analizadas por Intensity Fading para identificar la molécula responsable de esa actividad.

En el estudio de las interacciones proteasa-inhibidor, se realiza un ensayo de actividad enzimática con cada una de las fracciones, y así se determina en cuáles de ellas existe actividad inhibidora, intentándose establecer un vínculo entre señales de masa concretas y capacidad inhibidora (19). Es muy importante disolver las fracciones en un tampón adecuado para la formación de la interacción; habitualmente se utiliza el tampón que sea más adecuado para los ensayos de actividad enzimática. Se pueden estudiar varias proteasas diana al mismo tiempo realizándose ensayos en paralelo con cada una de las enzimas y determinando las fracciones de interés para cada proteasa. En estos casos, idealmente, los ensayos de actividad se realizan en lectores multiplaca, que facilitan los análisis.

3.3. Protocolo directo.

La aplicación de esta estrategia requiere la formación del complejo no covalente en la muestra de ensayo (en solución) y mantenerlo en estas condiciones nativas durante el proceso de cristalización con la matriz y su posterior ionización durante el análisis por MALDI MS.

3.3.1. Preparación de la matriz.

Las condiciones de preparación de la matriz son críticas para la formación del complejo no covalente. Durante el proceso de puesta a punto del protocolo con moléculas modelo se estudiaron los parámetros clave de la preparación de la matriz y su influencia en la formación/mantenimiento de los complejos (15,20). Las condiciones óptimas para una correcta cristalización y mantenimiento de los complejos son:

- Matriz. El ácido sinapínico (SA) y la 2,4,-dihidroxiacetofenona (DHAP) son las matrices con mejores resultados para la detección de las interacciones.

- Es importante que la concentración de disolvente orgánico no sea muy elevada (<30% en la matriz), para que se mantenga en valores por debajo del 15% después de la mezcla con la muestra.

- Variaciones de pH en el rango de 4 a 6 no afectan a la formación del complejo, pero valores muy bajos (<2) pueden causar la pérdida de la interacción.

- El método de preparación de la muestra por *dried droplet* (21) da mejores resultados que el de capa fina, debido a la importancia de la morfología de los cristales, que debe ser muy homogénea.

- Para los ensayos de utilizaron SA y DHAP a 5-10 mg/mL, en 20% acetonitrilo y 20mM citrato de amonio dibásico.

3.3.2. Incubación y análisis.

Se prepara una dilución seriada de la muestra problema, ya sea la disolución madre que hemos preparado o un extracto biológico (fraccionado o sin fraccionar). Para cada punto de la dilución se prepara una mezcla 1:1 en volumen con cada una de las matrices (DHAP y SA), se pipetea suavemente la mezcla unas 10 veces y se depositan 0.5 μL de la mezcla en una placa MALDI. El análisis de estas muestras nos da el espectro control.

Para cada punto de la dilución seriada se mezcla 1:1 con la disolución de la proteasa diana, pipeteando suavemente entre 5 y 10 veces. Se ensayan distintas concentraciones de la proteasa, entre 1 y 10 mg/ml y se incuba la mezcla durante unos pocos minutos (3-5 minutos). Finalmente se mezcla 1:1 con las dos matrices (DHAP y SA), pipeteando suavemente y se depositan 0.5 μL de la mezcla en una placa MALDI.

La comparación de este espectro con el espectro control nos permite determinar que iones moleculares desaparecen o ven disminuida su intensidad de señal (*Intensity fading*), indicando la existencia y masa de posibles ligandos.

3.3.3. Puesta a punto de la técnica: muestras simples.

Para la puesta a punto de la aproximación se realizan ensayos utilizando muestras simples e interacciones modelo. En nuestro caso se eligió modelos concretos de interacción proteasa-inhibidor, ya que se trata de sistemas muy bien caracterizados y con ensayos de actividad enzimática ya establecidos que permiten realizar el seguimiento y validación de los resultados.

En los primeros estudios se realizan ensayos utilizando distintas moléculas diana y ligandos, para los cuales las interacciones ya son conocidas. En nuestro caso utilizamos entre otras:

- Una serín-peptidasa (tripsina), usando el inhibidor pancreático básico de tripsina (BPTI) como ligando y el PCI como control.

- Una cisteína peptidasa (papaína), usando una molécula orgánica (E64) como ligando y otra molécula orgánica (3BP2) como control.

- Una metalo-peptidasa (CPA4) usando latexina (inhibidor proteico de ella) como ligando y tripsinógeno como molécula control.

Para los tres casos de estudio se prepararon disoluciones madre con la molécula control y la molécula ligando que permitieran obtener un espectro de masas MALDI-TOF con una intensidad de señal similar para ambas moléculas (Figura 3, paneles superiores) y se realizó una dilución seriada de esta disolución madre. Posteriormente se incubó cada uno de los puntos de la dilución seriada con la proteasa diana durante 3 minutos y se adquirió de nuevo el espectro de masas. Para cada pareja proteasa-inhibidor se elige la dilución en la cual se visualiza mejor el fenómeno de IF. En la figura 3 se presentan los resultados de los espectros control (paneles superiores) y espectros post-incubación (paneles inferiores) correspondientes a cada interacción modelo. En todos los casos se observa una clara disminución (fading) de la señal correspondiente a la molécula diana en relación a la molécula control. Este efecto se observa también en el caso del inhibidor E64, dejando patente que esta aproximación funciona tanto para la detección de interacciones proteína-proteína como molécula orgánica-proteína.

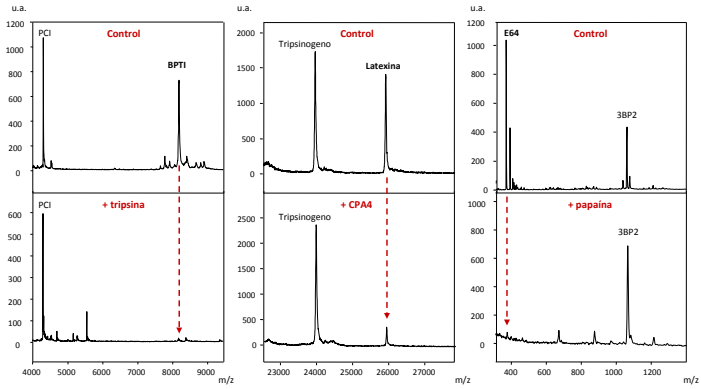


Figura 3. Intensity Fading MALDI-TOF MS en interacciones modelo. Espectros de masas MALDI-TOF de la interacción entre una proteasa modelo y una mezcla simple de ligandos. A: BPTI y PCI (control negativo de unión) frente a tripsina. B: Latexina y tripsinógeno (control negativo) frente a carboxypeptidasa A4. C: E64 y 3BP2 (control negativo) frente a papaína. Las mezclas fueron analizadas antes (paneles superiores) y después (paneles inferiores) de la adición de la proteasa.

3.3.4. Muestras complejas: Búsqueda de inhibidores de tripsina en un extracto de anémona marina.

Analizar muestras complejas mediante el protocolo directo de IF implica aceptar el hecho que un número importante de las especies presentes en nuestro extracto no se visualizarán (al menos con las actuales técnicas y protocolos) debido a efectos de supresión de señal. Cuando analizamos una muestra muy compleja de forma directa por MS MALDI-TOF sólo obtenemos señal de una parte de las especies presentes, debido a la competición que se establece entre todos los analitos para ser ionizados. Este fenómeno que se conoce como supresión de señal se ve influenciado por la concentración de la muestra, su complejidad, la presencia de sales u otros contaminantes, entre otros (19-22).

Para analizar la muestra de forma exhaustiva es muy conveniente realizar un fraccionamiento previo (normalmente cromatográfico) para reducir la complejidad de especies en cada fracción (y, con ello, minimizar la supresión de señal) y analizar cada una de las fracciones de forma individual (ver figura 2), o bien utilizar el protocolo indirecto (Figura 1b) que incorpora una etapa de incubación con la molécula diana inmovilizada. El protocolo directo sin prefraccionamiento puede utilizarse en algunos casos como método rápido de cribado para la identificación de ligandos que se encuentran a concentraciones relativamente altas en el extracto proteico.

Un ejemplo de esta aplicación lo encontramos en la búsqueda de inhibidores de proteasas en los extractos de organismos invertebrados marinos. Se partió de extractos de organismos marinos pertenecientes a distintos filos (Mollusca, Chordata, Chidaria y Annelida) y fueron analizados por IF-MALDI-TOF usando tripsina,

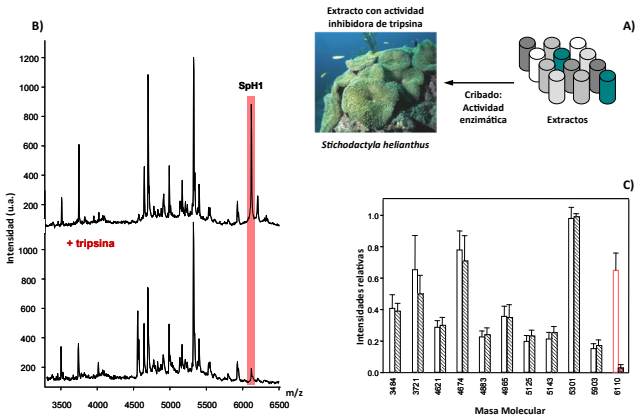


Figura 4. Intensity Fading en un extracto de *S. helianthus*. B: Espectro de masas MALDI-TOF del extracto antes (panel superior) y después (panel inferior) de la incubación con tripsina bovina. C: Intensidades relativas de las principales moléculas observadas en el espectro de masas del extracto. Las barras blancas y las barras cruzadas corresponden a los valores de intensidad antes y después, respectivamente, de la incubación con tripsina. Los resultados son la media de diez experimentos independientes.

CPA y papaína como moléculas diana. En el caso de la especie *Stichodactyla helianthus* se pudo identificar un inhibidor de tripsina de 6110 Da de masa molecular. En la figura 4B se presenta el espectro control del extracto (panel superior) y el espectro obtenido después de la incubación con tripsina bovina (panel inferior), dónde se observa una clara disminución de la intensidad de señal (fading) de la molécula de 6110 Da. Esta masa molecular coincide con la del inhibidor de tripsina SphI-1, un inhibidor tipo Kunitz que había sido previamente descrito por otro grupo de investigación (23). Para confirmar que se trataba del mismo inhibidor, se purificó la molécula por HPLC de fase reversa y se determinó su secuencia N-terminal por degradación de Edman, la cual cosa permitió la identificación de la molécula y la validación de la aproximación de IF. También se pudo observar la aparición de un pico de 4546 Da en el espectro inferior, pero se trata de un péptido resultante de la autolisis de la tripsina.

En la figura 4C se muestra un resumen estadístico de la media de diez experimentos independientes. El análisis estadístico muestra que en este extracto la especie de 6110 Da es la única que presenta una disminución de la intensidad relativa de señal estadísticamente significativa tras la adición de tripsina.

3.3.5. Otras aplicaciones: determinación semicuantitativa de constantes de afinidad.

Dados los buenos resultados de la aproximación de IF para la búsqueda y validación de interacciones se analizó la capacidad de esta metodología para evaluar las afinidades relativas de unión entre un inhibidor y una proteasa.

Se usó como modelo para el estudio la interacción entre la CPA y el inhibidor PCI, utilizando también un mutante de este inhibidor con los tres últimos residuos de la cola c-terminal delecionados (Δ3 PCI). Las constantes de inhibición de estas moléculas para la CPA están descritas (24):

PCI: $K_i = 1.5$ nM

Δ3 PCI: $K_i = 40$ μM

Se trata de dos moléculas muy parecidas a nivel de estructura primaria, pero con unas constantes de inhibición de cuatro órdenes de magnitud de diferencia.

Se prepara una disolución madre con las dos proteínas ligando a la misma molaridad (16 μM) y se obtiene el espectro de masas control, sin CPA (Figura 5, espectro superior). Se mezcla esta disolución con concentraciones decrecientes de CPA, partiendo de una concentración en exceso de la peptidasa de 10 veces (160 μM) hasta una concentración de 1 μM. Ante la presencia de un exceso de CPA las dos señales correspondientes a los inhibidores sufren una disminución casi total. Cuando la concentración de molécula diana va disminuyendo, la señal correspondiente al PCI sigue muy disminuida, mientras que la señal correspondiente al mutante de menor afinidad (Δ3 PCI) se recupera de manera rápida. Este efecto es debido a la competición que se establece entre estas dos moléculas para unirse a la CPA, de tal manera que cuando la concentración de CPA es limitante la mayor parte de la molécula diana está unida al ligando de mayor afinidad (PCI) y por lo tanto el ligando de afinidad inferior (Δ3 PCI) se libera y da señal en el espectro de masas.

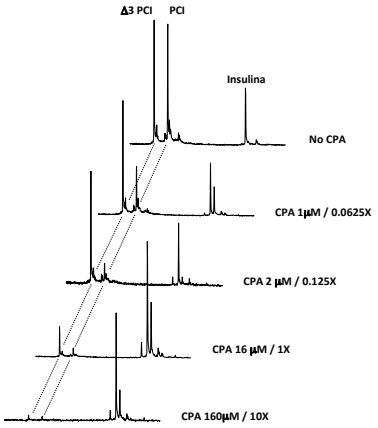


Figura 5. Determinación semicuantitativa de afinidades de unión por Intensity Fading. Espectros de masa MALDI-TOF de una mezcla de inhibidores de carboxipeptidasa A (CPA), PCI y Δ3PCI, y una molécula control (insulina) que no une CPA antes y después de la adición de concentraciones crecientes de CPA.

Este comportamiento también ha podido observarse para otros modelos de interacción proteína-proteína, por ejemplo entre la calmodulina y los péptidos de unión a esta proteína (25) o en el cálculo de constantes de afinidad relativas de distintos alcaloides para la glicoproteína ácida-alfa-1 (26) indicando que se trata de una buena metodología para obtener información sobre afinidad de unión de una forma rápida y con poca cantidad de proteína.

3.4. Protocolo indirecto.

El protocolo indirecto incorpora una etapa de captura por afinidad del ligando y su posterior elución y análisis por MS, de esta manera las muestras complejas pueden ser analizadas en la mayoría de los casos por IF sin necesidad de realizar un prefraccionamiento. Este protocolo presenta un inconveniente derivado del proceso de inmovilización en un soporte sólido: en algunos casos la interacción proteína-proteína puede verse modificada o limitada, dando lugar a falsos negativos de interacción. Pero este aspecto se ve compensado por una mayor eficiencia de análisis y una gran facilidad de interpretación de resultados, ya que la aplicación de este protocolo permite eluir de forma exclusiva o casi exclusiva, las moléculas ligando.

3.4.1. Preparación de la matriz.

Para el protocolo indirecto las condiciones de preparación de la matriz no son tan críticas, ya que la formación de los complejos es previa al análisis por MS, en consecuencia el porcentaje de solvente orgánico y de ácido trifluoroacético en las matrices puede aumentarse. Para los ensayos aquí descritos se utilizaron como matrices SA a 5-10 mg/mL, en 33% acetonitrilo y 0.1% àcido trifluoroacético y DHAP a 5-10 mg/mL, en 20% acetonitrilo y 20mM citrato de amonio dibásico.

3.4.2. Inmovilización.

La inmovilización de las proteínas diana puede realizarse utilizando diferentes soportes sólidos, siendo el de uso más habitual los “beads” o micropartículas de agarosa comerciales (con o sin núcleo magnético) o de material similar. El protocolo de inmovilización depende de la naturaleza del soporte sólido elegido y cada casa comercial especifica el protocolo adecuado para cada tipo de resina (Sigma, Pierce o GE Healthcare, entre otros).

Se describe a modo de ejemplo el protocolo utilizado en una de las aplicaciones que se detallan más adelante (ver apartado 3.5.6: muestras complejas: Búsqueda de un inhibidor de CP en un extracto de Ascaris), dónde se inmoviliza CPA en *Sepharose 4B* activada con bromuro de cianógeno, de GE Healthcare). Este soporte es una resina pre-activada utilizada para la inmovilización de proteínas a través de las aminas primarias.

El primer paso del proceso es la reacción de acoplamiento, para el caso aquí descrito se utilizó un tampón de acoplamiento $NaHCO_3$ 0.1M pH 8.3 con NaCl 0.5 M para disolver la molécula diana. Se mezcla la CPA con la resina, usando una proporción de 5 mg de CPA por cada mililitro de *Sepharose 4B* y se incuba durante una hora con agitación suave. El exceso de ligando es eliminado con varios lavados sucesivos con tampón de acoplamiento. Para bloquear los posibles sitios activos que no hayan reaccionado durante el proceso de acoplamiento se equilibra la resina con tampón TrisHCl 0.1 M pH 8.0 durante dos horas y posteriormente se somete a seis lavados consecutivos donde se alternan dos tampones: CH_3COOH 0.1 M con 0.5 M NaCl y Tris HCl 0.1 M pH 8.0 con 0.5 M NaCl. Estos lavados permiten eliminar el exceso de molécula diana que haya podido quedarse unido mediante interacciones iónicas. Es muy importante controlar el pH de la reacción de acoplamiento ya que la máxima eficiencia de reacción se obtiene a pH > 8.0. También debe tenerse en cuenta que no pueden utilizarse tampones con grupos amino primarios (por ej. Tris HCl) porqué compiten con la molécula diana para unirse a la *Sepharose 4B*. La presencia de sales en el tampón de acoplamiento también es imprescindible ya que minimiza la formación de agregados de proteína.

En el caso de la inmovilización de una enzima, es importante que ésta no modifique el ligando; si fuera así, la molécula eluída y la molécula presente inicialmente en el extracto presentarían masas moleculares distintas y eso dificultaría la interpretación de los espectros de MS. Si es posible, y especialmente cuando se trabaja con proteínas agresivas, como son las proteasas, es preferible trabajar con enzimas modificadas, ya sea por modificación química o por mutagénesis dirigida, para asegurar que mantengan su capacidad de unión pero no de hidrólisis (19).

Una vez inmovilizada la proteína de interés en la resina se continúa el protocolo con la incubación con el extracto problema, los lavados y la elución. Se puede trabajar en "batch" o bien utilizar otras estrategias, como el uso de microcolumnas (generalmente más efectivo). También existen alternativas a los beads de agarosa tradicionales: los beads o micropartículas magnéticas, con un núcleo magnético y una superficie derivatizada con grupos amino o carboxilo reactivos, para la unión covalente de los ligandos de forma rápida y eficiente (Invitrogen, Pierce o Dynal Biotech ente otras casas comerciales).

3.4.3. Unión / elución / análisis.

El fraccionamiento del extracto biológico de interés utilizando la resina de afinidad se realiza en un tampón de unión de baja fuerza iónica, por ejemplo Tris HCl 10-20 mM pH 7.5, e incubando la mezcla durante unos minutos (1-5 minutos) a temperatura ambiente. Posteriormente se lava la mezcla con un exceso de tampón de unión y finalmente se procede a la elución de la molécula de interés, que puede realizarse por cambio de pH, por incremento de fuerza iónica o por elución competitiva. En la aproximación de IF la elución por cambio de pH es la más adecuada (p.e. ácido fórmico 0.1%) ya que permite analizar de forma directa el resultado de la elución por MS, evitando los problemas de análisis que se generan si se utilizan concentraciones de sales elevadas.

3.4.4. Muestras complejas: Búsqueda de inhibidores de tripsina en un extracto del caracol marino *Cenchritis muricatus*.

Los invertebrados marinos representan una importante fuente de nuevas moléculas, las cuales pueden ser potencialmente utilizadas para investigación básica, aplicaciones industriales y en biomedicina. Entre estas moléculas, se cuenta con inhibidores de proteasas y otros polipéptidos bioactivos que han sido identificados en estos organismos marinos (27-29).

Teniendo en cuenta la creciente demanda de nuevas moléculas altamente selectivas y potentes frente a diferentes proteasas identificadas como dianas terapéuticas, se han analizado una serie de invertebrados marinos, los cuales representan los principales *phyla*. Como se observa en la figura 6, en la identificación de inhibidores de proteasas mediante *Intensity Fading* MALDI-TOF MS se han empleado especies pertenecientes a los *phyla* Cnidaria (corales, medusas, anémonas), Porifera (esponjas), Bryozoa (ectoproctos), Mollusca (caracoles, almejas), Annelida (gusanos segmentados), Echinodermata (estrellas y pepinos marinos) y Chordata (ascidias).

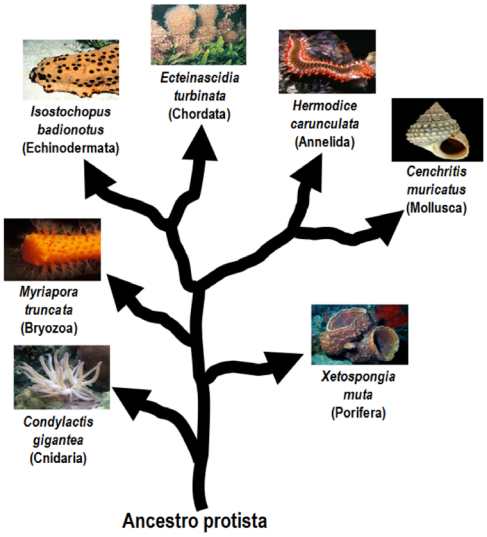


Figura 6. Árbol de clasificación con ejemplos de especies de invertebrados marinos utilizadas en la identificación de inhibidores de proteasas mediante Intensity Fading MALDI-TOF MS.

Los resultados obtenidos del análisis de *Intensity Fading* MALDI-TOF MS

del extracto crudo del caracol marino *Cenchritis muricatus* frente a tripsina inmovilizada en un soporte de glioxal Sepharose® CL-4B según el protocolo descrito por Guisan et al (30) (método indirecto) se presentan en la figura 7. El espectro de masas total (o espectro de masas control) es bastante heterogéneo, como es habitual para este tipo de muestras biológicas. En un segundo paso la muestra de extracto marino fue puesta en contacto con una matriz de tripsina-glioxal Sepharose. A continuación, como se muestra en el espectro MALDI de la fracción no retenida (NR, figura 7), se observó la desaparición de señales a m/z de 6732.7, 6005.1, 5947.7, 5482.8, 4748.7 y 3780.3, así como el decaimiento de señales a m/z de 4993.7 y 1519.6. En los espectros MALDI de los lavados subsiguientes de la matriz de afinidad se observaron señales entre 1 y 3 kDa, las cuales paulatinamente fueron disminuyendo en intensidad a lo largo de los diferentes lavados (LV-1 a LV-5, figura 7) y que habían sido observadas anteriormente tanto en el espectro del extracto total, como en la fracción no retenida. Este comportamiento parece indicar que estas señales podrían ser debidas a la eliminación de exceso de muestra, a interacciones débiles con la enzima inmovilizada o a interacciones inespecíficas con el soporte de glioxal Sepharose. En el paso de elución se observó la recuperación de señales MALDI a m/z de 6004.8, 5946.5 y 5483.5 (EL, figura 7), las cuales habían desaparecido completamente del espectro MALDI después de la interacción de la muestra con la matriz de afinidad. Estas señales fueron atribuidas a la presencia de inhibidores de tripsina en el extracto de *Cenchritis muricatus*, lo cual se confirmó mediante la determinación de actividad inhibidora de tripsina en esta fracción de elución. Para ello se usó un ensayo espectrofotométrico de cinética enzimática utilizando un sustrato cromogénico (31). En efecto, la señal MALDI en esta fracción de elución a m/z 5483.5 se corresponde con la masa molecular de CmPI-II (Uniprot ID: P84755), un inhibidor de serino proteasas que había sido aislado y caracterizado del extracto de *Cenchritis muricatus* (32). CmPI-II es un potente inhibidor de elastasa de neutrófilos humana y de tripsina, además es capaz de inhibir a la elastasa pancreática y a la subtilisina A. Este inhibidor de 50 residuos presenta un domino tipo Kazal no clásico (33). Las otras señales detectadas en la fracción de elución del experimento de *Intensity fading* pueden ser atribuidas a isoformas de CmPI-II, algunas de las cuales han sido descritas anteriormente: CmPI y CmPI-III (32-33) o bien podría tratarse de otros inhibidores de tripsina presentes en este molusco marino, no identificados y caracterizados hasta ahora.

Además de las señales mencionadas anteriormente, en la fracción de elución también se observaron otros picos de menor intensidad a 6715.5, 5875.6 y 2974.1 Da. que no habían sido detectados en ninguna de las fracciones anteriores, lo que podría indicar la presencia de señales suprimidas en el espectro total o productos de hidrólisis, ya que la tripsina inmovilizada mantiene su actividad proteolítica durante el experimento.

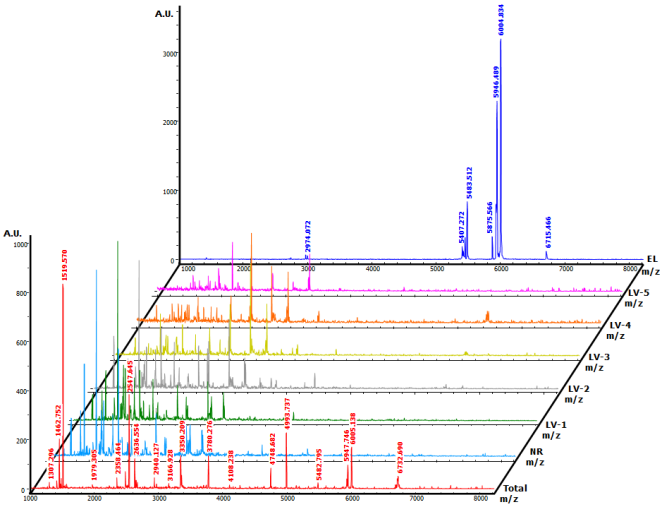


Figura 7. Intensity Fading de un extracto crudo de *Cenchritis muricatus* frente a tripsina-glioxal Sepharose® CL-4B. Total: espectro MALDI del extracto crudo de *Cenchritis muricatus* en tampón Tris-HCl 20 mM pH 8.0, NaCl 150 mM, $CaCl_2$ 20 mM. NR: espectro MALDI de las moléculas no retenidas, interacción a pH 8.0, 10 min. LV-1 a LV-5: espectros MALDI de las diferentes fracciones de lavado a pH 8.0. EL: espectro MALDI de la fracción de elución con TFA 0.5% v/v, 10 min. Todas las fracciones fueron desaladas mediante ZipTip®C_4 (millipore, USA). Cinco réplicas de cada fracción fueron analizadas por MS MALDI-TOF usando la matriz DHAP.

3.4.5. Muestras complejas: Búsqueda de CPs en el extracto del anélido marino *Sabellastarte magnifica.*

Los invertebrados marinos son organismos poco estudiados y que pueden ser fuente de nuevas proteasas e inhibidores, como vimos en el apartado anterior. En este trabajo se estudiaron 25 especies de invertebrados marinos y se realizó un primer cribado por actividad enzimática, dónde se analizó su actividad CPA utilizando N-(4-metoxifenilazormormil)-L-fenil-alanina) (AAFP) (34).

De las 25 especies estudiadas, tres de ellas presentaban una actividad específica positiva, y con una relación lineal entre la cantidad de extracto utilizado y la actividad detectada. El anélido marino *Sabellastarte magnifica* presentaba la mayor actividad específica y se eligió como fuente para caracterizar la actividad CP presente en el extracto.

En este caso el organismo presenta dos partes claramente diferenciadas, de manera que se obtuvieron dos extractos, uno del cuerpo y otro del plumero; pero tan solo se detectó actividad CPA en el cuerpo. Este extracto se mezcló con beads de agarosa con PCI inmovilizado y se incubó durante tres minutos. Se analizó por MS MALDI-TOF el extracto inicial, el lavado con tampón tris-HCl 10mM pH 7.5 y la elución posterior con ácido fórmico al 0.1% (Figura 8B). La perturbación del espectro MALDI-TOF del extracto causada por la incubación con los beads con PCI inmovilizado afecta principalmente a la señal de 34 kDa, que desaparece después de la incubación con el PCI y que posteriormente fue recuperada casi de forma pura tras la adición de ácido fórmico al 0.1%. Esta señal presenta una masa molecular similar a la de las otras metalocarboxipeptidasas descritas, cuando se encuentran en la forma activa (no de zimógeno).

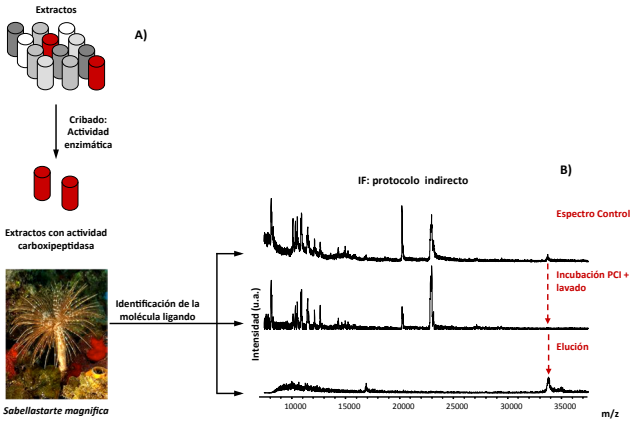


Figura 8. Identificación por Intensity Fading de una carboxipeptidasa en un extracto de *Sabellastarte magnifica*. A: Cribado de los extractos de ensayo por actividad enzimática. B: Espectros de masa MALDI-TOF del extracto antes (panel superior) y después (panel intermedio) de la incubación con PCI. La recuperación de la señal m/z correspondiente a la carboxipeptidasa se realizó eluyendo con ácido fórmico 0.1% (panel inferior).

Posteriormente se purificó esta proteína mediante dos pasos de cromatografía, un primer paso de cromatografía de afinidad y un segundo paso de intercambio inónico, y la proteína pura se caracterizó (35). Pudo confirmarse su actividad CP y determinarse sus constantes cinéticas así como obtener información secuencial por Edman y MS/MS, la cual cosa confirmó una elevada homología con otras metalo-carboxipeptidasas ya descritas.

3.4.6. Muestras complejas: Búsqueda de un inhibidor de CPs en un extracto de la lombriz intestinal *Ascaris suum*.

Ascaris suum es un parásito que infecta el tracto gastrointestinal de los cerdos causando la ascariasis, una enfermedad con un gran impacto económico en la producción porcina. Como estrategia defensiva este parásito produce un gran variedad de inhibidores de proteasas digestivas y del sistema inmunológico, como papaína, quimotripsina o catepsina entre otras. Se trata por lo tanto, de una fuente natural de inhibidores de proteasas, que se eligió para intentar identificar un inhibidor de metalo-carboxipeptidasa mediante la estrategia de IF.

En este caso, se partió de un extracto de gusanos adultos de ascaris. El homogenado se centrifugó y la fracción soluble se separó mediante una columna atmosférica C18 para desalar la muestra. Se descartó la fracción no unida y el posterior lavado con agua, y la elución se realizó en un único paso utilizando acetonitrilo al 60% .La muestra liofilizada se resuspendió en un tampón TrisHCl 20mM pH 7.5 y se analizó por MS MALDI-TOF obteniéndose el espectro control de la muestra. Se utilizó *Sepharose* 4B activada con bromuro de cianógeno para inmovilizar CPA1 bovina y obtener CPA-*Sepharose*. Se mezclaron a partes iguales el extracto y la CPA- *Sepharose* y se incubó la mezcla durante cinco minutos a temperatura ambiente. La muestra no unida se analizó por MS, posteriormente se lavó la mezcla con tampón Tris-HCl 10mM pH 7.5, se eluyó con ácido fórmico al 1% y se adquirió el espectro resultante de la elución. En la figura se observa una perturbación clara del espectro control (Figura 9B; espectro superior) al ser comparado con el espectro de la muestra no unida (Figura 9B; espectro intermedio); la señal correspondiente a la molécula de 7.7 kDa desaparece tras la incubación con la CPA- *sepharose*. En el espectro de la elución (Figura 9B; espectro inferior) se obtiene una única señal que coincide con la molécula desaparecida durante la incubación, indicando de forma clara la existencia de una molécula en el extracto de ascaris que interacciona con la CPA. Se realizaron diez experimentos independientes y se calculó la media de las intensidades relativas de todas las señales detectadas, la cual cosa permitió comprobar que hay una única señal que se ve perturbada por la incubación con CPA-*Sepharose* de forma estadísticamente significativa y corresponde a la señal de 7.7 kDa (Figura 9C). Posteriormente esta molécula fue purificada y caracterizada, demostrándose que se trata de un inhibidor fuerte de metalo-carboxipeptidasas tipo A/B (36).

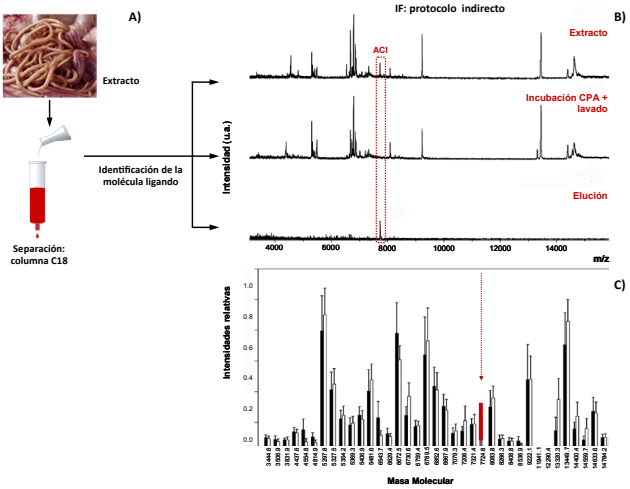


Figura 9. Identificación por Intensity Fading de un inhibidor de carboxipeptidasa en un extracto de *Ascaris suum*. A: Pre-fraccionamiento del extracto. B: Espectros de masa MALDI-TOF del extracto antes (panel superior) y después (panel intermedio) de la incubación con carboxipeptidasa A (CPA). La recuperación de la señal m/z correspondiente al inhibidor se realizó eluyendo con ácido fórmico (panel inferior). C: Intensidades relativas de las principales moléculas observadas en el espectro de masas del extracto. Las barras negras y blancas corresponden a los valores de intensidades antes y después, respectivamente, de la incubación con CPA. Los resultados son la media de diez experimentos independientes.

4. Conclusiones.

La caracterización de interacciones proteína-proteína y proteína-ligando es uno de los grandes retos de la biología celular y la biología de sistemas. La aproximación de IF MALDI-TOF MS facilita el estudio de este tipo de interacciones de forma fácil, rápida y usando poca cantidad de muestra, como lo indican los numerosos estudios en los que ya se ha aplicado esta técnica por parte de distintos investigadores con resultados muy positivos (25, 26, 32, 35-41). Por un lado la técnica permite visualizar interacciones proteína-ligando de forma directa por MS, por otro facilita la realización de cribados rápidos en múltiples extractos para identificar la presencia de los ligandos de interés, permitiendo establecer relaciones de alta sensibilidad entre medidas de actividad (p.e. enzimática) en dichos extractos y sus espectros de masas. O sea, genera información de proteómica e interactómica funcionales. Actualmente, entre otros, se está trabajando en la me-

jora de la sistemática del pre-fraccionamiento de los extractos, en la inmovilización de la proteína diana y en el análisis masivo y fiable de actividades y de espectros de MS, puntos críticos en el protocolo de la técnica, cuya estandarización permitirá aplicarla a estrategias de alto rendimiento.

Agradecimientos. Se agradece sinceramente la contribución y ayuda de distintos grupos que han colaborado en las distintas etapas del desarrollo de las aproximaciones aquí recogidas, y en especial al grupo de las Profs. María de los Ángeles Chavez y Maday Alonso del Ribero, del Centro de Proteínas de la Facultad de Biología de la Universidad de La Habana (Cuba) que proporcionaron muchos de los materiales y extractos biológicos aquí analizados, así como su valiosa colaboración científica para sus análisis e interpretaciones, como se indica en diversas publicaciones científicas previas conjuntas. También a diversos grupos europeos que han participado en el desarrollo de la IF MALDI-TOF MS, igualmente recogidos en publicaciones anteriores, así como a antiguos miembros de nuestro grupo de investigación en ello involucrados (Drs. J.Villanueva, O.Yanes, L.Sanglas, S.Trejo, J.L.Arolas…etc). Igualmente, al Plan Nacional de Investigación de I+D+i (área de biotecnología) por diversos proyectos que han soportado dicha investigación, asi como al Servicio de Proteómica y de Biología Estructural de la UAB en donde se han realizado muchos de dichos estudios.

REFERENCIAS.

1. **Feng X, Liu X, Luo Q, Liu BF. Mass spectrometry in systems biology: an overview.** Mass Spectrom Rev. 2008 Nov-Dec;27(6):635-60.

2. **Bader S, Kuhner S, Gavin AC. Interaction networks for systems biology.** FEBS Lett. 2008 Apr 9;582(8):1220-4.

3. **Guan H, Kiss-Toth E. Advanced Technologies for Studies on Protein Interactomes.** Protein – Protein Interaction: Springer Berlin / Heidelberg; 2008. p. 1-24.

4. **Sinz A. Investigation of protein-ligand interactions by mass spectrometry.** ChemMedChem. 2007 Apr;2(4): 425-31.

5. **Gavin A-C, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes.** Nature. 2002;415(6868):141-7.

6. **Konermann L, Tong X, Pan Y. Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches.** Journal of Mass Spectrometry. 2008;43(8):1021-36.

7. **Ganem B, Li YT, Henion JD. Detection of noncovalent receptor-ligand complexes by mass spectrometry.** Journal of the American Chemical Society. 1991;113(16):6294-6.

8. **Karas M, Bahr U, Ingendoh A, Nordhoff E, Stahl B, Strupat K, et al. Principles and applications of matrix-assisted UV-laser desorption/ionization mass spectrometry.** Analytica Chimica Acta. 1990;241(2):175-85.

9. **Karas M, Bahr U, Ingendoh A, Hillenkamp F. Laser Desorption/Ionization Mass Spectrometry of Proteins of Mass 100 000 to 250 000 Dalton.** Angewandte Chemie International Edition in English. 1989;28(6):760-1.

10. **Rosinke B, Strupat K, Hillenkamp F, Rosenbusch J, Dencher N, Krüger U, et al. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) of membrane proteins and non-covalent complexes.** Journal of Mass Spectrometry. 1995;30(10):1462-8.

11. **Cohen LRH, Strupat K, Hillenkamp F. Analysis of quaternary protein ensembles by matrix assisted laser desorption/ionization mass spectrometry.** Journal of the American Society for Mass Spectrometry. 1997;8(10):1046-52.

12. **Song F. A study of noncovalent protein complexes by matrix-assisted laser desorption/ionization. J Am Soc Mass Spectrom.** 2007 Jul;18(7):1286-90.

13. **Strupat K, Sagi D, Bonisch H, Schafer G, Peter-Katalinic J. Oligomerization and substrate binding studies of the adenylate kinase from by matrix-assisted laser desorption/ionization mass spectrometry.** Analyst. 2000;125(4):563-7.

14. **Woods AS, Buchsbaum JC, Worrall TA, Berg JM, Cotter RJ. Matrix-Assisted Laser Desorption/Ionization of Noncovalently Bound Compounds.** Analytical Chemistry. 1995;67(24):4462-5.

15. **Villanueva J, Yanes O, Querol E, Serrano L, Aviles FX. Identification of Protein Ligands in Complex Biological Samples Using Intensity-Fading MALDI-TOF Mass Spectrometry.** Analytical Chemistry. 2003;75(14):3385-95.

16. **Villanueva J, Yanes O, Querol E, Serrano L, Aviles FX. Analysis of protein-protein interactions in complex biological samples by MALDI TOF MS. Feasibility and use of hte intensity-fading (IF-) approach.** En: kamp R, Calvete JJ, Choli-Papadopoulou T, editors. Methods in proteome and protein analysis. Heidelberg: Springer; 2004. p. 183-202.

17. Yanes O, Aviles FX, Wenzel R, Nazabal A, Zenobi R, Calvete JJ. Proteomic profiling of a snake venom using high mass detection MALDI-TOF mass spectrometry. J Am Soc Mass Spectrom. 2007 Apr;18(4):600-6.

18. Yanes Ó, Villanueva J, Querol E, Aviles FX. Intensity-fading MALDI-TOF-MS: novel screening for ligand binding and drug discovery. Drug Discovery Today: TARGETS. 2004;3(2, Supplement 1):23-30.

19. Yanes O, Villanueva J, Querol E, Aviles FX. Functional screening of serine protease inhibitors in the medical leech Hirudo medicinalis monitored by intensity fading MALDI-TOF MS. Mol Cell Proteomics. 2005 Oct;4(10):1602-13.

20. Yanes O, Aviles FX, Roepstorff P, Jorgensen TJ. Exploring the "intensity fading" phenomenon in the study of noncovalent interactions by MALDI-TOF mass spectrometry. J Am Soc Mass Spectrom. 2007 Feb;18(2):359-67.

21. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Analytical Chemistry. 1988;60(20):2299-301.

22. Schlosser G, Pocsfalvi G, Huszár E, Malorni A, Hudecz F. MALDI-TOF mass spectrometry of a combinatorial peptide library: effect of matrix composition on signal suppression. Journal of Mass Spectrometry. 2005;40(12):1590-4.

23. Delfin J, González Y, Diaz J, Chavez M. Proteinase inhibitor from Stichodactyla helianthus: purification, characterization and immobilization. Arch Med Res. 1994 1994;25(2):199-204.

24. Marino-Buslje C, Venhudová G, Molina MA, Oliva B, Jorba X, Canals F, et al. Contribution of C-tail residues of potato carboxypeptidase inhibitor to the binding to carboxypeptidase A. European Journal of Biochemistry. 2000;267(5):1502-9.

25. Wang Z, Yu X, Cui M, Liu Z, Song F, Liu S. Investigation of calmodulin-Peptide interactions using matrix-assisted laser desorption/ionization mass spectrometry. J Am Soc Mass Spectrom. 2009;20(4):576-83.

26. Liu W, Liu S, Li H, Song F, Liu Z, Liu S. Binding of alpha 1-acid clycoprotein with aconitum alkaloids: an investigation using an intensity fading matrix-assisted laser desorption/ionization Fourier Transform mass spectrometry method. Rapid Commun Mass Spectrom. 2011 Apr;25(7):973-8.

27. Antuch W, Berndt KD, Chávez MA, Delfín J and Wuthrich K. The NMR solution structure of a Kunitz- type proteinase inhibitor from the sea anemone Stichodactyla helianthus. Eur J. Biochem. 1993;212:675-684.

28. Castañeda O, Sotolongo V, Amor AM, Stocklin R, Anderson A, Harvey AL and Karlsson E. Characterization of a potassium channel blocker from Caribbean sea anemone Stichodactyla helianthus. Toxicon. 1995;33:603-613.

29. Álvarez C, Tejuca M, Morera V, Besada V, Pazos F, Lanio ME and Padrón G. Novel primary structure of sticholysin and its interactions with membranes. Toxicon. 1996;34:301-301.

30. Guisan JM and Blanco RM. Stabilization of trypsin by multiple-point attachment to aldehyde-agarose gels. Ann N Y Acad Sci. 1987;501:67-72.

31. Erlanger BF, Kokowsky N and Cohen W. The preparation and properties of two new chromogenic substrates of trypsin. Arch. Biochem. Biophys. 1961; 95:271-278.

32. González Y, Tanaka YS, Hirata IY, Alonso del Rivero M, Oliva MLV, Araujo MS and Chávez MA. Purification and partial preliminar characterization of human neutrophile elastase from the marine snail Cenchritis muricatus (mollusca). Comp Biochem Physiol A Mol Integr Physiol. 2007;146(4):506-13.

33. González Y, Pons T, Gil J, Besada V, Alonso-del-Rivero, M, Tanaka AS, Araujo MS and Chávez MA. Characterization and comparative 3D modeling of CmPI-II, a novel 'non-classical' Kazal-type inhibitor from the marine snail Cenchritis muricatus (Mollusca). Biological Chemistry. 2007;388:1183–1194

34. Mock WL, Liu Y, Stanford DJ. Arazoformyl Peptide Surrogates as Spectrophotometric Kinetic Assay Substrates for Carboxypeptidase A. Analytical Biochemistry. 1996;239(2):218-

35. Alonso-del-Rivero M, Trejo SA, Rodriguez de la Vega M, Gonzalez Y, Bronsoms S, Canals F, et al. A novel metallocarboxypeptidase-like enzyme from the marine annelid Sabellastarte magnifica-a step into the invertebrate world of proteases. Febs J. 2009 Sep;276(17):4875-90.

36. Sanglas L, Aviles FX, Huber R, Gomis-Ruth FX, Arolas JL. Mammalian metallopeptidase inhibition at the defense barrier of Ascaris parasite. Proc Natl Acad Sci U S A. 2009 Feb 10;106(6):1743-7.

37. Gimenez-Oya V, Villacanas O, Fernandez-Busquets X, Rubio-Martinez J, Imperial S. Mimicking direct protein-protein and solvent-mediated interactions in the CDP-methylerythritol kinase homodimer: a pharmacophore-directed virtual screening approach. J Mol Model. 2009 Aug;15(8):997-1007.

38. Mishra M, Tamhane VA, Khandelwal N, Kulkarni MJ, Gupta VS, Giri AP. Interaction of recombinant CanPIs with Helicoverpa armigera gut proteases reveals their processing patterns, stability and efficiency. Proteomics. 2010 Aug;10(15):2845-57.

39. Shabab M, Kulkarni MJ, Khan MI. Study of papain-cystatin interaction by intensity fading MALDI-TOF-MS. Protein J. 2008 Jan;27(1):7-12.

40. Mireuta M, Hancock MA, Pollak M. Binding Between insulin-like growth factor 1 and insulin-like growth factor binding protein 3 is not influenced by clucose or 2-deoxy-D-glucose. J Biol Chem. 2011 May;286(19):16567-73

41. Ma L, Wang Z, Liu S, Song F, Liu Z, Liu S. Screening calmodulin-binding ligands using intensity-fading matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun Mass Spectrom. 2013 Jul;1527(13):1527-34.

Microarrays de proteínas

32

José Ignacio Casal Álvarez[1]

[1]Grupo de Proteómica Funcional. Centro de Investigaciones Biológicas (CIB-CSIC). , Madrid, Spain.

1. Introducción.

Un microarray[1] de proteínas es un pequeño dispositivo, del tamaño de un portaobjetos de microscopio, donde se han impreso cientos o miles de proteínas diferentes (incluyendo anticuerpos, dominios de proteínas, péptidos…), generalmente sobre una membrana de nitrocelulosa. Los microarrays de proteínas surgen como una evolución de los inmunoensayos miniaturizados (1) y constituyen la alternativa proteómica a los microarrays de DNA, desarrollados a partir de los años 90. Aunque se describieron inicialmente en el año 99 (2), solo recientemente se ha popularizado y extendido su uso debido a las mejoras técnicas introducidas tanto en la preparación como en el análisis de datos, así como a una mayor disponibilidad comercial. Los microarrays de proteínas representan una estrategia innovadora y versátil para, entre otras aplicaciones, i) la identificación de biomarcadores a nivel de proteoma completo (3), ii) estudiar interacciones proteína-proteína a gran escala (4) y iii) participar en el descubrimiento de nuevos fármacos (5). Las principales ventajas de los chips de proteínas sobre las técnicas tradicionales como el ELISA son un bajo consumo de muestra, la posibilidad de realizar múltiples ensayos simultáneamente (multiplex) y una fácil adaptación a la automatización.

El mayor obstáculo al rápido desarrollo de esta tecnología ha sido precisamente la preparación de su contenido impreso debido a las dificultades en la producción de proteínas y/o anticuerpos. Las diversas características moleculares de las proteínas hacen que la impresión y construcción de microarrays de proteínas sea mucho más difícil que la de microarrays de DNA. Actualmente no existe ninguna técnica similar a la PCR, que permita amplificar rápida y sencillamente una gran cantidad de proteína de la misma manera que ocurre con las moléculas de DNA. Los sistemas de producción y expresión de proteínas recombinantes son difíciles de automatizar y siempre impredecibles en cuanto a rendimiento. Además, las proteínas suelen ser más inestables que el DNA, lo que disminuye la vida media de los microarrays de proteínas almacenados. Respecto a su uso y en contraste con las sencillas técnicas de hibridación de ácidos nucleícos, la gran variedad química de las proteínas complica enormemente los estudios de reactividad, funcionalidad e interacción entre diversas proteínas. A continuación haremos un breve repaso de los tipos y aplicaciones más importantes de los microarrays de proteínas.

[1] Aunque existe la traducción castellana de micromatriz para "microarray" he preferido mantener la version inglesa dado su uso mucho mas frecuente.

2. Tipos de Microarrays de proteínas.

En este trabajo analizaremos los tres tipos básicos de microarrays: microarrays de anticuerpos, microarrays de fase reversa y microarrays de proteínas recombinantes (Figura 1).

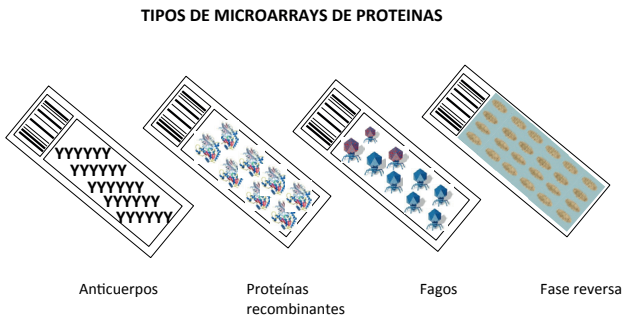


Figura 1. Tipos de microarrays de proteínas.

2.1. Microarrays de anticuerpos.

Los microarrays de anticuerpos pueden contener varios cientos o miles de anticuerpos impresos y se utilizan actualmente en la identificación de biomarcadores en cáncer y otras patologías, en la clasificación y/o estratificación de pacientes, así como en la caracterización de vías de señalización celular (6). Presentan como ventajas la gran estabilidad, selectividad y alta afinidad de los anticuerpos, así como el gran número de anticuerpos disponibles comercialmente, que aumenta día a día (7-8). Pese a ello, no todos lo anticuerpos son funcionales en el formato de microarray; esto es, pueden perder la capacidad de unión al antígeno cuando se fijan a una superficie sólida de vidrio o de nitrocelulosa. Anticuerpos teóricamente monoespecíficos pueden dar lugar a múltiples reactividades en un microarray cuando se les enfrenta a mezclas complejas de proteínas. De hecho, los microarrays de proteínas se utilizan también para el estudio de la especificidad de los anticuerpos, como en el caso del proyecto "Human Protein Atlas" (9). Aunque inicialmente se utilizaron exclusivamente anticuerpos completos, en los últimos años se han desarrollado arrays conteniendo fragmentos de anticuerpos recombinantes tipo scFv (Fragmento variable de cadena sencilla) (10-11). El formato scFv permite la producción de múltiples anticuerpos a partir de librerías de fagos y facilita su posterior impresión, dada la estructura similar de las moléculas de scFv.

Respecto a la detección, existen dos tipos de formatos basados bien i) en marcaje directo de las muestras con fluorocromo ó ii) en técnicas de "sándwich" con doble anticuerpo, uno de ellos marcado. En el marcaje directo, los extractos proteicos a analizar se marcan con uno o dos fluorocromos (Cy3 y Cy5 o Alexas 555 y 647), utilizando un procedimiento similar al descrito para DIGE y se incuban directamente con el microarray. El marcaje con dos fluorocromos se puede utilizar para el análisis de la expresión diferencial de proteínas en diferentes condiciones (p.ej. para el análisis comparativo de tejidos tumorales frente a tejidos normales) (Figura 2). Este tipo de ensayo competitivo presenta varias ventajas: 1) Se usan las dos muestras a comparar sobre el mismo array, 2) las condiciones de incubación son idénticas para las dos muestras y 3) presenta mayor linealidad y rango dinámico. Como problema se puede mencionar el potencial sesgo en el marcaje debido a la diferente eficiencia y estabilidad de los dos fluorocromos. Para corregir este sesgo deben realizarse siempre experimentos de alternancia de marcaje ("dye swapping") (12). Los microarrays de anticuerpos presentan ciertas ventajas frente a la electroforesis 2D-DIGE en la identificación de proteínas diferencialmente expresadas en extractos proteicos: i) No es necesario identificar las proteínas detectadas (dado que se utilizan anticuerpos de especificidad conocida); ii)se pueden detectar proteínas de baja abundancia y iii) se pueden detectar directamente modificaciones post-transduccionales (8).

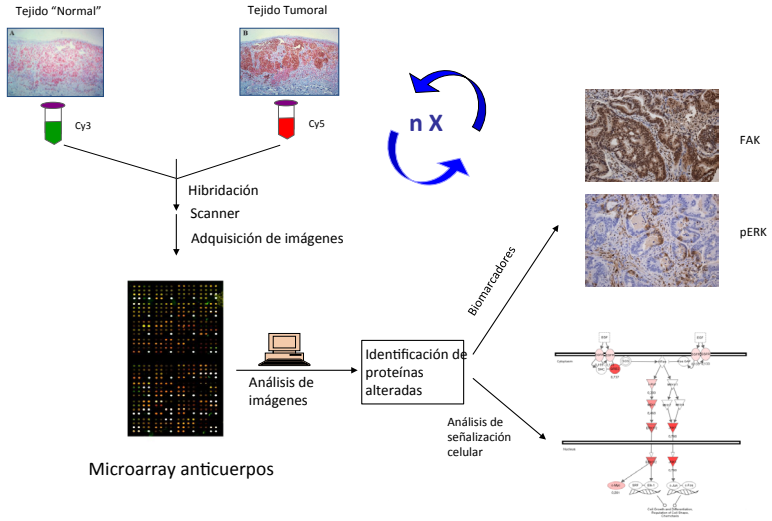


Figura 2. Aplicación de los microarray de anticuerpos en la identificación de biomarcadores y análisis de vías de señalización. El microarray Panorama (Sigma) fue hibridado con extractos proteicos correspondientes a tejidos pareados tumoral y normal de un paciente con cáncer colorrectal. Los lisados fueron marcados con Cy5 y Cy3 respectivamente. Tras escaneo, las dos imágenes fueron cuantificadas con GenePix. Las proteínas identificadas fueron validadas por western blot o inmunohistoquímica sobre tejidos o líneas celulares tumorales. Los perfiles de expresión fueron analizados mediante *Ingenuity Pathway Analysis* para comprobar el efecto a nivel de señalización global.

Los ensayos de detección basados en un ELISA tipo "sándwich" presentan una mayor especificidad (debido al uso de dos anticuerpos) y menor ruido de fondo, lo que les hace más apropiados para cuantificación absoluta. Sin embargo, son más difíciles de estandarizar y suelen estar limitados a unas pocas decenas de anticuerpos. Los arrays comerciales disponibles para la cuantificación de citoquinas (por ejemplo RayBiotech) pertenecen a este tipo.

2.2. Microarrays de fase reversa ó de lisados celulares.

Los microarrays de proteínas de fase reversa se preparan imprimiendo decenas o cientos de lisados celulares completos sobre un porta recubierto de nitrocelulosa (13-14). El nombre "fase reversa" viene dado por el hecho que se inmoviliza el analito (antígeno) en vez del anticuerpo de captura (15-16). Se utilizan para estudiar la presencia y los niveles de proteínas de interés en lisados celulares problema mediante anticuerpos específicos. Preparando copias seriadas de estos arrays es posible estudiar la evolución y alteraciones de múltiples proteínas simultáneamente. Originalmente, los microarrays de fase reversa se diseñaron para obtener perfiles de señalización celular en muestras de tejidos microdiseccionados a partir de tumores. Se pueden utilizar para el seguimiento de ciertas proteínas durante la progresión del cáncer u otras patologias (16), la validación de biomarcadores (17) o la identificación de las dianas de compuestos bioactivos con mecanismos de acción desconocidos (18). Una de las mayores limitaciones de uso de este tipo de microarrays viene dado por la falta de especificidad de los anticuerpos disponibles comercialmente, incluso cuando el anticuerpo produce una sola banda en western blot (18). Dado que la concentración local de las proteínas en el *spot* del array puede ser muy alta, pueden ocurrir falsas interacciones que deben ser controladas adecuadamente (19).

Se han usado microarrays de fase reversa para clasificar la composición proteica y vías de señalización en diferentes tipos de cáncer usando la colección de líneas celulares NCI-60 (20). Las líneas celulares fueron lisadas, impresas y testadas con más de 200 anticuerpos, para seleccionar finalmente 52 que dieron una señal adecuada.

Una variante de esta estrategia consiste en la preparación de microarrays de proteínas "naturales" e implica el fraccionamiento previo del extracto celular proteico mediante isoelectroenfoque

y cromatografía liquida como paso previo a la impresión. Posteriormente, las diferentes fracciones son impresas en formato de microarray e incubadas con los anticuerpos de interés o con los sueros de pacientes para la detección de autoanticuerpos que pueden ser usados como biomarcadores (21-22).

2.3. Microarrays de proteínas recombinantes.

Estos microarrays se basan en la impresión de cientos o miles de proteínas recombinantes, dominios de proteínas, péptidos o fagos sobre un portaobjetos en posiciones definidas. Se pueden utilizar para i) la búsqueda en suero ó plasma de autoanticuerpos específicos para una determinada enfermedad, ii) para el estudio de las propiedades bioquímicas y funcionalidad de las proteínas impresas (23), iii) para el estudio de las interacciones con otras proteínas (interactómica) y otras moléculas y iv) para el estudio de la actividad enzimática y la especificidad de sustrato de un determinado grupo de enzimas (p.ej. quinasas) (23). Las proteínas se deben imprimir en condiciones adecuadas para mantener su conformación, integridad y actividad. En algunos casos, para mantener dicha integridad y permanecer activa y funcional, la proteína se tiene que imprimir formando complejos con otras proteínas de su complejo multiproteico (24). En cualquier caso, la producción de proteínas en una buena cantidad y con la debida pureza en sistemas de expresión heterólogos como *E. coli*, levadura, baculovirus, etc. es difícil, laboriosa, lenta y costosa. Tampoco es equivalente, en cuanto a calidad y conformación de la proteína resultante, la expresión en *E. coli* que la expresión en sistemas eucarióticos como baculovirus, en cualquier caso los resultados obtenidos a partir de arrays de proteínas impresos con proteínas humanas expresadas en *E. coli* deben tomarse con cierta cautela por las posibles alteraciones en la conformación de la proteína (25). Como alternativa práctica puede resultar más sencillo estudiar algunos dominios de las proteínas que las proteínas completas, dado que los pequeños dominios son generalmente más sencillos de producir y purificar.

Estos arrays se pueden producir de tres maneras distintas: i) por expresión independiente y unión directa de la proteína al soporte, que generalmente suele ser de nitrocelulosa, ii) expresión y unión mediante *tags* de fusión (GST, 6xHis, etc.) y iii) mediante la tecnología NAPPAS (*Nucleic Acid Programmable Protein ArrayS*) realizando la síntesis *in situ* de la proteína y su captura posterior directamente en el microarray (26). Los microarrays NAPPAS son una alternativa interesante basada en la impresión de cDNAs sobre la superficie de los microarrays y la expresión de proteínas *in situ* utilizando un sistema de expresión basado en reticulocitos de conejo. La molécula naciente, marcada con GST, es capturada por un anticuerpo impreso adyacente anti-GST, de manera que quede retenida sobre la superficie del porta. Esta estrategia es prometedora en cuanto reúne en un solo porta la expresión, purificación y almacenamiento de las proteínas. Facilita por tanto la preparación de arrays a medida, a demanda del usuario y sólo con las proteínas deseadas. Por otro lado, la estabilidad del DNA facilita su preparación y conservación hasta su utilización. Sin embargo, la síntesis *in situ* puede ser problemática y poco reproducible entre arrays, incluso del mismo lote. Los niveles de expresión de proteínas pueden fluctuar significativamente haciendo difícil la comparación inter-arrays. Se necesitan asimismo extensas colecciones de clones de cDNA con el formato adecuado para la expresión de la proteína, que no son fácilmente accesibles a cualquier laboratorio. Todo ello ha contribuido a que su uso sea muy limitado y restringido a unos pocos laboratorios especializados.

Por último, el uso de bibliotecas de fagos T7 expresando cDNAs e impresas en formato *microarray* constituye otra alternativa para el descubrimiento de nuevos biomarcadores, ya que no hay sesgo hacia un determinado grupo de proteínas (27). Los péptidos codificados por los fragmentos de cDNA se exponen sobre la superficie de los fagos fusionado al extremo C-terminal de la proteína 10B de la cápsida. Los fagos desplegando péptidos específicos de cáncer sobre su superficie son seleccionados e impresos en portas de nitrocelulosa (Figura 3). Teóricamente, la tecnología permite la impresión de miles de péptidos relacionados con cáncer. Sin embargo, la identificación de los péptidos respectivos es difícil, dado que la mayoría de los péptidos

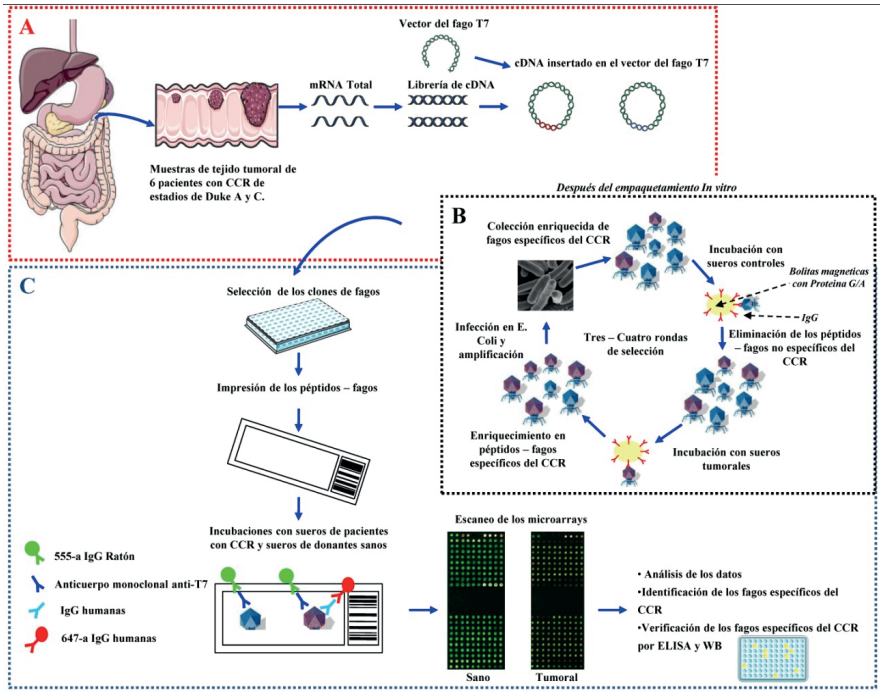


Figura 3. Esquema del proceso seguido para la obtención de biomarcadores potenciales de cáncer colorrectal (CCR) usando las técnicas de despliegue en fagos y *microarrays* de proteínas. A, Construcción de las librerías de fagos T7 a partir de RNA de tejido de CCR. B, Rondas de selección para enriquecer las colecciones en fagos específicos de IgGs de pacientes con CCR. C, Impresión e hibridación de los *microarrays* de fagos T7. Una vez obtenidos los datos, los resultados requieren un paso de validación mediante ELISA o *immunoblotting*.

desplegados por los fagos no presentan homología con secuencias de proteínas humanas conocidas, siendo la mayoría de ellos mimotopos (el péptido imita un epítopo conformacional de una proteína) (28). Por ello, las librerías de fagos T7 pueden ser reemplazadas por librerías combinatoriales de péptidos con resultados similares (29).

3. Metodología de impresión, hibridación y análisis de datos de Microarrays de proteínas.

A título de ejemplo se describe la metodología utilizada en nuestro laboratorio en el caso de microarrays de proteínas y fagos recombinantes para el cribado de sueros humanos en la detección de biomarcadores tumorales en cáncer colorrectal (30-31). Se discutirán asimismo los métodos más generales para la validación de dichos biomarcadores.

3.1 Impresión.

Una vez producidas, las proteínas ó fagos se transfieren a placas de 384 pocillos en forma de V (Genetix) y se imprimen a diferentes diluciones (sin diluir, 1:2, 1:3, 1:4, 1:5) en PBS o en PBS conteniendo 0.1% Tween 20 (PBST). La impresión se realiza habitualmente sobre portas recubiertos de nitrocelulosa (por ejemplo, FAST slides (Whatman)). La nitrocelulosa presenta una gran capacidad para la unión irreversible de proteínas sin alterar la estructura de las mismas. El ruido de fondo es relativamente bajo y se puede utilizar tanto para detección cromogénica como fluorescente. En nuestro laboratorio se han utilizado dos robots diferentes para hacer

las pruebas de impresión: el robot Microgrid II (BioRobotics) con 4 puntas sólidas (Arraylt) y el robot Omnigrid (GeneMachines) con 4 puntas con reservorio (TeleChem). Las puntas sólidas permiten un único contacto de la punta con el portaobjetos antes de retornar a la placa de muestra a recargarse. Las puntas con reservorio permiten imprimir hasta 100 puntos seguidos antes de volver a recargarse. El Microgrid II dispone de un lavador con flujo continuo de agua para lavar las puntas entre cada muestra y el Omnigrid realiza un paso de sonicación adicional al lavado entre diferentes muestras. Las muestras se imprimen con un diámetro de 150-750 µm, utilizando 5-30 nl de lisado por spot. Las muestras se imprimen por duplicado o triplicado. Es imprescindible un control de la temperatura y de la humedad durante la impresión. Todas las impresiones deben realizarse a una temperatura de 18-20ºC y una humedad relativa del 50%. Tras acabar la impresión, los portas de nitrocelulosa se dejan secar en el interior del robot un mínimo de 3h a una humedad relativa del 75%.

3.2 Hibridación.

Los *microarrays* se equilibran a 4ºC durante 15 min y se bloquean con 50 mL de tampón de bloqueo en tubos Falcón de 50 ml (BD) durante 1 h a 4ºC con agitación suave (en una noria a 40 rpm). Entre 150 µL y 2 ml de suero (diluido 1:50 a 1:500 en tampón de bloqueo) se aplican sobre la superficie del *array* y se incuba durante 90 min a 4ºC. Los *arrays* se lavan 5 veces durante 5 min con 50 mL de tampón de incubación. Los autoanticuerpos del suero unidos a las proteínas del *array* se pueden detectar utilizando un anticuerpo anti-IgG (H+L) humana marcado con AlexaFluor 647 (diluido 1:2000 en tampón de incubación) a 4ºC durante 90 min. Los *arrays* se lavaron 5 veces 5 min con 50 mL de PBST y se secaron por centrifugación a 1000 rpm durante 1 min. Finalmente, los portas se escanean usando láseres de 635 nm y 532 nm (Axon Laboratories). Para comprobar la uniformidad y la cantidad de proteína impresa en el array, los ProtoArrays (Invitrogen) se incubaron con un anticuerpo monoclonal anti-GST seguido de un anticuerpo anti-IgG de ratón marcado con AlexaFluor 555 ó, directamente, con el anticuerpo anti-IgG (H+L) humana marcado con AlexaFluor 647 para determinar los niveles de ruido en el ensayo. Tras escanear las portas se obtienen dos archivos de imagen en formato *.tif para cada *array,* que corresponden a la imagen escaneada con el láser 532 (verde) y a la imagen escaneada con el láser 645 (rojo), respectivamente. Antes de cuantificar las imágenes, es necesario crear un archivo con el formato *.gal con la información correspondiente al contenido y a la posición de cada punto del *array*. En el caso de los ProtoArrays, el fabricante nos proporciona este archivo *.gal que varía según el lote de *microarrays*.

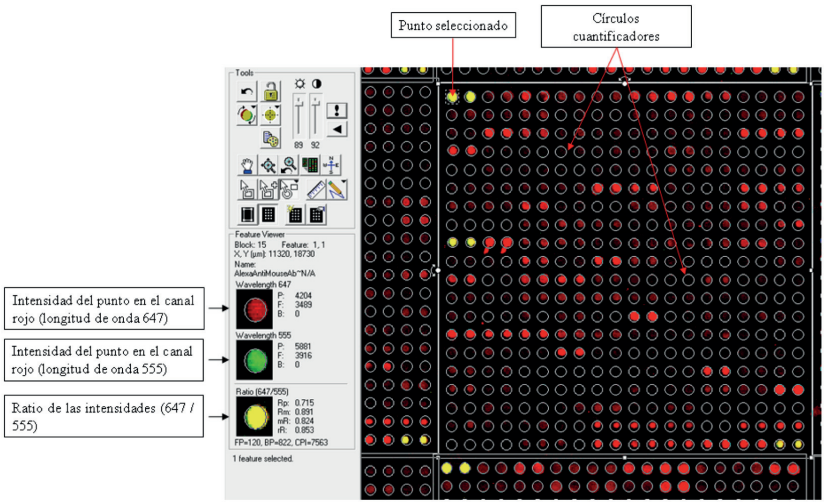


Figura 4. Cuantificación de las imágenes con el programa GenePix. Adaptado de Tesis Doctoral de I. Babel (2010).

La cuantificación de las imágenes se realiza con el programa GenePix introduciendo los 3 archivos (los 2 archivos *.tif y el archivo *.gal). El archivo *.gal permite al programa GenePix crear una plantilla con el nombre de cada punto. Posteriormente, para cuantificar cada punto de la imagen se ajusta cada círculo cuantificador alrededor de cada punto (proteína), lo que permite al programa dar un valor de la intensidad de cada pixel de la superficie del círculo para los 2 canales (rojo y verde) (Figura 4).

Una vez cuantificada la imagen, los resultados del análisis se obtienen en un archivo *.gpr donde encontramos los valores de la media de la intensidad de todos los píxeles que componen un punto para cada uno de los dos canales al igual que los valores correspondientes al ruido de fondo para cada punto y el ratio de las intensidades de los dos canales (647/555) para cada punto.

Para la cuantificación de los *microarrays* de fagos fue necesario generar el archivo *.gal. Este archivo se generó utilizando el programa del robot de impresión a partir del archivo *.txt que se introduce antes de imprimir los *arrays*. El archivo *.gal generado tiene el formato que se muestra en la Figura 5.

Block	Row	Column	ID	Name
1	1	1	1	555
1	1	2	1	555
.	.	.	.	.
.	.	.	.	.
4	5	7	1453	C415 2-D9
4	5	8	1454	C415 2-D9
.	.	.	.	.
.	.	.	.	.
12	10	9	4800	Vacio
12	10	10	4800	Vacio

Figura 5. Ejemplo de un archivo *.gal generado después de la impresión por el programa del robot de impresión. *Block,* es el número del bloque donde se encuentra el punto. *Row,* es el número de fila dentro del bloque donde se encuentra el punto. *Column,* es el número de columna dentro del bloque donde se encuentra el punto. *ID*, es el número de la muestra impresa en el punto. *Name,* es el nombre de la muestra impresa en cada punto. El nombre de la muestra de fago T7 se compone del nombre de la colección del fago (B415, C415, H415, HP415, B 10-3,...) seguido del número de la placa de donde viene la muestra seguido del número de pocillo de la placa.

3.3. Tratamiento estadístico de los datos de *microarrays.*

A medida que aumenta la complejidad de los microarrays aumenta la dificultad del análisis. Actualmente existen numerosos métodos bioestadísticos disponibles para analizar estos datos. En las aplicaciones para diagnóstico y pronóstico, se pueden agrupar en métodos supervisados y no supervisados. Los métodos no supervisados son útiles para la generación de hipótesis, donde las proteínas se categorizan exclusivamente por patrón de expresión e incluyen agrupamiento jerárquico, agrupamiento "K-means" y análisis de componente principal. Las técnicas supervisadas requieren algún conocimiento previo del tipo y la clasificación de las muestras e incluye ANOVA, t-test, Wilcoxon y el test de Mann-Whitney. En esta revisión no es posible por razones de espacio profundizar en la descripción de estos métodos y nos limitaremos a mencionar las que se han utilizado en cada ejemplo.

3.3.1. Análisis de microarrays de fagos T7.

Los datos obtenidos a partir de los microarrays de fagos se pueden normalizar y procesar mediante la interfaz Asterias (http://asterias.bioinfo.cnio.es/) para el uso de los paquetes informáticos Limma y marrayNorm de Bioconductor. Los archivos de cuantificación *.gpr se convierten en archivos *.txt utilizando el programa Microsoft Office Excel para cargarlos en la aplicación http://dnmad.bioinfo.cnio.es/. Tras introducir el número de columnas y filas que contiene el array principal y los subarrays se aplica una corrección de ruido de fondo y la normalización global de Loess. El archivo así generado se introduce en la aplicación http://prep.bioinfo.cnio.es/. Los datos se procesaron para filtrar los valores que faltaban o los puntos con una varianza demasiada alta. Además, se combinaron los duplicados para obtener un valor único para cada fago transformado en logaritmo en base 2. En el caso de ausencia de valores para un punto dado en más de un 30% de los arrays, dicho punto se elimina del análisis. En caso contrario, se introducen los valores que faltan usando el algoritmo KNNimpute. Finalmente, se obtiene un archivo *.txt que contiene los valores procesados para cada fago para cada microarray (31).

Los grupos de pacientes con cáncer e individuos sanos se pueden comparar realizando un t-test con el programa Pomelo II (http://pomelo2.bioinfo.cnio.es/) donde los p-valores se obtuvieron utilizando 200.000 permutaciones. El programa Pomelo II genera un gráfico que muestra los fagos con un valor de tasa de falsos resultados positivos (FDR) y un *p*-valor no ajustado a partir del archivo obtenido previamente y de un archivo que contiene la información de la clase de cada *microarrays* (a cada *array* se le asigna un valor 0 ó 1 que indica que pertenece al grupo normal o al grupo tumoral, respectivamente).

3.3.2. Análisis de Protoarrays.

En el caso de los Protoarrays (30), los resultados de la cuantificación de las imágenes se analizan con el ProtoArray *Prospector Analyser* (Invitrogen), que permite hacer un análisis de los perfiles de la respuesta inmune obtenidos. El algoritmo que usa este programa se basa en el principio de la desigualdad de Chebyshev (32). Tras cargar los archivos *.gpr de cada *array* en el programa, se normalizan los datos usando la normalización por cuantiles.

El algoritmo compara la señal de cada proteína a la señal de los controles negativos en el *array* y asigna un CI (intervalo de confianza) para cada *p*-valor (*p*-valor del principio de la desigualdad de Chebyshev) para cada proteína. El software identifica las señales significativas llamadas "hits", como positivas por encima del punto de corte que corresponde al ruido de fondo y calcula unos valores Z que reflejan la intensidad de la señal en comparación con todas las proteínas del *array*. El programa compara los grupos tumoral y normal e identifica las proteínas que tienen una señal aumentada en cada uno de los dos grupos calculando el *p*-valor para cada proteína basándose en la hipótesis de que no hay un aumento en la media global de la señal en un grupo comparado con el otro. Finalmente, calcula la prevalencia de las proteínas en cada grupo. Esto es, para cada proteína calcula el porcentaje de *arrays* tumorales y normales que tienen una señal mayor al punto de corte y pueden considerarse un "hit".

El "Multi Experiment Viewer" (MeV) (http://www.tm4.org/mev/) (33-34) es una aplicación gratuita para el análisis, la visualización y la extracción de datos de experimentos de *microarrays* a gran escala. Incorpora algoritmos para la agrupación, la visualización, la clasificación y el análisis estadístico. Este programa se puede utilizar para obtener los agrupamientos supervisados y no supervisados con los datos de *microarrays*. Los datos se cargan en un archivo *.txt con los valores obtenidos a partir de los *arrays* normalizados y procesados. Los clústeres se obtuvieron usando la distancia métrica y la correlación de Pearson para visualizar la discriminación entre los grupos normal y tumoral.

4. Aplicaciones de los Microarrays de proteínas en proteómica.

4.1 Aplicaciones de los microarrays de anticuerpos.

En general, los arrays de anticuerpos resultan especialmente apropiados para el estudio del proteoma de fluidos (suero, plasma, orina, etc) dado que las proteínas se encuentran en forma soluble, favoreciendo la correcta unión de las mismas a los anticuerpos. Los extractos celulares y de tejidos son generalmente más complejos, más difíciles de solubilizar y pueden contener más sustancias interferentes, aunque también se han utilizado de forma satisfactoria en múltiples estudios. En el caso de los arrays de anticuerpos y muestras de plasma/suero, las proteínas del plasma deben ser precipitadas previamente y marcadas con sondas fluorescentes (Cy3 o Cy5) antes de la hibridación (7).

En un estudio pionero, se utilizaron arrays de anticuerpos para caracterizar el efecto de la irradiación UV sobre la apoptosis en células de cáncer colorrectal (35). Los autores identificaron múltiples proteínas que fueron sobreexpresadas o reprimidas. Los cambios se confirmaron posteriormente por western blot. Recientemente, en nuestro laboratorio hemos utilizado los arrays de anticuerpos "Panorama" (Sigma) para identificar proteínas diferencialmente expresadas en muestras pareadas (tumoral/normal) de cáncer colorrectal y obtener una visión global de las vías de señalización implicadas (8). Se identificó un patrón de proteínas que permite distinguir claramente mucosa tumoral de mucosa sana y que nos permite agrupar los pacientes en diversos subtipos y localizar posibles marcadores de cáncer colorrectal (CHK1, MAPK3, MDM2 y calcineurina, entre otros). Asimismo, se observó una clara sobreexpresión en diferentes componentes de las rutas de EGFR y de Wnt/β-catenina, así como una inhibición de la vía de $p14^{ARF}$. El uso de este tipo de microarrays permite la detección de proteínas presentes en vías de señalización, que normalmente se expresan a unos niveles muy bajos y son muy difíciles de detectar por técnicas proteómicas convencionales. Otra ventaja de los arrays de anticuerpos consiste en la información directa obtenida sobre los estados de fosforilación de una proteína o familia de proteínas (8, 36). La disponibilidad cada vez mayor de anticuerpos específicos facilita este uso, aunque solo existe un conocimiento parcial de los posibles estados de fosforilación de las proteínas que componen el proteoma humano. El principal factor negativo lo constituye el elevado precio de estos microarrays.

Los microarrays basados en ensayos tipo sandwich se han utilizado fundamentalmente para la cuantificación de citoquinas y quimioquinas (37). Los anticuerpos específicos se imprimen sobre la membrana, se añade la muestra y a continuación un coctel de anticuerpos específicos marcados con biotina. La detección tiene lugar mediante luminiscencia reforzada. En un trabajo pionero se describió su aplicación al estudio proteómico del fluido intersticial de cáncer de mama (38), complementando satisfactoriamente los datos obtenidos con geles 2D. Actualmente se pueden detectar y cuantificar simultáneamente más de 100 citoquinas utilizando microarrays comerciales. Los datos de citoquinas se pueden utilizar conjuntamente con otros datos proteómicos para dilucidar las interacciones entre células tumorales y estromales en el medioambiente tumoral. En un trabajo reciente, nuestro grupo analizó el efecto del factor de transcripción Snail sobre la producción de citoquinas, quimioquinas y factores de crecimiento en células tumorales de cáncer de colon (39), observándose un importante efecto de Snail en la producción de citoquinas pro-inflamatorias y pro-angiogénicas que permiten la activación del estroma tumoral (Figura 6). Estos perfiles de citoquinas pueden presentar un valor diagnóstico importante en diferentes patologías.

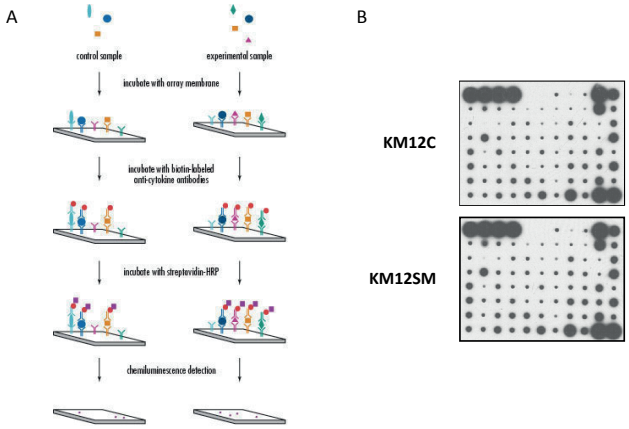


Figura 6. Esquema del proceso de hibridación y visualización en microarrays de anticuerpos específicos de citoquinas. A. Flujo de trabajo. Las muestras se añaden al microarray de anticuerpos específicos. Tras un lavado, se añade una mezcla de anticuerpos específicos marcados con biotina y posteriormente estreptavidina-peroxidasa para el revelado. La detección final se realiza por quimioluminiscencia. B. Análisis de citoquinas en medio condicionado por dos líneas celulares de cáncer de colon difiriendo en propiedades metastáticas.

4.2 Aplicaciones de los microarrays de proteínas/ dominios recombinantes.

Se han descrito múltiples aplicaciones para los microarrays de proteínas en el estudio de la funcionalidad, señalización, búsqueda de sustratos ó identificación de biomarcadores (40). Las primeras aplicaciones se describieron a nivel del proteoma completo de levadura, en arrays que contenían impresas 5800 proteínas de levadura (23) y permitieron analizar *in vitro* los sustratos o dianas para 87 quinasas (41). Más recientemente se han construido arrays conteniendo 2158 proteínas de *Arabidopsis thaliana,* que han permitido analizar los sustratos de interacción de las MAP quinasas en plantas (42). Algunas de las aplicaciones más interesantes descritas se han basado en el uso de dominios de interacción que retienen su capacidad de unión tras la impresión sobre el array. Se han usado microarrays de dominios "*coiled coil"* para estudiar la selectividad de unión con los factores de transcripción humanos bZIP (43) o dominios PDZ de ratón para determinar su interactoma (44). Estos estudios demostraron la impresionante capacidad de unión y alta selectividad de los dominios aislados. De la misma manera se han generado mapas de interacción para dominios WW en levaduras (45), o para el estudio de interacciones entre dominios SH3 y las proteínas básicas de mielina (46). Jones y col prepararon microarrays conteniendo dominios SH2 y PTB humanos e hibridaron los arrays con concentraciones diferentes de péptidos conteniendo fosfotirosina derivados de los 4 receptores humanos ErbB para un análisis cuantitativo de la interacción (47). El estudio demostró la promiscuidad de estos receptores tirosina-quinasa y su capacidad para unirse a múltiples receptores cuando se sobreexpresan en tumores. Los microarrays de proteínas se pueden usar también para determinar cinéticas enzimáticas, cribar posibles inhibidores enzimáticos en alto rendimiento y para obtener curvas de dosis-respuesta de compuestos seleccionados (48). El uso de microarrays de péptidos se ha extendido para la búsqueda de sustratos de proteínas en solución. Así, varios estudios han descrito el uso de microarrays conteniendo 1100 péptidos, sustratos de quinasas humanas, para el estudio del quinoma en el esófago de Barret (49), el efecto de COX-2 en la inhibición de la ruta Wnt en cáncer de colon (50) y otros.

Como extensión del uso de arrays de proteoma completo de levadura surge el desarrollo comercial de los Protoarrays (Invitrogen) que contienen impresas hasta 9000 proteínas recombinantes humanas expresadas en baculovirus en su última versión 5.0. Estos Protoarrays están siendo utilizados activamente para la búsqueda de biomarcadores basados en antígenos asociados a tumores (AATs), que son reconocidos por autoanticuerpos de los pacientes de cáncer o enfermedades autoinmunes (30, 32, 51). La respuesta inmune ocurre pronto durante el desarrollo tumoral y, como resultado, la respuesta de autoanticuerpos puede preceder a cualquier otra técnica en el diagnóstico del cáncer, suministrando un método efectivo para el cribado de la población y el diagnóstico pre-clínico. Utilizando Protoarrays que contenían 8000 proteínas humanas impresas hemos identificado un perfil de autoanticuerpos capaz de distinguir muestras tumorales de cáncer colorrectal de controles sanos y entre pacientes con diferentes tipos de metástasis en hígado, pulmón o medula ósea (30). Cuarenta y tres proteínas sirvieron de base para este reconocimiento diferencial con un valor $p < 0.04$, estadísticamente significativo. Basado en la prevalencia de los autoanticuerpos en el suero de los pacientes se seleccionaron 6 proteínas para un análisis en profundidad. La combinación de estas 6 proteínas: Pim-1, MAPKAPK3, STK4, SRC, FGFR4 y ACVR2B, fue capaz de predecir la presencia de cáncer con una especificidad y sensibilidad cercanas al 90%. El uso de autoanticuerpos para el diagnóstico y el pronóstico del cáncer constituye una herramienta fundamental para el descubrimiento de nuevos biomarcadores. Su aplicación en diagnóstico temprano puede ser particularmente relevante.

Existen también múltiples ejemplos de la aplicación de las librerías de fagos T7 en formato microarray a la detección de autoanticuerpos y TAAs en cáncer de pulmón, próstata, ovario y colorrectal (27-28, 31, 52-53). Se han llegado a obtener paneles detectores con una especificidad y sensibilidad cercanas al 90%. Uno de los problemas de esta estrategia es que los péptidos que utiliza para el diagnostico son mimotopos sin asignación a péptidos concretos, lo que complica la identificación de la proteína origen. Por otro lado, no permite detectar autoanticuerpos a posi-

bles modificaciones pos-transduccionales, por ejemplo glicosilaciones aberrantes presentes en las proteínas tumorales. En cualquier caso, esta aproximación puede complementar y enriquecer otras estrategias con vistas a obtener paneles diagnósticos más potentes.

4.3 Aplicaciones de los microarrays de fase reversa.

Los arrays de fase reversa se han empleado en numerosos estudios relacionados con cáncer (ver revisión (54)) así como enfermedades infecciosas. Entre otros ejemplos se han utilizado para analizar las interacciones entre proteínas y los efectos cuantitativos de RNAi combinatorial en un modelo de línea celular de cáncer de mama (55), examinar los efectos moleculares de la infección por ántrax en células de pulmón (56), estudio de los perfiles de anticuerpos tras la infección por Plasmodium (57), análisis de la interacción tumor-estroma (58), validación de la clusterina como un biomarcador potencial de carcinogénesis en sangre (59), identificación de IGFBP2 como biomarcador y diana terapéutica en glioblastoma (60), descifrado de la supresión de IL-8 en un modelo de fibrosis cística (61) y para la realización de un perfil molecular de la oxidación de grupos tiol en células de cáncer (62). En los últimos años han comenzado a utilizarse en diversas aplicaciones y ensayos clínicos, incluyendo predicción de respuesta a imatinib en pacientes de cáncer colorrectal con metástasis hepática (63), estudios de cáncer de mama refractarios a tratamiento, prevención de cáncer de mama invasivo por cloroquina, estudios de quimioterapia combinada de trastuzumab y/o lapatinib en pacientes de cáncer de mama Her2+, y aplicación en fase clínica II del tratamiento combinado de radiación, cisplatino y panitumumab en pacientes de cáncer de cabeza y cuello (54).

En cuanto a la búsqueda de biomarcadores, los microarrays de proteínas “naturales” han sido extensamente utilizados. Anticuerpos contra la L3 ubiquitin C-terminal hidrolasa fueron identificados en pacientes con cáncer de colon tras el fraccionamiento de lisados celulares en microarrays de nitrocelulosa (64). La misma aproximación se ha utilizado para detectar otros biomarcadores en cáncer de pulmón y próstata (65-66).

5. Conclusiones.

Existe actualmente un gran interés por los microarrays de proteínas. Sin embargo, se necesitan mayores esfuerzos en cuanto a producción y rendimiento para poder abaratar su coste y facilitar su uso de forma masiva en los laboratorios, como sucede con los chips de DNA. El número creciente de proteínas expresadas y disponibles, así como el interés de las empresas biotecnológicas y farmacéuticas en este campo, hace prever un aumento considerable tanto en el uso de los microarrays como en sus posibles aplicaciones.

Los microarrays de proteínas constituyen una interesante alternativa para diseccionar el proteoma, siendo particularmente relevantes en el estudio de proteínas diferencialmente expresadas con bajos niveles de expresión, estudio de las interacciones proteína-proteína a gran escala y en la búsqueda de biomarcadores basados en autoanticuerpos. Aunque muchos de los problemas técnicos relacionados con la impresión y producción de los microarrays se van solucionando en gran medida, existen numerosos problemas en su aplicación a la cuantificación de proteínas y, particularmente, en el análisis de datos en diferentes aplicaciones. Aunque la estrategia se benefició inicialmente de los avances en el campo de los microarrays de DNA resulta evidente que se necesitan mejores herramientas bioestadísticas, más adaptadas a las distintas modalidades de microarrays de proteínas para un análisis correcto de datos, especialmente en el campo de biomarcadores (67). El campo sigue creciendo a una gran velocidad, así como el número de posibles aplicaciones. Las aplicaciones clínicas son particularmente relevantes y es

presumible que surja una nueva generación de reactivos de diagnóstico a partir de las investigaciones actualmente en curso. La aplicación de los microarrays de proteínas y del conocimiento generado por el análisis individual de los circuitos de señalización molecular específicos de cada paciente con vistas a una terapia individualizada es factible en la práctica clínica rutinaria futura.

Agradecimientos.
Agradezco a Ingrid Babel y Rodrigo Barderas el trabajo realizado durante los últimos años en la búsqueda de biomarcadores de cáncer colorrectal y que nos ha permitido poner a punto muchas de las técnicas que aquí se describen. Este trabajo ha sido financiado en parte por el proyecto del Plan Nacional BIO2009-08818, el Programa COLOMICS2 de la CAM, la ayuda a grupos estables de investigación de la Asociación Española Contra el Cancer (AECC) y la plataforma Proteored.

REFERENCIAS.

1. Ekins RP, Chu FW. Multianalyte microspot immunoassay--microanalytical “compact disk” of the future. Clin Chem. 1991;37:1955-67.

2. Lueking A, Horn M, Eickhoff H, Bussow K, Lehrach H, Walter G. Protein microarrays for gene expression and antibody screening. Anal Biochem. 1999;270:103-11.

3. Casal JI, Barderas R. Identification of cancer autoantigens in serum: toward diagnostic/prognostic testing? Mol Diagn Ther. 2010;14:149-54.

4. MacBeath G, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. Science. 2000;289:1760-3.

5. Zhang Q, Bhola NE, Lui VW, et al. Antitumor mechanisms of combined gastrin-releasing peptide receptor and epidermal growth factor receptor targeting in head and neck cancer. Mol Cancer Ther. 2007;6:1414-24.

6. Haab BB. Antibody arrays in cancer research. Mol Cell Proteomics. 2005;4:377-83.

7. Miller JC, Zhou H, Kwekel J, et al. Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. Proteomics. 2003;3:56-63.

8. Madoz-Gurpide J, Canamero M, Sanchez L, Solano J, Alfonso P, Casal JI. A proteomic analysis of cell signaling alterations in colorectal cancer. Mol Cell Proteomics. 2007;6:2150-64.

9. Uhlen M, Bjorling E, Agaton C, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics. 2005;4:1920-32.

10. Ingvarsson J, Wingren C, Carlsson A, et al. Detection of pancreatic cancer using antibody microarray-based serum protein profiling. Proteomics. 2008;8:2211-9.

11. Steinhauer C, Wingren C, Hager AC, Borrebaeck CA. Single framework recombinant antibody fragments designed for protein chip applications. Biotechniques. 2002;Suppl:38-45.

12. Wingren C, Borrebaeck CA. Antibody microarray analysis of directly labelled complex proteomes. Curr Opin Biotechnol. 2008;19:55-61.

13. Speer R, Wulfkuhle JD, Liotta LA, Petricoin EF, 3rd. Reverse-phase protein microarrays for tissue-based analysis. Curr Opin Mol Ther. 2005;7:240-5.

14. Gulmann C, Sheehan KM, Kay EW, Liotta LA, Petricoin EF, 3rd. Array-based proteomics: mapping of protein circuitries for diagnostics, prognostics, and therapy guidance in cancer. J Pathol. 2006;208:595-606.

15. Liotta LA, Espina V, Mehta AI, et al. Protein microarrays: meeting analytical challenges for clinical applications. Cancer Cell. 2003;3:317-25.

16. Paweletz CP, Charboneau L, Bichsel VE, et al. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene. 2001;20:1981-9.

17. Boyd ZS, Wu QJ, O’Brien C, et al. Proteomic analysis of breast cancer molecular subtypes and biomarkers of response to targeted kinase inhibitors using reverse-phase protein microarrays. Mol Cancer Ther. 2008;7:3695-706.

18. Sevecka M, MacBeath G. State-based discovery: a multidimensional screen for small-molecule modulators of EGF signaling. Nat Methods. 2006;3:825-31.

19. Ambroz KL, Zhang Y, Schutz-Geschwender A, Olive DM. Blocking and detection chemistries affect antibody performance on reverse phase protein arrays. Proteomics. 2008;8:2379-83.

20. Nishizuka S, Charboneau L, Young L, et al. Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. Proc Natl Acad Sci USA. 2003;100:14229-34.

21. Hanash SM, Pitteri SJ, Faca VM. Mining the plasma proteome for cancer biomarkers. Nature. 2008;452:571-9.

22. Pereira-Faca SR, Kuick R, Puravs E, et al. Identification of 14-3-3 theta as an antigen that induces a humoral response in lung cancer. Cancer Res. 2007;67:12000-6.

23. Zhu H, Bilgin M, Bangham R, et al. Global analysis of protein activities using proteome chips. Science. 2001; 293:2101-5.

24. LaBaer J, Ramachandran N. Protein microarrays as tools for functional proteomics. Curr Opin Chem Biol. 2005;9:14-9.

25. Lueking A, Huber O, Wirths C, et al. Profiling of alopecia areata autoantigens based on protein microarray technology. Mol Cell Proteomics. 2005;4:1382-90.

26. Ramachandran N, Hainsworth E, Bhullar B, et al. Self-assembling protein microarrays. Science. 2004;305:86-90.

27. Zhong L, Hidalgo GE, Stromberg AJ, Khattar NH, Jett JR, Hirschowitz EA. Using protein microarray as a diagnostic assay for non-small cell lung cancer. Am J Respir Crit Care Med. 2005;172:1308-14.

28. Wang X, Yu J, Sreekumar A, et al. Autoantibody signatures in prostate cancer. The New England journal of medicine. 2005;353:1224-35.

29. Mintz PJ, Kim J, Do KA, et al. Fingerprinting the circulating repertoire of antibodies from cancer patients. Nat Biotechnol. 2003;21:57-63.

30. Babel I, Barderas R, Diaz-Uriarte R, Martinez-Torrecuadrada JL, Sanchez-Carbayo M, Casal JI. Identification of tumor-associated autoantigens for the diagnosis of colorectal cancer in serum using high density protein microarrays. Mol Cell Proteomics. 2009;8:2382-95.

31. Babel I, Barderas R, Diaz-Uriarte R, et al. Identification of MST1/STK4 and SULF1 Proteins as Autoantibody Targets for the Diagnosis of Colorectal Cancer by Using Phage Microarrays. Mol Cell Proteomics. 2011;10:M110 001784.

32. Hudson ME, Pozdnyakova I, Haines K, Mor G, Snyder M. Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays. Proc Natl Acad Sci U S A. 2007;104:17494-9.

33. Saeed AI, Bhagabati NK, Braisted JC, et al. TM4 microarray software suite. Methods Enzymol. 2006;411:134-93.

34. Saeed AI, Sharov V, White J, et al. TM4: a free, open-source system for microarray data management and analysis. Biotechniques. 2003;34:374-8.

35. Sreekumar A, Nyati MK, Varambally S, et al. Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. Cancer Res. 2001;61:7585-93.

36. Gembitsky DS, Lawlor K, Jacovina A, Yaneva M, Tempst P. A prototype antibody microarray platform to monitor changes in protein tyrosine phosphorylation. Mol Cell Proteomics. 2004;3:1102-18.

37. Huang RP, Huang R, Fan Y, Lin Y. Simultaneous detection of multiple cytokines from conditioned media and patient's sera by an antibody-based protein array system. Anal Biochem. 2001;294:55-62.

38. Celis JE, Gromov P, Cabezon T, et al. Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. Mol Cell Proteomics. 2004;3:327-44.

39. Pena C, Garcia JM, Larriba MJ, et al. SNAI1 expression in colon cancer related with CDH1 and VDR downregulation in normal adjacent tissue. Oncogene. 2009;28:4375-85.

40. Wolf-Yadlin A, Sevecka M, MacBeath G. Dissecting protein function and signaling using protein microarrays. Curr Opin Chem Biol. 2009;13:398-405.

41. Ptacek J, Devgan G, Michaud G, et al. Global analysis of protein phosphorylation in yeast. Nature. 2005;438:679-84.

42. Popescu SC, Popescu GV, Bachan S, et al. MAPK target networks in Arabidopsis thaliana revealed using functional protein microarrays. Genes Dev. 2009;23:80-92.

43. Newman JR, Keating AE. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science. 2003;300:2097-101.

44. Stiffler MA, Chen JR, Grantcharova VP, et al. PDZ domain binding selectivity is optimized across the mouse proteome. Science. 2007;317:364-9.

45. Hesselberth JR, Miller JP, Golob A, Stajich JE, Michaud GA, Fields S. Comparative analysis of Saccharomyces cerevisiae WW domains and their interacting proteins. Genome Biol. 2006;7:R30.

46. Polverini E, Rangaraj G, Libich DS, Boggs JM, Harauz G. Binding of the proline-rich segment of myelin basic protein to SH3 domains: spectroscopic, microarray, and modeling studies of ligand conformation and effects of posttranslational modifications. Biochemistry. 2008;47:267-82.

47. Jones RB, Gordus A, Krall JA, MacBeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature. 2006;439:168-74.

48. Funeriu DP, Eppinger J, Denizot L, Miyake M, Miyake J. Enzyme family-specific and activity-based screening of chemical libraries using enzyme microarrays. Nat Biotechnol. 2005;23:622-7.

49. van Baal JW, Diks SH, Wanders RJ, et al. Comparison of kinome profiles of Barrett's esophagus with normal squamous esophagus and normal gastric cardia. Cancer Res. 2006;66:11605-12.

50. Tuynman JB, Vermeulen L, Boon EM, et al. Cyclooxygenase-2 inhibition inhibits c-Met kinase activity and Wnt activity in colon cancer. Cancer Res. 2008;68:1213-20.

51. Gnjatic S, Ritter E, Buchler MW, et al. Seromic profiling of ovarian and pancreatic cancer. Proc Natl Acad Sci U S A. 2010;107:5088-93.

52. Chatterjee M, Mohapatra S, Ionan A, et al. Diagnostic markers of ovarian cancer by high-throughput antigen cloning and detection on arrays. Cancer Res. 2006;66:1181-90.

53. Ran Y, Hu H, Zhou Z, et al. Profiling tumor-associated autoantibodies for the detection of colon cancer. Clin Cancer Res. 2008;14:2696-700.

54. Mueller C, Liotta LA, Espina V. Reverse phase protein microarrays advance to use in clinical trials. Mol Oncol. 2010;4:461-81.

55. Sahin O, Lobke C, Korf U, et al. Combinatorial RNAi for quantitative protein network analysis. Proc Natl Acad Sci U S A. 2007;104:6579-84.

56. Popova T, Espina V, Bailey C, Liotta L, Petricoin E, Popov S. Anthrax infection inhibits the AKT signaling involved in the E-cadherin-mediated adhesion of lung epithelial cells. FEMS Immunol Med Microbiol. 2009;56:129-42.

57. Doolan DL, Mu Y, Unal B, et al. Profiling humoral immune responses to P. falciparum infection with protein microarrays. Proteomics. 2008;8:4680-94.

58. Iyengar P, Espina V, Williams TW, et al. Adipocyte-derived collagen VI affects early mammary tumor progression in vivo, demonstrating a critical interaction in the tumor/stroma microenvironment. J Clin Invest. 2005;115:1163-76.

59. Aguilar-Mahecha A, Cantin C, O'Connor-McCourt M, Nantel A, Basik M. Development of reverse phase protein microarrays for the validation of clusterin, a mid-abundant blood biomarker. Proteome Sci. 2009;7:15.

60. Moore LM, Holmes KM, Smith SM, et al. IGFBP2 is a candidate biomarker for Ink4a-Arf status and a therapeutic target for high-grade gliomas. Proc Natl Acad Sci U S A. 2009;106:16675-9.

61. Srivastava M, Eidelman O, Zhang J, et al. Digitoxin mimics gene therapy with CFTR and suppresses hypersecretion of IL-8 from cystic fibrosis lung epithelial cells. Proc Natl Acad Sci U S A. 2004;101:7693-8.

62. Seo YH, Carroll KS. Profiling protein thiol oxidation in tumor cells using sulfenic acid-specific antibodies. Proc Natl Acad Sci U S A. 2009;106:16163-8.

63. Belluco C, Mammano E, Petricoin E, et al. Kinase substrate protein microarray analysis of human colon cancer and hepatic metastasis. Clin Chim Acta. 2005;357:180-3.

64. Nam MJ, Madoz-Gurpide J, Wang H, et al. Molecular profiling of the immune response in colon cancer using protein microarrays: occurrence of autoantibodies to ubiquitin C-terminal hydrolase L3. Proteomics. 2003;3:2108-15.

65. Taylor BS, Pal M, Yu J, et al. Humoral response profiling reveals pathways to prostate cancer progression. Mol Cell Proteomics. 2008;7:600-11.

66. Madoz-Gurpide J, Kuick R, Wang H, Misek DE, Hanash SM. Integral protein microarrays for the identification of lung cancer antigens in sera that induce a humoral immune response. Mol Cell Proteomics. 2008;7:268-81.

67. Gnjatic S, Wheeler C, Ebner M, et al. Seromic analysis of antibody responses in non-small cell lung cancer patients and healthy donors using conformational protein arrays. J Immunol Methods. 2009;341:50-8.

Manual de
Proteómica

Sociedad Española de Proteómica