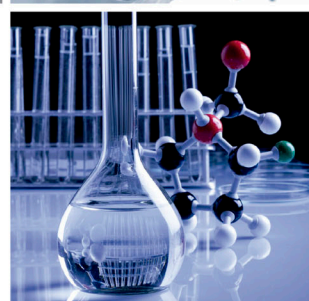
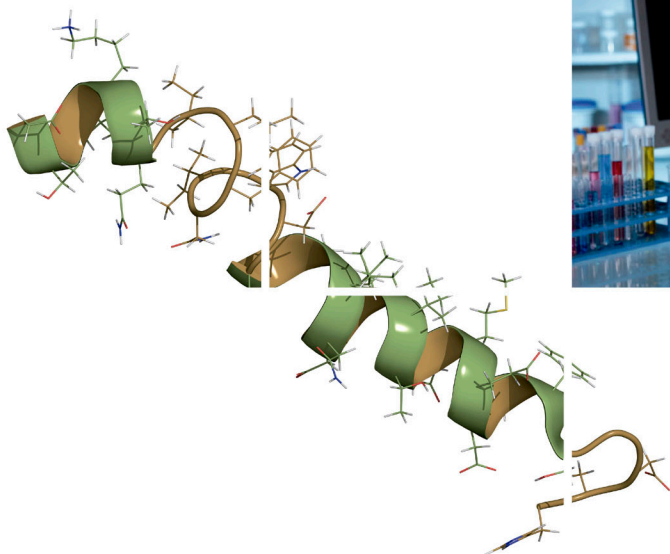
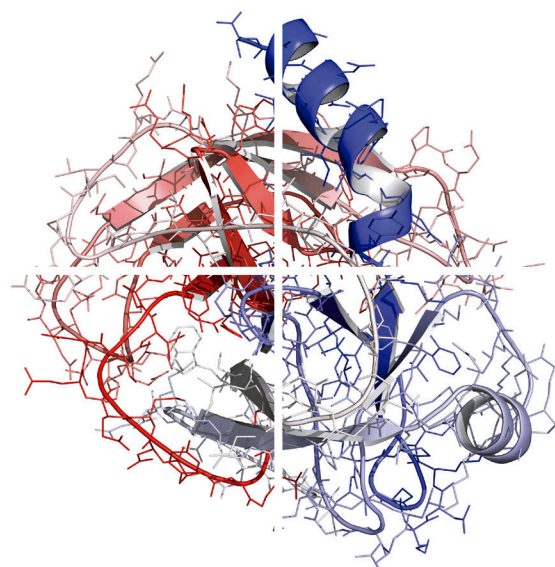




Manual de Proteómica

Volumen II



SEProt

Sociedad Española
de Proteómica

Manual de Proteómica

Volumen II



© **Sociedad Española de Proteómica** para la presente edición

C/ Jaime Roig 11

46010 Valencia

D.L.: NA-1731-2019

ISBN: 978-84-09-13534-9

Enmaquetado e impreso por Elías Rodríguez Toral. 23 pares (www.23pares.com)

Dedicado a
Juan Pablo Albar

Prólogo Manual de Proteómica - Volumen II

Hace ya cinco años que salía a la luz la primera edición del Manual de Proteómica editado por la SEProt, el cual nació con la intención de ser un referente didáctico en castellano para estudiantes y jóvenes científicos con interés por adquirir conocimientos básicos sobre proteómica. Esa primera edición, fundamentalmente introductoria y técnica, da paso ahora a una segunda muy enfocada a aspectos prácticos y metodológicos avanzados de gran aplicación, por ejemplo, en campos tan relevantes como el de la proteómica clínica. Se cubren ahora, por tanto, apartados no explorados anteriormente o que necesitaban ser actualizados. Así, en la presente edición se abordan en profundidad distintas aproximaciones de proteómica cuantitativa, incluyendo label-free, además de incluirse un capítulo dedicado a estudios top-down. El análisis de PTMs específicas más allá de las típicas fosforilaciones o glicosilaciones, como es el caso de ubiquitinación, acetilación y metilación, de gran relevancia en los sistemas biológicos, será abordado en detalle en otro capítulo que también se ocupará de análisis cuantitativos de PTMs basados en espectrometría de masas y anticuerpos. Los últimos avances en microarrays de proteínas y MS imaging también tienen su espacio en la presente edición, así como un capítulo dedicado a interactómica, todo ello de gran utilidad por ejemplo para la búsqueda de dianas farmacológicas y biomarcadores en condiciones patológicas. Esto último encuentra su complemento ideal en un capítulo dedicado a explicar la situación actual del Proyecto Proteoma Humano (HPP) y los principales logros conseguidos hasta la fecha, y en otro dedicado a la proteogenómica, poniendo énfasis precisamente en los estudios llevados a cabo en el contexto del HPP. El Manual se completa con un par de capítulos metodológicos dedicados a fragmentación peptídica y de proteínas por fotodisociación, y a métodos de adquisición de datos en espectrometría de masas (DIA vs. DDA). En resumen, esperamos que esta segunda edición del Manual de Proteómica de la SEProt sea de vuestro agrado y contribuya al objetivo de fomentar el interés y conocimiento sobre la investigación en el campo de la Proteómica en nuestro país y en otros muchos de habla hispana.

Ángel García e Ignacio Casal

Presidente y Vicepresidente de la SEProt

ÍNDICE



01.
Proteómica Cuantitativa



02.
Identificación de Modifica-
ciones post-traduccionales
mediante inmunopurificación
y espectrometría de masas



03.
Proteómica La estrategia
proteómica top-down
(Top-down proteomics)



04.
Fragmentación de Péptidos y
Proteínas por Fotodisociación



05.
Técnicas de adquisición
independiente de datos



06.
MALDI -TOF imaging mass
spectrometry (MALDI-IMS)



07.
Interactómica. Estrategias de
purificación por afinidad para
análisis del proteoma



08.
Proteogenómica



09.
Avances en microarrays de
proteínas: tipos, ensayos y
aplicaciones



10.
Proyecto Proteoma Humano
(Human Proteome Project,
HPP)

*Lucía Monteoliva Díaz¹, Inmaculada Jorge Cerrudo²,
Jesús Vázquez Cobos³, Manuel M. Sánchez del Pino⁴*

¹Profesora del Departamento de Microbiología, Universidad Complutense de Madrid.

²Científico Postdoctoral del Laboratorio de Proteómica Cardiovascular, Centro Nacional de Investigaciones Cardiovasculares.

³Director del Laboratorio de Proteómica Cardiovascular, Centro Nacional de Investigaciones Cardiovasculares.

⁴Profesor del Departamento de Bioquímica y Biología Molecular, Universidad de Valencia.

1. Introducción.

El reciente desarrollo de la proteómica permite a los investigadores identificar un elevado número de las proteínas presentes en un sistema biológico con relativa facilidad. Sin embargo, para conocer y entender con detalle los sistemas biológicos no basta con saber cuáles son sus componentes sino que también es necesario saber la abundancia de cada uno de ellos. Indudablemente, el análisis diferencial de proteínas es una estrategia experimental muy atractiva para los investigadores y ha estado presente desde las etapas iniciales de la proteómica. De hecho, el flujo de trabajo desarrollado inicialmente consistía en la separación y cuantificación de proteínas mediante electroforesis bidimensional (2-DE) seguido de la identificación de las proteínas diferenciales mediante huella peptídica.

La proteómica cuantitativa tiene por objeto determinar la abundancia de las proteínas presentes en el sistema biológico objeto de estudio. La cuantificación puede ser absoluta, si se proporciona la abundancia de las distintas proteínas presentes en una muestra, o relativa, si se proporciona el cambio de cada proteína en distintas muestras. La cuantificación de proteínas, sin embargo, no es trivial y supone un reto importante para cualquier técnica analítica. La dificultad deriva, principalmente, de dos factores. Por un lado, el amplio rango de concentraciones de proteínas presentes en una muestra. Se ha estimado que, en células de levadura, la abundancia de proteínas puede oscilar entre 50 y más de un millón de copias por célula (1, 2). Por otro lado, el número de proteínas distintas presentes en la muestra es elevado (3). El número estimado de proteínas expresadas en una célula de mamífero es de 10000 (4). Sin embargo, el número de variantes y proteínas modificadas puede aumentar considerablemente este número. Por tanto, para abordar la cuantificación del proteoma necesitamos técnicas analíticas capaces de discriminar entre miles de especies químicas distintas que se encuentran distribuidas en un rango de concentraciones que abarca varios órdenes de magnitud (5). Aunque la cobertura del proteoma que se obtiene con las técnicas de proteómica aumenta de manera constante, todavía no es completa. El número de proteínas que pueden ser cuantificadas es todavía menor ya que los requisitos para cuantificar son más exigentes que para identificar proteínas (6).



Para abordar la cuantificación de proteínas no hay una única metodología sino que disponemos de distintas estrategias experimentales. Como suele suceder, se trata de estrategias complementarias, cada una con sus ventajas y limitaciones. Como se ha mencionado anteriormente, la estrategia inicial de cuantificación más utilizada fue la 2-DE. En este caso se cuantifican las proteínas de la muestra mediante un análisis de las imágenes de los geles tras una tinción específica de proteínas. Esta estrategia lleva asociados algunos inconvenientes o limitaciones que han impulsado el desarrollo de nuevas formas de cuantificación. En la actualidad, la 2-DE es la principal estrategia basada en la separación y cuantificación de proteínas completas. La mayor parte de las estrategias alternativas se han desarrollado para analizar péptidos y obtener datos cuantitativos mediante espectrometría de masas. Existen múltiples métodos de cuantificación que se diferencian en su fundamento teórico, tratamiento experimental de la muestra, análisis de los datos o en la información que utilizan para la cuantificación. Así pues, podemos diferenciar los métodos sin etiqueta o label-free, que no introducen ningún tipo de modificación química, de los métodos que utilizan un marcaje de los péptidos con distintos isótopos estables. La introducción de etiquetas moleculares permite el multiplexado, es decir, la transmisión de varias señales independientes a través de un mismo canal. De esta forma, si marcamos los péptidos de distintas muestras con etiquetas distintas se pueden combinar, realizar un único tratamiento experimental y analizar finalmente los péptidos procedentes de cada muestra de manera individual. El multiplexado reduce considerablemente la variabilidad experimental. Sin embargo, excepto en el caso del marcaje metabólico, la introducción de la etiqueta implica un paso adicional en el procesamiento de la muestra y, por tanto, una posible fuente de variabilidad y error. Además, en ocasiones, las etiquetas comerciales pueden incrementar el coste del experimento considerablemente. Otro criterio de clasificación podría ser el tipo de señal que se utiliza para la cuantificación. En este sentido, hay métodos basados en la intensidad de las señales que llegan al detector del espectrómetro de masas mientras que otros se basan en el número de péptidos o espectros que se asignan a cada proteína identificada. Atendiendo al tipo de etiqueta que se utiliza, hay métodos de cuantificación que añaden etiquetas químicamente idénticas que pueden tener la misma masa (isobáricas) o distinta masa (no isobáricas). En el caso de las etiquetas isobáricas, la cuantificación se realiza con la información de los espectros de fragmentación, en los que se observan iones reporteros específicos de cada una de las etiquetas. Por el contrario, las etiquetas no isobáricas muestran dobletes o tripletes en los espectros de masas. Estas diferencias tienen sus consecuencias a la hora del análisis de los datos e incluso en cuanto a los instrumentos con los que se pueden analizar las muestras.

Cuando estamos interesados en cuantificar proteínas nos enfrentamos al problema de decidir cuál de las estrategias experimentales disponibles debemos utilizar. Puesto que ninguna de las opciones puede considerarse como definitiva, es necesario conocer los aspectos básicos de cada una de ellas, sus ventajas y limitaciones, su implementación en el flujo de trabajo del laboratorio o la información que pueden proporcionar antes de seleccionar la que mejor se adapte a nuestras necesidades. Con este objetivo en mente, en este capítulo se describen, sin pretender ser exhaustivos, algunas de las estrategias de proteómica cuantitativa más utilizadas.

2. Cuantificación mediante electroforesis bidimensional: 2D-DIGE.

2.1. Introducción

La electroforesis bidimensional (2-DE) ha sido durante muchos años la técnica de elección para la separación de proteínas en los estudios de expresión diferencial de proteínas. Esto se debe a su elevado poder de resolución que se consigue utilizando dos propiedades intrínsecas de las proteínas: en la primera dimensión, su punto isoeléctrico y en la segunda, su peso molecular. Para la separación por punto isoeléctrico se utiliza un gradiente de pH y en la segunda dimensión un gel desnaturante de poliacrilamida (SDS-PAGE). La técnica descrita por O'Farrell en 1975 (7) ha sufrido diversas modificaciones, como el desarrollo de los gradientes de pH inmovilizados o tiras de inmobilinas (IPGs) que han supuesto un aumento notable de la reproducibilidad y versatilidad y han permitido la generalización de uso en los diversos estudios proteómicos. Todos los detalles relativos a la electroforesis bidimensional han sido revisados recientemente por diversos autores (8, 9).

El flujo clásico de trabajo para el análisis de la expresión diferencial de proteínas mediante 2-DE es el siguiente: obtención de varias réplicas biológicas de los extractos proteicos de las condiciones de estudio, realización de un número suficiente de geles de cada una de las réplicas biológicas para evitar problemas de reproducibilidad, tinción de los geles, análisis de las imágenes para su comparación y la cuantificación relativa de las proteínas y, por último, identificación mediante espectrometría de masas de aquellas proteínas que varían su abundancia relativa en las distintas condiciones de estudio (10, 11). Este tipo de análisis presenta ciertas desventajas, algunas de ellas intrínsecas a la utilización de geles bidimensionales y otras relacionadas con la reproducibilidad de la técnica y las carencias de los métodos de detección o de análisis de imagen. Los problemas clásicamente relacionados con la 2-DE incluyen la dificultad de separación de proteínas muy grandes o muy pequeñas, con pl extremos o muy hidrofóbicas. Se han desarrollado algunas modificaciones que mejoran la separación de algunos de estos tipos de proteínas, como proteínas de membrana o de pls extremos (revisadas por Gorg (8, 12)). Los problemas relacionados con los sistemas de detección, como la baja sensibilidad de la tinción con azul de Coomassie o el limitado rango dinámico lineal de la tinción de plata, así como los relacionados con la reproducibilidad han sido solventados sustancialmente con el desarrollo de la tecnología 2D-DIGE.

2.2. Principios básicos de la tecnología DIGE

La tecnología 2D-DIGE (2-D Fluorescence Difference Gel Electrophoresis) es una tecnología "multiplex" que se basa en el marcaje diferencial con fluorocromos de muestras proteicas a comparar para su posterior separación en el mismo gel bidimensional. Después de la electroforesis, el gel se escanea con distintas longitudes de onda obteniendo imágenes correspondientes a cada una de las muestras. De esta manera se incrementa notablemente la reproducibilidad del experimento con respecto al flujo



clásico de trabajo comentado anteriormente (13). El uso de este marcaje fluorescente también aumenta la sensibilidad de la detección de proteínas (detectándose cantidades menores a 1 ng) y aumenta el rango dinámico lineal de la cuantificación de las proteínas (14).

Los fluorocromos están diseñados siguiendo unas características determinadas: reaccionan con el mismo aminoácido, tienen masas moleculares similares con lo que el aumento de tamaño que producen en las proteínas a las que se unen es el mismo, no modifican la carga del aminoácido con el que han reaccionado y además tienen espectros de absorción y emisión fácilmente separables. Esto permite que las proteínas (o isoformas proteicas) iguales, pero procedentes de distinta muestras, presenten la misma movilidad relativa en el gel bidimensional independientemente del fluorocromo con el que estén marcadas. Por tanto, estas proteínas pueden ser codetectadas en el mismo lugar del gel y a la vez pueden cuantificarse en cada una de las muestras por separado. Esto incrementa la reproducibilidad y mejora la cuantificación relativa de las proteínas en cada una de las muestras.

El primer conjunto de fluorocromos desarrollados son ésteres derivados de N-hidroxi-succinimida que reaccionan con el grupo ϵ -amino de los residuos de lisina de las proteínas para formar una amida. Se utilizan para el denominado “marcaje mínimo”. En este tipo de marcaje solo se marca de un 2 a un 5% del total de residuos de lisina, ya que la unión de muchas moléculas de fluorocromo a una proteína provocaría un aumento importante de su masa molecular y además produciría problemas de solubilidad. En un principio se desarrollaron dos fluorocromos de este tipo, Cy3 y Cy5 (13). Más recientemente se ha desarrollado un tercer fluorocromo para marcaje mínimo, Cy2, permitiendo separar tres extractos proteicos en el mismo gel (15). Este último fluorocromo se utiliza normalmente para incluir un “estándar interno” en todos los geles del experimento, lo que facilita su comparación y la normalización de las cantidades de proteína suponiendo una mejora con respecto a la reproducibilidad de los datos y su análisis estadístico. Se ha descrito que con la inclusión del estándar interno se aumenta el número de proteínas en las que se detectan cambios en su abundancia relativa en muestras con una inherente variabilidad biológica (15, 16). Los primeros fluorocromos Cy2, Cy3 y Cy5 para marcaje mínimo se han comercializado por GE Healthcare Life Sciences (CyDye® DIGE Fluors). También se han desarrollado otros fluorocromos con las mismas características, por ejemplo 3Dye™ 2D DIGE (Lumiprobe) o, más recientemente, una tecnología parecida denominada Refraction-2D® que se basa en la unión irreversible a las proteínas de un nuevo conjunto de fluorocromos: G-Dye 200, G-Dye 300, G-Dye 100 (NH DyeAGNOSTICS).

Otro conjunto de fluorocromos se ha diseñado para el denominado “marcaje a saturación”. Tienen un grupo reactivo maleimida que forma un enlace covalente con el grupo tiol de los residuos de cisteína de las proteínas. Debido a que marcan todos los residuos de cisteína de las proteínas se incrementa significativamente la sensibilidad del método (ver apartado 2.4) (17). En el caso del marcaje a saturación solamente están disponibles los fluorocromos, Cy3 y Cy5, también comercializados por GE Healthcare Life Sciences.

Los fundamentos y las aplicaciones de la tecnología DIGE han sido revisados por diversos autores y los detalles prácticos relacionados con el marcaje o el desarrollo de

los experimentos pueden encontrarse en los manuales de uso de las casas suministradoras (11, 18-23). En este capítulo se describen los aspectos fundamentales para la realización de un estudio de expresión diferencial mediante 2D-DIGE con marcaje mínimo, así como sus ventajas e inconvenientes y algunas de las aplicaciones más actuales.

2.3. Utilización de 2D-DIGE con marcaje mínimo en estudios de expresión diferencial de proteínas

El marcaje mínimo con los fluorocromos Cy2, Cy3 y Cy5 es el que se emplea en la mayoría de los experimentos de 2D-DIGE. En este caso, lo más habitual es utilizar Cy3 y Cy5 para marcar los extractos proteicos de las distintas condiciones a ensayar y Cy2 para el marcaje de un estándar interno que se carga en todos los geles para favorecer la normalización de los datos. Los pasos a seguir en un análisis de este tipo son los siguientes: a) diseño experimental, b) obtención de las muestras, marcaje mínimo y electroforesis bidimensional, c) cuantificación relativa de las proteínas: detección y análisis de imagen, d) tinción de los geles e identificación y validación de las proteínas de interés (Figura 1).

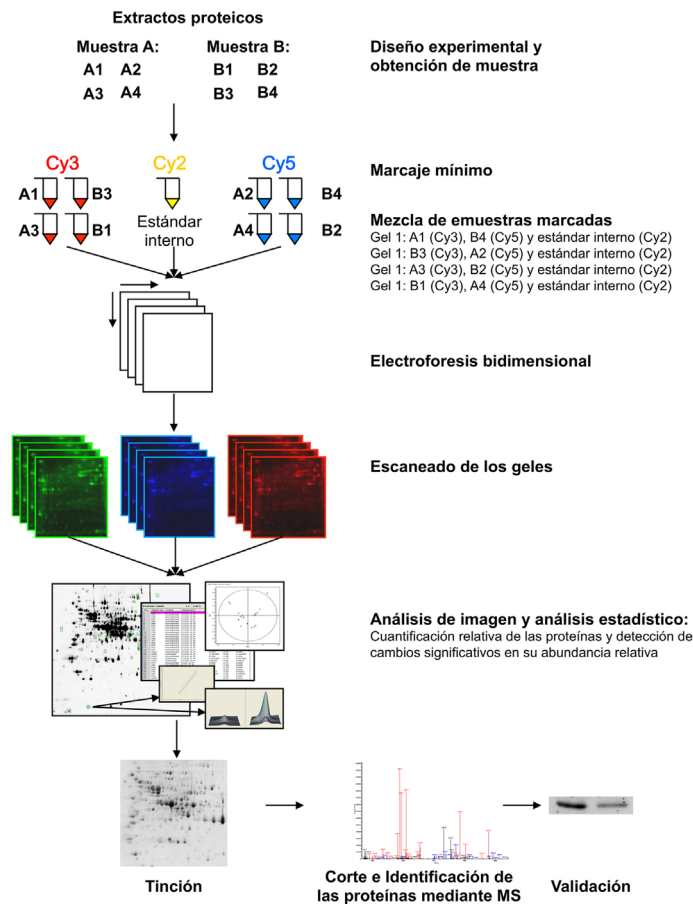


Figura 1. Esquema de un experimento de un experimento de 2D-DIGE utilizando marcaje mínimo.



2.3. a. Diseño experimental

Uno de los puntos claves para obtener resultados adecuados y significativos es un diseño experimental adecuado. Es importante seleccionar adecuadamente la muestra y las condiciones a comparar, así como el número de réplicas biológicas de cada condición y con qué fluorocromos se van a marcar.

Respecto a la muestra, en todos los estudios de expresión diferencial de proteínas es deseable la obtención de los extractos proteicos con protocolos sencillos, que no requieran muchos pasos de manipulación o enriquecimiento; ya que éstos pueden introducir variabilidad en el proceso. Según este razonamiento, el análisis de expresión diferencial de extractos celulares totales sería más adecuado que el de fracciones celulares. Sin embargo, el análisis de extractos celulares totales mediante electroforesis bidimensional presenta como importante problema el amplio rango dinámico de las proteínas en la muestra. Cuando existen proteínas de abundancias relativas muy diferentes dentro de una muestra, las proteínas poco abundantes no se detectan; a pesar de que se ha descrito que con 2D-DIGE se obtiene una respuesta lineal en un rango dinámico lineal de 4 órdenes de magnitud (14). Teniendo en cuenta la importancia de algunas de estas proteínas poco abundantes, como por ejemplo las proteínas reguladoras, en muchos casos es recomendable el fraccionamiento de la muestra (apartado 2.4).

El número de réplicas necesarias para tener resultados estadísticamente representativos no es fácil de establecer. Debido a que con el DIGE la reproducibilidad de los geles bidimensionales es bastante adecuada, no se realizan réplicas técnicas pero se necesita un número suficiente de réplicas biológicas. Se considera que entre 3 y 5 réplicas biológicas pueden ser suficientes, pero depende de la variabilidad intrínseca de la muestra (revisado en (23)). Así, cuando se trata de muestras de pacientes, el número de réplicas necesario puede ser mucho mayor. Se recomienda marcar la mitad de las réplicas biológicas de cada condición con Cy3 y la otra mitad con Cy5; evitando así, posibles preferencias de los fluorocromos por proteínas determinadas (revisado en (21)). El Cy2 se suele usar para el estándar interno que consiste en una muestra que se incluye en todos los geles y que suele estar formada por alícuotas de todas las muestras existentes en el experimento, de forma que incluye todas las proteínas a estudiar (15). Un ejemplo de diseño experimental se muestra en la Tabla 1.

Gel	Cy3	Cy5	Cy2
1	50 µg A1	50 µg B4	50 µg estándar interno
2	50 µg B3	50 µg A2	50 µg estándar interno
3	50 µg A3	50 µg B2	50 µg estándar interno
4	50 µg B1	50 µg A4	50 µg estándar interno

Tabla 1. Diseño experimental de un estudio de 2D-DIGE con marcaje mínimo. Se indica el fluorocromo con el cual se marca cada una de las 4 réplicas biológicas de las muestras A y B y cómo se mezclan las muestras marcadas para cargarlas en cada uno de los geles bidimensionales. El estándar interno está formado por alícuotas de A1, A2, A3, A4, B1, B2, B3 y B4.

2.3. b. Obtención de muestras, marcaje mínimo y electroforesis bidimensional

Además de realizar protocolos sencillos y reproducibles para la obtención de los extractos proteicos, hay que tener en cuenta la limpieza de las muestras y no incluir en el tampón de muestra productos que interfieran con el marcaje. Para la limpieza de los extractos proteicos se puede realizar un paso de precipitación en el que se eliminan

las sales y otros contaminantes que puedan interferir con el isoelectroenfoque de la 2-DE o con el marcaje. Es importante no incluir aminos primarias (como los anfolitos de la 2-DE) en el tampón para resolubilizar las proteínas, ya que como los fluorocromos reaccionan con el grupo ϵ -amino de las lisinas, las aminos interferirían con el marcaje. Además, tampoco se puede usar agentes reductores (24). Respecto a la cantidad de proteína, para un experimento de 2D-DIGE con marcaje mínimo se recomienda un mínimo de 75 μg de cada una de las muestras: 50 μg para marcar con Cy3 o Cy5 y 25 μg para su utilización en el estándar interno (ver Tabla 1). El proceso de marcaje es muy sencillo. Las condiciones óptimas descritas para el marcaje mínimo de las proteínas son: 50 μg de extracto proteico a pH 8,5 y 400 pmoles de fluorocromo, en frío y en oscuridad durante 30 minutos. La reacción se para con un exceso de lisina que reacciona con el fluorocromo libre. Con Cy2, se marcará suficiente cantidad de estándar como para incluir 50 μg en cada uno de los geles del experimento (24).

Una vez marcados, cada grupo de 3 extractos proteicos que van a ser separados en el mismo gel 2D según el diseño experimental realizado (Tabla 1) se mezclan y se cargan en una tira de isoelectroenfoque. Los geles bidimensionales se realizan de la manera habitual, simplemente hay que tener en cuenta que los sistemas de electroforesis que permiten correr muchos geles a la vez son más adecuados para los estudios comparativos ya que los geles serán más reproducibles. Además, si para detectar la fluorescencia se va a utilizar un sistema en que el gel no se separe de los cristales hay que tomar la precaución de utilizar cristales de baja fluorescencia.

2.3. c. Cuantificación relativa de las proteínas: detección y análisis de imagen

Una vez terminada la electroforesis, los geles se escanean para obtener las tres imágenes correspondientes a cada uno de tres fluorocromos. Se utilizan equipos con al menos 3 láseres (488 nm, 532nm y 633nm) que pueden excitar cada uno de los fluorocromos con uno de los láseres y separar sus espectros de emisión (19). Es importante la calidad de estas imágenes, ya que en ellas se basa la cuantificación de las manchas proteicas del gel; por tanto, hay que aplicar suficiente intensidad de láser como para detectar proteínas poco abundantes, pero evitar obtener fondo en la imagen y proteínas saturadas que no estarían dentro del rango dinámico lineal a la hora de cuantificar. Las imágenes de los geles se analizan mediante softwares de análisis de imagen, cuantificación y análisis estadístico como por ejemplo: DeCyder 2D (GE Healthcare Life Sciences), Progenesis SameSpots (TotalLab Ltd), Dymension 3 (Syngene, Imgen technologies), PD Quest 2D Analysis (Bio-Rad) o Delta2D (Decodon). Estos programas tienen algoritmos que detectan las proteínas o manchas proteicas de los geles, calculan su volumen según la fluorescencia y normalizan los volúmenes obtenidos. También comparan los volúmenes normalizados en el mismo gel y entre geles distintos, para lo que tienen que hacer coincidir los geles. El estándar interno facilita tanto este último proceso como el de normalización. Además, los softwares de análisis calculan las manchas proteicas que aumentan o disminuyen su abundancia relativa con una significancia estadística en las distintas condiciones de estudio. Se pueden realizar distintos tipos de análisis estadísticos cuyas ventajas y desventajas han sido revisadas recientemente por Minden y colaboradores (23).



2.3. d. Tinción de los geles e identificación y validación de las proteínas de interés

Para identificar las proteínas o isoformas proteicas, éstas se cortan de los geles. Debido a que con marcaje mínimo la mayor proporción de la proteína se encuentra sin marcar y la unión del fluorocromo provoca una pequeña diferencia en la movilidad electroforética, la mayor cantidad de proteína puede no co-localizar exactamente donde se detecta con los fluorocromos; por lo tanto, para cortar las proteínas los geles se tiñen después de su escaneado. Se puede utilizar cualquier tinción compatible con espectrometría de masas. Tras el corte del gel y digestión trípica, las proteínas se identifican mediante espectrometría de masas por los métodos habituales: huella peptídica (PMF) o fragmentación.

Una vez identificadas una serie de proteínas o isoformas proteicas cuya abundancia relativa varía en el proceso biológico que estamos estudiando, al igual que en otros tipos de estudios proteómicos de expresión diferencial, es de gran importancia validar los resultados mediante otras tecnologías. Si se dispone de anticuerpos frente a esa proteína, es frecuente la validación mediante Western Blot en SDS-PAGE. Esta técnica resulta muy adecuada excepto en el caso que se haya detectado incremento en la abundancia de proteína en unas isoformas y disminución en otras con distinto pI (Figura 2), ya que todas las isoformas se detectarían en la misma banda del gel monodimensional y se cuantificarían conjuntamente. En otros casos se validan los cambios proteicos detectados, analizando los procesos biológicos en los que están implicados con distintos tipos de ensayos funcionales.

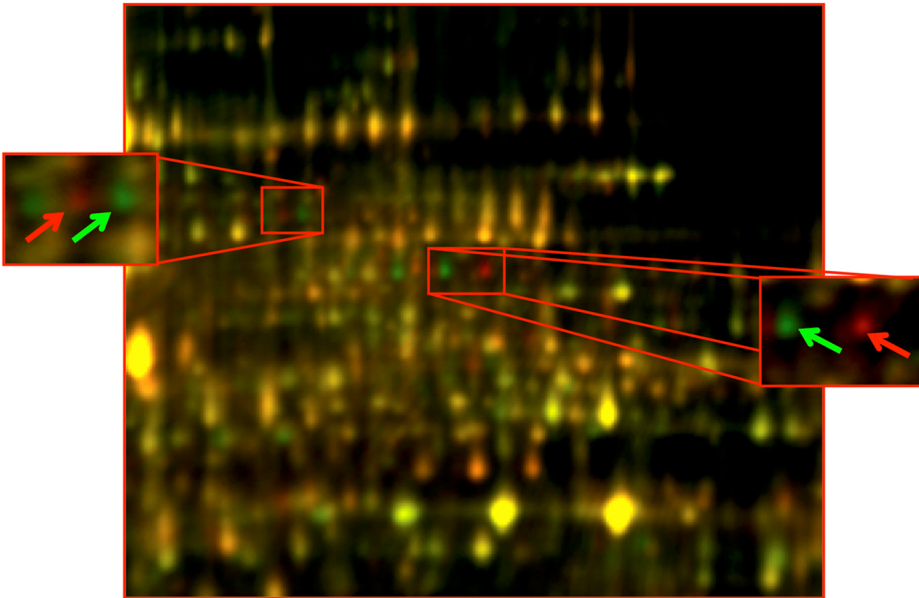


Figura 2. Imagen parcial de un gel de 2D-DIGE y detalle de proteínas que presentan isoformas con distintos cambios en su abundancia relativa. Se muestra el solapamiento de imágenes procedentes del marcaje con Cy3 y Cy5 de 2 extractos proteicos. Las isoformas proteicas detectadas como manchas proteicas amarillas presentan la misma abundancia en las muestras marcadas con los distintos fluorocromos, mientras que las manchas rojas y verdes indican mayor abundancia relativa en una de las muestras. En las zonas ampliadas se muestran 2 proteínas con al menos dos isoformas en las que se observan variaciones opuestas en su abundancia relativa: mientras que en una isoforma aumenta su abundancia relativa, en la otra disminuye. La ampliación de la derecha corresponde a dos isoformas de la enzima gliceraldehido-3-P-deshidrogenasa.

2.4. DIGE: Ventajas, desventajas y diferentes aproximaciones experimentales

Las ventajas más relevantes de la tecnología 2D-DIGE frente a las aproximaciones clásicas de los estudios de expresión diferencial basadas en geles bidimensionales y tinción son: disminución del número de geles, mejora en la reproducibilidad, aumento del rango dinámico lineal de cuantificación, mejora en la sensibilidad de detección y mejora en el análisis de imagen. Las limitaciones más importantes de esta tecnología son limitaciones intrínsecas a la 2-DE y relacionadas con el tipo de proteínas que se detectan. Normalmente, se detectan e identifican las proteínas solubles y más abundantes del proteoma de estudio que, en casos de extractos celulares, son proteínas citosólicas y metabólicas mayoritariamente. Para detectar otro tipo de proteínas se han descrito distintas estrategias de enriquecimiento. Por ejemplo, para disminuir la complejidad de los extractos de células eucariotas se han descrito enriquecimientos en orgánulos celulares específicos, como la mitocondria, seguidos de 2D-DIGE (25, 26). Otros estudios reducen la complejidad de las muestras a comparar con DIGE mediante técnicas de fraccionamiento como cromatografía de fase reversa de alta resolución (27) o isoelectroenfoque preparativo (28) permitiendo la detección e identificación de proteínas poco abundantes que sin enriquecimiento no habían sido identificadas. Además, como se ha comentado anteriormente, existe otro tipo de fluorocromos Cy3 y Cy5 que aumentan mucho la sensibilidad mediante el denominado marcaje a saturación. Estos fluorocromos reaccionan con todas las cisteínas de las proteínas y se usan para estudios de expresión diferencial de proteínas en casos en los que se dispone de cantidades muy pequeñas de muestra (17) como las obtenidas mediante microdissección por láser de células de tejidos neuronales o cancerosos (29, 30).

La comparación entre los estudios de expresión diferencial de proteínas mediante 2D-DIGE y técnicas de expresión diferencial sin gel es difícil. Se han realizado pocos estudios comparativos y, además, los resultados obtenidos de cada uno de estos estudios dependen de los protocolos concretos, la experiencia y la instrumentación de cada laboratorio. Uno de los estudios más importantes se ha realizado por el ABRF-PRG (31). Este y otros estudios han sido revisados por Timmis y Cramer (21). En general, la tecnología 2D-DIGE y las técnicas basadas en espectrometría de masas y marcaje isotópico presentan la misma precisión en la cuantificación de proteínas. Sin embargo, el 2D-DIGE presenta algunas ventajas como mayor rango dinámico de detección y cuantificación (21) y la capacidad de separar y cuantificar distintas isoformas de una proteína (Figura 2). Por otro lado, en cuanto al tipo de proteínas a analizar, las técnicas basadas en gel no son adecuadas para la identificación y cuantificación relativa de proteínas poco solubles. Además, las técnicas de cuantificación basadas en espectrometría de masas a nivel de péptidos tienen una cobertura muy superior, pudiéndose cuantificar varios miles de proteínas en un único experimento; en contraste las técnicas basadas en electroforesis bidimensional sólo permiten cuantificar, aproximadamente, un millar de proteínas. La mayoría de los autores coinciden en que ambas estrategias son complementarias.

Para el estudio de complejos proteicos, incluyendo complejos de proteínas de membrana, se utiliza una modificación de la tecnología DIGE: la combinación del marcaje con fluorocromos mediante DIGE y la separación de los complejos mediante "Blue native/SDS-PAGE". Esta aproximación permite la comparación sistemática y cuantificación relativa de las proteínas que forman parte de los complejos proteicos de distintas



fracciones u orgánulos celulares (32, 33).

2.5. Aplicaciones

La tecnología 2-DE DIGE se utiliza de manera rutinaria en estudios de expresión diferencial de proteínas en laboratorios de todo el mundo. Se ha aplicado a la mayoría de los campos de las ciencias biológicas o biomédicas. En abril de 2011 aparecían en Pubmed 976 citas, incluyendo 44 revisiones, que la utilizaban o comentaban. Por lo tanto, una revisión exhaustiva de todos los campos de aplicación sería demasiado larga. En distintas revisiones se comentan algunas de sus aplicaciones (11, 18, 21). Recientemente Minden y colaboradores han hecho un breve resumen de la utilización del 2D-DIGE centrándose en distintos organismos modelo, sobre todo en estudios de desarrollo o adaptación a distintas condiciones ambientales, en análisis de tejidos de enfermos y fluidos celulares para la identificación de nuevos biomarcadores y en proteómica subcelular (23).

En la actualidad, uno de los campos de aplicación en los que se continúan realizando muchos estudios con 2D-DIGE es el estudio de enfermedades humanas como cáncer, Alzheimer, enfermedades pulmonares o enfermedades infecciosas. Se han realizado muchos análisis de distintos tipos de cáncer mediante DIGE y algunos de los más recientes han identificado proteínas de interés clínico, como por ejemplo: proteínas clínicamente relevantes en cáncer de próstata (34), perfiles proteicos asociados con la diseminación de células tumorales en cáncer de cabeza y cuello (35) o un panel de proteínas biomarcadoras para predecir la respuesta a terapia en pacientes de mieloma (36). También se ha utilizado recientemente la tecnología 2D-DIGE para la identificación de nuevos biomarcadores para asma y enfermedad pulmonar obstructiva crónica (EPOC) (37), para el estudio de la enfermedad de Alzheimer en modelo de ratón (38) y para el estudio de enfermedades infecciosas como la tuberculosis pulmonar (39) o sus agentes etiológicos: bacterias como por ejemplo *Burkholderia cenocepacia* (40), hongos patógenos oportunistas como *Candida albicans* (28) y parásitos como *Leishmania chagasi* (41). Además de las aplicaciones biomédicas, se han descrito muchas otras, tanto en el campo de la biología básica, como por ejemplo en estudios de diferenciación celular (42) o en proteómica de plantas (43), así como en otros campos más aplicados como la contaminación medioambiental (44).

3. Cuantificación mediante espectrometría de masas.

3.1. Introducción

Como se ha mencionado, existen distintas estrategias para la cuantificación por espectrometría de masas. Se han desarrollado métodos específicos basados en el empleo de distintos isótopos estables que los espectrómetros de masas pueden discriminar con gran precisión y exactitud. La técnica de marcaje isotópico consiste en la incorporación de distintas etiquetas isotópicas al conjunto de péptidos o proteínas de

dos muestras que se quieren comparar. La introducción de estas etiquetas produce un cambio en la masa del péptido o de la proteína, sin afectar a sus propiedades bioquímicas y analíticas (45). Los marcajes se pueden realizar mediante la incorporación metabólica de isótopos estables en células en cultivo o en organismos enteros crecidos con medios complementados con aminoácidos (u otros nutrientes) marcados con ^2H , ^{13}C o ^{15}N (46), mediante reacción química de péptidos o proteínas con un reactivo de grupo marcado (ICAT (47), iTRAQ (48) o TMT (49), o mediante marcaje enzimático incorporando átomos de ^{18}O en el extremo C-terminal durante la protólisis con tripsina o proteasas análogas (50-52).

Una vez realizado el marcaje, el conjunto de péptidos marcados con el isótopo pesado se mezcla equimolarmente con el conjunto de péptidos marcados con el isótopo ligero. La cuantificación de los péptidos se realiza mediante espectrometría de masas en espectros MS o MS/MS, apareciendo como iones de la misma intensidad pero separados por la diferencia de masa que corresponde al marcaje isotópico. Los péptidos diferencialmente representados en cada muestra mostrarán ratios distintos al esperado 1:1, permitiendo la cuantificación relativa de los cambios a su correspondiente nivel de proteínas o la cuantificación absoluta utilizando péptidos marcados isotópicamente como estándares internos (53). Esta capacidad de multiplexado de las técnicas de marcaje isotópico reduce la variabilidad y el número de análisis considerablemente. Sin embargo, también tienen sus desventajas. Por ejemplo, excepto en el marcaje metabólico, es necesario someter las muestras a tratamientos químicos, más o menos complejos y costosos, que en menor o mayor medida van a introducir modificaciones no deseadas que alterarán la eficiencia de las identificaciones de los péptidos. Estos pasos adicionales en el procesamiento de la muestra necesariamente aumentan el tiempo de los experimentos y reducen el rendimiento global del proceso, lo que en numerosas circunstancias puede ser limitante. En la mayor parte de los casos, la cuantificación producida por estas técnicas, sobre todo cuando se realizan a gran escala, producen cuantificaciones relativas de proteínas individuales. Es decir, determinan la variación de cada proteína en distintas muestras. Sin embargo, no suelen proporcionar información sobre la abundancia absoluta de cada proteína respecto al total de proteínas de la muestra.

Paralelamente a las tecnologías de marcaje isotópico se están desarrollando estrategias de cuantificación *label-free* que no necesitan ningún tipo de marcaje, ampliando así las opciones de cuantificación en proteómica. En realidad, esta aproximación fue la primera que se utilizó en las etapas iniciales de la proteómica. En estos experimentos, las proteínas se cuantificaban en geles 2D, sin ningún tipo de marcaje, y después se comparaba la abundancia de cada proteína en los distintos geles. Una de las principales ventajas de estas técnicas es que no hay que tratar la muestra de ninguna manera especial. Una vez adquiridos los datos de los experimentos de LC-MS/MS, ya tenemos toda la información necesaria para realizar el análisis cuantitativo. Por tanto, los experimentos de *shotgun* que solían ser utilizados con fines únicamente descriptivos, pueden también ser analizados para obtener información cuantitativa.



3.2. Métodos sin marcaje o *Label-free*

3.2. a. Métodos basados en la intensidad de señal

Básicamente, hay dos estrategias generales para cuantificar mediante espectrometría de masas sin marcaje. La primera de ellas se basa en la intensidad de las señales. En estos métodos, la señal correspondiente a cada péptido (o más correctamente, a cada pico del espectro MS) se utiliza para construir un perfil cromatográfico o XIC (*eXtracted Ion Chromatogram*). Esta forma de cuantificar podría equipararse a la cuantificación convencional mediante HPLC de multitud de compuestos. La diferencia es que en este caso el detector conectado al HPLC es un espectrómetro de masas. El área del pico cromatográfico es proporcional a la cantidad de péptido que hay en la muestra en un amplio rango de concentraciones (54, 55). Tras una normalización adecuada, las áreas de los péptidos en distintas muestras pueden ser comparadas entre sí, o con un patrón si se quiere tener una cuantificación absoluta. La intensidad de señal con la que se detectan los péptidos en el espectrómetro de masas depende fundamentalmente de su concentración y sus propiedades fisicoquímicas. Por tanto, no pueden compararse las áreas de péptidos distintos, sino las áreas de cada péptido en distintas muestras. Si bien esto es así para los péptidos considerados individualmente, existe una relación entre la intensidad de la señal y la concentración de proteína. Se ha observado que la media de la señal de los tres péptidos trípticos más intensos por mol de proteína se mantiene constante con un coeficiente de variación menor del 10 % (56). Por tanto, si consideramos únicamente el promedio de las intensidades de estos péptidos, sí que podemos comparar entre distintas proteínas de una misma muestra. Si, además, introducimos en la muestra proteínas con concentración conocida podremos establecer la concentración absoluta aproximada de las proteínas identificadas en un experimento de *shotgun* (57, 58). Probablemente, la razón de esta relación es que los péptidos trípticos que generan una mayor respuesta de la señal MS tendrán una serie de propiedades fisicoquímicas similares, independientemente de las proteínas de las que provengan. Por tanto, las señales que produzcan serán comparables entre distintas proteínas y dependerán, fundamentalmente, de su concentración. Una variante de este procedimiento, denominada iBAQ (*Intensity-Based Absolute Quantification*) consiste en cuantificar las proteínas sumando las intensidades de todos los péptidos asignados a una proteína y normalizando este valor por el número de péptidos teóricos esperables en el rango de trabajo del espectrómetro (59).

El hecho de seleccionar y cuantificar picos de los espectros MS presenta ciertas ventajas. Por un lado, permite realizar el análisis cuantitativo sin necesidad de conocer la identidad de los péptidos a priori. Dado que los experimentos de LC-MS/MS contienen muchas más señales de las que finalmente se analizan (60), el análisis cuantitativo de dichas señales permite profundizar más en el proteoma e incluir péptidos con señales débiles que normalmente no serían identificados. Esta es la base de una estrategia de proteómica diferencial más racional denominada *directed proteomics* (61-65). Ésta consiste en realizar un primer experimento de LC-MS y comparar las diferencias de las áreas de todos los péptidos detectados. Tras analizar estadísticamente los datos, se seleccionan los péptidos diferenciales entre las distintas muestras. A continuación se generan listas de inclusión con estos péptidos diferenciales y se realiza un segundo experimento, esta vez de LC-MS/MS, para identificar únicamente los

péptidos diferenciales de interés. Por tanto, el proceso de selección e identificación se centra en los péptidos diferenciales, en lugar de hacerlo de manera indiscriminada en función de la señal. De esta manera se reduce considerablemente el sesgo hacia péptidos más abundantes de los experimentos de *shotgun*.

3.2. b. Métodos basados en el número de señales

La segunda estrategia de cuantificación sin marcaje, denominada recuento de espectros o *spectral counting*, se basa en la frecuencia con la que se detectan péptidos de cada proteína. Por tanto, en este caso, la cuantificación se realiza necesariamente después del proceso de identificación de proteínas. La idea básica en la que se fundamenta esta aproximación experimental es que en un experimento LC-MS/MS solo una fracción de los péptidos presentes en la muestra son seleccionados como precursores para ser fragmentados. Por tanto, es razonable pensar que la probabilidad de que un péptido sea fragmentado es mayor a medida que su concentración aumenta. Por supuesto, para este tipo de análisis es necesario corregir por el tamaño de la proteína ya que las proteínas grandes generan más péptidos que las pequeñas a igualdad de concentración molar. A diferencia de la intensidad de señal, la frecuencia de detección depende mucho menos de las propiedades fisicoquímicas que de la concentración de los péptidos. Es por ello que este tipo de análisis permite la comparación entre proteínas distintas de una misma muestra, proporcionando información sobre la abundancia de cada proteína respecto al conjunto de proteínas de la muestra. En realidad, es de los pocos métodos que nos permite tener información cuantitativa de una muestra de manera independiente, en lugar de cuantificaciones relativas entre varias muestras. En este sentido, se puede decir que estas técnicas proporcionan una cuantificación absoluta (66). La cuantificación la podemos realizar contando tanto los péptidos como los espectros asignados a cada proteína. Como acabamos de decir, una proteína de mayor tamaño genera más péptidos. Por ello, el número de péptidos asignados a cada proteína ha de normalizarse y hay varias maneras de hacerlo. La manera más sencilla es normalizar por el tamaño de la proteína, como se hace para calcular el NSAF (*Normalized Spectral Abundance Index*) (67, 68). En esta técnica, los espectros MS/MS (SpC) asignados a cada proteína k son normalizados por la longitud de la misma (L).

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

El valor de NSAF representa la abundancia de la proteína k respecto al total de las proteínas de la muestra (denominador, i). Tras la normalización, se ha demostrado que los valores de NSAF son proporcionales a la concentración de proteína (68). Otra forma de normalizar es la que se utiliza en el cálculo del emPAI (*Exponentially Modified Protein Abundance Index*) (69). En este método se tienen en cuenta la secuencia de las proteínas y las restricciones impuestas por la configuración del espectrómetro de masas y el gradiente cromatográfico utilizado. En función de estos últimos habrá ciertos péptidos observables y otros que no podrán ser detectados. Así, por ejemplo, no podrían



observarse péptidos menores de 700 Da si el rango de análisis del espectrómetro de masas es 350-2000 Th y seleccionamos iones con 2 o más cargas. De igual manera, los péptidos muy polares o muy hidrofóbicos no suelen eluir en los gradientes cromatográficos empleados normalmente. De esta forma, para calcular el emPAI se cuentan los precursores únicos de cada proteína y se normaliza por el número de péptidos potencialmente observables en nuestras condiciones de análisis (69):

$$PAI = \frac{N_{exp}}{N_{obs}}$$

$$emPAI = 10^{PAI} - 1$$

donde N_{exp} son los péptidos asignados a la proteína y N_{obs} son los péptidos observables de la misma. Este índice de abundancia de proteínas es fácil de calcular y hay distintos algoritmos de búsqueda, como Mascot y PeptideShaker, que lo calculan de manera rutinaria y lo proporcionan junto con los resultados de las identificaciones. Aunque la frecuencia de detección depende en mayor medida de la concentración de la proteína, las propiedades fisicoquímicas también influyen en el resultado final de estos análisis. Así, estas propiedades determinan el rendimiento de cada péptido durante las distintas etapas del procesamiento de las muestras hasta el análisis final en el espectrómetro de masas. Igualmente, la sensibilidad con la que se detecta cada péptido también está influenciado por las propiedades fisicoquímicas de los mismos. Por ello, los péptidos observables determinados en el cálculo del emPAI resultan a menudo una sobresimplificación que hace que los resultados pierdan precisión. Para tener en cuenta la dependencia de las condiciones experimentales en el número de péptidos observables se ha desarrollado un sistema alternativo más complejo, denominado APEX (*Absolute Protein Expression*) (70). Este método se ha implementado en el software de libre acceso *APEX Quantitative Proteomics Tool* (66). En éste, se utiliza un set de proteínas de referencia que estén en la muestra y que sean relativamente abundantes. Mediante un algoritmo de aprendizaje automático de clasificación se calcula la probabilidad de detección de los péptidos observados del set de referencia en función de distintas propiedades de los péptidos (70). A continuación se utilizan estos datos para calcular cuales son los péptidos observables más probables de cada proteína. Al final del proceso se determina un factor de corrección para cada proteína con el que normalizar los péptidos observados. De esta manera, se corrige el posible sesgo en la selección de los precursores como consecuencia del conjunto de condiciones experimentales utilizadas.

Más recientemente, se ha propuesto una metodología resultante de la combinación de las dos estrategias de cuantificación sin marcaje (71). Se evaluaron los principales parámetros derivados de espectrometría de masas, de forma aislada o combinada, así como varias formas de normalización. Al final se llegó a la conclusión de que la combinación del número de espectros (SC, el número de MS/MS por péptido), el número de péptidos únicos y la suma de las intensidades de los espectros de fragmentación (MS/MS) proporcionaban los mejores resultados. La normalización se realiza con la suma total de las intensidades de los espectros de fragmentación y la longitud de la proteína (L). Así, el parámetro cuantitativo propuesto es el índice espectral normalizado (SI_N , *Normalized Spectral Index*):

$$SI_N = \frac{\sum_{k=1}^{pn} \left(\sum_{j=1}^{sc} i_j \right)_k}{L_n \sum_{j=1}^{sc} i_j}$$

donde sc son los espectros de fragmentación, i es la intensidad de los iones de fragmentación del espectro j y k es un péptido de la proteína n . En resumen, se suman las intensidades de todos los iones (i) de todos los espectros MS/MS (j) de todos los péptidos (k) de la proteína (n) y se normalizan por la longitud de la proteína (L) y la suma total de las intensidades de los iones de fragmentación de todos los espectros MS/MS. De esta manera, la reproducibilidad entre replicas experimentales y biológicas así como la exactitud de los resultados parece mayor.

3.2. c. Mejoras y propiedades de los métodos sin marcaje

Un problema del *spectral counting* es que nos basamos en las identificaciones hechas, normalmente, mediante experimentos de descubrimiento o *shotgun*. En este tipo de experimentos se suelen adquirir los datos de LC-MS/MS en modo dependiente de los datos (DDA, *Data Dependent Acquisition*) en el que, como hemos mencionado, se pierde mucha información. Se ha demostrado que solo una fracción, relativamente pequeña, de los picos detectados por el espectrómetro de masas son seleccionados como precursores para fragmentar y obtener información de secuencia (60). Es más, de estos espectros MS/MS, solo una fracción se pueden utilizar para identificar péptidos. Es el mismo problema que tienen las técnicas de cuantificación por marcaje isobárico. Para aumentar la eficiencia de los experimentos *shotgun* se están desarrollando métodos de adquisición independientes de los datos (DIA, *Data Independent Acquisition*), que se describe en el capítulo 5. En este modo de adquisición se fragmentan todos los iones dentro de una ventana de selección de m/z predeterminada, en lugar de fragmentar precursores seleccionados del espectro MS. Posteriormente, se reconstruyen los espectros de fragmentación basándose en la co-elución cromatográfica de los fragmentos y los precursores presentes en la ventana de selección. Esta ventana de selección de precursores puede ir desde unas pocas unidades (72) a todo el rango de m/z (73, 74), pasando por valores intermedios (75-77). Los experimentos realizados de esta manera han mostrado, con mayor o menor éxito, un aumento de las proteína identificadas y una mayor cobertura de las mismas, una reducción del sesgo hacia proteínas más abundantes o señales más intensas y un aumento de la reproducibilidad entre experimentos. Todas estas ventajas mejoran significativamente la cuantificación tanto mediante *spectral counting* como mediante intensidad de señal. Sin embargo, su utilización no está muy extendida porque su implementación todavía no es trivial. La elección del método de adquisición depende del instrumento y del software que utilicemos. La velocidad de escaneo y la resolución son los principales determinantes del modo de adquisición. Aunque algunos grupos han desarrollado software de análisis propio (73, 75, 77), otras estrategias han sido desarrolladas por casas comerciales y sólo pueden utilizarse con sus instrumentos y su software.

Los métodos basados en *spectral counting* o en intensidades promedio son prácticamente los únicos que proporcionan información de cada proteína respecto al resto de proteínas de la muestra. Por ello, estas estrategias son ideales para determinar la



estequiometría de los componentes de complejos multiproteicos (78) y han sido empleadas con frecuencia para su caracterización (67, 68, 79-83).

Comparado con las técnicas de marcaje que permiten el multiplexado, en las técnicas sin marcaje cada muestra ha de analizarse por separado, lo que tiene ciertos inconvenientes. El tiempo de análisis y de utilización de los instrumentos se multiplica por el número de muestras analizadas. En ocasiones disponer de este tiempo adicional puede ser un problema o incrementar el coste del experimento significativamente si el análisis se está haciendo en un servicio. Este factor de multiplicación también afecta al procesamiento de los datos, que es relativamente intenso, sobre todo en los métodos basados en la intensidad de la señal. En este caso, el procesamiento de los datos incluye varias etapas para identificar los picos en los espectros, extraer su cromatograma y, finalmente, calcular el área para cada uno de ellos. Para ello es necesario el procesamiento de las señales, tanto a nivel de espectro como de cromatograma, para suprimir el ruido, suavizar los registros, discriminar entre señales solapadas, etc. La reproducibilidad de los experimentos es esencial para poder comparar los resultados y, aunque los experimentos se realicen en las mismas condiciones, siempre existe cierta variabilidad experimental entre las muestras que complica el diseño y análisis de los experimentos. La comparación de picos cromatográficos entre muestras requiere realizar un alineamiento. Dado que la m/z suele determinarse con mucha precisión y, además, podemos saber la identidad de muchos picos y utilizarlos como puntos de referencia, el alineamiento es menos complejo que en el caso de los spots de un gel 2D. Aun así, el alineamiento no es trivial y es esencial disponer de una buena reproducibilidad en la cromatografía, que es la parte más variable. Por ello, esta estrategia no es muy recomendable para experimentos complejos con separaciones multidimensionales que aumentan la variabilidad de las fracciones finales analizadas mediante LC-MS. Estas limitaciones no afectan a los métodos basados en el número de péptidos o espectros identificados. En estos casos el procesamiento de datos es más sencillo, sin necesidad de alinear los cromatogramas ni determinar las áreas. Por tanto, en casos de muestras complejas con separaciones más complejas, los métodos de *spectral counting* parecen más apropiados.

Estudios comparativos (31) han demostrado que, si bien las cuantificaciones realizadas por métodos sin marcaje se aproximan más a las concentraciones reales que los métodos con marcaje, la variabilidad entre réplicas es mayor. Asimismo, los errores de los métodos sin marcaje son más homogéneos con proteínas y concentraciones distintas. Por el contrario, la reproducibilidad de los métodos con marcaje parece ser mayor.

Por último, mencionar que aunque en ocasiones se consideran las técnicas *label-free* como opciones cuantitativas económicas por no requerir ninguna costosa etiqueta, no está claro que esto sea cierto. De hecho, si miramos las tarifas de los servicios (por ejemplo en www.proteored.org), es más costoso el análisis por LC-MS/MS que el marcaje. Por tanto, el mayor número de análisis de LC-MS/MS necesario en los experimentos *label-free* supera, en ocasiones, el coste de las etiquetas.

3.3. Proteómica dirigida: SRM/MRM

3.3. a. Introducción

Una de las técnicas proteómicas cuantitativas que más importancia está adquiriendo actualmente es la denominada SRM/MRM (*Selected/Multiple Reaction Monitoring*) (84). Conceptualmente, es distinta a la mayoría de estrategias cuantitativas que están basadas en experimentos *shotgun*. En estos últimos se analizan muchas proteínas simultáneamente (miles), aunque no se sabe de antemano cuáles. Por lo general, se trata de proteínas relativamente abundantes que, además, varían de un experimento a otro. Frente a esta aproximación experimental, los experimentos de SRM se centran en un número más reducido de proteínas (decenas) seleccionadas antes de realizar el experimento. Por ello, los experimentos de SRM se conocen como *targeted proteomics* o proteómica dirigida. La reducción del número de proteínas analizadas se compensa con la precisión, sensibilidad y rango dinámico en la cuantificación de las proteínas de interés. Estas propiedades convierten a los experimentos de SRM en la aproximación experimental más adecuada para realizar cierto tipo de experimentos (85). Así, por ejemplo, los análisis de SRM son uno de los pilares tecnológicos de proyectos como el del Proteoma Humano (*Human Proteome Project*), cuyo objetivo es la realización de un catálogo exhaustivo y sistemático de proteínas (86, 87). En áreas de estudio como la biología de sistemas, donde se requiere conocer con precisión la abundancia de las proteínas implicadas en determinados procesos de interés, los experimentos de SRM están mostrándose como herramientas de gran utilidad proporcionando datos cuantitativos de gran calidad (2, 88-90). El análisis y validación de biomarcadores es una aplicación particularmente apropiada para el empleo de las técnicas de SRM (63, 91-95). La exactitud y sensibilidad de la cuantificación de esta metodología permite sustituir eficientemente, al menos en determinadas ocasiones, a ciertas técnicas inmunológicas de uso más convencional, como ELISA (96) y western blot (97, 98). La cuantificación por SRM requiere la utilización de instrumentos tipo triple cuadrupolo (QQQ), ya que se basa en un modo particular de funcionamiento de estos instrumentos. Pese a que esta forma de análisis se está utilizando desde hace varias décadas para el análisis de moléculas pequeñas, su empleo en proteómica es más reciente. En esta técnica se cuantifican de manera específica una serie de péptidos, denominados péptidos proteotípicos, seleccionados por el usuario antes de empezar el experimento. Se trata, en definitiva, de desarrollar métodos analíticos de cuantificación específicos para cada proteína de interés. Hay excelentes revisiones explicando cómo establecer métodos de SRM (89, 99, 100).

3.3. b. Diseño de experimentos SRM

Para entender el funcionamiento de la cuantificación por SRM, es conveniente explicar brevemente el funcionamiento de los QQQ y su modo de funcionamiento en los experimentos de SRM. Los cuadrupolos son filtros de masas que dejan pasar al detector únicamente iones de una determinada m/z , la cual depende de los voltajes que se aplican. Por tanto, para hacer un espectro de masas es necesario hacer una rampa de voltaje de manera que la m/z de los iones que van llegando al detector va cambiando en función del voltaje concreto que se aplica en cada momento. La rampa



de voltaje que se aplica permite cubrir el rango de masas deseado. Este tipo de instrumentos puede funcionar obteniendo espectros MS (aplicando una rampa de voltajes) o monitorizando uno o varios iones (aplicando voltajes determinados) a lo largo de la cromatografía. Estas mismas opciones de funcionamiento se aplican a los cuadrupolos anterior y posterior a la cámara de colisión de los QQQ. En los experimentos de SRM, ambos cuadrupolos operan con voltajes discretos, en lugar de barridos. En estos casos, el primer cuadrupolo se fija en la m/z correspondiente al péptido que se desea analizar y el tercer cuadrupolo se fija en la m/z de uno de los fragmentos, generalmente de la serie y, que se producen en la cámara de colisión. A la combinación precursor/fragmento se le denomina transición. Una de las ventajas de este modo de funcionamiento es el aumento de sensibilidad que se obtiene (101) debido a que el instrumento está más tiempo acumulando señal correspondiente al ión seleccionado (comparado con la realización de un espectro). La contrapartida a este aumento de sensibilidad es cierta pérdida de información. No tenemos confirmación de la identidad del péptido seleccionado debido a que no hay un espectro de fragmentación que verifique la secuencia. Sin embargo, si se monitorizan varios fragmentos de cada péptido, la especificidad de la medida aumenta considerablemente. En la práctica, para cuantificar una proteína se analizan varios péptidos proteotípicos (de 2 a 5) y se monitorizan múltiples transiciones (de 2 a 5). El número de transiciones que podemos analizar depende de varios factores. Por un lado hay que considerar el tiempo dedicado a la detección de cada transición (normalmente entre 10 y 100 ms), del que depende la sensibilidad de la medida. Por otro lado hay que considerar las medidas con las que queremos definir el pico cromatográfico, que afectará a la precisión y exactitud en la cuantificación. Cuantas más medidas adquiramos mejor definiremos los picos y más fiable será la cuantificación. Sin embargo, como es lógico, cuantas más medidas realicemos de una transición, menos transiciones podremos analizar. Por tanto, el número de transiciones que podemos analizar dependerá del equilibrio que alcancemos entre el tiempo de medida dedicado a cada una y la frecuencia con la que las analizamos. Así, el número de medidas que utilicemos para definir el pico cromatográfico determinará la duración de los ciclos de funcionamiento del espectrómetro de masas mientras que el tiempo de medida que dediquemos a cada transición determinará las transiciones que se pueden analizar en cada ciclo.

Si consideramos un pico cuya duración es de 20 segundos y queremos definirlo con 10 puntos experimentales, necesitamos ciclos de 2 segundos. Entonces podemos analizar 200 transiciones de 10 ms (buena sensibilidad) o 20 transiciones de 100 ms (excelente sensibilidad). Estos tiempos de análisis son sensiblemente superiores al tiempo dedicado a una masa concreta durante la realización de un espectro de fragmentación. Si para cuantificar una proteína necesitamos analizar un mínimo de 2 péptidos y 2 transiciones por péptido, el cálculo anterior indica que podemos analizar de 5 a 50 proteínas en un mismo experimento de SRM. Sin embargo, si tenemos en cuenta que un pico cromatográfico tiene una duración de unos 20-30 segundos, resulta muy ineficiente monitorizar las mismas transiciones a lo largo de toda la cromatografía, cuya duración suele ser de varias decenas de minutos. Si disponemos de un sistema cromatográfico reproducible y conocemos los tiempos de elución de los péptidos que queremos analizar podemos optimizar el rendimiento de los instrumentos y aumentar el número de transiciones analizadas. Para ello, podemos programar el análisis de cada transición en una ventana de tiempo centrada en el tiempo de retención de cada péptido, aumentando considerablemente el número de transiciones que se pueden

analizar en un mismo experimento. Por ejemplo, si realizamos una cromatografía en la que los péptidos eluyen en un periodo de 30 min y programamos ventanas de 5 min para cada transición, estamos aumentando el número total de transiciones analizadas por un factor de 6. Indudablemente, para realizar un experimento de SRM programada (*Scheduled SRM*) debemos planificar cuidadosamente los experimentos y seleccionar los péptidos apropiados.

¿Qué proteínas y qué péptidos elegimos? La elección de la proteína depende, obviamente, del trabajo de cada investigador. Sin embargo, sí que hay ciertos factores que se deben tener en cuenta a la hora de seleccionar los péptidos proteotípicos que se van a utilizar para cuantificar. Como en cualquier técnica analítica, el objetivo es obtener la máxima especificidad y sensibilidad. En cuanto a la especificidad, hay que tener en cuenta que la cuantificación de proteínas por SRM se basa en la cuantificación de varios péptidos que van a “representar” a la proteína de interés. Por tanto, idealmente, deben seleccionarse péptidos únicos cuya secuencia no se repita en ninguna otra proteína en la especie que estemos analizando. Esto puede ser problemático en determinadas ocasiones como ante la presencia de isoformas y proteínas muy relacionadas, pudiéndose dar situaciones en las que no sea posible diferenciar entre ellas. En cuanto a la sensibilidad, se deben seleccionar aquellos péptidos que mejor se ionicen y proporcionen la máxima señal. Dado que es difícil predecir la respuesta de cada péptido, con frecuencia se utilizan datos de shotgun para identificar los péptidos de cada proteína que den las señales más intensas. Existen diversas bases de datos en Internet donde se puede buscar información sobre péptidos y sus propiedades (<http://www.srmatlas.org>, <http://www.peptideatlas.org>, <http://www.thegpm.org>). Además de estas consideraciones, hay una serie de propiedades que debemos tener en cuenta para seleccionar los péptidos. Puesto que perseguimos la máxima sensibilidad, resulta deseable que la mayor parte del péptido presente se concentre en una única especie iónica. Así, deberíamos seleccionar péptidos en los que la intensidad del ión secundario (generalmente triplemente cargado) represente menos del 10 % de la intensidad del ión mayoritario (generalmente doblemente cargado) (89). Si dividimos la cantidad de péptido presente en varios iones estamos disminuyendo la intensidad y, por tanto, la sensibilidad de la medida. Además, estamos complicando el análisis cuantitativo, ya que deberíamos computar todas las especies correspondientes a cada péptido. Por otro lado, no son aconsejables los péptidos susceptibles de sufrir modificaciones, ya sean de origen biológico o como consecuencia del tratamiento de la muestra, sobre todo porque se trata de un proceso poco controlado. Por ejemplo, un péptido que contenga metionina podría estar presente en su forma normal y oxidada. Por ello suelen excluirse péptidos que contengan metionina en cualquier posición o glutamina en posición N-terminal. En ocasiones, los péptidos flanqueados por pares de aminoácidos básicos (KK, RR, KR, RK) también son excluidos porque con frecuencia se producen cortes tripticos incompletos (102). Otro factor a tener en cuenta es el tamaño del péptido, que debe estar en el rango adecuado para una detección eficiente. Este rango puede estar entre 7 y 25 aminoácidos, aunque no hay un rango claro establecido. Si bien es aconsejable seguir estas recomendaciones, en la práctica es difícil cumplir todas ellas en todos los péptidos proteotípicos. Respecto a los fragmentos a utilizar, los más intensos y con mejor relación señal ruido proporcionarán mayor sensibilidad y, por lo general, suelen ser de la serie y. Los fragmentos cercanos a la m/z del precursor suelen mostrar más ruido. Por el contrario, los fragmentos con m/z mayores que la del ión precursor suelen ser más selectivos



porque eliminan posibles contribuciones de contaminantes monocargados. Como hemos mencionado, esta metodología implica el desarrollo de un método analítico específico para cada péptido. En este análisis individualizado debemos optimizar todos los parámetros necesarios para obtener la máxima señal en cada caso, sobre todo cuando necesitamos el máximo de sensibilidad. Entre estos parámetros, la energía de colisión debe ser optimizada para cada transición. Suele haber una relación lineal entre la energía de colisión óptima y la m/z del precursor a fragmentar que depende del tipo de instrumento. Los fabricantes suelen proporcionar esta información y se ha observado que la diferencia entre la energía de colisión calculada y la determinada experimentalmente suele ser menor del 10 % (103). Otros parámetros se optimizan para el conjunto del experimento, como es el caso de las condiciones cromatográficas y de ionización. En general, la ganancia conseguida con la optimización de estos parámetros es menor de un factor de 2 (85).

Una vez que se han seleccionado los péptidos y las transiciones, es necesario validarlas. Es decir, nos tenemos que asegurar de que realmente cuantifican lo que se supone que deben cuantificar. La utilización de péptidos sintéticos facilita mucho el desarrollo del método analítico, sobre todo si están marcados con isótopos estables. Su análisis permite caracterizar con precisión y facilidad las propiedades de ionización, fragmentación y cromatográficas de los péptidos de interés. De esta manera podemos saber con exactitud los tiempos de elución y, si es necesario, programar las ventanas de tiempo de análisis para cada uno de ellos. Si no disponemos de péptidos sintéticos tendremos que basarnos en otros parámetros. Un requisito imprescindible para validar la identidad del péptido analizado es que todas sus transiciones muestren una coelución perfecta en la cromatografía. La probabilidad de observar coelución de transiciones inespecíficas disminuye a medida que aumenta el número de transiciones analizadas para cada péptido, pero aumenta con la complejidad de muestra. Igualmente, la relación de las señales de las transiciones debería ser la misma o similar que la observada en los espectros de fragmentación. Puesto que en un experimento de SRM no se hacen espectros MS/MS de los péptidos analizados, es muy recomendable confirmar que la señal observada es debida al péptido seleccionado y así eliminar posibles falsos positivos. Si no es posible disponer de péptidos sintéticos, uno de los procedimientos habituales en los experimentos preliminares de puesta a punto es la realización de espectros MS/MS de los picos de interés. Los instrumentos suelen disponer de procedimientos para realizar espectros MS/MS cuando detectan la elución de las transiciones programadas (104). Sin embargo, si la señal es débil resulta complicado obtener espectros con información suficiente para confirmar la secuencia ya que la sensibilidad al realizar un espectro MS/MS es menor que el análisis por SRM. Si no es posible obtener un espectro MS/MS pero conocemos el tiempo de retención, podemos utilizarlo para confirmar, aunque sea de manera indirecta, el origen de las transiciones observadas. Para determinar la utilidad de los métodos desarrollados en muestras biológicas es importante determinar el efecto de la matriz en la detección de los péptidos. Los problemas de supresión de señal son frecuentes y pueden afectar a la sensibilidad del método. Además, la complejidad de la matriz aumenta las probabilidades de aparición de señales inespecíficas que puedan interferir con las medidas y dificultar la interpretación correcta de los datos. Por ello, es necesario realizar medidas con distintas matrices para determinar y cuantificar estos posibles efectos negativos.

En función de nuestras necesidades podremos diseñar los experimentos para realizar

una cuantificación relativa o absoluta. En el análisis de los experimentos de SRM se cuantifican las áreas de los picos cromatográficos de todas las transiciones que hayamos programado. Con el valor absoluto de las intensidades, directamente, podemos realizar una cuantificación relativa, de forma label-free. Obviamente, para corregir errores en la cantidad de muestra cargada en cada análisis, posibles variaciones en la estabilidad del spray a lo largo del experimento y otros factores difíciles de controlar, es necesario una normalización de los datos. Una opción es realizar una normalización frente a la señal total asumiendo que la mayor parte de los péptidos analizados no varía. También se puede añadir una misma cantidad de un estándar interno que nos sirva de referencia para cada experimento. Este tipo de procesos de normalización son efectivos para corregir desviaciones globales, pero no problemas individuales, como podría suceder con una interferencia de la matriz con un péptido concreto. La otra forma de cuantificar es utilizando estándares marcados con isótopos estables. En este caso, la cuantificación ya no depende de la intensidad absoluta de la señal sino de la relación entre las dos versiones de los péptidos analizados, marcados y sin marcar. Aparte del aumento del coste del experimento por la obtención de los estándares marcados, también se reduce el número de péptidos que podemos analizar a la mitad. Sin embargo, la calidad de la cuantificación es significativamente mejor porque elimina o reduce el efecto de muchas interferencias, sobre todo a nivel individual. La mayor parte de las posibles interferencias (spray, matriz, etc.) afectan de igual manera a los péptidos endógenos y a los estándares añadidos, que son químicamente iguales. Lo normal en este tipo de experimentos es normalizar por la cantidad de muestra de la que se parte, es decir, por la cantidad de proteínas en la mayor parte de los casos. Sin embargo, la cuantificación de proteínas no es un proceso que esté libre de errores y, en general, no proporciona medidas muy exactas. Una forma de normalizar de manera interna por la cantidad de proteína es cuantificar algunas proteínas, relativamente abundantes, que supuestamente se mantengan constantes (las denominadas *housekeeping*) (105). Si utilizamos unas 10 o más y tomamos la mediana de los valores de cuantificación, aunque alguna de ellas varíe no afectará al resultado. De esta manera tenemos un estándar de abundancia interno. Para realizar una cuantificación absoluta debemos saber con exactitud la cantidad del estándar interno marcado que añadimos. Realmente, este es el gran problema para obtener un valor preciso. La precisión de la cuantificación dependerá de la exactitud de la medida de la concentración del estándar utilizado. Esta cuantificación no es trivial. Si utilizamos péptidos sintéticos, estos pueden estar cuantificados de manera precisa mediante análisis de aminoácidos, lo cual encarece bastante los péptidos. Si usamos proteínas completas (106) o péptidos concatenados (QconCAT) (107), determinar la concentración con exactitud puede resultar más complicado.

La cuantificación de proteínas por SRM se realiza a partir de la cuantificación de unos pocos péptidos representativos. Por tanto, la abundancia de los péptidos que llegan al espectrómetro de masas debe ser una fiel representación de la abundancia de la proteína en la muestra. Para ello, hay que optimizar el proceso de preparación de muestras y cuantificar, lo mejor posible, el rendimiento de cada una de las etapas del proceso. Si queremos realizar una buena cuantificación debemos tener conocimiento y, a ser posible, cuantificar cualquier pérdida de material, péptido o proteína, que se haya podido producir en el procesamiento de las muestras. Si utilizamos estándares marcados, su introducción en el flujo de trabajo se debe realizar lo antes posible. Una vez que estén presentes en la muestra, las pérdidas de material afectarán por igual



a producto endógeno y al estándar y, por tanto, neutralizarán su efecto en la cuantificación final. Uno de los pasos más importantes en el procesamiento de la muestra es la obtención de péptidos, generalmente, mediante digestión triptica de las proteínas de la muestra. Una digestión incompleta implica una relación péptido proteotípico/proteína menor de uno y, por tanto, producirá una subestimación de la concentración de la proteína de interés. En principio, la mejor opción para contrarrestar el efecto de una digestión deficiente es la adición a la muestra de proteínas enteras marcadas isotópicamente (106, 108). Además, también corregirá en gran medida las posibles pérdidas de proteína en algunas de las etapas del procesamiento. Esta estrategia asume que el grado de modificación del estándar y de la proteína endógena es similar. Existen otras estrategias para tratar de corregir los efectos de una digestión ineficiente basadas en la adición de proteínas químicas con múltiples péptidos proteotípicos marcadas isotópicamente, denominada QconCAT (107), o la utilización de péptidos extendidos con las regiones flanqueantes de la proteína (109).

Los métodos de SRM han mostrado un amplio rango dinámico y una razonable sensibilidad, en el rango de attomoles (2). Sin embargo, el rango dinámico de las proteínas en plasma y la baja abundancia de algunas de ellas de interés clínico se escapan al análisis directo por SRM. Una estrategia para aumentar la sensibilidad de los análisis de SRM es combinarla con la inmunopurificación de los péptidos proteotípicos de interés, denominada SISCAPA (*Stable Isotope Standards and Capture by Anti-Peptide Antibodies*) (110, 111). Otras estrategias para mejorar la sensibilidad incluyen el enriquecimiento de subprotemas (112), como proteínas N-glicosiladas (113). Sin embargo, aumentar la complejidad del procesamiento puede añadir variabilidad y fuentes de error indeseables. Además, la necesidad de generar anticuerpos para cada péptido encarece, alarga el desarrollo del método analítico y reduce el número de péptidos/proteínas que se pueden analizar. Por otro lado, los desarrollos tecnológicos están aumentando continuamente los límites de detección de los instrumentos (114).

3.4. Métodos con marcaje

3.4. a. Marcaje metabólico: SILAC

Como todos los métodos de cuantificación mediante marcaje, pretende introducir una etiqueta diferencial en cada muestra para permitir el multiplexado en las etapas de preparación y análisis. El objetivo del marcaje metabólico es introducir esta etiqueta de manera natural en los aminoácidos constituyentes de las proteínas que se van a analizar. Inicialmente se realizó creciendo las células en medio mínimo conteniendo $^{15}\text{N}/^{13}\text{C}$ de manera similar a lo que se hace para RMN (115). Esta estrategia es sencilla y relativamente poco costosa. Sin embargo, este procedimiento tiene varias limitaciones. Por un lado, no todas las células pueden crecer en estas condiciones de medio mínimas ya que las condiciones de cultivo de las células eucariotas son más complejas. Por otro lado, la diferencia entre péptidos marcados y sin marcar será diferente en función de la composición de cada péptido y es desconocida hasta que se identifica el péptido, lo que dificulta el análisis. Más tarde, la estrategia desarrollada por el grupo de Matthias Mann se basó en introducir en el medio de cultivo aminoácidos esenciales marcados. El marcaje de las proteínas se realiza progresivamente a medida que las

células cultivadas en estas condiciones van incorporando aminoácidos marcados en las proteínas de nueva síntesis. El empleo inicial de leucina deuterada mostró algunos inconvenientes. Uno de ellos es que el marcaje de los péptidos que se produce es incompleto debido a que no todos los péptidos generados contienen leucina. Además, los análogos deuterados se separan en cromatografía de los compuestos no deuterados, lo que complica bastante el análisis (116). Más tarde se pasó a introducir la marca en lisina (117, 118) y arginina (116), e incluso otros aminoácidos (metionina: (119); tirosina: (120)) con lo que se asegura que todos los péptidos tripticos producidos tendrán marca. En ocasiones se han detectado problemas con la arginina y la prolina (121). Pese a los posibles problemas, el SILAC es una de las estrategias cuantitativas más utilizadas actualmente. Incluso se han llegado a marcar organismos enteros (122) como moscas e incluso ratones (123) o plantas (124). Cada vez se están desarrollando nuevas aplicaciones para el SILAC, como el Super-SILAC que se utiliza para crear miles de estándares para analizar muestras humanas (125).

La principal ventaja del SILAC es que la etiqueta introducida no es ningún grupo o modificación química que se introduzca en la proteína. Son los propios aminoácidos de la proteína los que contienen los isótopos por lo que no se altera ninguna de las propiedades físicas o químicas de los péptidos producidos. Por tanto, no se espera que haya ningún tipo de cambio cromatográfico o en sus propiedades de ionización o fragmentación. La otra gran ventaja de la técnica es que al realizarse el marcaje durante el cultivo, las muestras a comparar se pueden combinar justo al inicio de la preparación de las muestras por lo que se reduce al máximo las posibles diferencias debidas al procesamiento de la muestra. Sin embargo, es preciso optimizar el marcaje para cada tipo celular empleado para asegurarse de que las proteínas se marcan en su totalidad. Por ejemplo, es conveniente determinar si las células utilizadas son capaces de sintetizar arginina o no ya que este aminoácido no es esencial. En caso afirmativo, se debe establecer la concentración de arginina mínima necesaria para suprimir la síntesis *de novo* de este aminoácido (121). Lo normal es dializar los sueros empleados en los medios de cultivo para evitar contaminaciones de aminoácidos no marcados presentes en éstos. El problema es que hay células que no crecen bien en suero dializado. Una posible solución es mezclar un poco de suero no dializado con dializado. Aunque se puede producir una pequeña contaminación con isótopos ligeros en la muestra pesada, se puede corregir posteriormente si se cuantifica esta contaminación (126). El marcaje total suele conseguirse con un mínimo 5 pases. Finalmente, otro inconveniente del método SILAC es que el procedimiento de marcaje puede afectar al cultivo, ya que el medio conteniendo aminoácidos marcados nunca es exactamente igual al original, y los cultivos celulares con frecuencia son muy sensibles a pequeñas variaciones en la composición de los medios.

La metodología SILAC es un ejemplo de cuantificación en modo MS. Los espectros de los experimentos de SILAC muestran dobletes o tripletes para cada péptido, según el diseño utilizado. Hay ocasiones en las que se utiliza una combinación de átomos ligeros y pesados en el mismo aminoácido. Por ejemplo, la arginina puede tener 6 átomos de ^{13}C y 4 de ^{15}N , en cuyo caso la diferencia es de 10 Da. Otra opción es utilizar una arginina con sólo los átomos de C pesados de manera que la diferencia será de 6 Da. Se podría, por tanto, diseñar un experimento en el que se marcaran tres tipos celulares distintos y después observar tripletes separados 6 y 10 Da respecto al residuo sin marcar. En cualquier caso, el análisis es complejo porque hay bastantes



posibles combinaciones (sobre todo si no se usa tripsina) y porque se puede producir un solapamiento de las envolventes isotópicas. Por tanto, es necesario utilizar un instrumento de alta resolución para obtener buenos resultados. Existen distintos paquetes informáticos para analizar los experimentos de SILAC (ver Tabla 2). La cuantificación de los péptidos puede hacerse de dos maneras. Por un lado se pueden usar las intensidades (o las áreas) de los picos ligeros y pesados de los espectros de masas para calcular los ratios o, por otro lado, se puede obtener el XIC de cada uno de ellos e integrar el pico cromatográfico y tener así la información cuantitativa.

3.4. b. Marcaje químico: iTRAQ/TMT

Las estrategias de marcaje químico con isótopos estables tienen la ventaja de ser muy flexibles en el sentido de que cualquier compuesto capaz de reaccionar con grupos funcionales de una proteína y que sea susceptible de ser sintetizado con isótopos estables puede ser utilizado para cuantificar. Al igual que sucede con el SILAC, las estrategias de marcaje químico más convencionales introducen dos versiones de un mismo reactivo, pesada y ligera, que el espectrómetro de masas es capaz de diferenciar en modo MS. Sin embargo, el empleo de etiquetas de distinta masa (no isobáricas) tiene algunos inconvenientes. Por un lado, la complejidad de la muestra aumenta ya que existe una versión pesada y otra ligera de cada péptido. Además, esta duplicación afecta inversamente a la intensidad de los picos correspondientes a cada péptido ya que se divide a la mitad. Frente a esta estrategia se han desarrollado una serie de reactivos específicos para cuantificar por espectrometría de masas que evitan estos inconvenientes. Se trata de las etiquetas isobáricas, diseñadas para cuantificar en modo MS/MS o MS², como el iTRAQ (*isobaric Tag for Relative and Absolute Quantitation* (48)) o el TMT (*Tandem Mass Tags* (49)). El diseño de este tipo de compuestos es tal que las distintas versiones de los compuestos tienen todas virtualmente la misma masa (de ahí su denominación de isobáricas). Sin embargo, cuando el péptido etiquetado se fragmenta en el espectrómetro de masas, la etiqueta también se fragmenta generando iones reportero cuya masa es específica para cada versión del compuesto (Figura 3). La abundancia relativa que cada péptido fragmentado tiene en las distintas muestras se puede determinar midiendo las áreas de los distintos iones reportero en el espectro de fragmentación.

La cuantificación en MS/MS presenta una serie de ventajas e inconvenientes respecto a la cuantificación en MS que conviene tener en cuenta a la hora de diseñar los experimentos de cuantificación. Impone restricciones en cuanto a los instrumentos que se pueden utilizar. Así, los péptidos marcados con etiquetas isobáricas no pueden ser analizados con un MALDI-TOF o una trampa iónica tridimensional, pero sí en un MALDI-TOF/TOF o en un híbrido cuadrupolo-TOF ó un híbrido cuadrupolo-orbitrap. Si bien la cuantificación en modo MS tiene una mayor intensidad de señal, los espectros son más complejos y la relación S/N suele ser menor. Por el contrario, la cuantificación en MS/MS suele proporcionar una señal más sencilla y con una S/N mucho mayor. Por ello, la cuantificación es bastante sensible. Es cierto que puede haber impurezas o fragmentos que podrían solaparse con los iones reporteros, pero estos efectos introducen mucha menos variabilidad que la variabilidad biológica y no suponen desviaciones significativas (127, 128). Los instrumentos de alta resolución pueden eliminar los problemas de solapamiento de picos con m/z similares en modo MS y mejorar

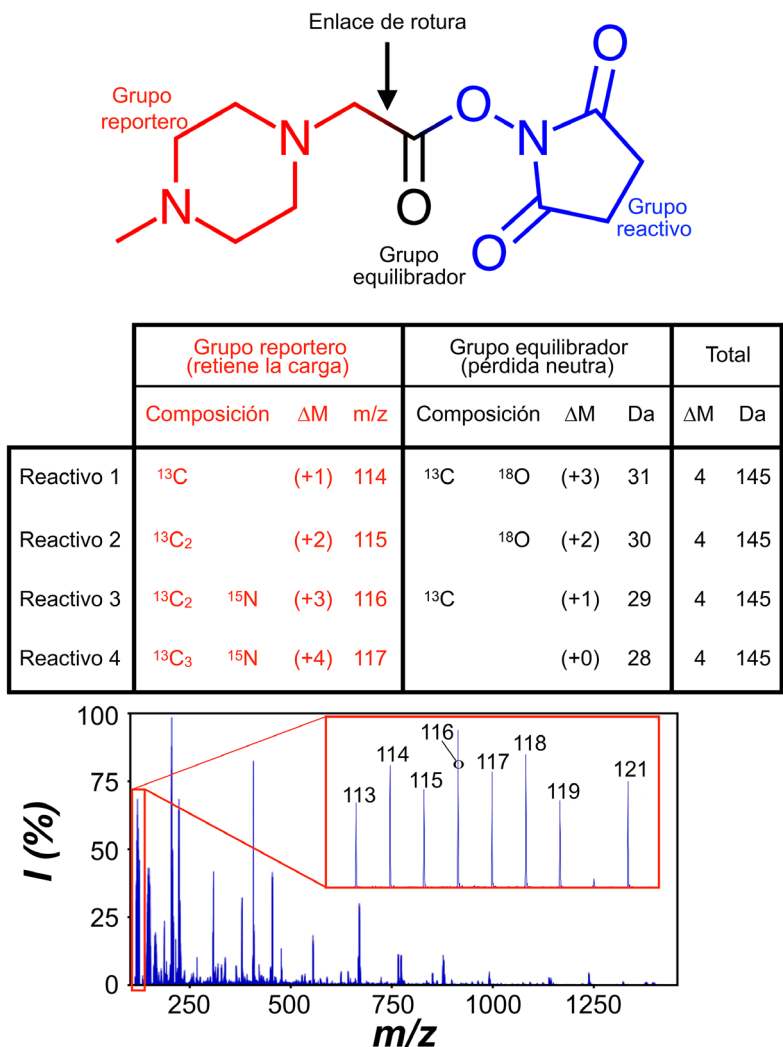


Figura 3. Marcaje isobárico iTRAQ. En el panel superior se muestra la estructura del reactivo iTRAQ-4plex. En rojo, negro y azul se indican el grupo reportero, equilibrador y reactivo, respectivamente. En el panel del medio se indica la composición isotópica de los grupos reportero y equilibrador. En el panel inferior se muestra un espectro de fragmentación de un péptido marcado con reactivos iTRAQ-8plex del que se puede deducir su secuencia. La zona del espectro correspondiente a los iones reportero se muestra ampliada en el recuadro rojo.

considerablemente la cuantificación. Sin embargo, la selección de los iones que se van a fragmentar no se puede hacer con tanta resolución. Por tanto, es frecuente que algunos iones contaminantes con m/z similares sean también seleccionados para fragmentar. Puesto que tanto el péptido seleccionado como el contaminante generan los mismos reporteros, la cuantificación puede verse comprometida. Dado que la mayoría de los péptidos no son diferenciales, esta contaminación tiende a amortiguar las diferencias. Este problema de contaminación parece ser la principal causa del apantallamiento de las diferencias experimentales esperadas (129). Obviamente, a medida que aumenta la complejidad de la muestra, el efecto de la contaminación es más aparente. Por ello, un mayor fraccionamiento contribuye a reducir el efecto de apantallamiento.



Se han propuesto distintas estrategias para corregir este efecto que van desde la adquisición de los espectros (129), trabajar en modo MS³ (130) a correcciones en el análisis de los datos (131).

Una aparente limitación de la cuantificación en MS/MS es el número de estos espectros que se obtienen en un experimento normal de LC-MS/MS. Si bien todos los picos observables en MS podrían ser susceptibles de ser cuantificados, bien mediante técnicas *label-free* o de marcaje isotópico no isobárico, sólo un porcentaje relativamente reducido de espectros MS/MS son utilizables tanto para identificación como para cuantificación (60). A pesar de ello, la comparación de iTRAQ y otras estrategias de marcaje no isobárico ha mostrado que el número de péptidos y proteínas identificados es igual o mayor utilizando iTRAQ (132). Esto se debe, fundamentalmente, a dos factores. Por un lado, los péptidos de todas las muestras contribuyen a la misma señal (132, 133), lo que supone una ventaja para identificar y cuantificar péptidos poco abundantes. Por otro lado, el aumento en la complejidad de la muestra al introducir un marcaje isotópico no isobárico hace que se produzca una mayor redundancia en la fragmentación. Es decir, se fragmenta más veces un mismo ión porque hay más versiones del mismo (pesada y ligera, por ejemplo), por lo que se produce un menor muestreo del proteoma (132), lo que no ocurre en el marcaje isobárico. Finalmente, tanto iTRAQ como TMT son reactivos diseñados para mejorar la fragmentación de los péptidos, al introducir aminas terciarias que confieren una carga positiva adicional a los péptidos.

A pesar de sus limitaciones, la cuantificación con iTRAQ es muy precisa y sensible aunque su exactitud sea inferior a la de otras estrategias (134). Además, actualmente es la estrategia experimental que permite un mayor multiplexado, analizando un mayor número de muestras simultáneamente, hasta 8 muestras. Asimismo, a diferencia de las técnicas de marcaje metabólico, la metodología iTRAQ permite su utilización tanto en muestras procedentes de experimentos *in vivo* como *in vitro*. Actualmente la tecnología TMT ha superado iTRAQ permitiendo un multiplexado de hasta 11 muestras.

Las etiquetas isobáricas son caras y los protocolos de marcaje son relativamente laboriosos. Por tanto, hay que trabajar de manera muy cuidadosa para obtener resultados buenos y reproducibles (127). Tanto los reactivos de iTRAQ como las nuevas versiones de los TMT utilizan derivados de ésteres de N-hidroxisuccinimida (NHS) que reaccionan con los grupos amino de los péptidos (N-terminal y lisina). Estos ésteres se hidrolizan espontáneamente en solución acuosa, sobre todo a pH básico. Puesto que la reacción de marcaje compite con la reacción de hidrólisis, se debe utilizar unas condiciones experimentales que favorezcan la reacción de marcaje sobre la de hidrólisis. Ésta es la razón por la que el reactivo iTRAQ se prepara en etanol o isopropanol y su concentración de trabajo no debe bajar de un 70 % ó 60 %, respectivamente. Asimismo, es importante mantener el pH de la mezcla de reacción entre 7,5 y 8,5 para que la reacción funcione correctamente. Este pH es una situación de compromiso que permite una reactividad del grupo amino adecuada (dependiente de la forma desprotonada) con una hidrólisis aceptable. A medida que las reacción de marcaje e hidrólisis avanzan, los productos que se van formando acidifican el medio. Por ello, es necesario que el medio de reacción tenga suficiente capacidad tamponadora para mantener el pH. Por razones obvias, hay que eliminar las sustancias que interfieran con la reacción de marcaje, fundamentalmente aminas primarias y secundarias, por

lo que es frecuente precipitar las proteínas en las etapas iniciales del proceso. Aunque la reacción es bastante específica para grupos amino, los grupos tioles a altas concentraciones pueden mostrar cierta reactividad. Por ello, se recomienda sustituir el DTT por TCEP (tris(2-carboxietil)fosfina) como agente reductor ya que no introduce tioles. Al final del proceso de marcaje, la mezcla de reacción tiene una elevada concentración de sales que puede interferir con los procesos de separación de péptidos como el isoelectroenfoque o la cromatografía de intercambio catiónico. Para evitar interferencias suele realizarse una etapa de desalado en un cartucho o punta de pipeta con C18 al concluir el marcaje. Los péptidos marcados, poseen unas características de hidrofobicidad diferentes a los no marcados. En consecuencia, puede ser necesario modificar los gradientes cromatográficos utilizados habitualmente con péptidos no marcados. A la hora de adquirir los datos, hay que aumentar la energía de colisión para obtener buena señal de los iones reporteros. Sin embargo, hay que evitar aplicar demasiada energía que pueda disminuir la señal de los fragmentos del péptido y se pierda eficiencia en la identificación de proteínas. En los instrumentos de ABSCIEX, que es la empresa que comercializa los reactivos iTRAQ, está bastante optimizado, pero se ha utilizado con éxito con instrumentos de otras casas comerciales. Los iones precursores se deben seleccionar con una ventana lo más estrecha posible que permita una adecuada transmisión. De esta manera se reduce en cierta medida la contaminación por otros iones con m/z similar, la principal causa del apantallamiento de los resultados.

3.4. c. Marcaje enzimático con $^{16}\text{O}/^{18}\text{O}$

El marcaje isotópico con ^{18}O es un marcaje enzimático realizado a nivel de péptido que consiste en el intercambio de dos átomos de ^{16}O por dos átomos de ^{18}O durante una reacción catalizada por tripsina, aumentando así el peso molecular del péptido en 4 Da (135). Durante el proceso de marcaje se generan dos nuevas especies químicas que contienen uno (B_1) o dos (B_2) átomos de ^{18}O en el extremo C-terminal del péptido, además de la especie química no marcada (B_0) (Figura 4). El marcaje se realiza incubando los extractos de proteínas con tripsina en condiciones que permiten la digestión, conteniendo el medio de reacción H_2^{18}O para la muestra a marcar y H_2^{16}O para la muestra control. Aunque originalmente el marcaje enzimático se llevaba cabo durante la digestión triptica, existen diversos factores, como valores extremos de pH,

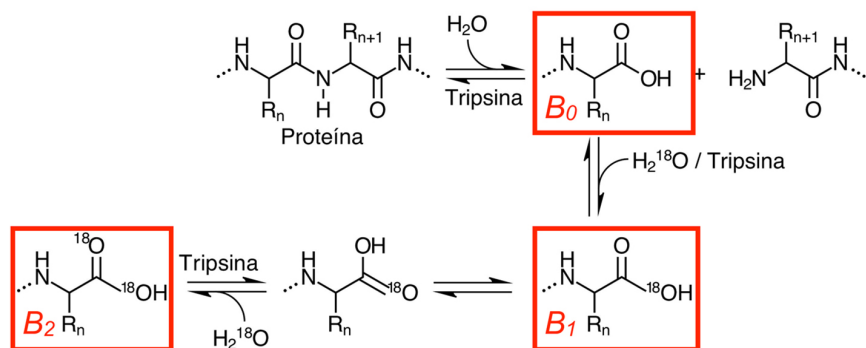


Figura 4. Esquema de la reacción enzimática de marcaje con $^{16}\text{O}/^{18}\text{O}$ catalizada por tripsina. En un primer paso se rompe el enlace peptídico mediante una reacción catalizada por tripsina en presencia de H_2^{16}O , generando la especie química B_0 , con dos átomos de ^{16}O en el extremo C-terminal de péptido. En un segundo paso, la tripsina permite la incorporación de uno o dos átomos de ^{18}O en el extremo C-terminal, generando dos nuevas especies químicas: B_1 , con un átomo de ^{18}O ; y B_2 , con dos átomos de ^{18}O .



que causan un intercambio de oxígenos con el medio (136). Hoy en día los procesos de digestión y marcaje se llevan a cabo de forma separada, cada uno de ellos en sus condiciones óptimas. (135, 137, 138).

Tradicionalmente la cuantificación se realizaba considerando que la proporción de péptidos en la muestra no marcada (A) corresponde a la fracción de especies químicas no marcadas, mientras que los péptidos de la muestra marcada (B), da lugar a especies mono- (B_1) y di- marcadas (B_2), calculándose el ratio como $A/(B_1 + B_2)$ (138-142). Pero durante el intercambio de oxígenos que tiene lugar en la reacción enzimática de marcaje, sólo la mitad de la fracción de especies mono-marcadas (B_1) da lugar a especies di-marcadas (B_2) o volverá al estado no marcado (B_0) (Figura 4) (52, 143-145), determinándose la abundancia relativa de un péptido en cada muestra como $(A+B_0)/(B_1+B_2)$.

Mediante consideraciones cinéticas puede demostrarse que la concentración total de péptidos de la muestra marcada, $B = B_0 + B_1 + B_2$, puede descomponerse en función de la concentración de cada una de las especies de la muestra marcada, utilizando un parámetro f que evalúa la eficiencia de marcaje (51)

$$\begin{aligned} B_2 &= B \cdot f^2 \\ B_1 &= B(B \cdot f)^2 \\ B_0 &= B \cdot 2 \cdot f(1-f) \end{aligned}$$

La relación entre las concentraciones de las tres especies químicas del péptido siempre sigue estas tres ecuaciones, independientemente del porcentaje de enriquecimiento con ^{18}O del agua utilizada y del posible intercambio de oxígenos en presencia de agua no marcada. En la Figura 5 se muestra un esquema de la contribución de cada una de las especies en un espectro real. El control riguroso de la eficiencia de marcaje permite determinar directamente las proporciones de A y B y controlar el marcaje de cada uno de los péptidos cuantificados en un mismo análisis.

La cuantificación de cada especie del péptido se calcula en función de la abundancia (intensidad/área) de cada uno de los picos correspondientes a las envolturas isotópicas determinadas por el detector del espectrómetro de masas. El ratio de los cinco primeros picos de cada envoltura isotópica se calcula como:

$$\text{ratio} \left(\frac{^{16}\text{O}}{^{18}\text{O}} \right) = \frac{I_0}{I_4 - \frac{M_4}{M_0} I_0 + I_2 \left(1 - \frac{M_2}{M_0} \right) - \left(1 - \frac{M_2}{M_0} \right) \frac{M_2}{M_0} I_0}$$

donde I_0 , I_2 y I_4 son las intensidades relativas medidas del primer, tercer y quinto pico de cada envoltura isotópica, y M_0 , M_2 y M_4 corresponden a las intensidades relativas teóricas de la envoltura isotópica del péptido. Este ajuste teórico (M_i), de acuerdo con la abundancia relativa de los isótopos C, O, H, N y S existentes en la naturaleza, se determina en función de la secuencia del péptido (52).

Un problema de la proteómica cuantitativa basada en MS es la resolución. La dife-

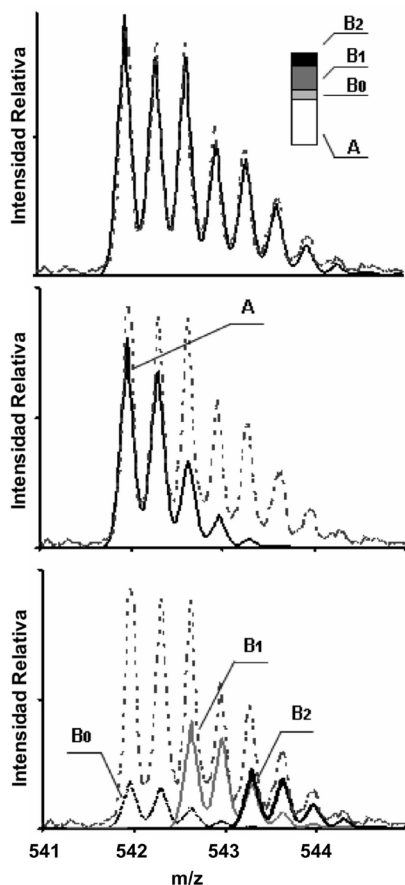


Figura 5. Estimación de la eficiencia de marcaje con $^{16}\text{O}/^{18}\text{O}$. Contribución a la envoltura isotópica de cada una de las especies químicas del péptido (A: especie proveniente de la muestra no marcada; B0, B1 y B2: especies provenientes de la muestra sometida a marcaje en la que se han incorporado ninguno, uno o dos átomos de ^{18}O , respectivamente). La envoltura isotópica se descompone en la suma de los cuatro componentes, cuyo valor está definido inequívocamente por la concentración de péptido en la muestra no marcada, A, la de péptido en la muestra marcada, B, y la eficiencia de marcaje, f .

rencia de masa entre la especie marcada y la especie no marcada es muy pequeña (4 Da para un marcaje completo y 2 Da para un marcaje incompleto). Además, cada especie peptídica presenta tres picos isotópicos principales separados $1/z$ unidades. Es por ello que normalmente se usan equipos de media y alta resolución, como los analizadores de masas FT-ICR (52, 135, 142, 143, 145, 146) o los analizadores de tiempo de vuelo (TOF) (147-153).

En los equipos de baja resolución, como la trampa iónica (IT), es difícil diferenciar y cuantificar la envoltura isotópica de los picos, sobre todo en péptidos iónicos doble o triplemente cargados (154). Aunque existen algunos trabajos donde usan este tipo de equipos (140, 150, 155, 156), no han sido utilizados de forma habitual en proteómica cuantitativa hasta la aparición de la trampa iónica lineal (LTQ), que presenta una velocidad de barrido supe-

rior a la trampa iónica tridimensional convencional (3D-IT) (157), y permite modos de barrido de alta resolución muy robustos en un rango de masas limitado (ZoomScan). En una trampa iónica lineal se pueden programar ciclos de barrido que realicen dos espectros, un ZoomScan centrado alrededor del doblete de envolturas isotópicas que se desea cuantificar y un espectro MS/MS para identificar el péptido (144, 158-161).

3.4. d. Estadística de cuantificación por marcaje isotópico

El número de datos que deben estudiarse en experimentos de proteómica diferencial a gran escala es formidable. Esto hace que uno de los problemas principales de la proteómica cuantitativa a gran escala sea el análisis de datos.

Uno de los cuellos de botella fundamentales es ¿cómo determinar si la diferencia en abundancia de una proteína entre las muestras que se comparan es real, o si se ha producido de forma artefactual? Los cambios de expresión en proteínas pueden considerarse significativos si son cambios mayores que un cierto umbral fijo que se establezca a priori, pero es preferible utilizar métodos estadísticos que establezcan la significatividad estadística basándose en criterios de probabilidad.



En este sentido, los métodos estadísticos aplicados en proteómica cuantitativa están claramente menos desarrollados que los que se utilizan en el campo de microarrays. La mayoría de las publicaciones de proteómica cuantitativa mediante espectrometría de masas han heredado los métodos ya establecidos en geles de electroforesis bi-dimensional, sin comprobar si estos métodos son realmente aplicables a las aproximaciones de MS/MS. Todos estos tests asumen un muestreo independiente y que los datos analizados siguen una distribución normal, suposiciones que no han sido demostradas en un contexto general.

Se emplean los Student's *t*-tests o ANOVA que calculan los valores *p* de la cuantificación de cada proteína de manera independiente. Es decir, la probabilidad de que la cuantificación de dicha proteína pertenezca a la distribución normal del resto de las proteínas del experimento solamente debido al azar (mientras mayor es la *p*, menos confiable es el cambio en abundancia de dicha proteína). Se consideran estadísticamente significativos aquellos resultados que tengan un *p* menor que un cierto umbral de error tolerable. Aunque dicho umbral se selecciona de forma bastante arbitraria, suele emplearse una $p < 0.05$. Esto significa que existe un 5% de probabilidad de que una proteína que se expresa de forma diferencial entre dos muestras, lo haga sólo por azar.

Sin embargo, en un experimento de proteómica de alto rendimiento, mientras mayor es el número de proteínas cuantificadas a las cuales se les aplica el test estadístico de forma independiente, mayor será el número de valores que cumplirán con el criterio de significatividad sólo por azar, y por tanto, mayor será la incidencia de los llamados falsos positivos (falsos cambios de abundancia de proteína). Por ejemplo, en un experimento con 10 000 proteínas cuantificadas, a una $p < 0.05$, el 5% de los cambios será sólo por azar, lo cual correspondería a 500 proteínas y sería un resultado inaceptable. Para evitar esto, cuando se trabaja con grandes poblaciones de datos, a los valores de *p* de cada cuantificación se les realiza una corrección de tests múltiples (*Multiple testing correction*) para mantener la tasa de error global por debajo de un umbral seleccionado por el investigador. Existen diferentes formas de corrección, pero la más aceptada es la FDR (*False Discovery Rate*) de Benjamini y Hochberg (162), la cual representa la proporción de valores esperados que se desvían del resto de la población de proteínas (*outliers*) sólo al azar frente al número total de *outliers* (163). Por ejemplo, en un experimento con 10 000 proteínas cuantificadas, al cortar con $p < 0.001$, se observan 200 proteínas que cambian en abundancia. Si se controla la FDR, se obtiene que el 5% de esos cambios se debe simplemente al azar ($10\,000 \times 0.001 = 10$; $10 \times 100\% / 200 = 5\%$), lo cual correspondería solamente a 10 proteínas.

Por otro lado, como se ha descrito anteriormente, en la mayoría de los métodos de marcaje isotópico, las mediciones en el espectrómetro de masas se realizan a nivel de péptido, a partir de los cuales se cuantifica la proteína que dio lugar a esos péptidos. Sin embargo, existen una serie de factores que pueden complicar este análisis, como por ejemplo, que una misma proteína se cuantifica a través de diferentes péptidos; un mismo péptido puede pertenecer a diferentes proteínas; a su vez un mismo péptido puede ser identificado con varias cargas, o con modificaciones post-traduccionales o en diferentes fracciones después de un proceso de separación multidimensional. Frecuentemente se presentan co-eluidos o el péptido en cuestión se identifica mal. Además, las diversas aproximaciones de marcaje isotópico pueden requerir el uso de

Software	Métodos	Resolución o MS	Buscador	Estadística	Publicación
APEX	Spectral Counting	Cualquiera (TPP)	Mascot, SEQUEST (TPP)	No comparable porque es label-free.	(66, 70)
ASAPratio (TPP)	ICAT, cICAT, SILAC	Cualquiera (TPP)	TPP	Utiliza el Interact de TPP para hacer el rolling-up, pero no tiene en cuenta los valores de cuantificación.	(172)
Libra (TPP)	iTRAQ	Cualquiera (TPP)			No publicado
pepXML2Excel	Exportador de resultados	Cualquiera (TPP)			No publicado
XPRESS (TPP)	ICAT	Baja resolución (datos exportados a TPP)			(176)
MaxQuant	SILAC	Orbitrap, FT (alta resolución)	Cualquiera	Rolling-up péptido-proteína por redundancia peptídica. Cuantificación por criterio FDR Benjamini-Hochberg.	(177)
MFPaQ	Label-free, SILAC, ICAT	No especificado (importa datos mgf)	Mascot	Criterios de exclusión basados en umbrales. Incluye un tratamiento de péptidos identificados en fracciones consecutivas de un gel.	(178)
MSQuant	SILAC, ¹⁵ N (con ayuda de otro SW)	Orbitrap, FT (alta resolución)	Mascot	Se calcula un X ² , y el rolling-up se hace con medianas de los X ² de los péptidos.	(165, 179)
MS-Spectre	Label-free	Alta resolución (datos importados: mzML, mzXML, MZData)	No integrado	No comparable porque es label-free.	No publicado
Multi-Q	iTRAQ	Cualquiera	Cualquiera, preferentemente SEQUEST (pRatio)	Su rolling-up no tiene en cuenta el error sistemático cometido en cada nivel.	(167)
muxQuant	¹⁵ N	FTICR	TPP	Uso de medianas para rolling-up.	(180)
ProRata	¹⁵ N, ¹³ C, SILAC, ICAT, ¹⁸ O	Thermo (LCQ-DECA, LTQ, LTQ-FT)	SEQUEST	Maximum likelihood. Basado en pesos estadísticos proporcional a SN ratio. No tiene en cuenta el error sistemático cometido en cada nivel. No controla la eficiencia.	(169)
PVIEW	Label-free (XIC)	Alta resolución	X!Tandem (y posiblemente otros, mzXML)	No comparable porque es label-free.	(181)
Quant	iTRAQ	Cualquiera	Mascot, SEQUEST	No hace realmente rolling-up hasta proteína. Cuantifica péptidos.	(174)
QUIL	cICAT, SILAC, ¹⁸ O	Thermo (LCQ-DECA, LTQ, LTQ-FT)	SEQUEST	Detección automática de anchos de pico, eliminación de background. Prevención de coeluidos. No estima la eficiencia en ¹⁸ O.	(182)
RAAMS	¹⁸ O	FT	SEQUEST (no especificado)	Estimación de la eficiencia similar a (51). Sin rolling-up.	(145)
RelEx	¹⁵ N, ICAT	Thermo (LCQ-DECA, LTQ, LTQ-FT)	SEQUEST	Test de Dixon para identificar cuantificaciones atípicas. Utiliza p-values, no válido para alto rendimiento.	(183)
Mascot Distiller	Cualquiera basado en marcaje isotópico	Cualquiera (no todos probados por los autores)	Mascot	Sin estadística visible.	No publicado
Proteome Discoverer (Thermo)	iTRAQ, TMT, SILAC, ICAT	Thermo (LCQ-DECA, LTQ, LTQ-FT)	SEQUEST, Mascot y Z-core. Combinación de todos los motores de búsqueda, aunque FDR mal calculada.	Uso de medallas para rolling-up. Diversos errores en desestimación manual.	No publicado
SanXoT	Cualquiera basado en marcaje isotópico	Cualquiera (probado en LTQ, Orbitrap y MALDI-TOF-TOF)	Cualquiera (SEQUEST y Mascot son transformados directamente, otros son mapeados por cualquier tabla de datos)	Ver texto.	(158)

Tabla 2. Comparativa de software de proteómica cuantitativa



espectrómetros de masa de diferente resolución y con diferentes modos de fragmentación. En la Tabla 2 se muestra una perspectiva del software que puede encontrarse actualmente en proteómica cuantitativa.

Para obtener la cuantificación a nivel de proteína, se ha empleado la media de las cuantificaciones de sus péptidos (164), así como la mediana (Proteome Discoverer, MSQuant (165)). También se ha utilizado la media ponderada (166, 167) y la correlación ponderada (53). La ponderación se realiza multiplicando la cuantificación de cada péptido por un peso estadístico, el cual generalmente depende de la intensidad del espectro. Sin embargo, para decidir la mejor forma de integrar la información cuantitativa desde el nivel de espectro, pasando por el nivel de péptido y hasta el nivel de proteína, es necesario analizar la distribución de datos a cada uno de los niveles. Es decir, el tipo de distribución, su simetría, si existen factores que alteran dicha distribución y la convierten en multimodal.

En la mayoría de los trabajos se asume que las cuantificaciones de los péptidos cumplen con una distribución normal, con una media y varianza homogénea. Esta asunción errónea de normalidad podría explicarse por el teorema del límite central: mientras mayor es el tamaño muestral, lo cual es típico en los experimentos de proteómica de alto rendimiento, más se asemeja la distribución de datos a una Gaussiana. Sin embargo, en un experimento de prueba, en el cual se comparó un proteoma contra sí mismo por marcaje con ^{18}O , se demostró que aunque la distribución es aparentemente simétrica, siguiendo la forma de una distribución normal, al analizar la llamada "gráfica de normalidad" (que representa el valor normalizado observado de la variable respecto al valor normalizado teórico y que en casos de distribución normal pura deben producir una recta) (158), se hace patente una clara desviación del comportamiento normal (Figura 6). De hecho, el comportamiento en los extremos de la función representada en la Figura 6C indica que la distribución estudiada podría ser fruto de la superposición de diversas distribuciones normales con distintas varianzas (168), cuyo resultado no es una distribución Gaussiana. La razón de esta heterogeneidad de varianza de las cuantificaciones se debe a que un mismo péptido puede cuantificarse

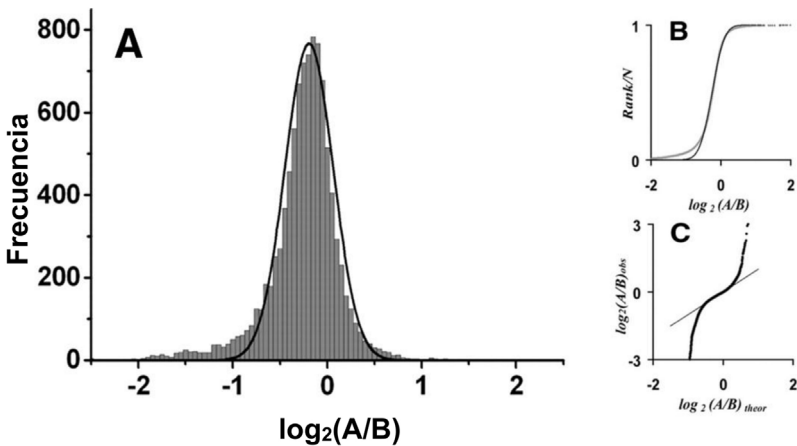


Figura 6. Análisis de los resultados de un experimento de prueba asumiendo que en la hipótesis nula los datos se distribuyen normalmente (158). (A) Histograma de frecuencias de los ratios (expresados en logaritmo en base 2), y ajuste del histograma a una distribución gaussiana. (B) Distribución acumulativa de frecuencias (en gris), mostrando la curva teórica correspondiente al mejor ajuste de los datos a una distribución normal (en negro). (C) Gráfica de normalidad (la línea delgada representa los resultados esperados para una distribución normal).

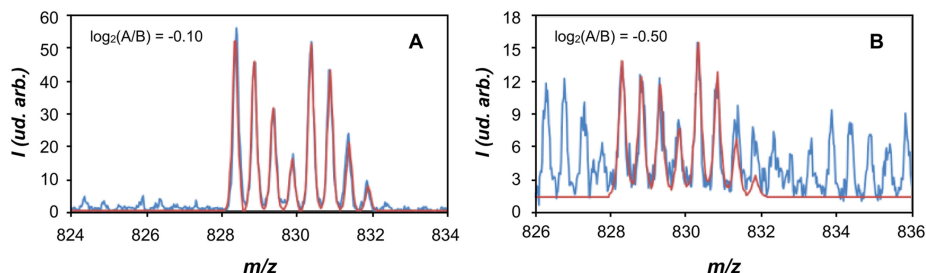


Figura 7. Calidad de los espectros. En el panel (A) se observa un espectro cuya una relación señal/ruido es alta, en el que la cuantificación es muy precisa. En (B) se observa una cuantificación del mismo péptido en un espectro con una relación señal/ruido muy inferior, en el que la cuantificación es mucho menos precisa y por tanto se lleva a cabo con mayor margen de error. Nótese que los valores de cuantificación difieren notablemente.

a través de espectros de diferente relación señal/ruido, y por tanto con diferente precisión (Figura 7) (131, 169, 170).

Aunque son pocos los autores que han abordado este problema de heterogeneidad de varianza a nivel de espectro, ya se han propuesto algunas soluciones. Por ejemplo, en el modelo WSPP (175) se ha introducido el concepto de peso estadístico del espectro (v_s), el cual aumenta con el cuadrado de la intensidad de la medida (reflejando

así el efecto intensidad) y disminuye con su desviación cuadrática media (reflejando así el efecto de la bondad de ajuste a la curva teórica). Una vez que se normalizan los espectros de acuerdo a su calidad, los datos agrupados en ventanas deslizantes de doscientas cuantificaciones siguen una distribución normal (Figura 8) (175). Por otro lado, en MultiQ (167) se intenta homogeneizar la varianza a nivel de espectro trabajando con cuentas espectrales por encima de un determinado umbral, lo cual a su vez compromete la profundidad de análisis del proteoma. Recientemente, a la intensidad de los espectros de iTRAQ se aplicó una transformación muy utilizada en el campo de los microarrays, como es la transformación logarítmica generalizada mediante el software VSN (*variance stabilization normalization*) (171). De esta forma se logró estabilizar la varianza, obteniendo homogeneidad de varianza en todo el rango de intensidades (131).

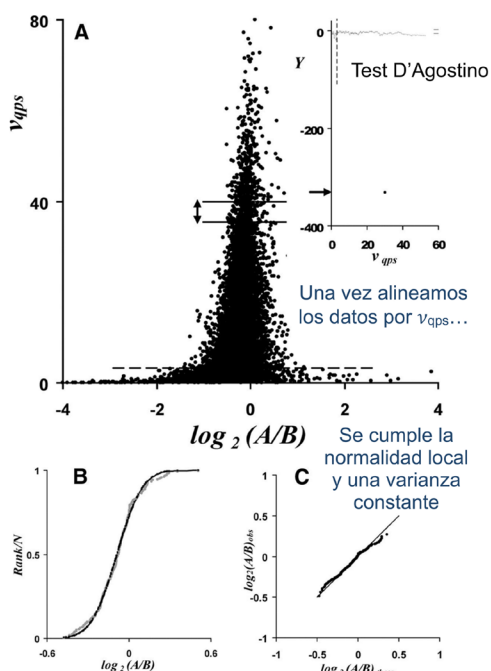


Figura 8. Estratificación de las medidas de cuantificación diferencial en poblaciones localmente normales de acuerdo a su peso estadístico (v_s). (A) Distribución de los pesos estadísticos en función de los $\log_2(\text{ratio})$. (A, interior) Resultados del test de normalidad de D'Agostino aplicado en ventanas deslizantes de 200 puntos ordenados por el peso estadístico; se muestran los valores del estadístico Y del test de D'Agostino, indicando los valores del 99% de confianza para una distribución normal de 200 datos (2.9 a -5.6); compárense estos valores con el valor de Y obtenido para el conjunto total de datos (indicado por un punto negro). (B) y (C) Gráficas de normalidad de una de las regiones de (A) tomada a modo de ejemplo (delimitada por barras horizontales).

Los demás modelos desarrollados para iTRAQ se centran sobre todo



en el análisis de la distribución de los datos a nivel de péptido. Así, ASAPratio (172) y GPS Explorer (173) asumen normalidad de datos después de eliminar *outliers* mediante diferentes filtros, mientras que ProQuant asume una distribución normal (174). Muy pocos trabajos tienen en cuenta posibles fuentes de artefactos como la eficiencia de marcaje (51, 145, 158, 167), la digestión parcial (158, 167) o la oxidación de metionina (158).

Actualmente no hay consenso en la literatura acerca de cómo tratar los datos provenientes de los péptidos. El modelo más avanzado descrito hasta la fecha es el modelo WSPP (175), el cual es el único que cuenta con una hipótesis nula a gran escala validada y se ha demostrado su aplicación en la práctica en centenares de experimentos usando el paquete SanXoT (158). En ese modelo jerárquico, el error final de la medición de una proteína se compone por la suma de los errores correspondientes a cada nivel de medición: espectro, péptido y proteína (Figura 9). La cuantificación a cada nivel se realiza a través de la media ponderada de las cuantificaciones correspondientes del nivel inmediato inferior, de manera que las mediciones más precisas

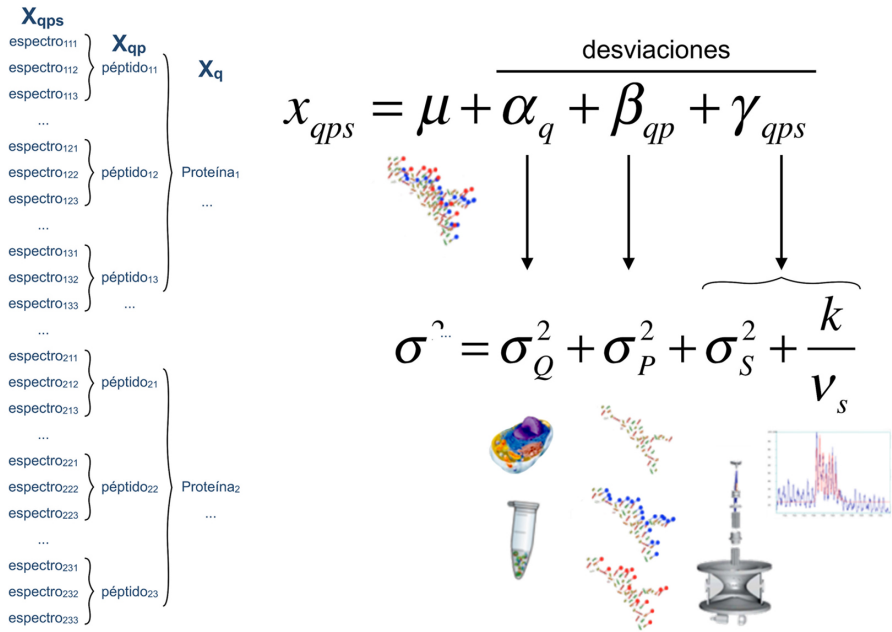


Figura 9. Consideración de las distintas fuentes de error de forma separada en el análisis de muestras marcadas por isótopos estables (158). X_{qps} , X_{qp} y X_q son las cuantificaciones a nivel de espectro, péptido y proteína, respectivamente.

pesan más en la media final (Figura 10).

Al expresar la medición a cada nivel en unidades de desviación estándar y al integrar a nivel de proteína, los datos siguen una distribución normal estandarizada, indicando que los errores se han estimado correctamente (Figura 11A, *inset*). En el experimento de hipótesis nula, de más de 1200 proteínas cuantificadas se obtuvo 1 cambio de expresión falso a una $FDR_{q < 5\%}$, mientras que la aplicación de los modelos estadísticos clásicos, donde se asume una distribución global normal de los datos, se obtenía un elevado número de falsos cambios de expresión, independientemente de los criterios

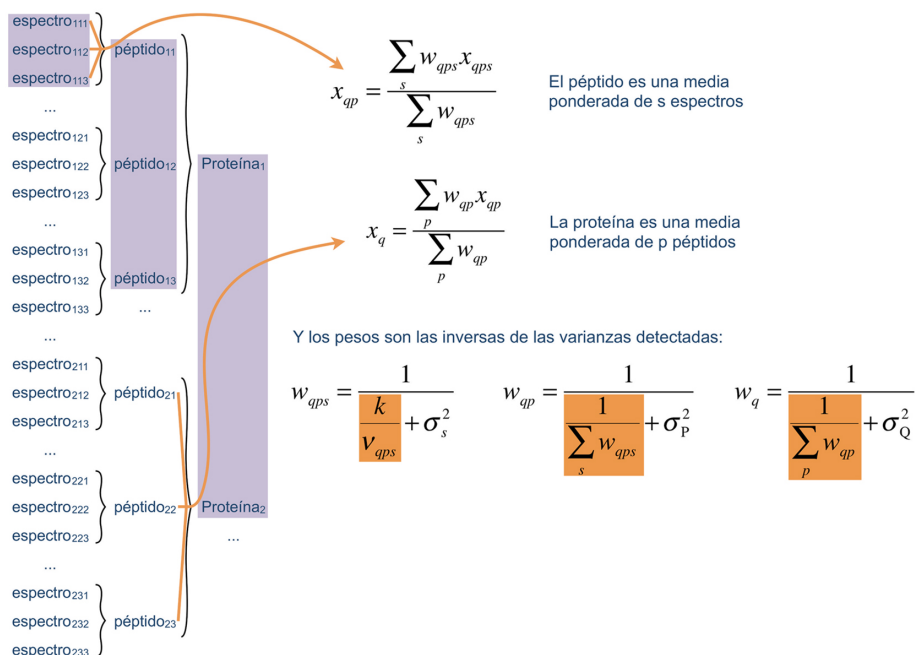


Figura 10. Integración de la información cuantitativa desde el nivel de espectro hasta el nivel de proteína según el modelo propuesto por (158). W_{qps} , W_{qp} y W_q son los pesos estadísticos (inversos de la varianza) de la cuantificación a nivel de espectro, péptido y proteína, respectivamente.

estadísticos de significatividad utilizados ($p < 0.05$, $FDR < 5\%$) (175).

Según el modelo WSPP (175), la varianza a nivel de espectro depende fundamentalmente del espectrómetro de masas en el que se ha realizado la medición, mientras que la varianza a nivel de proteína se ve influenciada por la forma de obtener la muestra proteica. Por otro lado, la dispersión a nivel de péptido depende del procedimiento de digestión de las proteínas y el marcaje de los péptidos resultantes. De esta forma es posible detectar artefactos debidos a la digestión parcial, a la oxidación de metioninas o a la eficiencia de marcaje, entre otros (175). Por ejemplo, como se observa en la Figura 12, los péptidos que tienen digestiones parciales o metioninas oxidadas tien-

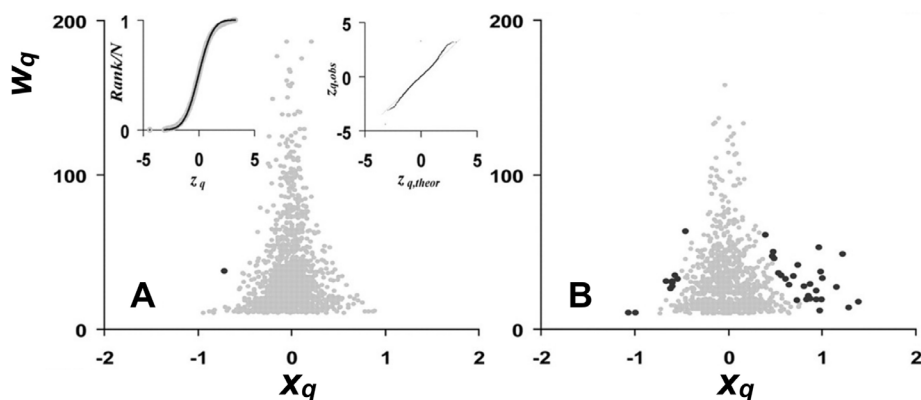


Figura 11. Análisis de (A) la hipótesis nula y (B) de una muestra tratada con VEGF durante 8 h con el modelo estadístico de (158). X_q y Z_q son los valores normal y estandarizado, respectivamente, de la cuantificación a nivel de proteína. W_q es el peso estadístico (inverso de la varianza) de la cuantificación de cada proteína. Los puntos negros representan las proteínas que cambian de expresión a una $FDR_q < 0.05$.



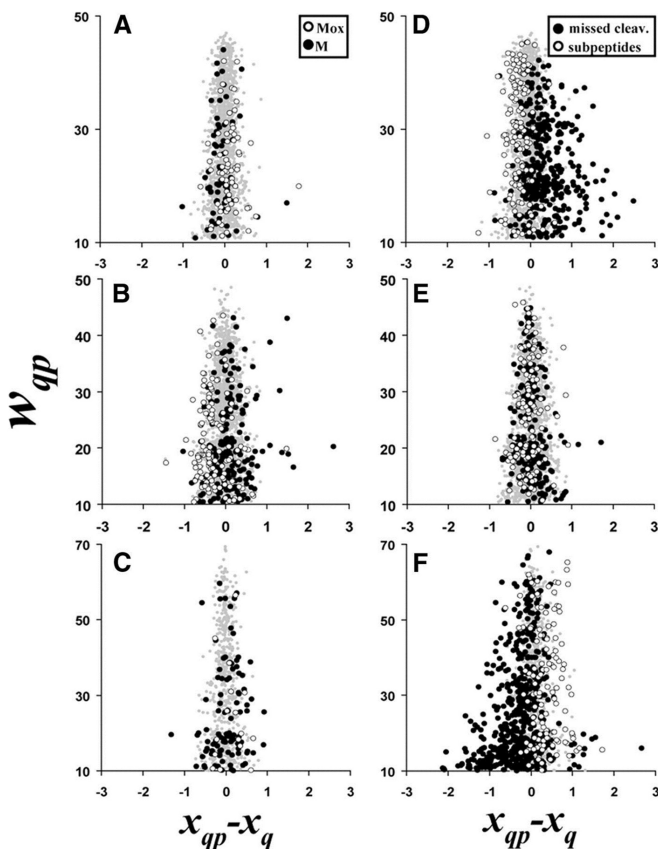


Figura 12. Efecto de (A-C) la oxidación de metioninas y de (D-F) la digestión parcial sobre la dispersión de los péptidos alrededor del valor de la proteína correspondiente (158).

den a desviarse de la relación 1:1 esperada, sugiriendo que estos péptidos no reflejan de manera confiable la concentración de la proteína de la que se derivaron. Por tanto, el análisis de los errores asociados a cada nivel de manera independiente permite controlar los posibles artefactos en el protocolo que van desde la preparación de la muestra hasta la obtención de datos en el espectrómetro de masas (175).

BIBLIOGRAFÍA

1. Ghaemmaghami S, Huh WK, Bower K, et al. Global analysis of protein expression in yeast. *Nature*. 2003;425:737-41.
2. Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell*. 2009;138:795-806.
3. Schneider LV, Hall MP. Stable isotope methods for high-precision proteomics. *Drug Discovery Today*. 2005;10:353-63.
4. Geiger T, Wehner A, Schaab C, Cox J, Mann M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics*. 2012;11:M111 014050.
5. Nilsson T, Mann M, Aebersold R, et al. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat Methods*. 2010;7:681-5.
6. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem*. 2007;389:1017-31.
7. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem*. 1975;250:4007-21.
8. Gorg A, Drews O, Luck C, Weiland F, Weiss W. 2-DE with IPGs. *Electrophoresis*. 2009;30 Suppl 1:S122-32.

9. Friedman DB, Hoving S, Westermeier R. Isoelectric focusing and two-dimensional gel electrophoresis. *Methods Enzymol.* 2009;463:515-40.
10. Graves PR, Haystead TA. Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev.* 2002;66:39-63.
11. Monteoliva L, Albar JP. Differential proteomics: an overview of gel and non-gel based approaches. *Brief Funct Genomic Proteomic.* 2004;3:220-39.
12. Gorg A, Weiss W, Dunn MJ. Current two-dimensional electrophoresis technology for proteomics. *Proteomics.* 2004;4:3665-85.
13. Unlu M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis.* 1997;18:2071-7.
14. Tonge R, Shaw J, Middleton B, et al. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics.* 2001;1:377-96.
15. Alban A, David SO, Bjorksten L, et al. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics.* 2003;3:36-44.
16. Friedman DB, Hill S, Keller JW, et al. Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics.* 2004;4:793-811.
17. Shaw J, Rowlinson R, Nickson J, et al. Evaluation of saturation labelling two-dimensional difference gel electrophoresis fluorescent dyes. *Proteomics.* 2003;3:1181-95.
18. Van den Bergh G, Arckens L. Fluorescent two-dimensional difference gel electrophoresis unveils the potential of gel-based proteomics. *Current Opinion in Biotechnology.* 2004;15:38-43.
19. Marouga R, David S, Hawkins E. The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Anal Bioanal Chem.* 2005;382:669-78.
20. Viswanathan S, Unlu M, Minden JS. Two-dimensional difference gel electrophoresis. *Nat Protoc.* 2006;1:1351-8.
21. Timms JF, Cramer R. Difference gel electrophoresis. *Proteomics.* 2008;8:4886-97.
22. Westermeier R, Scheibe B. Difference gel electrophoresis based on lys/cys tagging. *Methods Mol Biol.* 2008;424:73-85.
23. Minden JS, Dowd SR, Meyer HE, Stühler K. Difference gel electrophoresis. *Electrophoresis.* 2009;30 Suppl 1:S156-61.
24. GE Healthcare Bio-Sciences. 2-D Electrophoresis, Principles and Methods. Uppsala, Sweden: GE Healthcare Bio-Sciences, 2010 Contract No.: 80-6429-60 AD.
25. Richard E, Monteoliva L, Juarez S, et al. Quantitative analysis of mitochondrial protein expression in methylmalonic acidemia by two-dimensional difference gel electrophoresis. *J Proteome Res.* 2006;5:1602-10.
26. Ruiz-Romero C, Calamia V, Mateos J, et al. Mitochondrial dysregulation of osteoarthritic human articular chondrocytes analyzed by proteomics: a decrease in mitochondrial superoxide dismutase points to a redox imbalance. *Mol Cell Proteomics.* 2009;8:172-89.
27. Van den Bergh G, Clerens S, Vandesande F, Arckens L. Reversed-phase high-performance liquid chromatography prefractionation prior to two-dimensional difference gel electrophoresis and mass spectrometry identifies new differentially expressed proteins between striate cortex of kitten and adult cat. *Electrophoresis.* 2003;24:1471-81.
28. Monteoliva L, Martinez-Lopez R, Pitarch A, et al. Quantitative proteome and acidic subproteome profiling of *Candida albicans* yeast-to-hypha transition. *J Proteome Res.* 2011;10:502-17.
29. Wilson KE, Marouga R, Prime JE, et al. Comparative proteomic analysis using samples obtained with laser microdissection and saturation dye labelling. *Proteomics.* 2005;5:3851-8.
30. Sitek B, Sipos B, Kloppel G, et al. Application of fluorescence dye saturation labeling for differential proteome analysis of 1,000 microdissected cells from pancreatic ductal adenocarcinoma precursor lesions. *Methods Mol Biol.* 2008;425:1-14.
31. Turck CW, Falick AM, Kowalak JA, et al. The Association of Biomolecular Resource Facilities Proteomics Research Group 2006 study: relative protein quantitation. *Mol Cell Proteomics.* 2007;6:1291-8.
32. Reisinger V, Eichacker LA. How to analyze protein complexes by 2D blue native SDS-PAGE. *Proteomics.* 2007;7 Suppl 1:6-16.
33. Heinemeyer J, Scheibe B, Schmitz UK, Braun HP. Blue native DIGE as a tool for comparative analyses of protein complexes. *J Proteomics.* 2009;72:539-44.
34. Ummanni R, Mundt F, Pospisil H, et al. Identification of clinically relevant protein targets in prostate cancer with 2D-DIGE coupled mass spectrometry and systems biology network platform. *PLoS One.* 2011;6:e16833.
35. Bleijerveld OB, Brakenhoff RH, Schaaij-Visser TB, et al. Protein signatures associated with tumor cell dissemination in head and neck cancer. *J Proteomics.* 2011;74:558-66.
36. Rajpal R, Dowling P, Meiller J, et al. A novel panel of protein biomarkers for predicting response to thalidomide-based therapy in newly diagnosed multiple myeloma patients. *Proteomics.* 2011;11:1391-402.
37. Verrills NM, Irwin JA, He XY, et al. Identification of novel diagnostic biomarkers for asthma and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2011;183:1633-43.
38. Chou JL, Shenoy DV, Thomas N, et al. Early dysregulation of the mitochondrial proteome in a mouse model of Alzheimer's disease. *J Proteomics.* 2011;74:466-79.
39. Tanaka T, Sakurada S, Kano K, et al. Identification of tuberculosis-associated proteins in whole blood supernatant. *BMC Infect Dis.* 2011;11:71.
40. Madeira A, Santos PM, Coutinho CP, Pinto-de-Oliveira A, Sa-Correia I. Quantitative proteomics (2-D DIGE) reveals molecular strategies employed by *Burkholderia cenocepacia* to adapt to the airways of cystic fibrosis patients under antimicrobial therapy. *Proteomics.* 2011;11:1313-28.



41. Costa MM, Andrade HM, Bartholomeu DC, et al. Analysis of *Leishmania chagasi* by 2-D Difference Gel Electrophoresis (2-D DIGE) and Immunoproteomic: Identification of Novel Candidate Antigens for Diagnostic Tests and Vaccine. *J Proteome Res.* 2011;10:2172-84.
42. Dihazi H, Dihazi GH, Jahn O, et al. Multipotent adult germline stem cells and embryonic stem cells functional proteomics revealed an important role of eukaryotic initiation factor 5A (Eif5a) in stem cell differentiation. *J Proteome Res.* 2011;10:1962-73.
43. Mathesius U, Djordjevic MA, Oakes M, et al. Comparative proteomic profiles of the soybean (*Glycine max*) root apex and differentiated root zone. *Proteomics.* 2011;11:1707-19.
44. Zhang W, Liu Y, Zhang H, Dai J. Proteomic analysis of male zebrafish livers chronically exposed to perfluorononanoic acid. *Environ Int.* 2012;42:20-30.
45. Tonack S, Aspinall-O'Dea M, Jenkins RE, et al. A technically detailed and pragmatic protocol for quantitative serum proteomics using iTRAQ. *J Proteomics.* 2009;73:352-6.
46. 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.
47. Gygi SP, Rist B, Gerber SA, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol.* 1999;17:994-9.
48. Ross PL, Huang YN, Marchese JN, et al. Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Mol Cell Proteomics.* 2004;3:1154-69.
49. Thompson A, Schaefer J, Kuhn K, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem.* 2003;75:1895-904.
50. Mirgorodskaya OA, Kozmin YP, Titov MI, et al. Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (¹⁸O)-labeled internal standards. *Rapid Commun Mass Spectrom.* 2000;14:1226-32.
51. Ramos-Fernandez A, Lopez-Ferrer D, Vazquez 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.
52. Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C. Proteolytic ¹⁸O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem.* 2001;73:2836-42.
53. Bantscheff M, Boesche M, Eberhard D, et al. Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Mol Cell Proteomics.* 2008;7:1702-13.
54. Bondarenko PV, Chelius D, Shaler TA. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Analytical Chemistry.* 2002;74:4741-9.
55. Chelius D, Bondarenko PV. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of Proteome Research.* 2002;1:317-23.
56. Silva JC, Gorenstein MV, Li G-Z, Vissers JPC, Geromanos SJ. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics.* 2006;5:144-56.
57. Malmström J, Beck M, Schmidt A, et al. Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature.* 2009;460:762-5.
58. Silva JC, Denny R, Dorschel C, et al. Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Mol Cell Proteomics.* 2006;5:589-607.
59. Schwanhäusser B, Busse D, Li N, et al. Global quantification of mammalian gene expression control. *Nature.* 2011;473:337-42.
60. 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. *Journal of Proteome Research.* 2011;10:1785-93.
61. Domon B, Aebersold R. Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol.* 2010;28:710-21.
62. Domon B, Broder S. Implications of new proteomics strategies for biology and medicine. *J Proteome Res.* 2004;3:253-60.
63. Jaffe JD, Keshishian H, Chang B, et al. Accurate inclusion mass screening: a bridge from unbiased discovery to targeted assay development for biomarker verification. *Mol Cell Proteomics.* 2008;7:1952-62.
64. Schmidt A, Claassen M, Aebersold R. Directed mass spectrometry: towards hypothesis-driven proteomics. *Current Opinion in Chemical Biology.* 2009;13:510-7.
65. Schmidt A, Gehlenborg N, Bodenmiller B, et al. An integrated, directed mass spectrometric approach for in-depth characterization of complex peptide mixtures. *Mol Cell Proteomics.* 2008;7:2138-50.
66. Braisted JC, Kuntumalla S, Vogel C, et al. The APEX Quantitative Proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics.* 2008;9:529.
67. Paoletti A, Parmely T, Tomomori-Sato C, et al. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proceedings of the National Academy of Sciences of the United States of America.* 2006;103:18928-33.
68. Zybailov B, Mosley AL, Sardi ME, et al. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *Journal of Proteome Research.* 2006;5:2339-47.
69. Ishihama Y, Oda Y, Tabata T, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics.* 2005;4:1265-72.
70. Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol.* 2007;25:117-24.
71. Griffin NM, Yu J, Long F, et al. Label-free, normalized quantification of complex mass spectrometry data for proteomic

analysis. *Nat Biotechnol.* 2010;28:83-9.

72. Panchaud A, Scherl A, Shaffer SA, et al. Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. *Analytical Chemistry.* 2009;81:6481-8.
73. Geiger T, Cox J, Mann M. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol Cell Proteomics.* 2010;9:2252-61.
74. Plumb RS, Johnson KA, Rainville P, et al. UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Communications in Mass Spectrometry.* 2006;20:1989-94.
75. Carvalho PC, Han X, Xu T, et al. XDIA: improving on the label-free data-independent analysis. *Bioinformatics.* 2010;26:847-8.
76. Gillet LC, Navarro P, Tate S, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics.* 2012;11:O111.016717.
77. Venable JD, Dong M-Q, Wohlschlegel J, Dillin A, Yates JR. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature methods.* 2004;1:39-45.
78. Kuntumalla S, Braisted JC, Huang S-T, et al. Comparison of two label-free global quantitation methods, APEX and 2D gel electrophoresis, applied to the *Shigella dysenteriae* proteome. *Proteome Sci.* 2009;7:22.
79. Collins M, Choudhary J. Mapping multiprotein complexes by affinity purification and mass spectrometry. *Current Opinion in Biotechnology.* 2008;19:324-30.
80. Mosley AL, Sardi ME, Pattenden SG, et al. Highly reproducible label free quantitative proteomic analysis of RNA polymerase complexes. *Mol Cell Proteomics.* 2011;10:M110.000687.
81. Olinares PDB, Ponnala L, van Wijk KJ. Megadalton complexes in the chloroplast stroma of *Arabidopsis thaliana* characterized by size exclusion chromatography, mass spectrometry and hierarchical clustering. *Mol Cell Proteomics.* 2010;9:1594-615.
82. Sardi ME, Cai Y, Jin J, et al. Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proceedings of the National Academy of Sciences of the United States of America.* 2008;105:1454-9.
83. Sardi ME, Florens L, Washburn MP. Evaluation of clustering algorithms for protein complex and protein interaction network assembly. *Journal of Proteome Research.* 2009;8:2944-52.
84. Murray KK, Boyd RK, Eberlin MN, et al. Standard definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure Appl Chem.* 2013;85:1515-609.
85. Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Meth.* 2012;9:555-66.
86. Anderson NL, Anderson NG, Pearson TW, et al. A human proteome detection and quantitation project. *Mol Cell Proteomics.* 2009;8:883-6.
87. Paik Y-K, Omenn GS, Uhlen M, et al. Standard guidelines for the chromosome-centric human proteome project. *Journal of Proteome Research.* 2012;11:2005-13.
88. Costenoble R, Picotti P, Reiter L, et al. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in *Saccharomyces cerevisiae* under multiple conditions by targeted proteomics. *Molecular systems biology.* 2011;7:464.
89. Drabovich AP, Pavlou MP, Dimitromanolakis A, Diamandis EP. Quantitative analysis of energy metabolic pathways in MCF-7 breast cancer cells by selected reaction monitoring assay. *Mol Cell Proteomics.* 2012;11:422-34.
90. Wolf-Yadlin A, Hautaniemi 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;104:5860-5.
91. Chen Y-T, Chen H-W, Domanski D, et al. Multiplexed quantification of 63 proteins in human urine by multiple reaction monitoring-based mass spectrometry for discovery of potential bladder cancer biomarkers. *Journal of Proteomics.* 2012;75:3529-45.
92. Collins BC, Miller CA, Sposny A, et al. Development of a pharmaceutical hepatotoxicity biomarker panel using a discovery to targeted proteomics approach. *Mol Cell Proteomics.* 2012;11:394-410.
93. Conlon KP, Basrur V, Rolland D, et al. Fusion peptides from oncogenic chimeric proteins as putative specific biomarkers of cancer. *Mol Cell Proteomics.* 2013;12:2714-23.
94. Schiess R, Wollscheid B, Aebersold R. Targeted proteomic strategy for clinical biomarker discovery. *Molecular oncology.* 2009;3:33-44.
95. Whiteaker JR, Lin C, Kennedy J, et al. A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat Biotechnol.* 2011;29:625-34.
96. Hoofnagle AN, Wener MH. The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. *J Immunol Methods.* 2009;347:3-11.
97. Aebersold R, Burlingame AL, Bradshaw RA. Western blots versus selected reaction monitoring assays: time to turn the tables? *Mol Cell Proteomics.* 2013;12:2381-2.
98. Mann M. Can proteomics retire the western blot? *Journal of Proteome Research.* 2008;7:3065.
99. Bereman MS, MacLean B, Tomazela DM, Liebler DC, MacCoss MJ. The development of selected reaction monitoring methods for targeted proteomics via empirical refinement. *Proteomics.* 2012;12:1134-41.
100. Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular systems biology.* 2008;4:222.
101. Antignac JP, Le Bizet B, Monteau F, Poulain F, André F. Collision-induced dissociation of corticosteroids in electrospray tandem mass spectrometry and development of a screening method by high performance liquid chromatography/tandem mass spectrometry. *Rapid Communications in Mass Spectrometry.* 2000;14:33-9.
102. 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.



103. MacLean B, Tomazela DM, Abbatiello SE, et al. Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. 2010;82:10116-24.
104. Unwin RD, Griffiths JR, Whetton AD. A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple reaction monitoring-initiated detection and sequencing (MIDAS). *Nat Protoc.* 2009;4:870-7.
105. Lange V, Malmstrom J, Didion J, et al. Targeted quantitative analysis of *Streptococcus pyogenes* virulence factors by multiple reaction monitoring. *Mol Cell Proteomics.* 2008;7:1489-500.
106. Brun V, Dupuis A, Adrait A, et al. Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol Cell Proteomics.* 2007;6:2139-49.
107. Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature methods.* 2005;2:587-9.
108. Brun V, Masselon C, Garin J, Dupuis A. Isotope dilution strategies for absolute quantitative proteomics. *Journal of Proteomics.* 2009;72:740-9.
109. Neubert H, Muirhead D, Kabir M, et al. Sequential protein and peptide immunoaffinity capture for mass spectrometry-based quantification of total human β -nerve growth factor. 2013;85:1719-26.
110. Anderson N, Haines L, Hardie D, Olafson R, Pearson T. Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *Journal of Proteome Research.* 2004;3:235-44.
111. Berna M, Ackermann B. Increased throughput for low-abundance protein biomarker verification by liquid chromatography/tandem mass spectrometry. 2009;81:3950-6.
112. Shi T, Su D, Liu T, et al. Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. *Proteomics.* 2012;12:1074-92.
113. Stahl-Zeng J, Lange V, Ossola R, et al. High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics.* 2007;6:1809-17.
114. Shi T, Fillmore TL, Sun X, et al. Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. *Proceedings of the National Academy of Sciences of the United States of America.* 2012;109:15395-400.
115. Oda Y, Huang K, Cross FR, Cowburn 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;96:6591-6.
116. Ong S-E, Kratchmarova I, Mann M. Properties of ^{13}C -substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *Journal of Proteome Research.* 2003;2:173-81.
117. Ballif BA, Roux PP, Gerber SA, et al. Quantitative phosphorylation profiling of the ERK/p90 ribosomal S6 kinase-signaling cassette and its targets, the tuberous sclerosis tumor suppressors. *Proceedings of the National Academy of Sciences of the United States of America.* 2005;102:667-72.
118. Ibarrola N, Kalume DE, Gronborg M, Iwahori A, Pandey A. A proteomic approach for quantitation of phosphorylation using stable isotope labeling in cell culture. *Analytical chemistry.* 2003;75:6043-9.
119. Ong S-E, Mittler G, Mann M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nature methods.* 2004;1:119-26.
120. Ibarrola N, Molina N, Iwahori A, Pandey A. A novel proteomic approach for specific identification of tyrosine kinase substrates using [^{13}C]tyrosine. *The Journal of biological chemistry.* 2004;279:15805-13.
121. Ong S-E, Blagoev B, Kratchmarova I, et al. Stable Isotope Labeling by Amino Acids in Cell Culture for Quantitative Proteomics. In: Celis JE, editor. *Cell biology, a laboratory handbook*: Academic Press; 2006. p. 2141.
122. Gouw JW, Krijgsveld J, Heck AJ. Quantitative proteomics by metabolic labeling of model organisms. *Mol Cell Proteomics.* 2010;9:11-24.
123. Kruger M, Moser M, Ussar S, et al. SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell.* 2008;134:353-64.
124. Schütz W, Hausmann N, Krug K, Hampf R, Macek B. Extending SILAC to proteomics of plant cell lines. *The Plant cell.* 2011;23:1701-5.
125. Geiger T, Cox J, Ostasiewicz P, Wisniewski JR, Mann M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Meth.* 2010;7:383-5.
126. Ong SE, Mann M. Stable isotope labeling by amino acids in cell culture for quantitative proteomics. *Methods in molecular biology.* 2007;359:37-52.
127. Ohlund LB, Hardie DB, Elliott MH, et al. Standard Operating Procedures and Protocols for the Preparation and Analysis of Plasma Samples Using the iTRAQ Methodology. In: Ivanov AR, Lazarev AV, editors. *Sample Preparation in Biological Mass Spectrometry*. New York: Springer Verlag; 2011. p. 1089.
128. Pappin D. An iTRAQ Primer. Available from: <http://www.ushupo.org/TechTalk/tabid/101/Default.aspx>.
129. Wenger CD, Lee MV, Hebert AS, et al. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Methods.* 2011;8:933-5.
130. Ting L, Rad R, Gygi SP, Haas W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods.* 2011;8:937-40.
131. Karp NA, Huber W, Sadowski PG, et al. Addressing accuracy and precision issues in iTRAQ quantitation. *Mol Cell Proteomics.* 2010;9:1885-97.
132. Mertins P, Udeschi ND, Clauser KR, et al. iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. *Mol Cell Proteomics.* 2012;11:M111 014423.
133. Zhang Y, Askenazi M, Jiang J, et al. A robust error model for iTRAQ quantification reveals divergent signaling between oncogenic FLT3 mutants in acute myeloid leukemia. *Mol Cell Proteomics.* 2010;9:780-90.

134. Ow SY, Salim M, Noirel J, et al. iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J Proteome Res.* 2009;8:5347-55.
135. Yao X, Afonso C, Fenselau C. Dissection of proteolytic 18O labeling: endoprotease-catalyzed 16O-to-18O exchange of truncated peptide substrates. *Journal of Proteome Research.* 2003;2:147-52.
136. Schnolzer M, Jedrzejewski P, Lehmann WD. 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;17:945-53.
137. Fenselau C, Yao X. Proteolytic labeling with 18O for comparative proteomics studies: preparation of 18O-labeled peptides and the 18O/16O peptide mixture. *Methods Mol Biol.* 2007;359:135-42.
138. Wang YK, Ma Z, Quinn DF, Fu EW. Inverse 18O labeling mass spectrometry for the rapid identification of marker/target proteins. *Anal Chem.* 2001;73:3742-50.
139. Johnson KL, Muddiman DC. A method for calculating 16O/18O peptide ion ratios for the relative quantification of proteomes. *J Am Soc Mass Spectrom.* 2004;15:437-45.
140. Zang L, Palmer Toy D, Hancock WS, Sgroi DC, Karger BL. Proteomic analysis of ductal carcinoma of the breast using laser capture microdissection, LC-MS, and 16O/18O isotopic labeling. *J Proteome Res.* 2004;3:604-12.
141. Halligan BD, Slyper RY, Twigger SN, et al. ZoomQuant: an application for the quantitation of stable isotope labeled peptides. *J Am Soc Mass Spectrom.* 2005;16:302-6.
142. Qian WJ, Monroe ME, Liu T, et al. Quantitative proteome analysis of human plasma following in vivo lipopolysaccharide administration using 16O/18O labeling and the accurate mass and time tag approach. *Mol Cell Proteomics.* 2005;4:700-9.
143. Eckel-Passow JE, Oberg AL, Therneau TM, et al. Regression analysis for comparing protein samples with 16O/18O stable-isotope labeled mass spectrometry. *Bioinformatics.* 2006;22:2739-45.
144. López-Ferrer D, Ramos-Fernandez A, Martinez-Bartolome S, Garcia-Ruiz P, Vazquez J. Quantitative proteomics using 16O/18O labeling and linear ion trap mass spectrometry. *Proteomics.* 2006;6 Suppl 1:S4-11.
145. Mason CJ, Therneau TM, Eckel-Passow JE, et al. A method for automatically interpreting mass spectra of 18O-labeled isotopic clusters. *Mol Cell Proteomics.* 2007;6:305-18.
146. Stockwin LH, Blonder J, Bumke MA, et al. Proteomic analysis of plasma membrane from hypoxia-adapted malignant melanoma. *J Proteome Res.* 2006;5:2996-3007.
147. Bantscheff M, Dumpelfeld B, Kuster B. Femtomol sensitivity post-digest (18)O labeling for relative quantification of differential protein complex composition. *Rapid Commun Mass Spectrom.* 2004;18:869-76.
148. Brown KJ, Fenselau C. Investigation of doxorubicin resistance in MCF-7 breast cancer cells using shot-gun comparative proteomics with proteolytic 18O labeling. *J Proteome Res.* 2004;3:455-62.
149. Chi LM, Lee CW, Chang KP, et al. Enhanced interferon signaling pathway in oral cancer revealed by quantitative proteome analysis of microdissected specimens using 16O/18O labeling and integrated two-dimensional LC-ESI-MALDI tandem MS. *Mol Cell Proteomics.* 2009;8:1453-74.
150. Heller M, Mattou H, Menzel C, Yao X. Trypsin catalyzed 16O-to-18O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers. *Journal of the American Society for Mass Spectrometry.* 2003;14:704-18.
151. Staes A, Demol H, Van Damme J, et al. Global differential non-gel proteomics by quantitative and stable labeling of tryptic peptides with oxygen-18. *J Proteome Res.* 2004;3:786-91.
152. Stewart, II, Thomson T, Figeys D. 18O labeling: a tool for proteomics. *Rapid Commun Mass Spectrom.* 2001;15:2456-65.
153. Wang J, Gutierrez P, Edwards N, Fenselau C. Integration of 18O labeling and solution isoelectric focusing in a shotgun analysis of mitochondrial proteins. *J Proteome Res.* 2007;6:4601-7.
154. 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.
155. Hicks WA, Halligan BD, Slyper RY, et al. Simultaneous quantification and identification using 18O labeling with an ion trap mass spectrometer and the analysis software application "ZoomQuant". *J Am Soc Mass Spectrom.* 2005;16:916-25.
156. Sakai J, Kojima S, Yanagi K, Kanaoka M. 18O-labeling quantitative proteomics using an ion trap mass spectrometer. *Proteomics.* 2005;5:16-23.
157. Mayya V, Rezaul K, Cong YS, Han D. Systematic comparison of a two-dimensional ion trap and a three-dimensional ion trap mass spectrometer in proteomics. *Mol Cell Proteomics.* 2005;4:214-23.
158. Trevisan-Herraz M, Bagwan N, Garcia-Marques F, Rodriguez JM, Jorge I, Ezkurdia I, Bonzon-Kulichenko E and Vazquez J. SanXoT: a modular and versatile package for the quantitative analysis of high-throughput proteomics experiments. *Bioinformatics.* 2019;35:1594-1596.
159. Mirza SP, Olivier M. Methods and approaches for the comprehensive characterization and quantification of cellular proteomes using mass spectrometry. *Physiol Genomics.* 2008;33:3-11.
160. Jorge I, Casas EM, Villar M, et al. High-sensitivity analysis of specific peptides in complex samples by selected MS/MS ion monitoring and linear ion trap mass spectrometry: application to biological studies. *J Mass Spectrom.* 2007;42:1391-403.
161. Old WM, Meyer-Arendt K, Aveline-Wolf L, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics.* 2005;4:1487-502.
162. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Roy Statist Soc Ser B.* 1995;57:289-300.
163. Choi H, Nesvizhskii AI. False discovery rates and related statistical concepts in mass spectrometry-based proteomics. *J Proteome Res.* 2008;7:47-50.



164. Unwin RD, Pierce A, Watson RB, Sternberg DW, Whetton AD. Quantitative proteomic analysis using isobaric protein tags enables rapid comparison of changes in transcript and protein levels in transformed cells. *Mol Cell Proteomics*. 2005;4:924-35.
165. Andersen JS, Wilkinson CJ, Mayor T, et al. Proteomic characterization of the human centrosome by protein correlation profiling. *Nature*. 2003;426:570-4.
166. Gan CS, Chong PK, Pham TK, Wright PC. Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J Proteome Res*. 2007;6:821-7.
167. Lin WT, Hung WN, Yian YH, et al. Multi-Q: a fully automated tool for multiplexed protein quantitation. *J Proteome Res*. 2006;5:2328-38.
168. D'Agostino RB, Belanger A, D'Agostino RBJ. A Suggestion for Using Powerful and Informative Tests of Normality. *The American Statistician*. 1990;44:316-21.
169. Pan C, Kora G, McDonald WH, et al. ProRata: A quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation. *Anal Chem*. 2006;78:7121-31.
170. Pavelka N, Fournier ML, Swanson SK, et al. Statistical similarities between transcriptomics and quantitative shotgun proteomics data. *Mol Cell Proteomics*. 2008;7:631-44.
171. Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*. 2002;18 Suppl 1:S96-104.
172. Li XJ, Zhang H, Ranish JA, Aebersold R. Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Anal Chem*. 2003;75:6648-57.
173. Choe LH, Aggarwal K, Franck Z, Lee KH. A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in *Escherichia coli* cells. *Electrophoresis*. 2005;26:2437-49.
174. Boehm AM, Putz S, Altenhofer D, Sickmann A, Falk M. Precise protein quantification based on peptide quantification using iTRAQ. *BMC Bioinformatics*. 2007;8:214.
175. Navarro P, Trevisan-Herraz M, Bonzon-Kulichenko E, Nunez E, Martinez-Acedo P, Perez-Hernandez D, Jorge I, Mesa R, Calvo E, Carrascal M, Hernaez ML, Garcia F, Barcena JA, Ashman K, Abian J, Gil C, Redondo JM and Vazquez J. General statistical framework for quantitative proteomics by stable isotope labeling. *Journal of proteome research*. 2014;13:1234-47. Garcia-Marques F, Trevisan-Herraz M, Martinez-Martinez S, Camafeita E, Jorge I, Lopez JA, Mendez-Barbero N, Mendez-Ferrer S, Del Pozo MA, Ibanez B, Andres V, Sanchez-Madrid F, Redondo JM, Bonzon-Kulichenko E and Vazquez J. A Novel Systems-Biology Algorithm for the Analysis of Coordinated Protein Responses Using Quantitative Proteomics. *Molecular & cellular proteomics : MCP*. 2016;15:1740-60. Bonzon-Kulichenko E, Perez-Hernandez D, Nunez E, Martinez-Acedo P, Navarro P, Trevisan-Herraz M, Ramos Mdel C, Sierra S, Martinez-Martinez S, Ruiz-Meana M, Miro-Casas E, Garcia-Dorado D, Redondo JM, Burgos JS and Vazquez J. A robust method for quantitative high-throughput analysis of proteomes by 18O labeling. *Molecular & cellular proteomics : MCP*. 2011;10:M110 003335.
176. Han DK, Eng J, Zhou H, Aebersold R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol*. 2001;19:946-51.
177. 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;26:1367-72.
178. Bouyssié D, Gonzalez de Peredo A, Mouton E, et al. Mascot file parsing and quantification (MFPaQ), a new software to parse, validate, and quantify proteomics data generated by ICAT and SILAC mass spectrometric analyses: application to the proteomics study of membrane proteins from primary human endothelial cells. *Mol Cell Proteomics*. 2007;6:1621-37.
179. Mortensen P, Gouw JW, Olsen JV, et al. MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. *J Proteome Res*. 2010;9:393-403.
180. Palmblad M, Mills DJ, Bindschedler LV. Heat-shock response in *Arabidopsis thaliana* explored by multiplexed quantitative proteomics using differential metabolic labeling. *J Proteome Res*. 2008;7:780-5.
181. Khan Z, Bloom JS, Garcia BA, Singh M, Kruglyak L. Protein quantification across hundreds of experimental conditions. *Proc Natl Acad Sci U S A*. 2009;106:15544-8.
182. Wang G, Wu WW, Pisitkun T, et al. Automated quantification tool for high-throughput proteomics using stable isotope labeling and LC-MSn. *Anal Chem*. 2006;78:5752-61.
183. MacCoss MJ, Wu CC, Liu H, Sadygov R, Yates JR, 3rd. A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. *Anal Chem*. 2003;75:6912-21.

Identificación de modificaciones post-traduccionales mediante inmunopurificación y espectrometría de masas

Ana Martínez del Val¹, Javier Muñoz Peralta²,

¹Unidad de Proteómica. Centro Nacional de Investigaciones Oncológicas.

²PhD. Jefe de la Unidad de Proteómica, Centro Nacional de Investigaciones Oncológicas. Oncológicas.

1. Introducción.

1.1. Relevancia de las modificaciones post-traduccionales de proteínas en los sistemas biológicos

Las modificaciones post-traduccionales de las proteínas o PTMs (del inglés *post-translational modifications*) constituyen un mecanismo de regulación fundamental en múltiples procesos biológicos. Las PTMs permiten expandir la complejidad de las proteínas al modificar de manera covalente las cadenas laterales de sus aminoácidos, alterando su estructura y función. Hasta la fecha, se han descrito más de 300 tipos de modificaciones en proteínas(1). Entre todas estas modificaciones, la fosforilación es probablemente la PTM caracterizada a mayor escala a día de hoy. Esto se debe principalmente al temprano desarrollo de métodos de purificación altamente específicos para esta modificación. Así pues, numerosos trabajos han permitido determinar que la fosforilación afecta aproximadamente al 60% de las proteínas mediante la modificación covalente de más de 300.000 residuos cuyas funciones regulatorias controlan prácticamente todos los procesos celulares tales como la expresión génica, la señalización celular o el control del ciclo celular.

En lo que respecta al resto de modificaciones (metilaciones, acetilaciones, ubiquitinaciones...), se conocía poco sobre el impacto y las funciones regulatorias que tienen a nivel global en la célula, y las pocas funciones conocidas para las mismas habían sido descritas principalmente en histonas. Sin embargo, recientemente, el desarrollo de anticuerpos monoclonales altamente específicos para estas modificaciones ha permitido expandir el catálogo de proteínas modificadas post-traduccionalmente. Gracias a esto se han podido describir múltiples procesos biológicos regulados por acetilaciones, metilaciones y ubiquitinaciones de una forma similar a lo conocido hasta la fecha para los mecanismos de fosforilación. A lo largo de éste capítulo discutiremos brevemente las funciones biológicas de algunas de estas modificaciones, como son la ubiquitina-



ción, la acetilación y la metilación y explicaremos, en mayor detalle, las principales estrategias experimentales de inmuno-purificación que permiten su estudio a gran escala mediante técnicas proteómicas.

1.1.1. Ubiquitinación

La ubiquitinación es un proceso enzimático que consiste en la conjugación covalente de una (mono-ubiquitinación) o de varias (poli-ubiquitinación) cadenas de la molécula de ubiquitina en el extremo ϵ -amino de lisinas. La molécula de ubiquitina es un polipéptido compuesto por 76 aminoácidos. El mecanismo de ubiquitinación depende de la acción secuencial de tres enzimas: enzimas activadoras de ubiquitina (E1), enzimas de conjugación de ubiquitina (E2) y ubiquitin ligasas (E3). La ubiquitinación de proteínas es un proceso reversible puesto que puede ser regulado de manera transitoria mediante las enzimas de-ubiquitininas que eliminan el residuo de ubiquitina.

Asimismo, la ubiquitinación es una modificación muy versátil ya que puede modificar un mismo residuo con una o varias cadenas de ubiquitina. Las diferentes conformaciones que adoptan las cadenas poliméricas de ubiquitina permiten regular con mayor precisión las vías de señalización celular(2). Estas cadenas pueden ser reconocidas de manera específica por proteínas que posean dominios de unión a ubiquitina(3), de modo que la interacción puede desencadenar procesos como la degradación de la proteína(4), la señalización de las proteínas para su transporte (e.g. internalización por lisosomas(5)) o la activación de cascadas de señalización, como por ejemplo la ruta de NF-kappaB responsable de procesos inmunitarios e inflamatorios(6).

1.1.2. Acetilación

La acetilación consiste en la adición de un grupo acetilo al extremo ϵ -amino de las lisinas. Aunque se trata, principalmente, de un proceso enzimático catalizado por las lisina-acetil-transferasas (KATs, de *lysine acetyl transferases*), también se puede producir de manera no enzimática si la concentración de acetilCoA es muy elevada (e.g. mitocondrias). La acetilación es una reacción reversible catalizada de manera enzimática por las lisina-deacetilasas (KDAC, de *lysine de-acetylases*).

La acetilación de una lisina provoca la neutralización de su carga positiva e incrementa la hidrofobicidad de la cadena lateral. Este cambio modifica la forma en que la proteína interacciona con otras proteínas. Así, se ha descrito que esta modificación está implicada en la remodelación de la cromatina, ya que la de-acetilación de las histonas favorece la compactación de la cromatina y, por tanto, el silenciamiento de la transcripción. Por esta razón las KATs y las KDACs se consideran entre los principales represores epigenéticos. Además, se ha descrito la importancia de la acetilación en el control de ciclo celular o en la localización y transporte de proteínas nucleares(7–9).

1.1.3. Metilación

Mediante el proceso de metilación se transfiere uno o varios grupos metilo desde la S-adenosil-metionina (SAM) a las cadenas laterales de la lisina o de la arginina. La lisina se modifica por su extremo ϵ -amino, y puede incorporar uno, dos o tres grupos metilo. La metilación de arginina sucede en el grupo guanidinio de su cadena lateral, y puede tratarse de mono-metilación si solo se modifica un grupo amino, o dimetilación, la cual puede ser simétrica (si se modifican ambos grupos amino) o asimétrica (si se añaden dos grupos metilo a uno sólo de los grupos amino). Se trata de una reacción enzimática catalizada por las lisina-metiltransferasas (PKMTs de *protein lysine methyl transferases*) y las arginina-metiltransferasas (PRMTs de *protein arginine methyl transferases*). Igualmente, esta reacción es reversible, y la eliminación de los grupos metilo está catalizada por las demetilasas (PKDMs, de *protein lysine de-methylases* y PRDMs de *protein arginine de-methylases*).

A diferencia de la acetilación o la ubiquitinación; la metilación de lisinas no altera de manera significativa las propiedades de carga o hidrofobicidad de la proteína. No obstante, se ha descrito que su presencia puede incrementar la estabilidad de la proteína y la tolerancia a los shocks térmicos. La metilación de lisinas es principalmente conocida en histonas, en las cuales puede regular su interacción con el ADN, participando, por tanto, en mecanismos de regulación epigenéticos(10). Entre las proteínas no histónicas reguladas por metilación destaca también p53, para la cual se describió un mecanismo de metilación/demetilación que regula su interacción con sustratos(11) o su estabilidad (12). También se ha descrito una gran presencia de residuos metilados en proteínas que forman complejos con ácidos nucleicos, como las proteínas ribosomales(13).

La metilación de argininas, a diferencia de la metilación en lisinas, tiene un impacto en las características fisicoquímicas de la proteína. La arginina posee un grupo guanidinio susceptible de formar múltiples puentes de hidrógeno que le permite interactuar con otras argininas de su entorno, con otras proteínas y, principalmente, con ácidos nucleicos(14). Es por ello que la metilación de argininas está muy presente tanto en las proteínas que interactúan con el ARN y ADN, como en las que regulan el *splicing* alternativo(15).

1.2. Detección y análisis de PTMs mediante proteómica basada en espectrometría de masas

Debido a su alta sensibilidad y especificidad, la espectrometría de masas se ha posicionado como la tecnología con mayor potencial para el estudio de PTMs a gran escala. Sin embargo, y a pesar de su relevancia en múltiples procesos biológicos, las proteínas que contienen PTMs se encuentran normalmente en proporciones sub-estequiométricas respecto a su equivalente no modificado. Esto dificulta en gran medida la identificación global de residuos modificados mediante procedimientos estándar de proteómica *shotgun*. Por tanto, antes de su análisis por espectrometría de masas, es necesario implementar estrategias dirigidas a enriquecer aquellas modificaciones de interés. En ocasiones, este proceso de enriquecimiento se puede combinar con métodos de fraccionamiento para reducir la complejidad de la muestra y aumentar así



el rango dinámico, la cobertura y la sensibilidad del análisis.

Tradicionalmente, las modificaciones post-traduccionales se han estudiado de manera dirigida para una determinada proteína. Para ello, se empleaban métodos clásicos de biología molecular que dependían, en la mayoría de los casos, de la existencia de anticuerpos específicos frente a un aminoácido modificado en una posición concreta de la proteína. Aunque esta aproximación todavía es ampliamente utilizada, no permite abarcar la enorme complejidad que poseen las modificaciones post-traduccionales a nivel global, por ejemplo, durante la activación de vías de señalización. El reciente desarrollo de métodos de purificación para determinadas modificaciones ha sido clave para el estudio de PTMs desde una perspectiva proteómica. En este sentido, como se mencionaba anteriormente, la fosforilación es probablemente la modificación post-traduccional mejor caracterizada, en gran parte, consecuencia de la implementación de métodos que permiten purificar péptidos fosforilados en muestras complejas de manera muy específica y con una alta sensibilidad. El método de purificación de fosfopéptidos explota las características fisicoquímicas de los péptidos fosforilados: el grupo fosfato, cargado negativamente, aumenta la afinidad de los péptidos con resinas de óxidos metálicos (e.g.: TiO_2 (16)) o metales quelados (e.g.: Fe-IMAC(17–19)). No obstante, otro tipo de péptidos, como los que son muy ácidos(20) o los que contienen ácido siálico también poseen cierta afinidad por estas resinas(21) y pueden competir con los péptidos fosforilados durante los procedimientos de purificación.

Sin embargo, el estudio de otras modificaciones post-traduccionales más allá de la fosforilación se ha desarrollado a un ritmo más lento, principalmente debido a la ausencia de métodos de purificación altamente específicos. Por ello, el estudio de estas PTMs no despuntó hasta el desarrollo de los anticuerpos “pan-específicos”. Estos anticuerpos, a diferencia de los comúnmente usados en biología molecular, no reconocen una proteína o péptido concreto, sino que son capaces de identificar una modificación independientemente de la secuencia de aminoácidos de su entorno. El desarrollo de estos anticuerpos, junto con las mejoras de tecnologías de espectrometría de masas cada vez más sensibles, han permitido por primera vez describir en profundidad la magnitud y la compleja regulación de otras PTMs como son la acetilación, ubiquitinación y la metilación en sistemas celulares complejos.

Uno de los ejemplos más ilustrativos sobre el potencial del uso de anticuerpos para estudiar estas modificaciones a gran escala lo encontramos en Choudhary *et al.*(8) En este trabajo se realizó la purificación de péptidos acetilados en lisinas mediante anticuerpos pan-específicos y su análisis mediante espectrometría de masas permitiendo la identificación de 3.600 lisinas acetiladas en una línea celular humana. Este trabajo sentó las bases de una metodología que ha continuado optimizándose desde entonces. En 2012, Lundy *et al.*(22) describió 15.474 lisinas acetiladas en 4.541 proteínas en tejidos procedentes de rata. Este campo se encuentra en constante progreso y, recientemente, se ha hecho disponible de manera comercial una combinación de siete anticuerpos monoclonales complementarios, diseñados para aumentar la cobertura del acetiloma(23).

En el caso de la ubiquitinación, no fue hasta 2010 cuando se publicó el primer estudio en profundidad de un ubiquitinoma mediante cromatografía líquida acoplada a espectrometría de masas en tándem (LC-MS/MS)(24). En este trabajo, los autores

desarrollaron un anticuerpo monoclonal capaz de reconocer aquellos péptidos que contuviesen dos residuos de glicina (diGly) unidos a la cadena lateral de una lisina. Este residuo diGly resulta de digerir la molécula de ubiquitina con tripsina. De este modo, se consiguieron identificar 374 sitios de ubiquitinación en células HEK293. No obstante, en 2011, empleando la misma metodología, Kim *et al* identificaron 19.000 sitios de ubiquitinación en 5.000 proteínas, activados en respuesta a la inhibición del proteasoma en una línea celular humana(25).

A diferencia de la acetilación y la ubiquitinación, el estudio masivo de la metilación en proteomas complejos es mucho más reciente. Esto se debe a que inicialmente los estudios de MS dirigidos frente a esta modificación sufrían de muy baja eficiencia en el enriquecimiento de péptidos modificados. En 2004 se publicó el primer trabajo que identificaba sitios de metilación mediante espectrometría de masas(26) en el que los autores empleaban un marcaje metabólico con un grupo metilo 'pesado' que se incorporaría en aquellos aminoácidos susceptibles de ser metilados, permitiendo de esta forma identificar únicamente 59 sitios de metilación. Otro ejemplo es Uhlmann *et al*(27) en el que identificaban argininas metiladas mediante el uso de HILIC (*Hydrophilic interaction chromatography*) como forma de enriquecer en péptidos altamente básicos e hidrofílicos, características propias de los péptidos metilados. No obstante, esta aproximación sólo describió 249 sitios de metilación de argininas. La ausencia de métodos de purificación altamente específicos en estos estudios, hace que estas aproximaciones carezcan de sensibilidad y sean poco eficientes. Sin embargo, en 2014, el uso de marcaje isotópico con SILAC (*Stable isotope labeling by amino acids in cell culture*) junto con anticuerpos anti-monometil-arginina, lo cuales habían sido descritos previamente por primera vez ese mismo año(28), permitió describir hasta 2.502 sitios de metilación de argininas(10), aumentando en un orden de magnitud, el número de sitios descritos anteriormente. Más recientemente, Larsen *et al* describieron el que es hasta la fecha el metiloma de argininas más completo para células humanas, con 8.300 residuos descritos para 3.300 proteínas. Además de la información derivada de este estudio, los autores pudieron observar que la reacción de metilación por PRMT no es dependiente de la secuencia.

Para finalizar, en este primer bloque hemos descrito brevemente la evolución de la metodología experimental que nos permite actualmente estudiar de manera extensiva y sistemática el acetiloma, el ubitinoma y el metiloma entre otras modificaciones. Pero las posibilidades no finalizan aquí, la tecnología descrita en este capítulo permite también estudiar no solo cada modificación independientemente sino que también posibilita obtener información acerca de cómo estas modificaciones interactúan entre sí, pudiendo regular otras modificaciones diferentes y no sólo aquellas semejantes (*e.g*: cascadas de fosforilación). Este proceso, conocido como *cross-talk* entre modificaciones, aumenta la complejidad de los sistemas de señalización presentes en los sistemas biológicos. Por ejemplo, Swaney *et al*(29) emplearon espectrometría de masas para estudiar cómo se co-regulan los sitios de fosforilación y ubiquitinación de 466 proteínas de *Saccharomyces cerevisiae*. De este modo consiguieron identificar sitios de fosforilación co-regulados por sitios de ubiquitinación, los cuales están altamente conservados en levadura. Otro ejemplo destacable es el estudio realizado por Elia *et al*; en él evaluaron de manera conjunta el fosfoproteoma, el ubitinoma y el acetiloma en células sometidas a estrés replicativo. La metodología para estudiar varias PTMs conjuntamente puede variar. Mientras que en el caso de Elia *et al*(30) la purificación de

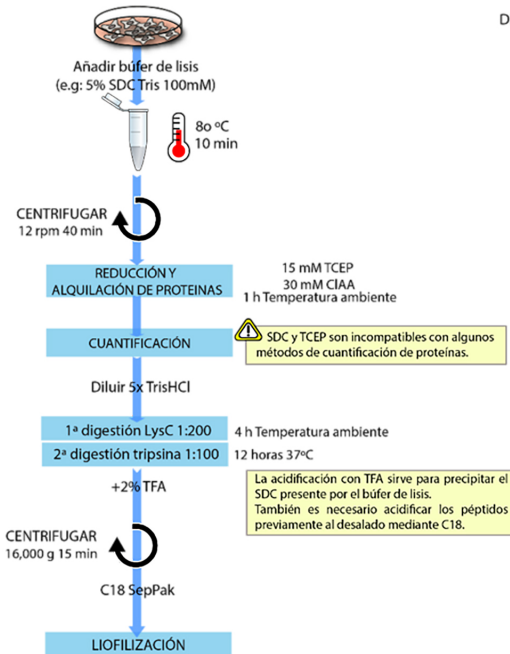


cada modificación se realizó de manera independiente, también es posible realizar un enriquecimiento secuencial en el sobrenadante del primer enriquecimiento que sirva de material de partida para el siguiente, y así sucesivamente(31).

2. Estrategias experimentales basadas en anticuerpos para el análisis de PTMS.

El análisis de cada modificación post-traducciona tiene unas particularidades que dependen en gran medida del tipo de muestra que se va a analizar y, por tanto, el protocolo experimental debe ser cuidadosamente evaluado en cada caso para maximizar la eficiencia de la inmunopurificación. En líneas generales, el proceso experimental descrito en este capítulo es fácilmente trasladable a todos los estudios en los que se emplea un anticuerpo para purificar PTMs (Figura 1). El protocolo de análisis de PTMs mediante espectrometría de masas consta de varias etapas: (i) lisis celular y extracción de las proteínas, (ii) digestión enzimática, (iii) inmunopurificación de los péptidos modificados, (iv) análisis mediante LC-MS/MS y (v) análisis de datos. En esta sección detallaremos las particularidades de cada una de estas etapas.

PREPARACIÓN DE MUESTRA



INMUNOPURIFICACIÓN

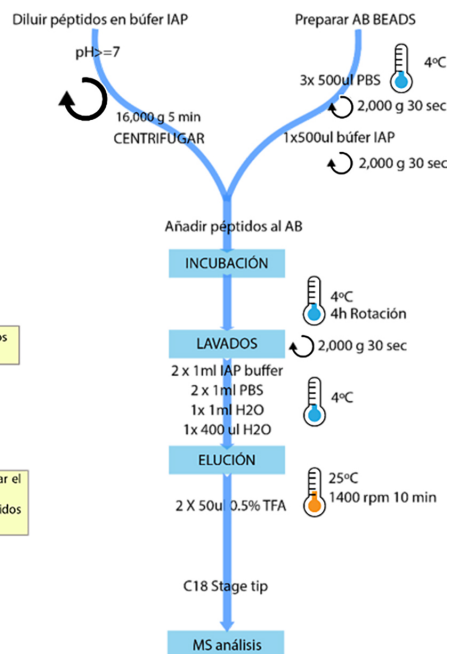


Figura 1. Protocolo experimental de un procedimiento estándar de purificación de péptidos modificados post-traduccionamente empleando anticuerpos.

2.1. Lisis y extracción de proteínas

Como en la mayoría de protocolos de proteómica *shotgun*, la lisis celular debe realizarse en condiciones desnaturalizantes para romper la estructura terciaria de las proteínas y así facilitar la digestión enzimática de las mismas. Por esta misma razón, los puentes disulfuro de las proteínas deben ser previamente reducidos y alquilados. Un punto fundamental en esta etapa consiste en evaluar la composición de los distintos búferes y reactivos que van a ser usados. Ciertos compuestos pueden inducir modificaciones artefactuales en la muestra que pueden provocar una disminución de la selectividad y/o sensibilidad de la inmunopurificación. Por ejemplo, en Martínez-Val et al se describe cómo la carbamilación artefactual de la muestra debido a la presencia de urea durante la preparación reduce en gran medida la eficiencia del enriquecimiento de péptidos acetilados(32). La carbamilación de lisinas es estéricamente muy similar a la acetilación. Por ello, en estas condiciones, la alta concentración de péptidos carbamilados competirá con los péptidos acetilados disminuyendo la eficiencia y especificidad de la inmunopurificación. Otro ejemplo de cómo las condiciones experimentales pueden afectar al resultado de una inmunopurificación se describe en Nielsen *et al*(33). En este trabajo se muestra cómo el empleo de iodo-acetamida (IAA) como agente alquilante puede generar un aducto de dos moléculas de IAA en lisinas, que posee exactamente la misma masa que el residuo diGly de la ubiquitina, y por tanto sería indistinguible para el motor de búsqueda. También se ha descrito que la presencia de ciertos alcoholes (i.e. metanol o alcohol isopropílico) durante la preparación de la muestra puede inducir la metilación artefactual en la misma, afectando principalmente a residuos de ácido glutámico y aspártico(34,35).

2.2. Digestión enzimática

Es importante destacar que el enriquecimiento de PTMs para su posterior análisis mediante espectrometría de masas se realiza normalmente a nivel de péptido. Esta estrategia favorece la accesibilidad del anticuerpo a la modificación, aumentando por tanto la sensibilidad de la técnica y permitiendo la localización exacta del residuo modificado en la secuencia proteica. Debido a su alta eficiencia y selectividad, la enzima más usada en estudios de proteómica es la tripsina. La tripsina digiere las proteínas por el extremo C-terminal de lisinas y argininas. En el caso de péptidos ubiquitinados, este paso de digestión triptica es esencial para su purificación, ya que la digestión con esta proteasa permite cortar la molécula de ubiquitina por la arginina en posición -3 y generar así el epítipo Lys-diGly reconocido por el anticuerpo.

No obstante, en un análisis de PTMs se debe considerar que la modificación en un residuo puede alterar la efectividad de la digestión, lo cual puede tener como resultado la generación de péptidos no óptimos para espectrometría de masas. Esto es especialmente relevante en las modificaciones tratadas en este capítulo, las cuales afectan a residuos de lisina y arginina, residuos tras los cuales digiere la tripsina. Es posible, por tanto, que la eficiencia de la tripsina se vea disminuida por la presencia de lisinas y argininas modificadas. Por ello, en ocasiones puede ser de utilidad emplear otro tipo de enzimas para digerir la muestra. Por ejemplo, la Asp-N, la Glu-C o la quimotripsina pueden ser útiles como sustitutos de la tripsina. Además de solventar la pérdida de eficiencia por un posible impedimento estérico, el uso de enzimas complementarias a la



tripsina, es decir, que digieran la proteína por otros residuos, puede permitir incrementar la cobertura de ciertas modificaciones, las que por su distribución en la secuencia de aminoácidos no puedan ser abarcadas con una digestión triptica.

2.3. Inmunopurificación

Como hemos mencionado anteriormente, la purificación de péptidos acetilados, metilados y ubiquitinados es posible gracias al desarrollo de anticuerpos pan-específicos capaces de reconocer dichas modificaciones. Algunos de los anticuerpos comerciales disponibles para la purificación de estas PTM están listados en la Tabla 1.

PTM	Referencia anticuerpo comercial	Empresa
Monometilación en arginina	Mono-Methyl Arginine Motif [mme-RG] Immunoaffinity Beads #12235	PTMScan® Cell Signaling Technologies
Dimetilación asimétrica en arginina	Asymmetric Di-Methyl Arginine Motif [adme-R] Kit #13474	PTMScan® Cell Signaling Technologies
Dimetilación simétrica en arginina	Symmetric Di-Methyl Arginine Motif [sdme-RG] Kit #13563	PTMScan® Cell Signaling Technologies
Mono, di y trimetilación en lisina	Pan-Methyl Lysine Kit #14809	PTMScan® Cell Signaling Technologies
Mono y di-metilación en lisina	Pan anti-mono/di-methyllysine antibody conjugated agarose beads	PTM Biolabs
Trimetilación en lisina	Pan anti-tri-methyllysine antibody conjugated agarose beads	PTM Biolabs
Residuo diGly en lisina	Ubiquitin Remnant Motif (K-ε-GG) Kit #5562	PTMScan® Cell Signaling Technologies
	Diglycyl-Lysine Antibody, clone GX41	LUCERNA
	Pan anti-diglycine lysine antibody conjugated agarose beads	PTM Biolabs
Acetilación de lisinas	Acetyl-Lysine Motif [Ac-K] Kit #13416	PTMScan® Cell Signaling Technologies
	Acetyl Lysine Antibody ICP0388-5MG	ImmuneChem
	Pan anti-acetylysine antibody conjugated agarose beads	PTM Biolabs

Tabla 1. Listado de anticuerpos comerciales pan-específicos para purificación de modificaciones post-traduccionales tales como la metilación, acetilación y ubiquitinación.

2.3.1. Eficiencia y selectividad en la inmunopurificación

Antes de comenzar esta sección es importante definir dos conceptos mediante los cuales se determina la calidad de una inmunopurificación: la selectividad y la eficiencia. La selectividad hace referencia a la especificidad del proceso, es decir, a la capacidad para purificar los péptidos que portan la modificación de interés. Por tanto, para lograr una alta selectividad hay que optimizar los parámetros que maximicen la identificación de la modificación de interés y reduzcan la co-purificación de especies no deseadas o las interacciones inespecíficas. Por otro lado, la eficiencia de la inmunopurificación

se refiere al número absoluto de péptidos modificados que han sido identificados en el experimento. Cuanto más eficiente sea la inmunopurificación, mayor número de sitios de modificación serán identificados.

En primer lugar, hay que tener en cuenta que la eficiencia de la inmunopurificación de PTMs debe hacer frente a la problemática del rango dinámico. La presencia de un grupo de proteínas muy abundante que porta la modificación de interés puede obstaculizar la detección del resto del proteoma modificado. Este es el caso de la metilación y la acetilación de histonas, cuyos péptidos modificados son muy abundantes y dificultan la detección del resto. Por ello, para mejorar el rango dinámico del análisis la muestra, se puede pre-fraccionar mediante diferentes técnicas (e.g. cromatográficas, isoelectroenfoque, purificación sub-celular...). De este modo se disminuye la complejidad de la muestra y se facilita la detección de especies poco abundantes. Para ello, es fundamental elegir un método de fraccionamiento que sea ortogonal al método de separación cromatográfica en fase reversa normalmente acoplado al espectrómetro de masas. En este sentido, la cromatografía de intercambio catiónico fuerte o SCX por sus siglas en inglés (*Strong Cation Exchange*)(9,30) y la cromatografía en fase reversa a pH básico han demostrado un alto grado de ortogonalidad(36,37) mejorando muy significativamente el rango dinámico del análisis. Comúnmente, el proceso de fraccionamiento precede al de inmunopurificación, por lo que hay que tener en cuenta que recurrir a este procedimiento supone incrementar el número de inmunopurificaciones a realizar, tantas como fracciones se obtengan. Esto, al igual que sucede al introducir más pasos en un protocolo experimental, puede suponer un aumento de la variabilidad debido al tratamiento de la muestra. Por ello, también puede barajarse la posibilidad de realizar el fraccionamiento de la muestra tras la inmunopurificación. El principal inconveniente de esta aproximación es que el material biológico resultante de una inmunopurificación es muy reducido, y someterlo a un proceso de fraccionamiento puede incrementar las pérdidas. Debe valorarse, por tanto, muy cuidadosamente si conviene realizar el procedimiento de esta manera, sólo aconsejable si se dispone de material de partida suficiente. Sin embargo, la implementación de métodos de fraccionamiento que minimizan la pérdida de muestra ha sido descrito recientemente(38), ofreciendo una posibilidad real de análisis proteómico para cantidades muy reducidas de muestra, como sucede en las inmunopurificaciones.

Finalmente, para maximizar la eficiencia y la selectividad de una inmunopurificación debe hacerse un esfuerzo en la optimización de varios parámetros experimentales: cantidad de material biológico (*i.e.*: proteína) de partida, ratio péptido-anticuerpo, tiempo de incubación, lavados y elución.

2.3.2. Material de partida

Según lo descrito en la literatura, el material de partida para purificar péptidos acetilados, metilados o ubiquitinados es muy variable, oscilando entre 2 y 20 mg de proteína en distintas publicaciones. Para obtener tales cantidades de proteína es necesario escalar los sistemas biológicos en los que se quiera realizar en estudio. Aunque esto puede suponer un obstáculo en muchas situaciones, esta necesidad surge con el fin de minimizar las pérdidas, lo cual es crucial en estos protocolos de inmunopurificación en los que la cantidad de péptido resultante es muy pequeña. Por tanto, maximizar la



cantidad de proteína puede ayudar a aumentar la eficiencia de la inmunopurificación. No obstante, si el sistema no permite ser escalado, el proceso también se puede realizar con cantidades de proteína de partida menores, extremando en este caso las medidas para evitar pérdidas bien por adhesión a los plásticos del material de trabajo o por precipitación, entre otras causas.

Estas cantidades de proteína tan elevadas, en el rango de los miligramos, contrastan con las empleadas en métodos de purificación de fosfopéptidos, los cuales se han implementado de tal modo que la purificación es eficiente con cantidades de partida tan pequeñas como 2.5 μg (39). Una posible razón de esta divergencia podría encontrarse en la naturaleza de estas modificaciones en un sistema biológico, dado que la fosforilación es más abundante que la acetilación, la ubiquitinación y la metilación(40).

2.3.3. *Ratio péptido-anticuerpo*

El procedimiento ideal para maximizar eficiencia y selectividad consiste en realizar curvas de titulación en las que se valore el resultado de la inmunopurificación a diferentes ratios de proteína/péptido y anticuerpo. El resultado óptimo será aquel que maximice la eficiencia y alcance una gran selectividad. Por un lado, un uso excesivo de anticuerpo puede provocar un incremento de las interacciones inespecíficas. Pero, por el contrario, cantidades limitantes de anticuerpo pueden resultar en una alta selectividad pero en un número final de péptidos modificados sub-óptimo. Aunque maximizar la selectividad es muy deseable, esto nunca debe perjudicar a la eficiencia del procedimiento(23).

La cantidad de anticuerpo que se debe emplear depende directamente de la cantidad de péptido de la que se disponga. Svinkina *et al*(23) determinan mediante curvas de titulación que para 2 mg de péptido la cantidad óptima de anticuerpo anti-acetil-lisina es 25 μg . No obstante, cuando se dispone de los anticuerpos conjugados a partículas en solución, la cantidad de anticuerpo a añadir suele indicarse en volumen. Por ejemplo, Martínez-Val *et al*(41) emplean 20 μl del anticuerpo conjugado por cada 12 mg de péptido. En cualquier caso, estos valores son orientativos y siempre es aconsejable implementar este parámetro del protocolo.

2.3.4. *Incubación*

La incubación del péptido junto con el anticuerpo debe realizarse en un búfer que favorezca la interacción entre el péptido y el anticuerpo. Dicha interacción es bastante estable, por lo que tan sólo se requiere un búfer estándar a un pH cercano a la neutralidad (comúnmente se emplea MOPS 50 mM pH 7). Es conveniente realizar el proceso de inmunopurificación a baja temperatura (i.e. 4°C), con el fin de minimizar, en la medida de lo posible, las interacciones inespecíficas de péptidos no modificados con la matriz de afinidad. A su vez, el uso de determinados detergentes durante la inmunopurificación también previene este tipo de interacciones inespecíficas y puede contribuir al aumento de la selectividad del ensayo(42). Por último, el tiempo de incubación recomendable varía entre diferentes publicaciones y fabricantes, oscilando entre 1 y 12 horas de incubación.

2.3.5. Lavado y elución de péptidos

Tras la incubación es necesario proceder a realizar varios lavados destinados a eliminar aquellos péptidos que hayan sido retenidos de manera inespecífica durante la incubación. Los lavados deben realizarse en una solución que no altere la conformación del anticuerpo para no romper las interacciones de interés, bien sea en el mismo tampón en el que se ha realizado la incubación, con PBS o con pequeñas concentraciones de detergentes o sales para eliminar interacciones inespecíficas. También puede resultar beneficioso cambiar de vial durante los lavados, dado que los péptidos pueden adherirse a los materiales plásticos de los que están fabricados los viales, y éstos péptidos pueden ser liberados durante la elución, lo cual es indeseable. Cambiar el vial tras la incubación y antes de los lavados puede ayudar, por tanto, a aumentar la selectividad del proceso.

Por último, la elución de los péptidos suele realizarse mediante la desnaturalización del anticuerpo, lo que provoca la pérdida de su conformación tridimensional y rompe la interacción antígeno-anticuerpo. La desnaturalización debe realizarse con búferes compatibles con el posterior análisis mediante espectrometría de masas. Esto descarta el uso de la mayor parte de los detergentes. Por el contrario, la elución en medio ácido (e.g: ácido fórmico) permite la desnaturalización del anticuerpo y es compatible con MS.

2.4. Análisis mediante LC-MS/MS

El campo de la proteómica basada en la espectrometría de masas ha evolucionado exponencialmente durante la pasada década. Las principales mejoras han ido dirigidas a incrementar la sensibilidad y la rapidez de los equipos con el fin último de incrementar la cobertura del proteoma a la vez que se reduce el tiempo de análisis. De este modo, en 2013 se mapeó el proteoma de la levadura en tan solo una hora de análisis(43), una cobertura similar a la que se había alcanzado años antes necesitando en ese caso de 144 horas de análisis(44). Desde entonces la tecnología ha continuado avanzando, y en 2017 se publicó la descripción más completa hasta la fecha, del proteoma de células HeLa en 34,5 horas (~11.292 proteínas y ~166.620 péptidos únicos) mediante el fraccionamiento previo de la muestra y su posterior análisis en un equipo Q Exactive HF capaz de funcionar a 20 Hz(45). Estas mejoras han repercutido de manera muy significativa, tanto en la identificación y cuantificación de proteomas completos, como en el análisis de modificaciones post-traduccionales. Esto es así porque, como hemos apuntado anteriormente, las muestras de ensayos de PTMs se caracterizan por necesitar una gran sensibilidad de análisis debido a la pequeña cantidad de muestra que se obtiene tras la inmunopurificación. La velocidad alcanzada por estos nuevos equipos también permite incrementar el número de las identificaciones y superar ciertas limitaciones impuestas por el rango dinámico de la muestra, el cual puede ser un gran obstáculo para alcanzar una adecuada profundidad de análisis en proteomas modificados.

Por otro lado, la resolución de masa que proporciona cada tipo de espectrómetro de masas influye en gran manera en nuestra habilidad para identificar modificaciones post-traduccionales. La resolución determina cual es la mínima diferencia de masas



entre dos péptidos para detectarlos como elementos independientes. Esto es fundamental para identificar péptidos modificados, ya que su identificación se basa en detectar el incremento en masa que la modificación aporta al péptido. Analizadores de masas de baja resolución, como las trampas iónicas, no son capaces de distinguir entre modificaciones que producen incrementos de masa muy similares, tales como la acetilación (42.010565 Da) y la trimetilación (42.046950 Da), que sólo se diferencian en 0.03638 Da o la fosforilación (79.966331 Da) y la O-sulfonación (79.956815 Da), sólo difieren en 0.00951 Da.

Finalmente, tras la identificación de la masa del péptido le sigue la fragmentación del precursor para identificar su secuencia y determinar el aminoácido donde se localiza la modificación. Por tanto, el método de fragmentación usado también puede determinar la capacidad para identificar determinadas modificaciones. Para modificaciones covalentes estables, como son la acetilación, la metilación y la ubiquitinación, los métodos de HCD (*high energy collisional dissociation*) y CID (*collision induced dissociation*) son los más empleados. No obstante, en modificaciones como la fosforilación, la localización del residuo fosforilado puede ser ambigua si se emplean los métodos de fragmentación anteriores. La energía liberada por HCD o CID incide sobre los enlaces más lábiles (los del grupo fosfato) y provoca patrones de fragmentación que dificultan la localización de la modificación. Este mismo efecto también se ha descrito en metilación de argininas, donde la fragmentación con CID provoca pérdida neutra, lo cual puede dificultar la localización del residuo modificado(46). Como alternativa para estos casos, los métodos de fragmentación basados en la captura o transferencia de electrones (*electron capture dissociation* o ECD y *electron-transfer dissociation* o ETD) pueden ser más convenientes ya que este tipo de fragmentaciones conservan las modificaciones que son más lábiles. Además, se ha descrito que la complementariedad de estas fragmentaciones puede ser útil para incrementar la identificación de residuos ubiquitinados, acetilados y metilados(47). Se ha observado que la fragmentación por ETD favorece aquellos péptidos con más carga(48,49) y los péptidos ubiquitinados tienen con frecuencia más carga que los no-ubiquitinados debido a la presencia del extremo N-terminal adicional aportado por el residuo diGly. Del mismo modo, los péptidos acetilados y metilados suelen estar más cargados por poseer residuos básicos internos en el péptido.

Por tanto, hay varios parámetros relacionados con la adquisición de datos que determinan el éxito del análisis de acetilaciones, metilaciones y ubiquitinaciones. En primer lugar, la rapidez y sensibilidad del equipo de espectrometría de masas puede influir de manera muy positiva en la profundidad de análisis. En segundo lugar, la resolución que alcance el analizador de masas es determinante a la hora de distinguir la modificación asociada a un determinado incremento de masa, lo cual puede ser fundamental para distinguir, por ejemplo acetilación de trimetilación en lisinas. Por último, aunque los métodos de fragmentación más empleados para estudiar las modificaciones tratadas en este capítulo son CID y HCD, puede ser interesante complementarlos con otros métodos de fragmentación (e.g. ETD).

2.5. Identificación de PTMs

Como hemos mencionado anteriormente, el incremento sobre el valor de la masa

del péptido no modificado (Tabla 2, Figura 2) es lo que permite la identificación de los péptidos modificados utilizando motores de búsqueda. Por ello, para analizar muestras enriquecidas en una determinada PTM hay que incluir dicha modificación como variable en los parámetros del motor de búsqueda. El motor de búsqueda contemplará por tanto la existencia de péptidos con el incremento de masa asociado a la modificación.

Es importante tener en cuenta que al realizar una búsqueda incluyendo modificaciones post-traduccionales, el espacio de búsqueda aumenta considerablemente ya que al conjunto de péptidos generados por la digestión *in silico* de todas las proteínas de la base de datos, se suman todos los posibles péptidos modificados. Esto provoca, no sólo un incremento en el tiempo de procesamiento de los datos, sino que

Modificación	Incremento de masa (Da)
Ubiquitinación (residuo diGly)	114.042927
Acetilación	42.010565
Mono-metilación	14.014565
Di-metilación	28.0313
Tri-metilación	42.05495

Tabla 2. Incremento de masa en Dalton que provocan las modificación del péptido por el residuo diGly de la ubiquitina, la acetilación o la metilación.

puede tener un impacto negativo sobre el FDR del análisis. Para intentar minimizar el impacto de este incremento del espacio de búsqueda se han propuesto algunas alternativas. Un ejemplo es la estrategia G-PTM(50) que propone reducir el espacio de búsqueda incluyendo en la base de datos sólo aquellos sitios de modificación que hayan sido previamente descritos. Algunos algoritmos, como Mascot Delta Score(51) o PTM Score implementado en Andrómeda(52) emplean la identificación de series de pérdida neutra o la presencia de iones diagnóstico para reforzar la identificación de péptidos modificados. La pérdida neutra es la eliminación de la modificación durante la fragmentación, esto puede dar como resultado que en el espectro de fragmentación aparezcan dos series de iones: con el incremento en masa debido a la modificación y sin él debido a la pérdida neutra. La pérdida neutra es un fenómeno asociado a modificaciones lábiles durante la fragmentación, lo cual no sucede habitualmente en las modificaciones tratadas en este capítulo. No obstante, se ha descrito que la trimetilación de lisinas da lugar a una pérdida neutra de 59 Da. En equipos de baja resolución que no permiten diferenciar entre acetilación (42.010565 Da) y trimetilación (42.05495 Da), la presencia de la serie de iones con pérdida neutra es clave para diferenciar entre ambas modificaciones(53). En otras ocasiones, la fragmentación puede dar lugar a iones imonio (Tabla 3), los cuales pueden ayudar a identificar la presencia de una determinada modificación en un péptido. Este fenómeno sí es común en péptidos acetilados (Figura 3) y metilados, pero no está descrito para péptidos ubiquitinados.

Además de la identificación de un péptido modificado, es importante determinar la localización exacta del residuo modificado dentro de la secuencia del péptido en cuestión. Esta es una de las ventajas de emplear la proteómica basada en espectrometría de masas para estudiar PTMs, que permite identificar el sitio de modificación en la proteína. Sin embargo, la correcta asignación del lugar de modificación depende en gran medida de los algoritmos informáticos encargados de la identificación de péptidos. Cuando un péptido sólo contiene un aminoácido susceptible de portar la modi-



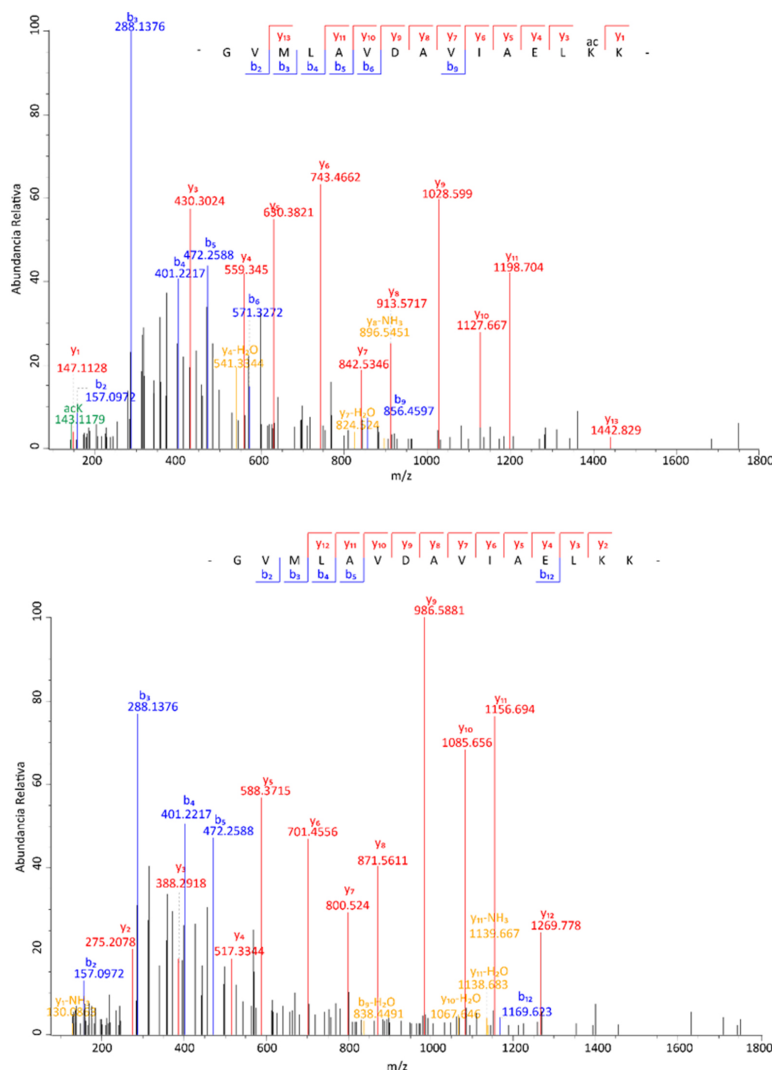


Figura 2. Espectros de fragmentación anotados del mismo péptido, el primero corresponde a su versión modificada por una acetilación y el segundo al péptido no modificado. Se puede observar que los iones de la serie y muestran un incremento en masa de 42 Da a partir del y₂. Los espectros han sido anotados por MaxQuant 1.6.0.16.

ficación, no cabe ningún error en la asignación del residuo modificado. No obstante, si un péptido contiene dos o más aminoácidos susceptibles de ser modificados, la localización de la modificación en el péptido debe realizarse en base a la información contenida en el espectro de fragmentación. Cuando en el espectro MS/MS se observa claramente uno o más iones que puedan explicar la modificación y que deriven de la fragmentación entre dos aminoácidos que puedan contenerla, la asignación es inequívoca. No obstante, en ocasiones estos iones no están presentes, o no pueden ser distinguibles del ruido. Existen varios algoritmos desarrollados por la comunidad proteómica para establecer la localización más posible en estos casos, un ejemplo es el A-score(54) o 'ambiguity score' que es una métrica que mide la probabilidad de que una serie de iones entre dos posiciones puedan explicar la modificación ocurra de manera aleatoria. Para ello el algoritmo determina qué picos del espectro de fragmentación serían necesarios para localizar la modificación. Estos algoritmos se

Modificación	Aminoácido	Ion diagnóstico (m/z)
Acetilación	K	143.11844, 126.09189
Metilación	R	143.1291, 115.0866, 112.0869, 74.0713, 70.0651
Metilación	K	115.12352, 98.09697, 84.08132
Di-metilación	R	157.1448, 115.0866, 112.0869, 88.0869, 71.0604
Di-metilación	K	129.13917, 84.08132
Trimetilación	K	143.15482, 84.08132

Tabla 3. Masa de los iones diagnóstico liberados durante la fragmentación de péptidos acetilados y metilados.

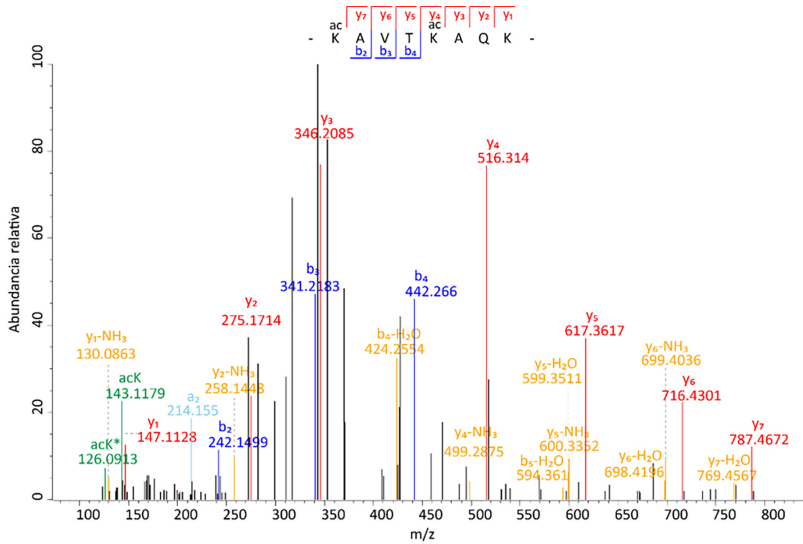


Figura 3. Espectro de fragmentación anotado correspondiente a un péptido doblemente acetilado. En la región de masas bajas se puede observar la presencia de dos iones diagnóstico característicos de esta modificación: 126.09 Da y 143.11 Da.

desarrollaron principalmente para determinar la localización de fosforilaciones, la cual puede ser compleja por la coexistencia de varios aminoácidos susceptibles de estar fosforilados en un mismo péptido. La determinación del lugar de modificación en péptidos metilados, acetilados o ubiquitinados, por el contrario, suele ser más sencilla. En una digestión triptica las proteínas se cortan en lisinas y argininas, salvo si éstas se encuentran modificadas. Por tanto, tras una digestión triptica, un péptido metilado, acetilado o ubiquitinado portará la modificación en la lisina o arginina interna del péptido. De hecho, muchos autores descartan como falsos positivos aquellas asignaciones de lisinas acetiladas, metiladas o ubiquitinadas en los extremos C-terminal de los péptidos. Por esta razón, los algoritmos de localización aplicados a este tipo de modificaciones reportan una tasa de residuos erróneamente localizados mucho menor que en casos de fosforilación.

2.6. Análisis cuantitativo de PTMs

El análisis de PTMs mediante proteómica basada en espectrometría de masas permite, además de identificar los residuos modificados, determinar los cambios en la abun-



dancia de una determinada modificación entre diferentes condiciones experimentales (e.g. un estado patológico y uno sano). Este proceso puede llevarse a cabo mediante diferentes aproximaciones: cuantificación sin marcaje o *label-free*, cuantificación con marcaje metabólico(55) (e.g. SILAC), cuantificación por marcaje químico (e.g. dimetil) y marcaje isobárico(56) (e.g. iTRAQ o TMT). La aproximación por *label-free* es la más sencilla experimentalmente pero es más sensible a la variabilidad técnica, ya que cada muestra se procesa y analiza de forma independiente. Por el contrario, el marcaje metabólico minimiza la variabilidad técnica ya que todas las muestras se procesan conjuntamente. No obstante, el marcaje mediante SILAC sólo permite comparar un máximo de tres condiciones. El marcaje isobárico por su parte permite incrementar el número de muestras a comparar hasta once. Sin embargo, tanto en TMT como el iTRAQ las muestras se lisan y digieren por separado antes del marcaje y, además, las mediciones derivadas de estos ensayos pueden sufrir una pérdida de exactitud debido a la compresión de ratios(57). Hay muchos ejemplos del uso de estas técnicas en estudios de PTMs aunque es importante tener en cuenta las particularidades de un ensayo de enriquecimiento con anticuerpos antes de elegir el método de cuantificación. Por ejemplo, si se opta por algún tipo de marcaje químico hay que decidir entre realizar la inmunopurificación antes o después del marcaje. Si se realiza antes sería necesario escalar la cantidad de reactivo de marcaje a las altas cantidades de proteína recomendadas en este tipo de experimento de enriquecimiento, lo cual puede incrementar mucho el coste económico del experimento.

Una diferencia clave entre un análisis cuantitativo de proteoma global y uno de PTM es que la cuantificación se realiza a nivel de péptido y no de proteína. Esto supone un reto añadido puesto que necesariamente la cuantificación se hace en base a uno o pocos péptidos, y normalmente de baja abundancia, lo cual puede disminuir la precisión de la medida.

Por otro lado, en los estudios cuantitativos de PTMs es fundamental tener en cuenta que los cambios en la abundancia de una determinada modificación pueden ser consecuencia del cambio en los niveles de expresión de la proteína. Para evitar caer en ese error es conveniente realizar un estudio del proteoma global y emplear los valores de expresión de la proteína para normalizar la abundancia de los residuos modificados(58). Hay herramientas, como IsobarPTM, que permiten realizar la cuantificación y normalización por cambios en el proteoma global de estudios de modificaciones con marcaje isobárico(59).

Además de la cuantificación relativa puede ser interesante determinar la estequiometría de una modificación, es decir, la fracción de la proteína que se encuentra modificada respecto al total. Aunque a día de hoy el estudio de la estequiometría de la acetilación, metilación y ubiquitinación sigue siendo un reto, algunos grupos han implementado métodos para medir la estequiometría de estas modificaciones. Una posible aproximación consiste en la comparación de los péptidos modificados de manera nativa frente al resto de péptidos, los cuales deberán haber sido modificados químicamente. Por ejemplo, Zhou *et al*(60) proponen marcar *in vitro* las lisinas no acetiladas con un reactivo (NHS-acetil ester), el cual incrementa la masa de los péptidos pero no altera sus comportamiento cromatográfico. De este modo eluirán conjuntamente el péptido acetilado y su pareja sin acetilar y se podrá cuantificar la diferencia de abundancia entre ambos.

3. Nuevas posibilidades en la identificación de modificaciones post-traduccionales

A lo largo de este capítulo hemos expuesto los avances alcanzados en el estudio de PTMs mediante espectrometría de masas gracias al uso de anticuerpos para la purificación de péptidos modificados. También hemos destacado el impacto que ha tenido el desarrollo de los equipos de espectrometría de masas en este campo. Sin embargo, parece ser que estas mejoras alcanzadas desde el plano tecnológico pueden implicar un cambio de paradigma en el futuro estudio de modificaciones post-traduccionales. Como hemos venido explicando en este capítulo, la purificación de los péptidos modificados era fundamental para reducir el rango dinámico de la muestra y poder así detectar especies que, de otro modo, quedarían ocultas por los péptidos no modificados mayoritarios. No obstante, los equipos de MS de última generación, gracias principalmente a su rapidez de análisis, pueden vencer este obstáculo. De hecho, en una publicación de 2017, Bekker-Jensen *et al*(45) proponen un método de análisis de proteomas complejos tan extensivo y con una cobertura del proteoma tan completa que permite detectar modificaciones post-traduccionales sin necesidad de enriquecimiento previo. Tan sólo incluyendo las modificaciones de interés en el motor de búsqueda pudieron identificar más de 10,000 sitios de fosforilación y 5,000 sitios de acetilación en N-terminal. Aunque estos números, para el caso de la fosforilación, están lejos de abarcar el fosfoproteoma completo(61), se trata de una posibilidad muy útil para complementar estudios de proteoma global.

Otra alternativa al estudio tradicional de PTMs proviene del campo de la bioinformática. Chick *et al*(62) propusieron un método de búsqueda imparcial de modificaciones en péptidos. Este método, al que denominaron 'Mass-tolerant database search' consiste

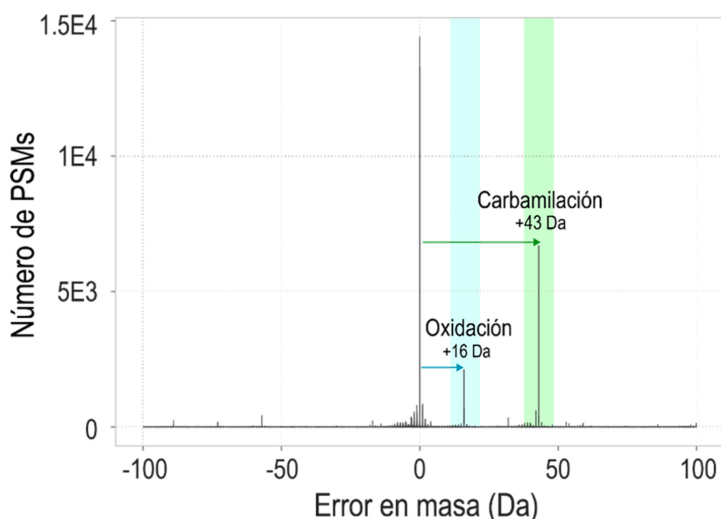


Figura 4. Histograma que representa la frecuencia de péptidos identificados en una búsqueda con tolerancia de masa de ± 100 Da. Los péptidos no modificados se encuentran representados en el valor 0 de error en masa del precursor. Sin embargo, se observan dos distribuciones significativas a +16 Da y +43 Da, las cuales pueden corresponder a péptidos con metioninas oxidadas y a péptidos carbamilados, respectivamente.



en realizar la búsqueda de los espectros frente a una base de datos sin indicar modificaciones, pero permitiendo un error de masa del precursor muy elevado. De este modo, se podrían identificar poblaciones de péptidos modificados situados en determinadas diferencias de masa (Figura 4). Esta estrategia requiere un paso posterior que permita emparejar los incrementos de masa observados con modificaciones químicas asociadas a los péptidos. Esta opción es especialmente útil cuando no se conoce la modificación presente en una determinada muestra. Su principal inconveniente es el empleo de grandes ventanas de tolerancia de precursor, lo cual ralentiza el proceso de búsqueda.

BIBLIOGRAFÍA

1. Creasy DM, Cottrell JS. Unimod: Protein modifications for mass spectrometry. *Proteomics*. 2004 Jun;4(6):1534–6.
2. Ye Y, Rape M. Building ubiquitin chains: E2 enzymes at work. *Nat Rev Mol Cell Biol*. 2009 Nov;10(11):755–64.
3. Ye Y, Blaser G, Horrocks MH, et al. Ubiquitin chain conformation regulates recognition and activity of interacting proteins. *Nature*. 2012 Dec 13;492(7428):266–70.
4. Xu P, Duong DM, Seyfried NT, et al. Quantitative Proteomics Reveals the Function of Unconventional Ubiquitin Chains in Proteasomal Degradation. *Cell*. 2009 Apr 3;137(1):133–45.
5. Lauwers E, Jacob C, André B. K63-linked ubiquitin chains as a specific signal for protein sorting into the multivesicular body pathway. *J Cell Biol*. 2009 May 4;185(3):493–502.
6. Deng L, Wang C, Spencer E, et al. Activation of the I κ B kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell*. 2000 Oct 13;103(2):351–61.
7. Van Dyke MW. Lysine deacetylase (KDAC) regulatory pathways: an alternative approach to selective modulation. *ChemMedChem*. 2014 Mar;9(3):511–22.
8. Choudhary C, Kumar C, Gnad F, et al. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* (80-). 2009 Aug 14;325(5942):834–40.
9. Schölz C, Weinert BT, Wagner SA, et al. Acetylation site specificities of lysine deacetylase inhibitors in human cells. *Nat Biotechnol*. 2015 Mar 9;33(4):415–23.
10. Geoghegan V, Guo A, Trudgian D, et al. Comprehensive identification of arginine methylation in primary T cells reveals regulatory enzyme complex and a unique polyubiquitin chain. *Cell*. 2015 Apr 7;6:6758.
11. Huang J, Perez-Burgos L, Placek BJ, et al. Repression of p53 activity by Smyd2-mediated methylation. *Nature*. 2006 Nov 15;444(7119):629–32.
12. Chuikov S, Kurash JK, Wilson JR, et al. Regulation of p53 activity through lysine methylation. *Nature*. 2004 Nov 18;432(7015):353–60.
13. Moore KE, Carlson SM, Camp ND, et al. A General Molecular Affinity Strategy for Global Detection and Proteomic Analysis of Lysine Methylation. *Mol Cell*. 2013 May 9;50(3):444–56.
14. Calnan BJ, Tidor B, Biancalana S, Hudson D, Frankel AD. Arginine-mediated RNA recognition: the arginine fork. *Science*. 1991 May 24;252(5009):1167–71.
15. Goulet I, Gauvin G, Boisvenue S, Côté J. Alternative Splicing Yields Protein Arginine Methyltransferase 1 Isoforms with Distinct Activity, Substrate Specificity, and Subcellular Localization. *J Biol Chem*. 2007 Nov 9;282(45):33009–21.
16. Pinkse MWH, Uitto PM, Hilhorst MJ, Ooms B, Heck AJR. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem*. 2004 Jul 15;76(14):3935–43.
17. Villén J, Gygi SP. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc*. 2008 Sep;3(10):1630–8.
18. Zhou H, Low TY, Hennrich ML, et al. Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. *Mol Cell Proteomics*. 2011 Oct;10(10):M110.006452.
19. Ruprecht B, Koch H, Medard G, Mundt M, Kuster B, Lemeer S. Comprehensive and Reproducible Phosphopeptide Enrichment Using Iron Immobilized Metal Ion Affinity Chromatography (Fe-IMAC) Columns. *Mol Cell Proteomics*. 2015 Jan;14(1):205–15.
20. Engholm-Keller K, Larsen MR. Titanium dioxide as chemo-affinity chromatographic sorbent of biomolecular compounds--applications in acidic modification-specific proteomics. *J Proteomics*. 2011 Dec 21;75(2):317–28.
21. Larsen MR, Jensen SS, Jakobsen LA, Heegaard NHH. Exploring the sialome using titanium dioxide chromatography and mass spectrometry. *Mol Cell Proteomics*. 2007 Oct;6(10):1778–87.
22. Lundby A, Lage K, Weinert BT, et al. Proteomic analysis of lysine acetylation sites in rat tissues reveals organ specificity and subcellular patterns. *Cell Rep*. 2012 Aug 30;2(2):419–31.
23. Svinkina T, Gu H, Silva JC, et al. Deep, Quantitative Coverage of the Lysine Acetylome Using Novel Anti-acetyl-ly-

sine Antibodies and an Optimized Proteomic Workflow. *Mol Cell Proteomics*. 2015 Sep;14(9):2429–40.

24. Xu G, Paige JS, Jaffrey SR. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat Biotechnol*. 2010 Aug;28(8):868–73.
25. Kim W, Bennett EJ, Huttlin EL, et al. Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome. *Mol Cell*. 2011;44(2):325–40.
26. Ong S-E, Mittler G, Mann M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat Methods*. 2004 Nov 21;1(2):119–26.
27. Uhlmann T, Geoghegan VL, Thomas B, Ridlova G, Trudgian DC, Acuto O. A Method for Large-scale Identification of Protein Arginine Methylation. *Mol Cell Proteomics*. 2012 Nov;11(11):1489–99.
28. Guo A, Gu H, Zhou J, et al. Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. *Mol Cell Proteomics*. 2014 Jan;13(1):372–87.
29. Swaney DL, Beltrao P, Starita L, et al. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat Methods*. 2013 Jun 9;10(7):676–82.
30. Elia AEH, Boardman AP, Wang DC, et al. Quantitative Proteomic Atlas of Ubiquitination and Acetylation in the DNA Damage Response. *Mol Cell*. 2015 Sep 3;59(5):867–81.
31. Mertins P, Qiao JW, Patel J, et al. Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat Methods*. 2013 Jul;10(7):634–7.
32. Martinez-Val A, Garcia F, Ximénez-Embún P, et al. Urea Artifacts Interfere with Immuno-Purification of Lysine Acetylation. *J Proteome Res*. 2017 Feb 3;16(2):1061–8.
33. Nielsen ML, Vermeulen M, Bonaldi T, Cox J, Moroder L, Mann M. Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat Methods*. 2008 Jun;5(6):459–60.
34. Chen G, Liu H, Wang X, Li Z. In vitro methylation by methanol: Proteomic screening and prevalence investigation. *Anal Chim Acta*. 2010 Feb;661(1):67–75.
35. Hart-Smith G, Yagoub D, Tay AP, Pickford R, Wilkins MR. Large Scale Mass Spectrometry-based Identifications of Enzyme-mediated Protein Methylation Are Subject to High False Discovery Rates. *Mol Cell Proteomics*. 2016 Mar;15(3):989–1006.
36. Batth TS, Francavilla C, Olsen J V. Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J Proteome Res*. 2014;13(12):6176–86.
37. Udeshi ND, Svinkina T, Mertins P, et al. Refined preparation and use of anti-diglycine remnant (K-ε-GG) antibody enables routine quantification of 10,000s of ubiquitination sites in single proteomics experiments. *Mol Cell Proteomics*. 2013 Mar;12(3):825–31.
38. Kulak NA, Geyer PE, Mann M. Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics. *Mol Cell Proteomics*. 2017 Apr;16(4):694–705.
39. Post H, Penning R, Fitzpatrick MA, et al. Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J Proteome Res*. 2017 Feb 3;16(2):728–37.
40. Khoury GA, Baliban RC, Floudas CA. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep*. 2011 Dec 13;1(1):90.
41. Martinez-Val A, Garcia F, Ximénez-Embún P, et al. Urea Artifacts Interfere with Immuno-Purification of Lysine Acetylation. *J Proteome Res*. 2017 Feb 3;16(2):1061–8.
42. Zhang G, Neubert TA. Use of detergents to increase selectivity of immunoprecipitation of tyrosine phosphorylated peptides prior to identification by MALDI quadrupole-TOF MS. *Proteomics*. 2006 Jan;6(2):571–8.
43. Hebert AS, Richards AL, Bailey DJ, et al. The One Hour Yeast Proteome. *Mol Cell Proteomics*. 2014 Jan;13(1):339–47.
44. de Godoy LMF, Olsen J V, Cox J, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*. 2008 Oct 30;455(7217):1251–4.
45. Bekker-Jensen DB, Kelstrup CD, Batth TS, et al. An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst*. 2017 Jun 28;4(6):587–99.e4.
46. Wang H, Straubinger RM, Aletta JM, et al. Accurate localization and relative quantification of arginine methylation using nanoflow liquid chromatography coupled to electron transfer dissociation and orbitrap mass spectrometry. *J Am Soc Mass Spectrom*. 2009 Mar;20(3):507–19.
47. Sobott F, Watt SJ, Smith J, Edelmann MJ, Kramer HB, Kessler BM. Comparison of CID Versus ETD Based MS/MS Fragmentation for the Analysis of Protein Ubiquitination. *J Am Soc Mass Spectrom*. 2009 Sep 1;20(9):1652–9.
48. Snijders APL, Hung M-L, Wilson SA, Dickman MJ. Analysis of Arginine and Lysine Methylation Utilizing Peptide Separations at Neutral pH and Electron Transfer Dissociation Mass Spectrometry. *J Am Soc Mass Spectrom*. 2010 Jan 1;21(1):88–96.
49. Porras-Yakushi TR, Sweredoski MJ, Hess S. ETD Outperforms CID and HCD in the Analysis of the Ubiquitylated Proteome. *J Am Soc Mass Spectrom*. 2015 Sep 21;26(9):1580–7.
50. Shortreed MR, Wenger CD, Frey BL, et al. Global Identification of Protein Post-translational Modifications in a Single-Pass Database Search. *J Proteome Res*. 2015 Nov 6;14(11):4714–20.
51. Savitski MM, Lemeer S, Boesche M, et al. Confident Phosphorylation Site Localization Using the Mascot Delta Score. *Mol Cell Proteomics*. 2011 Feb;10(2):M110.003830.
52. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen J V, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*. 2011 Apr 1;10(4):1794–805.
53. Zhang K, Yau PM, Chandrasekhar B, et al. Differentiation between peptides containing acetylated or tri-methylated lysines by mass spectrometry: An application for determining lysine 9 acetylation and methylation of histone H3. *Proteomics*. 2004 Jan;4(1):1–10.



54. Beausoleil SA, Villén J, Gerber SA, Rush J, Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol.* 2006 Oct 10;24(10):1285–92.
55. 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 May;1(5):376–86.
56. Thompson A, Schäfer J, Kuhn K, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem.* 2003 Apr 15;75(8):1895–904.
57. Martinez-Val A, Garcia F, Ximénez-Embún P, et al. On the Statistical Significance of Compressed Ratios in Isobaric Labeling: A Cross-Platform Comparison. *J Proteome Res.* 2016 Sep 2;15(9):3029–38.
58. Wu R, Dephoure N, Haas W, et al. Correct Interpretation of Comprehensive Phosphorylation Dynamics Requires Normalization by Protein Expression Changes. *Mol Cell Proteomics.* 2011 Aug;10(8):M111.009654.
59. Breitwieser FP, Colinge J. Isobar(PTM): a software tool for the quantitative analysis of post-translationally modified proteins. *J Proteomics.* 2013 Sep 2;90:77–84.
60. Zhou T, Chung Y, Chen J, Chen Y. Site-Specific Identification of Lysine Acetylation Stoichiometries in Mammalian Cells. *J Proteome Res.* 2016 Mar 4;15(3):1103–13.
61. Sharma K, D'Souza RCJ, Tyanova S, et al. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Rep.* 2014 Aug 11;8(5):1583–94.
62. Chick JM, Kolippakkam D, Nusinow DP, et al. A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat Biotechnol.* 2015;33(7):743–9. Introducción

La estrategia proteómica top-down (Top-down proteomics)

Marta Vilaseca¹, Marina Gay¹

¹Mass Spectrometry and Proteomics Core Facility, Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 10, 08028 Barcelona, Spain.

1. Introducción.

La proteómica es una herramienta esencial para caracterizar las proteínas y descifrar cómo interactúan en las células para desarrollar funciones biológicas específicas. La tecnología actual permite responder muchas preguntas en biología y a medida que avanzamos en el conocimiento de nuestros sistemas biológicos se originan preguntas aún más complejas que constituyen cada vez mayores retos analíticos. Comprender las funciones de los sistemas celulares en las enfermedades humanas es un desafío dado las características del proteoma humano, que incluyen, un alto rango dinámico de expresión de proteínas, que actúan de manera asociada y que poseen una complejidad extrema debido a la plétora de modificaciones y variaciones de secuencias existentes. Estimaciones recientes sugieren que el número total de posibles variantes de proteínas humanas está en algún punto entre los 10 millones [1] y 1 billón [2, 3].

Una gran parte de experimentos proteómicos emplean la estrategia de análisis de proteínas bottom-up - *proteómica clásica*- para responder múltiples preguntas respecto a la identidad de proteínas, modificaciones y abundancia relativa o absoluta en mezclas muy complejas y proteomas enteros. Es una estrategia altamente desarrollada, tanto a nivel técnico como de análisis de datos. De manera simplista, en la estrategia bottom-up, las proteínas son digeridas con proteasas y posteriormente, los péptidos resultantes, se analizan masivamente mediante análisis *highthroughput* o de alto rendimiento. La espectrometría de masas en tándem es usada para secuenciar los péptidos de proteomas enteros, y herramientas informáticas permiten identificar las secuencias peptídicas y asignarlas a proteínas concretas. Con esta misma estrategia se identifican las modificaciones post-traduccionales (PTMs) en los péptidos analizados. Esta estrategia es muy útil para identificar y cuantificar proteínas, así como para identificar sus modificaciones, pero presenta ciertas limitaciones derivadas de trabajar a nivel de péptido. Amplias regiones de las proteínas suelen no ser detectadas y, por tanto, pueden dejar atrás información importante respecto a truncamientos u otras variaciones de secuencia, sin dar la asignación correcta de los péptidos a una isoforma proteica concreta. Por otro lado, la relación entre los péptidos se pierde en la digestión con las proteasas y no es posible deducir la conectividad entre las modificaciones que componen las múltiples proteoformas presentes en las células. El término *proteoforma* designa todas las distintas formas moleculares en que podemos encontrar a la



proteína producto de un único gen, incluyendo todas las variantes genéticas, *splicing* alternativo de transcritos de RNA y modificaciones post-traduccionales (Figura 1) [2].

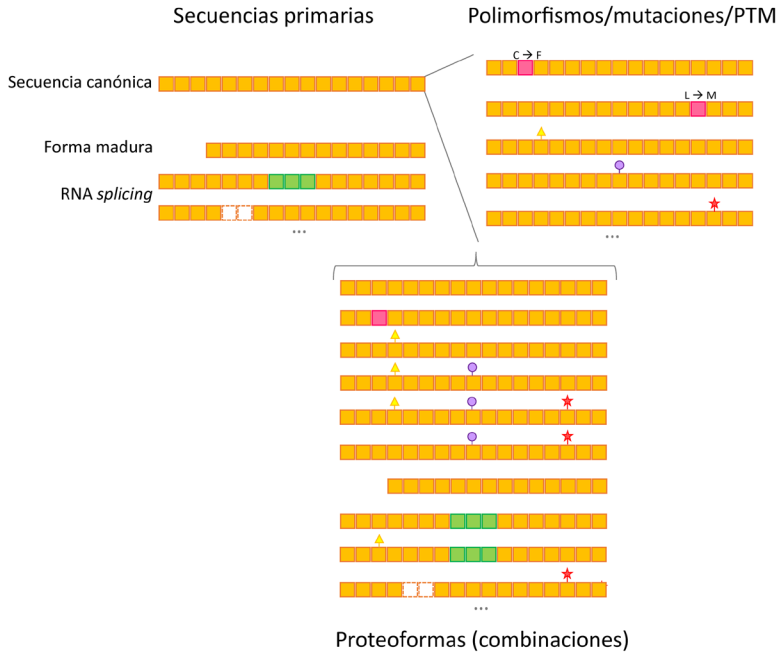


Figura 1. El término proteoforma describe la variabilidad biológica a nivel de estructura primaria de una proteína. Comprende la variabilidad de secuencia derivada de: provenir de más de un gen, o de un solo gen pero generando distintas variantes alternativas de *splicing*, modificaciones de RNA, múltiples sitios de iniciación translacionales en transcritos mRNA, o distintos sitios de deslizamiento, así como modificaciones post-traduccionales (PTMs). Estos mecanismos de diversificación generan millones de formas diferentes de proteínas, denominadas proteoformas. Estimaciones recientes sugieren un total de posibles variantes de proteínas humanas entre 10 millones y 1 billón [1, 2].

Pero, ¿cuál es la importancia de poder caracterizar las proteínas a nivel de molécula o proteoforma? Precisamente, la sofisticación funcional de los organismos complejos se basa en la habilidad de usar un número limitado de secuencias de genes en muchas formas diferentes. Conocer la proteoforma o proteoformas precisas implicada en una función concreta es clave para comprender los mecanismos moleculares usados por los organismos eucariotas [4].

La proteómica top-down o top-down proteomics se centra en el estudio de las proteoformas. Es la única estrategia que proporciona este tipo de información, dado que analiza las proteínas sin digestión previa, fragmentándolas directamente en fase gas. Las proteínas se introducen enteras en el espectrómetro de masas y se mide tanto su masa intacta como la masa de los iones fragmento mediante espectrometría de masas en tándem (Figura 2).

Los avances en prefraccionamiento de proteínas, instrumentación de MS y técnicas de disociación de iones de proteína intacta, han contribuido a que la proteómica top-down (TDP) este irrumpiendo como una modalidad accesible en sus distintas variantes. Es una estrategia en constante desarrollo, usada muchas veces de manera complementaria a la proteómica clásica, pero con un potencial disruptivo y un valor translacional creciente.

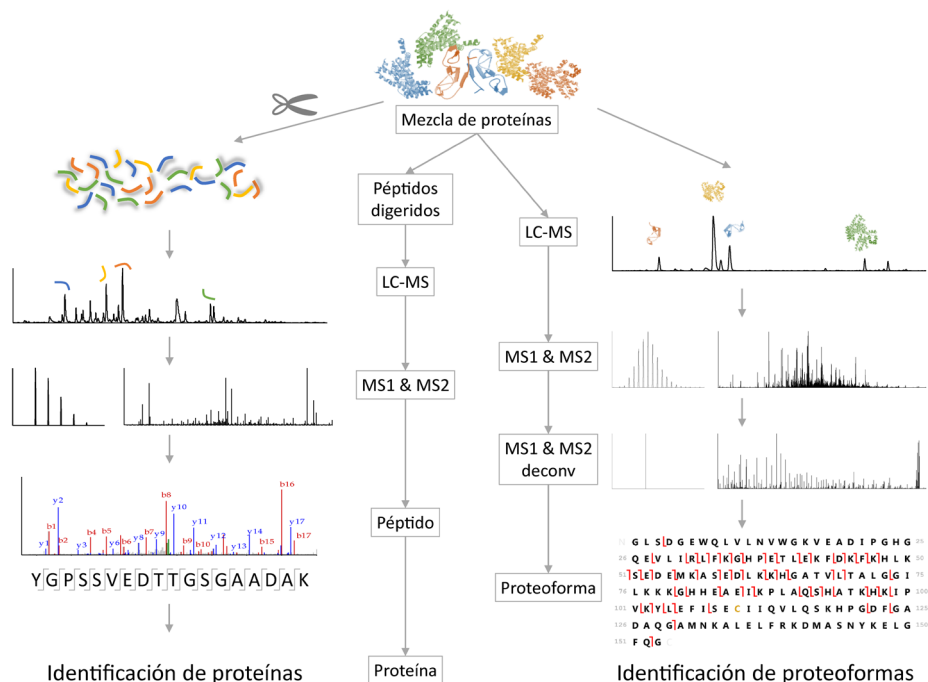


Figura 2. Representación esquemática de las aproximaciones top-down (derecha) vs bottom-up (izquierda).

A lo largo de este capítulo se describen diferentes estrategias de análisis top-down.

Por un lado, el análisis top-down en condiciones desnaturalizantes, cuyo objetivo se centra en la identificación y, a ser posible, caracterización de la secuencia proteica completa. Como se verá, esta estrategia, ha evolucionado desde el análisis de una sola proteína (top-down MS) al análisis high-throughput (top-down proteomics) de mezclas proteicas complejas de hasta 50-80 kDa. El reto analítico no alcanzado todavía mediante esta metodología reside en la posibilidad de analizar el proteoma entero y situarse al mismo nivel que la proteómica bottom-up. Se explicarán cuáles son las limitaciones de la técnica en este sentido, el estado actual y las nuevas tendencias.

Por otro lado, se describe la emergente estrategia native-top-down, que usa de manera sinérgica las habilidades de la establecida espectrometría de masas nativa-capaz de trabajar en condiciones de preparación de muestra e instrumentales que preservan las débiles interacciones no covalentes, y por tanto, la estructura y los complejos de proteínas-, con el top-down, para obtener información sobre la composición, estequiometría, empaquetamiento de subunidades de los complejos, y además, identificar y caracterizar la secuencia, modificaciones o variantes de cada subunidad en el complejo.

Es de destacar también la existencia de la denominada estrategia middle-down, que muestra semejanzas entre la estrategia top-down y bottom-up [5]. Esta técnica surge para facilitar el análisis top-down, sobretudo en proteínas de gran tamaño molecular, como los anticuerpos monoclonales (mAbs) [6, 7]. Las proteínas son digeridas a péptidos largos o de tamaño intermedio previamente al análisis MS. Normalmente se usan



proteasas específicas, que selectivamente rompen entre dos aminoácidos básicos, como el caso de la OmpT proteasa [8], o por encima o por debajo de la región bisagra, en el caso de los mAbs, como las proteasas GingiskHAN o IdeS respectivamente [9, 10]. Los reactivos químicos reductores de puentes disulfuro se usan frecuentemente en el caso de los anticuerpos (para simplificar el análisis MS a las cadenas pesadas y ligeras), y también otros reactivos, como el bromuro de cianógeno, que corta específicamente entre residuos de metionina generando péptidos largos con el residuo de homoserina lactona en N-terminal [11]. Entre las proteasas más usadas para proporcionar péptidos más largos que los que se obtienen con la tradicional tripsina se encuentran la Asp-N, Lys-C y Glu-C. Los fragmentos proteicos resultantes se analizan posteriormente por estrategias top-down para su identificación y caracterización [5, 12].

2. Fundamentos de la proteómica top-down.

La espectrometría de masas top-down analiza las proteínas, y más concretamente, las proteoformas, introducidas enteras, en el espectrómetro de masas. Las proteínas se ionizan generalmente a presión atmosférica por la técnica de electrospray (ESI) [13-16], formándose en este proceso un spray de gotas con iones moleculares múltiplemente cargados que son desolvatados y finalmente eyectados a la fase gas a presión reducida. Durante la ionización no se produce fragmentación de la molécula. El proceso de ionización puede realizarse también por desorción iónica laser asistida por una matriz (MALDI) [17, 18], aunque, en ese caso, el experimento MS no se realiza habitualmente acoplado en línea a métodos cromatográficos.

Una vez la proteína o mezcla de proteínas son ionizadas, el espectrómetro de masas, usando un sistema complejo de ópticas iónicas, filtros de masas y analizadores, mide la relación m/z de los iones múltiplemente cargados de la proteína/s. Algoritmos de deconvolución de estos datos permiten determinar la masa de la proteína a carga cero. Además, en el espectrómetro, se seleccionan los iones de carga múltiple (iones precursores) y se fragmentan mediante distintas técnicas como el CID (disociación por colisión de iones), ETD (disociación por transferencia electrónica), HCD (disociación por alta colisión), UPVD (fotodisociación ultravioleta), etc. La relación m/z de los iones fragmento proporciona información de secuencia de la proteína. Este tipo de experimento se denomina espectrometría de masas en tándem o MS/MS o MS²¹. Mediante el uso de motores de búsqueda y software especializados (ver apartado de análisis de datos), las masas de los fragmentos de las proteínas obtenidos en fase gas, juntamente con la masa total de la proteína, obtenida en el ciclo MS1, identifican y caracterizan la secuencia de la proteoforma, aunque, en muchas ocasiones, todavía se requiere también un análisis final manual para su validación (Figura 3).

La detección e identificación de proteínas intactas, especialmente a nivel de proteoma, depende del alto rendimiento de los espectrómetros de masas. La alta resolución y

¹El término tándem MS actualmente engloba la primera etapa de espectrometría de masas (MS1) y los distintos ciclos de aislamiento de iones precursores m/z y fragmentación consecutivos (MS2 hasta MSn), tanto si estos ciclos se producen en el espacio (en analizadores de sector magnético, cuadrupolo o tiempo de vuelo (TOF)), como en el tiempo (en analizadores ion trap, de resonancia ciclotrónica (ICR) y orbitrap).

exactitud en la masa son críticas para separar y asignar de manera precisa picos en el espectro provenientes de espectros de precursores complejos, conteniendo múltiples proteoformas intactas o espectros de fragmentación conteniendo centenares de iones. Se requiere alta resolución para distinguir puentes disulfuro ($\Delta m = 2$ Da), deamidación ($\Delta m = 1$ Da), trimetilación vs acetilación ($\Delta m = 39$ mDa) y fosforilación vs sulfatación ($\Delta m = 10$ mDa) [19, 20] (ver apartado instrumentación y técnicas de fragmentación).

Un valor clave de la proteómica top-down es que teóricamente puede llegar a la elucidación completa de la secuencia primaria de aminoácidos de las proteínas, incluyendo sus PTMs, si se consigue una fragmentación eficiente mediante las distintas técnicas de disociación actualmente al alcance. Combinando diferentes técnicas de fragmentación, por ejemplo, se ha llegado a tener disociación completa de la secuencia en proteoformas de entre 30-80 kDa ([21] [22]) y hasta de 150kDa, como en el caso de anticuerpos, disponiendo de instrumentación modificada ([23]). No obstante, y a pesar de los avances instrumentales, por lo general, a medida que la masa de las proteínas a analizar aumenta (>30 kDa), y/o su proporción en la muestra es baja, la eficiencia de la fragmentación en fase gas disminuye y no se consigue una fragmentación completa de la molécula trabajando a tiempos cromatográficos.

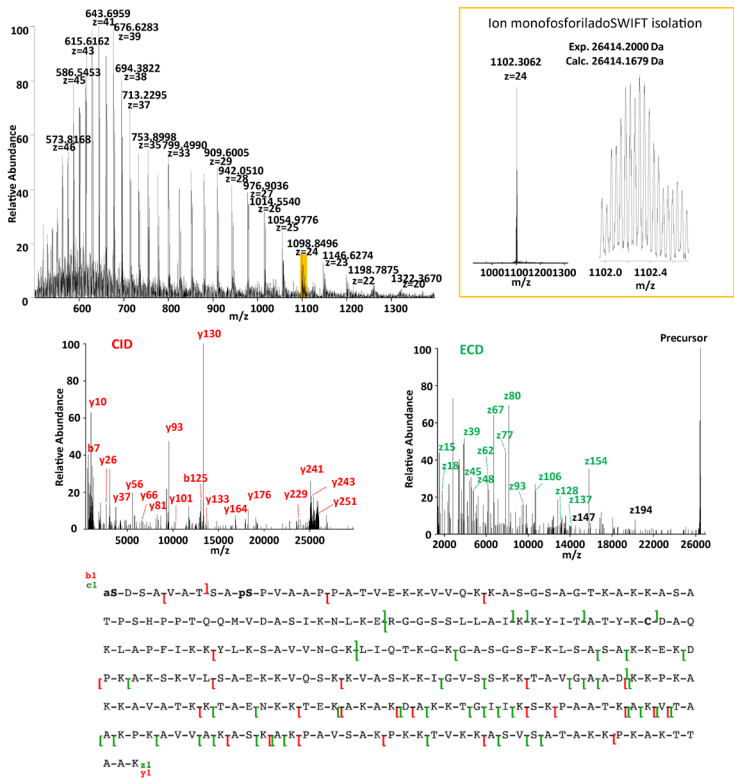


Figura 3. Espectros MS1 y MS2 de la Histona 1 de *Drosophila melanogaster* (dH1). Mapeo de PTMs. a) Espectro MS1 de dH1 adquirido en un LTQ-FT Ultra de 12 T, b) espectro del aislamiento del ion m/z 1102 correspondiente a la carga z=24, correspondiente a la proteoforma monofosforilada de dH1 con acetilación en el N-terminal, oxidación de M52 a metionina sulfona y de C81 a ácido cisteico resultantes de un tratamiento MPA (*mild performic acid*), c) fragmentaciones CID (izquierda) y ECD (derecha) del ion con carga z=24. Se indican los fragmentos b- e y- (CID) y c- y z- más relevantes, d) mapa de fragmentación mostrando la fosforilación en S11 y acetilación N-terminal en negra. La oxidación en C81 se indica también en negra. Los paréntesis rojos indican la cobertura de iones b- e y- N- y C-terminales. Los paréntesis verdes indican la cobertura de iones c- y z- N- y C-terminales [3]



3. Breve Historia del Top-down.

El descubrimiento que catalizó la era moderna de la proteómica fue el descubrimiento del electrospray por Fenn et al. [14] en 1989. En su revista fundacional presentó los primeros espectros de masas por electrospray y baja resolución de tres proteínas enteras: el citocromo, la mioglobina y la albumina. La última figura del artículo mostró también, en coordinación con los laboratorios de McLaferty y Hunts, los espectros de alta resolución de mioglobina y citocromo c adquiridos en un espectrómetro FTICR (Transformada de Fourier-resonancia ciclotrónica iónica) [14, 24].

Aunque el top-down está asociado a instrumentos FT-MS (Transformada de Fourier acoplado a espectrometría de masas), trabajos iniciales mostraron fragmentación top-down de ribonucleasa A (14 kDa) en instrumentos de triple cuadrupolo [25]. También en triples cuadrupolos y usando electrospray, Feng y Konishi demostraron la misma aproximación en mioglobina y un anticuerpo monoclonal de 150 kDa ([26],[27] [28]). Se sucedieron diversos artículos de análisis de proteína intacta a alta resolución vía FTMS-MS [29] y se extendió la demostración de la técnica top-down sobre proteínas puras o semi-puras en una variedad de instrumentos MS, usando MALDI-TOF/TOF, ESI-QTOF, ESI-ion trap, ESI-cuadrupolo-FTMS, etc [27, 30, 31] .

A mediados de los 90, se extendió el uso de la técnica MALDI-TOF con ISD (ion-source decay) al ver que se podían generar amplias series de iones tipo $-c$ y $-y$ a alto potencial de láser, disponiendo de cantidades grandes de muestra (mg) ([32], [30]). Estudios en el laboratorio de McLaferty a mitad de los noventa, también ayudaron a formalizar el top-down y demostraron la utilidad de la FTMS para polipéptidos largos [33]. En un estudio dirigido de la *thiaminasa I* (42 kDa), Kelleher hizo uso de la alta resolución y la fragmentación nozzle-skimmer para asignar heterogeneidades no esperadas en la secuencia y para localizar el sitio activo del enzima [34]. El ESI-FTMS fue proyectado como la opción a escoger para la ionización y fragmentación directa de proteínas intactas en mezclas.

En los métodos top-down la determinación de la carga de los fragmentos es crítica para obtener masas fiables. Por ello, el top-down quedó durante años reservado a los instrumentos existentes de alta resolución, es decir, los instrumentos con grandes imanes [35], considerándose una técnica muy exclusiva y de la que pocos laboratorios podían disponer. Avances, como la aparición de instrumentos Q-FTMS híbridos y nuevas técnicas de disociación como el ECD (electron capture dissociation), desarrollada por Roman Zubarev mientras trabajaba en el laboratorio de McLafferty, mejoraron el rango dinámico de MS/MS para iones de alta m/z [36, 37], aunque seguían fundamentándose en el uso de instrumentación ICR. Posteriormente, la aparición del analizador Orbitrap, desarrollado por Zubarev y Makarov en 2013 [38] revolucionó la proteómica en general, y ha contribuido en gran medida a la expansión de la técnica top-down pudiéndose aplicar en instrumentación más asequible en los laboratorios actuales.

No obstante, desde el desarrollo del analizador orbitrap y analizadores TOF con resoluciones > 50.000 , el top-down ha continuado siendo, hasta hace relativamente poco

tiempo, un reducto no explorado por la gran mayoría de los especialistas en proteómica. Por ello, mientras se producía la explosión proteómica y la técnica bottom-up aumentaba exponencialmente su potencial de análisis cualitativo y cuantitativo sobre el proteoma entero, el top-down jugaba un papel importante exclusivamente en la determinación de proteoformas únicas, aunque, generalmente, sobre proteínas purificadas. El motivo por el cual el top-down, a pesar de las ventajas que ofrece a nivel de la caracterización de las proteínas, sigue no siendo de uso general por la comunidad proteómica, es por la dificultad de la técnica y la no existencia de métodos generales útiles para el análisis de todas las proteínas. Existen *reviews* que explican las limitaciones de top-down y los retos que plantea a la instrumentación para el análisis de proteínas de alto peso molecular. [39-42]. Las limitaciones básicas se centran en cuatro puntos principales: a) el prefraccionamiento o “*front end*”, sobre todo cuando el objetivo es llevar a cabo un análisis *high-throughput*, b) la instrumentación MS, que requiere afrontar el reto que supone la detección y aislamiento de especies de alto peso molecular para su fragmentación, teniendo en cuenta que estas conllevan cargas muy elevadas derivadas de la ionización por ESI y que por tanto, proteoformas de peso muy similar son difícilmente separables [39], c) la obtención de espectros MS2 con la suficiente información de secuencia a escala cromatográfica y c) el análisis de datos que no está todavía totalmente automatizado cuando se trata de analizar muestras complejas.

4. Estrategias de prefraccionamiento de mezclas complejas de proteínas.

La base de un experimento top-down MS reside en la generación y medida de iones de proteínas en fase gas, su activación para efectuar la fragmentación y análisis de los fragmentos generados. El proteoma es extraordinariamente complejo; el número de proteínas y la diferente combinatoria de proteoformas derivadas que pueden existir en una célula, tejido u organismo en un momento determinado complica enormemente el análisis top-down a gran escala. El rango dinámico del proteoma celular es, además, muy amplio (aproximándose a los siete órdenes de magnitud) y en cambio, el de los espectrómetros de masas, limitado (de entre 4 y 5 órdenes de magnitud), de manera que sólo las proteínas más abundantes son identificadas si no se reduce la complejidad de la muestra antes de la medida en el espectrómetro [43]. Las técnicas de fraccionamiento de proteínas son, por tanto, esenciales para reducir la complejidad de la muestra, mejorar el rango dinámico para su detección y aumentar la especificidad del análisis.

En proteómica top-down se han explorado y siguen explorando diferentes estrategias para prefraccionar proteínas, así como combinaciones multidimensionales para optimizar la cobertura del proteoma. Las estrategias de separación se realizaban en los inicios del top-down de manera off-line, independientemente del espectrómetro de masas, colectando las fracciones eluidas e infusionando después en el espectrómetro para disponer de más tiempo de colección de datos. Este es el caso habitual del estudio centrado en proteínas puras concretas o mezclas sencillas. La evolución hacia la proteómica top-down, fuerza necesariamente la tendencia hacia la separación on-line,



para tener un mayor rendimiento y menor manipulación de la muestra. De todos modos, teniendo en cuenta la gran complejidad de las muestras, se requiere, la mayoría de veces, una separación previa off-line seguida de una separación con acoplamiento on-line al espectrómetro.

La integración de la separación a nivel de proteínas y el acoplamiento on-line es complicada a pesar de los distintos avances tecnológicos. En primer lugar, la solubilidad de las proteínas es menos predecible que la de los péptidos y, en ocasiones, tienen tendencia a pegarse en el material cromatográfico. Por otro lado, hasta la aparición de instrumentación MS lo suficientemente rápida, a la vez que de alta resolución, el tiempo requerido para generar suficientes iones fragmento para tener una caracterización completa de la secuencia era también limitante. Actualmente, las distintas técnicas de fragmentación CID, ETD, ETDhCD, UPVD, etc. que se trataran en las próximas secciones, permiten tener información de secuencia (no necesariamente completa) en tiempos cromatográficos reales y en instrumentación comercial y no modificada para proteínas de hasta 50 kDa con optimización previa de los métodos.

4.1. Cromatografía Líquida

La cromatografía líquida es uno de los métodos más comunes para el prefraccionamiento de mezclas complejas de proteínas intactas. El principio de separación se basa en la partición diferencial de los analitos entre la fase móvil y la fase estacionaria. En general, usando disolventes cromatográficos compatibles, la cromatografía líquida, en sus diferentes escalas de flujo (micro, nano, etc..) se acopla eficientemente a la ionización por electrospray (ESI) [44]. Aunque hay una gran variedad de métodos de cromatografía líquida (LC), la cromatografía en fase reversa (RP-LC), la de interacción hidrofóbica (HILIC) y la de intercambio iónico son las tres más comúnmente usadas para la separación de proteínas intactas [45].

La cromatografía de fase reversa (RPLC) usa fases móviles polares y fases estacionarias no polares, lo que permite a los analitos más hidrofílicos eluir primero. Las fases estacionarias más comunes son las cadenas alquilo (C4, C5, C8 y C18) unidas a partículas de sílice. Las cadenas alquilo cortas son las preferidas para las separaciones de proteína intacta de peso molecular mayor, dado que son menos retentivas y ofrecen una mayor recuperación [45]. De manera similar, las fases reversas poliméricas (PLRP) ofrecen también buenas recuperaciones, hidrofobicidad uniforme y mayor fuerza mecánica para la separación de proteínas intactas [21, 44].

A diferencia de la RPLC, la interacción HILIC usa fases estacionarias polares y gradientes crecientes del contenido de agua, lo cual resulta en la elución de las especies más hidrofóbicas primero [46, 47]. Se han descrito, por ejemplo, separaciones de histonas con múltiples PTMs usando HILIC previamente al análisis top-down [48, 49]. Existen incluso ejemplos donde, usando cromatografía HILIC off-line, se han separado isómeros posicionales de histonas humanas H4 y posteriormente cuantificado a partir de las abundancias relativas de sus fragmentos [50].

La cromatografía HILIC ha demostrado ser muy eficiente para la separación con alta resolución de proteínas intactas, aunque en muchos casos, las sales usadas con-

vencionalmente son incompatibles con la MS. En este sentido, trabajos incorporando tartrato amónico como una sal compatible con MS, han demostrado una gran selectividad y ortogonalidad frente a la cromatografía de fase reversa, siendo su rendimiento comparable a la separación que ofrece el convencional sulfato amónico [51].

Mientras la RPLC y HILIC se basan en la separación de las proteínas por hidrofobicidad, la cromatografía de intercambio iónico usa diferencias en la carga del analito. Incrementando la fuerza iónica de la fase móvil se eluye el analito de la fase estacionaria cargada. Tradicionalmente, la cromatografía de intercambio catiónico ha sido usada extensamente, y tiene una alta capacidad de carga.

Sin embargo, la resolución de las proteínas por cromatografía líquida es limitada usando un único tipo de separación cromatográfica. Más concretamente, cuando se trabaja con proteínas altamente modificadas (con muchas y distintas PTMs), uno de los mayores retos es resolver cromatográficamente las distintas proteoformas de una proteína (que potencialmente pueden comprender diferentes combinaciones de la misma modificación o modificaciones de peso molecular similares sobre la misma molécula).

Considerando los efectos de las diferentes PTMs en la estructura y las propiedades fisicoquímicas de una proteína, el intercambio iónico y la hidrofiliidad/hidrofobicidad son los principales factores a tener en cuenta en el desarrollo de técnicas de fraccionamiento de isoformas/proteoformas de proteínas. Por ello, las combinaciones de diferentes tipos de cromatografía en serie son las que ofrecen mayor cobertura de las proteínas. Existen muchos ejemplos de aproximaciones bidimensionales basadas exclusivamente en técnicas cromatográficas. La cromatografía de exclusión molecular (SEC) y de intercambio iónico se usan para llevar a cabo la primera dimensión [52]. Trabajos destacables son los realizados por García et al. [48] en 2007 y Pesavento et al. [53] en 2008 donde se separan y cuantifican proteoformas de histonas, H3.2 i H4, combinando intercambio aniónico débil (WAX) y cromatografía HILIC. En 2010 el grupo de Ljiljana Paša-Tolić presentó la primera aproximación top-down online 2D usando HILIC y RPLC para la caracterización de mezclas de núcleos de histonas [54]. Kellie et al., también describieron en 2010 el uso de una combinación de cromatografía de intercambio iónico off-line y de fase reversa identificando por la estrategia top-down 133 proteínas de células blancas de la sangre [55]. Similarmente, otra aproximación top-down usó una cromatografía líquida tridimensional de intercambio iónico, HILIC (en tartrato amónico) y RPLC separando e identificando 640 proteínas de 293 fracciones en un lisado de células HEK [56].

El *cromatofocusing* y el enfoque isoelectrico capilar o en solución, son consideradas métodos “front-end” alternativos al gradiente de sal de la cromatografía de intercambio iónico para aplicaciones top-down [55, 57]. Ambos métodos son capaces de separar proteínas en función de su punto isoelectrico usando gradientes de pH para la elución. El *cromatofocusing* se ha usado también acoplado off-line a RPLC para fraccionar proteínas <60kDa de *Methanosarcina acetivorans* [58] y proteínas integrales de membrana [59]. El enfoque isoelectrico, aunque efectivo para el fraccionamiento, es hasta el momento una técnica con una deficiencia significativa respecto a la pobre recuperación de proteínas [60, 61]



4.2. Electroforesis

La electroforesis capilar es una aproximación muy popular de separación de proteínas intactas basada en la migración diferencial de proteínas en un campo eléctrico [44, 62]. El método electroforético más común es el SDS PAGE, en el que las proteínas envueltas en SDS migran en una matriz de gel de poliacrilamida en un campo eléctrico, consiguiéndose una separación de proteínas por tamaño molecular [63]. Este es un método muy usado en la estrategia proteómica bottom-up. Una vez las proteínas separadas en el gel, se recortan las bandas correspondientes, se digieren directamente en el gel y se extraen los péptidos y analizan por cromatografía líquida de fase reversa acoplada a MS. Esta aproximación se puede extender a la electroforesis en gel de poliacrilamida bidimensional (2D-PAGE) en la que el isoelectrofocusing es usado en la primera dimensión de la separación y la SDS-PAGE es la segunda dimensión de separación de las proteínas intactas [64].

2D-PAGE es la técnica más conocida para la separación de proteínas intactas debido a su capacidad de pico. Sin embargo, la extracción de proteínas intactas de los geles resulta muy ineficiente y se deben buscar técnicas alternativas de separación en solución para llevar a cabo una estrategia top-down MS.

Mientras la separación convencional basada en geles no es aplicable por top-down proteomics, se han desarrollado distintas estrategias tales como la electroforesis en gel de tubo (*tube gel electrophoresis*). La electroforesis en gel de elución continua utiliza una columna de gel de tubo que puede separar proteínas intactas que son colectadas a medida que van eluyendo del final de la columna. Meng et al. demostraron la eficacia de esta aproximación para el fraccionamiento del proteoma de levadura usando un surfactante lábil al ácido como sustituto del SDS (dodecil sulfato de sodio), el cual se degradaba, vía acidificación, evitando interferencias, seguido de RPLC en columnas C4 [65].

La electroforesis en gel de tubo evolucionó a la invención del GELFrEE (*gel-eluted liquid fraction entrapment electrophoresis*). Este método de separación, descrito en 2008 por Tran & Doucette [66], facilitó enormemente la separación multidimensional “front-end” para análisis top-down de proteomas complejos, siendo una estrategia robusta de fraccionamiento por peso molecular de proteínas en fase líquida. El sistema consta de una cámara separada para cargar y coleccionar la muestra y de un cartucho de separación, que se introduce en un instrumento de sobremesa GELFrEE, desde donde se automatizan los métodos de separación. Se aplica un voltaje constante entre los reservorios del ánodo y del cátodo durante la separación de las proteínas. Cada mezcla de proteínas es conducida electroforéticamente desde la cámara de carga a una columna de gel. Las proteínas se concentran en una apretada banda en el *stacking gel* y son separadas, basadas en sus respectivas movilidades electroforéticas en el gel resolutivo. A medida que las proteínas eluyen de la columna son atrapadas y concentradas, en fase líquida, en la cámara de colección. Esta tecnología permite la separación de proteínas de un amplio rango de masas (5-100kDa) en base a su peso molecular. El atrapamiento en fase líquida elimina la necesidad de elución o recorte de la banda y convierte al proceso en altamente reproducible. El sistema GELFrEE fue aplicado por primera vez en un estudio top-down por Lee et al. en 2009 y la eliminación del SDS utilizado durante la separación se llevó a cabo por precipitación con

mezclas de metanol/cloroformo antes del análisis nanoLC [67, 68]. Múltiples estudios han combinado el GELFrEE con otras técnicas de fraccionamiento para desarrollar flujos de trabajo para el análisis de proteoformas intactas a gran escala [69-73]. Una limitación importante del GELFrEE es precisamente el uso de SDS, un detergente que interfiere enormemente con la ionización por electrospray y que debe ser, por tanto, eliminado en su totalidad de las muestras. El proceso de “*clean-up*” se realiza por precipitación con metanol/cloroformo o con columnas de spin, aunque ninguno de los dos procedimientos esta automatizado ni es necesariamente reproducible.

Una innovación reciente ha solucionado este problema acoplando fraccionamiento de flujo de campo asimétrico directamente entre el GELFrEE y la MS, en una matriz para eliminar en-línea el SDS en sólo 5 min [74]. A pesar de su eficiencia, su uso general, en cambio, no ha sido todavía implementado.

Similamente al GELFrEE, nuevas aproximaciones de separación basadas en el enfoque isoelectrico o atrapamiento por pH son adecuadas para el top-down. Concretamente, los sistemas MSWIFT (*membrane separated wells for isoelectric focusing and trapping*) [75, 76] y los sistemas MJ-CIEF (*multi-junction capillary isoelectric focusing*) de multiensamblaje o atrapamiento por pH (pI-TRAP) [77]. Estos últimos, comercializados en sistemas denominados Biomotif, desarrollados por R. Zubarev, integran el cargado de muestra, el desalado on-line, el fraccionamiento por pH (IEF) y la colección de fracciones y permiten separar tanto péptidos como proteínas. Se han usado recientemente de manera eficiente para realizar separaciones isoelectricas directamente de proteínas de plasma humano para el descubrimiento de biomarcadores [77]. Este fraccionamiento se usa previamente a la cromatografía RPLC-MS y aumenta la cobertura del análisis proteómico. En concreto puede ayudar a la identificación de proteínas que cambian su pI en las enfermedades (e.g, incorporando ciertas PTMs que hacen variar su pI) y que no son normalmente detectadas con otro tipo de prefraccionamiento (Figura 4).

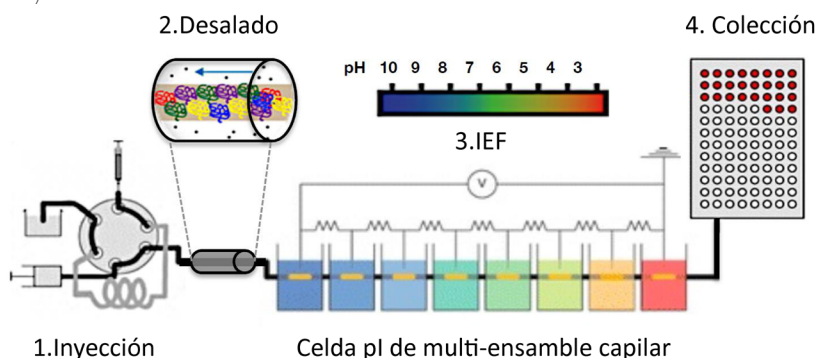


Figura 4. Esquema del sistema Biomotif multi-junction capillary pI-CELL o de multiensamblaje o atrapamiento por pH (pI-TRAP).

4.3. Movilidad iónica-MS

La movilidad iónica acoplada al masas (IM-MS) no se puede considerar estrictamente una técnica de separación “*front-end*” puesto que no se realiza previamente al fenómeno de ionización, sino una vez creados los iones en fase gas. De todos modos, se incluye en la sección de prefraccionamiento puesto que es una técnica muy útil en los casos en los que se pretende la separación de proteoformas, algunas isobáricas (o de



igual m/z) pero con PTMs en posiciones distintas y que son susceptibles de presentar, por tanto, conformaciones distintas, siendo estas difícilmente separables por cromatografía líquida [78] [79]. La técnica de movilidad iónica está basada en el tiempo que los iones tardan en atravesar una cámara llena de gas y bajo la aceleración de un campo eléctrico. La movilidad de los iones depende de su área de colisión cruzada (CCS), que es definida como el área del ion proyectada en un plano y que efectivamente colisiona con las moléculas del gas. Por IM-MS se consigue simultáneamente, no sólo separar iones en función de su relación masa/carga, sino también de su forma y tamaño. Dependiendo de la arquitectura del espectrómetro que se utilice, la separación por movilidad iónica puede aplicarse tanto a los iones precursores, como a los iones fragmento, después que las proteoformas precursoras hayan sido fragmentadas por CID o ETD [80]. En este último caso, se consigue mejorar otra de las grandes limitaciones del top-down, el rango dinámico de la abundancia de los iones fragmento, facilitando el aumento de la cobertura de la secuencia proteica.

Hemos visto que la maximización de la separación, en el *front end*, mediante diferentes técnicas de prefraccionamiento, facilita la posterior identificación, caracterización y asignación correcta de las múltiples proteoformas existentes. De todos modos, la resolución completa de muchas proteoformas es difícilmente alcanzable antes de la espectrometría de masas y, como se verá en los siguientes apartados, es la nueva instrumentación MS la que tiende a salvar este déficit. Concretamente, se está evolucionando en este sentido para poder resolver iones altamente cargados muy próximos en el espacio m/z y, a la vez, fragmentarlos de manera eficiente en tiempos cromatográficos.

5. Instrumentación y técnicas de fragmentación.

5.1. General

La caracterización, tanto de proteínas, como de péptidos, por espectrometría de masas en tándem requiere básicamente tres pasos principales: a) la ionización, que permite el paso del polipéptido de líquido a fase gas en forma de iones cargados que van a ser manipulados en el espectrómetro, b) la determinación de la masa del polipéptido intacto o MS1 (masa intacta) i c) la determinación de la secuencia peptídica a través de la adquisición de espectros MS2 resultantes de la activación y fragmentación de los iones precursores seleccionados. En el caso del análisis de proteínas, las tres operaciones incrementan en dificultad a medida que aumenta su peso molecular.

Aunque existen ejemplos de ionización por MALDI, en general, la ionización por ESI es la escogida en la proteómica top-down. Del mismo modo que en la proteómica bottom-up, la ionización en modo positivo es la más usada y la polaridad de los iones a analizar se escoge seleccionando el sesgo (bias) del voltaje del capilar. Una característica de la técnica de ionización por ESI es la formación de iones moleculares intactos del analito. En el caso de desear fragmentación durante la ionización, esta puede inducirse en la región de transporte de los iones de la fuente ESI, incrementando el

voltaje del *sampling cone*. Este proceso es conocido como *in-source collision induced dissociation* or *nozzle-skimmer (NS) dissociation*.

La ionización por ESI de proteínas y, de especies poliméricas en general, proporciona espectros que contienen series en forma de campana de Gauss de iones múltiplemente cargados para una misma masa M de proteína $(M+nH)^{n+}$. En condiciones desnaturalizantes, las proteínas tienen mucha polidispersidad de carga y esto se traduce en campanas de Gauss muy amplias comparado con los péptidos. El espectro de la Histona H1 de *Drosophila melanogaster* en la Figura 3 ilustra un ejemplo de espectro ESI típico. Cada pico en el espectro difiere del contiguo en una carga. Dado que en el espectrómetro de masas analizamos iones según su relación m/z y no m , el fenómeno de ionización por ESI con carga múltiple hace posible el análisis de compuestos de alto peso molecular con espectrómetros de rango de masa limitado.

A diferencia del análisis bottom-up, para el análisis top-down se requiere instrumentación de alta resolución, y más especialmente, para el análisis de los iones fragmentos creados en el experimento MS/MS. En el procesamiento de los datos MS1 de proteínas intactas, la resolución isotópica de los clústeres de iones creados es útil, aunque no estrictamente necesaria para determinar la carga de los iones y en último término, su masa MS1. En cambio, los espectros de fragmentación de proteoformas intactas son complejos, conteniendo una variedad de iones producto de diferentes estados de carga que se pueden solapar. Por tanto, un poder de resolución > 50.000 es mínimamente necesario para deconvolucionar correctamente a carga cero e interpretar el espectro MS2. Los niveles de resolución requeridos para hacer top-down se alcanzan actualmente con los analizadores TOF modernos (50.000 k) y los analizadores FTMS (Orbitrap y ICR-FTMS) que van desde los 100.000 K hasta $1.000.000$ K o incluso más en grandes imanes.

Existen tres métodos básicos para la adquisición de datos top-down, que se basan precisamente en conseguir una alta resolución en los iones fragmento. Son los llamados métodos high-high, medium-high y low-high, diferenciándose estos en la resolución usada en el análisis MS1. En el apartado de análisis de datos se explica cómo se tratan los datos adquiridos dependiendo de los métodos usados.

Otro de los parámetros a tener en cuenta en el análisis MS de proteína intacta es la sensibilidad, que marcará en último término la relación señal/ruido (S/N). Dado que de una proteína entera se generan por ESI una amplia diversidad de iones, correspondiente a las distintas especies cargadas de la misma, esto significa que se produce una división de la señal a distintas señales con intensidades inversamente proporcionales al número de las mismas [39]. Si, además, la proteína está en una baja proporción en la muestra, la cantidad de iones es muy limitada. Si hablamos de proteoformas, la dispersión de señales es múltiple y en posiciones m/z que difieren sólo en milésimas o centésimas de unidad de masa. Los analizadores TOF, que operan bajo el principio de recuento de iones individuales, solucionan el problema de la S/N acumulando más iones que en las aplicaciones bottom-up. En los instrumentos FTMS, que operan con una cantidad fija de iones o cargas para evitar efectos de carga y pérdidas resolutivas, se promedian múltiples scans tanto para el MS1 como para el MS2. En las aproximaciones top-down, ambas soluciones enlentecen los ciclos experimentales en relación al bottom-up.



Por otro lado, se debe asegurar en top-down la selección precisa de iones precursores que difieren en pocas unidades (centésimas o milésimas teniendo en cuenta su carga). Por ejemplo, Zabrouskov et al., propusieron la deamidación de asparaginas o glutaminas en RNase A ($\Delta m = +1$ Da) como un reloj biológico [81]. Este nivel de resolución y exactitud en la masa fuerza el uso de instrumentación de alta resolución. En espectrometría de masa top-down es útil la combinación de analizadores de alta resolución en arquitecturas híbridas. La presencia de múltiples analizadores permite mucha versatilidad en la manipulación de los iones y además ofrece una mayor posibilidad de selección de las técnicas de fragmentación combinables para poder aplicar en la resolución de cuestiones biológicas específicas. En los siguientes apartados se resumen la instrumentación y las diferentes técnicas de fragmentación top-down, su evolución en los últimos años y sus aplicaciones más destacables.

5.2. Instrumentación por Transformada de Fourier-Resonancia de iones ciclotrónica

El nivel resolutivo y de exactitud de masa requerido para llevar a cabo estudios top-down hizo que originalmente los instrumentos más usados y populares fueran los analizadores FTICR (*Fourier transform ion cyclotron resonance*). Estos analizadores se basan en la excitación de iones a su frecuencia ciclotrónica dentro de un campo magnético fuerte [82]. Los iones son atrapados en la celda de resonancia por un campo magnético coaxial y voltajes aplicados. Estos rotan a una frecuencia determinada por su masa y carga y la fuerza del campo. Su radio de órbita aumenta hasta que pasan al plato de detección y provocan que electrones fluyan a la placa de detección monitorizando así su corriente imagen. La señal detectada (también llamada *transient*) se convierte desde el dominio tiempo al dominio de frecuencia a través del algoritmo de transformada de Fourier y a m/z a través de la calibración de la masa.

El ESI-FTICR-MS se usó para análisis de proteína intacta por primera vez en 1989 [24] y se fueron sucediendo en instrumentos de 2.8 y hasta 12 Teslas [83, 84] con modificaciones que permitían mayor sensibilidad, rango dinámico y resolución.

En 2004 se describió un nuevo instrumento híbrido ion-trap-FTICR-MS [85] de manera que permitía el almacenaje de iones y manipulación en el ion trap, desde una fuente de iones continua, hasta su inyección en paquetes de iones al ICR. La exactitud en la masa mejoró a través del control preciso del número de iones o cargas (uso del *automatic gain control* (AGC)) permitidas a entrar en la celda ICR, incluso provenientes de flujos variables de iones, como es el caso de la LC-MS. En este tipo de instrumentos, la fragmentación se produce en el ion trap por CID y los fragmentos pueden ser detectados a alta resolución en la ICR o en el ion trap. Este instrumento se comercializó en la versión de 7 T (teslas) y llega a una resolución de 1000.000 (resolución definida a m/z 400) y exactitudes de < 2 ppm sin calibración interna. El uso del 7 T LTQ-FTICR, se ha sido descrito para diversas aplicaciones proteómicas top-down [86-88], así como la de su versión con 12 T [89, 90] y 14.5 T [91], todos ellos, todavía en uso. Los instrumentos FTICR, suelen tener incorporados también celdas ECD (*electron capture dissociation*, ver apartado siguiente), que permiten la fragmentación de iones por métodos alternativos y su detección en el ICR. Aunque muy resolutiva, este tipo de instrumentación es lenta para los requerimientos de trabajos top-down de mues-

tras complejas trabajando a tiempos cromatográficos y está siendo sobrepasada por la nueva instrumentación Orbitrap, o nueva generación TOF en versiones híbridas, que actualmente combinan además nuevas técnicas de fragmentación más eficientes como el HCD, el ETD o el UVPD.

5.3. Instrumentación Orbitrap.

El analizador Orbitrap se ha convertido en una alternativa FTMS muy popular, tanto para aplicaciones top-down como bottom-up, dado que su coste es inferior al de los grandes imanes FT y se puede aplicar en diversas arquitecturas híbridas [92-94].

El analizador Orbitrap fue descrito en 2000 [95]. Es una trampa iónica que cuenta con un par de electrodos axialmente simétricos. En este campo eléctrico, los iones giran alrededor del electrodo central mientras oscilan a lo largo de la longitud de electrodo. La frecuencia de estas oscilaciones es proporcional a $(m/z)^{-1/2}$. La corriente imagen en los electrodos externos se monitoriza y la señal de dominio de tiempo resultante se convierte a frecuencia y luego a m/z como en la FTICR.

Como en el caso de los anteriormente citados instrumentos FTICR, los primeros Orbitraps también se hibridaron inicialmente con trampas LTQ, permitiendo el uso de iones continuos desde la fuente de ionización y aumentando la precisión de masa con la AGC, que permitía además una fragmentación eficiente [93]. El acoplamiento de los dos analizadores se logró mediante el uso de un octapolo de transferencia después de la trampa de iones en un cuadrupolo curvo solo de radiofrecuencia RF (trampa C) usado para expulsar iones axialmente hacia el analizador Orbitrap. Este nuevo instrumento fue capaz de obtener 60.000 poder de resolución (m/z 400) usando *transients* de un segundo y logrando resoluciones isotópicas de mioglobina y anhidrasa carbónica. Ya en 2006, se describió el uso de LTQ-Orbitrap para un análisis más extenso de proteínas intactas modelo, de masas inferiores a 30 kDa [96]

De manera destacable, aplicaciones en top-down del grupo de Kelleher en instrumentación Orbitrap incluían una profunda caracterización de la compleja matriz de modificaciones post-traduccionales en histonas [97-99] y en proteomas semicomplejos de células humanas [100]). También Fenselau analizó en Orbitrap biomarcadores proteicos de organismos bacterianos no secuenciados, colocándolos consecuentemente en su contexto filogenético [101, 102].

Otros ejemplos de uso de Orbitraps híbridos (LTQ-Orbitrap) son el estudio de proteínas de bajo peso molecular en sangre humana, incluyendo la cuantificación de proteoformas de apolipoproteína [103, 104] y trabajos de proteómica venómica de serpiente del grupo de J. Calvete [105].

Los instrumentos LTQ-Orbitrap comerciales no modificados se han usado para distinguir diversas glicoformas de anticuerpos intactos (150 kDa) y se han fragmentado sus cadenas ligeras y pesadas usando CID, ETD, o HCD [106, 107].

El LTQ-Orbitrap Velos ha sido usado en el análisis de enfermedades causadas por variantes de hemoglobina a partir de gotas de sangre seca, teniendo esta aplicación un



importante potencial de uso clínico [108]. Usando un instrumento Orbitrap modificado de alto-campo acoplado a un Orbitrap Velos Pro con trampa de iones dual y procesamiento de la señal avanzada, Michalski et al. describieron la identificación de 690 proteínas únicas por top-down en la línea celular de cáncer H1299 [70].

El instrumento híbrido Orbitrap Fusion Lumos, ha representado en los últimos años un gran avance para las aplicaciones top-down. Este instrumento combina tres analizadores: un cuadrupolo, una trampa de iones lineal segmentada y un Orbitrap, [109]. Esta configuración única, ofrece gran versatilidad, dado que permite fragmentar los iones por distintas técnicas: CID, ETD, HCD de manera separada o incluso combinando ETD con métodos de colisión HCD o CID, en lo que se denomina EThcD y ETcID, incrementando así la cobertura de la secuencia de proteínas. Un ejemplo es el estudio de Bruner et al [110], donde se consigue una mayor cobertura del regulador mitótico Bora, permitiendo la localización específica de los sitios de fosforilación en las distintas proteoformas. La precisión en la selección del precursor con el nuevo cuadrupolo segmentado ofrece la posibilidad de fragmentación exclusiva de proteoformas que están muy cerca en el espacio m/z . Además, este instrumento, presenta una tecnología avanzada, incluso respecto a su anterior versión (Orbitrap Fusion), ofreciendo la posibilidad de regulación del vacío óptimo en la región del *Ion Routing Multipole* dependiendo de la estrategia que se utiliza (e.g, bottom-up (8 mbar), top-down (2 mbar)). Con la modificación del sistema de vacío en esta región del instrumento, se ha conseguido minimizar la pérdida de iones prematura de proteína por *transient decay* al colisionar con moléculas de gas neutro (típico de las altas presiones) y por tanto, incrementado la relación S/N, a la vez que beneficiado de las altas presiones para una más eficiente fragmentación posterior, por ejemplo por HCD o ETDHcD. La mejor transferencia de iones de alta m/z en las distintas regiones del instrumento (C-trap, IRM, Orbitrap y ion-trap) juntamente con la multiplicidad de técnicas de fragmentación disponibles y la precisión en la selección del precursor, ha permitido recientemente, la caracterización de numerosas proteoformas de Histona H3, caracterizando hasta 8 metilaciones en una misma cadena polipeptídica [111]. La evolución de los instrumentos Orbitrap híbridos es vertiginosa. Justo antes de la publicación de este manuscrito, se ha presentado en la ASMS 2019 el equipo análogo de nueva generación. El nuevo instrumento Orbitrap Eclipse supera el rendimiento y la flexibilidad del equipo anterior. La incorporación de las opciones de: PTCR (reducción de carga por transferencia de protón), separación FAIMS (*High-Field Asymmetric Waveform Ion Mobility Spectrometry*) antes de la interficie, monitorización de la relación masa/carga hasta 8000 y medición a 1 millón de resolución entre otros, son realmente una gran promesa para que se efectúe la implementación efectiva de la metodología top-down.

5.4. Instrumentación TOF

La instrumentación de tiempo de vuelo (TOF-MS; *Time of flight mass spectrometry*) ha sido una opción atractiva durante muchos años para el análisis de proteína intacta debido a su rango m/z potencialmente ilimitado, adquisición de alta velocidad, capacidad de medición de masa exacta y sensibilidad. El analizador TOF fue comercializado a finales de 1950 por la Corporación Bendix [112], aunque los primeros instrumentos presentaban limitaciones tanto en el poder de resolución como de exactitud en la medida de la masa. Desde principios de los años setenta hasta principios de los noventa,

se superaron estas limitaciones. Inicialmente, la llegada del reflectron TOF-MS superó la propagación de la energía iónica y, por lo tanto, aumentó en resolución de masa. La combinación posterior con aceleración ortogonal (aa) se puede ver como el catalizador para la amplia gama de instrumentación TOF que está disponible hoy en día, con una potencia de resolución de masas y una precisión de masa muy mejoradas, llegando actualmente a los hasta 80.000 de poder resolutivo y precisiones de 1 ppm [113, 114]. Estos analizadores se combinan también en configuraciones de múltiples analizadores y configuraciones híbridas y llevan incorporados distintas técnicas de fragmentación (ver apartado siguiente).

Aunque el FTMS se ha convertido en dominante para el análisis top-down, existen estudios recientes que demuestran la eficiencia de la tecnología TOF para el análisis de proteína intacta y top-down. Como ejemplo, instrumentos híbridos cuadrupolo-TOF, combinados con métodos de fragmentación radicalaria (ETD) han demostrado ser eficientes para analizar proteínas de tamaño medio y largo [22], a la vez que útiles para resolver proteoformas biológicamente relevantes en *Neisseria meningitidis invasiveness* [115]. En el campo farmacéutico instrumentación TOF de última generación está siendo muy eficiente para el análisis de anticuerpos monoclonales intactos y reducidos, ADCs (antibody drug conjugates) [116] en sistemas con analizadores múltiples TOF [117] y híbridos QTOF [118].

5.5. Métodos de fragmentación

Tal y como se ha comentado, los espectrómetros de masas usados para la técnica top-down combinan la alta resolución de los analizadores con las arquitecturas híbridas (combinación de analizadores en serie que además incorporan distintas técnicas de fragmentación).

La activación y fragmentación de iones de proteína intacta se puede llevar a cabo por distintas técnicas de fragmentación. La disociación de iones en fase gas o, en general, la transformación al espectro de masas tándem (MS²), es una etapa clave para generar información de estructura y de secuencia. La fragmentación tándem MS/MS implica el aislamiento del ion padre, seguido de la exposición a un estímulo externo para inducir la fragmentación controlada. Es necesario fragmentar el ion precursor de una manera secuencia-específica para obtener información sobre la proteína, igualmente que ocurre en el caso de los péptidos (i. e. rotura del enlace entre cada aminoácido con una sola rotura por péptido intacto o ión proteico). Existen diferentes técnicas de fragmentación: *collision-induced dissociation* (CID), *electron capture dissociation* (ECD), *electron transfer dissociation* (ETD), *ultra violet photo dissociation* (UVPD), *high collision dissociation* (HCD), *infrared multiphoton dissociation* (IRMPD), y nuevas combinaciones de métodos de colisión con ETD (ET_hcD, ET_ciD).

Cabe distinguir, a este nivel, la diferencia en la ejecución de la fragmentación top-down de proteínas simples y la fragmentación top-down de mezclas complejas. El punto clave se centra en llegar a obtener información de secuencia a través de la fragmentación MS/MS mediante las distintas técnicas disponibles. Mientras el análisis top-down de proteínas simples se ha llevado a cabo generalmente por infusión de la proteína purificada, el reto se encuentra en el análisis eficiente de mezclas complejas, inclusive



del proteoma entero, en tiempos cromatográficos.

El análisis top-down a gran escala se empezó originalmente usando CID [71]. Por CID los iones padre seleccionados colisionan con átomos o moléculas de gas neutro (típicamente helio, nitrógeno o argón) [119], originando iones tipo “b” y “y” (Figura 5). En general, el CID no genera fragmentación completa para caracterizar la proteína (sobre todo a medida que aumenta el tamaño de la misma), pero sí suficiente para producir “tags” de secuencia para su identificación.

En general el CID es superado en eficiencia por la técnica de fragmentación HCD

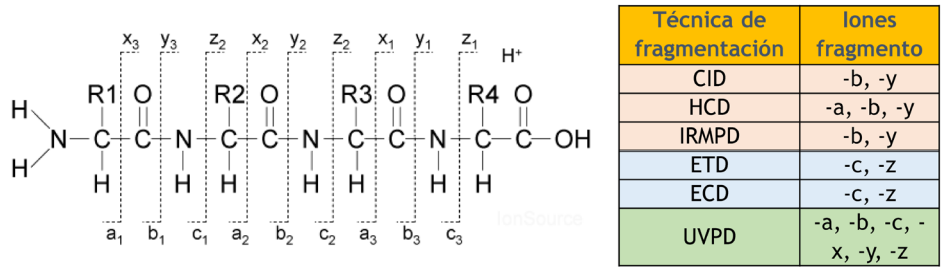


Figura 5. Nomenclatura de iones fragmento de péptidos/proteínas (Roepstorff, *Biomed Mass Spectrom.*, 1984 11, 601). Tabla con técnicas de fragmentación y tipo de iones generados.

(higher-energy collisional dissociation). El HCD se basa en la activación de iones por colisión con un haz, y se efectúa en dispositivos de cuadrupolo o multipolo originando también iones “b” e “y” y en menor medida iones tipo “a”.

Aunque es cierto que el HCD es muy eficiente para el análisis peptídico, por la estrategia top-down, se obtienen mayores grados de cobertura de secuencia proteica por técnicas de fragmentación activadas por la formación de un ion radical, como son, el ECD y el ETD, también llamadas técnica no ergódicas (*non-ergodic*). El ECD, desarrollado por Zubarev en 1998 [37], se basa en la captura de electrones por iones precursores positivos del analito para formar un catión radical inestable, que fragmentará por roturas subsecuentes específicas por el enlace n-C α . Se producen iones tipo “z”, “c”. Generalmente, la fuente de electrones es un cátodo de Cs u otro metal blando. En el caso del ETD, implementado posteriormente por un equipo de investigadores de la Universidad de Virginia [120], los electrones son transferidos desde un anión reactivo (donador de electrones), que suele ser un compuesto de fluoranteno, antraceno o azobenceno excitado. El ETD, al igual que el ECD, genera iones fragmento tipo “c” y “z”. Estas técnicas son sólo eficientes para fragmentar iones de carga mayor que 2, aunque en comparación con el CID presentan ventajas sobre la fragmentación de péptidos largos o proteínas. A nivel de péptido, el ECD y ETD retienen las modificaciones posttraduccionales lábiles, a diferencia de la fragmentación CID de péptidos trípticos, que habitualmente rompe PTMs como fosforilaciones y glicosilaciones. Sin embargo, cuando los métodos de fragmentación clásicos como el CID son aplicados en top-down, sí suelen retener también las modificaciones habitualmente lábiles. Esto es debido a que la estructura de orden superior que tienen los iones gaseosos mayores a 8 kDa conduce la fragmentación hacia el esqueleto amida preferencialmente sobre la eyección de las modificaciones (PTMs) (Figura 6). De manera destacable,

mientras los puentes disulfuro permanecen estables a la fragmentación CID, por ECD y ETD pueden ser parcialmente fragmentados [121].

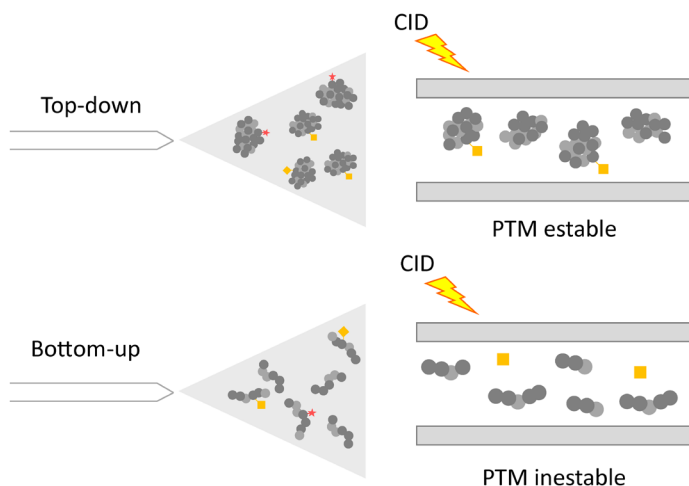


Figura 6. Estabilidad de las PTMs según el tipo de técnica de fragmentación usada. Diferencias entre las aproximaciones top-down y bottom-up.

Aunque con alto grado de rotura a lo largo de la secuencia peptídica, estas dos últimas técnicas se caracterizan por su baja eficiencia de fragmentación, entendida como la intensidad total de iones producto en relación a la intensidad del ion precursor. Por ello, el ECD y ETD se usan generalmente para la fragmentación de proteínas simples, y de gran longitud, pero no en mezclas complejas. Además, y debido a los requerimientos instrumentales, el ECD no es adecuado para todo tipo de instrumentación, quedando restringido a instrumentación ICR. El ETD sin embargo, se ha incorporado con facilidad a instrumentación de más bajo coste (tanto QTOF, como Orbitrap) y es el más usado actualmente. Concretamente, se ha extendido el uso de fragmentación ETD combinado con otras técnicas de colisión.

Como se ha mencionado en el apartado anterior, el nuevo instrumento híbrido Orbitrap, ofrece actualmente la posibilidad de fragmentar los iones por CID, ETD, HCD de manera separada pero también combinando ETD con métodos de colisión HCD o CID, en lo que se denomina ETHcD y ETcID, incrementando de esta manera la cobertura de la secuencia de proteínas. En el campo biofarmacéutico, por ejemplo, la combinación de ETD, HCD y CID en un flujo de trabajo MS3 consiguió la caracterización del 95% de la secuencia del factor de estimulación de una colonia de granulocitos [122].

Recientemente, se están desarrollando nuevas técnicas de activación de iones de proteína intacta compatibles con los requisitos instrumentales y las restricciones y rapidez requeridas para obtener suficiente información de secuencia en los tiempos de análisis cromatográfico. Concretamente, se han desarrollado técnicas basadas en el uso de fotones para depositar energía en los cationes de proteína enteros. Por un lado, la llamada "activated-ETD" (aiETD) [123, 124], usa fotones infrarrojos de baja energía para incrementar la eficiencia del ETD en nuevos instrumentos. Esta técnica es análoga al IRMPD (infrared multiphoton dissociation) asociado a ECD en instrumentos ICR.



La activación por métodos fotónicos favorece la desnaturalización de las estructuras proteicas de alto orden en la fase gas y resulta en una mayor relación S/N de los iones producto [123].

Por otro lado, se están haciendo también grandes avances por la técnica UVPD (fotodisociación ultravioleta)[125, 126]. En este método se aprovecha la absorción de fotones UV por parte del esqueleto amida para iniciar la rotura de la cadena peptídica, la cual ocurre a lo largo de toda la secuencia aminoacídica, permitiendo una cobertura de la secuencia muy alta [127, 128]. Las ventajas del UVPD se destacaron recientemente en estudios sobre variantes de proteínas fluorescentes verdes, donde este método de disociación produjo una diversidad mucho más amplia de fragmentos que mediante CID, HCD o ETD [129]. Los fragmentos que se obtienen por esta técnica, que se distribuyen a lo largo de toda la secuencia peptídica son del tipo "a", "b", "c", "x", "y", "z", "v", "w", y permiten tener una alta cobertura de la secuencia. La cobertura de secuencia casi completa que se obtiene por UVPD facilita la interrogación profunda de la estructura primaria para localizar con precisión las modificaciones y truncamientos que pueden ocurrir en las proteínas. Este nuevo tipo de fragmentación es factible a escala cromatográfica y se prevé como la tendencia para el análisis de proteínas por la estrategia top-down. Por ejemplo, Shaw et al. [130] usaron UVPD en un instrumento Orbitrap para caracterizar inequívocamente una proteoforma dioxida de la proteína Pin1 humana, una proteína relevante en los modelos de Alzheimer y enfermedad del cáncer. Asimismo, fotones UV de 193-nm se han usado recientemente para caracterizar proteínas ribosomales en LC-MS/MS sobre mezclas complejas[125] y para el análisis high-throughput de proteínas humanas en un instrumento Orbitrap [126].

6. Análisis del proteoma por Top-down Highthroughput.

Sin duda, el bottom-up es todavía la estrategia MS que se escoge para el análisis del proteoma a gran escala. Los avances tecnológicos han facilitado la identificación y cuantificación de miles de proteínas según esta estrategia a partir de muy poca cantidad de muestra [131], estableciendo así una barrera para que el top-down avance más lentamente. De todos modos, el valor añadido que proporciona poder resolver proteoformas por top-down está forzando a que se invierta en el desarrollo de esta estrategia, no sólo en el análisis de proteínas aisladas, como se ha llevado a cabo desde su introducción, sino también para profundizar a nivel de cobertura del proteoma, a pesar de la disminución de la sensibilidad que conlleva trabajar con proteínas enteras respecto al trabajo con péptidos. Sólo en los últimos años los avances de las plataformas de separación *front-end*, la mejor instrumentación, las nuevas estrategias de adquisición de datos y procesamiento están facilitando el aumento de la amplitud y profundidad de la cobertura del proteoma por esta estrategia.

Aunque han habido destacadas aplicaciones top-down a gran escala en microorganismos [65, 89, 132], el primer ejemplo de top-down proteomics a gran escala aplicado a tejidos primarios humanos fue descrito por Roth et. al usando una plataforma de

separación bidimensional en leucocitos [90]. Se caracterizaron 133 proteoformas de 67 genes únicos. Posteriormente, Tran et. al usaron una separación tetradimensional compuesta de isoelectrofocusing, seguida de GELFrEE y LCMSMS consiguiendo la identificación de 3000 proteoformas de hasta 150 kDa en líneas celulares humanas senescentes de HeLa S3, B16F10 y H1299 [71]. Este mismo sistema se analizó también por top-down usando un fraccionamiento extensivo subcelular de organelos, mejorando la cobertura del proteoma a 5000 proteoformas asociadas a 1220 proteínas [70].

Recientemente, el uso de UVPD y HCD en aplicaciones high-throughput top-down han demostrado la eficiencia de este tipo de fragmentación a escala cromatográfica en un análisis comparativo en lisados de células HeLa. La cobertura de secuencia demostró ser superior por UVPD que por HCD ($28 \pm 20\%$ para UVPD frente a $17 \pm 8\%$ por HCD), mientras se detectaron 153 proteínas y 489 proteoformas usando UVPD y 271 proteínas y 982 proteoformas por HCD [126]. Sin embargo, la utilización de UVPD por highthroughput top-down está limitada por la bioinformática a causa de la complejidad de los espectros MS2, añadida a la intrínseca complejidad del análisis top-down en proteínas.

Podemos concluir que la proteómica top-down *high-throughput*, en condiciones desnaturalizantes, (siguiendo flujos de trabajo análogos y originalmente desarrollados para la proteómica bottom-up) ha evolucionado lo suficiente para permitir el análisis top-down de proteínas <50 kDa con un grado de cobertura del proteoma cualitativo y cuantitativo similar al del bottom-up. Como hemos visto, la identificación fiable y caracterización de varios miles de proteoformas es factible [133-135]. Sin embargo, el número de proteoformas identificadas por top-down en condiciones desnaturalizantes de rutina cae a pesos moleculares superiores a 50 kDa a causa de problemas técnicos, tanto de limitaciones de la espectrometría de masas como de la necesidad de mejoras en las separaciones cromatográficas o *front-end*, como también en la dificultad del análisis automatizado de los datos. Soluciones para extender el rango de pesos moleculares de las proteoformas identificadas durante experimentos de top-down desnaturalizantes incluyen el uso de nuevas generaciones de instrumentos de 21 T FT-ICR [136], Orbitraps modificados (mass extended) o nuevas estrategias de adquisición de datos, como las desarrolladas por Y. O. Tsybin, dirigidas a incrementar la relación S/N espectral de proteoformas largas mediante la grabación de *transients* de corto tiempo de dominio promediadas a lo largo de varios *runs* cromatográficos (*short time-domain transients averaging*) [23].

Aunque se requiere todavía una cobertura más profunda del proteoma por top-down proteomics las oportunidades de expansión del top-down a gran escala pasan también por la cuantificación de alto rendimiento de los proteomas intactos.

7. Hacia la proteómica top-down cuantitativa.

La habilidad de cuantificar la abundancia relativa de proteoformas concretas entre colecciones de muestras es uno de los objetivos clave para que la metodología top-down



se extendía en proteómica clínica y especialmente en el campo de descubrimiento de biomarcadores. Ya en 2006, Pesavento et al. [50] usaron simplemente la medida de los ratios de intensidad de múltiples proteoformas de histona H4 que coeluyen y de sus fragmentos, para tener una cuantificación relativa entre espectros en una sola muestra. Su trabajo introdujo el PIRR (*protein ion relative ratio*) y el FIRR (*fragment ion relative ratio metrics*) calculado a partir de la integración de los isótopos más abundantes de estados de carga escogidos de iones precursores intactos y de iones fragmento [50]. Esta técnica ha sido usada recientemente en una serie de estudios para cuantificar proteoformas fosforiladas en troponin I cardíaca (cTnI), el biomarcador más típico de daño cardíaco agudo. El estudio se ha realizado en: tejido de corazón humano [137] ratón WT y cTnI-mutado [138], modelos de rata hipertensas con fallos cardíacos [139], modelos porcinos [140] y monos Rhesus [141]. Aunque la técnica PIRR tiene gran potencial, cuando se aplica a múltiples muestras en proteomas enteros, tiene todavía limitaciones provenientes de su inherente variabilidad asociada al análisis proteómico a gran escala y requiere de mayor desarrollo.

Tal como describe Neil Kelleher en su reciente artículo de revisión "*Progress in Top-Down proteomics and the analysis of proteoforms*" [142], se han diseñado diferentes aproximaciones top-down cuantitativas. Para mejorar la variabilidad en la cuantificación de proteoformas en experimentos a gran escala en proteomas enteros, existen, en primer lugar, estrategias de marcaje isotópico *in vitro* e *in vivo*. Por ejemplo, el marcaje de cisteínas y el marcaje metabólico $^{14}\text{N}/^{15}\text{N}$ se han aplicado como prueba de concepto en proteomas de levadura [86, 143]. Estas técnicas, sin embargo, presentan limitaciones de rango dinámico y de señal/ruido (S/N). También se ha probado SILAC top-down [144], aunque se han descrito limitaciones de marcaje completo de la arginina pesada a medida que aumenta el número de argininas y el peso molecular de las proteínas [144]. Para intentar paliar esta limitación, el mismo grupo que desarrolló esta aproximación determinó el número de aminoácidos marcados que se podían marcar en proteoformas determinadas y la eficiencia de incorporación de aminoácidos y empleó de manera eficiente esta metodología top-down SILAC a células madre de embriones humanos [145]. Hung et al. [146] han desarrollado también estudios piloto para aplicar *Tandem Mass Tags* (TMT) para la cuantificación top-down. Esta aproximación es muy atractiva dado que es una aproximación de marcaje *in vitro* dependiente de la fragmentación en tándem y los pares marcados isotópicamente mantienen los mismos perfiles cromatográficos. Otros métodos previos como el marcaje de cisteínas [143] alteraban los tiempos de retención y por tanto impedían la cuantificación *intra-spectrum*.

Los métodos *label-free* para cuantificar relativamente proteomas enteros a resoluciones de proteoformas son los que aparentemente tienen más garantías de éxito en la aproximación top-down. Son métodos de menor dificultad en la preparación de las muestras, mucho más baratos al evitar reactivos de marcaje y aplicables a muestras humanas primarias. La espectrometría de masas diferencial, desarrollada por Yates y su grupo [147] para el análisis peptídico, representa una opción viable para el análisis de proteína intacta. Esta técnica consiste en un flujo proteómico automatizado que proporciona cuantificación relativa de conjuntos de datos de LC-MS, requiriendo solamente información de m/z , tiempos de retención e intensidad de picos de interés, así como numerosos replicados para disminuir el ruido [147]. El método encuentra selectivamente diferencias estadísticamente significativas en la intensidad de los iones

de alta y baja abundancia, contabilizando la variabilidad de las intensidades medidas y el hecho de que las verdaderas diferencias persistirán en el tiempo. Como ejemplo, Mazur et al. publicaron en 2010 la cuantificación de diversas proteoformas de apolipoproteína in HDL (lipoproteínas de alta densidad) aislado de pacientes con colesterol alto y bajo [103].

En el contexto de la cuantificación top-down *label free*, Whitelegge y su grupo han desarrollado el concepto IMT (intact mass tag) [148, 149]. IMT corresponde a una señal de masa significativa, produciendo un peso molecular determinado que puede ser comparado en una base de datos, asignado e identificado. Trabajos en cistatinas y otras proteoformas secretoras han demostrado que, una vez definido el IMT, estos pueden usarse para monitorizar cambios en su abundancia en diferentes muestras, de manera cuantitativa. Este tipo de estrategias son muy posiblemente las que marquen el rumbo de la proteómica cuantitativa top-down a gran escala, dado que el marcaje metabólico no es posible en muestras primarias humanas. Por ejemplo, Castagnola et al. [150] han usado abundancias de iones extraídos de IMTs de interés en saliva humana para demostrar la hiperfosforilación de proteínas salivares en niños autistas, lo cual sugiere asincronías moleculares que podrían estar relacionadas en el sistema nervioso central, el desarrollo y la patogénesis del autismo. Una aplicación reciente que combina el concepto IMTs (usando concretamente bases de datos con masas exactas específicas y etiquetas de tiempo) y MS diferencial ha demostrado ser útil para la cuantificación de 83 proteoformas de 20 IDs en parótida humana y secreciones de la glándula submandibular [151].

Es indudable que la cuantificación de proteoformas en proteomas enteros, incluso con los flujos *label-free* que se están desarrollando, sufre todavía de muchas limitaciones. Las evaluaciones estadísticas, sin ir más lejos, están limitadas al test de Student's, que son insuficientes para abordar las diversas fuentes de variación técnica inherentes de estudios comparativos complejos a múltiples niveles [73]. En este sentido, Ntai et al. [73] aplicaron en un estudio piloto en un mutante de delección de levadura un modelo estadístico jerárquico lineal que permitía efectos anidados usando dos análisis separados ANOVA para probar el significado estadístico de los cambios de intensidad y estimar los efectos de tamaño para la cuantificación con QMTs (*quantitation mass targets*). Este estudio detectó un total de 838 QMTs. 120 QMTs demostraron tener un cambio de intensidad significativo entre líneas mutadas y naturales (WT), un 63% fueron identificados con fiabilidad por tándem MS o por la aproximación IMT [73]. Esta aproximación top-down *label free*, diseñada para corregir la alta variabilidad en pacientes y tratamientos, representa una herramienta muy interesante en investigación traslacional, puesto que es aplicable a diversas muestras primarias, que como se ha mencionado con anterioridad, no pueden ser marcadas metabólicamente.

8. Espectrometría de masas top-down nativa.

El campo de la espectrometría de masas nativa está basada en la habilidad de mantener intacta la estructura de las proteínas y los complejos proteicos dentro del espectrómetro de masas, conservando las interacciones no covalentes débiles, tanto



intramoleculares como entre subunidades proteicas u otros ligandos como ADN, co-factores, etc... [5, 152]. La proteína o complejos proteicos se introducen disueltos en tampones acuosos, sin adición de ácidos ni disolventes orgánicos desnaturizantes. Se trabaja en condiciones de fuentes optimizadas y vacíos modificados (*ion-cooling*) para no romper la estructura tridimensional original ni las interacciones no covalentes [153].

La espectrometría de masas nativa es una herramienta poderosa en biología estructural. En combinación con la espectrometría de masas top-down proporciona información sobre la composición, estequiometría, arquitectura de las subunidades, organización topológica, sitios de unión y ofrece una manera de desentrañar la inherente heterogeneidad de los ensamblajes de proteínas. Actualmente la espectrometría de masas nativa top-down permite investigar la diversidad multinivel de los complejos proteicos. El reto se centra en el desarrollo de métodos capaces de capturar estados moleculares específicos de complejos proteicos o *multiproteoform complexes* (MPCs) (que incluyen diferentes proteoformas, composición y estequiometría) [154, 155] en fuentes endógenas. Sin duda, esta información sobre las proteínas, todavía más específica que la que ofrecen las proteoformas a nivel individual, proporciona un mayor conocimiento de los procesos que dirigen la biología humana y la evolución de enfermedades.

Native top-down ha sido usada tradicionalmente para el análisis dirigido de complejos de proteínas purificadas, permitiendo el análisis de ensamblados macromoleculares con masas en el rango de los megadalton, como el caso del ribosoma intacto [156, 157] o ensamblados de cápsides virales [158]. Además, native top-down puede proporcionar información única de las perturbaciones proteoforma-proteoforma que regulan procesos celulares específicos [159].

Recientemente, en un Orbitrap modificado, se ha descrito la heterogeneidad del complejo FBP1, el enzima que limita la velocidad en la glucogénesis, usando un nuevo método native-top-down realizando un análisis a tres niveles (MS3), empezando a partir del tetrámero intacto, analizando sus subunidades de composición y finalmente la secuencia primaria de cada subunidad, todo en tres etapas dentro del espectrómetro [160].

También hay ejemplos de nuevos métodos de espectrometría de nativa para la rápida caracterización de proteínas y complejos proteicos sobreexpresados directamente de lisados de células de *E. coli*. Estos nuevos métodos de preparación de muestra evitan las tediosas etapas de prepurificación para compatibilizar con la espectrometría de masas nativa. Combinándolos con el top-down, que efectúa la eyección gradual de subunidades intactas en el espectrómetro y su posterior fragmentación, se ha podido caracterizar recientemente de forma precisa los MPCs en diversos sistemas [161].

Alternativamente a su papel complementario en biología estructural [162], las altas capacidades de análisis de proteínas de alto peso molecular de la proteómica top-down nativa, quizás son usadas algún día en-línea con técnicas de separación adecuadas, tales como la cromatografía de exclusión molecular, incrementando potencialmente el análisis top-down *high-throughput* y dirigiendo la espectrometría de masa nativa hacia el "Discovery mode" [163]. Junto con el desarrollo de la instrumentación, esta posible

dirección de la espectrometría de masas nativa fomentaría la creación de nuevas herramientas informáticas, particularmente cuando el análisis se centra en los complejos multiproteoformas, requiriendo algoritmos de búsqueda en base de datos que consideraran tanto la estructura o ensamblaje intacto como la de las subunidades [154].

La espectrometría de masas nativa top-down tiene el potencial de acceder a las regiones de alto peso molecular del proteoma humano, actualmente inexploradas por la técnica *highthroughput* top-down, así como de proporcionar información mejorada y detalles moleculares.

9. Análisis de datos top-down.

La proteómica top-down es una potente tecnología que nos permite el análisis directo de todas las proteoformas, ya sean derivadas de splicing alternativo, de mutaciones puntuales o de PTMs y que nos ofrece una información más precisa y complementaria a la proteómica bottom-up. El análisis de los datos top-down obtenidos a partir de espectrometría de masas representa un desafío, tanto por la deconvolución de los espectros como su posterior búsqueda en bases de datos. A diferencia de la proteómica bottom-up, donde se encuentran soluciones robustas para la identificación, caracterización y cuantificación de péptidos y proteínas, en la proteómica top-down el análisis de los datos no está aún bien establecido.

Una de las mayores limitaciones en el análisis de proteínas intactas se encuentra a la hora de lidiar con la complejidad de los espectros obtenidos a alta resolución. La ionización por electrospray de péptidos y proteínas produce espectros de masas que contienen series de iones múltiplemente cargados con su masa carga (m/z) asociada. Cuanto mayor es la secuencia de amino ácidos más probabilidad hay de encontrar múltiples estados de cargas por lo que el espectro resultante es complejo y difícil de interpretar. Por este motivo, el primer paso en el análisis de los datos de proteómica top-down es la deconvolución de los espectros adquiridos en alta resolución de MS y MS/MS. Con la deconvolución, la lista compleja de señales m/z con múltiples isótopos de un espectro de masas se traduce a una lista de masas monoisotópicas con carga uno o zero mucho más simple y fácil de analizar. Existen diferentes algoritmos que permiten la deconvolución de espectros de masas de alta resolución. El primero y uno de los más utilizados es el Thorough High-Resolution Analysis of Spectra by Horn (THRASH), que automáticamente detecta la distribución de los isotopómeros experimentales y la compara con la distribución teórica [165]. El algoritmo Xtract [165], muy parecido al TRASH, realiza la deconvolución de espectros resueltos isotópicamente, es decir, espectros en los que es posible distinguir entre picos de diferente composición isotópica de un mismo componente. El MS-deconv es otro algoritmo muy utilizado que usa una aproximación combinatorial para la deconvolución de espectros [166]. A diferencia de los algoritmos anteriores, el ReSpect (Positive Probability, Ltd) deconvolucionan espectros que no están resueltos isotópicamente. Una de las limitaciones de este tipo de algoritmos para la deconvolución de espectros de alta resolución es que algunos de los picos experimentales quedan sin asignar, con la consecuente falta de identificación de algunas proteoformas.



Existen dos grandes tipos de aproximaciones para la identificación de proteoformas a partir del análisis de datos top-down: a) el análisis basado en la búsqueda en bases de datos anotadas, donde se incluyen todas las posibles proteoformas con las distintas modificaciones post-traduccionales definidas en una determinada posición y b) el análisis basado en algoritmos de alineamiento de espectros [167], que contrastan los espectros de masas en tándem de las proteínas con la base de datos sin anotar, donde no se requiere un conocimiento previo de las posibles PTMs ni su lugar de localización.

Actualmente, existen una gran variedad de herramientas para el análisis de datos de proteómica top-down como el MS-Align+ [168], el MSPathFinder (<https://omics.pnl.gov/software/mspathfinder>) o el Mash Suite Pro [169]. En el web del Consortium for Top-down proteomics (<http://www.topdownproteomics.org/resources/software/>) se recogen los principales softwares para tratar este tipo de datos. Vamos a poner como ejemplo dos softwares que utilizan estrategias de búsqueda distintas. El ProSight [170], como ejemplo de software que realiza las búsquedas sobre bases de datos anotadas y el TopPic, que basa su búsqueda en algoritmos de alineamiento de espectros.

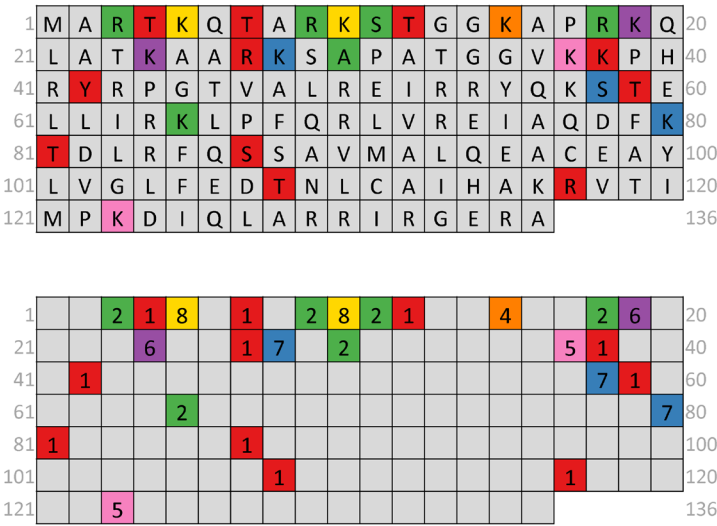
El ProSight PTM [170] fue el primer software desarrollado específicamente para la identificación de proteínas a partir del análisis de proteómica top-down. Actualmente también existe el ProSight PD, que es un nodo integrado dentro del Proteome Discoverer (Thermo Scientific). La integración con el Proteome Discoverer facilita su uso y la visualización posterior de los datos. Además, los resultados están integrados con el ProSight Lite [171] para una validación posterior de una determinada proteína en caso necesario.

El ProSight usa el Xtract o el TRASH como algoritmos para deconvolucionar espectros o el Respect cuando se trata de espectros sin suficiente resolución. La base de datos que se utiliza para las búsquedas puede ser en formato fasta, por lo que no se tienen en cuenta las distintas proteoformas, o puede ser una base de datos anotada donde se incluyen todas las posibles proteoformas ya sean splicing alternativo, mutaciones puntuales, puentes disulfuro o PTM. El formato de entrada de la base de datos es un fichero .txt o .xml que puede descargarse directamente de UniProt. Este tipo de búsquedas tiene dos evidentes limitaciones. En primer lugar y como ya hemos comentado, se requiere un conocimiento previo de la muestra para incluir en la base de datos todas las proteoformas esperables. Si estas proteoformas no están anotadas en la base de datos el usuario debe introducirlas para que se tengan en cuenta en la búsqueda y actualmente no existen herramientas para anotar las bases de datos de forma sistemática y automatizada. La segunda limitación es que el número de entradas en la base de datos crece exponencialmente con el número de modificaciones o variantes a tener en cuenta para cada una de las secuencias canónicas. Podemos calcular el número de proteoformas teóricas para una determinada base de datos mediante la siguiente fórmula:

$$Proteoformas = \sum_{j=1}^P \left[\prod_{i=1}^n (m_i + 1) \right]_j$$

donde P es el número de proteínas en la base de datos, n el número de amino ácidos en una determinada secuencia y m el número de posibles modificaciones para cada uno de los amino ácidos [4].

Para hacernos una idea de la magnitud de estos números vamos a calcular el número posible de proteoformas para la histona H3.1 (P68431). Esta proteína tiene anotados en UniProt 86 residuos modificados, algunos de ellos compartidos en el mismo amino ácido. El número total de proteoformas para solo esta entrada es de $5.46 \cdot 10^{14}$ (Figura 7). De esta manera el tamaño de la base de datos hace que los tiempos de búsqueda se hagan inalcanzables. Por este motivo, en este tipo de búsquedas se limita el número máximo de modificaciones para una misma secuencia.



$$Proteoformas\ H3.1 = \prod_{i=1}^n (m_i + 1) = (2 + 1)(1 + 1)(8 + 1) \cdots (1 + 1)(5 + 1) = 5.46 \cdot 10^{14}$$

Figura 7. Secuencia de la histona H3.1 (P68431). En color se indican los amino ácidos que tienen alguna modificación anotada en UniProt (arriba). Abajo se indican el número de modificaciones anotadas para cada uno de los amino ácidos. La ecuación calcula el número total de proteoformas para esta secuencia teniendo en cuenta las modificaciones anotadas para cada uno de los residuos.

Si salvamos el problema de la magnitud de la base de datos y del conocimiento previo de las PTMs en las que estamos interesados, el ProSight ofrece una búsqueda a tres niveles que es muy interesante para la identificación y caracterización de proteínas. En primer lugar, se realiza una búsqueda de masa absoluta en la que se buscan aquellas proteoformas definidas como tal en la base de datos. En este caso, el error de masa del ion precursor se define a pocos Da, en función de la resolución a la que se haya adquirido el espectro de MS1 y el error de los fragmentos a pocos ppm. En el segundo nivel, se efectúa una búsqueda de biomarcadores sobre los espectros que no han sido identificados en el primer nivel. El objetivo es encontrar formas truncadas de las proteoformas, se buscan tags de secuencia en los espectros de MS/MS y se ajusta la masa experimental a la masa de la cadena truncada con muy poco error. En el tercer nivel se analizan los espectros que no se han identificado en los otros dos niveles y se realiza una búsqueda también de masa absoluta, pero en este caso la tolerancia



del ion precursor se define a centenaes de Da. El objetivo aquí es encontrar modificaciones no descritas en la base de datos. En este tercer nivel hay que seleccionar la opción de Delta M Mode, que permite identificar PTMs desconcidas. El delta m es la diferencia entra la masa del precursor teórica y la experimental.

El TopPIC Suite [172], a diferencia del MS Align + o el MSPathFinder (dos herramientas basadas también en la alineación de espectros), que sólo pueden usarse mediante línea de comandos, está también disponible con interfaz gráfica, lo que lo hace sencillo de utilizar. Está compuesto de dos herramientas: el TopFD (Top-down mass spectral Feature Detection) y el TopPic (Top-down mass spectrometry based Proteoform Identification and Characterization).

El TopFD deconvoluciona los espectros generando una lista de masas monoisotópicas. Los ficheros .msalign de salida del TopFD son, a su vez, los ficheros de entrada para el TopPIC.

El TopPIC es el motor de búsqueda propiamente dicho y es capaz de identificar y caracterizar las proteoformas de un proteoma entero mediante la búsqueda top-down frente a una base de datos fasta sin previo conocimiento de las posibles modificaciones o PTMs existentes en tiempos de búsqueda razonables. Para ello realiza una búsqueda en tres pasos para la identificación de PrSM (proteoform spectrum matches). En primer lugar, filtra las proteínas para reducir el número de proteínas candidatas. Luego usa un algoritmo de alineamiento de espectros. Este algoritmo encuentra el alineamiento óptimo entre un espectro de masas y la secuencia de una proteína, usando un número especificado de desplazamiento de masas, que se corresponden con las modificaciones no esperadas. Finalmente calcula el E value para cada PrSM candidato. Los PrSM con mejor E value son los que se reportan en última instancia.

El TopPIC tiene un segundo componente que usa el MISScore, basado en modelos Bayesianos [173], para encontrar las modificaciones que mejor explican este desplazamiento de masa en los distintos PrSM.

La principal limitación del TopPIC suite es que en la tabla de salida las modificaciones se muestran como una desviación de masa y es el usuario quien debe decidir a que corresponde esa desviación, si a una PTM, combinaciones de PTMs o mutaciones de aa en la secuencia. A priori, puede parecer algo sencillo, pero cuando se analizan proteínas con diferentes posibles variaciones es una tarea complicada que requiere de validación manual.

El análisis de datos top-down es todavía unos de los cuellos de botella para la estrategia proteómica top-down. Uno de los principales problemas es que en muchas ocasiones y, en último término, la validación final de las identificaciones pasa por una confirmación o inspección manual de las asignaciones. Sin embargo, en los últimos años, se han producido numerosos avances y se han desarrollado distintas herramientas para la identificación y caracterización de proteoformas de forma automatizada y con interfaz gráfica que facilitan su utilización a nivel de usuario. Incluso existen herramientas que dan un paso más allá y analizan datos de proteómica nativa top-down para elucidación de complejos macromoleculares proteicos [154].

10. Aplicaciones Top-down.

La proteómica top-down se aplica en diversos campos tales como: venómica [174, 175]), análisis de proteína intacta [176], cross-linking y análisis estructural de proteínas [177, 178], investigación biomédica, clínica etc. A continuación, se describen las aplicaciones destacadas en investigación biomédica y clínica:

10.1. Top-down en investigación biomédica y clínica

El análisis de las proteoformas a nivel de proteína intacta da un valor añadido a la investigación clínica y traslacional proporcionando información muy valiosa de la complejidad molecular que no puede ser abordada únicamente por el análisis a nivel de péptido. Gregorich et al. [179] describieron los retos y ventajas de las estrategias top-down centradas en la compresión y diagnosis de enfermedades humanas. Esta aproximación única, puede proporcionar la cuantificación de formas modificadas de proteínas, el mapa completo de las modificaciones junto a las secuencia completa, el descubrimiento de modificaciones no esperadas, la identificación y cuantificación de isómeros posicionales y la determinación del orden y jerarquía de múltiples modificaciones [180].

Zhang et al. usaron top-down, por ejemplo, para el estudio de la fosforilación de la troponin I cardiaca en muestras post-mortem y de trasplante humano de tejido de corazón como potencial biomarcador de la insuficiencia cardiaca crónica [141]. El mismo grupo de investigación reportó por top-down también, nuevas posiciones de fosforilación y variantes de *splice*, presentes en tejidos cardiacos humanos y de ratón [181]. Ying Ge y colaboradores pudieron relacionar según esta estrategia la alteración de modificaciones posttraduccionales en troponin-I cardiaca (cTNI) a fallo cardiaco (HF heart failure) asociado a disfunción contráctil, tanto en animales modelo con HF como en muestras clínicas [141].

Pero el uso de la espectrometría de masas top-down para medir proteoformas clínicamente relevantes no es nuevo [182]. Ya fue usada, por ejemplo, en 2013, cuando Bruker recibió una beca para desarrollar esta tecnología con un sistema MALDI Bio-typer CA, para la identificación de bacterias Gram-negativas cultivadas de pacientes [18]. En este sistema, una colonia microbiana se dispone directamente sobre una placa de MALDI y se proporciona mediante la estrategia top-down el perfil de las proteínas más abundantes. Comparando el perfil proteico con librerías, se identifican las bacterias a nivel de género, especie e incluso subespecie. Esta nueva capacidad no solo cambia el cálculo de negocios en relación con el desarrollo de antibióticos específicos sino también puede reducir el problema de resistencia a los antibióticos. Al operar a nivel de proteoforma (es decir, descifrando la variación de secuencia de proteínas ribosómicas enteras) este ensayo se hace económicamente viable y funcional.

Nuevos ensayos clínicos top-down han demostrado recientemente la monitorización efectiva de proteoformas marcadoras claves en diabetes con cuantificación precisa y con niveles de detección altos de PTMs partiendo de solo 5 ul de sangre [183]. Un



número creciente de laboratorios trabajando en la interfaz de la tecnología y la salud humana están mostrando el uso emergente del top-down [184, 185].

Otros ejemplos destacables del uso de top-down para el diagnóstico de enfermedades son, la detección de cambios en la secuencia de transtieretina en plasma para diagnosticar amiloidosis hereditaria [186, 187], o el análisis de variantes de hemoglobina humana A2 en eritrocitos para diagnosticar enfermedades de la sangre tales como la talasemia [188, 189]. Estos métodos top-down de cuantificación de proteínas y variantes concretas son aplicables en ion traps y han demostrado ser muy precisos analíticamente y factibles para el uso en laboratorios diagnósticos. De igual manera, métodos clínicos para la medida del factor de crecimiento de insulina en suero se usan como indicativos de anomalías en el crecimiento [190].

La tendencia del top-down en clínica, apunta hacia análisis más dirigidos (targeted) [142]. Es por tanto muy factible que se empiecen a diseñar modos de adquisición *targeted* MS/MS para proteoformas enteras (como en los métodos SRM, single-reaction –monitoring) con el fin de cuantificar la cantidad de proteoformas minoritarias determinadas en muestras complejas (ver apartado de cuantificación de proteomas intactos). En este sentido, se han desarrollado ya recientemente nuevos métodos top-down dirigidos, a nivel de iones fragmento MS2, para el análisis de variantes de hemoglobina causadas por polimorfismos de nucleótido único, para la diagnosis de la talasemia [191].

Por otro lado, en clínica, la combinación de la espectrometría de masas *imaging*, capaz de mapear la distribución molecular de las especies directamente sobre tejidos biológicos, con la espectrometría de masas top-down es también una herramienta en desarrollo. Esta combinación permite estudiar la clasificación molecular de tejidos e identificar biomarcadores proteicos, generando potencialmente un considerable valor diagnóstico y pronóstico [192] [193]. Históricamente, en la espectrometría de masas *imaging* se ha usado la técnica de ionización por MALDI. De manera destacable, LESA (Liquid Extraction Surface Analysis), es otra técnica de análisis similar sobre tejidos u otras superficies, que se usa acoplada a la espectrometría de masas usando ionización por electrospray. LESA-MS ha demostrado ser útil para el diagnóstico y detección de ciertas proteínas biomarcadoras de enfermedades no alcohólicas del hígado (NASH, non alcoholic steatohepatitis) [194] en secciones de tejidos de biopsias, y también en el determinados tipos de análisis de sangre.

Para que el top-down sea de uso general en diagnosis clínica, los espectrómetros de masas deben evolucionar aún más para realizar experimentos de análisis de proteína intacta de manera rutinaria. Aunque hemos visto que la tendencia está cambiando, la mayoría de instrumentos todavía están desarrollándose con el análisis de péptidos en mente. La cantidad de muestra requerida y el grado de manipulación de la muestra deben también disminuir cuando se trabaja con muestras únicas, como por ejemplo, las biopsias. Además, la incorporación del acoplamiento LCMS en técnicas como el LESA, y una mayor automatización del proceso, son obligatorias para su uso eficiente y aplicación en diagnosis.

11. Conclusiones y perspectivas.

A día de hoy, el bottom-up y el top-down son técnicas complementarias. Mientras el bottom-up permanece inigualable en profundidad de cobertura del proteoma y cuantificación de péptidos, la proteómica top-down caracteriza las proteoformas y sus variantes directamente. Por otro lado, el top-down nativo puede estudiar directamente complejos de proteínas y complejoformas [195].

A medida que la técnica top-down continua evolucionando, tanto en condiciones desnaturalizantes, como de manera nativa, se prevé una tendencia o punto de inflexión donde el análisis de proteína intacta se extenderá aún más. La razón es clara. El análisis top-down y las medidas encaradas a resolver y caracterizar proteoformas tienen un valor de gran especificidad molecular, lo cual ofrece fuertes conexiones mecanísticas a las patologías.

La capacidad de resolver, caracterizar y relativamente cuantificar proteoformas concretas son el verdadero potencial de la estrategia top-down. Esta información no es asequible por lo general mediante la proteómica clásica, que trabaja a nivel de péptido. Se han nombrado numerosos ejemplos y aplicaciones top-down en clínica y biomedicina, aunque en la actualidad las aplicaciones se han extendido y son de diversa naturaleza. Van desde la elucidación de distintas proteoformas de alérgenos en mostaza [196], la determinación de la calidad del pan de harina [197], hasta el análisis de células mononucleares de sangre periférica humana [198], usando tanto instrumentación de cuadrupolo-TOF como Orbitrap.

En clínica, se espera que cada vez se encuentren biomarcadores más específicos, resultado de poder concretar en proteoformas funcionales y no tanto en péptidos tripticos, que pueden haber perdido la información característica principal de la proteína a la que han estado asignados [199, 200]. Estamos en el inicio de una era, la de las proteoformas, en la cual, su principal herramienta de análisis, la proteómica top-down, debe evolucionar aún más para alcanzar las promesas que engloba esta técnica.

En muy poco tiempo se han ido sucediendo diferentes aplicaciones que van en esta línea:

En 2009, la ionización SELDI (Surface –enhanced laser desorption/ionization) TOF-MS se usó ya para correlacionar proteoformas S100A6 de pacientes con supervivencia en estado I de células cancerígenas de pulmón [201]. En 2011, proteoformas fosforiladas fueron apuntadas también como candidatos biomarcadores en insuficiencia cardíaca crónica en FTICR-MS [141]. A continuación, y con instrumentación Orbitrap, se han llegado a identificar biomarcadores intactos en fluido cerebrospinal para la prognosis pediátrica de tumores cerebrales [202] y múltiples proteoformas en saliva potencialmente asociadas a complicaciones del síndrome de Down, incluyendo la aparición temprana de la enfermedad de Alzheimer [203]. Top-down se ha usado en estudios para encontrar nuevos marcadores de cirrosis, conteniendo proteoformas ubiquitinadas [204], de esquizofrenia y trastornos bipolares [184], desarreglos del neurodesarrollo (en combinación con MALDI-Imaging) [205], la enfermedad de Parkinson (pro-



teoformas de α -sinucleína) [206] y en meduloblastoma y astrocitoma pilocítico [207].

Estando de acuerdo en el gran potencial del top-down y de las ventajas a nivel de la información que es capaz de proporcionar, la pregunta que uno se formula es cuando se producirá la inflexión en la adopción más generalizada de la técnica y despliegue de la tecnología proteómica top-down, tanto en investigación académica como clínica. Este es el quid de la cuestión.

Hay una afirmación común en el hecho que el top-down está varios años por detrás del bottom-up en términos de aceptación e implementación. Siendo esto cierto, el incremento en número de usuarios de esta metodología y los avances mencionados y en constante evolución serán los agentes de conducción hacia un cambio de rumbo en la proteómica aplicada.

Es cierto que el top-down ha sido un reto tecnológicamente y todavía existe esta percepción histórica en la mente de expertos en el campo. Como hemos visto, los avances en instrumentación por la mayor parte de fabricantes durante la última década han hecho factibles el top-down dirigido (*targeted*) y a gran escala (*high-throughput*) incluso en espectrómetros más asequibles económicamente. Por otro lado, una serie de practicantes de la Proteómica han fundado el "Consortium for Top-down proteomics" (CTDP, <http://topdownproteomics.org/>), que tiene el objetivo de acelerar la investigación colaboradora [208], de incrementar la visibilidad de la proteómica top-down dentro de la comunidad proteómica y la difusión de conocimiento y de las mejores prácticas para recién llegados en el campo. Es de destacar que, el primer manuscrito del CTDP centrado en el análisis de las proteoformas tiene actualmente más de 250 citas [2]. Como un reconocimiento del progreso acelerado en top-down proteomics, el Uniport Knowledgebase ha empezado a cruzar referencias de sus números de acceso con los identificadores permanentes (PFR) en el Proteoform Atlas alojado por el CTDP (<http://repository.topdownproteomics.org/>).

La idea principal del top-down es que la información que se obtiene tiene una conexión mucho más directa con los fenotipos complejos que dictan la salud y la enfermedad que la que se obtiene por el análisis bottom-up. Por tanto, estudiar el proteoma a nivel de proteína intacta y de proteoforma, va a proporcionar una mejor traducción de los datos proteómicos a la comprensión de los fenotipos y la detección precoz de las enfermedades. La estrategia top-down irá siendo empujada a la vanguardia de la proteómica y la investigación biomédica a medida que el análisis de las proteoformas ayude a entender la complejidad de los fenotipos y esto produzca una presión y motivación superior sobre los grupos de investigación y la industria para focalizarse aún más en su desarrollo.

REFERENCIAS

1. Munoz, J. and A.J. Heck, From the human genome to the human proteome. *Angew Chem Int Ed Engl*, 2014. 53(41): p. 10864-6.
2. Smith, L.M., N.L. Kelleher, and P. Consortium for Top Down, Proteoform: a single term describing protein complexity. *Nat Methods*, 2013. 10(3): p. 186-7.
3. Kelleher, N.L., A cell-based approach to the human proteome project. *J Am Soc Mass Spectrom*, 2012. 23(10): p. 1617-24.
4. Aebersold, R., et al., How many human proteoforms are there? *Nat Chem Biol*, 2018. 14(3): p. 206-214.
5. Tsiatsiani, L. and A.J. Heck, Proteomics beyond trypsin. *FEBS J*, 2015. 282(14): p. 2612-26.
6. Cotham, V.C. and J.S. Brodbelt, Characterization of Therapeutic Monoclonal Antibodies at the Subunit-Level using Middle-Down 193 nm Ultraviolet Photodissociation. *Anal Chem*, 2016. 88(7): p. 4004-13.
7. Resemann, A., et al., Full validation of therapeutic antibody sequences by middle-up mass measurements and middle-down protein sequencing. *MAbs*, 2016. 8(2): p. 318-30.
8. Wu, C., et al., A protease for 'middle-down' proteomics. *Nat Methods*, 2012. 9(8): p. 822-4.
9. Fornelli, L., et al., Middle-down analysis of monoclonal antibodies with electron transfer dissociation orbitrap fourier transform mass spectrometry. *Anal Chem*, 2014. 86(6): p. 3005-12.
10. Moelleken, J., et al., GingisKhan protease cleavage allows a high-throughput antibody to Fab conversion enabling direct functional assessment during lead identification of human monoclonal and bispecific IgG1 antibodies. *MAbs*, 2017. 9(7): p. 1076-1087.
11. Weinglass, A.B., et al., Exploring the role of a unique carboxyl residue in EmrE by mass spectrometry. *J Biol Chem*, 2005. 280(9): p. 7487-92.
12. Cristobal, A., et al., Toward an Optimized Workflow for Middle-Down Proteomics. *Anal Chem*, 2017. 89(6): p. 3318-3325.
13. Dole M., M.L.L., Hines R. L., Mobley R. C., Ferguson, L.D, Alice M. B, Macromolecular beams of macroions. *J. Che. Phys.*, 1968. 49: p. 2240-2249.
14. Fenn, J.B., et al., Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 1989. 246(4926): p. 64-71.
15. Meng, C.K. and J.B. Fenn, Analyzing organic molecules with electrospray mass spectrometry. *Am Biotechnol Lab*, 1990. 8(4): p. 54-60.
16. Fenn, J.B., Mann, C. K., meng, S. F., Wong, S. F., Whitehouse, C. M. , Electrospray ionization: principles and practice. *Mass Spectrom. Rev.*, 1990. 9: p. 37-70.
17. Karas, M. and F. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*, 1988. 60(20): p. 2299-301.
18. Tanaka, K., Waki, H, et al. , Protein and polymer analysis up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.*, 1988. 2: p. 151-153.
19. Catherman, A.D., O.S. Skinner, and N.L. Kelleher, Top Down proteomics: facts and perspectives. *Biochem Biophys Res Commun*, 2014. 445(4): p. 683-93.
20. Tipton, J.D., et al., Analysis of intact protein isoforms by mass spectrometry. *J Biol Chem*, 2011. 286(29): p. 25451-8.
21. Vellaichamy, A., et al., Size-sorting combined with improved nanocapillary liquid chromatography-mass spectrometry for identification of intact proteins up to 80 kDa. *Anal Chem*, 2010. 82(4): p. 1234-44.
22. Fornelli, L., et al., Top-down analysis of 30-80 kDa proteins by electron transfer dissociation time-of-flight mass spectrometry. *Anal Bioanal Chem*, 2013. 405(26): p. 8505-14.
23. Fornelli, L., et al., Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol Cell Proteomics*, 2012. 11(12): p. 1758-67.
24. Henry, K.D., et al., Fourier-transform mass spectrometry of large molecules by electrospray ionization. *Proc Natl Acad Sci U S A*, 1989. 86(23): p. 9075-8.
25. Loo, J.A., C.G. Edmonds, and R.D. Smith, Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. *Science*, 1990. 248(4952): p. 201-4.
26. Feng, R., Konishi, Y, *Anal. Chem*, 1993. 65: p. 645-649.
27. Reid, G.E. and S.A. McLuckey, 'Top down' protein characterization via tandem mass spectrometry. *J Mass Spectrom*, 2002. 37(7): p. 663-75.
28. Feng, R. and Y. Konishi, Stepwise refolding of Acid-denatured myoglobin: Evidence from electrospray mass spectrometry. *J Am Soc Mass Spectrom*, 1993. 4(8): p. 638-45.
29. Senko, M.W., S.C. Beu, and F.W. McLafferty, High-resolution tandem mass spectrometry of carbonic anhydrase. *Anal Chem*, 1994. 66(3): p. 415-8.
30. Reiber, D.C., T.A. Grover, and R.S. Brown, Identifying proteins using matrix-assisted laser desorption/ionization in-source fragmentation data combined with database searching. *Anal Chem*, 1998. 70(4): p. 673-83.
31. Nemeth-Cawley, J.F., B.S. Tangarone, and J.C. Rouse, "Top Down" characterization is a complementary technique to peptide sequencing for identifying protein species in complex mixtures. *J Proteome Res*, 2003. 2(5): p. 495-505.
32. Bonet, C., et al., The GAGA protein of Drosophila is phosphorylated by CK2. *J Mol Biol*, 2005. 351(3): p.



33. Senko, M.W. and F.W. McLafferty, Mass spectrometry of macromolecules: has its time now come? *Annu Rev Biophys Biomol Struct*, 1994. 23: p. 763-85.
34. Kelleher, N.L., et al., Thiaminase I (42 kDa) heterogeneity, sequence refinement, and active site location from high-resolution tandem mass spectrometry. *J Am Soc Mass Spectrom*, 1995. 6(10): p. 981-4.
35. McLafferty, F.W., High-resolution tandem FT mass spectrometry above 10 kDa. *Acc. Chem. Res.*, 1994. 27: p. 379-386.
36. Zubarev, R.A., Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol*, 2004. 15(1): p. 12-6.
37. Zubarev, R.A., et al., Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem*, 2000. 72(3): p. 563-73.
38. Zubarev, R.A. and A. Makarov, Orbitrap mass spectrometry. *Anal Chem*, 2013. 85(11): p. 5288-96.
39. Compton, P.D., et al., On the scalability and requirements of whole protein mass spectrometry. *Anal Chem*, 2011. 83(17): p. 6868-74.
40. Zhou, H., et al., Advancements in top-down proteomics. *Anal Chem*, 2012. 84(2): p. 720-34.
41. Ahlf, D.R., P.M. Thomas, and N.L. Kelleher, Developing top down proteomics to maximize proteome and sequence coverage from cells and tissues. *Curr Opin Chem Biol*, 2013. 17(5): p. 787-94.
42. Armirotti, A. and G. Damonte, Achievements and perspectives of top-down proteomics. *Proteomics*, 2010. 10(20): p. 3566-76.
43. Zubarev, R.A., The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics*, 2013. 13(5): p. 723-6.
44. Capriotti, A.L., et al., Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics. *J Chromatogr A*, 2011. 1218(49): p. 8760-76.
45. Hiraoaka, K., Kudaka, I., Electrospray interface for liquid chromatography-mass spectrometry. *Rapid Commun Mass Spectrom*, 1990. 4: p. 519-526.
46. Alpert, A.J., Hydrophobic interaction chromatography of peptides as an alternative to reversed-phase chromatography. *J Chromatogr*, 1988. 444: p. 269-74.
47. Alpert, A.J. and P.C. Andrews, Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica. *J Chromatogr*, 1988. 443: p. 85-96.
48. Garcia, B.A., et al., Pervasive combinatorial modification of histone H3 in human cells. *Nat Methods*, 2007. 4(6): p. 487-9.
49. Pesavento, J.J., et al., Mild performic acid oxidation enhances chromatographic and top down mass spectrometric analyses of histones. *Mol Cell Proteomics*, 2007. 6(9): p. 1510-26.
50. Pesavento, J.J., C.A. Mizzzen, and N.L. Kelleher, Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: human histone H4. *Anal Chem*, 2006. 78(13): p. 4271-80.
51. Xiu, L., et al., Effective protein separation by coupling hydrophobic interaction and reverse phase chromatography for top-down proteomics. *Anal Chem*, 2014. 86(15): p. 7899-906.
52. Sharma, S., et al., Proteomic profiling of intact proteins using WAX-RPLC 2-D separations and FTICR mass spectrometry. *J Proteome Res*, 2007. 6(2): p. 602-10.
53. Pesavento, J.J., et al., Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. *J Biol Chem*, 2008. 283(22): p. 14927-37.
54. Tian, Z., et al., Two-dimensional liquid chromatography system for online top-down mass spectrometry. *Proteomics*, 2010. 10(20): p. 3610-20.
55. Kellie, J.F., et al., The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput. *Mol Biosyst*, 2010. 6(9): p. 1532-9.
56. Valeja, S.G., et al., Three dimensional liquid chromatography coupling ion exchange chromatography/hydrophobic interaction chromatography/reverse phase chromatography for effective protein separation in top-down proteomics. *Anal Chem*, 2015. 87(10): p. 5363-5371.
57. Lubman, D.M., et al., Two-dimensional liquid separations-mass mapping of proteins from human cancer cell lysates. *J Chromatogr B Analyt Technol Biomed Life Sci*, 2002. 782(1-2): p. 183-96.
58. Patrie, S.M., et al., Top down mass spectrometry of < 60-kDa proteins from *Methanosarcina acetivorans* using quadrupole FRMS with automated octopole collisionally activated dissociation. *Mol Cell Proteomics*, 2006. 5(1): p. 14-25.
59. Whitelegge, J., Tandem mass spectrometry of integral membrane proteins for top-down proteomics. *Trac-Trends Anal. Chem.*, 2005. 24: p. 576-582.
60. Tran, J.C. and A.A. Doucette, Rapid and effective focusing in a carrier ampholyte solution isoelectric focusing system: a proteome prefractionation tool. *J Proteome Res*, 2008. 7(4): p. 1761-6.
61. Ouvry-Patat, S.A., et al., Free-flow electrophoresis for top-down proteomics by Fourier transform ion cyclotron resonance mass spectrometry. *Proteomics*, 2008. 8(14): p. 2798-808.
62. Rabilloud, T., Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics*, 2002. 2(1): p. 3-10.
63. Laemmli, U.K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 1970. 227(5259): p. 680-5.
64. Gygi, S.P., et al., Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A*, 2000. 97(17): p. 9390-5.

65. Meng, F., et al., Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. *Anal Chem*, 2002. 74(13): p. 2923-9.
66. Tran, J.C. and A.A. Doucette, Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. *Anal Chem*, 2008. 80(5): p. 1568-73.
67. Lee, J.E., et al., A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome. *J Am Soc Mass Spectrom*, 2009. 20(12): p. 2183-91.
68. Tran, J.C. and A.A. Doucette, Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Anal Chem*, 2009. 81(15): p. 6201-9.
69. Li, Y., et al., Optimizing capillary electrophoresis for top-down proteomics of 30-80 kDa proteins. *Proteomics*, 2014. 14(10): p. 1158-64.
70. Catherman, A.D., et al., Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Mol Cell Proteomics*, 2013. 12(12): p. 3465-73.
71. Tran, J.C., et al., Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*, 2011. 480(7376): p. 254-8.
72. Catherman, A.D., et al., Top down proteomics of human membrane proteins from enriched mitochondrial fractions. *Anal Chem*, 2013. 85(3): p. 1880-8.
73. Ntai, I., et al., Applying label-free quantitation to top down proteomics. *Anal Chem*, 2014. 86(10): p. 4961-8.
74. Kim, K.H., et al., Online matrix removal platform for coupling gel-based separations to whole protein electrospray ionization mass spectrometry. *J Proteome Res*, 2015. 14(5): p. 2199-206.
75. Lim, P., R. North, and G. Vigh, Rapid isoelectric trapping in a micropreparative-scale multicompartment electrolyzer. *Electrophoresis*, 2007. 28(12): p. 1851-9.
76. Pai, P.J., et al., Efficient electrophoretic method to remove neutral additives from protein solutions followed by mass spectrometry analysis. *Anal Chem*, 2011. 83(7): p. 2814-8.
77. Pirmoradian, M., J. Astorga-Wells, and R.A. Zubarev, Multijunction Capillary Isoelectric Focusing Device Combined with Online Membrane-Assisted Buffer Exchanger Enables Isoelectric Point Fractionation of Intact Human Plasma Proteins for Biomarker Discovery. *Anal Chem*, 2015. 87(23): p. 11840-6.
78. Guevremont, R., High-field asymmetric waveform ion mobility spectrometry: a new tool for mass spectrometry. *J Chromatogr A*, 2004. 1058(1-2): p. 3-19.
79. Thalassinos K., S.S., Jennings KR, Scrivens JH, Giles K, Wildgoose J, Hoyes J, Bateman RH, Bowers MT., Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer. *Int. J. Mass Spectrom.*, 2004. 236: p. 55-63.
80. Zinnel, N.F., P.J. Pai, and D.H. Russell, Ion mobility-mass spectrometry (IM-MS) for top-down proteomics: increased dynamic range affords increased sequence coverage. *Anal Chem*, 2012. 84(7): p. 3390-7.
81. Zabrouskov, V., et al., Stepwise deamidation of ribonuclease A at five sites determined by top down mass spectrometry. *Biochemistry*, 2006. 45(3): p. 987-92.
82. Marshall, A.G., C.L. Hendrickson, and G.S. Jackson, Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev*, 1998. 17(1): p. 1-35.
83. Patrie, S.M., et al., 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(7): p. 1099-108.
84. Tolmachev, A.V., et al., FT-ICR MS optimization for the analysis of intact proteins. *Int J Mass Spectrom*, 2009. 281(1-3): p. 32-38.
85. Roth, M.J., et al., Precise and parallel characterization of coding polymorphisms, alternative splicing, and modifications in human proteins by mass spectrometry. *Mol Cell Proteomics*, 2005. 4(7): p. 1002-8.
86. Parks, B.A., et al., Top-down proteomics on a chromatographic time scale using linear ion trap fourier transform hybrid mass spectrometers. *Anal Chem*, 2007. 79(21): p. 7984-91.
87. Ryan, C.M., et al., Post-translational modifications of integral membrane proteins resolved by top-down Fourier transform mass spectrometry with collisionally activated dissociation. *Mol Cell Proteomics*, 2010. 9(5): p. 791-803.
88. Bonet-Costa, C., et al., Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of *Drosophila melanogaster* linker histone H1. *J Proteomics*, 2012. 75(13): p. 4124-38.
89. Ferguson, J.T., et al., Top-down proteomics reveals novel protein forms expressed in Methanosarcina acetivorans. *J Am Soc Mass Spectrom*, 2009. 20(9): p. 1743-50.
90. Roth, M.J., et al., "Proteotyping": population proteomics of human leukocytes using top down mass spectrometry. *Anal Chem*, 2008. 80(8): p. 2857-66.
91. Schaub, T.M., et al., High-performance mass spectrometry: Fourier transform ion cyclotron resonance at 14.5 Tesla. *Anal Chem*, 2008. 80(11): p. 3985-90.
92. Hu, Q., et al., The Orbitrap: a new mass spectrometer. *J Mass Spectrom*, 2005. 40(4): p. 430-43.
93. Makarov, A., et al., Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem*, 2006. 78(7): p. 2113-20.
94. Michalski, A., et al., Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics*, 2011. 10(9): p. M111 011015.
95. Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem*, 2000. 72(6): p. 1156-62.
96. Macek, B., et al., Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. *Mol Cell Proteomics*, 2006. 5(5): p. 949-58.



97. Siuti, N., et al., Gene-specific characterization of human histone H2B by electron capture dissociation. *J Proteome Res*, 2006. 5(2): p. 233-9.
98. Boyne, M.T., 2nd, et al., Precise characterization of human histones in the H2A gene family by top down mass spectrometry. *J Proteome Res*, 2006. 5(2): p. 248-53.
99. Thomas, C.E., N.L. Kelleher, and C.A. Mizzen, Mass spectrometric characterization of human histone H3: a bird's eye view. *J Proteome Res*, 2006. 5(2): p. 240-7.
100. Ahlf, D.R., et al., Evaluation of the compact high-field orbitrap for top-down proteomics of human cells. *J Proteome Res*, 2012. 11(8): p. 4308-14.
101. Wynne, C., et al., Top-down identification of protein biomarkers in bacteria with unsequenced genomes. *Anal Chem*, 2009. 81(23): p. 9633-42.
102. Wynne, C., N.J. Edwards, and C. Fenselau, Phyloproteomic classification of unsequenced organisms by top-down identification of bacterial proteins using capLC-MS/MS on an Orbitrap. *Proteomics*, 2010. 10(20): p. 3631-43.
103. Mazur, M.T., et al., Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *Proc Natl Acad Sci U S A*, 2010. 107(17): p. 7728-33.
104. Cheon, D.H., et al., Comprehensive Analysis of Low-Molecular-Weight Human Plasma Proteome Using Top-Down Mass Spectrometry. *J Proteome Res*, 2016. 15(1): p. 229-44.
105. Petras, D., et al., Venom Proteomics of Indonesian King Cobra, *Ophiophagus hannah*: Integrating Top-Down and Bottom-Up Approaches. *J Proteome Res*, 2015. 14(6): p. 2539-56.
106. Bondarenko, P.V., et al., 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 Spectrom*, 2009. 20(8): p. 1415-24.
107. Zhang, J., H. Liu, and V. Katta, Structural characterization of intact antibodies by high-resolution LTQ Orbitrap mass spectrometry. *J Mass Spectrom*, 2010. 45(1): p. 112-20.
108. Edwards, R.L., et al., Top-down proteomics and direct surface sampling of neonatal dried blood spots: diagnosis of unknown hemoglobin variants. *J Am Soc Mass Spectrom*, 2012. 23(11): p. 1921-30.
109. Senko, M.W., et al., Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Anal Chem*, 2013. 85(24): p. 11710-4.
110. Brunner, A.M., et al., Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteoform characterization. *Anal Chem*, 2015. 87(8): p. 4152-8.
111. Zheng, Y., et al., Unabridged Analysis of Human Histone H3 by Differential Top-Down Mass Spectrometry Reveals Hypermethylated Proteoforms from MMSET/NSD2 Overexpression. *Mol Cell Proteomics*, 2016. 15(3): p. 776-90.
112. Wiley WC, M.I., Time-of-flight mass spectrometer with improved resolution. *Rev Sci Instrum*, 1955. 26: p. 1150-1157.
113. Dawson JHJ, G.M., Orthogonal acceleration time-of-flight mass spectrometer. *Rapid Commun in Mass Spectrom*, 1989. 3(5): p. 155-159.
114. Coles J, G.M., Orthogonal acceleration – a new direction for time-of-flight mass spectrometry: fast, sensitive mass analysis for continuous ion sources. *Trends in Analytical Chemistry*, 1993. 12(5): p. 203-213.
115. Chamot-Rooke, J., et al., Posttranslational modification of pili upon cell contact triggers *N. meningitidis* dissemination. *Science*, 2011. 331(6018): p. 778-82.
116. Marcoux, J., et al., Native mass spectrometry and ion mobility characterization of trastuzumab emtansine, a lysine-linked antibody drug conjugate. *Protein Sci*, 2015. 24(8): p. 1210-23.
117. Huang, L., et al., Impact of variable domain glycosylation on antibody clearance: an LC/MS characterization. *Anal Biochem*, 2006. 349(2): p. 197-207.
118. Ayoub, D., et al., Correct primary structure assessment and extensive glyco-profiling of cetuximab by a combination of intact, middle-up, middle-down and bottom-up ESI and MALDI mass spectrometry techniques. *MAbs*, 2013. 5(5): p. 699-710.
119. McLafferty, F.W., Bryce T. A., Metastable-ion characteristics: characterization of isomeric molecules. *Chem. Commun*, 1967: p. 1215-1217.
120. Hunt, D., Coon JJ, Syka JEP, Marto JA. , Electrom transfer dissociation for biopolymer sequence mass spectrometric analysis. 2009-05-19. US patent 7534622.
121. Wu, S.L., et al., Mass spectrometric determination of disulfide linkages in recombinant therapeutic proteins using online LC-MS with electron-transfer dissociation. *Anal Chem*, 2009. 81(1): p. 112-22.
122. Levy, M.J., A.C. Gucinski, and M.T. Boyne, 2nd, Primary Sequence Confirmation of a Protein Therapeutic Using Top Down MS/MS and MS(3). *Anal Chem*, 2015. 87(14): p. 6995-9.
123. Riley, N.M., M.S. Westphall, and J.J. Coon, Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal Chem*, 2015. 87(14): p. 7109-16.
124. Riley, N.M., et al., The Negative Mode Proteome with Activated Ion Negative Electron Transfer Dissociation (AI-NETD). *Mol Cell Proteomics*, 2015. 14(10): p. 2644-60.
125. Cannon, J.R., et al., Ultraviolet photodissociation for characterization of whole proteins on a chromatographic time scale. *Anal Chem*, 2014. 86(4): p. 2185-92.
126. Cleland, T.P., et al., High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. *J Proteome Res*, 2017. 16(5): p. 2072-2079.
127. Warnke, S., G. von Helden, and K. Pagel, Analyzing the higher order structure of proteins with conformer-selective ultraviolet photodissociation. *Proteomics*, 2015. 15(16): p. 2804-12.

128. Brodbelt, J.S., Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem Soc Rev*, 2014. 43(8): p. 2757-83.
129. Dang, X. and N.L. Young, Ultraviolet photodissociation enhances top-down mass spectrometry as demonstrated on green fluorescent protein variants. *Proteomics*, 2014. 14(10): p. 1128-9.
130. Shaw, J.B., et al., Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J Am Chem Soc*, 2013. 135(34): p. 12646-51.
131. Wisniewski, J.R., et al., Universal sample preparation method for proteome analysis. *Nat Methods*, 2009. 6(5): p. 359-62.
132. Bunger, M.K., et al., Automated proteomics of *E. coli* via top-down electron-transfer dissociation mass spectrometry. *Anal Chem*, 2008. 80(5): p. 1459-67.
133. Durbin, K.R., et al., Quantitation and Identification of Thousands of Human Proteoforms below 30 kDa. *J Proteome Res*, 2016. 15(3): p. 976-82.
134. Anderson, J.C., et al., Decreased abundance of type III secretion system-inducing signals in *Arabidopsis* mkp1 enhances resistance against *Pseudomonas syringae*. *Proc Natl Acad Sci U S A*, 2014. 111(18): p. 6846-51.
135. Vorontsov, E.A., et al., Abundant Lysine Methylation and N-Terminal Acetylation in *Sulfolobus islandicus* Revealed by Bottom-Up and Top-Down Proteomics. *Mol Cell Proteomics*, 2016. 15(11): p. 3388-3404.
136. Shaw, J.B., et al., 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer Greatly Expands Mass Spectrometry Toolbox. *J Am Soc Mass Spectrom*, 2016. 27(12): p. 1929-1936.
137. Zabrouskov, V., et al., Unraveling molecular complexity of phosphorylated human cardiac troponin I by top down electron capture dissociation/electron transfer dissociation mass spectrometry. *Mol Cell Proteomics*, 2008. 7(10): p. 1838-49.
138. Ayaz-Guner, S., et al., In vivo phosphorylation site mapping in mouse cardiac troponin I by high resolution top-down electron capture dissociation mass spectrometry: Ser22/23 are the only sites basally phosphorylated. *Biochemistry*, 2009. 48(34): p. 8161-70.
139. Dong, X., et al., Augmented phosphorylation of cardiac troponin I in hypertensive heart failure. *J Biol Chem*, 2012. 287(2): p. 848-57.
140. Zhang, J., et al., Deciphering modifications in swine cardiac troponin I by top-down high-resolution tandem mass spectrometry. *J Am Soc Mass Spectrom*, 2010. 21(6): p. 940-8.
141. Zhang, J., et al., Top-down quantitative proteomics identified phosphorylation of cardiac troponin I as a candidate biomarker for chronic heart failure. *J Proteome Res*, 2011. 10(9): p. 4054-65.
142. Toby, T.K., L. Fornelli, and N.L. Kelleher, Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu Rev Anal Chem (Palo Alto Calif)*, 2016. 9(1): p. 499-519.
143. Du, Y., et al., Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. *Anal Chem*, 2006. 78(3): p. 686-94.
144. Collier, T.S., et al., Top-down identification and quantification of stable isotope labeled proteins from *Aspergillus flavus* using online nano-flow reversed-phase liquid chromatography coupled to a LTQ-FTICR mass spectrometer. *Anal Chem*, 2008. 80(13): p. 4994-5001.
145. Collier, T.S., et al., Quantitative top-down proteomics of SILAC labeled human embryonic stem cells. *J Am Soc Mass Spectrom*, 2010. 21(6): p. 879-89.
146. Hung, C.W. and A. Tholey, Tandem mass tag protein labeling for top-down identification and quantification. *Anal Chem*, 2012. 84(1): p. 161-70.
147. Wiener, M.C., et al., Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Anal Chem*, 2004. 76(20): p. 6085-96.
148. Gomez, S.M., et al., The chloroplast grana proteome defined by intact mass measurements from liquid chromatography mass spectrometry. *Mol Cell Proteomics*, 2002. 1(1): p. 46-59.
149. Ryan, C.M., et al., Confident assignment of intact mass tags to human salivary cystatins using top-down Fourier-transform ion cyclotron resonance mass spectrometry. *J Am Soc Mass Spectrom*, 2010. 21(6): p. 908-17.
150. Castagnola, M., et al., Hypo-phosphorylation of salivary peptidome as a clue to the molecular pathogenesis of autism spectrum disorders. *J Proteome Res*, 2008. 7(12): p. 5327-32.
151. Wu, S., et al., Quantitative analysis of human salivary gland-derived intact proteome using top-down mass spectrometry. *Proteomics*, 2014. 14(10): p. 1211-22.
152. Heck, A.J., Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods*, 2008. 5(11): p. 927-33.
153. Lopez, A., et al., Analyzing slowly exchanging protein conformations by ion mobility mass spectrometry: study of the dynamic equilibrium of prolyl oligopeptidase. *J Mass Spectrom*, 2016. 51(7): p. 504-11.
154. Skinner, O.S., et al., An informatic framework for decoding protein complexes by top-down mass spectrometry. *Nat Methods*, 2016. 13(3): p. 237-40.
155. Melani, R.D., et al., Mapping Proteoforms and Protein Complexes From King Cobra Venom Using Both Denaturing and Native Top-down Proteomics. *Mol Cell Proteomics*, 2016. 15(7): p. 2423-34.
156. Rostom, A.A., et al., Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proc Natl Acad Sci U S A*, 2000. 97(10): p. 5185-90.
157. Loo, J.A., et al., Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex. *J Am Soc Mass Spectrom*, 2005. 16(7): p. 998-1008.
158. Snijder, J., et al., Studying 18 MDa virus assemblies with native mass spectrometry. *Angew Chem Int Ed Engl*, 2013. 52(14): p. 4020-3.
159. Snijder, J., et al., Insight into cyanobacterial circadian timing from structural details of the KaiB-KaiC interac-

tion. *Proc Natl Acad Sci U S A*, 2014. 111(4): p. 1379-84.

160. Ben-Nissan, G., et al., Triple-Stage Mass Spectrometry Unravels the Heterogeneity of an Endogenous Protein Complex. *Anal Chem*, 2017. 89(8): p. 4708-4715.

161. Gan, J., et al., Native Mass Spectrometry of Recombinant Proteins from Crude Cell Lysates. *Anal Chem*, 2017. 89(8): p. 4398-4404.

162. Dyachenko, A., et al., Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proc Natl Acad Sci U S A*, 2013. 110(18): p. 7235-9.

163. Muneeruddin, K., M. Nazzaro, and I.A. Kaltashov, Characterization of intact protein conjugates and biopharmaceuticals using ion-exchange chromatography with online detection by native electrospray ionization mass spectrometry and top-down tandem mass spectrometry. *Anal Chem*, 2015. 87(19): p. 10138-45.

164. Horn, D.M., R.A. Zubarev, and F.W. McLafferty, Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J Am Soc Mass Spectrom*, 2000. 11(4): p. 320-32.

165. Zabrouskov, V., et al., New and automated MSn approaches for top-down identification of modified proteins. *J Am Soc Mass Spectrom*, 2005. 16(12): p. 2027-38.

166. Liu, X., et al., Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Mol Cell Proteomics*, 2010. 9(12): p. 2772-82.

167. Frank, A.M., et al., Interpreting top-down mass spectra using spectral alignment. *Anal Chem*, 2008. 80(7): p. 2499-505.

168. Liu, X., et al., Protein identification using top-down. *Mol Cell Proteomics*, 2012. 11(6): p. M111 008524.

169. Cai, W., et al., MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics. *Mol Cell Proteomics*, 2016. 15(2): p. 703-14.

170. LeDuc, R.D., et al., ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res*, 2004. 32(Web Server issue): p. W340-5.

171. Fellers, R.T., Greer, J. B., Early, B. P., Yu, X., LeDuc, R. D., Kelleher, N. L., and Thomas, P. M., ProSight Lite: Graphical software to analyze top-down mass spectrometry data. *Proteomics*, 2014. 15(1253-1238).

172. Kou, Q., L. Xun, and X. Liu, TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics*, 2016. 32(22): p. 3495-3497.

173. Kou, Q., et al., Characterization of Proteoforms with Unknown Post-translational Modifications Using the MIScore. *J Proteome Res*, 2016. 15(8): p. 2422-32.

174. Melani, R.D., F.C.S. Nogueira, and G.B. Domont, It is time for top-down venomics. *J Venom Anim Toxins Incl Trop Dis*, 2017. 23: p. 44.

175. Pla, D., et al., Transcriptomics-guided bottom-up and top-down venomics of neonate and adult specimens of the arboreal rear-fanged Brown Treesnake, *Boiga irregularis*, from Guam. *J Proteomics*, 2018. 174: p. 71-84.

176. Chen, Y.C., et al., Effective top-down LC/MS+ method for assessing actin isoforms as a potential cardiac disease marker. *Anal Chem*, 2015. 87(16): p. 8399-8406.

177. Sinz, A., Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev*, 2006. 25(4): p. 663-82.

178. Novak, P., et al., A top-down approach to protein structure studies using chemical cross-linking and Fourier transform mass spectrometry. *Eur J Mass Spectrom (Chichester)*, 2003. 9(6): p. 623-31.

179. Gregorich, Z.R. and Y. Ge, Top-down proteomics in health and disease: challenges and opportunities. *Proteomics*, 2014. 14(10): p. 1195-210.

180. Zhang, H. and Y. Ge, Comprehensive analysis of protein modifications by top-down mass spectrometry. *Circ Cardiovasc Genet*, 2011. 4(6): p. 711.

181. Zhang, J., et al., Phosphorylation, but not alternative splicing or proteolytic degradation, is conserved in human and mouse cardiac troponin T. *Biochemistry*, 2011. 50(27): p. 6081-92.

182. Savaryn, J.P., et al., The emergence of top-down proteomics in clinical research. *Genome Med*, 2013. 5(6): p. 53.

183. Mao, P. and D. Wang, Top-down proteomics of a drop of blood for diabetes monitoring. *J Proteome Res*, 2014. 13(3): p. 1560-9.

184. Iavarone, F., et al., Characterization of salivary proteins of schizophrenic and bipolar disorder patients by top-down proteomics. *J Proteomics*, 2014. 103: p. 15-22.

185. Barnidge, D.R., et al., Using mass spectrometry to monitor monoclonal immunoglobulins in patients with a monoclonal gammopathy. *J Proteome Res*, 2014. 13(3): p. 1419-27.

186. Theberge, R., et al., Top-Down Analysis of Small Plasma Proteins Using an LTQ-Orbitrap. Potential for Mass Spectrometry-Based Clinical Assays for Transthyretin and Hemoglobin. *Int J Mass Spectrom*, 2011. 300(2-3): p. 130-142.

187. Vila-Rico, M., et al., Quantitative analysis of post-translational modifications in human serum transthyretin associated with familial amyloidotic polyneuropathy by targeted LC-MS and intact protein MS. *J Proteomics*, 2015. 127(Pt B): p. 234-46.

188. Acosta-Martin, A.E., et al., Quantitative mass spectrometry analysis of intact hemoglobin A2 by precursor ion isolation and detection. *Anal Chem*, 2013. 85(16): p. 7971-5.

189. Coelho Graca, D., et al., Electron transfer dissociation mass spectrometry of hemoglobin on clinical samples. *J Am Soc Mass Spectrom*, 2012. 23(10): p. 1750-6.

190. Kam, R.K., C.S. Ho, and M.H. Chan, Serum Insulin-like Growth Factor I Quantitation by Mass Spectrometry: Insights for Protein Quantitation with this Technology. *EJIFCC*, 2016. 27(4): p. 318-330.

191. Coelho Graca, D., et al., Identification of hemoglobin variants by top-down mass spectrometry using selected diagnostic product ions. *Anal Bioanal Chem*, 2015. 407(10): p. 2837-45.
192. Delcourt, V., et al., Combined Mass Spectrometry Imaging and Top-down Microproteomics Reveals Evidence of a Hidden Proteome in Ovarian Cancer. *EBioMedicine*, 2017. 21: p. 55-64.
193. Ait-Belkacem, R., et al., MALDI imaging and in-source decay for top-down characterization of glioblastoma. *Proteomics*, 2014. 14(10): p. 1290-301.
194. Sarsby, J., et al., Top-down and bottom-up identification of proteins by liquid extraction surface analysis mass spectrometry of healthy and diseased human liver tissue. *J Am Soc Mass Spectrom*, 2014. 25(11): p. 1953-61.
195. Ntai, I., et al., Integrated Bottom-Up and Top-Down Proteomics of Patient-Derived Breast Tumor Xenografts. *Mol Cell Proteomics*, 2016. 15(1): p. 45-56.
196. Hummel, M., T. Wigger, and J. Brockmeyer, Characterization of mustard 2S albumin allergens by bottom-up, middle-down, and top-down proteomics: a consensus set of isoforms of Sin a 1. *J Proteome Res*, 2015. 14(3): p. 1547-56.
197. Lagrain, B., et al., Identification of intact high molecular weight glutenin subunits from the wheat proteome using combined liquid chromatography-electrospray ionization mass spectrometry. *PLoS One*, 2013. 8(3): p. e58682.
198. Toby, T.K., et al., A comprehensive pipeline for translational top-down proteomics from a single blood draw. *Nat Protoc*, 2019. 14(1): p. 119-152.
199. Vialaret, J., et al., Identification of multiple proteoforms biomarkers on clinical samples by routine Top-Down approaches. *Data Brief*, 2018. 18: p. 1013-1021.
200. Wang, Z., et al., Top-down Mass Spectrometry Analysis of Human Serum Autoantibody Antigen-Binding Fragments. *Sci Rep*, 2019. 9(1): p. 2345.
201. De Petris, L., et al., Tumor expression of S100A6 correlates with survival of patients with stage I non-small-cell lung cancer. *Lung Cancer*, 2009. 63(3): p. 410-7.
202. Desiderio, C., et al., Cerebrospinal fluid top-down proteomics evidenced the potential biomarker role of LVV- and VV-hemorphin-7 in posterior cranial fossa pediatric brain tumors. *Proteomics*, 2012. 12(13): p. 2158-66.
203. Cabras, T., et al., Significant modifications of the salivary proteome potentially associated with complications of Down syndrome revealed by top-down proteomics. *Mol Cell Proteomics*, 2013. 12(7): p. 1844-52.
204. Laouirem, S., et al., Progression from cirrhosis to cancer is associated with early ubiquitin post-translational modifications: identification of new biomarkers of cirrhosis at risk of malignancy. *J Pathol*, 2014. 234(4): p. 452-63.
205. Ye, H., et al., Top-down proteomics with mass spectrometry imaging: a pilot study towards discovery of biomarkers for neurodevelopmental disorders. *PLoS One*, 2014. 9(4): p. e92831.
206. Kellie, J.F., et al., Quantitative measurement of intact alpha-synuclein proteoforms from post-mortem control and Parkinson's disease brain tissue by intact protein mass spectrometry. *Sci Rep*, 2014. 4: p. 5797.
207. Martelli, C., et al., Integrated proteomic platforms for the comparative characterization of medulloblastoma and pilocytic astrocytoma pediatric brain tumors: a preliminary study. *Mol Biosyst*, 2015. 11(6): p. 1668-83.
208. Fornelli, L., et al., Top-down proteomics: Where we are, where we are going? *J Proteomics*, 2018. 175: p. 3-4.

Fragmentación de Péptidos y Proteínas por Fotodisociación

Julia Bauzá-Martínez^{1,2}, Eliandre de Oliveira³

¹Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

²Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

³Plataforma de Proteómica - Parc Científic de Barcelona. C/Baldiri Reixac, 10-12, 08028, Barcelona, España

1. Introducción.

La espectrometría de masas (EM) viene experimentando un constante desarrollo debido a su capacidad de analizar diversos tipos de moléculas y, en particular, debido a su aplicación cada vez más acentuada en el campo de las ciencias de la vida para el análisis de biomoléculas tales como son las proteínas y péptidos.

La EM es una técnica esencial en el campo de la proteómica y el avance de este campo se debe en gran parte a la constante evolución de los espectrómetros de masas. En los últimos años estos instrumentos han experimentado un aumento en sensibilidad, en la velocidad de adquisición de los datos y en la resolución, con el consecuente incremento de exactitud en la detección de las masas moleculares. Estos avances, junto con el constante desarrollo en los sistemas de ionización y de fragmentación de iones, han conllevado al perfeccionamiento y ampliación de aplicaciones de EM en el campo de la proteómica. El desarrollo en los métodos de fragmentación, en particular la fragmentación por fotodisociación será discutida en el presente capítulo.

La EM, además de facilitar la determinación de la masa molecular de moléculas previamente ionizadas mediante la detección y medición de la relación masa/carga (m/z), puede facilitar información estructural cuando se realizan análisis de masas en tandem. En este último caso, tienen lugar dos eventos: en un primer análisis (MS) se mide la masa molecular del péptido o proteína (ion padre), y una vez este ion está disociado, en un segundo evento (MS/MS) se miden las masas moleculares de los fragmentos formados a partir del ion padre (iones hijos), para obtener información relativa a su secuencia de aminoácidos. Esta información estructural puede ser utilizada para diferentes estudios que van desde la identificación¹ y cuantificación de proteínas presentes en bases de datos hasta estudios de secuenciación de novo², además de caracterización de modificaciones postraduccionales (PTMs)³ o de interacciones proteína-proteína⁴ y proteína-ligando. De hecho, los análisis estructurales de péptidos y proteínas mediante espectrometría de masas dependen fundamentalmente de la capacidad de ionización y de la calidad de la fragmentación de dichos iones en estado



gaseoso. La eficiencia de la activación y fragmentación se puede evaluar a través del espectro de MS/MS. Cuanto mayor la eficiencia de la fragmentación, más completo será el espectro de MS/MS, lo que permitirá una asignación completa de una determinada secuencia peptídica.

A su vez, la eficiencia de la ionización y de fragmentación dependen de la estructura primaria de los péptidos y proteínas. El tamaño de la cadena peptídica, así como su composición aminoacídica, son factores determinantes en la eficiencia de la fragmentación. En este sentido, cabe destacar que los estudios de proteómica pueden ser divididos en tres grandes grupos: i) "Bottom-up", ii) "Middle-down" y iii) "Top-down".

i) **"Bottom-up"**, constituye la mayor parte de los estudios y experimentos de proteómica. Esta técnica consiste en digerir proteínas enzimáticamente, y analizar los péptidos obtenidos en la proteólisis por cromatografía líquida acoplada a la espectrometría de masas (LC-MS/MS). La enzima más utilizada es la tripsina debido a su elevada especificidad. La hidrólisis ocurre en el extremo carboxiterminal de los residuos de arginina (R) y lisina (K), que están presentes en la mayoría de proteínas, facilitando la obtención de péptidos con una medida adecuada para el análisis por LC-MS/MS. Además, los péptidos tripticos que son aquellos formados por la hidrólisis con tripsina, al tener un residuo de R o K en el extremo C-terminal facilitan la obtención de péptidos con al menos dos cargas positivas a pH ácido, lo que favorece el análisis por LC-MS/MS. Con esta técnica se analizan tanto muestras sencillas, de proteínas separadas previamente por técnicas de electroforesis (SDS-PAGE o bidimensional) o de afinidad; como muestras complejas o muy complejas, como son los extractos de proteínas de tejidos, o biofluidos o de cultivos celulares.

Esta es la técnica empleada para la mayoría de aplicaciones de proteómica mediante espectrometría de masas. Es la técnica más eficiente para los análisis de identificación de proteínas en muestras complejas, identificaciones masivas o a gran escala (shot-gun), donde la digestión con tripsina genera miles de péptidos que son fragmentados y posteriormente identificados mediante la comparación con base de datos. También es la técnica más utilizada para los estudios de cuantificación del nivel de expresión de proteínas, ya sea mediante el marcaje con isótopos estables (como son el "Stable Isotope Labeling by/with Amino acids in Cell culture", SILAC; Isobaric tags for relative and absolute quantitation, iTRAQ; Tandem mass tag, TMT; Inductively coupled plasma, ICP; etc) como utilizando métodos sin marcaje o "label-free". Sin embargo, aunque la técnica de "bottom-up" sea la más utilizada, los resultados obtenidos en los estudios de PTMs o de interacciones proteína-proteína muchas veces son incompletos o poco informativos. Esto se debe a que muchos de los péptidos provenientes de la digestión enzimática no son observados en el análisis de LC-MS/MS. Ya sea por el tamaño de los péptidos formados en la hidrólisis o por la hidrofobicidad de los mismos, la detección de todos los péptidos de una proteína es muy poco frecuente. Este hecho supone una gran dificultad para los análisis de secuenciación de isoformas, que requieren una máxima cobertura de la secuencia de las proteínas en estudio, o la observación y detección de todos o casi todos los péptidos formados en la digestión enzimática.

Con los sistemas de LC-MS/MS disponibles hasta los últimos años, los estudios realizados con péptidos presentaban rendimiento muy superior a los estudios realizados con proteínas intactas. Sin embargo, esta tendencia está cambiando progresivamente

y es cada vez mayor el número de publicaciones de proteómica en que se realiza el análisis de péptidos de alto peso molecular y proteínas intactas, presentados en los dos apartados a continuación.

ii) Middle-down, consiste en digerir las proteínas parcialmente y analizar péptidos de mayor tamaño por LC-MS/MS. En este caso, se utilizan enzimas menos convencionales⁵ para obtener péptidos de mayor tamaño. Con esta técnica el número de péptidos formados es menor y, por tanto, el análisis de LC-MS/MS es menos complejo. Sin embargo, los péptidos de mayor tamaño tienen cargas más altas y los métodos de fragmentación utilizados en la técnica de "middle-down" son aquellos más eficientes para la fragmentación de péptidos con cargas altas ($Z > +3$). Por otro lado, los péptidos de mayor tamaño son portadores de más información, tanto aquellas modificaciones relacionadas con los cambios debido a PTMs como las modificaciones a nivel de estructura primaria, relativas a las diferentes isoformas de una determinada proteína.

iii) Top-down, consiste en el análisis por LC-MS/MS de proteínas intactas, sin digestión enzimática previa. Al facilitar el análisis de proteínas intactas, esta técnica es la que permite realizar una caracterización más precisa y completa de las posibles modificaciones de una proteína. Recientemente, se ha designado el término "proteoforma"⁶ para describir el conjunto de formas moleculares en que se puede encontrar una proteína producto de un solo gen, incluyendo variaciones genéticas, de transcripción o modificaciones postraduccionales. No obstante, pocos grupos están llevando a cabo análisis proteómicos utilizando la técnica "top-down". Esto es debido en parte a que los sistemas de espectrometría de masas convencionales, que son óptimos para los análisis de tipo "bottom-up", no están optimizados y no son eficientes para el análisis de proteínas intactas. La fragmentación de proteínas intactas requiere de mayor energía y los métodos más adecuados serán discutidos en este capítulo.

2. Métodos de fragmentación.

En líneas generales, la fragmentación de un péptido tiene lugar por la ruptura de la cadena peptídica principal, generando dos fragmentos: un N- y un C-terminal. La

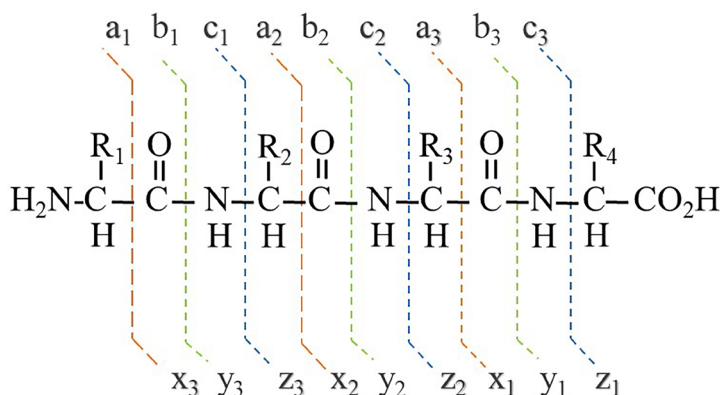


Figura 1: Patrón de los iones hijos formados a partir de la fragmentación de la cadena peptídica.



nomenclatura de los fragmentos generados o iones hijos fue propuesta por Roepstorff & Fohlman⁷ en 1984 (Figura 1).

La técnica de fragmentación más comúnmente utilizada en los espectrómetros de masas es la denominada **disociación inducida por colisión (CID)**⁸. Esta es una técnica de disociación de baja energía, en la cual los iones son activados al colisionar con moléculas de un gas inerte (Ar, He o N₂) que es introducido en la cámara de colisión. Esta técnica es del tipo de activación "slow-heating"⁹ y la fragmentación ocurre por la acumulación de energía interna proveniente de la energía cinética derivada de las múltiples colisiones. El patrón de fragmentación corresponde a la ruptura del enlace más débil de la cadena peptídica, el enlace amida, generando iones "b" e "y" (Fig. 1). Esto se debe a que en los iones peptídicos activados ocurre la relocalización de los protones hacia el nitrógeno amida o el oxígeno carboxílico del enlace peptídico, lo que debilita el enlace amida facilitando su ruptura. Detalles sobre el mecanismo molecular de la fragmentación por CID se encuentran muy bien descritos en el Manual de Proteómica Vol. 1 (PM Acedo y colaboradores. Mecanismos de fragmentación de iones peptídicos mediante disociación inducida por colisión (CID). Manual de Proteómica capítulo 17, 395-407). Aunque sea el método más ampliamente utilizado en los análisis proteómicos, la fragmentación por CID es dependiente de la secuencia de aminoácidos, la estructura primaria, así como del estado de cargas de los iones padre. Así mismo, esta técnica de fragmentación presenta menor eficiencia cuando los iones tienen cargas elevadas, como ocurre en la ionización por electrospray (ESI) de péptidos de mayor tamaño o proteínas analizados con las técnicas de "middle- o top-down" que forman iones con múltiples cargas positivas. Además, cuando se realiza la fragmentación CID en analizadores del tipo trampa iónica, los fragmentos de masas bajas se pierden o no se observan. Esto es debido a que el alto voltaje de radio frecuencia (RF) requerido para que la disociación de los iones padres tenga lugar es incompatible con los valores de voltaje necesarios para atrapar los fragmentos o iones de baja masa molecular formados. Esto genera espectros de MS/MS con poca información en las zonas de masas moleculares bajas. En general, masas moleculares menores que aproximadamente 28% de la masa del ion padre no son observadas. Este problema quedó solucionado con la introducción de la técnica de **disociación inducida por colisión de alta energía (HCD)**¹⁰. El mecanismo de fragmentación de los iones de esta técnica es idéntico al de CID, pero se realiza en el interior de un octapolo. Este es un tipo de analizador de masas que permite seleccionar y retener iones en todo el rango de masas moleculares, permitiendo así observar aquellos fragmentos de baja masa molecular. Los iones hijos son luego acelerados y almacenados en la C-Trap para su consecuente análisis en el analizador de masas de tipo Orbitrap. Los espectros de MS/MS resultantes de una fragmentación HCD suelen contener iones "b" e "y", al igual que en CID, aunque los espectros MS/MS de ésta última son ricos en iones de baja masa molecular, ayudando así a obtener una mayor cobertura de secuencia y siendo preferida para estudios de secuenciación de novo.

La necesidad de métodos de fragmentación complementarios a los de disociación por colisión, con mayor eficiencia para fragmentar moléculas que necesitan de mayor energía o capacidad para acceder a diferentes patrones de fragmentación, ha dado lugar al desarrollo de métodos de fragmentación basados en la interacción con electrones, **disociación por captura de electrones (ECD)**¹¹ y **disociación por transferencia de electrones (ETD)**¹² y a métodos de fragmentación basados en la interacción

con fotones, conocido como **fotodisociación (PD)**¹³. El método de disociación por captura de electrones se basa en la interacción entre iones con varias cargas positivas (múltiple-cargados) y electrones de baja energía (< 1 eV). De la interacción se forman especies radicales que rápidamente resultan en la ruptura de la cadena peptídica. En la disociación por ETD, la interacción ocurre entre los iones múltiple-cargados positivamente con aniones radicales de antraceno o fluoranteno, los cuales son excelentes donadores de electrones. A diferencia de la fragmentación por CID/HCD que ocurre principalmente en CO-N, la fragmentación por ECD y ETD ocurre preferentemente entre los enlaces amino-carbono alfa (N-C α), formando fragmentos o iones hijos del tipo "c" y "z" (Figura 1) y dejando la cadena lateral del aminoácido intacta. Por esta razón, en la fragmentación por ETD, la mayoría de las modificaciones lábiles son conservadas, lo que hace que esta técnica sea muy atractiva para los análisis de PTMs. Sin embargo, la eficacia de las técnicas de fragmentación anteriormente comentadas es altamente dependiente del estado de cargas de los iones analizados.

Finalmente, la fotodisociación es un método de activación iónica que utiliza fotones de luz para facilitar la disociación o ruptura de una molécula. Al interactuar una población de iones en fase gaseosa con fotones, se produce un aumento de la energía interna de los iones que resultará en la desestabilización de los enlaces moleculares y la consecuente fragmentación de la molécula. Las técnicas de fotodisociación usadas se pueden dividir fundamentalmente en dos grupos: la que utiliza fotones del espectro de luz infrarroja (IR), también conocida por "fotodisociación multifotónica infrarroja" (IRMPD), y la que utiliza fotones del espectro de luz ultravioleta (UV), llamada "fotodisociación ultravioleta" (UVPD). Ambas técnicas se diferencian principalmente en la energía de los iones, siendo los fotones infrarrojos de baja energía (0.1 - 1 eV por fotón) y los ultravioletas de alta energía (>3 eV por fotón). Teniendo en cuenta que la energía requerida para disociar un enlace peptídico, generando iones "b" e "y", es de 4 eV, se puede observar que usando IRMPD, múltiples fotones serán necesarios para acumular dicha energía y disociar el enlace amida. En cambio, utilizando UVPD, unos pocos fotones o la absorción de un único fotón es suficiente para inducir la ruptura del enlace amida. Debido a la diferencia de energía entre ambas técnicas, los patrones de fragmentación derivados de cada una de ellas también difieren. En IRMPD, los iones hijos obtenidos son generalmente aquellos en que el enlace amida se rompe, es decir ocurre la formación de iones "b" e "y". Por lo contrario, la alta energía proporcionada por los fotones ultravioletas se acumula rápidamente y no se observan tan solo los iones "b" e "y" sino que también se observan iones "a", "x", "c" y "z"¹⁴. Incluso, si la energía y el tiempo de interacción son suficientes, iones correspondientes a la fragmentación secundaria de los iones hijos ("d", "v" o "w") pueden ser generados y observados en el espectro MS/MS.

El mecanismo de fragmentación de ambas técnicas de fotodisociación difiere notablemente. En la técnica IRMPD, los péptidos o proteínas se fragmentan en su estado electrónico fundamental, siendo estas rutas de fragmentación de baja energía que generan los mismos iones que las técnicas basadas en colisiones. En cambio, en la técnica UVPD, los fotones de mayor energía permiten generar estados electrónicos excitados de estas moléculas, accediendo así a diversas vías de fragmentación y generando una mayor colección de fragmentos¹⁵, sin embargo, la relajación del estado excitado puede ocurrir, lo cual explica la generación de iones "b" e "y" también mediante esta técnica.



3. Acoplamiento de sistemas de fotodisociación a espectrómetros de masas.

Los EM más punteros del mercado suelen estar equipados de serie para realizar una o varias de las técnicas clásicas de fragmentación (CID, HCD, ECD o ETD). Sin embargo, la fragmentación por fotodisociación aún se encuentra en fase de desarrollo. De hecho, el único EM que lleva acoplado de fábrica el sistema de fragmentación por fotodisociación es el Orbitrap Fusion Lumos, el más reciente instrumento de la generación de los Orbitraps que, además de CID, HCD y ETD, se puede adquirir con un módulo de UVPD a 213 nm desde el año 2017. Los sistemas de fotodisociación requieren del uso de láseres como fuente de fotones. Estos láseres generalmente se instalan de forma manual en el propio EM¹⁶ (Figura 2) mediante ventanas de selenuro de zinc (ZnSe) o fluoruro de bario (BaF₂), de forma que el láser pueda incidir en la celda HCD o en la trampa iónica donde tiene lugar la fotodisociación, manteniendo el EM aislado y en condiciones de vacío. Los láseres usados difieren para ambas técnicas de fotodisociación.

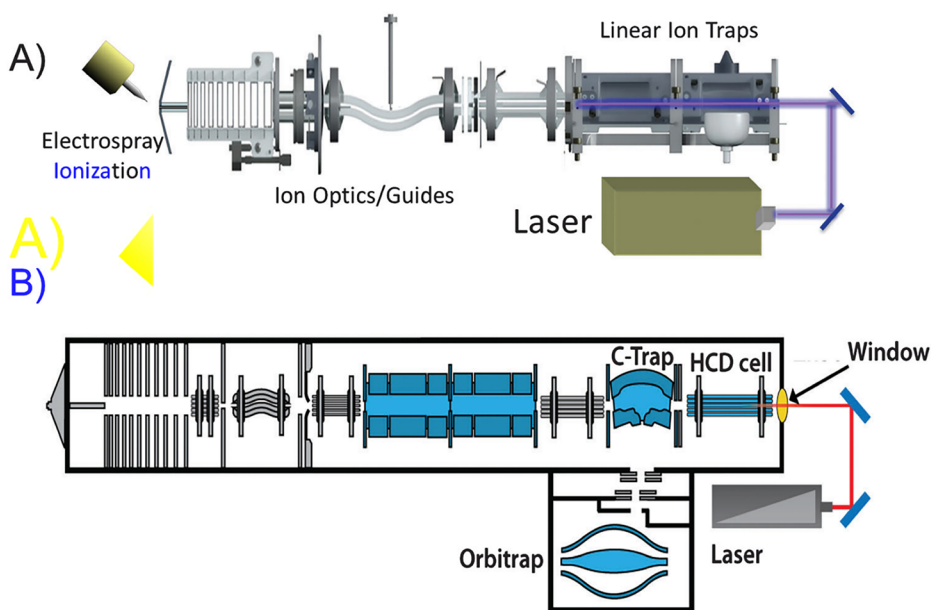


Figura 2: Implementación de sistema de fotodisociación en una trampa iónica (A) y en la celda HCD de un Orbitrap (B) (Reproducido con permiso de Ref. 16.)

En el caso de IRMPD, normalmente se utilizan láseres de CO₂ que emiten a longitudes de onda de baja energía en el espectro del infrarrojo (de 0,7 a 1000 μm), siendo 10,4 μm la longitud de onda generalmente usada. Como ya se ha comentado anteriormente, al emitir fotones de baja energía, cuando se usan láseres de CO₂ el tiempo de exposición requerido es mayor, permitiendo así a los iones interactuar con suficientes fotones para ser activados y disociarse. Cuando se trabaja en el rango del ultravioleta,

el tiempo de exposición puede minimizarse, ya que los fotones de alta energía activan rápidamente a los iones. En este caso, los láseres utilizados pueden ser de dos clases:

i) Láseres de excímeros (dímeros excitados) cómo el láser de Fluoruro de argón (ArF) que emite fotones de 193 nm, o el láser de flúor (F2) emitiendo fotones de 157 nm. Ambos láseres han sido ampliamente usados en experimentos de fotodisociación de péptidos y proteínas, ya que la energía de ambas longitudes de onda es bien absorbida por estas moléculas. Cabe destacar que el láser de F2 requiere de condiciones de vacío para evitar la absorción de los fotones por las moléculas de aire.

ii) Láseres de granate de itrio y aluminio dopado con neodimio (Nd:YAG) que emiten a 1064 nm, pero pueden ser modulados para emitir armónicos de menor longitud de onda que serán de interés para ciertas aplicaciones en alguna de las técnicas explicadas más adelante. Normalmente, los armónicos más utilizados son el cuarto (266 nm) y el tercero (355 nm). Sin embargo, la energía de estas longitudes de onda no es absorbida por los péptidos y proteínas, a excepción de los residuos aromáticos que pueden absorber luz a 266 nm. Por lo tanto, el uso de estas longitudes de onda va normalmente acompañado del acoplamiento de grupos cromóforos a los péptidos o proteínas para permitir la absorción de fotones.

4. Aplicación de la fotodisociación en el ámbito de la proteómica.

Como ya se ha comentado previamente, las técnicas de fragmentación basadas en colisiones son las más ampliamente utilizadas en estudios de proteómica. En concreto, y para experimentos de proteómica “bottom-up”, el uso CID en trampas iónicas está bastante generalizado. Esto se debe a que, en dichos experimentos, normalmente el ion padre es analizado en alta resolución en un analizador de masas del tipo Orbitrap, mientras que los iones hijos se fragmentan y analizan en la trampa iónica, un analizador de masas de menor resolución, pero que permite realizar el análisis de masas de los fragmentos a alta velocidad. En cambio, cuando se utiliza HCD, los fragmentos resultantes se analizan en el Orbitrap, lo cual incrementa el tiempo de análisis. En este sentido, la fotodisociación mediante luz del espectro infrarrojo ha sido explorada como una técnica alternativa a la fragmentación por CID en trampas iónicas.

La técnica IRMPD ofrece múltiples ventajas sobre CID: i) al ser la fragmentación independiente del voltaje de radiofrecuencia, mediante IRMPD se puede obtener un espectro MS/MS rico en masas moleculares bajas; ii) la trampa iónica permite focalizar los iones en un punto del espacio para mejorar la eficiencia de interacción con los fotones infrarrojos. Así pues, el uso de IRMPD en trampas iónicas ha demostrado ser superior al uso de CID para péptidos multicargados¹⁷ (hasta +5) y en estudios de muestras complejas¹⁸. Utilizando esta técnica de fragmentación, se ha demostrado que es posible analizar el proteoma de levadura marcado isobáricamente en una trampa iónica, lo cual es extremadamente complejo utilizando CID, debido a la pérdida



de los fragmentos de baja masa molecular, entre los cuales se encontrarían los iones diagnósticos del marcaje. Para más información sobre marcaje isobárico y otras técnicas de marcaje referirse a la revisión de Rauniyar and Yates¹⁹.

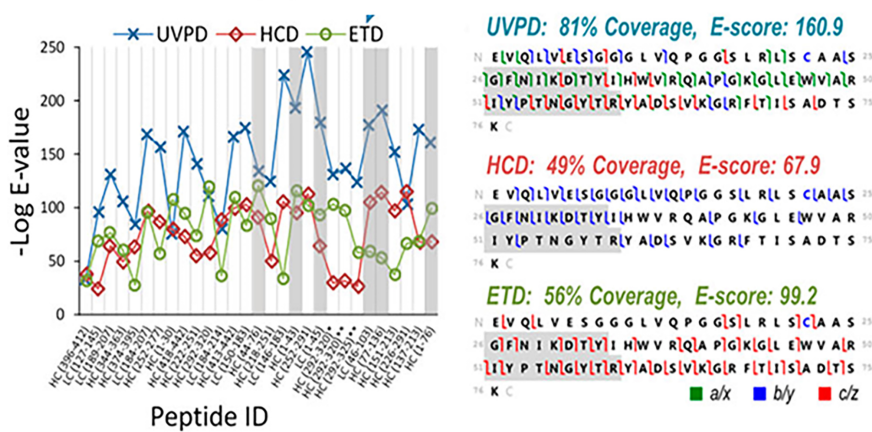
Otra de las interesantes aplicaciones de IRMPD es en la exploración de la parte ácida del proteoma. Las proteínas de naturaleza ácida o, en su lugar, aquellas que tienen un alto contenido de PTMs ácidas como la fosforilación, suelen ionizarse mejor por deprotonación que por protonación. Como para los análisis de LC-MS/MS en proteómica normalmente se utiliza el método de ionización por “electrospray” en modo positivo, los péptidos tripticos generalmente contendrán como mínimo dos sitios de protonación: el extremo N-terminal y el aminoácido básico tras el que corta la tripsina en el extremo carboxiterminal. Estos péptidos se ionizarán por ESI en medio ácido para generar cationes en fase gaseosa que serán normalmente fragmentados por CID o HCD. Como ya hemos comentado, las técnicas de fragmentación basadas en colisiones son altamente dependientes de la disponibilidad de protones móviles, ya que la relocalización de estos es fundamental para provocar la ruptura del enlace amida. Por lo tanto, estas técnicas de fragmentación no serán idóneas para fragmentar aniones cuando se trabaje usando ESI en modo negativo. En este campo se han desarrollado técnicas de fragmentación dirigidas a trabajar con aniones, como es el caso de “negative electron transfer dissociation” (NETD), y se ha explorado también las técnicas de fotodisociación. Sin embargo, la combinación entre la activación de los aniones con luz infrarroja y la disociación usando NETD ha sido, hasta la fecha, la técnica que ha permitido identificar el mayor número de proteínas en modo negativo. Utilizando esta técnica a que se denominó AI-NETD, se identificaron 3467 proteínas de un digerido triptico de levaduras²⁰. Por otro lado, la fotodisociación por UVPD usando fotones de longitud de onda de 193 nm también ha demostrado ser prometedora en el estudio de proteínas ácidas. Se publicó el resultado de identificación de 3663 y 2350 péptidos tripticos en el proteoma de células Halo y HeLa respectivamente²¹. Además, la fragmentación de aniones peptídicos por UVPD ha demostrado ser más eficiente que NETD²² siendo capaz de proporcionar una alta cobertura de secuencia²³.

La fotodisociación mediante UVPD ha sido ampliamente explorada en múltiples ámbitos de la proteómica. Esto se debe principalmente a su capacidad para obtener un patrón de fragmentos completo. Seguidamente explicaremos algunas de las aplicaciones de la UVPD en diversas áreas de la proteómica.

Una de las aplicaciones más prometedoras de la fragmentación por UVPD se encuentra en el campo de la secuenciación de novo de péptidos y proteínas, en el cual se necesita obtener información completa de la secuencia aminoacídica directamente a través del espectro MS/MS. Esta aplicación es especialmente importante en los estudios de organismos cuyo genoma y proteoma está poco caracterizado, además de los análisis de proteínas con secuencia hipervariables, como es el caso de los anticuerpos. En ambos casos es fundamental disponer de espectros MS/MS con alta cobertura de secuencia y ricos en fragmentos informativos, que permitan caracterizar la secuencia de la proteína. En este sentido, UVPD ha proporcionado resultados muy prometedores que superan claramente a otras técnicas de fragmentación. Por ejemplo, en el estudio realizado por Cotham y colaboradores en 2017²⁴ se demostró la habilidad de esta técnica de fragmentación para secuenciar de novo múltiples péptidos parcialmente digeridos (> 4.5 kDa, característicos de la técnica “middle-down”) del

anticuerpo Trastuzumab. No sólo todos los péptidos fragmentados obtuvieron mayor cobertura de secuencia por UVPD cuando se comparó con HCD o ETD (Figura 3 A y B), sino que esta técnica permitió además caracterizar el CDR3 (región de secuencia hipervariable) de tres anticuerpos contra el virus de la gripe, incluidos en el análisis (Figura 3C).

A) UVPD, HCD and ETD performance: B) UVPD, HCD and ETD performance:
E-Score for 29 peptides CDR1/CDR2



C) Antibody mixture UVPD fragmentation: CDR3 containing peptide

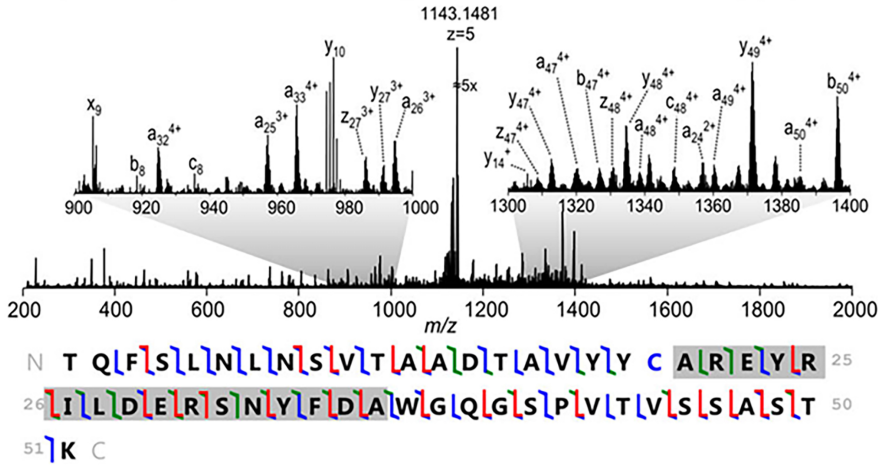


Figura 3: Comparación de la eficiencia de fragmentación de proteínas utilizando UVPD, HCD y ETD. (a) Valores de “E-score” para la identificación de 29 secuencias peptídicas utilizando los tres métodos de activación; (b) Cobertura obtenida en la fragmentación de la región determinante de la complementariedad (CDR) CDR1/CDR2 con los tres métodos de activación; (c) Espectro obtenido de la fragmentación por UVPD para el péptido correspondiente a CDR3. La región hipervariable está sombreada en gris. (Reproducido con permiso de Ref. 24.)

Otra aplicación muy prometedora de las técnicas de fragmentación por fotodisociación es el análisis de proteínas intactas, “top-down”, que es la técnica que permite la caracterización completa de una secuencia proteica. A pesar del gran potencial de la técnica, el top-down no es utilizado como técnica de rutina debido a las características de los sistemas de LC-MS/MS y de las herramientas bioinformáticas que, en general, están mejor desarrolladas para el análisis de péptidos que para el de proteínas intactas.



tas. Además, dentro de la técnica “top-down” se encuentra la espectrometría de masas nativa que permite analizar interacciones proteína-proteína o proteína-ligando, así como descifrar la estequiometría de complejos protéicos. Para realizar el análisis por EM de proteínas intactas y, en concreto, cuando se quieren estudiar las proteínas en su estado nativo, es necesario que la ionización se haga de forma suave y utilizando un solvente adecuado, para evitar la ruptura de enlaces no covalentes en el sistema a estudiar. A la vez, estas técnicas requieren de un método de disociación de alta energía, que permita obtener alta cobertura de secuencia (a nivel de estructura primaria) o que permita disociar las interacciones entre proteínas (a nivel de estructura cuaternaria). Cabe destacar, que cuando se trabaja con proteínas intactas por técnicas “top-down”, el número de cargas de los iones formados es elevado, en especial si se utilizan solventes que provocan la desnaturalización de las proteínas, lo cual generará una amplia distribución de cargas durante el proceso de ionización por ESI. La técnica de fragmentación por CID, al ser de baja energía, facilita la ruptura de los enlaces más débiles, los enlaces peptídicos o enlaces amida, pero no permite formación de suficientes fragmentos para obtener alta cobertura de secuencia cuando se trabaja con proteínas intactas. Los métodos clásicos que fragmentan con energías más elevadas (HCD, ETD o ECD) son los más utilizados para los análisis “top-down”, pero estos métodos fallan en muchos casos debido a su dependencia del estado de carga de los iones padre^{25,26}. Por ello, algunos grupos están dedicando esfuerzos en optimizar e implementar métodos de fragmentación por fotodisociación para el análisis de proteínas intactas. En 2013, se publicó el análisis de proteínas modelo de hasta 29 kDa utilizando un espectrómetro de masas Orbitrap Elite modificado para ser dotado de un láser de ArF (193 nm)²⁷. Los resultados obtenidos con la fragmentación por UVPD de proteínas intactas son muy prometedores. La comparación entre las fragmentaciones con los métodos CID, HCD y ETD demostraron que, mientras los métodos clásicos varían en la eficiencia y, por tanto, en la cobertura de secuencia según el estado de cargas del ion padre, la fragmentación de los mismos iones padres rindió aproximadamente 100% de cobertura de la secuencia en todos los casos (se analizaron iones de mioglobina con cargas de +7 a +13). Con el mismo instrumento se consiguió la mayor cobertura de secuencia de la proteína anhidrasa carbónica (87%) descrita hasta el momento. De acuerdo con la explicación de los autores, esto es debido a que el exceso de energía adquirida por la absorción de fotones de longitud de onda de 193 nm (cerca de 3 eV) facilita tanto la fragmentación de los extremos como de la región central de la proteína, proporcionando un amplio rango de tipos de fragmentos (a, b, y, c, x y z) y posibilitando una caracterización más completa de la secuencia. Además de analizar proteínas puras, la fragmentación por UVPD puede ser aplicada para el análisis de muestras complejas. Algunos estudios realizados con extractos proteicos de células Hela, demuestran la viabilidad de realizar fragmentación por UVPD en la escala de tiempo de la separación por cromatografía líquida (LC-MS/MS)²⁸. Las técnicas de fragmentación están en constante desarrollo, debido la necesidad de obtener fragmentos que faciliten la máxima cobertura de las secuencias proteicas y la fotodisociación, además de ser utilizada de forma individual, ha sido combinada con la fragmentación ETD formando la AI-ETD (activated-ion ETD)²⁹. En este caso, los iones son primeramente activados por fotones de baja intensidad (en el espectro de luz infrarroja) y luego reciben energía suplementaria de la activación por transferencia de electrones, consiguiendo así que la fragmentación por ETD se vea menos afectada por la dependencia del estado de cargas del ion padre.

Por su capacidad para retener las modificaciones postraduccionales lábiles en los fragmentos generados, la técnica de UVPD también viene siendo explorada en el estudio de PTMs, como la fosforilación y la glicosilación. La fosforilación es una modificación ácida de las proteínas que tiene un papel muy importante en la regulación de un gran número de procesos biológicos y patológicos. Esta modificación, por su naturaleza ácida, puede ser estudiada por EM en el modo negativo. En estudios llevados a cabo por dos grupos se ha observado que, para el estudio de fosfopéptidos en el modo negativo, la fragmentación por UVPD es más eficaz comparada con la fragmentación por HCD^{30,31}. Por lo que respecta a la glicosilación, se han implementado tanto métodos de disociación por IRMPD como por UVPD. En el primer caso se utilizó un espectrómetro de masas de tipo FT-ICR equipado con un láser de CO₂ para caracterizar sitios de O-glicosilación en proteínas modelo³², y para la disociación por UVPD se utilizó láser de 193 nm acoplado a un equipo LTQ Velos para estudiar O-glicosilaciones, comparando la eficiencia de la fragmentación de glicopéptidos por UVPD y CID, tanto en modo positivo como en modo negativo³³.

Por lo que respecta a estudios "bottom-up", la técnica de fotodisociación por UVPD ha sido mayoritariamente explorada empleando láseres de longitudes de onda mayores de 193 nm, dado que la alta energía de los fotones de esta longitud de onda causa una fragmentación excesiva de los péptidos. Sin embargo, cuando se usan longitudes de onda mayores a 193 nm es necesario acoplar cromóforos a los péptidos y proteínas, dado que las moléculas de naturaleza protéica no pueden absorber energía en dichas longitudes de onda y requieren de la presencia de moléculas cromóforas que incrementen la eficiencia de absorción de estos fotones. La derivatización de péptidos con moléculas cromóforas ha servido para estudiar, por ejemplo, la fracción del proteoma que contiene cisteína. Este aminoácido responsable por la formación de puentes disulfuro y requerido para establecer la estructura terciaria de las proteínas se haya presente en más de un 90% de las proteínas humanas³⁴. Por esta razón, estudiar sólo aquellos péptidos que contengan cisteínas permite reducir la complejidad del análisis proteómico sin perder la representación de la mayoría de proteínas presentes. Con la derivatización de los péptidos en los residuos de cisteína se consigue una fragmentación selectiva por UVPD de esta fracción del proteoma, como se ha demostrado en algunos estudios^{35,36,37}.

5. Conclusiones.

En los últimos años, las técnicas de fragmentación, esenciales para el estudio de péptidos y proteínas mediante espectrometría de masas han evolucionado notablemente abriendo un abanico de aplicaciones diseñadas para cubrir las necesidades de diversos campos de la proteómica. En este sentido, la fotodisociación ha demostrado en los últimos años su potencial como técnica de fragmentación de péptidos y proteínas. La fotodisociación por IRMPD, pese a presentar un patrón de fragmentación muy similar al de otras técnicas de disociación de baja energía, ha permitido superar el problema de detección de iones de bajo peso molecular en trampas iónicas. Además, ha demostrado su potencial en combinación con la técnica de fragmentación ETD, tanto en el modo de ionización positiva (AI-ETD) como negativa (AI-NETD).



La contribución de la fotodisociación por UVPD al campo de la proteómica parece ser todavía más prometedora. Al ser capaz de generar una gran variedad de fragmentos, aplicando esta técnica de fragmentación se han obtenido resultados con muy elevada cobertura de secuencia lo que es fundamental en el estudio de PTMs, o de forma más global, la caracterización de proteoformas y en el campo de la secuenciación de novo, tanto en el análisis de péptidos como de proteínas. La actual implementación de esta técnica de fragmentación a un EM comercial facilitará su utilización por un mayor número de grupos de investigación lo que conllevará al aumento y desarrollo de nuevas aplicaciones.

REFERENCIAS

1. Steen H, Mann M. The abc's (and xyz's) of peptide sequencing. *Nat Rev Mol Cell Biol* 2004; 5: 699–711
2. Horn DM, Zubarev RA, McLafferty FW. Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc Natl Acad Sci U S A* 2000; 97: 10313–7
3. 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: 4431–6
4. Smits AH, Vermeulen M. Characterizing Protein–Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends Biotechnol* 2016; 34: 825–34
5. Cristobal A, Marino F, Post H, Van Den Toorn HWP, Mohammed S, Heck AJR. Toward an Optimized Workflow for Middle-Down Proteomics. *Anal Chem*. 2017; 89(6):3318–3325
6. Smith LM, Kelleher NL, Proteomics TC for TD, et al. Proteoform: a single term describing protein complexity. *Nat Methods* 2013; 10: 186–7
7. Roepstorff P, Fohlman J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed Mass Spectrom* 1984; 11: 601
8. Mitchell Wells J, McLuckey SA. Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol* 2005; 402: 148–85
9. McLuckey SA, Goeringer DE. Slow heating methods in tandem mass spectrometry. *J Mass Spectrom* 1997; 32: 461–74
10. Olsen J V, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* 2007; 4: 709–12
11. Zubarev RA. Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol* 2004; 15: 12–6
12. 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* 2004; 101: 9528–33
13. Woodin RL, Bomse DS, Beauchamp JL. Multiphoton dissociation of gas phase ions using low intensity CW laser radiation. *Chem Biochem Appl Lasers* 1979. p. 355–88
14. Julia B-M, de Oliveira E. Photodissociation Mass Spectrometry of Peptides and Proteins. *Encycl Anal Chem* 2018; 1–22
15. Halim MA, Girod M, MacAleese L, Lemoine J, Antoine R, Dugourd P. Combined Infrared Multiphoton Dissociation with Ultraviolet Photodissociation for Ubiquitin Characterization. *J Am Soc Mass Spectrom* 2016; 27: 1435–42
16. Brodbelt JS. Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem Soc Rev* 2014; 43: 2757–83
17. Madsen JA, Brodbelt JS. Comparison of infrared multiphoton dissociation and collision-induced dissociation of supercharged peptides in ion traps. *J Am Soc Mass Spectrom* 2009; 20: 349–58
18. Ledvina AR, Lee MV, McAlister GC, Westphall MS, Coon JJ. Infrared multiphoton dissociation for quantitative shotgun proteomics. *Anal Chem* 2012; 84: 4513–9
19. Rauniyar N, Yates JR. Isobaric labeling-based relative quantification in shotgun proteomics. *J. Proteome Res. American Chemical Society* 2014. p. 5293–309
20. Riley NM, Rush MJP, Rose CM, et al. The Negative Mode Proteome with Activated Ion Negative Electron Transfer Dissociation (AI-NETD). *Mol Cell Proteomics* 2015; 14: 2644–60
21. Madsen JA, Xu H, Robinson MR, et al. High-throughput database search and large-scale negative polarity liquid chromatography-tandem mass spectrometry with ultraviolet photodissociation for complex proteomic samples. *Mol Cell Proteomics* 2013; 12: 2604–14
22. Shaw JB, Madsen JA, Xu H, Brodbelt JS. Systematic comparison of ultraviolet photodissociation and electron transfer dissociation for peptide anion characterization. *J Am Soc Mass Spectrom* 2012; 23: 1707–15
23. Madsen JA, Kaoud TS, Dalby KN, Brodbelt JS. 193-nm photodissociation of singly and multiply charged pep-

tide anions for acidic proteome characterization. *Proteomics* 2011; 11: 1329–34

24. Cotham VC, Horton AP, Lee J, Georgiou G, Brodbelt JS. Middle-Down 193-nm Ultraviolet Photodissociation for Unambiguous Antibody Identification and its Implications for Immunoproteomic Analysis. *Anal Chem* 2017; 89: 6498–504
25. Rožman M, Gaskell SJ. Charge state dependent top-down characterisation using electron transfer dissociation. <https://onlinelibrary-wileycom.sire.ub.edu/doi/pdf/10.1002/rcm.5330>
26. Greer SM, Holden DD, Fellers R, Kelleher NL, Brodbelt JS. Modulation of Protein Fragmentation Through Carbamylation of Primary Amines. *J Am Soc Mass Spectrom* 2017; 28: 1587–99
27. Shaw JB, Li W, Holden DD, et al. Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J Am Chem Soc* 2013; 135: 12646–51
28. Cleland TP, DeHart CJ, Fellers RT, et al. High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. *J Proteome Res* 2017; 16: 2072–9
29. Riley NM, Westphall MS, Coon JJ. Sequencing Larger Intact Proteins (30–70 kDa) with Activated Ion Electron Transfer Dissociation. *J Am Soc Mass Spectrom Journal of The American Society for Mass Spectrometry*; 2018; 29: 140–9
30. Fort KL, Dyachenko A, Potel CM, et al. Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics. *Anal Chem* 2016; 88: 2303–10
31. Robinson MR, Taliaferro JM, Dalby KN, Brodbelt JS. 193 nm Ultraviolet Photodissociation Mass Spectrometry for Phosphopeptide Characterization in the Positive and Negative Ion Modes. *J Proteome Res* 2016; 15: 2739–48
32. Seipert RR, Dodds ED, Lebrilla CB. Exploiting Differential Dissociation Chemistries of O-Linked Glycopeptide Ions for the Localization of Mucin-Type Protein Glycosylation. *J Proteome Res* 2009; 8: 493–501
33. Chandler KB, Pompach P, Goldman R, Edwards N. Exploring Site-Specific N-Glycosylation Microheterogeneity of Haptoglobin Using Glycopeptide CID Tandem Mass Spectra and Glycan Database Search. *J Proteome Res* 2013; 12: 3652–66
34. Giron P, Dayon L, Sanchez J-C. Cysteine tagging for MS-based proteomics. *Mass Spectrom Rev* 2011; 30: 366–95
35. Parker WR, Holden DD, Cotham VC, Xu H, Brodbelt JS. Cysteine-Selective Peptide Identification: Selenium-Based Chromophore for Selective S-Se Bond Cleavage with 266 nm Ultraviolet Photodissociation. *Anal Chem* 2016; 88: 7222–9
36. Agarwal A, Diedrich JK, Julian RR. Direct elucidation of disulfide bond partners using ultraviolet photodissociation mass spectrometry. *Anal Chem* 2011; 83: 6455–8
37. Enjalbert Q, Simon R, Salvador A, et al. Photo-SRM: laser-induced dissociation improves detection selectivity of selected reaction monitoring mode. *Rapid Commun Mass Spectrom* 2011; 25: 3375–81



Técnicas de adquisición independiente de datos

Emilio Camafeita, José Manuel Rodríguez, Enrique Calvo y Jesús Vázquez

Laboratorio de Proteómica Cardiovascular, Centro Nacional de Investigaciones Cardiovasculares, Madrid

1. El problema del submuestreo en proteómica.

En proteómica convencional, o *bottom-up*, se lleva a cabo una digestión enzimática del extracto proteico que se quiere analizar seguida del análisis de los péptidos resultantes mediante cromatografía líquida acoplada a espectrometría de masas (*liquid chromatography coupled with mass spectrometry*, LC-MS). En el método convencional de barrido, denominado adquisición dependiente de datos (*data-dependent acquisition*, DDA), los iones cuya señal supera un umbral de ruido preestablecido en el espectro de masas se aíslan y se fragmentan dando lugar a un espectro de masas en tándem o de segundo orden (MS/MS) (Fig. 1), característico de la secuencia aminoacídica del péptido. Los espectros MS/MS permiten la identificación de los péptidos comparándolos con los patrones teóricos generados a partir de una base de datos. Los métodos DDA (también llamados *shotgun*) son muy eficientes a la hora de identificar el mayor número posible de proteínas en muestras biológicas; no obstante, en muestras muy complejas el número de iones precursores que coeluyen por unidad de tiempo puede superar la velocidad de barrido del espectrómetro, de manera que sólo los iones más intensos son sometidos a fragmentación, produciendo un submuestreo de la población de péptidos sesgado hacia las proteínas más abundantes. Este problema de submuestreo del que adolecen los métodos DDA se traduce en una disminución del rango dinámico que limita el análisis de especies proteicas poco abundantes(1). Por otro lado, debido a la naturaleza estocástica del proceso de selección de iones y su posterior fragmentación, no siempre se fragmentan exactamente los mismos iones cuando se analizan muestras diferentes en el mismo equipo; ello complica los estudios comparativos que requieren un análisis cuantitativo sistemático de las mismas proteínas en muchas muestras diferentes.

Debido a estas limitaciones se han desarrollado ensayos proteómicos dirigidos en los que los iones que se van a fragmentar se especifican *a priori*, sin tener en cuenta su intensidad previa. Entre estos ensayos destaca la monitorización selectiva de reacciones (*selected reaction monitoring*, SRM). Este abordaje concentra los recursos del espectrómetro de masas en el análisis de un conjunto restringido de iones precursores, a definir de manera previa a la adquisición de datos, que representan a determinadas proteínas de interés (Fig. 1). Gracias a su elevada sensibilidad y su sobresaliente precisión a la hora de cuantificar los péptidos escogidos, el ensayo SRM resulta espe-



cialmente adecuado para la cuantificación de un conjunto reducido de proteínas en un número elevado de muestras, por lo que es el método más usado para la validación de biomarcadores. Sin embargo, dado lo limitada que resulta la fracción del proteoma que abarca el análisis, el método SRM no resulta aplicable en buena parte de los experimentos proteómicos, cuyo fin es explotar la potencia de la tecnología DDA para identificar el máximo número posible de proteínas y revelar así en detalle la composición proteica de la muestra.

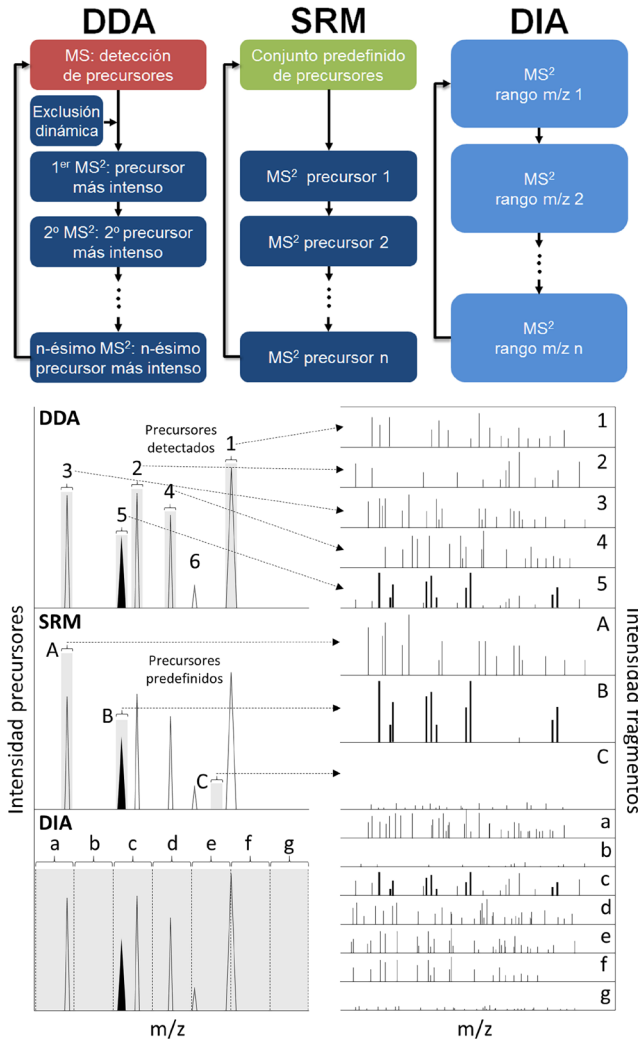


Figura 1. Esquema de los distintos modos de adquisición de datos en un espectrómetro de masas: adquisición dependiente de datos (data dependent acquisition, DDA), monitorización selectiva de reacciones (selected reaction monitoring, SRM) y adquisición independiente de datos (data independent acquisition, DIA). Tanto en DDA como en SRM los iones precursores son aislados, fragmentados y analizados individualmente, ya sea eligiéndolos en el transcurso del análisis en función de su intensidad (DDA) o tras prefijarlos con antelación (SRM). En DIA, sin embargo, se omiten tanto la detección como la predefinición de precursores para aislar y fragmentar en paralelo la totalidad de los péptidos que aparecen en un determinado rango de m/z.

A lo largo de los últimos años se han desarrollado técnicas de barrido alternativas, como la adquisición independiente de datos (*data-independent acquisition*, DIA), que trata de aunar la gran cobertura de proteínas que permiten los métodos DDA con la ca-

pacidad de los métodos SRM de fragmentar consistentemente las mismas especies. Los métodos DIA, que suscitan un interés creciente(2), llevan a cabo la fragmentación de la totalidad de los péptidos comprendidos en un determinado rango de masas que se divide en ventanas, sin necesidad de establecer previamente los iones que se van a fragmentar (Fig. 1). Ello permite fragmentar todas las especies y obtener sistemáticamente los mismos espectros de fragmentación en todas las carreras. En los métodos DIA convencionales, sin embargo, al no depender la ventana de aislamiento de la identificación previa de un precursor, se pierde el vínculo entre los iones fragmento y los iones precursores que los originan, generándose espectros de fragmentación compuestos, producto de la fragmentación simultánea de varios péptidos de masa similar que se aíslan en la misma ventana, cuya complejidad obstaculiza su análisis posterior.

2. Implementaciones del método DIA.

El método DIA ha sido concebido desde distintos planteamientos en diversas plataformas tecnológicas. Así mismo, el problema de la complejidad de los espectros de fragmentación compuestos generados por este método se ha abordado desde distintos enfoques, ya sea recurriendo a librerías espectrales o mediante la deconvolución de los espectros de fragmentación. Todo ello ha dado lugar a diversas implementaciones(3-14) (Tabla 1), entre las cuales SWATH y el barrido independiente de datos (*data independent scanning*, DiS) merecen mención especial y se describen en secciones aparte.

Método	Instrumentación	Anchura de la ventana de aislamiento	Año	Referencia
Shotgun CID	Q-TOF	Rango de m/z completo	2003	3
DIA	Trampa iónica	10 m/z	2004	4
MS ^E	QqTOF	Rango de m/z completo	2005	5
p ² CID	Q-TOF	Rango de m/z completo	2006	6
PAcIFIC	Trampa iónica	2,5 m/z	2009	7
AIF	Orbitrap	Rango de m/z completo	2010	8
XDIA	Trampa iónica	20 m/z	2010	9
SWATH	QqTOF	25 m/z	2012	10
FT-ARM	LTQ-orbitrap	12 m/z, 100 m/z	2012	11
MSX	Híbrido Q-orbitrap	4 m/z	2013	12
HRM	Híbrido Q-orbitrap	Variable	2015	13
DiS	Híbrido Q-orbitrap	2 m/z	2016	14

AIF: all ions fragmentation; CID: collision-induced dissociation; DIA: data independent acquisition; DiS: data independent scanning; ETD: electron-transfer dissociation; FT-ARM: Fourier transform-all reaction monitoring; HRM: hyper reaction monitoring; p²CID: parallel collision-induced dissociation; PAcIFIC: precursor acquisition independent from ion count; XDIA: extended data-independent acquisition.

Tabla 1. Métodos de adquisición independiente de datos.



2.1. Descubrimiento del método DIA

En su trabajo pionero de 2000, cuando el método DDA empezaba a asentarse, Mas-selon *et al.*(15) demostraron que era posible identificar péptidos en una sencilla mezcla de siete especies mediante su fragmentación simultánea en un espectrómetro de masas de resonancia ion-ciclotrón con transformada de Fourier, gracias en buena medida a su sobresaliente exactitud de masa y a su resolución. Posteriormente, reconocidas ya las limitaciones de las que adolece el método DDA(16), Purvine *et al.*(3) lograron identificar péptidos ionizados en un instrumento ESI-TOF (es decir, que emplea ionización por electrospray, ESI) acoplada a un analizador de tiempo de vuelo (*time-of-flight*, TOF)) a partir de la fragmentación simultánea de la totalidad de los iones precursores que coeluyen en un determinado momento de la separación cromatográfica. Un año más tarde, Venable *et al.*(4) dieron a conocer su método DIA, basado en el aislamiento y fragmentación secuenciales de iones precursores en ventanas de 10 m/z en una trampa iónica lineal. Con este método se cubre en unos segundos el rango completo de m/z , de forma que cada péptido es fragmentado varias veces a lo largo de su tiempo de elución. En comparación con el método DDA convencional, los autores encontraron que el método DIA incrementa en un 20% el número de espectros de fragmentación. Sin embargo, los dos métodos de adquisición dieron lugar a números comparables de péptidos y proteínas identificados, ya que no se aplicó ninguna estrategia específica para la identificación de péptidos a partir de los complejos espectros de fragmentación compuestos obtenidos mediante DIA. De este modo, la búsqueda en bases de datos proporcionó un único candidato por espectro, como ocurre en DDA, obviando buena parte de la información contenida en dicho espectro. El método XDIA (*extended data-independent acquisition*, adquisición independiente de datos ampliada), presentado como una prolongación del método DIA, combina medidas de alta resolución en modo MS en orbitrap con fragmentaciones consecutivas en ventanas de 20 m/z con 1 m/z de solapamiento(9). La fragmentación se lleva a cabo alternando disociación por transferencia electrónica y disociación inducida por colisión (*collision-induced dissociation*, CID).

2.2. El método MS^E

En 2005 Silva *et al.*(5) desarrollaron para la plataforma cuadrupolo-TOF el método MSE, basado en la aplicación alterna de baja y alta energía de colisión a fin de registrar dos canales de datos en las medidas de MS: uno con las señales de masa provenientes de péptidos intactos (a 10 eV) y otro con las de los correspondientes fragmentos (a 28-35 eV). A continuación, los precursores son asociados a sus correspondientes fragmentos atendiendo a sus propiedades cromatográficas, para lo cual se emplea un algoritmo que modela las desviaciones de los tiempos de retención. A la hora de efectuar comparaciones cuantitativas entre distintas muestras, se emplean estándares internos (endógenos o exógenos) para corregir la variabilidad asociada al proceso de inyección y a la cantidad de proteína cargada. De esta forma, el complejo conjunto de datos inicial queda reducido a un número limitado de componentes cuya abundancia relativa cambia de forma estadísticamente significativa en una determinada comparativa, de modo que las proteínas implicadas son identificadas mediante búsqueda en bases de datos. En los últimos años la evolución del abordaje MSE, que ha gozado de gran popularidad entre los métodos DIA(17), ha derivado en las versiones de alta y

ultra definición (HDMS^E y UDMS^E respectivamente). En HDMS^E los iones precursores son separados en función de su movilidad mediante espectrometría de masas de movilidad iónica y a continuación fragmentados empleando energías de colisión incrementales(18), mientras que UDMS^E aplica energías de colisión específicas en función del tiempo de deriva del ion(19).

2.3. Disociación paralela inducida por colisión (p²CID)

La energía de colisión necesaria para fragmentar un determinado ion depende de su carga y de su masa(20), de tal manera que los iones de mayor tamaño y menor carga requieren energías más altas. Frente al método DDA, en el que la energía de colisión se ajusta para cada uno de los iones precursores aislados, en DIA se emplea un único valor, o a lo sumo un conjunto reducido de valores, de dicha energía para la fragmentación simultánea de la totalidad de los péptidos coeluyentes. Desarrollada en la plataforma cuadrupolo TOF, la disociación paralela inducida por colisión (*parallel collision-induced dissociation*, p²CID) alterna dos procesos de fragmentación CID sin selección de ion precursor: uno en la fuente y otro en la celda de colisión(6). Para la fragmentación paralela en la fuente se emplea una energía elevada que por una parte genera iones grandes monocargados y por otra redistribuye la carga de los péptidos para generar fundamentalmente iones con una y dos cargas. A continuación, los iones precursores intactos y los productos de su fragmentación en la fuente son conducidos a la celda de colisión CID para su fragmentación en paralelo.

2.4. Adquisición de precursores independiente del número de iones (PACIFIC)

El método PACIFIC (*precursor acquisition independent from ion count*, adquisición de precursores independiente del número de iones)(7) recurre a múltiples inyecciones de la muestra para fragmentar los péptidos, sin selección previa de precursor, en 10 intervalos contiguos de 1,5 m/z de anchura con ventanas de aislamiento de 2,5 m/z. Con cada nueva inyección, hasta totalizar 67, las ventanas se desplazan hasta cubrir el rango de iones precursores 400-1400 m/z. A los espectros de fragmentación así obtenidos se les asigna como ion precursor el valor central de la correspondiente ventana de aislamiento. De este modo, en la búsqueda en bases de datos puede emplearse una tolerancia de $\pm 3,75$ Da, correspondiente a $\pm 1,25$ m/z para un precursor de carga +3, para la masa del ion precursor. En comparación con el método DIA original(4, 9), el uso de ventanas de aislamiento significativamente más estrechas que propone PACIFIC rebaja notablemente la complejidad de los espectros de fragmentación compuestos resultantes, lo cual redundará en una mayor eficacia de identificación en la búsqueda en bases de datos. No obstante, el método requiere como contrapartida un elevado número de inyecciones que cuestionan no solo su reproducibilidad, sino también su viabilidad por el gran consumo de muestra que necesita.

2.5. El método AIF

El método AIF (*all ion fragmentation*, fragmentación de todos los iones) se beneficia de la posibilidad que brinda la celda de colisión HCD (*higher-energy collisional*



dissociation, disociación por colisión de alta energía) del espectrómetro de masas *Exactive* de fragmentar péptidos eficientemente sin selección previa de iones precursores(8), de forma análoga a la metodología previamente descrita para la plataforma cuadrupolo-TOF (véanse los métodos MS^E y p²CID más arriba). Una vez realizada la fragmentación, los iones fragmento son separados en el analizador de alta resolución orbitrap, lo cual facilita su asignación a iones precursores de acuerdo con sus perfiles de coelución. Para ello, los autores llevaron a cabo medidas alternas en modo MS y AIF, a alta resolución (100.000 a m/z 200) en ambos casos, con una frecuencia de una medida por segundo, lo cual garantiza que todos los péptidos separados cromatográficamente sean detectados y fragmentados varias veces. En comparación con instrumentos anteriores basados en el analizador orbitrap, el *Exactive* exhibe la capacidad única de llenar la trampa C (el dispositivo que inyecta los iones en el orbitrap, o *C-trap*) mientras el orbitrap analiza los iones inyectados en el evento anterior; esto permitió a los autores ajustar el tiempo máximo de inyección a 1 s, a fin de maximizar la sensibilidad, sin que se viese afectado el tiempo de ciclo. Para la interpretación de los datos los autores recurrieron al programa *MaxQuant*(21), que primero asigna los picos AIF (iones fragmento) a señales MS (iones precursores) en función de sus perfiles de elución para después llevar a cabo la identificación de péptidos mediante búsqueda en bases de datos de proteínas.

2.6. Monitorización de todas las reacciones con transformada de Fourier (FT-ARM)

El método FT-ARM (*Fourier transform-all reaction monitoring*, monitorización de todas las reacciones con transformada de Fourier) se basa en una combinación de DIA y análisis de correlación espectral(11). El método lleva a cabo la fragmentación en paralelo de la totalidad de los iones precursores presentes en el rango 500-1500 m/z empleando ventanas de aislamiento de 12 o 100 m/z en un instrumento híbrido trampa iónica lineal-orbitrap. La exactitud de masa con que se miden los fragmentos proporciona la selectividad necesaria para llevar a cabo la identificación de péptidos en muestras complejas, sin haber determinado previamente la masa de los iones precursores, mediante una búsqueda convencional en bases de datos. La puntuación de los candidatos así obtenidos se calcula mediante el producto escalar entre el espectro de fragmentación teórico y cada uno de los espectros experimentales, de tal manera que la puntuación es elevada cuando las señales del espectro de referencia se alinean con las señales del espectro a comparar con una tolerancia de 5 ppm. En comparación con DDA, FT-ARM identifica un número sensiblemente menor de péptidos cuando se emplea una ventana de aislamiento de 100 m/z. Esta diferencia a favor de DDA disminuye rápidamente a medida que la ventana se estrecha, lo cual apunta a una menor eficiencia del aislamiento y fragmentación en ventanas más anchas. Por otra parte, la cuantificación es lineal y muy sensible; y es posible para cada uno de los péptidos analizados mediante FT-ARM.

2.7. El método MSX

Con el fin de evitar la pérdida de selectividad que supone el empleo de ventanas de aislamiento anchas, manteniendo al mismo tiempo una frecuencia de muestreo

aceptable, el método MSX(12) se inspira en el multiplexado basado en la transformada de Hadamard(22), dividiendo el rango 500-900 m/z en 100 ventanas de 4 m/z, de las cuales cinco son seleccionadas de forma aleatoria para la fragmentación en paralelo de la totalidad de los iones que abarcan en un espectrómetro de masas *Q-Exactive*. De esta forma, MSX alcanza la frecuencia de muestreo de un método DIA que emplease 20 ventanas de 20 m/z con la selectividad de un abordaje basado en 100 ventanas de 4 m/z. Los espectros de fragmentación resultantes se analizan con *Skyline*(23) para su procesamiento, lo que reduce el ruido químico y resulta en una mejora notable de la selectividad, ya que elimina de los espectros de fragmentación buena parte de los fragmentos interferentes de péptidos coeluyentes aislados en otras ventanas de aislamiento.

2.8. Monitorización de hiperreacciones (HRM)

El empleo de librerías espectrales normalizadas por tiempo de retención, construidas a partir de medidas repetitivas en modo DDA, constituye la esencia de la monitorización de hiperreacciones (*hyper reaction monitoring*, HRM)(13). El método, desarrollado en un espectrómetro de masas *Q-Exactive*, comprende una medida MS seguida de 19 medidas MS² realizadas en ventanas de aislamiento que varían en función de la densidad de iones precursores coeluyentes, un parámetro que cambia en el transcurso de la separación cromatográfica. De esta forma se rebaja sensiblemente el tiempo de ciclo, que se reduce casi a la mitad en comparación con AIF (véase el Apartado 2.5 más arriba), acercándose su duración a la del método SWATH (véase el Apartado siguiente). Para el análisis de los espectros DIA así obtenidos, HRM emplea *Spectronaut*(24), que incorpora algoritmos basados en *machine learning* para la definición y evaluación de picos. A continuación, los datos son confrontados con la librería espectral. El análisis comparativo de DDA y HRM puso de manifiesto la superioridad del método DIA, no solo en cuanto al número de péptidos identificados, sino también en la reproducibilidad de la cuantificación de proteínas(13).

3. SWATH.

Basándose en el concepto de fragmentación de precursores aislados secuencialmente en ventanas amplias propuesto por Venable *et al.*(4), en 2012 Gillet *et al.*(10) presentaron SWATH, que combina la elevada especificidad del método DIA con una novedosa estrategia dirigida de extracción de datos a partir de espectros de fragmentación compuestos. En su implementación original, SWATH opera a alta resolución con ventanas de aislamiento de 25 m/z, denominadas *swaths**, para cubrir en 3,2 s el intervalo 400-1.200 m/z en $(1200 - 400) / 25 = 32$ etapas secuenciales que generan sendos espectros de fragmentación compuestos (Fig. 2a), de tal manera que cada ventana es analizada varias veces en el tiempo de elución de un determinado péptido. Este ciclo se repite a lo largo de todo el cromatograma para generar complejos mapas iónicos a partir de la fragmentación de la práctica totalidad de los precurso-

*El término *swath* alude a la faja de hierba o mies que es segada en un golpe de guadaña.



res detectables. A continuación, dichos mapas de fragmentación son rastreados en busca de péptidos concretos, generalmente empleando la información disponible *a priori* en librerías espectrales, e.g. valores de m/z , intensidades relativas y tiempos de retención (Fig. 2b). De este modo quedan delimitados conjuntos de señales de masa coeluyentes que correlacionan de forma precisa con la información disponible para un determinado péptido, el cual resulta así inequívocamente identificado en el complejo mapa de fragmentación. Al igual que SRM (véase el Apartado 1 más arriba), SWATH aprovecha la elevada especificidad de la fragmentación peptídica para llevar a cabo la identificación y cuantificación de péptidos con precisión; no obstante, al abarcar la totalidad de los precursores y sus fragmentos, SWATH prescinde del complejo proceso de selección de péptidos diana y optimización de transiciones. En comparación con DDA, SWATH arroja mejoras en la cobertura de detección y en la reproducibilidad; sin embargo, el papel esencial que juegan en SWATH las librerías espectrales supone una limitación del método, ya que es necesario emplear tiempo y muestra para la preparación de dichas librerías en idénticas condiciones experimentales a fin de determinar empíricamente el patrón de fragmentación de cada péptido de forma precisa. Asimismo, no es fácil estimar la tasa de falsos positivos que se obtienen en este abordaje, ya que no resulta evidente cómo incorporar un proceso equivalente a la búsqueda en

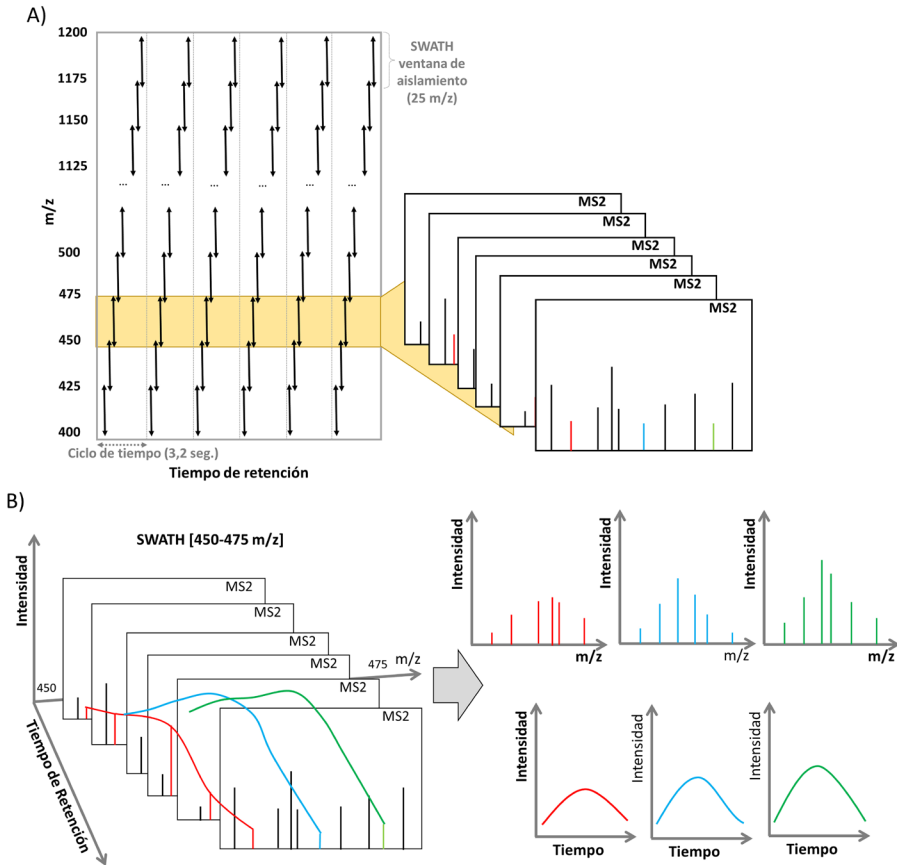


Figura 2. Esquema de la tecnología SWATH A) El rango de masas 400-1200 m/z se recorre fragmentándose los iones peptídicos en ventanas de aislamiento de 25 m/z . Para cada ventana de aislamiento, los iones fragmentados se registran en un espectro de masas de alta resolución (MS2) durante 100 ms; con lo que para 32 ventanas de aislamiento que cubren 400-1200 m/z , se requiere un tiempo de ciclo de 3,2 seg. B) Las intensidades de los iones fragmentados se extraen de los espectros de fragmentación de alta resolución para producir los correspondientes cromatogramas de iones extraídos.

bases de datos invertidas a algoritmos de puntuación basados en librerías, para los cuales la intensidad, un parámetro ausente en las bases de datos comunes, constituye un factor esencial.

3.1. Herramientas bioinformáticas para el análisis de datos SWATH

Los algoritmos de búsqueda que se emplean convencionalmente para la identificación de péptidos en los abordajes DDA resultan inapropiados para el análisis de datos obtenidos mediante un método DIA como SWATH por una doble razón: en primer lugar, suponen, incorrectamente, que las señales de masa de cada espectro de fragmentación provienen de un único péptido; en segundo lugar, obvian los patrones dinámicos de los perfiles de elución asociados a los complejos mapas de fragmentación generados por SWATH. Por ello se han desarrollado diversas herramientas computacionales orientadas específicamente a la identificación de péptidos a partir de la información obtenida mediante este abordaje. Dichas herramientas se agrupan en dos categorías principales: herramientas dirigidas (basadas en librerías) y herramientas no dirigidas (independientes de librerías).

3.1.1. Análisis dirigido de datos provenientes de SWATH

El análisis dirigido, basado en librerías, constituye la estrategia más común para la identificación de péptidos a partir de los datos generados por SWATH. Este abordaje se basa en rastrear la correlación entre señales de masa y la información disponible a priori para dichos péptidos. Dicha información se almacena en librerías espectrales que generalmente especifican, para cada péptido, los siguientes parámetros: i) valor de m/z del precursor; ii) valor de m/z de una selección de los correspondientes iones fragmento, así como sus intensidades relativas; y iii) tiempo de retención cromatográfica. De este modo funcionan herramientas como *Skyline*(23), *Spectronaut*(24), *OpenSWATH*(25) y *MSPLIT-DIA*(26).

Implementado en la plataforma libre OpenMS(27), OpenSWATH(25) automatiza el proceso de análisis simplificado descrito originalmente por Gillet *et al.*(10), basado en el rastreo de los cromatogramas de los fragmentos más intensos de los distintos precursores en los correspondientes espectros de fragmentación. Para ello, en una primera etapa convierte a un formato apropiado tanto los datos SWATH originales, que pueden provenir de distintos tipos de espectrómetros de masas, como las librerías a emplear. A continuación, se emplean péptidos de referencia para alinear los distintos cromatogramas de entrada a fin de normalizar los datos para extraer los cromatogramas de los iones fragmento. El algoritmo señala entonces en qué posiciones de los cromatogramas coeluyen los distintos fragmentos y les asigna una puntuación en función de los perfiles de elución de los fragmentos y el grado de coincidencia del tiempo de retención y las intensidades de los fragmentos experimentales con los valores esperados. Finalmente, el programa calcula una FDR por comparación de los resultados con los obtenidos en ensayos *decoy* puntuados de forma idéntica a la descrita. *Spectronaut* implementa una estrategia similar, si bien en este caso el nivel de significación estadística de la identificación de los péptidos se evalúa a través de *mProphet*(24), un programa que incorpora varios algoritmos de puntuación que atien-



den a parámetros como la coincidencia temporal y forma de los picos coeluyentes, así como la semejanza global entre estos y los correspondientes a péptidos de referencia tanto endógenos como exógenos.

A fin de evitar el tedioso proceso de obtención de librerías espectrales específicas de cada experimento SWATH, a lo largo de los últimos años han surgido diversas librerías de acceso público(28). Algunas de estas librerías comprenden los proteomas completos de organismos como *Saccharomyces cerevisiae*(29), *Streptococcus pyogenes*(30) y *Mycobacterium tuberculosis*(31), mientras que otras abarcan subproteomas humanos de relevancia en distintas patologías como el glicoproteoma humano(32) y proteínas relacionadas con cáncer(33). Las librerías están disponibles en repositorios como ProteomeXChange(34), PeptideAtlas(35) y SWATHAtlas(36); este último contiene información suficiente para cuantificar mediante SWATH cerca de la mitad de las proteínas humanas anotadas en la base de datos UniProtKB/Swiss-Prot(37).

3.1.2. Análisis no dirigido de datos provenientes SWATH

Si bien el abordaje dirigido, basado en librerías, de los datos proporcionados por SWATH cuantifica péptidos con buena sensibilidad, es evidente que su ámbito está limitado a los analitos acerca de los cuales existe información en las librerías empleadas. Por ello, se han desarrollado programas como *DeMux*(38) y *DIA-Umpire*(39) y *Group-DIA*(40) para la identificación de péptidos en este tipo de aproximación DIA que prescinden del uso de librerías. Dichos programas analizan detalladamente la correlación entre clústeres de iones con el fin de calcular los correspondientes perfiles de elución, de modo que el posterior análisis de la covarianza de los perfiles de elución permite elaborar espectros sintéticos que vinculan iones fragmento a precursores únicos. Los espectros sintéticos así obtenidos se emplean en búsquedas convencionales en bases de datos de proteínas como si de datos DDA se tratase. Tanto *DeMux*(38) como *DIA-Umpire*(39) superan sensiblemente a la búsqueda en bases de datos convencional en el número de péptidos identificados a partir de datos DIA; sin embargo, el método convencional DDA supera a DIA con *DeMux* o *DIA-Umpire* en número de péptidos identificados.

3.2. Aplicaciones del método SWATH

SWATH se ha aplicado en proyectos de muy diversa índole empleando distintos tipos de muestras como células, tejidos, plasma, fluido del lavado broncoalveolar, orina, etc. Así, Williams et al. llevaron a cabo un estudio a gran escala del papel de la mitocondria en el metabolismo hepático del ratón mediante el análisis del genoma, el transcriptoma, el metaboloma y el proteoma de cientos de individuos pertenecientes a la población de referencia genética BXD(41). Gracias en parte a la precisa cuantificación de 2622 proteínas llevada a cabo mediante este abordaje, los autores descubrieron diversos vínculos entre variantes genéticas. Entre éstas destacan las variantes de secuencia de *Cox7a2l* (también llamada SCAF1, o *supercomplex assembly factor 1*) que alteran la actividad de la correspondiente proteína y son en última instancia las responsables de las diferencias encontradas en lo que respecta a la formación de supercomplejos mitocondriales.

En plasma, SWATH se ha empleado, por ejemplo, para estudiar las variaciones del proteoma en un estudio longitudinal con mellizos (gemelos dicigóticos) y gemelos idénticos (monocigóticos)(42). Los autores tomaron muestras de sangre en dos tiempos distintos a 22 parejas de mellizos y 36 parejas de gemelos idénticos y elaboraron una librería espectral, que contaba con 43000 péptidos pertenecientes a 1667 proteínas, a partir de la información disponible sobre el proteoma de plasma en PeptideAtlas(43) y experimentos DDA adicionales. Mediante OpenSWATH(25) fue posible rastrear en los 232 mapas de datos DIA resultantes 1904 péptidos correspondientes a 342 proteínas, una gran mayoría de las cuales fueron cuantificadas con coeficientes de variación menores del 25%, mientras que se identificaron 13 polimorfismos de un solo nucleótido(42). Estos resultados superan ampliamente a los de abordajes previos basados en *arrays* de anticuerpos(44).

SWATH ha sido utilizado igualmente para la búsqueda de biomarcadores en diversos tipos de cáncer. Así, un estudio reciente utilizó la cuantificación de glicoproteínas mediante esta técnica para la diferenciación de distintos tipos de cáncer: normal, no agresivo, agresivo y metastásico. El análisis de biopsias de próstata humana(45) condujo a la cuantificación de unos 1400 sitios de glicosilación y el descubrimiento de algunos marcadores de agresividad tumoral. SWATH ha permitido asimismo elaborar un panel de posibles biomarcadores de adenocarcinoma en el fluido del lavado broncoalveolar(46). Lambert et al. acoplaron la purificación por afinidad con SWATH y DDA para analizar la modulación de la red de interacciones proteína-proteína originada por mutaciones asociadas a melanoma en la quinasa humana CDK4(47). OpenSWATH(25) logró cuantificar más de 5000 péptidos, superando en 2300 a DDA. Asimismo, la fracción de péptidos cuantificados con coeficientes de variación inferiores al 20% fue mayor en el caso de SWATH (82%) en comparación con DDA (75%).

Kulkarni *et al.*(48) estudiaron las proteínas transportadas por exosomas aislados tanto de sangre como de orina de ratones expuestos a radiación. El análisis exhaustivo mediante SWATH permitió establecer sendos paneles de biomarcadores en tejido de la respuesta sistémica a altas dosis de radiación ionizante en ambos fluidos. Un estudio de enzimas metabolizadoras y transportadores de fármacos en microsomas humanos de origen intestinal, hepático y renal llevó a cabo una comparación entre SWATH y métodos DIA dirigidos. Los autores encontraron muy buen acuerdo entre los valores cuantitativos proporcionados por ambos abordajes, si bien SWATH arrojó menores coeficientes de variabilidad(49). En el terreno de las enfermedades neurológicas, la técnica ha servido para caracterizar el proteoma sináptico en pacientes de Alzheimer(50). De esta forma se encontraron 30 proteínas, implicadas en procesos como la transducción de señal, la autofagia y el estrés oxidativo, cuya abundancia se ve alterada de forma significativa en los pacientes respecto de los controles.

La reproducibilidad del método ha sido analizada exhaustivamente por Collins *et al.*(51). Empleando una librería espectral generada por un laboratorio ajeno al estudio, los autores consiguieron cuantificar varios miles de proteínas de forma reproducible mediante un ensayo SWATH multicéntrico en el que participaron once laboratorios y que arrojó asimismo límites de detección y rangos dinámicos muy homogéneos, para lo cual resultó esencial el control estricto de la FDR tanto a nivel de proteína como de péptido(52).



4. El método DiS.

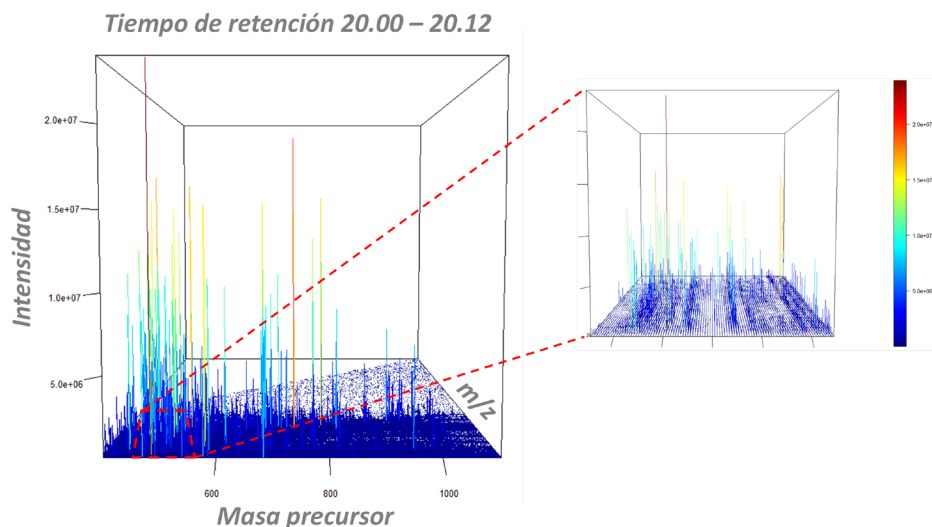
Como ya se ha comentado anteriormente, los espectros de fragmentación que se generan usando ventanas de aislamiento relativamente grandes (como las que se usan en SWATH) contienen mezclas complejas de fragmentos que provienen de varios iones precursores. Cuanto mayores sean las ventanas de aislamiento y la complejidad de la muestra mayor será la complejidad del espectro de fragmentación resultante, haciendo así más complicada la validación de múltiples secuencias en un mismo espectro.

Por tanto, los métodos DIA deberían idealmente trabajar con ventanas de aislamiento lo más pequeñas posible para minimizar el problema de la interferencia. Sin embargo, cuanto menor es el tamaño de la ventana de aislamiento mayor es el tiempo de ciclo (*cycle-time*), es decir, el tiempo que transcurre entre dos fragmentaciones de la misma ventana de aislamiento. Y por tanto, menor es el número de puntos que contiene el pico cromatográfico, lo que se traduce en una menor calidad de cuantificación. No obstante, es posible buscar un compromiso entre la anchura de la ventana de aislamiento y la calidad de la cuantificación, sobre todo si la cuantificación de las proteínas no se basa estrictamente en el pico cromatográfico.

Con este razonamiento en mente, en el laboratorio de Vázquez y cols. se ha desarrollado un método DIA que parte de una configuración más parecida a la de la proteómica shotgun e intenta aproximarse hacia la configuración de la proteómica dirigida. En este método, llamado DiS (o *data-independent scanning*), se alcanza un compromiso entre la anchura de la ventana de aislamiento, el tiempo de ciclo y la anchura del pico cromatográfico. El método DiS usa una ventana de aislamiento de 2 Da, que es idéntica a la que se usa en los métodos DDA que se emplean en proteómica shotgun, y que por tanto genera espectros de fragmentación menos complejos que pueden analizarse usando motores de búsqueda convencionales. Esta anchura de ventana de aislamiento genera el mismo nivel de interferencia que en un análisis convencional y requiere un tiempo de ciclo de fragmentación de unos 18 segundos (en equipos como el Q-Exactive HF). Considerando que la anchura media del pico cromatográfico en carreras medias/largas es de aproximadamente un minuto se obtienen una media de 3.3 espectros de fragmentación por pico cromatográfico. Hay que tener en cuenta que en la práctica, el tiempo de ciclo se acorta considerablemente (a menos de 12 segundos) cuando se analizan muestras complejas a una concentración óptima, ya que entonces la acumulación de iones máxima, establecida por el control automático de ganancia (AGC), se consigue antes de que transcurra el tiempo máximo de apertura de las lentes (IT) que llevan los iones al analizador.

Con el análisis DiS se obtiene información tetradimensional de los péptidos incluyendo la masa del ion parental, el tiempo de retención en columna, y la intensidad y masa de los fragmentos del espectro correspondiente (Figura 3A). El tratamiento de datos, como ya se ha comentado, se puede realizar con los mismos motores de búsqueda y parámetros que se usan en las aproximaciones DDA, por lo que no requiere ningún tipo de librería espectral ni base de datos específica, pudiendo emplear bases de datos convencionales. Por otro lado, la identificación de péptidos puede llevarse a cabo

A)



B)

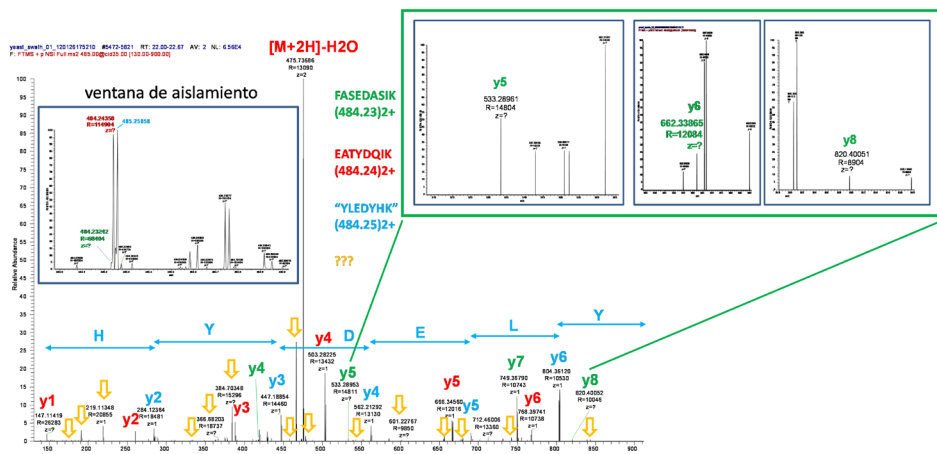


Figura 3. Esquema de las gráficas tetradsimensionales que se obtienen usando la tecnología DiS y características de la fragmentación. A) Distribución de señales de fragmentación correspondientes a los 350 iones precursores que barren todo el rango de masas en un intervalo de tiempo determinado. El recuadro a la derecha muestra una ampliación de la zona de bajo peso molecular. La escala colorimétrica determina la intensidad de los fragmentos. B) Espectro de fragmentación característico DiS en la ventana de m/z 485. En la ventana de aislamiento (recuadro a la izquierda) hay una mezcla de iones parentales con distintas intensidades (marcados con colores verde, rojo y azul). El espectro de fragmentación contiene fragmentos pertenecientes a todos los iones parentales de la ventana con distinta intensidad, dependiendo de la abundancia de cada uno de ellos. En los recuadros de la derecha se muestran ampliaciones donde se detectan fragmentos del ión parental de menor intensidad (ion verde). Los fragmentos marcados en amarillo son señales no asignadas a ninguno de los tres iones parentales de la ventana, que podrían corresponder a otros iones parentales adicionales.

estimando la FDR mediante la estrategia target/decoy usada en DDA. El método DiS permite confirmar las identificaciones por inspección de los espectros de fragmentación, gracias al reducido ruido de fondo característico de las fragmentaciones analizadas en alta resolución en el Orbitrap. De este modo se pueden validar secuencias pertenecientes a especies muy poco abundantes que dejan una intensidad muy reducida en el espectro de fragmentación, aunque suficiente para dar una buena cobertura de las series principales de fragmentación (Figura 3B). Como puede apreciarse en la figura, el espectro de fragmentación de la ventana correspondiente a la masa 485



incluye dos masas muy similares con una alta intensidad junto con una tercera masa de un ion cientos de veces menos intenso. Sin embargo, en el espectro se pueden derivar sin dificultad las tres secuencias e incluso quedan aún masas de fragmentos sin asignar que podrían corresponder a alguna otra secuencia. Esto hace que los análisis DiS tengan no solo una mejor sensibilidad, sino también un rango dinámico mucho más amplio que un análisis shotgun convencional en el que el ion minoritario raramente hubiera sido elegido para su fragmentación.

Los perfiles tetradimensionales característicos del análisis DiS permiten incrementar en un 25% el porcentaje de péptidos identificados en una muestra compleja cuando se comparan los resultados con los obtenidos mediante aproximaciones clásicas de proteómica DDA a la misma FDR. Por otro lado, el acoplamiento de esta metodología de trabajo con técnicas compatibles de separación de especies proteicas tales como la electroforesis en condiciones nativas (BNGE) permite ampliar el número de dimensiones analizadas hasta 6, si incluimos la movilidad electroforética del complejo proteico y el tejido o línea celular u organismo del que proviene. Esta tecnología, conocida como Blue-DiS, junto con todas las soluciones de análisis de datos desarrolladas en el grupo de Vazquez y cols., ya ha sido exitosamente aplicada en varios estudios biológicos. Hemos aplicado la metodología Blue-DiS para el análisis detallado del proteoma de los complejos mitocondriales en varios tejidos y líneas celulares de ratón. Gracias a esta tecnología hemos podido mapear cómo se estratifican los complejos multiproteicos que constituyen las diferentes maquinarias biológicas en los tejidos. Como se puede observar en la Figura 4, algunas de las funciones celulares se organizan en complejos multiproteicos de distinto tamaño molecular, mientras que hay otras que se concentran en la parte inferior del BNG y corresponden a proteínas libres. Algunas de las categorías funcionales de los distintos tejidos tienen una distribución similar en los perfiles (líneas de contornos en la Figura 4A), como es el caso de las proteínas que coordinan el metabolismo de ROS o la actividad de acetilasa. Sin embargo, otras categorías funcionales muestran una distribución específica de tejido, como es el caso de la biosíntesis de fosfolípidos (únicamente presente en cerebro), o el metabolismo de esteroides que solo puede detectarse en hígado (líneas de contornos en la Figura 4A). Una de las categorías de mayor tamaño molecular es la de la fosforilación oxidativa (OXPHOS) responsable de coordinar la respiración mitocondrial en la célula. En la Figura 4B se pueden apreciar los perfiles de contorno para cada una de las proteínas correspondientes a los complejos respiratorios mitocondriales en tres tejidos independientes (corazón, hígado y cerebro) y en dos líneas celulares (una línea control FC57, y una línea que carece de DNA mitocondrial Rho0). Como puede observarse, en la zona de muy alto peso molecular, las proteínas de los distintos complejos respiratorios se combinan en superestructuras que reciben el nombre de supercomplejos respiratorios, mientras que en la línea control negativo Rho0 no puede apreciarse ningún superensamblaje (Figura 4B). Este estudio nos ha permitido conocer con detalle cuáles son los supercomplejos que se ensamblan de forma natural en la célula, e incluso hemos sido capaces de predecir y descubrir tanto estructuras como factores de superensamblaje gracias a la aplicación de modelos matemáticos basados en los datos proteómicos.

Además, la tecnología Blue-Dis ya ha sido usada en varios trabajos publicados. Entre estos estudios podemos destacar un análisis detallado de cómo la expresión de algunas subunidades de complejo IV regula el ensamblaje de algunas superestructuras respiratorias mitocondriales(53), un estudio que determina con precisión cuáles son

los residuos más sensibles a daño oxidativo en las proteínas del complejo I y cómo se relaciona este fenómeno con la degradación y desensamblaje de superestructuras(14), el análisis de interacciones con la conexina 43 in mitocondria(54), así como estudios de la expresión de proteínas de distintos complejos respiratorios para la regulación de la dinámica de las crestas mitocondriales(55, 56). Finalmente, cabe indicar que esta tecnología de adquisición de datos tiene la ventaja adicional de que permite estudiar el comportamiento de cualquier proteína *in silico* y *a posteriori* en cualquier set de datos previamente adquirido, ya que esta metodología permite obtener mapas teóricos de todas las secuencias del proteoma dentro del rango de masas analizado y no requiere para su análisis del uso de librerías peptídicas.

Abreviaturas

AIF, fragmentación de todos los iones (*all ion fragmentation*)

CID, disociación por colisión (*collision-induced dissociation*)

DDA, adquisición dependiente de datos (*data-dependent acquisition*)

DIA, adquisición independiente de datos (*data-independent acquisition*)

DiS, barrido independiente de datos (*data independent scanning*)

ESI, electronebulización (*electrospray*)

FT-ARM, monitorización de todas las reacciones con transformada de Fourier (*Fourier transform-all reaction monitoring*)

HCD, disociación por colisión de alta energía (*higher-energy collisional dissociation*)

HDMS^E, MS^E de alta definición (*high definition MS^E*)

HRM, monitorización de hiperreacciones (*hyper reaction monitoring*, HRM)

LC-MS, cromatografía líquida acoplada a espectrometría de masas (*liquid chromatography coupled with mass spectrometry*)

m/z, relación masa-carga

MS, espectrometría de masas (*mass spectrometry*)

MS², espectrometría de masas en tándem

p2CID, disociación paralela inducida por colisión (*parallel collision-induced dissociation*)

PACIFIC, adquisición de precursores independiente del número de iones (*precursor acquisition independent from ion count*)

SRM, monitorización selectiva de reacciones (*selected reaction monitoring*)

TOF, tiempo de vuelo (*time-of-flight*)

UDMS^E, MS^E de ultra definición (*ultra definition MS^E*)

XDIA, adquisición independiente de datos ampliada (*extended data-independent acquisition*)

REFERENCIAS

1. 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(4):1785-93.
2. Doerr A. DIA mass spectrometry. *Nature Methods.* 2014;12:35.
3. Purvine S, Eppel JT, Yi EC, Goodlett DR. Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics.* 2003;3(6):847-50.
4. Venable JD, Dong MQ, Wohlschlegel J, Dillin A, Yates JR. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods.* 2004;1(1):39-45.
5. Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ, et al. Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem.* 2005;77(7):2187-200.
6. Ramos AA, Yang H, Rosen LE, Yao X. Tandem parallel fragmentation of peptides for mass spectrometry. *Anal Chem.* 2006;78(18):6391-7.
7. Panchaud A, Scherl A, Shaffer SA, von Haller PD, Kulasekara HD, Miller SI, et al. Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. *Anal Chem.* 2009;81(15):6481-8.
8. Geiger T, Cox J, Mann M. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol Cell Proteomics.* 2010;9(10):2252-61.
9. Carvalho PC, Han X, Xu T, Cociorva D, Carvalho Mda G, Barbosa VC, et al. XDIA: improving on the label-free data-independent analysis. *Bioinformatics.* 2010;26(6):847-8.
10. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis., (2012).
11. Weisbrod CR, Eng JK, Hoopmann MR, Baker T, Bruce JE. Accurate peptide fragment mass analysis: multiplexed peptide identification and quantification. *J Proteome Res.* 2012;11(3):1621-32.
12. Egerton JD, Kuehn A, Merrihew GE, Bateman NW, MacLean BX, Ting YS, et al. Multiplexed MS/MS for improved data-independent acquisition. *Nat Methods.* 2013;10(8):744-6.
13. Bruderer R, Bernhardt OM, Gandhi T, Miladinovic SM, Cheng LY, Messner S, et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics.* 2015;14(5):1400-10.
14. Guaras A, Perales-Clemente E, Calvo E, Acin-Perez R, Loureiro-Lopez M, Pujol C, et al. The CoQH2/CoQ Ratio Serves as a Sensor of Respiratory Chain Efficiency. *Cell Rep.* 2016;15(1):197-209.
15. Masselon C, Anderson GA, Harkewicz R, Bruce JE, Pasa-Tolic L, Smith RD. Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptide identification from mixtures. *Anal Chem.* 2000;72(8):1918-24.
16. Yi EC, Marelli M, Lee H, Purvine SO, Aebersold R, Aitchison JD, et al. Approaching complete peroxisome characterization by gas-phase fractionation. *Electrophoresis.* 2002;23(18):3205-16.
17. Souza GH, Guest PC, Martins-de-Souza D. LC-MS(E), Multiplex MS/MS, Ion Mobility, and Label-Free Quantitation in Clinical Proteomics. *Methods Mol Biol.* 2017;1546:57-73.
18. Distler U, Kuharev J, Navarro P, Tenzer S. Label-free quantification in ion mobility-enhanced data-independent acquisition proteomics. *Nat Protoc.* 2016;11(4):795-812.
19. Distler U, Kuharev J, Navarro P, Levin Y, Schild H, Tenzer S. Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat Methods.* 2014;11(2):167-70.
20. Smith RD, Loa JA, Barinaga CJ, Edmonds CG, Udseth HR. Collisional activation and collision-activated dissociation of large multiply charged polypeptides and proteins produced by electrospray ionization. *J Am Soc Mass Spectrom.* 1990;1(1):53-65.
21. 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(12):1367-72.
22. Williams ER, Loh SY, McLafferty FW, Cody RB. Hadamard transform measurement of tandem Fourier-transform mass spectra. *Anal Chem.* 1990;62(7):698-703.
23. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics.* 2010;26(7):966-8.
24. Reiter L, Rinner O, Picotti P, Huttenhain R, Beck M, Brusniak MY, et al. mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods.* 2011;8(5):430-5.
25. Rost HL, Rosenberger G, Navarro P, Gillet L, Miladinovic SM, Schubert OT, et al. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol.* 2014;32(3):219-23.
26. Wang J, Tucholska M, Knight JD, Lambert JP, Tate S, Larsen B, et al. MSPLIT-DIA: sensitive peptide identification for data-independent acquisition. *Nat Methods.* 2015;12(12):1106-8.
27. Rost HL, Sachsenberg T, Aiche S, Bielow C, Weissner H, Aicheler F, et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods.* 2016;13(9):741-8.
28. Schubert OT, Gillet LC, Collins BC, Navarro P, Rosenberger G, Wolski WE, et al. Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nat Protoc.* 2015;10(3):426-41.
29. Picotti P, Clement-Ziza M, Lam H, Campbell DS, Schmidt A, Deutsch EW, et al. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature.* 2013;494(7436):266-70.
30. Karlsson C, Malmstrom L, Aebersold R, Malmstrom J. Proteome-wide selected reaction monitoring assays for



the human pathogen *Streptococcus pyogenes*. *Nat Commun.* 2012;3:1301.

31. Schubert OT, Ludwig C, Kogadeeva M, Zimmermann M, Rosenberger G, Gengenbacher M, et al. Absolute Proteome Composition and Dynamics during Dormancy and Resuscitation of *Mycobacterium tuberculosis*. *Cell Host Microbe.* 2015;18(1):96-108.

32. Huttenhain R, Surinova S, Ossola R, Sun Z, Campbell D, Cerciello F, et al. N-glycoprotein SRMAtlas: a resource of mass spectrometric assays for N-glycosites enabling consistent and multiplexed protein quantification for clinical applications. *Mol Cell Proteomics.* 2013;12(4):1005-16.

33. Huttenhain R, Soste M, Selevsek N, Rost H, Sethi A, Carapito C, et al. Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci Transl Med.* 2012;4(142):142ra94.

34. Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol.* 2014;32(3):223-6.

35. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, et al. The PeptideAtlas project. *Nucleic Acids Res.* 2006;34(Database issue):D655-8.

36. Rosenberger G, Koh CC, Guo T, Rost HL, Kouvonen P, Collins BC, et al. A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci Data.* 2014;1:140031.

37. UniProt Consortium T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2018;46(5):2699.

38. Bern M, Finney G, Hoopmann MR, Merrihew G, Toth MJ, MacCoss MJ. Deconvolution of mixture spectra from ion-trap data-independent-acquisition tandem mass spectrometry. *Anal Chem.* 2010;82(3):833-41.

39. Tsou CC, Avtonomov D, Larsen B, Tucholska M, Choi H, Gingras AC, et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods.* 2015;12(3):258-64, 7 p following 64.

40. Li Y, Zhong CQ, Xu X, Cai S, Wu X, Zhang Y, et al. Group-DIA: analyzing multiple data-independent acquisition mass spectrometry data files. *Nat Methods.* 2015;12(12):1105-6.

41. Williams EG, Wu Y, Jha P, Dubuis S, Blattmann P, Argmann CA, et al. Systems proteomics of liver mitochondria function. *Science.* 2016;352(6291):aad0189.

42. Liu Y, Buil A, Collins BC, Gillet LC, Blum LC, Cheng LY, et al. Quantitative variability of 342 plasma proteins in a human twin population. *Mol Syst Biol.* 2015;11(1):786.

43. Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, et al. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol Cell Proteomics.* 2011;10(9):M110006353.

44. Kato BS, Nicholson G, Neiman M, Rantalainen M, Holmes CC, Barrett A, et al. Variance decomposition of protein profiles from antibody arrays using a longitudinal twin model. *Proteome Sci.* 2011;9:73.

45. Liu Y, Chen J, Sethi A, Li QK, Chen L, Collins B, et al. Glycoproteomic analysis of prostate cancer tissues by SWATH mass spectrometry discovers N-acyethanolamine acid amidase and protein tyrosine kinase 7 as signatures for tumor aggressiveness. *Mol Cell Proteomics.* 2014;13(7):1753-68.

46. Ortea I, Rodriguez-Ariza A, Chicano-Galvez E, Arenas Vacas MS, Jurado Gamez B. Discovery of potential protein biomarkers of lung adenocarcinoma in bronchoalveolar lavage fluid by SWATH MS data-independent acquisition and targeted data extraction. *J Proteomics.* 2016;138:106-14.

47. Lambert JP, Ivoisev G, Couzens AL, Larsen B, Taipale M, Lin ZY, et al. Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nat Methods.* 2013;10(12):1239-45.

48. Kulkarni S, Koller A, Mani KM, Wen R, Alfieri A, Saha S, et al. Identifying Urinary and Serum Exosome Biomarkers for Radiation Exposure Using a Data Dependent Acquisition and SWATH-MS Combined Workflow. *Int J Radiat Oncol Biol Phys.* 2016;96(3):566-77.

49. Nakamura K, Hirayama-Kurogi M, Ito S, Kuno T, Yoneyama T, Obuchi W, et al. Large-scale multiplex absolute protein quantification of drug-metabolizing enzymes and transporters in human intestine, liver, and kidney microsomes by SWATH-MS: Comparison with MRM/MS and HR-MRM/MS. *Proteomics.* 2016;16(15-16):2106-17.

50. Chang RY, Etheridge N, Nouwens AS, Dodd PR. SWATH analysis of the synaptic proteome in Alzheimer's disease. *Neurochem Int.* 2015;87:1-12.

51. Collins BC, Hunter CL, Liu Y, Schilling B, Rosenberger G, Bader SL, et al. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat Commun.* 2017;8(1):291.

52. Rosenberger G, Bludau I, Schmitt U, Heusel M, Hunter CL, Liu Y, et al. Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nat Methods.* 2017;14(9):921-7.

53. Cogliati S, Calvo E, Loureiro M, Guaras AM, Nieto-Arellano R, Garcia-Poyatos C, et al. Mechanism of super-assembly of respiratory complexes III and IV. *Nature.* 2016;539(7630):579-82.

54. Denuc A, Nunez E, Calvo E, Loureiro M, Miro-Casas E, Guaras A, et al. New protein-protein interactions of mitochondrial connexin 43 in mouse heart. *J Cell Mol Med.* 2016;20(5):794-803.

55. Quintana-Cabrera R, Quirin C, Glytsou C, Corrado M, Urbani A, Pellattiero A, et al. The cristae modulator Optic atrophy 1 requires mitochondrial ATP synthase oligomers to safeguard mitochondrial function. *Nat Commun.* 2018;9(1):3399.

56. Glytsou C, Calvo E, Cogliati S, Mehrotra A, Anastasia I, Rigoni G, et al. Optic Atrophy 1 Is Epistatic to the Core MICOS Component MIC60 in Mitochondrial Cristae Shape Control. *Cell Rep.* 2016;17(11):3024-34.

MALDI -TOF imaging mass spectrometry (MALDI-IMS)

*Ibon Iloro¹, Beatriz Rocha², Francisco Blanco²,
Cristina Ruiz-Romero² y Felix Elortza¹*

¹ Plataforma de Proteómica, CIC bioGUNE, ProteoRed-ISCIII, CIBERehd, Parque Científico Tecnológico de Bizkaia, Edificio 800, 48160, Derio.

² Unidad de Proteómica-ProteoRed/ISCIII, Grupo de Investigación de Reumatología (GIR), CIBER-BBN, Instituto de Investigación Biomédica de A Coruña (INIBIC-CHUAC), A Coruña.

1. Introducción.

Actualmente, las técnicas proteómicas permiten los estudios a gran escala del contenido proteico/peptídico de muestras biológicas de cualquier tipo, sean biofluidos, células o tejidos, y pueden abarcar desde organismos completos a complejos subcelulares. Estas técnicas representan el siguiente paso a la genómica y la transcriptómica, y son ahora una punta de lanza en la investigación biomédica, como estudios patológicos, diagnóstico, pronóstico, prognosis de enfermedades, etc. (1-5). El conocimiento detallado de la distribución y localización proteica a lo largo de un tejido puede ofrecer información vital para la investigación clínica. Lamentablemente, la inmensa mayoría de los métodos proteómicos (electroforesis, LC-MS...) requieren la rotura y homogenización de los tejidos, destruyendo la relación entre las regiones morfológicas de los tejidos y la localización específica de proteínas/péptidos. A finales de los noventa, el profesor Caprioli fue uno de los pioneros en introducir el análisis proteómico directo mediante imagen sobre cortes de tejidos intactos utilizando técnicas de MALDI-MS. Es sólo a partir de una metodología desarrollada en 1997 por el profesor Caprioli, entre otros grupos, utilizando técnicas de MALDI-MS (6-8), cuando ha sido posible el análisis proteómico directo mediante imagen sobre cortes de tejidos intactos. Este método es capaz de detectar y localizar un gran número de señales de manera directa sobre cortes de tejido en una única adquisición. Además, la técnica, denominada comúnmente MALDI-IMS o MALDI imaging, permite el análisis y mapeo no solamente de proteínas/péptidos (9-11), sino también de otros tipos de analitos, como lípidos (12), metabolitos (13) o fármacos (5).

Desde su introducción, la tecnología MALDI-TOF (MALDI-time-of-flight) MS se ha convertido en una importante herramienta, especialmente en proteómica (14), que cuenta con una serie de ventajas respecto a otros métodos de espectrometría de masas. Destaca sobre todo por su robustez, diseño sencillo, una amplia ventana de detección, y la gran sensibilidad del analizador TOF. Además, la interpretación de los espectros es muy sencilla, ya que genera principalmente iones monocargados (por lo que la



extrapolación m/z – Da es directa), y en el rango habitualmente utilizado en el análisis de péptidos/proteínas, la interferencia de las matrices comúnmente utilizadas es baja. Aparte de estas ventajas, la técnica MALDI-TOF MS está especialmente bien diseñada para realizar estudios de imagen MALDI-IMS. Así, las muestras de tejido se pueden fijar de manera rutinaria sobre portaobjetos de vidrio, tal y como se realiza en estudios de análisis de imagen por microscopía óptica, sobre los que se pulveriza directamente la matriz MALDI, que finalmente pueden incorporarse, mediante una placa adaptadora, a un equipo MALDI-TOF convencional, sin necesidad de realizar ninguna modificación al equipo de masas.

Por todo esto, la técnica MALDI-IMS ofrece considerables ventajas en el análisis directo de los cambios fisiopatológicos de muestras de tejido, manteniendo siempre el contexto histopatológico, vital en muchas circunstancias. Permite el mapeado simultáneo de cientos de señales de péptidos/proteínas a lo largo de una sección de tejido, mostrando la distribución espacial de cada una de estas señales, así como de su abundancia relativa en cada punto del tejido. Además, las imágenes del mapeo de estos analitos pueden correlacionarse directamente con la arquitectura tisular subyacente sin ningún tipo de distorsión geométrica, permitiendo la aparición de una nueva área de análisis clínico que ha pasado a denominarse “Histología Molecular” (15, 16).

Lamentablemente, debido en muchos casos a la relativa novedad de esta técnica, el MALDI-IMS presenta en la actualidad una serie de limitaciones. Algunas de estas limitaciones pueden incluso que sean de difícil solución hasta que no aparezcan actualizaciones substanciales de tipo instrumental. Así, existen limitaciones en forma de resolución lateral alcanzada, sobre todo en MALDI-IMS de proteínas/péptidos (no tanto con otros analitos, como lípidos), no llegándose aún a nivel celular (40/50 μm es el límite aproximado en IMS de proteína/péptido, no así en lípido, que ha llegado a resoluciones de 10 μm de manera rutinaria). También en estudios proteicos, el rango de masas es limitado (hasta 70 – 100 kDa de forma rutinaria), y las señales son de difícil identificación. No obstante, algunas de las limitaciones iniciales de la técnica se han ido solventando con el tiempo. Por ejemplo, los largos tiempos de adquisición de las imágenes se han ido solventando con la aparición de láseres ultra-rápidos de última generación (17), así como de la electrónica asociada. La reproducibilidad inter muestral también ha mejorado ostensiblemente con la aparición y posterior desarrollo de los depositadores de matriz automatizados (18).

Pero a pesar de estas limitaciones, la técnica MALDI-IMS tiene actualmente muchas aplicaciones potenciales en análisis clínico y biomedicina, como la histopatología (19), detección precoz (20), medicina individualizada (21), seguimiento de la eficacia terapéutica (22) y estudios de toxicidad (23), entre otros. Mediante el MALDI-IMS, los patólogos pueden correlacionar la distribución de señales específicas con estructuras significativamente patológicas (24-26). Recientemente, algunos investigadores han recreado imágenes moleculares tridimensionales de tejidos mediante la técnica MALDI-IMS y co-registrado estas imágenes volumétricas con resultados provenientes de otras técnicas de imagen como la resonancia magnética nuclear (MRI) (27-29).

Esta es por lo tanto una técnica con una enorme potencialidad, que abre las puertas de un nuevo área de conocimiento denominada “Histología molecular”, con interesantes aplicaciones en biomedicina y farmacología.

2. Metodología estándar en MALDI-IMS.

La preparación de muestra es una etapa crítica en MALDI-IMS (30). La mayor parte de la calidad final y reproducibilidad de un análisis de imagen va a venir dado por la calidad de la preparación.

De manera rutinaria, el material de partida en un análisis consiste en cortes finos de tejido montados y fijados sobre un portaobjetos de microscopía óptica (estos portaobjetos pueden ser convencionales o con una superficie conductiva, en función del tipo de fuente MALDI). Dependiendo del tipo de analito a examinar, el tejido debe ser tratado para eliminar o minimizar la presencia de sustancias interferentes, y preparar la muestra para recibir la matriz MALDI (23, 31, 32). Tras este paso de pre-tratamiento, es cuando se aplica la matriz MALDI apropiada al analito a estudiar, ya sea de manera manual o automatizada. Tras esta aplicación, el portaobjetos con los tejidos puede ser insertado en el instrumento MALDI-TOF. El espectrómetro de masas rastrea los tejidos con el láser generando una matriz de disparos, grabando la distribución espacial de los analitos moleculares, ya sean péptidos, proteínas o pequeñas moléculas para cada posición. Finalmente, un software de procesamiento de imagen importa los datos crudos de masas y los procesa con el fin de obtener la visualización de los componentes moleculares y su co-registro con la imagen óptica de la muestra. Para comparación entre diferentes condiciones experimentales (análisis de dos o más tejidos) o de diferentes regiones (en un mismo tejido), se emplean paquetes estadísticos específicos. El diagrama de flujo de todo el procedimiento se muestra en la Figura 1.

2.1. Preservación de muestra y almacenamiento

En esta técnica es crucial que el tejido extraído se preserve lo más rápidamente posible para reducir la degradación molecular y conservar la integridad espacial. Usualmente se utilizan técnicas de criopreservación (33) a bajas temperaturas para detener de manera eficiente las actividades químicas y enzimáticas que puedan dañar el material de partida. El método de criopreservación utilizado ha de ser correctamente formulado para llegar a las bajas temperaturas de corte necesarias (-20 a -30 °C) evitando la formación de cristales de hielo que puedan dañar las estructuras. Hay que tener en cuenta además que a veces el uso de crioprotectores pueden causar alteraciones químicas. Por todo esto, el protocolo más utilizado en la mayoría de estudios de MALDI-IMS es el congelado rápido mediante nitrógeno líquido (LN₂) inmediatamente tras la extracción del tejido (34). Tras este paso, las muestras pueden almacenarse a -80 °C hasta proceder a su corte previo al análisis por MALDI-IMS.

Un problema común en este punto es que los tejidos obtenidos de cirugías o procedimientos diagnósticos se preservan de manera rutinaria en bloques mediante fijación por formalina y embebido en parafina (formalin-fixed, paraffin-embedded; FFPE). Esto es debido a que estos bloques son tremendamente estables y pueden almacenarse a temperatura ambiente, sin ningún otro requisito. En la actualidad se están desarrollando protocolos y llevando a cabo estudios sobre este tipo de tejidos parafinados FFPE (35, 36). A pesar de que el procedimiento de embebido preserva la morfología celular,



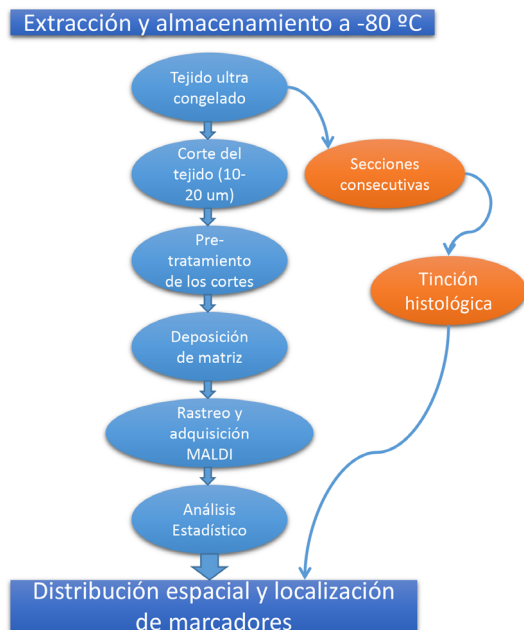


Figura 1. Representación esquemática del protocolo MALDI-IMS para muestra crio-preservada. Las muestras son recogidas e inmediatamente congeladas a -80 °C. Posteriormente, el tejido es embebido en una solución de corte (OCT, agua...) y cortado usando un criostato. Posteriormente, en función del analito a analizar, el tejido puede necesitar un pretratamiento previo (como la deslipidización, por ejemplo, en el caso de MALDI-IMS de proteínas o péptidos). Una vez adquiridos los espectros, los datos son procesados y analizados estadísticamente. Se pueden realizar secciones consecutivas para realizar tinción histológica, o bien esta tinción se puede realizar al mismo tejido sometido a MALDI-IMS, tras su análisis.

jido. A continuación, para revertir el entrecruzamiento de las proteínas inducido por la fijación con formalina, las muestras se calientan en ácido cítrico durante 30 min a 98 °C. Actualmente, gracias a la optimización de este protocolo se pueden medir péptidos y metabolitos en muestras FFPE mediante MALDI-IMS (42).

2.2. Corte de muestras

Debido a la naturaleza del material criopreservado a temperatura de hielo seco (-50 °C), el procedimiento de corte más apropiado es el corte por criostato o criotomo (que consiste básicamente en un micrótopo por rotación en una cámara a baja temperatura). Existen diversos protocolos de embebido para asegurar la integridad del tejido durante el proceso de corte. Habitualmente, estos protocolos hacen uso del polímero denominado "Optimal Cutting Temperature" (OCT) o Tissue-Tek, gelatina (43) o agarosa (44) como solución de embebido. Nuestra recomendación es evitar en la medida de lo posible estas soluciones, ya que pueden causar supresión iónica (especialmente el OCT), o alterar la integridad química de la muestra al esparcirse sobre la muestra durante el proceso de corte. Para algunos tejidos, puede intentarse el corte directo sin embebido, aunque una solución más global consiste en embeber las muestras en hielo. El protocolo es muy sencillo, y consiste en colocar la muestra sobre el disco de corte del criotomo dentro de la cámara (que estará a entre -20 y -30 °C) e ir añan-

el tratamiento con formalina resulta en la formación de puentes de metileno entre los aminoácidos de las proteínas (37), creando una malla de proteína entrecruzada que hay que romper utilizando una proteasa para que la muestra pueda analizarse mediante MALDI-IMS. Estos métodos requieren la adopción de pasos adicionales en la preparación de la muestra para revertir el entrecruzamiento entre las proteínas y eliminar la parafina que es incompatible con MS (38). En los últimos años, se han desarrollado diferentes protocolos para la desparafinación y eliminación del entramado proteico en los tejidos FFPE para su análisis mediante MALDI-IMS (39-41). Estos protocolos sugieren que las muestras deben calentarse durante 1 hora a 60 °C y transferirse directamente a una solución de xileno (100%, dos lavados de 5-10 min), etanol (70%, tres lavados de 2 min) y agua (dos lavados de 5 min) para conseguir una buena desparafinación del te-



Figura 2. Fotografía y diagrama esquemático que muestra la deposición de agua para formar un recubrimiento de hielo como soporte de corte alrededor del tejido. Este procedimiento asegura un corte óptimo, evitando el uso de soluciones de embebido problemáticas, como puede ser el OCT. Es necesario añadir diversas capas de agua hasta que el tejido queda completamente embebido.

diendo agua destilada sobre el tejido de manera que esta se congela formando un recubrimiento de hielo que soporta al tejido (Figura 2). Este recubrimiento de hielo es suficiente para mantener la integridad del tejido durante el corte utilizando una cuchilla estándar de criotomo a una temperatura menor a -15°C (20, 45). El grosor de las secciones debe ser de entre 10 y 20 μm . En gran parte de las publicaciones, se utiliza de manera rutinaria secciones de 12 - 14 μm . Secciones más finas de 10 μm son difíciles de manipular y montar, mientras que se ha reportado que secciones más gruesas de 20 μm pueden afectar a la posterior señal en el espectrómetro de masas (debido a que las secciones de tejido actúan como aislantes eléctricos) (46).

Los cortes se montan sobre portaobjetos de vidrio convencionales, en algunos casos de fuentes MALDI atmosféricas (AP-MALDI), o más comúnmente, sobre portaobjetos de vidrio recubiertos de ITO (indium-tin oxide). Este tipo de portaobjetos es conductivo, necesario para permitir la ionización en la mayor parte de fuentes MALDI. Como también son transparentes, estos portaobjetos pueden utilizarse en microscopía óptica convencional, permitiendo la posterior correlación de la imagen MALDI-IMS con la imagen óptica (15). Cabe recordar que estos portaobjetos ITO solo están recubiertos generalmente por una de las caras, que es por la que hay que montar las muestras de tejido cortadas. Además, las propiedades de esta superficie son diferentes a las del vidrio, lo que puede dar problemas de fijación del tejido sobre el portaobjetos (algunos tipos de tejido presentan una baja fijación en este tipo de superficies).

Normalmente, para producir una suficiente fijación, basta con calentar con la mano durante unos segundos el portaobjetos con el tejido. No obstante, para algunos tipos de tejido esto no es suficiente necesitándose algo más de calor. A modo de ejemplo, nosotros utilizamos el calor de un centro de parafinado (60°C durante 30 segundos) para fijar correctamente la mayor parte de tejidos (Figura 3). Pero hay ciertos tipos de muestra que presentan una casi nula adsorción a los portaobjetos recubiertos por ITO. En estos casos, es necesario un tratamiento previo a la superficie, como puede ser la poli-lisinación (recubrimiento de poli-lisina), pudiéndose utilizar cualquiera de los métodos disponibles.

Para obtener un correcto solapamiento de la imagen por MALDI-IMS y la imagen óptica, sería necesario utilizar el mismo tejido para obtener ambas imágenes. Normalmente, tras coleccionar la imagen mediante MALDI-IMS, la matriz es eliminada mediante





Figura 3. Fotografía mostrando un ejemplo de "dispositivo de fijación". Consiste en un centro de parafinizado que proporciona algo de calor (60 °C aprox) para fijar el tejido sobre la superficie de ITO del portaobjetos. Hemos usado este dispositivo ya que es de uso común en laboratorios de preparación de cortes. Los cortes se dejan secar por un tiempo de unos 30 s aprox.

baño en etanol, y los tejidos se tiñen (rutinariamente con Hematoxilina-eosina, o HE) para obtener las imágenes ópticas (28). Lamentablemente, a veces esto no es posible debido a la importante degradación que sufre el tejido bajo la irradiación del láser. Por esto, es una buena práctica recoger cortes consecutivos/adyacentes sobre diferentes portaobjetos: uno para realizar la imagen MALDI-IMS, y el otro para realizar la tinción para la imagen óptica. También existe un protocolo para realizar una tinción óptica previa a la colección de la imagen MALDI-IMS (15), aunque se ha visto que la tinción por HE estándar no es compatible con el análisis posterior de MALDI-IMS, por lo que solo es válido para otro tipo de tinciones.

2.3. Preparación de cortes

Tras el corte, las muestras han de desecarse previo a su almacenado. La desecación se realiza mediante la colocación del portaobjetos en una cámara desecadora conectada a una línea de vacío durante 20 – 30 minutos. Para las imágenes de proteína/péptido, es necesario también un protocolo de deslipidizado, ya que los lípidos interferirán en la ionización de las proteínas/péptidos, reduciendo sus señales. Hay varios protocolos de deslipidizado: los dos más utilizados son el lavado por escalera de etanol (de 70% a 100% de etanol, en tres pasos de 30 segundos cada uno) y lavado de Carnoy (28). Para muestras con un alto contenido lipídico, se recomienda el protocolo de Carnoy. Este protocolo consiste en varias etapas: baño en 70% etanol por 1 min, baño en 100% etanol por 1 min, baño en solución Carnoy (etanol:cloroformo:acético, 6:3:1) por 3 min, baño en 100% etanol por 1 min, inmersión en 0.2% TFA por 30 seg y baño en 100% etanol por 1 min. Tras el deslipidizado, las muestras han de desecarse de nuevo durante 30 min para poderse analizar mediante MALDI-IMS o almacenarse (en desecador conectado a vacío por un máximo de 2 semanas).

2.4. Digestión de las muestras

Otra opción para el análisis MALDI-IMS de proteínas es la digestión enzimática del tejido a analizar. La digestión de las muestras produce péptidos característicos que nos permitirán identificar y localizar las proteínas correspondientes. Se deben optimizar diferentes parámetros, incluido el tampón utilizado, la concentración de la enzima o el tiempo y la temperatura de incubación de acuerdo al tipo de tejido que se quiera analizar. Entre las diferentes enzimas, la tripsina es la enzima más comúnmente empleada en MALDI-IMS. Otras enzimas como la quimotripsina, elastasa, pepsina o N-glicosidasa F también se han utilizado satisfactoriamente en estudios de MALDI-imaging, ya sea solas o en combinación con la tripsina (47). Especialmente para muestras FFPE, la adición de detergentes compatibles con MS, como el octilglucósido, mejora la detección e identificación de proteínas tales como histonas (48). Estos detergentes rompen las membranas celulares y exponen a las proteínas a una mejor interacción con la tripsina. Además, se ha demostrado que surfactantes como RapiGest (Waters) mejoran la eficiencia de la digestión sin interferir con el análisis de MS. La digestión de las muestras se realiza generalmente a 37°C durante un período de incubación de 16-18 h, manteniendo los portaobjetos en una atmósfera húmeda (generalmente en una caja con 50% metanol) para evitar la deshidratación del tejido/enzima. Finalmente, las muestras se cubren con matriz.

2.5. Deposición de matriz

La deposición de la matriz constituye el paso más delicado en un experimento MALDI-IMS. Hay disponibles varias matrices MALDI para este tipo de estudios, así como diferentes tipos de soluciones de dilución para estas matrices. Para el análisis combinado de proteína/péptido, la matriz SA (ácido sinapínico) proporciona las mejores señales en el rango de 2000 – 30000 Da, típico de estudios MALDI-IMS (44), mientras que la matriz α -cyano-4-hydroxycinnamic acid (HCCA) es mejor para estudios a nivel peptídico (1000 – 8000 Da) (49). La concentración de matriz varía entre 10 y 20 mg/ml, en una solución de un 70 a un 90% de acetonitrilo (ACN) con un 0.3% de ácido trifluoroacético (TFA). De manera rutinaria, se suele utilizar matriz SA (15 mg/ml, en un 70% ACN/0.3% TFA). En la Tabla 1 se resumen las principales matrices utilizadas actualmente en MALDI-IMS.

La deposición de la matriz se puede realizar de manera manual o automatizada. Los métodos manuales emplean bien aerógrafos estándar, o nebulizadores TLC (Figura 4A) (49). Si bien éste es un método simple y rápido, tiene una desventaja fundamental, que es la mínima reproducibilidad debido a la imposibilidad de controlar el tamaño de las gotas, su homogeneidad, y su deposición. Debido a esto, es siempre mejor utilizar un método automatizado de deposición.

Básicamente, hay dos tipos de dispositivos de deposición automatizada: Spotters y Sprayers:

- Los spotters (por ejemplo, Shimadzu ChiP or Leaptec TmiD) aplican diminutas gotas de la solución de matriz sobre el tejido, una a una formando una trama. Su resolución depende del tamaño de gota (a menudo unos 50 μ m), aunque su



disposición es muy precisa y uniforme.

- Los sprayers (Leaptec TM Sprayer, Bruker Imageprep or Ideko/bioGUNE Langartech(18), Figura 4B) pulverizan la solución de matriz sobre el tejido, de manera que el tamaño de gota es mucho menor, permitiendo resoluciones por debajo de 100 μm . Cada dispositivo sprayer aplica la matriz de maneras diferentes (vaporización, nebulización...), por lo que cada uno tiene protocolos propios para alcanzar una deposición apropiada para el análisis.

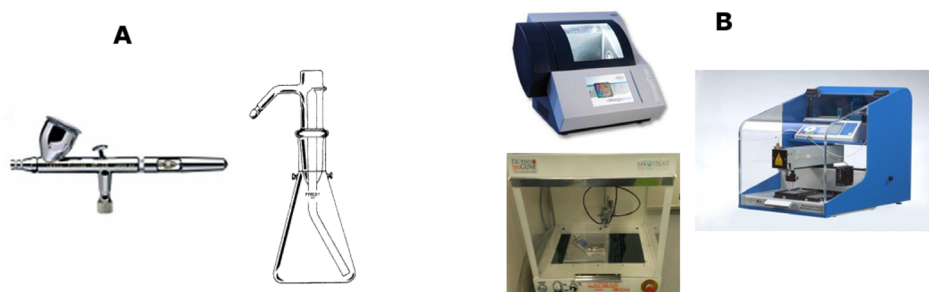


Figura 4. Tipos diferentes de nebulizadores que se pueden usar en la fase de spray. Los dos primeros (A, de izquierda a derecha) se usan manualmente: pistola de modelismo y spray de TLC (thin layer chromatography). Desaconsejamos el uso de estos dispositivos, por carecer de reproducibilidad y proporcionar imágenes poco resolutivas. Los dispositivos automatizados (B) ofrecen buena reproducibilidad y resolución, pero, lógicamente, a un mayor coste.

En estudios de proteína/péptido por MALDI-IMS, los analitos han de ser extraídos del tejido, mezclados con la matriz y co-cristalizados con ella. La extracción se lleva a cabo durante la fase acuosa, por lo que si el tamaño de gota es muy grande, o si las gotas se fusionan entre ellas, se producirá un exceso de difusión, perdiéndose la información espacial. Por otro lado, si la fase acuosa dura muy poco, debido a una rápida evaporación o por un tamaño de gota excesivamente pequeño, la extracción será insuficiente, lo que producirá una señal muy poco intensa. Por todo esto, hay un balance entre las etapas de humedecido y secado al aplicar la matriz, al cual se ha de llegar para obtener una buena señal de los analitos por un lado, y una buena resolución lateral por el otro. Llegar a este balance puede requerir de múltiples pruebas y errores.

2.6. Obtención de imágenes

Como ya se ha mencionado, no hay necesidad de equipamiento MALDI-TOF especializado o especialmente acondicionado para realizar un análisis de MALDI-IMS. Cualquiera de los instrumentos MALDI-TOF disponibles en la actualidad se pueden usar para realizar este tipo de análisis (aunque ya se están presentando los primeros MALDI-TOF especialmente diseñados para realizar análisis de MALDI-IMS). El único requisito que deben cumplir es que puedan realizar una adquisición automatizada, y disponer de un software de visualización y rastreo. Disponer de un láser rápido (con alta tasa de repetición) es también muy importante para minimizar los tiempos de adquisición. Usualmente, todo el proceso de adquisición es llevado a cabo de manera automatizada por el software proporcionado por el fabricante. Thermo, Bruker y Sciex, entre otros, han desarrollado este tipo de software para sus instrumentos. El software de adquisición crea una trama de rastreo sobre el tejido (Figura 5A) sobre el cual se realiza la adquisición vía ionización MALDI, grabándose un espectro completo por

Matriz	Abreviatura	Moléculas diana	Tipo	Referencias
Ácido alfa-ciano-4-hidroxicinámico	CHCA	Péptidos, proteínas de bajo PM, glicopéptidos	Cristalina	(50)
Ácido 3,5-dimetoxi-4-hidroxicinámico	SA	Proteínas de alto PM, péptidos, lípidos	Cristalina	(51)
Ácido 2,5-dihidroibenzoico	DHB	Azúcares, proteínas, péptidos, nucleótidos, fosfolípidos, fármacos	Cristalina	(52)
2,6-dihidroxiacetofenona	DHA	Proteínas, fosfolípidos, fármacos	Cristalina	(53)
CHCA/anilina		Péptidos	Iónica sólida	(54)
CHCA/N,N-dimetilanilina		Péptidos	Iónica sólida	(54)
CHCA/2-amino-4-metil-5-nitropiridina		Péptidos	Iónica líquida	(54)
CHCA/n-butilamina	CHCAB	Péptidos, lípidos	Iónica líquida	
2,4,6-Trihidroxiacetofenona	THAP	Nucleótidos, lípidos	Cristalina	(55)
Ácido 3-hidroxicolínico	3-HPA	Nucleótidos, péptidos, glicoproteínas	Cristalina	(55, 56)
9-aminoacridina	9-AA	Fosfolípidos, metabolitos	Cristalina	(57)
4-paranitroanilina	PNA	Lípidos	Cristalina	(58)
N-(1-Naftil)-etilendiamina diclorohidrato	NEDC	Iones metálicos, metabolitos, lípidos	Cristalina	(59, 60)
Norharman		Lípidos	Cristalina	(61)

Tabla1. Tipos de matrices utilizadas actualmente en MALDI-IMS.

cada punto de rastreo. Las imágenes se construyen representando la intensidad de cada señal respecto a la posición relativa de la señal en el tejido (Figura 5B). Las intensidades de cada señal para un valor específico m/z se visualizan como una escala de color (o gradiente de grises) sobre la totalidad del área del tejido. Debido a que los analitos son directamente ionizables mediante el MALDI, no son necesarios métodos de etiquetado o derivatización, por lo que la distribución molecular representada no posee ningún sesgo.

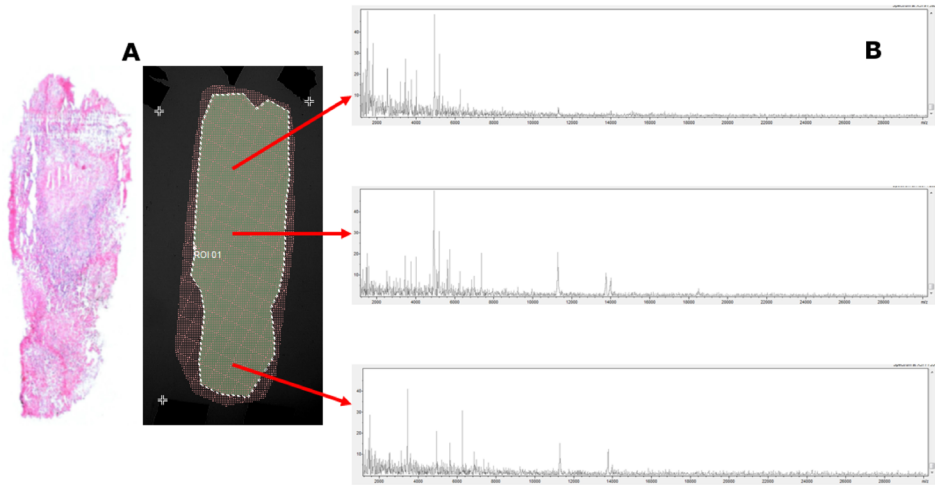


Figura 5. El software de MALDI-IMS (Fleximaging 3.0 en este caso) debe tener herramientas para poder superponer la imagen óptica del tejido con la matriz de disparos a realizar sobre el tejido (**Figura 5A**). Por cada punto de muestreo se obtiene un espectro completo, en el rango de masas seleccionado, que forma un pixel para cada masa (**Figura 5B**). Las imágenes obtenidas consisten en la intensidad de cada masa versus su posición relativa en el tejido.



2.7. Eliminación de la matriz y tinción de las muestras

Las imágenes de MALDI-IMS resultantes deben correlacionarse con una imagen óptica que proporcione la información histológica. En esta etapa hay dos posibilidades alternativas:

- Una posibilidad es usar el mismo tejido ya rasterizado mediante MALDI-IMS para realizar la tinción óptica. Para poder hacer esto, la matriz ha de eliminarse por completo antes de realizar la tinción. La matriz puede eliminarse fácilmente mediante lavado en etanol al 95% en una placa Petri sometida a agitación suave durante 1 min. Tras esto, se realiza un protocolo estándar de tinción (usualmente HE) para obtener la información histológica sobre la que superponer la imagen molecular MALDI-IMS. Este es el mejor método para la superposición, ya que consiste en el mismo tejido para ambas imágenes. Lamentablemente, en algunos casos, debido a la naturaleza del tejido, o a elevados niveles de irradiación láser necesarios para producir la ionización durante la adquisición, el tejido queda gravemente dañado, y no es posible realizar correctamente la tinción, o la imagen resultante es de muy baja calidad.
- La otra alternativa, que solventa el anterior problema, consiste en teñir un tejido obtenido en un corte adyacente al tejido rastreado por MALDI-IMS, y montado sobre un portaobjetos diferente durante el crioseccionado. Por otro lado, en esta aproximación ambas imágenes son el resultado de diferentes tejidos, por lo que la superposición puede no ser perfecta. Este método sería válido para estructuras histológicas suficientemente grandes, aunque siempre existirán diversas imprecisiones en estructuras más pequeñas.

De manera general, desde este artículo recomendamos realizar ambas aproximaciones de manera simultánea, para poder disponer siempre de la imagen óptica a correlacionar.

2.8. Procesado de datos y análisis estadístico

Una vez colectados los espectros, las imágenes se pueden obtener de manera inmediata, mostrándose la distribución espacial de cada analito (Figura 6). Normalmente, puede hacerse seguimiento de un elevado número de señales de proteína/péptido (100–300 señales, por término medio, Figura 6) en una ventana de masas típica de 1000 a 30000 thomson (Th). Para poder realizar un estudio de las estructuras moleculares subyacentes es necesario reducir toda esta complejidad. La técnica más utilizada para realizar esta reducción de información presente es el análisis de componentes principales o PCA (62). El PCA permite la proyección de datos multidimensionales en un menor número de dimensiones reteniendo toda la información esencial del sistema. Para ello realiza una transformación lineal del Sistema de coordenadas original de manera que la primera coordenada del nuevo sistema de coordenadas apunta en la misma dirección a la mayor varianza de los datos del sistema, y así sucesivamente hacia menores varianzas. Así, el PCA reemplaza las variables originales por unas nuevas variables, llamadas Componentes Principales (PC), con varianzas en orden decreciente (Figura 7). En una preparación de muestra homogénea, los resultados de

la PCA deberían corresponder con las estructuras histológicas presentes en el tejido, de manera que si la muestra posee diversas estructuras, estas deberían aparecer reflejadas en la PCA resultante.

Si bien el PCA es capaz de reducir la dimensionalidad de un set de espectros, no puede clasificar los diferentes grupos de espectros presentes. Esto es útil para ver las semejanzas de las estructuras a nivel molecular. Una forma de clasificación no supervisada de los espectros de masa por su similitud espectral en un análisis de imagen se consigue mediante la Clusterización Jerárquica (63). Estos resultados se representan a modo de dendrograma conteniendo la totalidad de espectros presen-

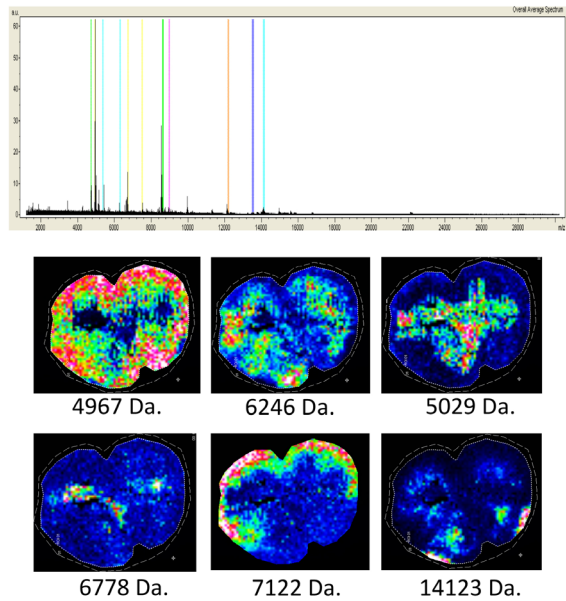


Figura 6. Una vez colectados todos los espectros, puede pasarse a visualizar las imágenes adquiridas. Se muestran las intensidades de señal específicas para cada valor m/z , utilizando una escala de color o de grises sobre toda el área de tejido analizada, mostrando por lo tanto la distribución de cada analito. La ventana de masas habitual en experimentos de MALDI-IMS de proteínas/péptidos va de 1000 a 30000 thomson, pudiéndose en ocasiones ampliar un poco más.

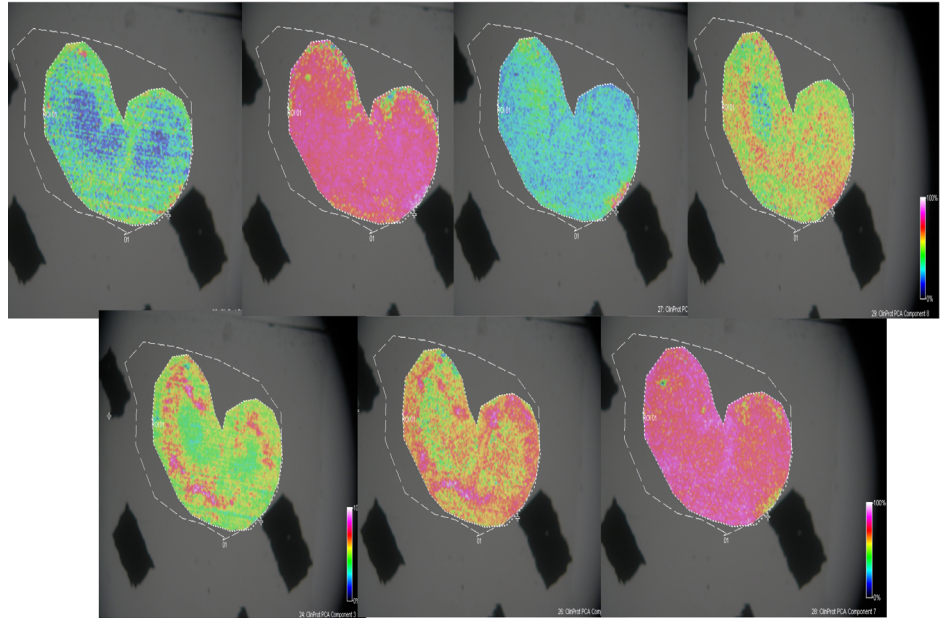


Figura 7. Para realizar un análisis lo menos sesgado posible, es común realizar algún tipo de análisis no supervisado, como por ejemplo el PCA (Análisis de componentes principales). El PCA es un método muy usado para describir la varianza del set de datos reduciendo su dimensionalidad. En esta figura se muestran las 7 componentes principales mas informativas de un corte renal. En este caso, los colores muestran las diferentes regiones encontradas por el sistema.

tes. Cada rama de este dendrograma puede considerarse como una clase espectral, de forma que cada nodo refleja una estructura molecular determinada (Figura 8). Dos nodos muy próximos compartirán la mayoría de las señales espectrales, por lo que serán similares a nivel molecular, mientras que dos nodos muy separados serán muy diferentes a nivel molecular. La dimensionalidad del dendrograma puede ser muy elevada, por lo que suele ser recomendable utilizar los datos del PCA para realizar este análisis de clusterizado.

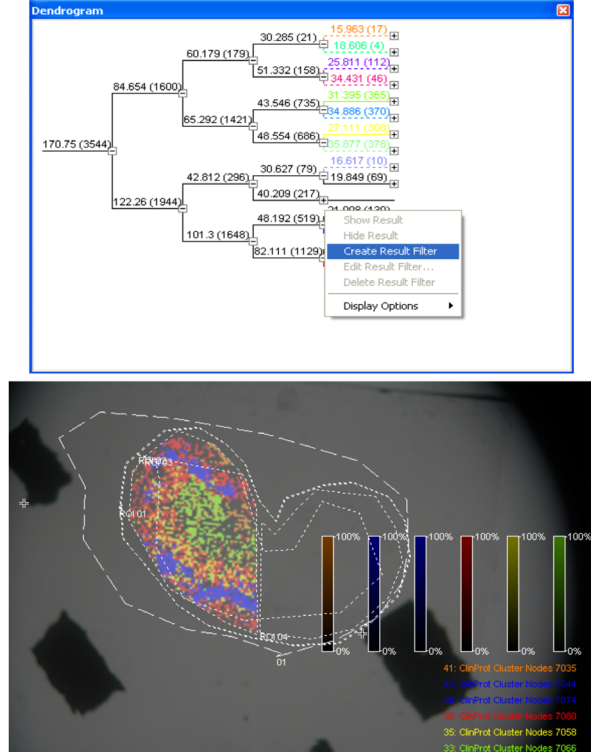


Figura 8. Método de clasificación no supervisada de clusterización jerárquica. Este tipo de análisis hace posible seleccionar nodos de un dendrograma que reflejen determinadas regiones o características histológicas. Esto permite la extracción de información extra, como la similitud molecular entre diferentes regiones. Este grado de similitud se muestra en forma de dendrograma, con aquellas regiones mas similares compartiendo ramas, y aquellas mas diferentes mas alejadas en la representación.

Finalmente, sets de tejidos de diferentes condiciones clínicas (por ejemplo, controles vs. pacientes enfermos), o incluso diferentes regiones de un mismo tejido (por ejemplo, tejido conectivo sano vs. tejido canceroso) se pueden clasificar basándose en los perfiles de masa que generen los diferentes modelos o condiciones (conjuntos de picos e intensidades que separen las diferentes condiciones). Existen multitud de algoritmos diferentes de generación de modelos disponibles para este tipo de estudios, como el algoritmo genético (64), red neural supervisada (65), máquina vectorial (66), etc. La elección de un algoritmo sobre los demás está dictada por el tipo de tejido y el tipo de clases esperadas del sistema.

2.9 Identificación y validación de los resultados

La identificación de las moléculas detectadas por IMS se puede lograr mediante mediciones de masa exacta o mediante espectrometría de masas en tándem (MS/MS), ya sea en la misma sección de tejido, después de realizar el experimento de MALDI-imaging, o en una sección de tejido consecutiva.

En la primera aproximación, las mediciones se realizan en espectrómetros de masas de alta resolución, tipo FT-ICR u orbitrap, con el fin de obtener la masa exacta de los precursores o valores m/z correspondientes a las moléculas de interés. A continuación, estos valores m/z experimentales se comparan con bases de datos especializadas en los diferentes tipos de moléculas con una tolerancia de masa máxima de

5-7 ppm. Actualmente, las bases de datos más utilizadas para la identificación de lípidos incluyen LipidBank (67, 68), Lipid Maps (68, 69), Human Metabolome Database (HMDB) (70) y LipidBlast (71). Para la identificación de metabolitos y fármacos las bases de datos más utilizadas son METLIN (72), KEGG (73) y HMDB. Por otro lado, MSiMass list (74) y MaTisse (75) son dos de las bases de datos empleadas para la identificación de proteínas y péptidos detectados por MSI.

La identificación de las moléculas también se puede realizar comparando los espectros de MS/MS obtenidos para cada precursor con espectros de MS/MS obtenidos a partir de un estándar, o comparando dichos espectros con las bases de datos mencionadas anteriormente. Sin embargo, no todos los valores m/z o precursores (especialmente los que se corresponden con moléculas menos abundantes) se pueden seleccionar y fragmentar usando esta estrategia. De hecho, en el caso de péptidos, no más de unas pocas decenas se identifican directamente en el tejido por MS/MS, debido a la supresión iónica o la alta complejidad de las muestras (76).

La detección e identificación de proteínas de alto peso molecular mediante MALDI-MSI todavía representa un desafío importante. El análisis directo de proteínas por MSI se limita típicamente a aquellas por debajo de 100 kDa (51). Para la detección de proteínas por encima de los 100 kDa, es necesario aplicar tripsina u otras enzimas, con el fin de digerir directamente las proteínas presentes en el tejido. En este método, la tripsina se deposita directamente sobre la superficie del tejido antes de la aplicación de la matriz como se explicó en el apartado 2.4. Para la aplicación de la tripsina se pueden utilizar los dispositivos automáticos descritos en la sección 2.5. A continuación, se lleva a cabo la fragmentación directa por MS/MS de los péptidos detectados para identificar la proteína original. Este método ofrece la ventaja de identificar las proteínas sin una separación previa de las mismas u homogenización de la muestra, a la vez que se preserva la distribución espacial de las proteínas en la muestra de tejido. Sin embargo, esta aproximación permite únicamente la identificación directa de las proteínas más abundantes del tejido. Actualmente, MALDI-MSI se puede combinar con otras técnicas como LC-MS/MS, LESA o la microdissección por captura laser (LMD) para mejorar la identificación de las proteínas. A continuación, se explican las diferentes estrategias actuales empleadas para aumentar el número de identificaciones:

a) Combinación de diferentes enzimas.

Además de la tripsina, las secciones histológicas se pueden tratar con otro tipo de enzimas para desentrañar el peptidoma de interés. Una de las enzimas más utilizadas es la péptido N-glicosidasa F, que rompe el enlace N-glicosídico de los oligosacáridos presentes en las glicoproteínas. Recientemente, se ha publicado un protocolo en el que la misma sección de tejido se puede digerir in situ secuencialmente con N-glicosidasa F y tripsina, incrementando así el número total de proteínas identificadas mediante IMS (77). Además, el uso de tripsina con otras enzimas como colagenasas, mejora la extracción y detección de proteínas de la matriz extracelular de muchos tejidos. De hecho, se ha probado que el uso de colagenasa 3 y metaloproteínasa 12 permite localizar colágenos y otras proteínas de matriz extracelular como la elastina en diferentes tejidos humanos mediante MALDI-IMS (78).



b) Microdissección por captura láser (LMD) acoplada a LC-MS/MS.

La combinación de MALDI-IMS con proteómica basada en LC es una estrategia alternativa para la identificación de proteínas. En esta aproximación, una sección de tejido se analiza mediante MSI, mientras que una sección consecutiva se homogeniza y se analiza por LC-MS/MS. La identificación por LC-MS/MS de péptidos generados a partir de secciones de tejido paralelas nos permite generar una base de datos de péptidos específica que puede ayudar en la identificación de las proteínas detectadas en los experimentos de IMS. Sin embargo, uno de los inconvenientes de utilizar secciones histológicas consecutivas es que, al homogenizar completamente el corte tisular, se pierde el perfil de proteínas específico de cada región de interés definida por IMS (por ejemplo, tumor vs estroma). Con el fin de mantener la firma proteica de cada región, lo ideal es realizar la identificación de las proteínas a partir del mismo corte histológico analizado previamente mediante MALDI-IMS. En esta estrategia, las regiones de interés específicas identificadas por MALDI-IMS pueden aislarse del mismo tejido utilizando la microdissección por captura láser (LMD) y luego analizarse mediante LC-MS/MS para aumentar la cobertura del proteoma (79). El microdisector láser es básicamente un microscopio óptico acoplado a un láser y un sistema de captura de las muestras diseccionadas. Adicionalmente, el microscopio posee una cámara conectada a un ordenador con el que se realiza la selección de las regiones del tejido que han de aislarse para su disección. El láser empleado suele ser un láser infrarojo o UV diseñados para evitar daños a las muestras capturadas. El sistema de captura es distinto en cada equipo, puede ser mediante un film adhesivo (Arcturus, MMI), por una catapulta de onda de presión inducida por un láser (Zeiss) o simplemente por gravedad (Leica). Además, los portaobjetos para LMD poseen una membrana sobre la que se disponen las secciones histológicas y que al realizarse la microdissección se desprende del portaobjetos junto con la muestra. El tipo de polímero de la membrana varía dependiendo de si los portaobjetos son de cristal o si tienen marcos de acero. El uso de un tipo de polímero u otro depende de la aplicación final del LMD ya que el material puede interferir en los procesos posteriores. Justo antes de la microdissección se pueden teñir las secciones histológicas para poder identificar las áreas de interés que se desean separar del tejido. Se pueden hacer tinciones convencionales, como HE o incluso inmunohistoquímicas.

c) Análisis de superficie de extracción líquida (LESA).

MALDI-IMS se puede combinar con técnicas de ionización ambiental, como el análisis de superficie de extracción líquida (LESA), para mejorar la identificación de proteínas de una región localizada de tejido. LESA implica la deposición de una gota de disolvente sobre la superficie del tejido. El disolvente está en contacto con la superficie durante unos segundos formando una microunión líquida entre la superficie y la punta de la pipeta, lo que promueve la extracción de las proteínas en el disolvente. Finalmente, la muestra se aspira y se introduce en el espectrómetro de masas vía ionización por electrospray (ESI). Además, LESA se puede combinar con una etapa de digestión proteolítica seguidamente a la extracción a través de la microunión líquida (80). De este modo, las proteínas se extraen de las secciones de tejido mediante LESA, se someten a una digestión automatizada con tripsina, y la mezcla de péptidos resultante se analiza por LC-ESI-MS/MS (9, 81).

Finalmente, técnicas como inmunohistoquímicas, western blot e incluso RT-PCR cuantitativa ofrecen métodos alternativos para confirmar y validar la abundancia y distribución de las moléculas diana en el tejido (82).

Otros métodos de ionización acoplados a IMS

Además de MALDI, hay otros métodos de ionización que se han acoplado a IMS. Cada uno de estos métodos confiere a la técnica de IMS unas particularidades que la hacen adecuada para un tipo de estudio en concreto. En este apartado comentaremos de forma resumida los métodos más comúnmente empleados: SIMS y DESI.

a) Espectrometría de masas de iones secundarios (SIMS)

Entre las diferentes tecnologías de ionización MSI, la espectrometría de masas de iones secundarios (SIMS) es probablemente una de las más empleadas después de MALDI-IMS. SIMS utiliza un haz de iones primarios que se puede enfocar con precisión en la superficie a analizar. Cuando el haz molecular incide en la superficie del tejido, la alta energía de los iones primarios se transfiere a las moléculas del tejido produciéndose una fragmentación extensa de las mismas. Es por eso que con SIMS, las moléculas detectables están normalmente limitadas a un rango de masa estrecho de solo unos pocos cientos de daltons. Sin embargo, el uso de nuevos haces de iones (por ejemplo, C_{60}^{+} y Bi_3^{+}) ha extendido satisfactoriamente este límite de rango de masa a ~ 2 kDa (26). La principal ventaja de SIMS frente a los demás métodos de ionización es la alta resolución espacial de las imágenes. Con las capacidades ofrecidas recientemente se pueden alcanzar los 100 nm (83). La mayoría de los estudios realizados con SIMS se han centrado en el análisis de fármacos, lípidos e iones elementales en el campo de la caracterización de materiales y superficies.

b) Ionización por electrospray de desorción (DESI)

Además de los métodos MSI estándar que funcionan en vacío que se han mencionado anteriormente (MALDI-IMS y SIMS), también se han desarrollado técnicas de ionización ambiental, que operan en condiciones de presión atmosférica. La ventaja de este tipo de técnicas es que no requieren la aplicación de matriz para la ionización (requiriendo por tanto una mínima preparación de la muestra) y son más fáciles de implementar. Entre ellos, la ionización por electrospray de desorción (DESI) usa un haz de gotas cargadas para desorber e ionizar las moléculas de la superficie de la muestra. La resolución espacial obtenida con DESI es aproximadamente de 100-500 μm , pero este valor puede disminuir a 40 μm usando condiciones de operación particulares (nano-DESI) (84). Si bien este método ha demostrado tener éxito en el análisis de moléculas de bajo peso molecular, tales como fármacos y metabolitos, se ha demostrado que es más difícil adaptar la técnica para el análisis de proteínas (85).



Agradecimientos

CIC bioGUNE recibe ayudas del Departamento de industria, turismo y comercio del Gobierno Vasco mediante los programas Etortek y Elkartek, y también de la Diputación Foral de Bizkaia; La plataforma de Proteómica del CIC bioGUNE es parte de ProteoRed (PRB3 IPT17/0019), CIBERehd y Severo Ochoa (SEV-2016-0644).

El trabajo de la Unidad de Proteómica del GIR está respaldado con ayudas del Instituto de Salud Carlos III (PI17/00404, PT17/0019/0014, PI16/02124, CIBER-CB06/01/0040, RETIC-RIER-RD12/0009/0018, CPII15/00013) y de la Xunta de Galicia (IN606B-2016/004).

REFERENCIAS

1. Longuespee R, Casadonte R, Kriegsmann M, Pottier C, Picard de Muller G, Delvenne P, et al. MALDI mass spectrometry imaging: A cutting-edge tool for fundamental and clinical histopathology. *Proteomics Clin Appl.* 2016;10(7):701-19.
2. Cole LM, Clench MR. Mass spectrometry imaging for the proteomic study of clinical tissue. *Proteomics Clin Appl.* 2015;9(3-4):335-41.
3. Rodrigo MA, Zitka O, Krizkova S, Moullick A, Adam V, Kizek R. MALDI-TOF MS as evolving cancer diagnostic tool: a review. *J Pharm Biomed Anal.* 2014;95:245-55.
4. Zaima N, Goto-Inoue N, Moriyama T. Matrix-assisted laser desorption/ionization imaging mass spectrometry: new technology for vascular pathology. *J Vasc Res.* 2014;51(2):144-8.
5. Vegvari A. Drug localizations in tissue by mass spectrometry imaging. *Biomark Med.* 2015;9(9):869-76.
6. Chaurand P, Norris JL, Cornett DS, Mobley JA, Caprioli RM. New developments in profiling and imaging of proteins from tissue sections by MALDI mass spectrometry. *J Proteome Res.* 2006;5(11):2889-900.
7. Stoekli M, Chaurand P, Hallahan DE, Caprioli RM. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nat Med.* 2001;7(4):493-6.
8. Stoekli M, Farmer TB, Caprioli RM. Automated mass spectrometry imaging with a matrix-assisted laser desorption ionization time-of-flight instrument. *J Am Soc Mass Spectrom.* 1999;10(1):67-71.
9. Fournier I, Wisztorski M, Salz M. Tissue imaging using MALDI-MS: a new frontier of histopathology proteomics. *Expert Rev Proteomics.* 2008;5(3):413-24.
10. Hardesty WM, Caprioli RM. In situ molecular imaging of proteins in tissues using mass spectrometry. *Anal Bioanal Chem.* 2008;391(3):899-903.
11. Beine B, Diehl HC, Meyer HE, Henkel C. Tissue MALDI Mass Spectrometry Imaging (MALDI MSI) of Peptides. *Methods Mol Biol.* 2016;1394:129-50.
12. Yalcin EB, de la Monte SM. Review of matrix-assisted laser desorption ionization-imaging mass spectrometry for lipid biochemical histopathology. *J Histochem Cytochem.* 2015;63(10):762-71.
13. Miura D, Fujimura Y, Wariishi H. In situ metabolomic mass spectrometry imaging: recent advances and difficulties. *J Proteomics.* 2012;75(16):5052-60.
14. Hillenkamp F, Karas M. Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods Enzymol.* 1990;193:280-95.
15. Chaurand P, Schwartz SA, Billheimer D, Xu BJ, Crecelius A, Caprioli RM. Integrating histology and imaging mass spectrometry. *Anal Chem.* 2004;76(4):1145-55.
16. Caprioli RM, Farmer TB, Gile J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal Chem.* 1997;69(23):4751-60.
17. Ogrinc Potocnik N, Porta T, Becker M, Heeren RM, Ellis SR. Use of advantageous, volatile matrices enabled by next-generation high-speed matrix-assisted laser desorption/ionization time-of-flight imaging employing a scanning laser beam. *Rapid Commun Mass Spectrom.* 2015;29(23):2195-203.
18. Iloro I, Bueno A, Calvo J, Urreta H, Elortza F. Langartech: A Custom-Made MALDI Matrix Sprayer for MALDI Imaging Mass Spectrometry. *J Lab Autom.* 2016;21(2):260-7.
19. Lou S, Balluff B, Cleven AH, Bovee JV, McDonnell LA. An experimental guideline for the analysis of histologically heterogeneous tumors by MALDI-TOF mass spectrometry imaging. *Biochim Biophys Acta.* 2016.
20. Mourino-Alvarez L, Iloro I, de la Cuesta F, Azkargorta M, Sastre-Oliva T, Escobes I, et al. MALDI-Imaging Mass Spectrometry: a step forward in the anatomopathological characterization of stenotic aortic valve tissue. *Sci Rep.* 2016;6:27106.
21. Reyzer ML, Caldwell RL, Dugger TC, Forbes JT, Ritter CA, Guix M, et al. Early changes in protein ex-

- pression detected by mass spectrometry predict tumor response to molecular therapeutics. *Cancer Res.* 2004;64(24):9093-100.
22. Atkinson SJ, Loadman PM, Sutton C, Patterson LH, Clench MR. Examination of the distribution of the bioreductive drug AQ4N and its active metabolite AQ4 in solid tumours by imaging matrix-assisted laser desorption/ionisation mass spectrometry. *Rapid Commun Mass Spectrom.* 2007;21(7):1271-6.
 23. Khatib-Shahidi S, Andersson M, Herman JL, Gillespie TA, Caprioli RM. Direct molecular analysis of whole-body animal tissue sections by imaging MALDI mass spectrometry. *Anal Chem.* 2006;78(18):6448-56.
 24. Norris JL, Caprioli RM. Imaging mass spectrometry: a new tool for pathology in a molecular age. *Proteomics Clin Appl.* 2013;7(11-12):733-8.
 25. Schwamborn K, Caprioli RM. Molecular imaging by mass spectrometry--looking beyond classical histology. *Nat Rev Cancer.* 2010;10(9):639-46.
 26. Aichler M, Walch A. MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice. *Lab Invest.* 2015;95(4):422-31.
 27. Seeley EH, Caprioli RM. 3D imaging by mass spectrometry: a new frontier. *Anal Chem.* 2012;84(5):2105-10.
 28. Deutschens F, Yang J, Caprioli RM. High spatial resolution imaging mass spectrometry and classical histology on a single tissue section. *J Mass Spectrom.* 2011;46(6):568-71.
 29. Sinha TK, Khatib-Shahidi S, Yankeeelov TE, Mapara K, Ehtesham M, Cornett DS, et al. Integrating spatially resolved three-dimensional MALDI IMS with in vivo magnetic resonance imaging. *Nat Methods.* 2008;5(1):57-9.
 30. McDonnell LA, Heeren RM. Imaging mass spectrometry. *Mass Spectrom Rev.* 2007;26(4):606-43.
 31. Agar NY, Kowalski JM, Kowalski PJ, Wong JH, Agar JN. Tissue preparation for the in situ MALDI MS imaging of proteins, lipids, and small molecules at cellular resolution. *Methods Mol Biol.* 2010;656:415-31.
 32. Agar NY, Yang HW, Carroll RS, Black PM, Agar JN. Matrix solution fixation: histology-compatible tissue preparation for MALDI mass spectrometry imaging. *Anal Chem.* 2007;79(19):7416-23.
 33. Bank HL, Brockbank KG. Basic principles of cryobiology. *J Card Surg.* 1987;2(1 Suppl):137-43.
 34. Mainini V, Lalowski M, Gotsopoulos A, Bitsika V, Baumann M, Magni F. MALDI-imaging mass spectrometry on tissues. *Methods Mol Biol.* 2015;1243:139-64.
 35. Oetjen J, Lachmund D, Palmer A, Alexandrov T, Becker M, Boskamp T, et al. An approach to optimize sample preparation for MALDI imaging MS of FFPE sections using fractional factorial design of experiments. *Anal Bioanal Chem.* 2016;408(24):6729-40.
 36. Aoki Y, Toyama A, Shimada T, Sugita T, Aoki C, Umino Y, et al. A novel method for analyzing formalin-fixed paraffin embedded (FFPE) tissue sections by mass spectrometry imaging. *Proc Jpn Acad Ser B Phys Biol Sci.* 2007;83(7):205-14.
 37. Kitamoto Y, Maeda H. Reevaluation of the reaction of formaldehyde at low concentration with amino acids. *J Biochem.* 1980;87(5):1519-30.
 38. Casadonte R, Caprioli RM. Proteomic analysis of formalin-fixed paraffin-embedded tissue by MALDI imaging mass spectrometry. *Nat Protoc.* 2011;6(11):1695-709.
 39. De Sio G, Smith AJ, Galli M, Garancini M, Chinello C, Bono F, et al. A MALDI-Mass Spectrometry Imaging method applicable to different formalin-fixed paraffin-embedded human tissues. *Mol Biosyst.* 2015;11(6):1507-14.
 40. Diehl HC, Beine B, Elm J, Trede D, Ahrens M, Eisenacher M, et al. The challenge of on-tissue digestion for MALDI MSI- a comparison of different protocols to improve imaging experiments. *Anal Bioanal Chem.* 2015;407(8):2223-43.
 41. Gustafsson OJ, Eddes JS, Meding S, McColl SR, Oehler MK, Hoffmann P. Matrix-assisted laser desorption/ionization imaging protocol for in situ characterization of tryptic peptide identity and distribution in formalin-fixed tissue. *Rapid Commun Mass Spectrom.* 2013;27(6):655-70.
 42. Buck A, Ly A, Balluff B, Sun N, Gorzalka K, Feuchtinger A, et al. High-resolution MALDI-FT-ICR MS imaging for the analysis of metabolites from formalin-fixed, paraffin-embedded clinical tissue samples. *J Pathol.* 2015;237(1):123-32.
 43. Altelaar AF, van Minnen J, Jimenez CR, Heeren RM, Piersma SR. Direct molecular imaging of *Lymnaea stagnalis* nervous tissue at subcellular spatial resolution by mass spectrometry. *Anal Chem.* 2005;77(3):735-41.
 44. Kruse R, Sweedler JV. Spatial profiling invertebrate ganglia using MALDI MS. *J Am Soc Mass Spectrom.* 2003;14(7):752-9.
 45. Martin-Lorenzo M, Balluff B, Maroto AS, Carreira RJ, van Zeijl RJ, Gonzalez-Calero L, et al. Molecular anatomy of ascending aorta in atherosclerosis by MS Imaging: Specific lipid and protein patterns reflect pathology. *J Proteomics.* 2015;126:245-51.
 46. Caldwell RL, Caprioli RM. Tissue profiling by mass spectrometry: a review of methodology and applications. *Mol Cell Proteomics.* 2005;4(4):394-401.
 47. Cillero-Pastor B, Heeren RM. Matrix-assisted laser desorption ionization mass spectrometry imaging for peptide and protein analyses: a critical review of on-tissue digestion. *J Proteome Res.* 2014;13(2):325-35.
 48. Djidja MC, Francese S, Loadman PM, Sutton CW, Scriven P, Claude E, et al. Detergent addition to tryptic digests and ion mobility separation prior to MS/MS improves peptide yield and protein identification for in situ proteomic investigation of frozen and formalin-fixed paraffin-embedded adenocarcinoma tissue sections. *Proteomics.* 2009;9(10):2750-63.
 49. Schwartz SA, Reyzer ML, Caprioli RM. Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation. *J Mass Spectrom.* 2003;38(7):699-708.
 50. Herring KD, Oppenheimer SR, Caprioli RM. Direct tissue analysis by matrix-assisted laser desorption ionization mass spectrometry.



zation mass spectrometry: application to kidney biology. *Semin Nephrol.* 2007;27(6):597-608.

51. Mainini V, Bovo G, Chinello C, Gianazza E, Grasso M, Cattoretti G, et al. Detection of high molecular weight proteins by MALDI imaging mass spectrometry. *Mol Biosyst.* 2013;9(6):1101-7.
52. Stoyanovsky DA, Sparvero LJ, Amoscato AA, He RR, Watkins S, Pitt BR, et al. Improved spatial resolution of matrix-assisted laser desorption/ionization imaging of lipids in the brain by alkylated derivatives of 2,5-dihydroxybenzoic acid. *Rapid Commun Mass Spectrom.* 2014;28(5):403-12.
53. Ogrinc Potočnik N, Porta T, Becker M, Heeren RM, Ellis SR. Use of advantageous, volatile matrices enabled by next-generation high-speed matrix-assisted laser desorption/ionization time-of-flight imaging employing a scanning laser beam. *Rapid Commun Mass Spectrom.* 2015;29(23):2195-203.
54. Lemaire R, Tabet JC, Ducoroy P, Hendra JB, Salzet M, Fournier I. Solid ionic matrixes for direct tissue analysis and MALDI imaging. *Anal Chem.* 2006;78(3):809-19.
55. Stübiger G, Belgacem O. Analysis of lipids using 2,4,6-trihydroxyacetophenone as a matrix for MALDI mass spectrometry. *Anal Chem.* 2007;79:3206-13.
56. Tholey A, Heinzel E. Ionic (liquid) matrices for matrix-assisted laser desorption/ionization mass spectrometry-applications and perspectives. *Anal Bioanal Chem.* 2006;386(1):24-37.
57. Cerruti CD, Benabdellah F, Laprévotte O, Touboul D, Brunelle A. MALDI imaging and structural analysis of rat brain lipid negative ions with 9-aminoacridine matrix. *Anal Chem.* 2012;84(5):2164-71.
58. Steven RT, Race AM, Bunch J. para-Nitroaniline is a promising matrix for MALDI-MS imaging on intermediate pressure MS systems. *J Am Soc Mass Spectrom.* 2013;24(5):801-4.
59. Wang J, Qiu S, Chen S, Xiong C, Liu H, Zhang N, et al. MALDI-TOF MS imaging of metabolites with a N-(1-naphthyl) ethylenediamine dihydrochloride matrix and its application to colorectal cancer liver metastasis. *Anal Chem.* 2015;87(1):422-30.
60. Wang J, Wang C, Han X. Enhanced coverage of lipid analysis and imaging by matrix-assisted laser desorption/ionization mass spectrometry via a strategy with an optimized mixture of matrices. *Anal Chim Acta.* 2018;1000:155-62.
61. Scott AJ, Flinders B, Cappell J, Liang T, Pelc RS, Tran B, et al. Norharmaline Matrix Enhances Detection of Endotoxin by MALDI-MS for Simultaneous Profiling of Pathogen, Host, and Vector Systems. *Pathog Dis.* 2016.
62. Van de Plas R, Ojeda F, Dewil M, Van Den Bosch L, De Moor B, Waelkens E. Prospective exploration of biochemical tissue composition via imaging mass spectrometry guided by principal component analysis. *Pac Symp Biocomput.* 2007:458-69.
63. McCombie G, Staab D, Stoeckli M, Knochenmuss R. Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. *Anal Chem.* 2005;77(19):6118-24.
64. Holland JH. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*: MIT Press; 1992. 228 p.
65. Hammer B, Strickert M, Villmann T. Supervised Neural Gas with General Similarity Measure. *Neural Process Lett.* 2005;21(1):21-44.
66. Vapnik VN. *The nature of statistical learning theory*: Springer-Verlag New York, Inc.; 1995. 188 p.
67. Yasugi E, Watanabe K. [LIPIDBANK for Web, the newly developed lipid database]. *Tanpakushitsu Kakusan Koso.* 2002;47(7):837-41.
68. Taguchi R, Nishijima M, Shimizu T. Basic analytical systems for lipidomics by mass spectrometry in Japan. *Methods Enzymol.* 2007;432:185-211.
69. Fahy E, Sud M, Cotter D, Subramaniam S. LIPID MAPS online tools for lipid research. *Nucleic Acids Res.* 2007;35(Web Server issue):W606-12.
70. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res.* 2013;41(Database issue):D801-7.
71. Kind T, Liu KH, Lee dY, DeFelice B, Meissen JK, Fiehn O. LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat Methods.* 2013;10(8):755-8.
72. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, et al. METLIN: a metabolite mass spectral database. *Ther Drug Monit.* 2005;27(6):747-51.
73. Kanehisa M. The KEGG database. *Novartis Found Symp.* 2002;247:91-101; discussion -3, 19-28, 244-52.
74. McDonnell LA, Walch A, Stoeckli M, Corthals GL. MSiMass list: a public database of identifications for protein MALDI MS imaging. *J Proteome Res.* 2014;13(2):1138-42.
75. Maier SK, Hahne H, Gholami AM, Balluff B, Meding S, Schoene C, et al. Comprehensive identification of proteins from MALDI imaging. *Mol Cell Proteomics.* 2013;12(10):2901-10.
76. Debois D, Bertrand V, Quinton L, De Pauw-Gillet MC, De Pauw E. MALDI-in source decay applied to mass spectrometry imaging: a new tool for protein identification. *Anal Chem.* 2010;82(10):4036-45.
77. Heijs B, Holst S, Briaire-de Bruijn IH, van Pelt GW, de Ru AH, van Veelen PA, et al. Multimodal Mass Spectrometry Imaging of N-Glycans and Proteins from the Same Tissue Section. *Anal Chem.* 2016;88(15):7745-53.
78. Angel PM, Comte-Walters S, Ball LE, Talbot K, Mehta A, Brockbank KGM, et al. Mapping Extracellular Matrix Proteins in Formalin-Fixed, Paraffin-embedded Tissues by MALDI Imaging Mass Spectrometry. *J Proteome Res.* 2017.
79. DiIillo M, Pellegrini D, Ait-Belkacem R, de Graaf EL, Caleo M, McDonnell LA. Mass Spectrometry Imaging, Laser Capture Microdissection, and LC-MS/MS of the Same Tissue Section. *J Proteome Res.* 2017;16(8):2993-3001.
80. Wisztorski M, Fatou B, Franck J, Desmons A, Farré I, Leblanc E, et al. Microproteomics by liquid extraction surface analysis: application to FFPE tissue to study the fimbria region of tubo-ovarian cancer. *Proteomics Clin*

Appl. 2013;7(3-4):234-40.

81. Quanico J, Franck J, Dauly C, Strupat K, Dupuy J, Day R, et al. Development of liquid microjunction extraction strategy for improving protein identification from tissue sections. *J Proteomics*. 2013;79:200-18.

82. Rebours V, Le Faouder J, Laouirem S, Mebarki M, Albuquerque M, Camadro JM, et al. In situ proteomic analysis by MALDI imaging identifies ubiquitin and thymosin- β 4 as markers of malignant intraductal pancreatic mucinous neoplasms. *Pancreatology*. 2014;14(2):117-24.

83. Gamble LJ, Anderton CR. Secondary Ion Mass Spectrometry Imaging of Tissues, Cells, and Microbial Systems. *Micros Today*. 2016;24(2):24-31.

84. Campbell DI, Ferreira CR, Eberlin LS, Cooks RG. Improved spatial resolution in the imaging of biological tissue using desorption electrospray ionization. *Anal Bioanal Chem*. 2012;404(2):389-98.

85. Douglass KA, Venter AR. Protein analysis by desorption electrospray ionization mass spectrometry and related methods. *J Mass Spectrom*. 2013;48(5):553-60.



Interactómica. Estrategias de purificación por afinidad para análisis del proteoma

Rubén A. Bartolomé y J. Ignacio Casal

Departamento de Biomedicina Molecular. Centro de Investigaciones Biológicas (CIB-CSIC). Madrid. España.

1. Introducción.

La mayoría de los procesos biológicos en los seres vivos requieren grandes complejos multiproteicos para llevarse a cabo. Habitualmente, las proteínas actúan como componentes de complejos proteicos dinámicos para realizar sus funciones. De manera que el número de interacciones entre diferentes proteínas es enorme. El estudio in vitro de dichas interacciones es una rama de la biología relativamente reciente pero de gran interés denominado interactómica. En un organismo multicelular los interactomas difieren según el tejido, el estado biológico, etc. La dinámica del interactoma hace que su determinación sea muy compleja. En ocasiones se intenta determinar los interactomas cambiantes de determinadas funciones o localizaciones celulares relevantes. Otro abordaje consiste en generar una mayor visión de conjunto de un interactoma estático. Partes sustanciales del interactoma ya han sido mapeadas en levadura, plantas, nematodos, mosca y humano (ver revisiones [1, 2]). La compleja dinámica de los interactomas supone que la información sobre dichos interactomas sea extremadamente importante para entender los complicados mecanismos moleculares que intervienen en la actividad de los sistemas biológicos. De hecho, la alteración de interacciones entre proteínas puede tener consecuencias muy severas, tales como el desarrollo de ciertas enfermedades [3, 4]. Por otra parte, los estudios de interactómica son de gran interés, ya que permiten un abordaje innovador: el desarrollo de nuevos fármacos que alteren la regulación de las interacciones proteicas (por ejemplo, bloqueando la formación de complejos funcionales) en lugar de afectar a proteínas individuales [5]. Adicionalmente, los datos de interactómica obtenidos mediante ensayos proteómicos proporcionan una herramienta fundamental en la caracterización de la función de proteínas poco estudiadas [6]. Si en el interactoma de dicha proteína desconocida se identifican proteínas de función conocida puede inferirse la función de la primera. Esto se basa en la asunción de que la función de una proteína está conectada con las funciones de las proteínas con las que interacciona.

2. Técnicas en interactómica.

La interactómica usa herramientas bioinformáticas, bioquímicas, genómicas, de biología estructural y celular. Existen muchísimas técnicas para detectar interacciones entre



proteínas, algunas sólo permiten testar la interacción entre dos proteínas (técnicas binarias) mientras otras permiten el aislamiento e identificación de complejos proteicos. Centrándonos en aquellas que pueden adaptarse a un estudio proteómico, podemos clasificar los diferentes abordajes en la caracterización de los interactomas en dos categorías: técnicas bioinformáticas y técnicas experimentales [1, 7].

Las técnicas bioinformáticas se basan en el análisis de datos previos sobre transcripción, evolución y estructura de las proteínas. El análisis de perfiles de expresión de mRNA en diferentes sistemas (diferentes organismos, tejidos, estados biológicos, etc.) permite hallar correlaciones en la expresión de varios genes. A menudo, genes con expresiones similares codifican para proteínas que interactúan entre sí, aunque tales inferencias han de ser verificadas por otros métodos. Los métodos basados en la relación evolutiva de las proteínas tratan de inferir la función y el interactoma de una proteína a través de sus homólogos en otros organismos, como las levaduras, estudiados en profundidad. Del mismo modo, proteínas pertenecientes a una misma familia, dentro de un mismo organismo, suelen tener funciones relacionadas e incluso supletorias. En general, proteínas con más de un 30% de homología en su secuencia se encuentran formando parte de los mismos complejos o muy similares y por tanto con las mismas interacciones proteína-proteína. La estructura de una proteína proporciona importantes pistas sobre su función. El conocimiento de la estructura espacial de un complejo de proteínas permite estimar la posibilidad de la formación de un complejo análogo por proteínas de estructura tridimensional análoga. La presencia de dominios característicos en una proteína permite predecir las interacciones con otras proteínas que tengan dominios interactuantes con dicho dominios. Finalmente, la secuencia de la proteína desconocida permite, mediante el análisis de los motivos presentes, asociarla a funciones e interacciones que poseen las proteínas con dichos motivos.

Entre las técnicas experimentales se halla el sistema de doble híbrido de levaduras (*yeast two-hybrid system*). El sistema de doble híbrido proporciona un método muy empleado para la detección de nuevas interacciones entre proteínas debido a su simplicidad y su alto rendimiento [8, 9]. Esta técnica se basa en un factor de transcripción separado en dos fragmentos, cada uno de los cuales se fusiona mediante clonaje a una de las proteínas cuya interacción se quiere analizar. Si ambas proteínas interactúan entre sí, los dos fragmentos del factor de transcripción se reúnen y el gen reportero se transcribe, detectándose un cambio fenotípico. El uso de genotecas en este sistema proporciona un método para identificar nuevas proteínas interactivas mediante la posterior secuenciación de las levaduras con cambio fenotípico [10].

Otras técnicas experimentales tienen como fundamento la cromatografía de afinidad en fase sólida. Entre estas destacan la inmunoprecipitación específica de anticuerpos, precipitación de proteínas de fusión con la glutatión-S-transferasa y/o cola de histidinas cuando no existen buenos anticuerpos, la biotinilación dependiente de proximidad y la purificación de afinidad en tándem. Relacionada con estas técnicas bioquímicas está la técnica de “pesca molecular” (*molecular fishing*) en un biosensor óptico, consistente en la inmovilización de una proteína problema en el canal de medida de la superficie de un chip óptico por el cual se hace pasar una solución proteica. El biosensor óptico registrará una serie de medidas cuantitativas basadas en el efecto de resonan-

cia plasmónica de superficie cuando se produce la unión a la proteína inmovilizada [7, 11]. Todas estas técnicas bioquímicas no identifican las proteínas interactuantes con la proteína problema, para lo cual se usan en combinación con técnicas proteómicas.

A continuación describiremos con más detalle algunas de las técnicas más utilizadas.

2.1. Inmunoprecipitación por anticuerpos

La técnica de inmunoprecipitación consiste en usar anticuerpos específicos para la proteína estudiada, que a su vez mantendrá las interacciones con diversas proteínas con las que forma un complejo y, posteriormente, analizar dichos complejos por espectrometría de masas. El paso clave es encontrar un anticuerpo lo suficientemente específico y selectivo para minimizar el número de falsos positivos y falsos negativos. Brevemente, las células o tejidos se lisan en solución de lisis (150 mM NaCl, 50 mM Hepes, 1% NP-40, inhibidores de proteasas), se incuban con microesferas (*beads*) de agarosa conjugadas con proteína A o G para eliminar las proteínas inespecíficas (*precleaning*) y a continuación se incuban con el anticuerpo primario (1-5 μ g por mg de lisado) en presencia de microesferas de agarosa. Una vez formado el inmunoprecipitado, se lavan las microesferas varias veces con la solución de lisis, se resuspenden en buffer de Laemmli y se hierven para liberar las proteínas de las microesferas, que se resuelven en un gel de poliacrilamida. Dicho gel se puede analizar por *Western blot* para verificar las interacciones con proteínas.

Para el estudio proteómico, el gel se corre 1-2 centímetros, y cada carril se divide en las fracciones deseadas que serán digeridas con tripsina en el gel. Los péptidos extraídos se analizan en el espectrómetro de masas, de manera que se identifican las proteínas asociadas a la proteína estudiada. Siguiendo esta estrategia, en nuestro laboratorio se realizó la identificación del interactoma de la cadherina 17 en células de cáncer de colon [12]: se identificaron 389 proteínas coimmunoprecipitadas (que no aparecieron con el anticuerpo control), de las cuales 67 se seleccionaron como interactores específicos de dicha proteína. Entre las proteínas específicas, el grupo más nutrido (26 proteínas) estaban relacionadas con las adhesiones focales, lo cual, junto con el hecho de que cadherina 17 posee un motivo RGD, nos permitió identificarla como un nuevo ligando de integrinas [13].

El análisis mediante espectrometría de masas se realizó de manera habitual. Los péptidos se cargaron en una precolumna C18-A1 y se separaron en una columna Biosphere C18 usando un flujo de 250 nL/min y un gradiente de 100 min en un HPLC nanoEasy acoplado a un fuente de iones por nanoelectrospray. Los espectros se adquirieron en una trampa iónica lineal LTQ Orbitrap Velos en modo dependiente de datos usando el método *top 15*. Los barridos continuos se adquirieron con un rango de masas de entre 400 y 1200 Th, un valor diana de 10^6 iones y una resolución de 30.000 (a una relación m/z de 400). Los 15 iones más intensos fueron sometidos a disociación inducida por colisión usando una energía de colisión normalizada del 35% y un valor diana de 10^4 iones. La exclusión dinámica se estableció con una duración de 30 s. Los espectros de masas fueron identificados mediante el motor de búsqueda SEQUEST con Proteome Discoverer usando la base de datos Uniprot. Las tolerancias de masa en precursores y fragmentos fueron respectivamente de 10 ppm y 0.8 Da.



Los péptidos así identificados se validaron con Percolator usando un $q < 0.01$. Posteriormente, se realizaron experimentos similares para determinar el interactoma de VE-cadherina en células de cáncer de mama y melanoma [14].

Para aumentar la identificación de interactores se puede realizar un fraccionamiento previo celular, seleccionando él o los compartimentos celulares de interés (secretoma, membrana, núcleo, mitocondrias, etc.) [15]. Asimismo, se puede cambiar el detergente (o su concentración) usado en la solución de lisis para detectar exclusivamente interacciones fuertes (como Tritón X-100, Tween 20) o por el contrario mantener interacciones más débiles (como CHAPS, Brij-35, Digitonina), con la contrapartida de una menor especificidad. Siempre han de usarse detergentes no desnaturalizantes. En el caso de proteínas de membrana no es conveniente usar detergentes especialmente débiles dado que se corre el riesgo de no liberar la proteína de interés. En este caso es recomendable usar la solución de lisis GST-FISH [12] o bien las soluciones del método Karginov-Agaphonov [16] que permiten extraer las proteínas de membrana manteniendo mejor sus asociaciones.

Para evitar la presencia del anticuerpo entre las proteínas coimmunoprecipitadas, el cual al ser muy abundante puede enmascarar otras proteínas en un abordaje proteómico, el anticuerpo se puede conjugar covalentemente a microesferas de proteína A o G, previamente a la incubación con el lisado. Para ello se incuba el anticuerpo con las microesferas (1 $\mu\text{g}/10 \mu\text{L}$) y a continuación se conjuga durante 1 h con 20 mM de DMP (dihidrocloreto de dimetil-pimelimidato) en 0.1 M borato sódico. Las microesferas

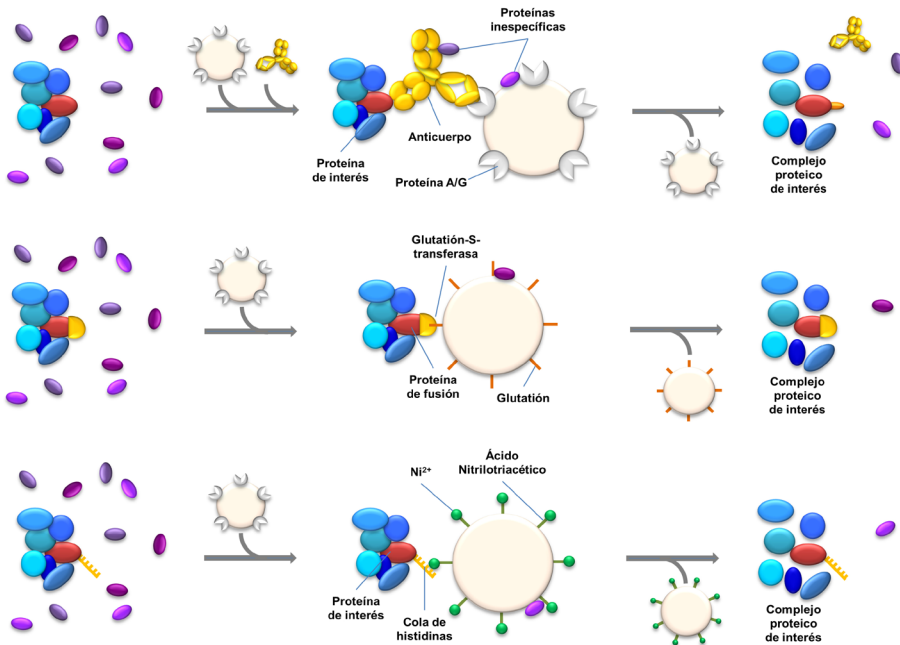


Figura 1. Inmunoprecipitación, precipitación por GST y purificación por cola de histidinas. En la parte superior se muestra el proceso de inmunoprecipitación, donde un lisado celular o tisular es atrapado por un anticuerpo y unas esferas acopladas a proteína A o G, liberándose el complejo proteico de interés tras hervir el inmunocomplejo. En la parte central, un lisado conteniendo una proteína de fusión con la Glutación-S-transferasa es capturada con una resina acoplada a glutatión, liberándose nuevamente mediante hervor. En la parte inferior se muestra la purificación de una proteína recombinante que posee una cola de 5 o 6 histidinas mediante su unión a una resina conjugada con ácido nitrilotriacético y cationes de níquel. La purificación se consigue mediante elución con imidazol.

se lavan para eliminar la proteína no conjugada con 50 mM glicina-HCl pH 2.5 y se resuspenden en la solución deseada.

2.2. Precipitación de proteínas de fusión con la glutatión-S-transferasa (*pull down*)

Esta técnica conocida comúnmente como *pull down* se basa en el uso de proteínas de fusión acopladas con la glutatión-S-transferasa y unas microesferas de agarosa conjugadas con glutatión [17]. Brevemente, se produce la proteína de fusión de la forma habitual, y los lisados celulares se incuban con microesferas de agarosa conjugada con glutatión. Tras la incubación se realizan lavados con la misma solución de lisado (20% sacarosa, 10% glicerol, 50 mM Tris pH 8.0, 0.2 mM $\text{Na}_2\text{S}_2\text{O}_5$, 2 mM MgCl_2 , 2 mM DTT). Dichas microesferas así conjugadas se incuban con células lisadas en solución GST-FISH (10% glicerol, 50 mM Tris pH 7.4, 100 mM NaCl, 2 mM MgCl_2 , 1% NP-40) y, tras los lavados, las microesferas se resuspenden en buffer de Laemmli y se hierven. Las proteínas así liberadas se cargarán en un gel de poliacrilamida que se puede analizar cómo se ha descrito en la inmunoprecipitación. Como control se suelen emplear microesferas de agarosa, para eliminar las proteínas inespecíficas que se unan a las propias microesferas.

2.3. Precipitación de proteínas con cola de histidinas

Esta técnica se basa en la purificación de la proteína recombinante (y de los complejos en los que esté presente) por afinidad de la etiqueta de poli-His de la proteína problema con metal inmovilizado (generalmente Ni^{2+}) [18]. Existen matrices de níquel-ácido nitrilotriacético (Ni^{2+} -NTA) o cobalto-carboximetilaspártato (Co^{2+} -CMA) comerciales acopladas a resinas sólidas (columnas). El protocolo consiste en pasar lentamente por las columnas los extractos celulares o tisulares previamente lisados en solución de carga (300 mM NaCl, 50 mM NaH_2PO_4 , pH 8.0). Tras los lavados con la misma solución (pero conteniendo 10 mM imidazol), la columna se eluye con un gradiente entre 10 y 250 mM imidazol en solución de carga. Las fracciones se pueden cargar en un gel de poliacrilamida o directamente someter a un cambio de solución y digestión con tripsina.

La gran ventaja de este sistema es la gran cantidad de muestra que se puede purificar. Los problemas con esta técnica surgen debido a la contaminación con proteínas inespecíficas que contengan al menos dos histidinas contiguas, por lo que es conveniente usar como control lisados sin la presencia de la proteína marcada. Otro problema es la purificación de proteínas de membrana, ya que estas matrices de afinidad toleran cantidades limitadas de detergentes. Evidentemente los lisados no pueden contener quelantes de cationes divalentes como EDTA.

2.4. Purificación por afinidad en tándem

Esta aproximación bioquímica se desarrolló para incrementar la eficacia en el aislamiento de complejos proteico mediante el uso de proteínas conjugadas con dos eti-



quetas (*tags*), una de las cuales suele ser CMP (*Calmodulin-Binding Peptide*) y la otra un fragmento de la Proteína A, separados ambos por un sitio de corte de la proteasa TEV (*Tobacco Etch Virus*) del virus del jaspeado del tabaco, cuya secuencia es ENLY-FQ\S para crear el *tag* de purificación por afinidad en tándem (TAP) [19]. Inicialmente se utilizó en levaduras pero posteriormente se adaptó a células de mamífero [20]. Esta tecnología ha sido muy utilizada para obtener una visión global de la biología de sistemas de interactomas celulares [21]. En una primera fase, el complejo formado por la proteína marcada es aislado mediante inmunoprecipitación con anticuerpos que se unen a la Proteína A. A continuación, en lugar de hervir la muestra, se corta con la proteasa TEV, liberándose el complejo de las microesferas de agarosa. El sobrenadante se incuba entonces con microesferas conjugadas con calmodulina, puesto que CMP queda entonces expuesto. Tras la incubación y los lavados se añade EGTA que secuestra los iones de calcio causando la destrucción de la interacción CMP-Calmodulina y la liberación del complejo proteico que se analiza por espectrometría de masas utilizando diferentes estrategias. Existen diversas variaciones de esta técnica usando diversos *tags*, como *streptavidin-binding peptide*, etc [22]. Las ventajas de la doble purificación son que elimina muchos contaminantes que quedan en el primer aislamiento del complejo y a la vez se usan lavados suaves que mantiene la estructura de los complejos proteicos [23]. Uno de los inconvenientes de esta técnica es debido a que la calmodulina endógena o las proteínas que se unen a ella pueden dificultar una unión eficiente en el segundo paso de purificación; para evitar este problema se han desarrollado diferentes *tags*. Otras limitaciones son el tamaño relativamente grande del *tag*, que puede ocasionar problemas para ser expresado, así como la necesidad de múltiples pasos de lavado y cambios de solución, que puede causar la pérdida de las interacciones más débiles.

2.5. Biotinilación dependiente de proximidad (BioID)

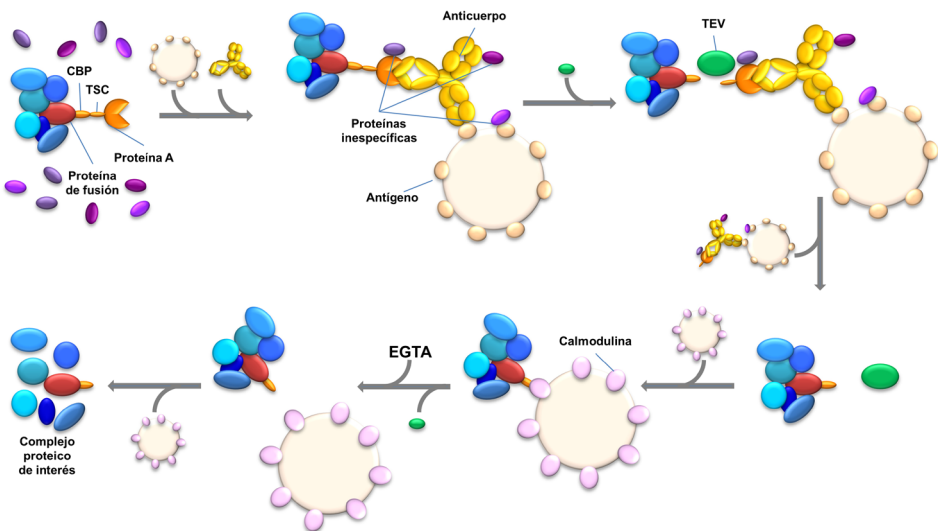


Figura 2. Purificación en tándem. Un lisado celular o tisular conteniendo una proteína de fusión con las regiones CBP-TSC-Proteína A es incubado con anticuerpos y microesferas conjugadas con el antígeno correspondiente. La adición de la proteasa TEV libera el complejo proteico de interés que es capturado con una resina conjugada con calmodulina. Finalmente, mediante un compuesto quelante de calcio se libera purificado el complejo de interés.

La biotinylación dependiente de proximidad se ha convertido en uno de los sistemas más eficientes para el estudio de la interactómica de proteínas individuales basado en purificación por afinidad y espectrometría de masas. En este método, la proteína de interés se fusiona a una forma mutada de la biotin-ligasa de E.coli, denominada BirA* [24]. Esta biotin-ligasa puede aún catalizar la formación de biotina activada pero rápidamente se disocia del intermediario y continúa la reacción. De esta manera, el gen fusionado a BirA* genera una nube de biotina activada in vivo que puede reaccionar con el grupo amino épsilon de los residuos de lisina de forma covalente. Estas uniones tan estables permiten el empleo de condiciones más drásticas de lisis celular dado que no es necesario mantener el complejo. A continuación, las proteínas marcadas con biotina se pueden purificar utilizando resinas de estreptavidina. Las proteínas purificadas son analizadas posteriormente por espectrometría de masas. Esta estrategia se ha utilizado para proteínas localizadas en distintos compartimentos celulares y vías de señalización [25, 26].

Una descripción completa de la técnica se puede encontrar en Lambert et al (2014) [27]. Como en otros experimentos de interactómica es muy importante la inclusión de controles negativos apropiados para evitar la detección de falsos positivos (por ejemplo, proteínas biotiniladas endógenamente o proteínas marcadas que resulten de una biotinylación promiscua) para ello se utilizan células transfectadas con los vectores sin el gen de interés o sin el gen de interés y sin BirA*. Generalmente se utilizan vectores tipo Gateway para facilitar el clonaje de la proteína de interés fusionada a BirA* bien en posición N-terminal o C-terminal. Asimismo, los vectores disponen de un tag como FLAG para poder llevar a cabo un “pull down” complementario, lo que permite llevar a cabo una comparativa con el método de biotinylación. De hecho aunque se observa cierto solapamiento en la población de proteínas identificadas por ambas técnicas, existen diferencias importantes que aconsejan el uso de ambas estrategias para obte-

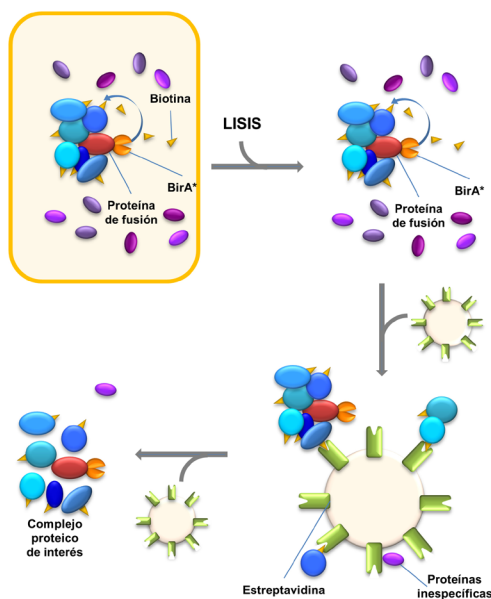


Figura 3. Biotinylación dependiente de proximidad. La expresión de una proteína de fusión con la biotinilasa BirA* en un contexto celular produce el marcaje con biotina de las proteínas próximas a dicha enzima. Tras la lisis, las proteínas biotiniladas son capturadas mediante una resina conjugada con estreptavidina. La purificación del complejo se logra en condiciones denaturizantes como alta temperatura o SDS.



ner una representación más completa del interactoma.

Uno de los aspectos fundamentales de la técnica reside en el empleo de un sistema inducible de expresión en células de mamífero derivado del sistema Flp-In T-REx, en el cual se pueden establecer líneas isogénicas y la expresión de proteínas se puede controlar mediante un promotor inducible por tetraciclina. Esta estrategia resulta en la integración del gen a expresar en una localización cromosómica constante, una población celular homogénea de expresión regulada bajo el control del promotor de tetraciclina. Sin embargo, esta ventaja constituye también una limitación en si misma dado que limita el empleo de esta técnica a un solo tipo celular y complica su uso en las líneas celulares (p.e. líneas establecidas de carcinomas) donde se han visto alteraciones en la expresión de dicha proteína. Para solucionar esta limitación, los mismos autores han desarrollado una nueva colección de vectores lentivirales para biotilación proximal que pueden ser utilizados en cualquier línea celular sin restricción [28], lo que facilita el estudio de la proteína de interés en su modelo celular más adecuado [29].

2.6. Enriquecimiento de factores de transcripción

El aislamiento e identificación por técnicas proteómicas de factores de transcripción ha sido especialmente difícil debido a la baja abundancia de estos, incluso previo aislamiento de los núcleos celulares [30]. Un sistema conocido como catTFRE, consistente en un ADN sintético que contiene una colección concatenada en tándem de secuencias consenso reconocidas por diversos factores de transcripción se ha empleado con éxito para la purificación de dichos factores [31]. Recientemente hemos utilizado este método para realizar un “pull down” de extractos nucleares en dos pares diferentes de células de cáncer colorrectal, KM12SM/KM12C y SW620/480, relacionadas genéticamente pero que difieren en la capacidad metastática. Los extractos nucleares de KM12 y SW se extrajeron con el fragmento de DNA catTFRE biotilado, recuperado después de la amplificación por PCR, se digirieron con Lys-C y tripsina, se analizaron mediante LC-MS “label-free” y se cuantificaron con MaxLFQ. Encontramos 240 proteínas que muestran una desregulación significativa en células KM12SM altamente metastáticas en relación con células KM12C no metastáticas y 257 proteínas alteradas en SW620 metastáticas vs SW480 [32]. Aunque el enfoque catTFRE se propuso inicialmente para el enriquecimiento de los factores de transcripción debido a la presencia de regiones de unión al concatémero, en nuestras manos también se enriquecieron muchas clases de proteínas de unión a ARN/ADN, unión a cromatina y daño al ADN en el procedimiento de extracción. Por lo tanto, la estrategia no fue particularmente selectiva para los factores de transcripción. En cualquier caso, identificamos un número significativo de factores de transcripción alterados en ambas líneas celulares.

2.7. Otros sistemas de identificación de complejos proteicos

El sistema CL7-Im7 tiene una gran afinidad entre sus constituyentes (*Colicin E7 DNase* y su inhibidor *Immunity protein 7*) proporcionando un sistema de purificación de alto rendimiento en un solo paso. La proteína a estudiar se marca con CL7 y los lisados se incuban con microesferas acopladas a Im7 [33].

El sistema de trampa vírica (*virotrap*) consiste en la fusión de la proteína de interés con el dominio C-terminal de la proteína gag del virus de la inmunodeficiencia humana (HIV) [34]. La presencia de gag en la célula provoca (incluso en ausencia del resto del virus) la formación de vesículas a partir de la membrana plasmática. La purificación de estas vesículas semejantes a virus se hace muy sencilla con la coexpresión de la proteína G del virus de la estomatitis vesicular (VSV-G) marcada con Flag, que se expresa en la membrana plasmática. Una vez purificadas, las vesículas se rompen y su contenido se analiza mediante las habituales técnicas proteómicas. La ventaja de este método es que elimina la necesidad de lisar las células y de lavados intensos, lo cual permite eliminar muchas interacciones falsas entre proteínas separadas espacialmente en distintos orgánulos celulares y a la vez mantener interacciones relativamente débiles. Sin embargo, actualmente este sistema sólo se ha empleado en la línea HEK-293T; adicionalmente, sólo sería útil para proteínas de citoplasma o el lado citoplasmático de la membrana plasmática. Posiblemente las proteínas más abundantes también aparezcan por azar en las vesículas sin que constituyan interacciones con la proteína de interés.

Para el aislamiento de proteínas de membrana plasmática celular, un sistema clásico es el marcaje de dichas proteínas con biotina mediante sulfo-NHS-SS-biotina. Tras parar la reacción, las células se lisan y se procede a la purificación de las proteínas marcadas con microesferas o geles de sepharosa conjugados con avidina o streptavidina [35]. Se utilizó SILAC para comparar dos líneas celulares de cáncer colorrectal: KM12C y KM12SM, de diferente capacidad metastática, para encontrar y cuantificar las diferencias en la expresión de proteínas, principalmente a nivel de la superficie celular. Después de la biotinilación seguida de purificación por afinidad, las proteínas de la membrana se separaron mediante SDS-PAGE y se analizaron utilizando nanoflujo LC-ESI-LTQ. Se identificaron un total de 291 proteínas de membrana y proteínas asociadas con un valor de $p < 0,01$, de las cuales se encontró que 60 proteínas se expresaban de forma diferencial por más de 1,5 veces. Identificamos una cantidad de señalización celular, CD, integrinas y otras moléculas de adhesión celular (CDH17, JUP) entre las proteínas más desreguladas [35].

Otros métodos de marcaje por proximidad relacionados con el sistema BioID son el EMARS, SPPLA y el APEX. EMARS (activación mediada por enzima de fuentes radi-cales) es un sistema que acopla la peroxidasa de rábano picante (HRP) a la proteína a estudiar o a una proteína a la que se une específicamente [36]. La adición de peróxido de hidrógeno y de aril-azida conjugada con biotina permite el marcaje con biotina de proteínas hasta 200-300 nm de la HRP, que serán purificadas por los métodos habituales (matrices de estreptavidina). El inconveniente de esta técnica se debe sobre todo al gran número de falsos positivos en la identificación de interactores debido a la activación de aril-azida por enzimas intracelulares. En parte esto se puede solventar mediante el uso de aril-azida conjugada con fluoresceína, dado que este compuesto es mucho menos activado por enzimas endógenos. Este sistema se usa con gran éxito en estudio de lipid rafts, donde la HRP se conjuga con la toxina colérica, que une gangliósidos específicos de estos dominios de membrana [37]. SPPLA (*selective proteomic proximity labeling assay using tyramide*) es muy similar a EMARS, pero usa como reactivo de marcaje tiramida conectada a biotina por un brazo de 12 carbonos con un puente disulfuro, lo cual permite una mayor accesibilidad de la estreptavidina a la biotina y la recuperación de las proteínas purificadas en la resina mediante un agente



reductor [38]. En contraste, APEX usa la ascorbato peroxidasa derivada del guisante o de la soja (denominada APEX, de ahí el nombre del método), la cual es más adecuada para estudios intracelulares dado que es muy poco sensible a cambios en pH o concentración de Ca^{2+} . APEX puede fusionarse con una proteína de interés o bien a un péptido señal para localizarla en un orgánulo y marcar así las proteínas de un compartimento celular exclusivo. Como en los métodos anteriores, la reacción comienza con la adición de peróxido de hidrogeno y en este caso de fenoles acoplados con biotina. Actualmente se usa una versión mejorada de la peroxidasa denominada APEX2, que tiene una actividad mayor y más duradera [39, 40].

2.8. Tecnologías de unión covalente (*cross-linking*)

Las tecnologías de "*cross-linking*" completan los métodos anteriores para aumentar la detección de interacciones débiles o transitorias [41]. Para ello, las interacciones entre proteínas se estabilizan mediante una unión covalente. Estos métodos se clasifican generalmente en función de la naturaleza de la molécula empleada como conector: (i) compuestos que reaccionan con el N-terminal de la proteína y los residuos de lisina, (ii) compuestos que reaccionan con el C-terminal de la proteína y los residuos de aspártico y glutámico, (iii) compuestos que reaccionan con los grupos tioles y (iv) compuestos foto-reactivos. Estas técnicas son muy dependientes de varios factores. Es clave la longitud de la molécula conectora empleada: cuanto más larga sea, mayor será el número de interacciones detectadas. Pero este aumento de la sensibilidad viene aparejado con una menor especificidad. La concentración y el tiempo de exposición a las moléculas conectoras tienen un gran impacto en el número y la especificidad de interacciones posibles detectadas.

Entre las moléculas conectoras se hallan el glutaraldehído y el DSSO (disuccinimidil-sulfóxido). Una de las más empleadas es el formaldehído que une proteínas a distancias muy cortas, permitiendo una gran especificidad, y es reversible mediante hervido de la muestra [42]. Usualmente se emplea con proteínas marcadas con estreptavidina. Tras la unión covalente con formaldehído se purifica mediante una matriz específica conjugada con biotina y tras la elución con biotina, la muestra se hierve y se analiza mediante el método usual. Esta técnica es conocida como SPINE (*strep-protein interaction experiment*) [43]. Similar es la técnica de inmunoprecipitación mediante unión covalente reversible (ReCLIP), donde se usan moléculas conectoras con tiol que pueden ser procesadas [44]. Esta técnica permite realizar la inmunoprecipitación en condiciones muy estrictas de manera que se aumente la especificidad, eluyendo el inmunoprecipitado mediante el cambio de pH.

Otra posibilidad en estas técnicas de *cross-linking* consiste en el enriquecimiento específico de los péptidos marcados con el conector (mediante cromatografía de exclusión, cromatografía de afinidad, marcaje isotópico, intercambio catiónico, etc.). Este abordaje se emplea en las técnicas PIR (*protein interaction reporter*), donde la molécula conectora va marcada a su vez con un tag, como por ejemplo biotina [45]. Especialmente diseñado para glicoproteínas de membrana que actúan como receptores se ha desarrollado la técnica LRC-TriCEPS (*Ligand-receptor capture-trifunctional chemoproteomics reagents*), que usa conectores con la doble capacidad de unir aminoácidos (para el ligando) y azúcares de las glicosilaciones (para el receptor), estando marcado

con biotina [46]. Tras la aplicación del conector, el lisado celular y la digestión proteica, los péptidos marcados son purificados con una resina conjugada con estreptavidina.

Un problema severo de las técnicas de *cross-linking* consiste en que muchos de los algoritmos de identificación usuales en espectrometría de masas no son compatibles con los marcajes covalentes, debido a la complejidad de las moléculas conectoras.

3. Análisis bioinformático. Eliminación de contaminantes.

El análisis del interactoma de una proteína comienza con el descarte de las proteínas inespecíficas existiendo para ello múltiples herramientas bioinformáticas (ver revisión [47]). De hecho, la mayoría de las proteínas identificadas mediante diferentes técnicas proteómicas no forman interacciones específicas con la proteína problema [1, 2, 48]. Para seleccionar los auténticos interactores, una vez tengamos los péptidos y proteínas identificados por el motor de búsqueda y la base de datos empleados, se eliminan las proteínas que aparecen en el control utilizado inmunoprecipitado con anticuerpo irrelevante, precipitado con proteína GST sin fusionar, etc.. Estas proteínas no interactuantes o bien se unen inespecíficamente a la matriz de purificación o bien son tan abundantes que siguen apareciendo tras los pasos de purificación. Se puede establecer un baremo para no eliminar proteínas que aunque se identifican en el control son mucho más abundantes en la muestra problema; por ejemplo, una diferencia de 3 veces en el número de péptidos identificados para una proteína o un determinado ratio en las intensidades de los péptidos. Asimismo se pueden descartar los contaminantes comunes en estos ensayos proteómicos utilizando la web *CRAPome* [49] (<http://crapome.org>). Otra web similar para eliminar falsos interactuantes se denomina *Straightforward Filtering Index*, SFINX [23] (<http://sfinx.ugent.be/>). Respecto al criterio de selección de proteínas, en caso de un único experimento se puede establecer que cada proteína haya sido identificada por al menos dos péptidos. Si se cuenta con réplicas nos podemos ceñir a las proteínas identificadas en varios replicados.

En ocasiones se utiliza la técnica de SILAC en combinación con los métodos de interactómica para determinar interacciones “bona fide” y proporcionar un sistema de identificación de interacciones cuantitativo. Aunque se puede aplicar a las diversas técnicas resulta especialmente útil cuando se intentan observar diferencias dinámicas en los complejos proteicos. Por ejemplo, el método QUICK (*quantitative immunoprecipitation combined with knockdown*) consiste en una inmunoprecipitación de una proteína en dos lisados provenientes de dos poblaciones de células marcadas con los isótopos ligeros y pesados, una de las cuales ha sido silenciada para la proteína a estudiar e inmunoprecipitar [50]. De esta manera se pueden eliminar muchas de las proteínas inespecíficas purificadas en la inmunoprecipitación. Otras técnicas de marcaje como ICAT o iTRAQ también pueden ser utilizadas para este tipo de estudios [51].

Otra fuente de contaminación en algunas estrategias se produce durante la lisis celular. Este paso puede desestabilizar las interacciones más débiles, generando falsos



positivos al poner en contacto proteínas que no podrían interactuar en condiciones fisiológicas al estar espacialmente separadas en diferentes localizaciones celulares [1]. Si se conoce la localización de la proteína estudiada, se pueden eliminar proteínas que in vivo no deberían interactuar con la proteína en cuestión. Por ejemplo, una proteína de membrana plasmática o secretada no debería interactuar con proteínas nucleares o mitocondriales, aunque conviene tener en cuenta que muchas proteínas tienen varias localizaciones diferentes donde pueden ejercer distintas funciones. Para conocer la localización de proteínas es especialmente útil la web COMPARTMENTS [52] (<https://compartments.jensenlab.org/Search>). Por último, se pueden eliminar o cuestionar aquellas proteínas cuya función implique la interacción transitoria con otras proteínas, tales como: proteínas ribosomales, chaperonas y proteínas involucradas en modificación posttraduccional o involucradas en la degradación (ubiquitinasa, proteínas del proteosoma). En caso de proteínas de membrana o secretadas se incluirían aquí también las proteínas implicadas en transporte de vesículas entre diferentes compartimentos. Estas proteínas pueden ser relevantes para conocer como la proteína es transportada, glicosilada o degradada, es decir como es regulada.

El resto de proteínas identificadas en la inmunoprecipitación por anticuerpos se pueden considerar que interactúan específicamente con la proteína a estudiar, aunque deben ser validadas posteriormente utilizando técnicas ortogonales. Las interacciones ya publicadas se pueden consultar en BioGRID (*Biological General Repository for Interaction Datasets*) [<https://thebiogrid.org/>]. Para determinar las funciones de las proteínas identificadas existen numerosas herramientas bioinformáticas que podemos emplear. Una de las más empleadas es STRING (*Search Tool for the Retrieval of Interacting Genes/Proteins*), que proporciona una visualización de las interacciones entre proteínas conocidas o predichas por diferentes fuentes [<https://string-db.org/>]. Adicionalmente proporciona información sobre los procesos biológicos, funciones moleculares y componentes celulares. Muy similar es DAVID (*Database for Annotation, Visualization and Integrated Discovery*) que además proporciona la identificación de vías de señalización, asociación con enfermedades y expresión por tejidos [<https://david.ncifcrf.gov/tools.jsp>]. Una descripción más detallada de cada proteína se puede consultar en la web GeneCards [<https://www.genecards.org/>], mientras que para rutas moleculares es muy útil KEGG (*Kyoto Encyclopedia of Genes and Genomes*) [<https://www.genome.jp/kegg/>].

4. Conclusiones

Se han revisado aquí algunas de las técnicas de purificación por afinidad más utilizadas en la caracterización del interactoma, indicando algunas de las limitaciones y dificultades asociadas. En este tipo de análisis, más que en ningún otro, son esenciales los controles de calidad para evitar la detección de proteínas inespecíficas, tanto de falsos positivos como de falsos negativos. Siempre es necesario llevar a cabo validaciones de tipo ortogonal mediante experimentos de coinmunoprecipitación, microscopía confocal, afinidad, etc. que permitan validar inequívocamente las observaciones obtenidas por espectrometría de masas. Por otro lado, hay que tener en cuenta que la mayoría de las técnicas descritas ofrecen fundamentalmente un punto de vista estáti-

co dando lugar a solapamientos incompletos entre las interacciones proteína-proteína capturadas por las diferentes estrategias [53, 54]. Sin embargo, las redes de interacción proteica son extremadamente dinámicas y responden de forma dispar frente a estímulos externos. En resumen, las redes de proteínas presentan una extraordinaria complejidad que requiere múltiples estudios para llegar a las conclusiones correctas. En este sentido, estrategias como BioID permiten obtener representaciones más completas y una imagen mucho más dinámica de las diferentes interacciones de una proteína en un periodo largo de tiempo que la foto fija obtenida por estrategias como la inmunoprecipitación o el “pull down”.

REFERENCIAS

1. Titeca, K., I. Lemmens, J. Tavernier, and S. Eyckerman, Discovering cellular protein-protein interactions: Technological strategies and opportunities. *Mass Spectrom Rev*, 2019. 38(1): p. 79-111.
2. Gingras, A.C., M. Gstaiger, B. Raught, and R. Aebersold, Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol*, 2007. 8(8): p. 645-54.
3. Kuzmanov, U. and A. Emili, Protein-protein interaction networks: probing disease mechanisms using model systems. *Genome Med*, 2013. 5(4): p. 37.
4. Ryan, D.P. and J.M. Matthews, Protein-protein interactions in human disease. *Curr Opin Struct Biol*, 2005. 15(4): p. 441-6.
5. Hopkins, A.L., Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*, 2008. 4(11): p. 682-90.
6. Bartolome, R.A., I. Garcia-Palmero, S. Torres, M. Lopez-Lucendo, I.V. Balyasnikova, and J.I. Casal, IL13 Receptor alpha2 Signaling Requires a Scaffold Protein, FAM120A, to Activate the FAK and PI3K Pathways in Colon Cancer Metastasis. *Cancer Res*, 2015. 75(12): p. 2434-44.
7. Ivanov, A.S., V.G. Zgoda, and A.I. Archakov, [Protein interactomics technologies]. *Bioorg Khim*, 2011. 37(1): p. 8-21.
8. Young, K.H., Yeast two-hybrid: so many interactions, (in) so little time. *Biol Reprod*, 1998. 58(2): p. 302-11.
9. Fields, S. and R. Sternglanz, The two-hybrid system: an assay for protein-protein interactions. *Trends Genet*, 1994. 10(8): p. 286-92.
10. Parrish, J.R., K.D. Gulyas, and R.L. Finley, Jr., Yeast two-hybrid contributions to interactome mapping. *Curr Opin Biotechnol*, 2006. 17(4): p. 387-93.
11. Florinskaya, A., P. Ershov, Y. Mezentsev, L. Kaluzhskiy, E. Yablokov, A. Medvedev, and A. Ivanov, SPR Biosensors in Direct Molecular Fishing: Implications for Protein Interactomics. *Sensors (Basel)*, 2018. 18(5).
12. Bartolome, R.A., R. Barderas, S. Torres, M.J. Fernandez-Acenero, M. Mendes, J. Garcia-Foncillas, M. Lopez-Lucendo, and J.I. Casal, Cadherin-17 interacts with alpha2beta1 integrin to regulate cell proliferation and adhesion in colorectal cancer cells causing liver metastasis. *Oncogene*, 2014. 33(13): p. 1658-69.
13. Bartolome, R.A., A. Pelaez-Garcia, I. Gomez, S. Torres, M.J. Fernandez-Acenero, B. Escudero-Paniagua, J.I. Imbaud, and J.I. Casal, An RGD motif present in cadherin 17 induces integrin activation and tumor growth. *J Biol Chem*, 2014. 289(50): p. 34801-14.
14. Bartolome, R.A., S. Torres, S. Isern de Val, B. Escudero-Paniagua, E. Calvino, J. Teixido, and J.I. Casal, VE-cadherin RGD motifs promote metastasis and constitute a potential therapeutic target in melanoma and breast cancers. *Oncotarget*, 2017. 8(1): p. 215-227.
15. Mendes, M., A. Pelaez-Garcia, M. Lopez-Lucendo, R.A. Bartolome, E. Calvino, R. Barderas, and J.I. Casal, Mapping the Spatial Proteome of Metastatic Cells in Colorectal Cancer. *Proteomics*, 2017. 17.
16. Karginov, A. and M. Agaphonov, A simple enrichment procedure improves detection of membrane proteins by immunoblotting. *Biotechniques*, 2016. 61(5): p. 260-261.
17. Smith, D.B. and K.S. Johnson, Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*, 1988. 67(1): p. 31-40.
18. Bornhorst, J.A. and J.J. Falke, Purification of proteins using polyhistidine affinity tags. *Methods Enzymol*, 2000. 326: p. 245-54.
19. Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, and B. Seraphin, The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, 2001. 24(3): p. 218-29.
20. Burckstummer, T., K.L. Bennett, A. Preradovic, G. Schutze, O. Hantschel, G. Superti-Furga, and A. Bauch, An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat Methods*, 2006. 3(12): p. 1013-9.
21. Gavin, A.C., P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L.J. Jensen, S. Bastuck, B.



Dumpelfeld, A. Edelmann, M.A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A.M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J.M. Rick, B. Kuster, P. Bork, R.B. Russell, and G. Superti-Furga, Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 2006. 440(7084): p. 631-6.

22. Li, Y., S. Franklin, M.J. Zhang, and T.M. Vondriska, Highly efficient purification of protein complexes from mammalian cells using a novel streptavidin-binding peptide and hexahistidine tandem tag system: application to Bruton's tyrosine kinase. *Protein Sci*, 2011. 20(1): p. 140-9.

23. Titeca, K., P. Meysman, K. Gevaert, J. Tavernier, K. Laukens, L. Martens, and S. Eyckerman, SFINX: Strai-ghtforward Filtering Index for Affinity Purification-Mass Spectrometry Data Analysis. *J Proteome Res*, 2016. 15(1): p. 332-8.

24. Roux, K.J., D.I. Kim, M. Raida, and B. Burke, A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol*, 2012. 196(6): p. 801-10.

25. Kim, D.I., K.C. Birendra, W. Zhu, K. Motamedchaboki, V. Doye, and K.J. Roux, Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc Natl Acad Sci U S A*, 2014. 111(24): p. E2453-61.

26. Couzens, A.L., J.D. Knight, M.J. Kean, G. Teo, A. Weiss, W.H. Dunham, Z.Y. Lin, R.D. Bagshaw, F. Sicheri, T. Pawson, J.L. Wrana, H. Choi, and A.C. Gingras, Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. *Sci Signal*, 2013. 6(302): p. rs15.

27. Lambert, J.P., M. Tucholska, C. Go, J.D. Knight, and A.C. Gingras, Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J Proteomics*, 2015. 118: p. 81-94.

28. Samavarchi-Tehrani, P., H. Abdouni, R. Samson, and A.C. Gingras, A Versatile Lentiviral Delivery Toolkit for Proximity-dependent Biotinylation in Diverse Cell Types. *Mol Cell Proteomics*, 2018. 17(11): p. 2256-2269.

29. Knight, J.F., V.Y.C. Sung, E. Kuzmin, A.L. Couzens, D.A. de Verteuil, C.D.H. Ratcliffe, P.P. Coelho, R.M. Johnson, P. Samavarchi-Tehrani, T. Grussio, H.W. Smith, W. Lee, S.M. Saleh, D. Zuo, H. Zhao, M.C. Guiot, R.R. Davis, J.P. Gregg, C. Moraes, A.C. Gingras, and M. Park, KIBRA (WWC1) Is a Metastasis Suppressor Gene Affected by Chromosome 5q Loss in Triple-Negative Breast Cancer. *Cell Rep*, 2018. 22(12): p. 3191-3205.

30. Giamb Bruno, R., F. Grebien, A. Stukalov, C. Knoll, M. Planyavsky, E.L. Rudashevskaya, J. Colinge, G. Superti-Furga, and K.L. Bennett, Affinity purification strategies for proteomic analysis of transcription factor complexes. *J Proteome Res*, 2013. 12(9): p. 4018-27.

31. Ding, C., D.W. Chan, W. Liu, M. Liu, D. Li, L. Song, C. Li, J. Jin, A. Malovannaya, S.Y. Jung, B. Zhen, Y. Wang, and J. Qin, Proteome-wide profiling of activated transcription factors with a concatenated tandem array of transcription factor response elements. *Proc Natl Acad Sci U S A*, 2013. 110(17): p. 6771-6.

32. Torres, S., I. Garcia-Palmero, C. Marin-Vicente, R.A. Bartolome, E. Calvino, M.J. Fernandez-Acenero, and J.I. Casal, Proteomic Characterization of Transcription and Splicing Factors Associated with a Metastatic Phenotype in Colorectal Cancer. *J Proteome Res*, 2018. 17(1): p. 252-264.

33. Vassilyeva, M.N., S. Klyuyev, A.D. Vassilyev, H. Wesson, Z. Zhang, M.B. Renfrow, H. Wang, N.P. Higgins, L.T. Chow, and D.G. Vassilyev, Efficient, ultra-high-affinity chromatography in a one-step purification of complex proteins. *Proc Natl Acad Sci U S A*, 2017. 114(26): p. E5138-E5147.

34. Eyckerman, S., K. Titeca, E. Van Quickelberghe, E. Cloots, A. Verhee, N. Samyn, L. De Ceuninck, E. Timmerman, D. De Sutter, S. Lievens, S. Van Calenberg, K. Gevaert, and J. Tavernier, Trapping mammalian protein complexes in viral particles. *Nat Commun*, 2016. 7: p. 11416.

35. Luque-Garcia, J.L., J.L. Martinez-Torrecuadrada, C. Epifano, M. Canamero, I. Babel, and J.I. Casal, Differential protein expression on the cell surface of colorectal cancer cells associated to tumor metastasis. *Proteomics*, 2010. 10(5): p. 940-52.

36. Jiang, S., N. Kotani, T. Ohnishi, A. Miyagawa-Yamguchi, M. Tsuda, R. Yamashita, Y. Ishiura, and K. Honke, A proteomics approach to the cell-surface interactome using the enzyme-mediated activation of radical sources reaction. *Proteomics*, 2012. 12(1): p. 54-62.

37. Miyagawa-Yamaguchi, A., N. Kotani, and K. Honke, Expressed glycosylphosphatidylinositol-anchored horseradish peroxidase identifies co-clustering molecules in individual lipid raft domains. *PLoS One*, 2014. 9(3): p. e93054.

38. Rees, J.S., X.W. Li, S. Perrett, K.S. Lilley, and A.P. Jackson, Selective Proteomic Proximity Labeling Assay Using Tyramide (SPPLAT): A Quantitative Method for the Proteomic Analysis of Localized Membrane-Bound Protein Clusters. *Curr Protoc Protein Sci*, 2015. 80: p. 19 27 1-18.

39. Martell, J.D., T.J. Deerinck, Y. Sancak, T.L. Poulos, V.K. Mootha, G.E. Sosinsky, M.H. Ellisman, and A.Y. Ting, Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat Biotechnol*, 2012. 30(11): p. 1143-8.

40. Rhee, H.W., P. Zou, N.D. Udeshi, J.D. Martell, V.K. Mootha, S.A. Carr, and A.Y. Ting, Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*, 2013. 339(6125): p. 1328-1331.

41. Sinz, A., Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev*, 2006. 25(4): p. 663-82.

42. Sutherland, B.W., J. Toews, and J. Kast, Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions. *J Mass Spectrom*, 2008. 43(6): p. 699-715.

43. Herzberg, C., L.A. Weidinger, B. Dorrbecker, S. Hubner, J. Stulke, and F.M. Commichau, SPINE: a method for the rapid detection and analysis of protein-protein interactions in vivo. *Proteomics*, 2007. 7(22): p. 4032-5.

44. Smith, A.L., D.B. Friedman, H. Yu, R.H. Carnahan, and A.B. Reynolds, ReCLIP (reversible cross-link immuno-precipitation): an efficient method for interrogation of labile protein complexes. *PLoS One*, 2011. 6(1): p. e16206.

45. Tang, X. and J.E. Bruce, A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies. *Mol Biosyst*, 2010. 6(6): p. 939-47.

46. Frei, A.P., O.Y. Jeon, S. Kilcher, H. Moest, L.M. Henning, C. Jost, A. Pluckthun, J. Mercer, R. Aebersold,

- E.M. Carreira, and B. Wollscheid, Direct identification of ligand-receptor interactions on living cells and tissues. *Nat Biotechnol*, 2012. 30(10): p. 997-1001.
47. Nesvizhskii, A.I., Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. *Proteomics*, 2012. 12(10): p. 1639-55.
48. Dunham, W.H., M. Mullin, and A.C. Gingras, Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics*, 2012. 12(10): p. 1576-90.
49. Mellacheruvu, D., Z. Wright, A.L. Couzens, J.P. Lambert, N.A. St-Denis, T. Li, Y.V. Miteva, S. Hauri, M.E. Sardi, T.Y. Low, V.A. Halim, R.D. Bagshaw, N.C. Hubner, A. Al-Hakim, A. Bouchard, D. Faubert, D. Fermin, W.H. Dunham, M. Goudreault, Z.Y. Lin, B.G. Badillo, T. Pawson, D. Durocher, B. Coulombe, R. Aebersold, G. Superti-Furga, J. Colinge, A.J. Heck, H. Choi, M. Gstaiger, S. Mohammed, I.M. Cristea, K.L. Bennett, M.P. Washburn, B. Raught, R.M. Ewing, A.C. Gingras, and A.I. Nesvizhskii, The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods*, 2013. 10(8): p. 730-6.
50. Selbach, M. and M. Mann, Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK). *Nat Methods*, 2006. 3(12): p. 981-3.
51. Ranish, J.A., E.C. Yi, D.M. Leslie, S.O. Purvine, D.R. Goodlett, J. Eng, and R. Aebersold, The study of macromolecular complexes by quantitative proteomics. *Nat Genet*, 2003. 33(3): p. 349-55.
52. Binder, J.X., S. Pletscher-Frankild, K. Tsafou, C. Stolte, S.I. O'Donoghue, R. Schneider, and L.J. Jensen, COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database (Oxford)*, 2014. 2014: p. bau012.
53. von Mering, C., R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, and P. Bork, Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 2002. 417(6887): p. 399-403.
54. Bader, J.S., A. Chaudhuri, J.M. Rothberg, and J. Chant, Gaining confidence in high-throughput protein interaction networks. *Nat Biotechnol*, 2004. 22(1): p. 78-85.



Proteogenómica

*José González-Gomariz^{1,2}, Víctor Segura^{1,2},
Elizabeth Guruceaga^{1,2}*

¹Plataforma de Bioinformática del Centro para la Investigación Médica Aplicada (CIMA).

²Instituto de Investigación Sanitaria de Navarra (IdiSNA).

1. Introducción.

La secuenciación completa del genoma humano en el año 2000 desentrañó el primer nivel de complejidad de la biología humana y supuso la aparición de nuevos métodos analíticos en el ámbito de la investigación biomédica. Desde entonces se han desarrollado una gran cantidad de proyectos cuyo objetivo es la identificación de patrones de expresión génica asociados a los distintos tipos y subtipos de tumores, y la identificación de mutaciones somáticas incluyendo cambios en el número de copias (deleciones y amplificaciones de ADN) y mutaciones puntuales (SNVs). Todos estos resultados están siendo incorporados a la investigación en oncología y se han obtenido diversas evidencias de que este conocimiento puede servir para mejorar los tratamientos y ayudar a la detección precoz de enfermedades (1). En aplicaciones como la clasificación de tumores empleando sus características moleculares, se ha incrementado la capacidad para distinguir subtipos dentro de un mismo tipo tumoral permitiendo un pronóstico y un tratamiento más adecuado gracias a la estratificación de los pacientes (2). Por ejemplo, la presencia o ausencia de una mutación en un gen determinado puede tener un efecto determinante en la respuesta del paciente al tratamiento (3).

Tras los mencionados avances en genómica, el estudio de las proteínas codificadas por el genoma (proteoma) es el siguiente paso para comprender la fisiología humana, tanto en individuos sanos como enfermos. El estudio del proteoma, a pesar de presentar similitudes con el estudio del genoma, supone un gran desafío para la comunidad científica. La Organización del Proteoma Humano (HUPO, <http://www.hupo.org>) está coordinando las iniciativas para describir el proteoma humano de una forma sistemática, así como su implicación en las distintas enfermedades. En concreto, a través del Proyecto Proteoma Humano (HPP), un esfuerzo internacional que aglutina a centros de investigación de 22 países del mundo y del que forma parte el consorcio español del HPP (spHPP). En esta tarea se está empleando la tecnología que se considera actualmente más potente para la realización de experimentos de proteómica a gran escala, la espectrometría de masas (MS) (4).

La Proteogenómica, un área de estudio que integra la Genómica y la Proteómica, tiene mucho que aportar en el proceso de caracterización del proteoma humano al igual



que ya ha demostrado ser de utilidad en otros ámbitos (5,6). Su objetivo principal es lograr la anotación detallada de cada gen mediante la asignación de péptidos a los espectros detectados en experimentos de MS utilizando datos de secuencia de genes o transcritos a nivel de genoma completo, DNA-Seq o RNA-Seq (7). Esta metodología se ha aplicado con éxito en organismos no utilizados como modelo y que por ello carecen de bases de datos de referencia para sus proteínas (8). Sin embargo, la integración de grandes cantidades de datos de RNA-Seq y MS supone un gran reto, como por ejemplo la generación eficiente de bases de datos no redundantes a partir de los datos de RNA-Seq (9). Estas dificultades se han abordado con éxito permitiendo la integración de la cuantificación del proteoma humano a gran escala con la información del transcriptoma y de los polimorfismos del ADN, y revelar así los múltiples y diversos mecanismos reguladores de la expresión génica (10).

El integrar datos de proteómica con datos de transcriptómica y genómica a gran escala proporciona información valiosa con distintas aplicaciones, ya que los péptidos así identificados proporcionan información única para la anotación de genes, la confirmación de la traducción de un transcrito o la predicción de nuevos genes (11). Además, los patrones de expresión génica específicos de tejido o tipo celular pueden guiar la búsqueda de proteínas hasta ahora no detectadas mediante MS o la identificación de variantes de proteínas resultantes del proceso de corte y empalme alternativo, polimorfismos de aminoácidos o modificaciones postraduccionales (12). En el caso del HPP se ha utilizado la información de la expresión génica para diseñar experimentos de proteómica dirigidos hacia la identificación de proteínas para las que aún no hay evidencias experimentales fiables (13–17). La detección de polimorfismos de aminoácidos (SAPs) es otra aplicación relevante, ya que en el ámbito clínico los SAPs pueden ser responsables, por ejemplo, de la resistencia a antibióticos de una bacteria o de su virulencia (18). También se han descrito SAPs relacionados con el cáncer colorrectal (19), asociados a obesidad y diabetes (20), o como biomarcadores de determinadas enfermedades (21). Además empleando esta estrategia es posible detectar péptidos generados a partir de la traducción de transcritos que no se habían descrito previamente o que eran considerados no codificantes (péptidos noveles) en todo tipo de muestras biológicas. En los siguientes apartados profundizaremos en las aplicaciones de la Proteogenómica ya mencionadas.

2. Proteogenómica en el Proyecto Proteoma Humano (HPP).

El HPP es un proyecto diseñado en 2010 con dos iniciativas principales: caracterizar y definir métodos analíticos para cada una de las proteínas del proteoma humano utilizando las técnicas emergentes disponibles en ese momento (C-HPP), y proporcionar los medios para poder identificar y cuantificar las proteínas en distintas condiciones experimentales y patologías, entre ellas cáncer, enfermedades metabólicas, enfermedades reumatoideas, etc. (BD-HPP). La estrategia utilizada para desarrollar el proyecto C-HPP consistió en dividir el proteoma de manera que grupos científicos de diferentes nacionalidades caracterizaran cada uno de los cromosomas (12,13). Los 24

cromosomas más el proteoma codificado por el genoma mitocondrial se repartieron entre 25 equipos científicos de 22 países diferentes. A pesar de utilizar la tecnología más puntera, tanto la identificación de proteínas de baja abundancia como distinguir las isoformas de las proteínas presentes en una célula determinada suponen un gran reto. La gran heterogeneidad de las proteínas y su amplio rango dinámico, entre otras razones, restringen la capacidad de detección y cuantificación a gran escala de las proteínas (4). Por ello las proteínas que todavía no han sido detectadas mediante experimentos de MS han sido denominadas "*missing proteins*". Los grupos que participan en el C-HPP están utilizando la Proteogenómica para integrar datos transcriptómicos y proteómicos, confiando en los métodos de secuenciación de ARN para guiar los análisis de proteómica a gran escala (13). De esta forma, mediante la Transcriptómica es posible sugerir líneas celulares preferenciales donde detectar determinadas proteínas. Esta actividad del consorcio ha generado nuevos métodos proteogenómicos y herramientas bioinformáticas (22,23).

El spHPP está contribuyendo a esta iniciativa enfocándose en la detección y anotación biológica de los genes y proteínas del cromosoma 16. El consorcio español estableció su estrategia general y objetivos durante una primera reunión celebrada en Madrid (España) el 2 de abril de 2012. Se basa en la plataforma ProteoRed perteneciente al Instituto de Salud Carlos III (ISCIII), un consorcio de proteómica integrado por 21 laboratorios de proteómica con más de 11 años de experiencia en la coordinación de actividades multicéntricas, compartiendo tecnología de vanguardia, estandarización de datos, bioinformática e investigación. Tanto los datos públicos o generados en el propio consorcio, como los métodos bioinformáticos desarrollados durante estos años para la detección y caracterización de las proteínas "*missing*" se detallan a continuación.

2.1. Fuentes de información para la anotación del genoma y el proteoma en el contexto proteogenómico del C-HPP

Aunque se han publicado ya varios borradores del proteoma humano basados en MS en los que se identificaron más del 95% de las proteínas humanas (24,25), aproximadamente el 15% del proteoma humano no ha sido detectado de forma fiable. Es decir, evitando asignaciones de espectros a secuencias de péptido (PSM) ambiguas que provocan identificaciones dudosas a nivel de proteína. La iniciativa C-HPP se ha propuesto utilizar únicamente información de detección de proteínas que cumpla criterios estrictos de asignación para evitar dichas identificaciones erróneas. NeXtProt es la base de datos de referencia para la anotación de las proteínas humanas que sigue estos principios (26). Esta base de datos asigna evidencia experimental para cada proteína humana distinguiendo cinco niveles: PE1 (evidencia experimental a nivel de proteína), PE2 (evidencia experimental a nivel de transcrito), PE3 (proteína inferida por homología), PE4 (proteína predicha a partir de la secuencia de un gen) y PE5 (proteína incierta por la dudosa posibilidad de que se traduzca). En la Tabla 1 se muestra la información de la versión 2018-01-17 de neXtProt para cada cromosoma.

Para generar la clasificación mencionada, neXtProt integra información de UniProtKB, genómica (dbSNP, COSMIC), transcriptómica (BGee), detección de proteínas con anticuerpos (Human Protein Atlas, HPA) y en base a experimentos de espectrometría



Cromosoma	Entradas neXtProt	Genes codificantes proteínas	Nivel de proteína (PE1)	Nivel de transcrito (PE2)	Nivel de homología (PE3)	Proteínas predichas (PE4)	Inciertas (PE5)	Proteínas "missing"
1	2063	2041	1793	149	71	8	51	228
2	1273	1249	1160	81	18	1	16	100
3	1077	1072	960	81	18	3	15	102
4	762	752	677	50	20	1	20	71
5	872	881	776	76	9	3	10	88
6	1111	1040	1001	89	14	6	29	109
7	939	907	791	93	9	4	43	106
8	703	668	598	59	14	2	41	75
9	809	776	672	70	21	14	34	105
10	750	726	660	65	8	1	16	74
11	1325	1305	1106	181	97	2	40	280
12	1031	1036	923	70	15	2	22	87
13	329	321	292	22	2	1	12	25
14	677	613	588	44	27	3	15	74
15	600	597	507	48	14	1	32	63
16	837	860	745	67	9	1	23	77
17	1166	1183	1038	89	13	4	23	106
18	276	269	248	15	1	2	10	18
19	1433	1472	1259	114	27	2	33	143
20	550	543	478	51	3	5	13	59
21	252	230	191	22	15	1	23	38
22	497	442	426	44	5	1	21	50
X	838	842	731	86	22	6	29	114
Y	47	45	37	12	1	0	7	13
MT	15	13	15	0	0	0	0	0
Todo	20232	19883	17672	1678	453	74	578	2205

Tabla 1. Número de proteínas de cada cromosoma y su correspondiente evidencia experimental según la versión 2018-01-17 de neXtProt (26).

de masas (PeptideAtlas), información de estructura 3D (Protein Data Bank, PDB), y modificaciones post-traduccionales extraídas de publicaciones (<https://www.nextprot.org>). Las proteínas de los niveles PE2, PE3 y PE4 son las consideradas como proteínas “missing”, cuyo número ha ido modificándose en los últimos años debido a la evolución de los criterios utilizados para dar como detectada una proteína y al trabajo realizado para la detección fiable y robusta de estas proteínas cumpliendo dichos criterios (27). En concreto, las evidencias MS de una proteína son consideradas válidas cuando: 1) la tasa de falsos positivos o FDR es menor del 1% a nivel de PSM, péptido y proteína; 2) se ha detectado más de 1 péptido de al menos 9 aminoácidos; 3) por lo menos 2 de dichos péptidos son proteotípicos, péptidos únicos de dicha proteína que no son compartidos con otras proteínas teniendo en cuenta las distintas posibles variantes o SNVs de los mismos.

Una vez bien definidas las proteínas “missing”, los esfuerzos se han dirigido hacia su detección. Se ha analizado una cantidad ingente de datos, incluyendo nuevos experimentos y datos públicos de líneas celulares, tejidos normales y tumorales tanto de microarrays o de RNA-Seq para el transcriptoma como de proteómica *shotgun* para el proteoma (14–17,28). El gran número y la diversidad de las muestras disponibles facilitan el uso de la Proteogenómica para localizar el tejido específico o la condición

experimental en la que se expresa una proteína “missing” determinada. Hay que agradecer a iniciativas como ENCODE (Encyclopedia of DNA Elements), HBM (Human BodyMap 2.0 de Illumina), CCLE (29), expO (Expression Project for Oncology) y NCI60 (30), el que pongan a disposición de la comunidad científica datos de tanto valor. En la Tabla 2 se resume el tipo, número y disponibilidad de los datos.

PROYECTO	FUENTE	MUESTRA	DISPONIBILIDAD	TIPO DE EXPERIMENTO	REFERENCIAS
TRANSCRIPTOMA					
spHPP	Líneas celulares	MCF7, CCD18, Jurkat	GSE40168	Microarray	(14)
HBM	Tejidos humanos normales	16 tejidos humanos	GSE30611	RNA-seq	(15)
ENCODE	Líneas celulares	MCF7, k562, HepG2	http://genome.org.es	RNA-seq	(23)
CCLE	Líneas celulares	917 líneas celulares	GSE36133	microarray, RNA-seq	(16),(17)
expO	Tejidos humanos tumorales	2158 muestras de 156 tejidos distintos	GSE2109	microarray	(16)
Roth et al. (2006)	Tejidos humanos normales	352 muestras de 65 tejidos distintos	GSE3526	microarray	(16)
PROTEOMA					
spHPP	Líneas celulares	MCF7, CCD18, Jurkat	PXD27330-PXD27369	Shotgun	(14)
spHPP	Líneas celulares	MCF-7, Jurkat, RAMOS, CCD-18Co	PXD000442, PXD000443, PXD000447, PXD000449	Shotgun	(15)
PRIDE	Líneas celulares y tejidos	8 tejidos humanos	PXD001468, PXD002367, PXD001242, PXD000754, PXD000605, PXD000004, PXD000269, PXD002145	Shotgun	(28)
NCI-60	Líneas celulares	59 líneas celulares	http://129.187.44.58:7070/NCI60/main/index	Shotgun	(17)

Tabla 2. Datos públicos utilizados para identificar proteínas “missing” mediante la integración de información transcriptómica y proteómica. spHPP (spanish Human Proteome Project), CCLE (Cancer Cell Line Encyclopedia), expO (expression project for Oncology), PRIDE (The PRoteomics IDentifications database) NCI60 (Cancer cell line panel of National Cancer Institute). Los datos con identificación “PXD” proceden de la base de datos PRIDE.

2.2. Métodos bioinformáticos desarrollados para la detección de proteínas “missing”

La información mencionada en el apartado anterior ha sido analizada en profundidad mediante diversos flujos de análisis para definir el conjunto de genes que se expresan en cada una de las miles de muestras consideradas y así dirigir la búsqueda de las proteínas de interés. Además se han desarrollado métodos bioinformáticos para aumentar la eficacia en la selección de la muestra adecuada (16) y para mejorar la cobertura obtenida en los experimentos de proteómica (17). Se han conseguido predecir las muestras biológicas en las que es más probable la detección de las proteínas



“missing” aplicando un método de clasificación Bayesiano sobre datos de expresión génica de más de 3400 experimentos de microarrays, y de las características de los genes codificantes de proteínas (Figura 1) (16).

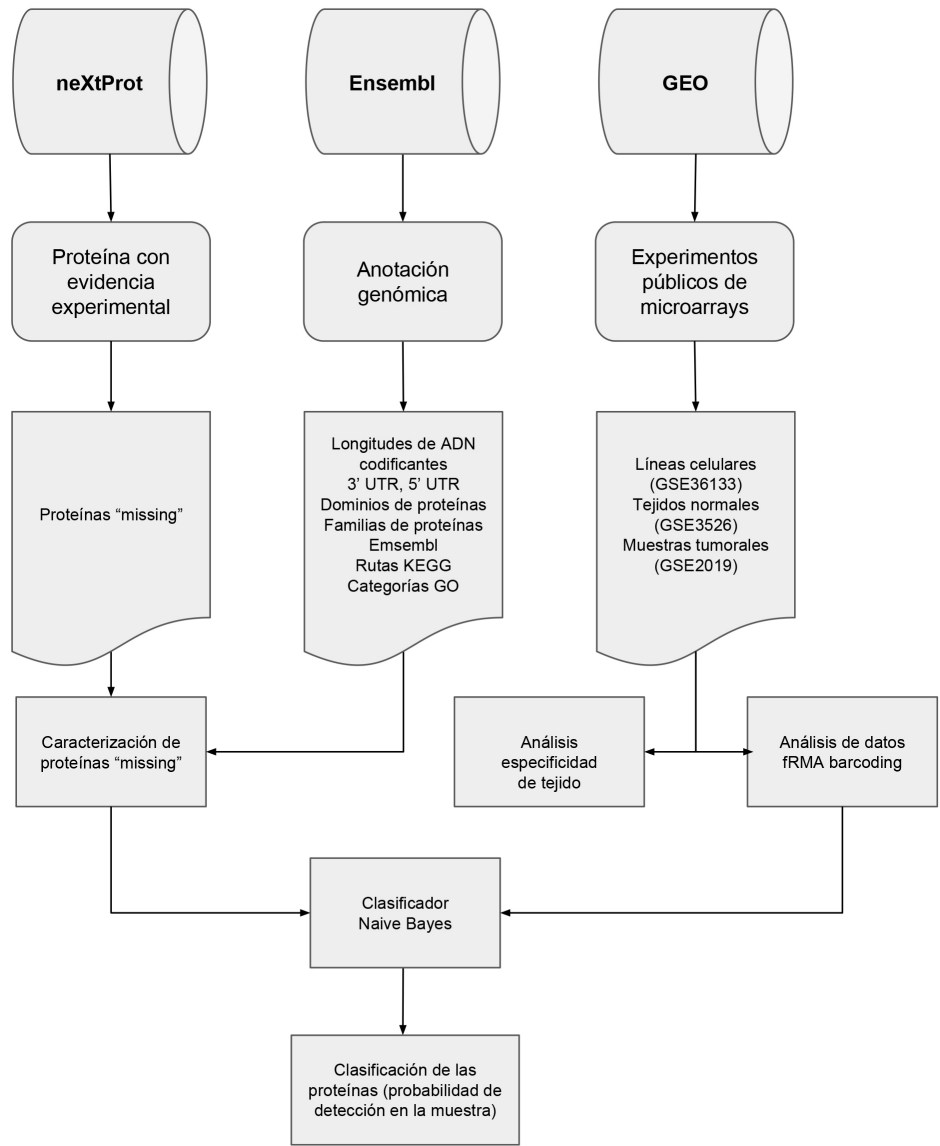


Figura 1. Flujo de análisis bioinformático desarrollado para la caracterización y localización de las proteínas "missing". Los experimentos públicos de microarrays de expresión utilizados en el análisis se detallan en la figura.

La caracterización de las proteínas "missing" (Figuras 2a y 2b) nos ha permitido descubrir que los genes codificantes de estas proteínas tienen una menor expresión y son más cortas que los genes codificantes de las proteínas que sí son detectadas por MS. Además, el clasificador desarrollado es capaz de proponer la muestra en la que será más fácil detectar una proteína determinada (Figura 2c), y señala los tejidos en los

que es posible detectar un mayor número de proteínas “missing”, que son testículo, cerebro y músculo esquelético. Estos resultados fueron incluidos en la actualización del estado del proyecto HPP en 2015 (31).

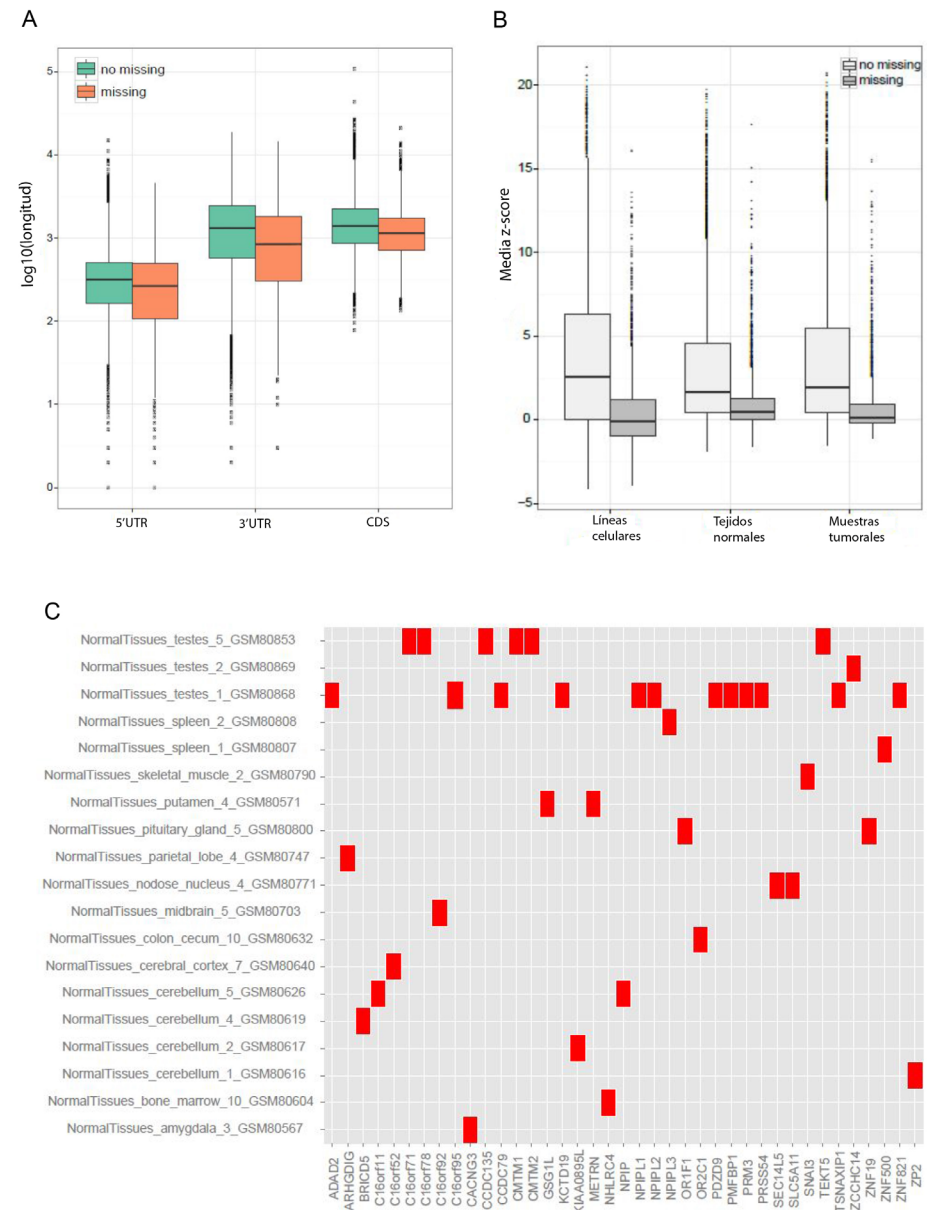


Figura 2. Caracterización y localización de las proteínas “missing”. (A) Distribución de la longitud de las regiones 5' UTR, 3' UTR y codificante (CDS) para los genes codificantes de proteínas “missing” y de proteínas detectables (“nonmissing”). (B) Distribución de los valores medios del z-score de expresión para las líneas celulares, tejidos normales y muestras tumorales. (C) Proteínas “missing” del cromosoma 16 y tejidos normales donde es más probable que se estén expresando.

Por otra parte, se ha propuesto también la integración de los resultados de diferentes motores de búsqueda para aumentar la cobertura del proteoma que somos capaces de detectar en un experimento de proteómica. En el caso del spHPP se han incluido los motores de búsqueda Comet, Mascot, OMSSA y X!Tandem en el análisis proteo-



genómico de las líneas tumorales caracterizadas por la iniciativa NCI60 (Figura 3) (17).

Los resultados de estos motores de búsqueda se complementan aumentando la posibilidad de detectar péptidos que cumplan todos los criterios establecidos por el C-HPP para identificar proteínas “missing”. En total se realizaron más de 3600 búsquedas y los péptidos detectados con $FDR < 1\%$ a nivel de PSM, péptido y proteína se muestran en la Figura 4a. Además se ha desarrollado un clasificador para evaluar la detectabilidad de un péptido en base a su secuencia. Para ello se utilizó la información de GPMDB (32) sobre el número de veces que se ha detectado cada péptido y las propiedades físico-químicas de esos mismos péptidos almacenadas en la base de datos AAindex (Figura 4b) (33). Se comprobó que el porcentaje de péptidos detectables es mayor en las proteínas ya identificadas por MS mientras que en las proteínas “missing” es menor (Figura 4c). Este clasificador tiene una aplicación de especial relevancia a la hora de

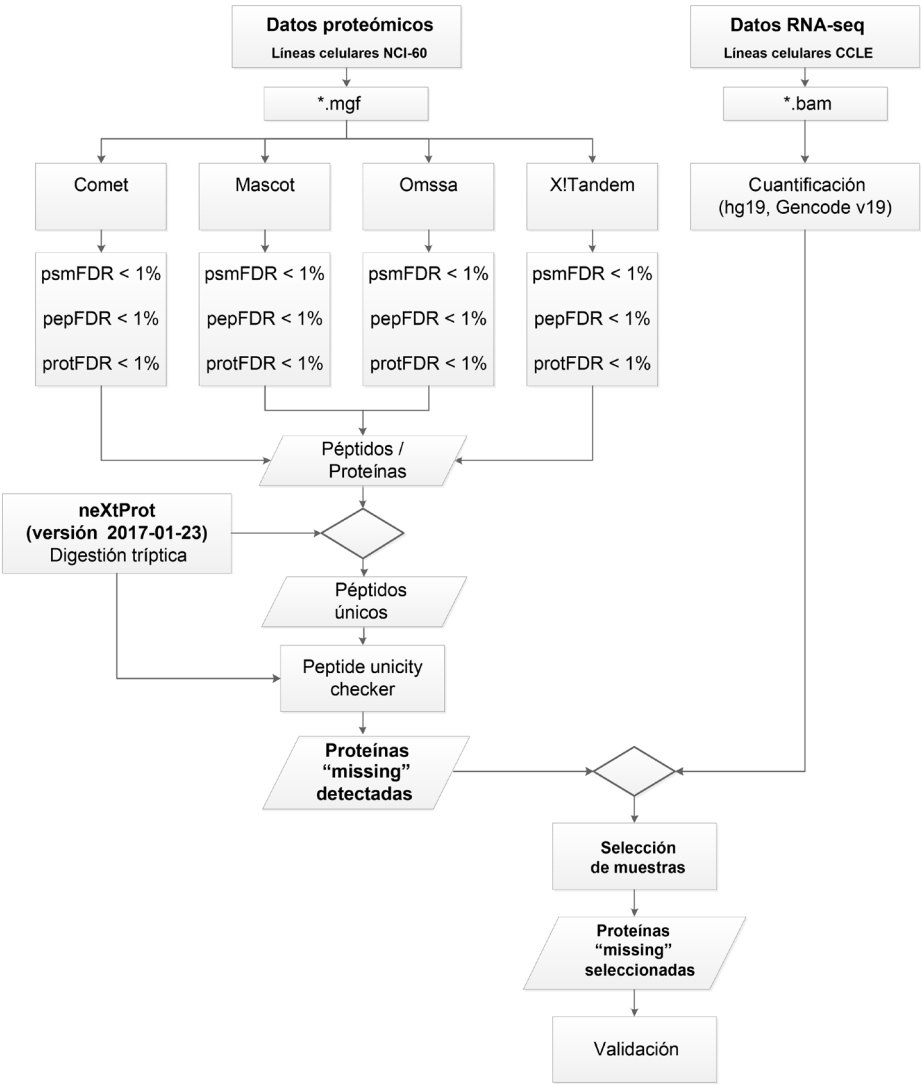


Figura 3. Esquema del análisis integrado de experimentos *shotgun* de proteómica (NCI60) y experimentos transcriptómicos de RNA-Seq (CCLE) basado en los resultados de 4 motores de búsqueda para identificar proteínas “missing”.

validar proteínas “missing” detectadas con un solo péptido, o “one hit wonders”. En este caso hay que diseñar experimentos de proteómica dirigidos (MRM) para detectar más péptidos únicos de dicha proteína y el clasificador puede identificar aquellos péptidos que son detectables por espectrometría facilitando así su validación (Figura 4c)

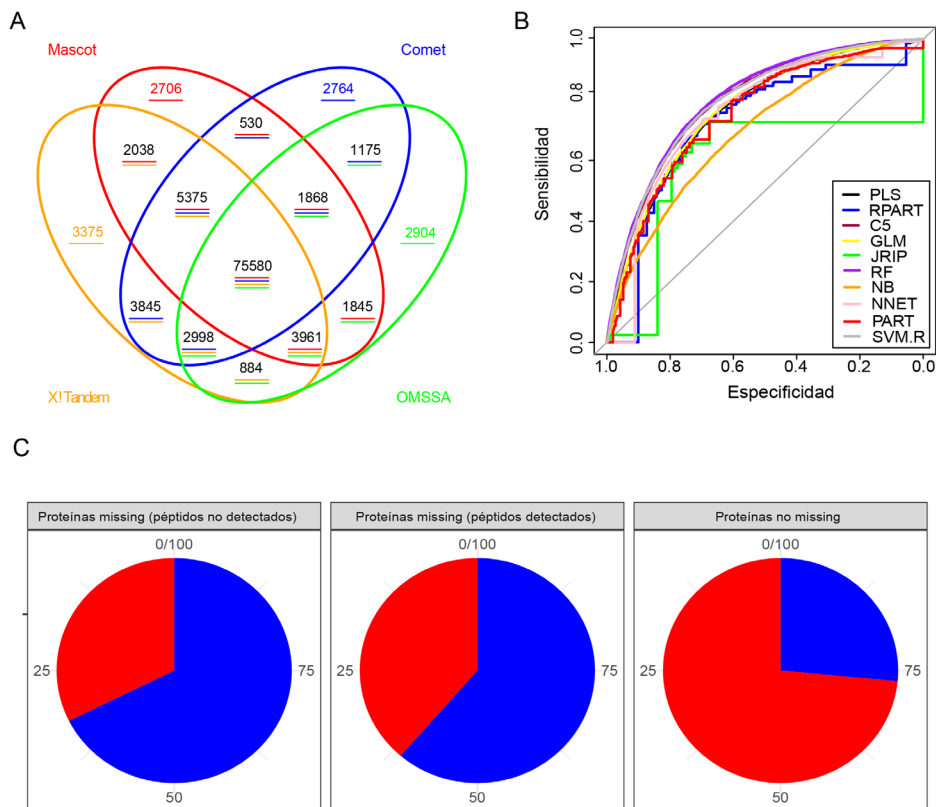


Figura 4. Identificación de proteínas “missing” en líneas celulares tumorales y estudio de la detectabilidad de sus péptidos. (A) Número de péptidos únicos detectados con cada motor de búsqueda utilizado y teniendo en cuenta todos los experimentos analizados. (B) Evaluación mediante curvas ROC de los distintos clasificadores desarrollados para distinguir la detectabilidad de un péptido. Métodos de clasificación utilizados: RPART, C5, JRIP, Random Forest (RF), PART, regresión parcial por mínimos cuadrados (PLS), modelo lineal general (GLM), Naïve Bayes (NB), red neuronal (NNET), y máquinas de soporte vectorial (SVM.R). (C) Porcentaje de péptidos detectables predichos, representado en color rojo, para distintos conjuntos de péptidos: péptidos no detectados de proteínas “missing”, péptidos detectados de proteínas “missing”, y péptidos detectados de proteínas que no son “missing”.

3. Análisis proteogenómico para la detección de péptidos mutados.

El cáncer es una de las principales causas de muerte en todo el mundo, y representa el 15% del número total de muertes anuales. Además, en las próximas décadas, se espera que la tasa de mortalidad debida al cáncer se duplique. Hay más de 200 tipos de cáncer y cada uno se puede clasificar en varios subtipos con diferentes características moleculares y clínicas (34). Esta complejidad explica la heterogeneidad observada en la respuesta a terapia y la tasa de supervivencia de los pacientes (35). Las



mutaciones de la secuencia de ADN impulsan la transformación neoplásica y causan, entre otros efectos, el crecimiento celular descontrolado (36). Por lo tanto, la identificación del catálogo completo de aberraciones de ADN es imprescindible para mejorar no solo la prevención del cáncer y su detección precoz sino también su tratamiento.

El desarrollo de tecnologías de secuenciación masiva (NGS) ha acelerado el descubrimiento de mutaciones genéticas como sustituciones, deleciones, inserciones, o amplificaciones (37,38). La secuenciación del exoma o la del genoma completo han demostrado ser los métodos más valiosos para identificar mutaciones relevantes para el diagnóstico y tratamiento de enfermedades humanas, incluido el cáncer (39). La identificación de mutaciones en la línea germinal (40) y la identificación de mutaciones somáticas en cáncer (41) son el principal objetivo de estos experimentos. Para cumplir dicho objetivo se han llevado a cabo proyectos de caracterización a gran escala como el Proyecto del Genoma del Cáncer (CGP, <http://www.sanger.ac.uk/genetics/CGP>), el Consorcio Internacional del Genoma del Cáncer (ICGC, <https://dcc.icgc.org>) y The Cancer Genome Atlas (TCGA, <http://cancergenome.nih.gov>). Desde su inicio en 2006, el proyecto TCGA ha caracterizado más de 10,000 muestras tumorales de 34 tipos diferentes de cáncer, utilizando secuenciación de alto rendimiento. Los análisis bioinformáticos de esta gran cantidad de datos han identificado más de diez millones de mutaciones (42).

Por otro lado, los principales avances en Proteogenómica se han realizado en la investigación del cáncer y, más específicamente, en la detección de cambios específicos del proteoma en el tumor. Los SAPs pueden provocar el inicio de la enfermedad, su progresión o cambios en la respuesta al tratamiento (43,44). La etapa crítica en el análisis proteogenómico de muestras tumorales es la generación personalizada de las bases de datos de secuencias de péptidos necesarias para realizar las búsquedas de MS. Se han descrito diferentes enfoques basados en el procesamiento de datos genómicos, tanto a partir de experimentos de DNA-Seq como de RNA-Seq (9,45,46). En este contexto, la gran cantidad de experimentos disponibles en proyectos como la TCGA suponen un recurso particularmente valioso para consolidar esta metodología en el entorno de la oncología clínica.

3.1. Generación de bases de datos de péptidos mutados

Para identificar SAPs en muestras tumorales a partir de las mutaciones somáticas descritas en el proyecto TCGA proponemos el flujo de análisis mostrado en la Figura 5. En este análisis se ha tenido en cuenta la información de 10183 exomas pertenecientes a pacientes de 31 tipos de cáncer diferentes, incluyendo algunos de los más prevalentes en la población mundial (47).

La información de mutaciones somáticas se obtuvo en el portal de la TCGA y los ficheros MAF correspondientes se procesaron utilizando código de R propio. De esta forma se identificaron 3,009,480 variantes únicas sin tener en cuenta ningún tipo de filtrado. A partir de los SNVs mencionados se generó una base de datos que incluía la predicción del efecto causado por dichas mutaciones para lo cual se reanotaron con VeIP (48) todos los exomas. Este programa convierte cada SNV a las coordenadas de aminoácidos correspondientes y proporciona las predicciones sobre su posible efecto

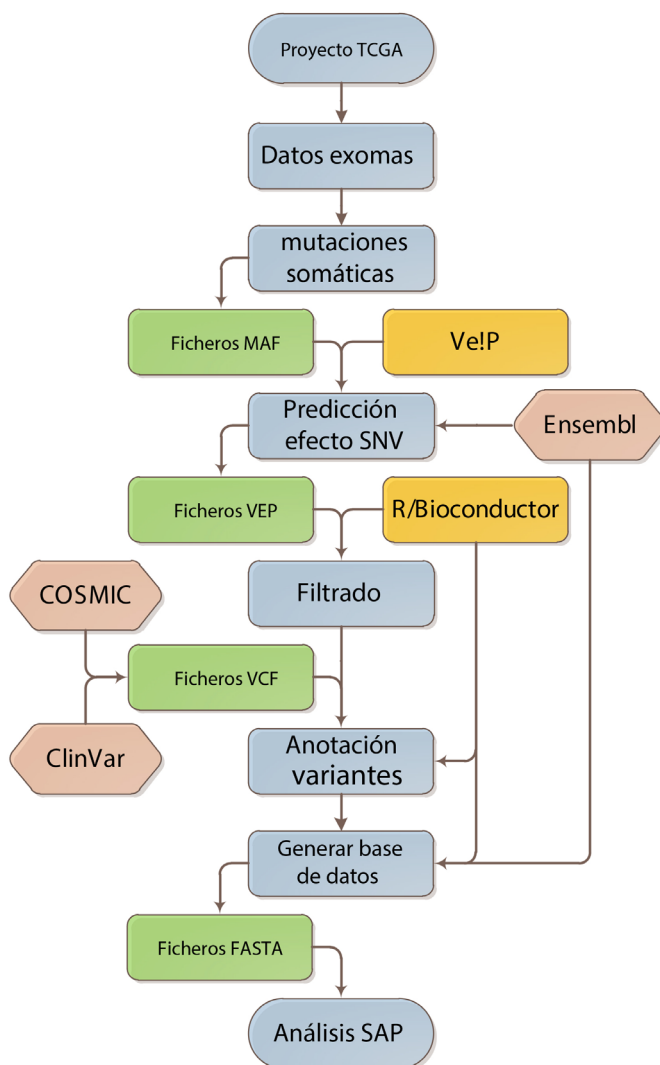


Figura 5. Esquema general de la creación de las bases de datos personalizadas para la identificación de SAPs. Se muestran las distintas fases del análisis en color azul, los formatos de ficheros en verde, las bases de datos en rojo y las herramientas bioinformáticas en naranja.

calculadas por SIFT (49) y PolyPhen-2 (50). Esta información fue utilizada para filtrar los SNV sinónimos utilizando el entorno estadístico R / Bioconductor de manera que sólo conservamos aquellos SNV clasificados como “*missense*” y que por tanto producen cambio en la secuencia de la proteína resultante. En total se identificaron 1,161,751 variantes genómicas únicas y no sinónimas cuya distribución según los distintos tipos de cáncer analizados se muestra en la Figura 6a. La relación entre el número de genes y el número de mutaciones detectadas en cada gen se representa en la figura 6b. Cabe destacar que hay más de 12,000 genes con al menos 50 variantes.

La estrategia utilizada para generar la base de datos de SAPs a partir de las variantes no sinónimas identificadas consistió en: 1) extraer del archivo FASTA de secuencias proteicas de Ensembl (versión 75) la secuencia de 80 aminoácidos adyacente



al aminoácido mutado; 2) comprobar el aminoácido de referencia y la posición de la mutación para filtrar aquellos casos donde se detecte una discrepancia (51). De esta forma sólo se almacenan en la base de datos aquellos péptidos mutados que no se dan en la referencia. En este caso, la base de datos incluía 1,525,055 SAPs únicos pertenecientes a 92,412 proteínas y 19,925 genes. Estas variantes genéticas, como se muestra en la Figura 5, fueron anotadas con información clínica procedente de COSMIC (52) y ClinVar (53) para facilitar la posterior interpretación de los resultados. Además de esta base de datos con las variantes encontradas en los pacientes de la TCGA, también se prepararon 3 bases de datos adicionales específicas para hepatocarcinoma (LIHC), cáncer de colon (COAD) y cáncer de mama (BRCA). En este caso se incluyeron en cada base de datos solamente las variantes observadas en los tipos tumorales correspondientes.

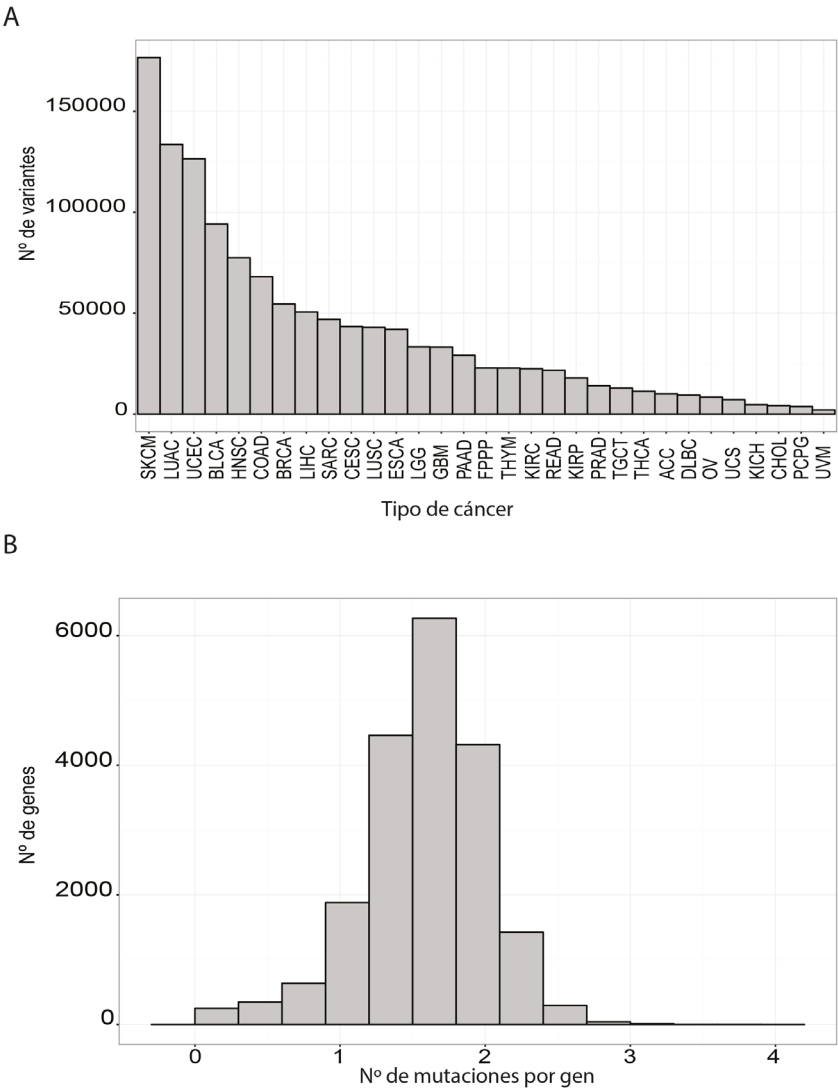


Figura 6. Variantes no sinónimas identificadas en la TCGA. (A) Número de variantes en cada tipo tumoral. (B) Número de genes en función del número de variantes detectadas en cada gen. (Ver Apéndice para las siglas de los distintos tumores).

3.2. Análisis de experimentos de proteómica para la detección de péptidos mutados

Las bases de datos en formato FASTA generadas en el apartado anterior se utilizaron en el análisis de experimentos proteómicos de alto rendimiento para detectar péptidos mutados en líneas celulares (47). En concreto, se emplearon experimentos de proteómica *shotgun* generados por el spHPP de 4 líneas celulares tumorales cuyos datos están almacenados en PRIDE (54), y se demostró que la aproximación proteo-genómica propuesta es eficaz y consistente (Figura 7). Las líneas celulares analizadas fueron Jurkat (línea celular de linfoblastos de células T humanas), CCD18 (línea celular de fibroblastos de colon humano), MCF7 (línea celular de adenocarcinoma de mama humano) y Huh7 (línea celular de hepatocarcinoma humano), y los correspondientes identificadores del repositorio ProteomeXchange (<http://www.proteomexchange.org>) son: PXD000039, PXD000442, PXD000443 y PXD000449.

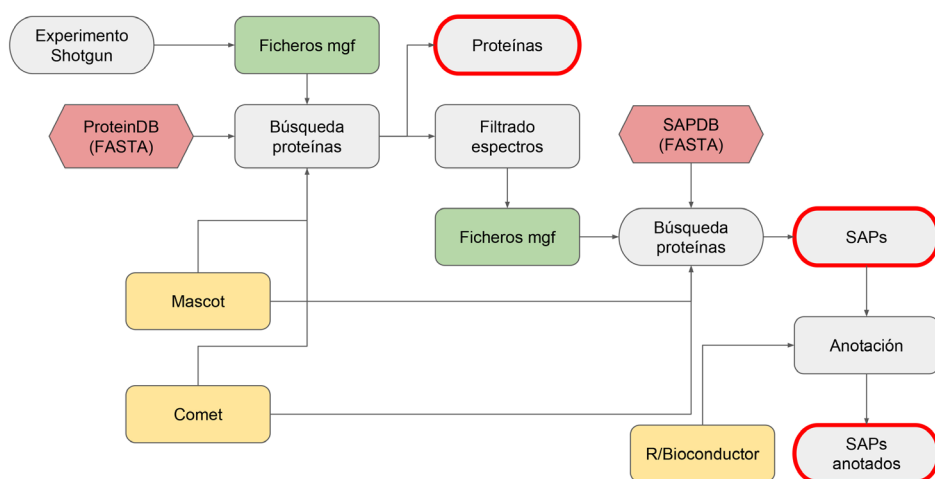


Figura 7. Flujo de análisis bioinformático para la identificación de SAPs en experimentos *shotgun* de proteómica. Se muestran las distintas fases del análisis en color azul, los formatos de ficheros en verde, las bases de datos en rojo y las herramientas bioinformáticas en naranja. Los resultados del análisis se han remarcado en rojo.

Para realizar el estudio se emplearon los ficheros *mgf* y se utilizaron varias combinaciones de motores de búsqueda (*Mascot* o *Comet*) y bases de datos FASTA (general de la TCGA o específica de tumor). Para cada muestra se realizaron dos búsquedas consecutivas empleando la estrategia “*target-decoy*” utilizando primero la base de datos de referencia de proteínas (*ProteinDB*), que en este caso es la base de datos de proteínas *Ensembl* (versión 75), y en segundo lugar la base de datos de péptidos mutados (*SAPDB*) previamente generada como se muestra en la Figura 7. En la primera búsqueda identificamos las proteínas presentes en las muestras aplicando un criterio de FDR a nivel de PSM y proteína ($FDR < 1\%$). Después de eliminar los espectros asignados de los archivos *mgf* realizamos una nueva búsqueda para detectar péptidos mutados usando la base de datos de SAPs, tanto la base de datos general de la TCGA como la específica para cada tipo tumoral. El experimento de la línea celular MCF7 se analizó con la base de datos de SAPs de BRCA, la línea celular CCD18 con la información de COAD y la línea Huh7 con LIHC. En estas búsquedas, los análisis estadísticos de los resultados se llevaron a cabo solo a nivel de PSM ($FDR < 1\%$).



Se realizó una búsqueda con las bases de datos de péptidos ("target") y otra para las proteínas o péptidos "decoy" generados mediante el método de secuencia pseudo-reversa. Los parámetros de búsqueda se establecieron de la siguiente manera: carbamidometilación de cisteína como modificación fija y oxidación de metionina como modificación variable; tolerancia de error para la masa del precursor y del fragmento de 10 ppm y 0,05 Da respectivamente; y se permitieron hasta 2 roturas omitidas o "missed cleavages". Para el cálculo de FDR se utilizó el método Mayu (55). La inferencia de las proteínas a partir de los resultados de Mascot se realizó mediante el algoritmo PAnalyzer (56), y los grupos de proteínas no concluyentes se descartaron. En el caso de Comet, se consideraron todas las proteínas proporcionadas por el motor de búsqueda. El número de SAPs detectados se muestra en la Tabla 3.

Los PSM obtenidos y el número de identificaciones de péptidos, proteínas y genes mutados fueron muy escasos. Estos resultados fueron consistentes con la baja abundancia esperada de péptidos mutados en las muestras y la naturaleza aleatoria de los experimentos de MS. Para cada selección de motor de búsqueda (Mascot o Comet) y base de datos SAP (TCGA o tumor específico), comparamos los resultados (péptidos

Motor de búsqueda	Base de datos SAP	Muestra	Espectros asignados	PSMs	SAPs	Proteínas mutadas	Genes mutados
Comet	BRCA	MCF7_1	83	84	66	185	65
		MCF7_2	136	137	122	352	115
	COAD	CCD18_1	44	52	14	63	14
		CCD18_2	467	475	223	571	179
	LIHC	HUH7_1	1	1	1	4	1
		HUH7_2	446	446	260	704	235
	TCGA	CCD18_1	909	941	241	578	200
		CCD18_2	2649	2781	1051	1474	481
		HUH7_1	37	38	18	44	18
		HUH7_2	553	567	318	770	242
		JURKAT_1	407	426	234	552	193
		JURKAT_2	121	123	82	216	78
		MCF7_1	426	435	183	454	151
		MCF7_2	5	5	5	14	5
Mascot	BRCA	MCF7_1	81	115	97	218	68
		MCF7_2	18	22	8	24	5
	COAD	CCD18_1	35	65	19	64	10
		CCD18_2	52	82	33	78	21
	LIHC	HUH7_1	4	5	5	9	4
		HUH7_2	6	6	5	9	5
	TCGA	CCD18_1	596	907	316	581	179
		CCD18_2	893	1238	466	841	256
		HUH7_1	90	134	50	107	35
		HUH7_2	117	153	105	225	70
		JURKAT_1	656	1111	690	1150	332
		JURKAT_2	398	554	364	795	239
		MCF7_1	565	1972	1627	1093	343
		MCF7_2	16	27	22	63	12

Tabla 3. Número de péptidos, proteínas y genes mutados identificados con un FDR a nivel de PSM menor del 1% empleando la base de datos personalizada de SAPs, tanto general (TCGA) como específica (BRCA, COAD o LIHC), y los motores de búsqueda Mascot y Comet.

o genes mutados) para cada línea celular (Tabla 3). Es notoria la disminución del número de genes mutados en relación con el número correspondiente de SAPs, lo cual probablemente se deba al número de variantes por gen. Además, es importante destacar la mayor cantidad de identificaciones obtenidas utilizando el catálogo completo de SAPs que cuando se utilizan bases de datos específicas de cada tipo tumoral. Este resultado sugiere que la calidad de las identificaciones de péptidos obtenidas empleando la base de datos general de la TCGA es mejor. Para confirmar esta hipótesis representamos las distribuciones de las puntuaciones de asignación para cada motor de búsqueda, "*ion score*" para Mascot y *XCorr* para Comet. Efectivamente, en la Figura 8a puede observarse que las asignaciones correspondientes a SAPs obtienen puntuaciones más elevadas al utilizar en la búsqueda la base de datos general.

En resumen, identificamos 4,826 péptidos mutados en las cuatro líneas celulares, 2,916 con Mascot y 2,419 con Comet. Esto puede traducirse en la detección de

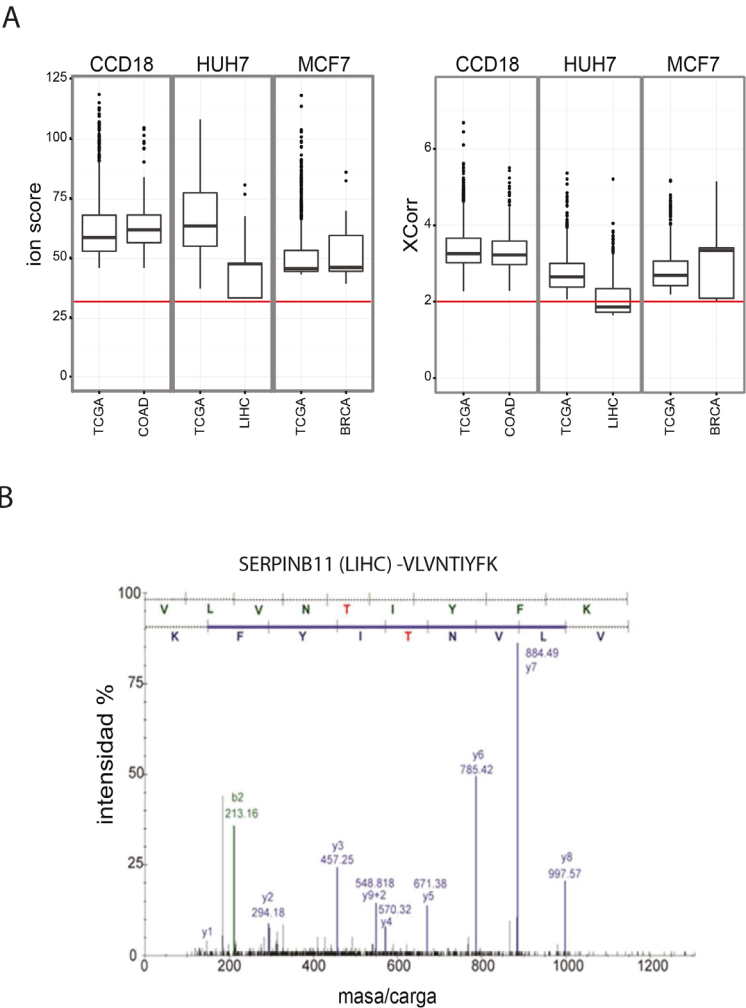


Figura 8. SAPs identificados en líneas celulares tumorales. (A) Distribución de las puntuaciones obtenidas para las asignaciones de espectros de Mascot ("*ion score*") y Comet (*XCorr*) utilizando la base de datos general (TCGA) o las específicas de cada tipo tumoral (COAD, LIHC o BRCA). (B) Espectro e iones de las series b e y asignados al péptido mutado VLVNTIYFK del gen SERPINB11 identificado en hepatocarcinoma (LIHC).



1,580 genes mutados, 877 genes detectados con Mascot y 1,161 genes detectados con Comet. Pudimos anotar 2,422 variantes genéticas identificadas utilizando COSMIC (58.04% de todas las detecciones), mientras que solo 30 variantes genéticas detectadas (0.69%) se habían descrito previamente en ClinVar. La lista de genes mutados se analizó con Ingenuity (<https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>) para comprobar su implicación en el desarrollo y progresión del cáncer. El análisis de enriquecimiento de categorías relacionadas con enfermedades no sólo relacionó los genes con cáncer ($p < 1e-30$), sino que además se obtuvieron enriquecimientos en genes relacionados con hepatocarcinoma, cáncer de colon, cáncer de mama y neoplasia hematológica coincidiendo con el origen de las líneas celulares estudiadas.

Por último se evaluó la detección de los péptidos mutados mediante la inspección manual de los espectros correspondientes. Se seleccionaron los mejores candidatos y dos expertos en MS clasificaron las asignaciones en calidad alta, media o baja considerando 3 características: 1) la calidad de las series de iones y o b, 2) la intensidad de los picos y 3) la presencia de picos no asignados (58). Una proporción considerable, el 65.71%, fueron clasificados como buenos candidatos para una posible validación utilizando otro tipo de aproximación como los ensayos de monitorización de reacción múltiple (SRM/MRM). En la Figura 8b se muestra uno de los péptidos mutados perteneciente a la proteína SERPINB11 y detectado en LIHC. A partir de la señal original del espectro se han reasignado sus iones evaluando así la identificación obtenida. La asignación de ambas series de iones ha permitido obtener la secuencia del péptido detectada por el motor de búsqueda y hay que señalar que la secuencia obtenida incluye el aminoácido mutado validando el resultado.

4. Proteogenómica para la identificación de péptidos noveles.

La Proteogenómica permite aplicar estrategias para caracterizar y comprender la base molecular de una enfermedad, así como para identificar nuevos biomarcadores clínicos de diagnóstico, pronóstico y selección del tratamiento más adecuado. Dichas estrategias tratan de detectar péptidos que no están presentes en las bases de datos de referencia, por ejemplo, SAPs y péptidos correspondientes a genes definidos como no codificantes o péptidos noveles. Como ya hemos visto, estos estudios requieren la generación de bases de datos personalizadas y aproximaciones bioinformáticas específicas.

Al igual que en el apartado anterior centramos el estudio de SAPs en los distintos tipos de cáncer, para el estudio de péptidos noveles vamos a analizar el transcriptoma humano tumoral. Dicho transcriptoma ha sido objeto de un estudio detallado y se ha generado un catálogo de todos los transcritos expresados, incluidos los ARN largos no codificantes (lncRNA). En concreto, dicho transcriptoma, denominado MiTranscriptome (59), se obtuvo a partir del análisis de 7,256 experimentos de RNA-Seq, que incluyen 5,847 muestras tumorales de la TCGA. El catálogo completo contiene

384,066 transcritos diferentes, algunos de ellos no descritos en bases de datos como Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>), Ensembl (<https://www.ensembl.org>) o Gencode (<https://www.gencodegenes.org/>).

En este apartado, proponemos un análisis bioinformático basado en MiTranscriptome y la información almacenada en Gencode v19 para detectar péptidos novedales asociados con cáncer. En primer lugar se generan las bases de datos personalizadas: estructuras anotadas en Gencode con traducción conocida (secuencia de aminoácidos de la proteína resultante), estructuras anotadas pero con secuencia de proteína desconocida y estructuras no anotadas. En estos dos últimos casos la secuencia de aminoácidos correspondiente se ha inferido a partir de la secuencia de ARN observada. Estas bases de datos se utilizan para el estudio del conjunto de experimentos públicos de *shotgun* de NCI60 previamente mencionado (30). De esta forma identificamos péptidos novedales en las distintas líneas celulares tumorales y demostramos la viabilidad de nuestro método bioinformático (60).

4.1. Generación de bases de datos de péptidos novedales

La iniciativa MiTranscriptome trata de describir por primera vez toda la complejidad del transcriptoma humano, especialmente en el caso de muestras tumorales (59). En este estudio se procesaron un total de 25 conjuntos de datos de RNA-Seq correspondientes a 7,256 muestras, procedentes de los proyectos TCGA y ENCODE entre otros, para definir un conjunto final de 6,503 experimentos de alta calidad. El análisis bioinformático basado en Tophat2 (61) y Cufflinks2 (62,63) generó un ensamblado al inicio del transcriptoma usando hg19 como genoma de referencia. Como resultado, el ensamblado humano MiTranscriptome contiene 384,066 transcritos predichos disponibles en un archivo en formato GTF en su página web (<http://mitranscriptome.org/>).

Una vez obtenidas las estructuras de este ensamblado de muestras tumorales, anotamos primero los transcritos con el fin de identificar todos los genes conocidos, tanto codificantes como no codificantes (Figura 9). Para ello, utilizamos la anotación del genoma humano de Gencode versión 19, que incluye 57,820 genes (196,520 transcritos), 20,345 de los cuales codifican proteínas (81,814 transcritos). La anotación se realizó con el programa Cuffcompare del paquete de análisis Cufflinks y se identificaron un total de 311,361 transcritos. Sin embargo, solo 165,020 de estas estructuras tenían una secuencia de aminoácidos conocida en la base de datos Gencode. De esta forma definimos 3 bases de datos FASTA diferentes: (1) ProteinDB, incluye 26,264 secuencias de aminoácidos obtenidas de Gencode que representan 165,020 transcritos de MiTranscriptome; (2) NonCodingDB, con información sobre 146,341 transcritos de MiTranscriptome anotados utilizando Gencode pero cuyas secuencias se obtuvieron traduciendo los 6 posibles marcos de lectura a partir de la secuencia del transcrito (35,379,040 secuencias de aminoácidos); (3) NovelTranscriptDB, contiene transcritos detectados en el ensamblado que no están presentes en la base de datos de Gencode. En este caso, las secuencias de aminoácidos de los 72,597 transcritos también se obtuvieron mediante la traducción de los 6 posibles marcos de lectura de dichas estructuras (7,934,008 secuencias).

Uno de los principales problemas estadísticos a tratar en los estudios proteogenó-



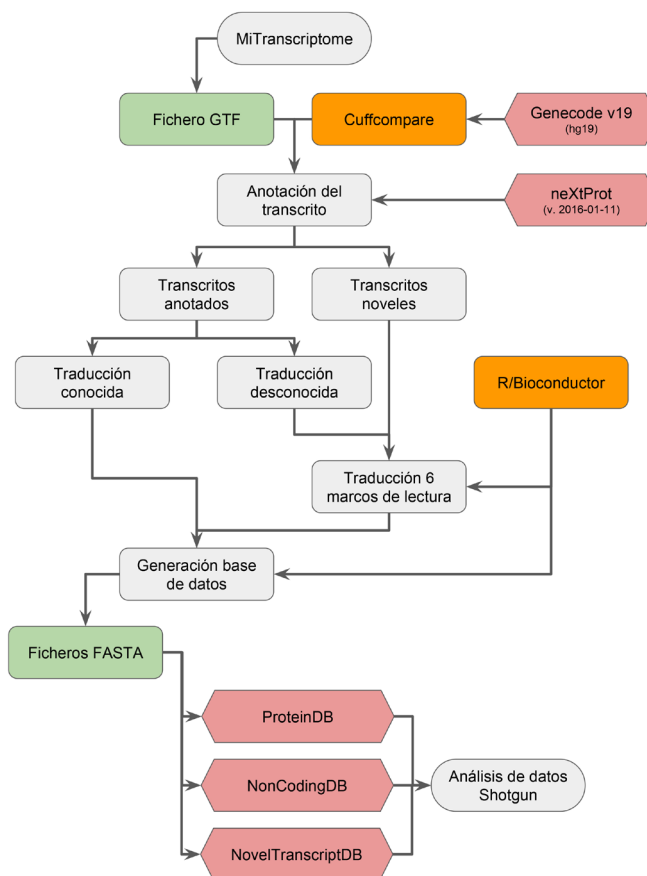


Figura 9. Esquema general de la creación de las bases de datos personalizadas para la identificación de péptidos novedales. Se muestran las distintas fases del análisis en color gris, los formatos de ficheros en verde, las bases de datos en rojo y las herramientas bioinformáticas en naranja.

micos es el inevitable aumento del tamaño de las bases de datos personalizadas al emplear la traducción de todos los posibles marcos de lectura para inferir las secuencias de aminoácidos de los péptidos derivados de los experimentos genómicos o transcriptómicos. La principal consecuencia de utilizar estas bases de datos en las búsquedas de proteómica es la dificultad para controlar el FDR y, por lo tanto, la alta probabilidad de falsas asignaciones de péptidos. Este tipo de error es particularmente importante cuando el objetivo del estudio proteogenómico es la detección de nuevos biomarcadores clínicos para su uso en medicina personalizada. Para reducir el tamaño de las bases de datos personalizadas aplicamos un clasificador similar al que utilizamos para la detección de proteínas “missing” previamente mencionado (17), que se basa en la información sobre el número de observaciones para cada péptido almacenada en la base de datos GPMDB (32) y la evaluación de más de 550 propiedades fisicoquímicas y bioquímicas de dichos péptidos utilizando el paquete seqinr de R (64). Primero, filtramos las características que distinguen de forma significativa los péptidos con mayor número de observaciones de los péptidos con pocas observaciones. Para ello utilizamos técnicas de muestreo y un test estadístico con el que seleccionamos 302 propiedades de las cuales 106 son consideradas propiedades no redundantes. El clasificador Random Forest se entrenó con dichas características y se evaluó (Figura 4b) mediante el paquete caret de R (65). Este clasificador se ha empleado para

descartar los péptidos predichos de las bases de datos personalizadas que son considerados no detectables.

4.2. Análisis de experimentos de proteómica para la detección de péptidos nuevos

Una vez creadas las bases de datos, analizamos los experimentos de *shotgun* disponibles en el proyecto NCI60 (30) correspondientes a 43 líneas celulares para las cuales también hay experimentos de RNA-Seq (29). Se descargaron los archivos de los datos crudos y utilizando la herramienta MSConvertGUI se convirtieron en archivos *mgf* para su posterior análisis. El análisis bioinformático consiste en tres búsquedas sucesivas utilizando el motor de búsqueda Mascot y eliminando los espectros asignados del fichero *mgf* antes de la siguiente búsqueda (Figura 10). Este flujo de análisis se puede generalizar fácilmente para incluir motores de búsqueda adicionales.

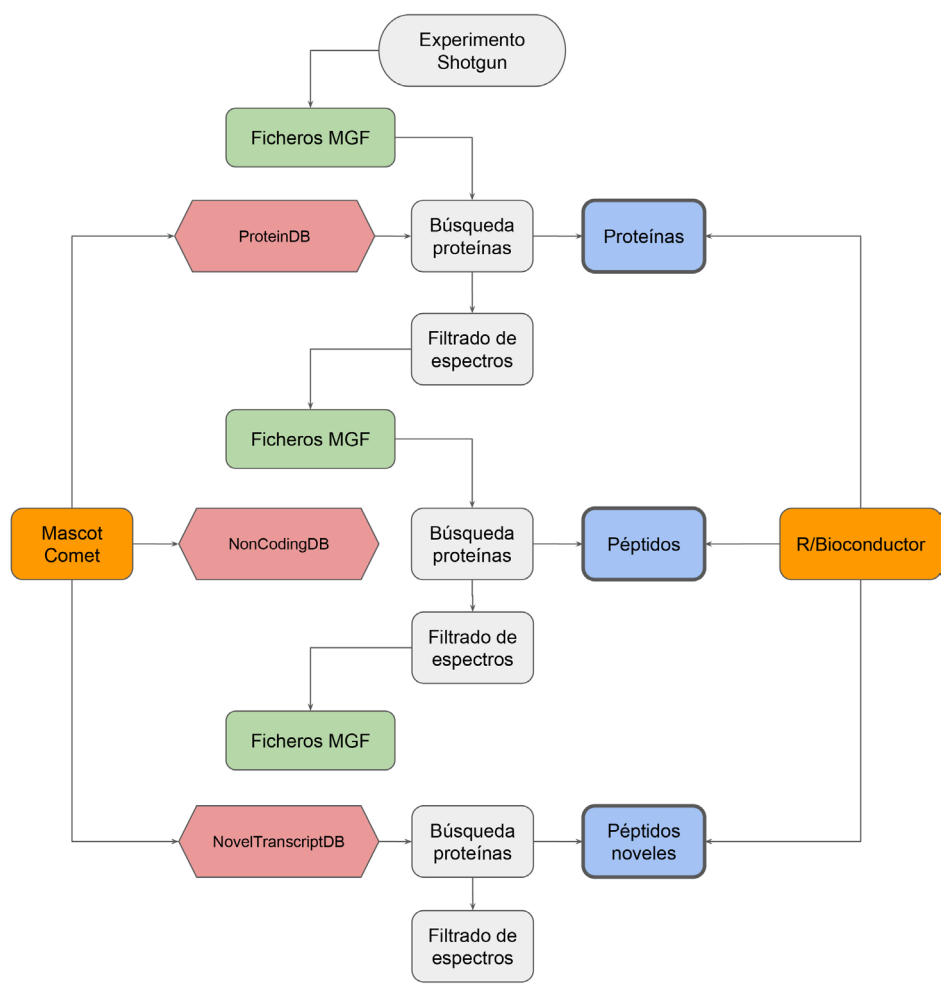


Figura 10. Flujo de análisis bioinformático para la identificación de péptidos nuevos en experimentos *shotgun* de proteómica. Se muestran las distintas fases del análisis en color gris, los formatos de ficheros en verde, las bases de datos en rojo y las herramientas bioinformáticas en naranja. Los resultados del análisis aparecen en color azul.

Para las búsquedas con Mascot se generó una base de datos “*decoy*” utilizando el método de secuencia pseudoreversa, y se realizaron búsquedas separadas para las bases de datos “*target*” y “*decoy*”. Los parámetros de búsqueda se establecieron de la siguiente manera: carbamidometilación de cisteína como una modificación fija y oxidación de metionina como modificación variable. La tolerancia de masa del precursor y del fragmento se estableció en 10 ppm y 0,05 Da respectivamente, y se permitió una rotura omitida o “*missed cleavage*”. Se calculó el FDR mediante Mayu (55) y se aplicó un FDR <1% tanto a nivel de PSM como de proteína. El proceso de identificación de proteínas se simplificó utilizando péptidos únicos para garantizar identificaciones fiables. Se definieron como péptidos únicos los péptidos de entre 9 a 30 aminoácidos generados tras la digestión *in silico* de la base de datos neXtProt con Proteogest (66), que no son compartidos por las proteínas de dicha base de datos.

Los resultados obtenidos demuestran la utilidad del flujo de análisis desarrollado, ya que se consiguen asignaciones con todas las bases de datos, incluida la base de datos de péptidos noveles. Como cabía esperar hay un mayor número de identificaciones de péptidos pertenecientes a proteínas conocidas pero las asignaciones a péptidos noveles también pasan los criterios estadísticos de asignación (Figura 11a). La distribución de “*ion score*” para las identificaciones de cada base de datos se muestra en la Figura 11b.

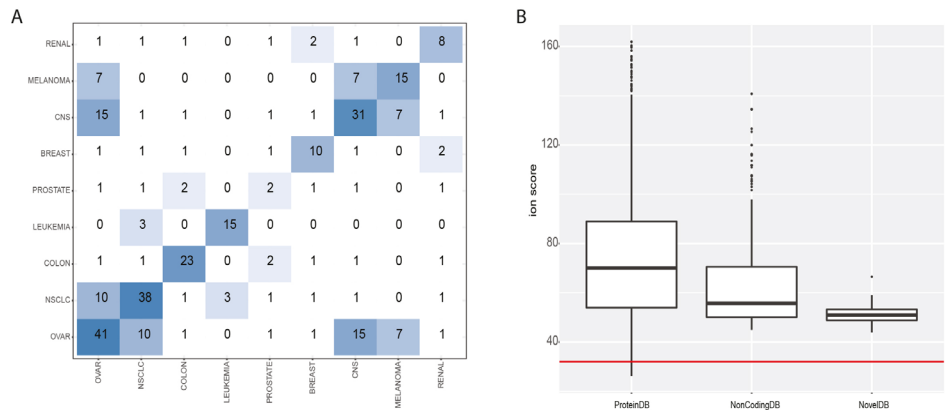
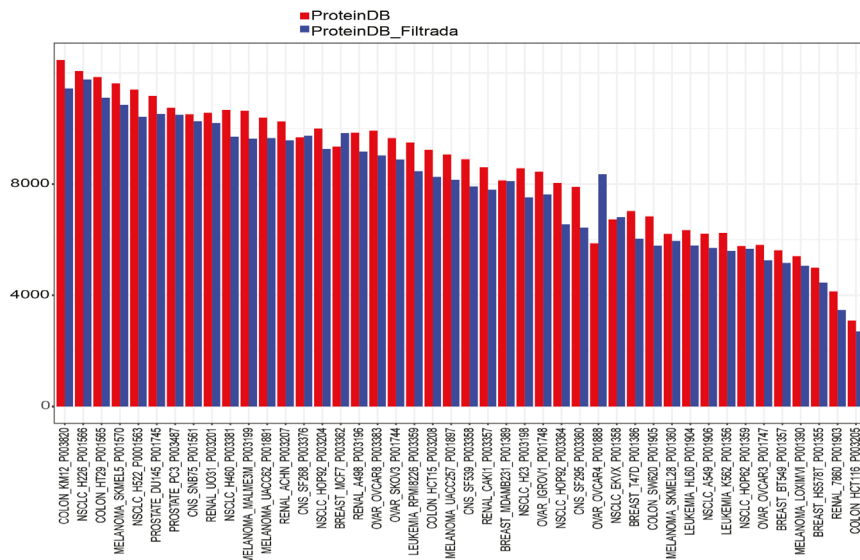


Figura 11. Péptidos noveles detectados en líneas celulares tumorales. (A) Número de péptidos noveles detectados en cada tipo tumoral teniendo en cuenta los resultados de todas las líneas celulares analizadas. (B) Distribución de las puntuaciones de Mascot para las asignaciones obtenidas con las distintas bases de datos personalizadas.

El resultado obtenido con la base de datos ProteinDB filtrada y sin filtrar con el clasificador de péptidos detectables permite evaluar si introducimos un sesgo en el análisis al modificar la base de datos. Como podemos ver en la Figura 12 los péptidos detectados son prácticamente los mismos por lo que el uso del filtro propuesto reduciría el tamaño de la base de datos sin modificar el resultado. De esta forma, facilitamos en gran medida el uso de bases de datos personalizadas al controlar el efecto del gran tamaño de este tipo de bases de datos.



B

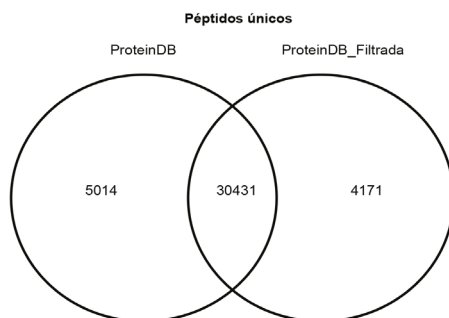


Figura 12. Efecto de reducir el tamaño de las bases de datos personalizadas eliminando los péptidos que predecimos que no son detectables. (A) Número de péptidos identificados en cada experimento utilizando la base de datos filtrada (ProteinDB_Filtrada) y sin filtrar (ProteinDB). (B) Número de péptidos únicos identificados teniendo en cuenta todos los experimentos analizados utilizando la base de datos filtrada (ProteinDB_Filtrada) y sin filtrar (ProteinDB).

5. Conclusiones.

La comprensión de la biología humana se basa en la definición precisa de todos los componentes moleculares y sus interacciones para configurar mapas a nivel celular, de tejido y órgano. La anotación detallada de los genes y proteínas surge entonces como un tema central que debe abordarse mediante enfoques innovadores, como lo es la Proteogenómica, que integra datos de experimentos genómicos, transcrip-tómicos y proteómicos a gran escala. En este capítulo se describen diferentes flujos bioinformáticos de análisis que utilizando experimentos públicos de grandes iniciativas internacionales de caracterización molecular (HBM, TCGA, NCI60, CCLE...) propor-cionan información relevante en distintas áreas de conocimiento.



Las aproximaciones bioinformáticas descritas han contribuido enormemente a (1) los avances en la descripción del proteoma humano y en concreto la identificación de las proteínas “missing”, (2) el desarrollo de nuevos métodos de análisis de datos para mejorar los resultados obtenidos en un experimento *shotgun* de proteómica y (3) la caracterización del transcriptoma humano de muestras tumorales definiendo SAPs y péptidos noveles relacionados con el cáncer. La innovación en métodos proteogenómicos tiene el objetivo de, integrando información de diferentes ómicas, contribuir a la comprensión de la complejidad de la maquinaria celular y molecular.

Agradecimientos

La Plataforma de Bioinformática del CIMA es miembro de ProteoRed-ISCI. Este trabajo ha sido financiado por PRBB-ISCI (PT13/0001/0002), PRB3-ISCI (PT17/0019/0013), el Ministerio de Ciencia e Innovación (DPI2015-68982-R) y el Dpto. de Salud del Gobierno de Navarra (33/2015). Agradecemos también a todas las iniciativas públicas que dan acceso a conjuntos de datos transcriptómicos y proteómicos con la correspondiente anotación clínica, y en particular a IGC y expO por sus esfuerzos para obtener muestras de tejido en condiciones estándar y realizar análisis de expresión génica en un conjunto clínicamente anotado de muestras tumorales.

Apéndice

Acrónimos utilizados para cada tipo tumoral analizado de la TCGA:

ACC Carcinoma Adrenocortical
BLCA Carcinoma Urotelial
BRCA Carcinoma de Mama Invasivo
CESC Carcinoma Escamoso de Cuello Uterino y Adenocarcinoma Endocervical
CHOL Colangiocarcinoma
COAD Adenocarcinoma de Colon
DLBC Linfoma Difuso de Células B Grandes
ESCA Cáncer de Esófago
GBM Glioblastoma Multiforme
HNSC Carcinoma Escamoso de Cabeza y Cuello
KIRC Carcinoma Renal de Células Cromóforas
KIRC Carcinoma Renal de Células Claras
KIRP Carcinoma Renal de Células Papilares
LAML Leucemia Mieloide Aguda
LGG Glioma de Bajo Grado
LIHC Hepatocarcinoma
LUAD Adenocarcinoma Pulmonar
LUSC Carcinoma Pulmonar de Células Escamosas
MESO Mesotelioma
OV Cistoadenocarcinoma Seroso de Ovario
PAAD Adenocarcinoma Pancreático
PCPG Feocromocitoma y Paraganglioma
PRAD Adenocarcinoma de Próstata
READ Adenocarcinoma de Recto
SARC Sarcoma

SKCM Melanoma Cutáneo
 STAD Adenocarcinoma de Estómago
 TGCT Tumor de Células Germinales de Testículo
 THCA Carcinoma Tiroideo
 THYM Timoma
 UCEC Carcinoma Endometrial del Cuerpo Uterino
 UCS Carcinosarcoma uterino
 UVM Melanoma Uveal

REFERENCIAS

1. Blanchette PS, Spreafico A, Miller FA et al. Genomic testing in cancer: Patient knowledge, attitudes, and expectations. *Cancer*. 2014;120(19):3066–73.
2. Wang Y, Wang Y, Klijn JGM et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Erasmus*. 2005;671–9.
3. Mok TS, Wu Y, Thongprasert S et al. Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma. *N Engl J Med*. 2009;361(10):947–57.
4. Nilsson T, Mann M, Aebersold R, Yates JR, Bairoch A, Bergeron JJM. Mass spectrometry in high-throughput proteomics: Ready for the big time. *Nat Methods*. 2010;7(9):681–5.
5. Nagaraj N, Wisniewski JR, Geiger T et al. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol*. 2011;7(548):1–8.
6. Yates JR 3rd, Eng JK, McCormack AL. Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal Chem*. 1995 Sep;67(18):3202–10.
7. Ansong C, Purvine SO, Adkins JN, Lipton MS, Smith RD. Proteogenomics: Needs and roles to be filled by proteomics in genome annotation. *Briefings Funct Genomics Proteomics*. 2008;7(1):50–62.
8. Evans VC, Barker G, Heesom KJ, Fan J, Bessant C, Matthews DA. De novo derivation of proteomes from transcriptomes for transcript and protein identification. *Nat Methods*. 2012;9(12):1207–11.
9. Woo S, Cha SW, Merrihew G et al. Proteogenomic database construction driven from large scale RNA-seq data. *J Proteome Res*. 2014 Jan 3;13(1):21–8.
10. Wu L, Candille SI, Choi Y et al. Variation and Genetic Control of Protein Abundance in Humans. *Nature*. 2013 Jul 4;499(7456):79–82.
11. Castellana N, Bafna V. Proteogenomics to discover the full coding content of genomes: a computational perspective. *J Proteome Res*. 2010 Oct 10;73(11):2124–35.
12. Paik YK, Jeong SK, Omenn GS et al. The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat Biotechnol*. 2012;30(3):221–3.
13. Paik YK, Hancock WS. Uniting ENCODE with genome-wide proteomics. *Nat Biotechnol*. 2012;30(11):1065–7.
14. Segura V, Medina-Aunon JA, Guruceaga E et al. Spanish human proteome project: Dissection of chromosome 16. *J Proteome Res*. 2013;12(1):112–22.
15. Segura V, Medina-Aunon JA, Mora MI et al. Surfing transcriptomic landscapes. A step beyond the annotation of chromosome 16 proteome. *J Proteome Res*. 2014;13(1):158–72.
16. Guruceaga E, Sanchez Del Pino MM, Corrales FJ, Segura V. Prediction of a missing protein expression map in the context of the Human Proteome Project. *J Proteome Res*. 2015;14(3):1350–60.
17. Guruceaga E, Garin-Muga A, Prieto G et al. Enhanced Missing Proteins Detection in NCI60 Cell Lines Using an Integrative Search Engine Approach. *J Proteome Res*. 2017 Dec 1;16(12):4374–90.
18. de Souza GA, Arntzen MØ, 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(1):M110.002527.
19. Halvey PJ, Wang X, Wang J et al. Proteogenomic analysis reveals unanticipated adaptations of colorectal tumor cells to deficiencies in DNA mismatch repair. *Cancer Res*. 2014;74(1):387–97.
20. Su Z-D, Sun L, Yu D-X et al. Quantitative detection of single amino acid polymorphisms by targeted proteomics. *J Mol Cell Biol*. 2011;3(5):309–15.
21. Pusch W, Flocco T, Leung S-M. Mass spectrometry-based clinical proteomics. *Pharmacogenomics*. 2003;4(4):463–76.
22. Nagaraj SH, Waddell N, Madugundu AK et al. PGTools: A Software Suite for Proteogenomic Data Analysis and Visualization. *J Proteome Res*. 2015 May 1;14(5):2255–66.
23. Tabas-Madrid D, Alves-Cruzeiro J, Segura V et al. Proteogenomics Dashboard for the Human Proteome Project. *J Proteome Res*. 2015 Sep 4;14(9):3738–49.
24. Kim M-S, Pinto SM, Getnet D et al. A draft map of the human proteome. *Nature*. 2014 May 29;509(7502):575–



25. Wilhelm M, Schlegl J, Hahne H et al. Mass-spectrometry-based draft of the human proteome. *Nature*. 2014 May 28;509:582.
26. Gaudet P, Michel P-A, Zahn-Zabal M et al. The neXtProt knowledgebase on human proteins: 2017 update. *Nucleic Acids Res*. 2017 Jan 4;45(Database issue):D177–82.
27. Deutsch EW, Overall CM, Van Eyk JE et al. Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. *J Proteome Res*. 2016 Nov 4;15(11):3961–70.
28. Garin-Muga A, Odriozola L, Martínez-Val A et al. Detection of Missing Proteins Using the PRIDE Database as a Source of Mass Spectrometry Evidence. *J Proteome Res*. 2016 Nov 4;15(11):4101–15.
29. Barretina J, Caponigro G, Stransky N et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012 Mar 28;483:603.
30. Park ES, Rabinovsky R, Carey M et al. Integrative analysis of proteomic signatures, mutations and drug responsiveness in the NCI 60 cancer cell line set. *Mol Cancer Ther*. 2010 Feb 2;9(2):257–67.
31. Horvatovich P, K Lundberg E, Chen Y-J et al. Quest for Missing Proteins: Update 2015 on Chromosome-Centric Human Proteome Project. *J Proteome Res*. 2015 Jun 15;14.
32. Craig R, Cortens JP, Beavis RC. Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *J Proteome Res*. 2004 Dec 1;3(6):1234–42.
33. Kawashima S, Pokarowski P, Pokarowska M, Kolinski A, Katayama T, Kanehisa M. AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res*. 2008 Jan 12;36(Database issue):D202–5.
34. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol*. 2015 Jan 20;19(1A):A68–77.
35. McDermott U, Downing JR, Stratton MR. Genomics and the Continuum of Cancer Care. *N Engl J Med*. 2011 Jan 26;364(4):340–50.
36. Hanahan D, Weinberg RA. The Hallmarks of Cancer. *Cell*. 2000;100(1):57–70.
37. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*. 2010 Sep 17;11:685.
38. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2013 Jan 9;31(1):10.1038/nbt.2450.
39. Pabinger S, Dander A, Fischer M et al. A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform*. 2014 Mar 21;15(2):256–78.
40. Kurian AW, Hare EE, Mills MA et al. Clinical Evaluation of a Multiple-Gene Sequencing Panel for Hereditary Cancer Risk Assessment. *J Clin Oncol*. 2014 Jul 1;32(19):2001–9.
41. Tamborero D, Gonzalez-Perez A, Perez-Llamas C et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci Rep*. 2013 Oct 2;3:2650.
42. Chin L, Hahn WC, Getz G, Meyerson M. Making sense of cancer genomic data. *Genes Dev*. 2011 Mar 15;25(6):534–55.
43. Alfaro JA, Sinha A, Kislinger T, Boutros PC. Onco-proteogenomics: cancer proteomics joins forces with genomics. *Nat Methods*. 2014 Oct 30;11:1107.
44. Zhang B, Wang J, Wang X et al. Proteogenomic characterization of human colon and rectal cancer. *Nature*. 2014 Sep 18;513(7518):382–7.
45. Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. *Nat Methods*. 2014 Oct 30;11:1114.
46. Wang X, Zhang B. customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. *Bioinformatics*. 2013 Dec 15;29(24):3235–7.
47. Garin-Muga A, Corrales FJ, Segura V. Proteogenomic Analysis of Single Amino Acid Polymorphisms in Cancer Research BT. *Proteogenomics*. In: Végvári Á, editor. Cham: Springer International Publishing; 2016. p. 93–113.
48. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*. 2010 Aug 15;26(16):2069–70.
49. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009 Jun 25;4:1073.
50. Adzhubei IA, Schmidt S, Peshkin L et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010 Apr;7(4):248–9.
51. Yang X, Lazar IM. XMAAn: A Homo sapiens Mutated-Peptide Database for the MS Analysis of Cancerous Cell States. *J Proteome Res*. 2014 Dec 5;13(12):5486–95.
52. Forbes SA, Beare D, Gunasekaran P et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*. 2015 Jan 28;43(Database issue):D805–11.
53. Landrum MJ, Lee JM, Riley GR et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014 Jan 1;42(Database issue):D980–5.
54. Vizcaino JA, Côté RG, Csordas A et al. The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res*. 2013 Jan 29;41(Database issue):D1063–9.
55. Reiter L, Claassen M, Schrimpf SP et al. Protein Identification False Discovery Rates for Very Large Proteomics Data Sets Generated by Tandem Mass Spectrometry. *Mol Cell Proteomics*. 2009 Nov 16;8(11):2405–17.
56. Prieto G, Aloria K, Osinalde N, Fullaondo A, Arizmendi JM, Matthiesen R. PAnalyzer: A software tool for protein inference in shotgun proteomics. *BMC Bioinformatics*. 2012;13(1):288.
57. Roth RB, Hevezi P, Lee J et al. Gene expression analyses reveal molecular relationships among 20 regions of the

human CNS. *Neurogenetics*. 2006;7(2):67–80.

58. Jumeau F, Com E, Lane L et al. Human Spermatozoa as a Model for Detecting Missing Proteins in the Context of the Chromosome-Centric Human Proteome Project. *J Proteome Res*. 2015 Sep 4;14(9):3606–20.

59. Iyer MK, Niknafs YS, Malik R et al. The Landscape of Long Noncoding RNAs in the Human Transcriptome. *Nat Genet*. 2015 Mar 19;47(3):199–208.

60. Garin-Muga A, Ezcurdia I, Corrales FJ, Vázquez J S V. Proteogenomics Analysis of Novel Transcripts and Isoforms Using the Human MiTranscriptome Assembly. Poster Communication to the 14th Human Proteome Organisation (HUPO) World Congress. Vancouver, Canadá; 2015.

61. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14(4):R36.

62. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2012 Dec 9;31:46.

63. Ghosh S, Chan C-KK. Analysis of RNA-Seq Data Using TopHat and Cufflinks. *Methods Mol Biol*. 2016;1374:339–61.

64. Charif D, Lobry J. SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. Vol. 1, Structural approaches to sequence evolution. 2007. 207-232 p.

65. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Software*; Vol 1, Issue 5. 2008 Nov 10

66. Cagney G, Amiri S, Premawaradena T, Lindo M, Emili A. In silico proteome analysis to facilitate proteomics experiments using mass spectrometry. *Proteome Sci*. 2003 Aug 13;1:5.



Avances en microarrays de proteínas: tipos, ensayos y aplicaciones

Manuel Fuentes^{1,2}, Rodrigo Barderas³

¹Proteomics Unit, Cancer Research Centre (IBMCC/CSIC/USAL/IBSAL), 37007 Salamanca, Spain.

²Department of Medicine and Cytometry General Service-NUCLEUS. Cancer Research Centre. (IBMCC/CSIC/USAL/IBSAL/ CIBERONC), 37007 Salamanca, Spain.

³Functional Proteomics Unit, Chronic Disease Programme, Instituto de Salud Carlos III, 28222 Majadahonda, Madrid, Spain.

1. Introducción.

Actualmente, la caracterización exhaustiva y sistemática de los proteomas es una de las tareas más complejas en investigación biomédica; principalmente, porque las proteínas que los componen presentan características muy diversas en cuanto a sus niveles de expresión y localización, y, además, la misma proteína puede presentar múltiples isoformas o modificaciones postraduccionales -PTMs- que modifican su función, localización subcelular y/o estructura. Junto a esta inherente complejidad de los proteomas, el amplio rango dinámico en los niveles de expresión proteica constituye otro desafío a considerar en la caracterización sistemática y exhaustiva de un proteoma concreto.

Más allá de la proteómica basada en espectrometría de masas, otra posible solución para la caracterización de proteomas reside en el uso de microarrays de proteínas (1-6). La proteómica basada en microarrays de proteínas representa una herramienta valiosa para mejorar nuestro conocimiento en los mecanismos fisiológicos y/o patológicos, ya sea para la caracterización de proteomas, la identificación de alteraciones proteicas o bien para la caracterización inmunológica humoral (Fig. 1), con la particular ventaja sobre la proteómica basada en espectrometría de masas en que se conoce de antemano la proteína o el anticuerpo de estudio, lo cual facilita la verificación subsiguiente de los resultados. Además, dentro de las aproximaciones proteómicas disponibles para análisis masivos, los arrays de proteínas se han convertido en una poderosa herramienta para la medicina personalizada, el descubrimiento de biomarcadores y el subsiguiente análisis funcional; gracias a la principal ventaja que presentan los arrays de proteínas al permitir el análisis de miles de proteínas en un único experimento con un consumo mínimo de muestras y reactivos (7).



Los microarrays de proteínas consisten generalmente en microspots distribuidos geoméricamente ordenados en un soporte sólido que contiene reactivos de captura que pueden analizarse con muestras donde se presentan las moléculas de unión. En este sentido, los arrays de proteínas, constituyen una plataforma miniaturizada que permiten ensayos de alto rendimiento. Además del amplio espectro de aplicaciones de los arrays de proteínas, con el fin de diseñar un experimento masivo o utilizar concretamente un tipo u otro que afecte directamente los resultados finales de cualquier estudio biomédico y biológico, resulta necesario conocer las características y peculiaridades de los diferentes tipos de arrays de proteínas.

Tipos de Microarrays de Proteínas: Usos y Aplicaciones

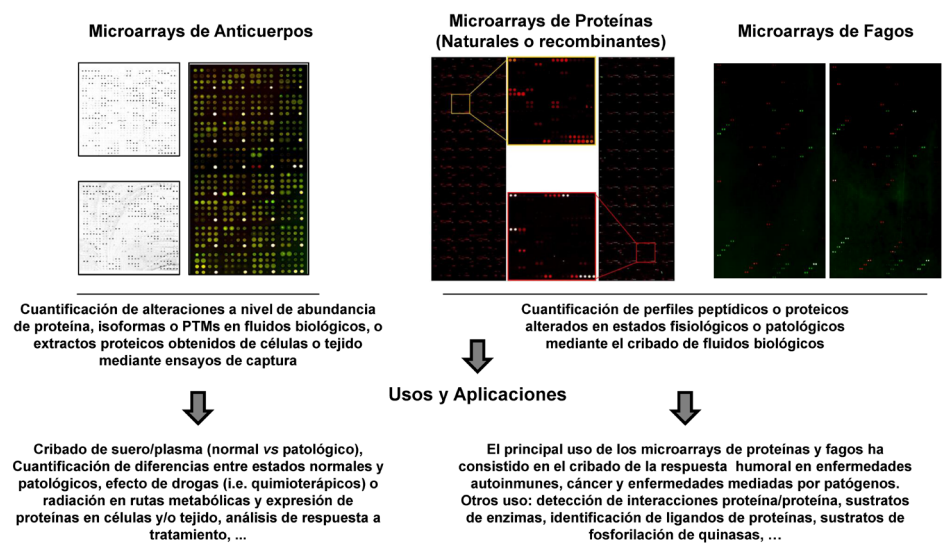


Figura 1. Estrategias proteómicas basadas en microarrays de proteínas. Detalle de los usos y aplicaciones de los microarrays de proteínas, anticuerpos y fagos. Se observan imágenes representativas de los diferentes tipos de microarrays de proteínas disponibles para la caracterización de la respuesta inmune humoral utilizando microarrays de proteínas y fagos, y la identificación de proteínas expresadas diferencialmente utilizando microarrays de anticuerpos. Izquierda: Imagen representativa de Microarrays de Anticuerpos representativos de un ensayo de captura con las muestras control y patológica marcadas con diferentes fluorocromos (derecha) y Microarray de anticuerpos representativos de un ensayo de captura con una muestra marcada con biotina seguido de streptavidina marcada con Cy3. Derecha: Imágenes representativas de microarrays de proteínas recombinantes y microarrays de fagos cribados con sueros control y patológicos.

En este capítulo, procedemos a describir los tipos, características y principales usos de los diferentes tipos de microarrays de proteínas de acuerdo con el contenido, el formato y los métodos de detección, para posteriormente, proporcionar una breve descripción de los métodos de impresión, hibridación y análisis de los datos de acuerdo a si los microarrays de proteínas son comerciales o personalizados (8), para concluir con las aplicaciones más recientes y relevantes en el campo.

2. Clasificación de los arrays de proteínas.

En la actualidad, existen varias clasificaciones de arrays de proteínas según su contenido (microarrays de proteínas, anticuerpos y fagos), método de obtención (arrays

ensamblados o auto-ensamblados), método de detección utilizado (libre de etiquetas o basados en etiquetas fluorescentes), o su formato (impresos/depositados en portas –formato plano- o en microarrays de esferas –beads arrays-).

2.1. Arrays de proteínas según su contenido

Globalmente, los microarrays de proteínas se dividen en tres grandes grupos atendiendo a su contenido: microarrays de proteínas (proteínas recombinantes expresadas en sistemas heterólogos o proteínas naturales aisladas de lisados celulares), microarrays de anticuerpos (si las biomoléculas impresas son anticuerpos o fragmentos de anticuerpos) y microarrays de fagos (las biomoléculas depositadas en los microarrays son fagos que despliegan en su superficie péptidos o proteínas asociados habitualmente a una determinada patología) (9-14).

Los microarrays de proteínas y de fagos se pueden utilizar para la identificación de interacciones proteína-proteína, identificación de pequeños ligandos, identificación de sustratos de quinasas... (1, 2). Sin embargo, la principal aplicación de los microarrays de proteínas y fagos está relacionada con la caracterización de la respuesta humoral en múltiples enfermedades (10, 15-23).

Actualmente, existen microarrays de proteínas de alta densidad disponibles comercialmente que contienen más de 9500 proteínas humanas recombinantes (HUProt y ProtoArray), y que hacen que su empleo esté únicamente limitado al presupuesto del laboratorio (1, 9, 20, 21, 23, 24). Además de los microarrays de proteínas completas, actualmente podemos encontrar microarrays de fragmentos de proteínas que contienen prácticamente el proteoma completo considerando un fragmento de proteína por cada gen (25). Esta última alternativa ha sido desarrollada en los últimos años por la plataforma Human Protein Atlas (www.proteinatlas.org) que ha producido anticuerpos y fragmentos de proteínas –PreSTs- para prácticamente cada proteína humana (sin distinguir entre sus isoformas) (26, 27). Por otra parte, los microarrays de fagos constituyen una alternativa económica y casera a los microarrays de proteínas comerciales para el descubrimiento de antígenos/auto-antígenos y anticuerpos/auto-anticuerpos específicos de patología. Esta última aproximación requiere del enriquecimiento de bibliotecas de fagos, usualmente del fago T7, que despliegan en su superficie péptidos de una determinada patología mediante la utilización de sueros de pacientes patológicos respecto de controles. Finalmente, fagos individuales se imprimen en portas de nitrocelulosa y se hibridan con sueros de pacientes y sueros control para identificar aquellos fagos que presentan los péptidos inmunorreactivos específicos de una determinada patología. Esta aproximación se ha utilizado con éxito en diferentes enfermedades, y especialmente en cáncer (10, 12, 15-17).

Finalmente, el otro gran grupo de microarrays de proteínas lo conforman los microarrays de anticuerpos o fragmentos de anticuerpos, cuya principal aplicación consiste en identificar perfiles de proteínas alteradas en una determinada patología (11, 28-32). Al igual que para los microarrays de proteínas, actualmente se pueden encontrar microarrays de anticuerpos comerciales que contienen más de mil anticuerpos frente a proteínas humanas (RayBiotech) (33, 34), lo que permite identificar en un único experimento proteínas alteradas en una determinada patología a un precio razonable



(34-36). Adicionalmente, también podemos encontrar microarrays de anticuerpos que contienen fragmentos de anticuerpos (scFvs, Fabs ...) (37, 38), aunque en este caso dichos microarrays de fragmentos de anticuerpos han tenido en los últimos años menos recorrido que los microarrays de anticuerpos completos, probablemente por tener entre 10-100 veces menor afinidad que los anticuerpos completos y por tener que encontrar un “framework” específico a nivel de scFvs para su correcta inmovilización en la superficie de los portas (38, 39).

2.2. Métodos de generación de microarrays de proteínas

De acuerdo a su método de obtención, los microarrays de proteínas los podemos clasificar en dos grupos principales de acuerdo al método utilizado para generar el contenido del array: arrays ensamblados y arrays auto-ensamblados (40, 41).

2.2.1. Arrays Ensamblados

Estas matrices se basan principalmente en la utilización de biomoléculas purificadas (como anticuerpos, proteínas, péptidos, etc...) inmovilizadas en una superficie funcionalizada en una distribución espacial geométrica. Los subtipos más importantes de arrays ensamblados son:

Arrays de captura: Se emplean biomoléculas específicas de analitos (como reactivos de afinidad-REA), generalmente anticuerpos, las cuales se depositan en la superficie y luego, mediante la interacción con la muestra biológica, es posible identificar específicamente y cuantificar la presencia del analito correspondiente. Estos arrays se utilizan en el descubrimiento de biomarcadores y/o en el establecimiento de perfiles diferenciales de expresión de proteínas. La calidad (especificidad y afinidad) de los REA está vinculada a la calidad de los resultados finales obtenidos. Además, los métodos de detección empleados para revelar el reconocimiento del analito con el reactivo de captura también juegan un papel importante en los resultados finales. Normalmente, marcaje fluorescente directo (por ejemplo, fluorocromos o fluoróforos Cy3/Cy5) o marcaje indirecto (por ejemplo, biotina-estreptavidina marcada con fluorescencia o HRP, anticuerpos conjugados con HRP, etc...).

Arrays de fase reversa: dichos arrays se basan en un concepto opuesto, donde las muestras se depositan en la superficie; y posteriormente la interacción de las muestras se evalúa simultáneamente para todos ellos con un único ligando (REA). Esta aproximación permite analizar muchas muestras simultáneamente utilizando una única molécula de consulta. Sin embargo, el principal problema asociado a dichos arrays consiste en conseguir optimizar los ensayos para evitar la reactividad cruzada intra-array e inter-array. Este subtipo de arrays, aunque requiere de mucho tiempo, es un buen enfoque para evaluar experimentalmente las rutas de señalización intracelular de proteínas en un formato de alto rendimiento; sin embargo, analizar y detectar ligandos de baja abundancia en muestras complejas puede ser problemático (42).

2.2.2. Arrays auto-ensamblados

Este tipo de arrays (Fig. 2) parece más complejo debido a que su característica principal se basa en la generación *in situ* de la proteína de interés en cada spot del array. Con dicho fin, todos los componentes requeridos para la expresión de la proteína *in situ* se tienen que inmovilizar en la superficie del array, para posteriormente, incubar el array con los reactivos necesarios para la transcripción-traducción *in vitro* (IVTT). La proteína nascente se captura mediante un reactivo de captura de afinidad (REA es habitualmente un anticuerpo), o mediante una etiqueta codificada en el extremo carboxi- o amino-terminal de la proteína.

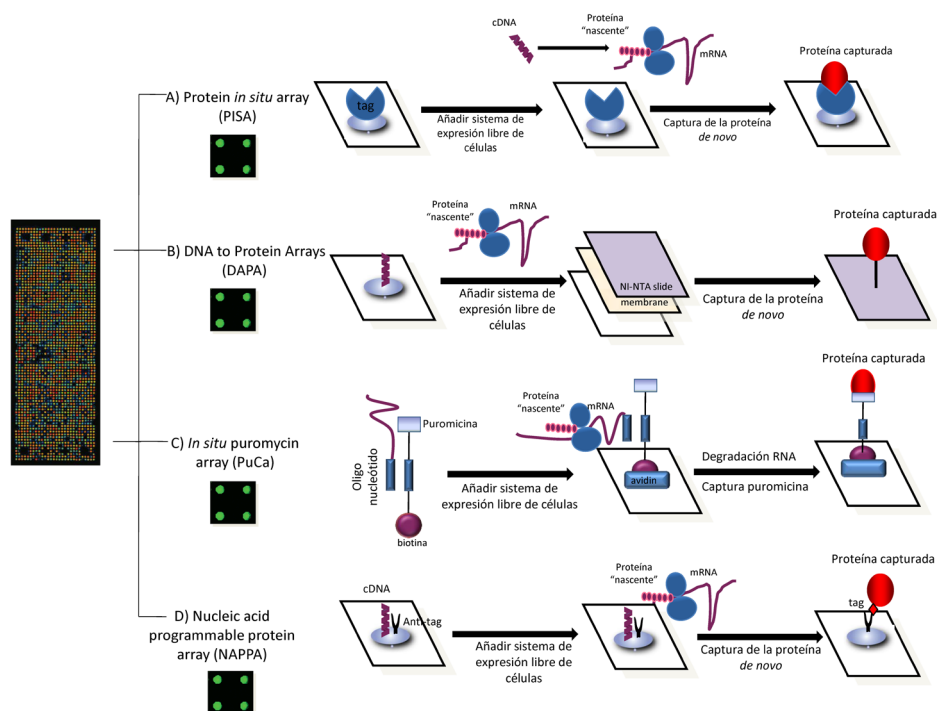


Figura 2. Representación esquemática de diferentes microarrays auto-ensamblados. A.-Protein In situ microarrays (PISA), B.-DNA-to-Protein Microarray (DAPA); C.-In situ Puromycin-Arrays (PuCa); D.-Nucleic Acids Programmable Protein Arrays (NAPPA).

La ventaja principal de estos microarrays es que ofrecen nuevas perspectivas de uso en múltiples aplicaciones, pues permiten la caracterización funcional y la identificación de PTMs. Aunque existen varias subclases (como PISA, DAPA, NAPPA, etc.) (43, 44), los arrays auto-ensamblados que más se han optimizado en los últimos años son los arrays tipo NAPPA (2, 13, 43-47) (Fig. 2), que incluso han permitido el desarrollo de biosensores auto-ensamblados (48). Los arrays NAPPA auto-ensamblados se han utilizado con éxito en los últimos años para el análisis de interacciones proteína-proteína y para detectar el interactoma de factores de transcripción (49) y, principalmente, para la identificación de biomarcadores (2, 13).



2.3. Formatos de los microarrays de proteínas

Los arrays de proteínas también se pueden clasificar en cuanto a la estructura física de su superficie, esférica o plana (43, 50):

2.3.1. Arrays planos

En este formato las biomoléculas se encuentran impresas en la superficie de un porta, que puede ser de nitrocelulosa o de vidrio recubierto de grupos químicos funcionales (amino-, epoxi-,...). Este tipo de microarrays son los convencionales ya que permiten la impresión de hasta aproximadamente 25000 spots en su superficie.

2.3.2. Arrays de microesferas

En este formato, los REA están unidos covalentemente a microesferas codificadas con diferentes colores, con un diámetro de 0.02 a 0.1 μm en el caso de las nanopartículas y de 0.1 a 6 μm en el caso de las microesferas (50). En este formato, hay al menos una distribución geométrica codificada mediante un color único por cada ligando y muestra. Para distinguirlos, las microesferas con diferentes ligandos se marcan con diferentes fluorocromos o combinaciones de fluorocromos. Mediante este código de color, las poblaciones de microesferas diferentes, cada una de un único código de color, se detectan mediante citometría de flujo; y utilizando otro fluorocromo como reportero, es posible identificar y cuantificar las biomoléculas diana.

Dichos arrays de microesferas han tenido un auge importante en los últimos años, gracias principalmente a encontrar microesferas (por ejemplo: Luminex®) de color único comerciales que permiten la inmovilización de hasta 384 moléculas diferentes, en un mismo ensayo. Por ello, dichos arrays se han convertido en una herramienta única para validar los resultados derivados de un cribado masivo realizado con microarrays de tipo plano.

2.4. Métodos de detección para microarrays de proteínas

Dependiendo del método de detección de la interacción en los microarrays, podemos hablar de detección basada en etiquetas –radioisótopos o fluorocromos–, o bien, mediante una detección libre de marcaje (51).

2.4.1. Métodos basados en marcaje con etiquetas

Marcaje fluorescente convencional

Los más utilizados son los fluoróforos convencionales como fluoresceína, BODIPY o cianinas, por su requerimiento mínimo en cuanto a seguridad y al desarrollo de escáneres de microarrays que permiten la lectura de un gran número de fluorocromos de diferente excitación y emisión. Entre los diferentes fluoróforos, las cianinas son las más utilizadas en los arrays de proteínas debido a su interacción mínima con otras biomo-

lécúlas y su alta intensidad. Además, se pueden combinar más de dos fluorocromos para analizar diferentes proteomas o identificar varios biomarcadores simultáneamente. Es el usado habitualmente para microarrays de tipo plano.

Microarrays de esferas: detección mediante citometría de flujo

La citometría de flujo se puede usar para detectar proteínas solubles presentes en fluidos corporales o en lisados celulares usando microesferas (de diferentes tamaños y con diferentes códigos de color). La citometría de flujo tiene dos láseres para el análisis cuantitativo y cualitativo que permiten conocer la unión de las proteínas diana en un tiempo mínimo. Comparando con técnicas fluorescentes convencionales, la citometría de flujo permite una evaluación cuantitativa-cualitativa rápida y precisa de un alto número de proteínas, con bajo coste y alta sensibilidad.

Sistemas basados en microesferas magnéticas

Estos sistemas se basan en el empleo de microesferas magnéticas conjugadas con anticuerpos que permiten la detección de proteínas gracias a su capacidad para capturar proteínas solubles con alta reproducibilidad y sensibilidad, asociadas con bajo ruido de fondo, un amplio rango dinámico y bajo coste.

Sistemas de marcaje fluorescente basados en puntos cuánticos

Los “puntos cuánticos” (QD) son nanocristales, formados por un núcleo de semiconductor y material fluorescente recubierto con otro semiconductor, que tienen propiedades ópticas muy estables y una gran diferencia entre las longitudes de onda de emisión y excitación. Estas partículas nanométricas podrían funcionar como etiquetas unidas con péptidos o anticuerpos para el reconocimiento de componentes celulares. Tienen una alta fluorescencia, mayor foto-estabilidad, excitación multicolor, un espectro de emisión ajustable y estrecho, y un mayor rendimiento cuántico, en comparación con los fluorocromos orgánicos. También tienen varias desventajas, como su alta susceptibilidad a la oxidación y la fotólisis, así como algunos riesgos para la salud humana y el medio ambiente. Se usan para localizar biomarcadores tumorales.

Marcaje con nanopartículas

Las nanopartículas de oro se utilizan en microarrays de proteínas debido a sus propiedades ópticas, eficiencia cuántica y compatibilidad con una amplia gama de longitudes de onda.

2.4.2. Métodos de detección libres de marcaje

Resonancia de Plasmón superficial (SPR)

Esta tecnología se basa en la generación de plasmones en una superficie, que son oscilaciones de electrones libres propagados paralelamente a la interfaz metal/medio, lo que permite medir los cambios en el índice de refracción de la superficie del sensor. La variación de intensidad de la luz reflejada y/o el ángulo de incidencia nos dicen si hay asociaciones o disociaciones moleculares, lo que permite determinar la relación entre la asociación y la disociación de biomoléculas entre sí. Además, la tecnología de resonancia de plasmón superficial también permite evaluar la afinidad y la especificidad de estas interacciones, facilitando la medida de interacciones biomoleculares en tiempo real con alta sensibilidad.



Microcantilevers

Microcantilevers son láminas delgadas de silicio recubiertas con una superficie de oro asociada con sistemas nanomecánicos de reconocimiento biomolecular. Por esta razón, los anticuerpos o proteínas están inmovilizados en dichas láminas, de modo que cuando existe una interacción entre estas moléculas y sus posibles ligandos, podemos medir la variación en la posición de la lámina usando diferentes sistemas ópticos o electrónicos.

Microscopía de Fuerza Atómica

Es un análisis microscópico que observa solo las variaciones topológicas de la superficie de un objeto. Se ha utilizado para la evaluación de la inmovilización de biomoléculas con una resolución micrométrica, y para la detección de interacciones de proteínas en la superficie de los arrays.

2.5. Tipos de arrays según su aplicación

Teniendo en cuenta la aplicación final, los arrays de proteínas se clasifican en dos categorías: arrays analíticos y arrays funcionales (Fig. 3).

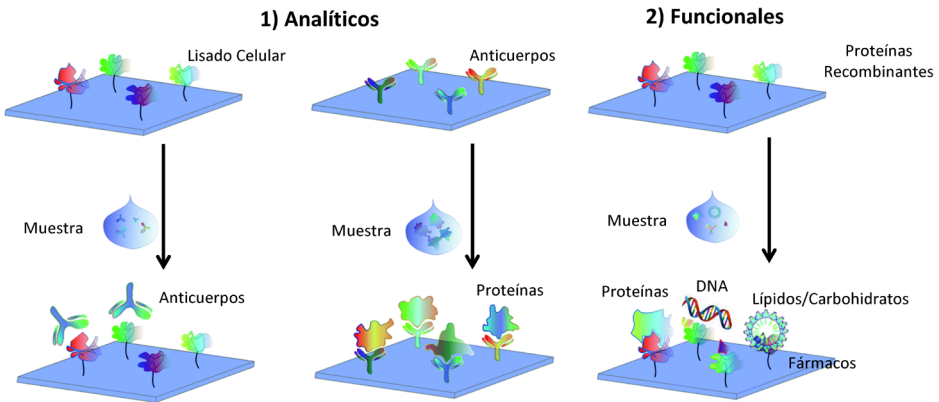


Figura 3. Esquema de arrays de proteínas según su función. Cuando se depositan anticuerpos o lisados celulares se clasifican en 1.-Arrays analíticos, Cuando se imprimen proteínas y se evalúa su interacción con otras biomoléculas se clasifican en 2.-Arrays Funcionales.

2.5.1. Arrays Analíticos

Estos microarrays se usan normalmente para identificar cuantitativamente la presencia de proteínas múltiples en un único ensayo. Comúnmente, la aplicación principal es la detección de proteínas expresadas de forma diferente y su abundancia en diferentes muestras. Generalmente, los microarrays analíticos se emplean para determinar biomarcadores, que son mediciones biométricas, que incluyen firmas moleculares, que predicen una afección biológica o clínica (por ejemplo, normal frente a patológica), a menudo con un potencial valor diagnóstico o pronóstico (43, 50, 52).

2.5.2. Arrays funcionales

Se utilizan principalmente para la caracterización e identificación de la función específica de las proteínas, así como sus interacciones con otras moléculas (incluyendo proteínas, péptidos, moléculas pequeñas/drogas, enzimas/sustratos o ácidos nucleicos,...) (2, 43, 53). Además, los arrays de proteínas funcionales también permiten la detección e identificación de PTMs, tales como glicosilación, fosforilación, acetilación,... que típicamente modulan la función, regulación y/o renovación de las proteínas. Como consecuencia, esta tecnología es altamente atractiva para la industria farmacéutica, pues el descubrimiento de estas interacciones fomenta el desarrollo de nuevos objetivos farmacéuticos, diagnósticos y terapéuticos.

3. Impresión, hibridación y análisis de datos.

Aunque los microarrays de anticuerpos, proteínas y fagos permiten el análisis en paralelo de miles de proteínas y múltiples muestras simultáneamente, sus diversas aplicaciones y su metodología de uso presentan algunas diferencias importantes.

En el caso de los microarrays de anticuerpos, el análisis en paralelo de múltiples proteínas en pequeños volúmenes de muestra se utiliza para medir alteraciones en la abundancia de cientos a miles de proteínas con el principal fin de identificar perfiles proteicos alterados en muestras biológicas. Para ello, la metodología para usarlos requiere de extracción de proteína, cuantificación de extractos proteicos obtenidos a partir de células o tejidos, o cuantificación de proteínas de fluidos biológicos, para estudiar posteriormente las alteraciones de proteínas a través de microarrays de anticuerpos utilizando las muestras marcadas con etiquetas o en ensayos en sandwich utilizando métodos libres de marcaje.

En el caso de los microarrays de fagos y microarrays de proteínas cuyo uso principal es el descubrimiento de biomarcadores asociados a la respuesta humoral requiere de una metodología de hibridación similar con modificaciones menores relacionadas con las diluciones de suero óptimas.

Sin embargo todos los microarrays comparten una metodología de impresión/deposición similar y en la que podemos encontrar arrayers o spotters que permiten la impresión por contacto -utilizando puntas (lisas o con reservorio)-, o sin contacto -utilizando métodos basados en fotoquímica, láser, deposición por electrospray y/o tecnologías de inyección de tinta- (54).

3.1. Impresión

La impresión por contacto tiene como principales ventajas la alta reproducibilidad de los puntos impresos y un fácil y más económico mantenimiento de los arrayers. Sin embargo, entre los inconvenientes se incluye el gran consumo de tiempo para imprimir los microarrays y el reducido número de portas (decenas a aproximadamente



una centena) que se pueden imprimir simultáneamente. Las técnicas de impresión sin contacto surgen de otras aplicaciones y como gran ventaja tienen el aumento considerable tanto de la velocidad de impresión como del número de portas que se pueden imprimir simultáneamente (llegando a poder imprimir, simultáneamente varias centenas e incluso miles de ellos). Sin embargo, algunas de las técnicas sin contacto tienen problemas derivados de la interferencia entre gotas satélite y la desnaturalización de las biomoléculas durante la impresión. Destacar asimismo, que actualmente hay equipos piezoeléctricos que permite la impresión simultánea de cientos de portas con una reproducibilidad equivalente a los arrays de contacto.

3.2. Hibridación

Los estudios proteómicos basados en microarrays de proteínas, anticuerpos o microarrays de fagos permiten la identificación de proteínas y péptidos específicamente alterados dentro de un individuo. Por otro lado, el empleo de muestras agrupadas de individuos con la misma patología en comparación con muestras agrupadas del grupo control, permite la identificación de péptidos y proteínas cuya alteración es común en la patología sin encontrarnos con problemas asociados a variabilidad intra-individuo. En cualquiera de los dos casos es recomendable hibridar los microarrays utilizando muestras de individuos dentro de un rango de edad similar y manteniendo tanto como sea posible la misma proporción entre mujeres y hombres.

A título de ejemplo describimos a continuación la metodología utilizada en nuestros laboratorios en el caso de utilizar microarrays de proteínas y fagos para el cribado de sueros para la detección de biomarcadores basados en la respuesta humoral, y por otra parte microarrays de anticuerpos para identificar proteínas alteradas en situaciones patológicas (9-12, 23). En ambas metodologías antes de comenzar la hibridación es importante atemperar los portas a 4°C durante 15 minutos para posteriormente mantenerlos a temperatura ambiente durante otros 15 minutos para evitar la desnaturalización de las proteínas impresas en su superficie y aclimatar los microarrays a la temperatura de trabajo.

3.2.1. Hibridación de microarrays de anticuerpos

Tanto si se trata de microarrays de anticuerpos comerciales, personalizados o impresos en el laboratorio en microarrays de nitrocelulosa o en portas de vidrio para ensayos en sándwich, es importante preparar todos los reactivos y filtrarlos justo antes de comenzar el ensayo. Además, se recomienda equilibrar los microarrays en tampón de lavado durante 1 minuto antes de comenzar el ensayo, realizando todos los pasos del protocolo en agitación y en oscuridad.

De forma resumida, el protocolo habitual de manejo de los microarrays de anticuerpos en nuestros laboratorios es el siguiente. Una vez equilibrado el array se incuba durante 2 horas a temperatura ambiente en un recipiente quadriPERM con 2 ml de tampón de bloqueo (PBS, Tween-20 0.1%, BSA 1%, leche en polvo 5%). Después de las 2 horas, se aspira el tampón de bloqueo y se añaden 1,6 ml de la muestra de proteína en una cámara de hibridación (Corning) incubando durante toda la noche a 4°C. Al

día siguiente se retiran con cuidado las muestras de cada pocillo y se lava cada porta 6 veces con 2 ml de tampón de lavado en un recipiente quadriPERM incubando cada vez durante 5 minutos a temperatura ambiente. Tras el último lavado se incuba nuevamente con 1,6 ml de cóctel de anticuerpos (conjugados, o no, con biotina) durante 2 horas a temperatura ambiente en una cámara de hibridación y se lava 6 veces cada porta como se indicó anteriormente. Finalmente, se incuba cada array con, alternativamente, 2 ml de estreptavidina conjugada a un fluorocromo o un anticuerpo IgG anti-ratón marcado con fluorescencia durante 2 h a temperatura ambiente (en un recipiente quadriPERM), para terminar el ensayo lavando nuevamente como se indica anteriormente. Finalmente, se realiza un lavado de 1 min en agua MilliQ para eliminar restos de sales que puedan provocar auto-fluorescencia y ruido de fondo durante el escaneo del array y se seca el array mediante centrifugación a 1500 rpm durante 10 min. Finalmente se escanea el porta en un escáner de microarrays con el láser adecuado y a la potencia de láser y ganancia de multiplicador indicada por el fabricante, o a la máxima posible, evitando la saturación de la fluorescencia de los puntos.

3.2.2. Hibridación de microarrays de proteínas y fagos

El flujo de trabajo para el cribado de la respuesta inmune humoral de pacientes y controles con microarrays de proteínas y fagos es casi equivalente con modificaciones mínimas entre ellos relacionadas con las diluciones de suero óptimas para el descubrimiento de biomarcadores.

Tras atemperar los microarrays de proteínas o fagos, se incuban los portas durante 1 h a temperatura ambiente, en agitación a 30 rpm, con 4 ml de tampón de bloqueo (PBS, BSA 1%, leche 3% y Tween 20 0.1% a pH 7.3) centrifugado a 4000 rpm durante 15 min para eliminar sustancias insolubles en un recipiente quadriPERM. Tras desechar el exceso de tampón de bloqueo, los arrays se incuban en una cámara de hibridación con 1,6 ml de suero diluido 1:300 en tampón de bloqueo durante 90 minutos a temperatura ambiente en agitación a 30 rpm. Para los microarrays de fagos se incuban la misma solución incluyendo una dilución 1:200 de anticuerpo anti-T7. Posteriormente, se lavan los microarrays 6 veces con PBS, Tween-20 al 0,1% durante 5 min a temperatura ambiente en un recipiente quadriPERM a 30 rpm y se incuban los arrays con 2 ml de una dilución 1:2000 de anticuerpo anti-IgG humana Alexa Fluor 647 en el tampón de bloqueo. En el caso de los microarrays de fagos se incuban la misma solución conteniendo una dilución 1:10000 de anticuerpo IgG anti-ratón marcado con Alexa Fluor 555 en un recipiente quadriPERM a 30 rpm. El protocolo se termina con un nuevo ciclo de lavados con PBS, Tween-20 al 0,1%, seguido de un lavado final en PBS para eliminar las trazas de detergente, y un último lavado en agua MilliQ para eliminar las sales y, finalmente, se secan los arrays mediante centrifugación a 1500 rpm durante 10 minutos.

3.3. Escaneo y análisis de datos

Para el escaneo de microarrays, se necesita un escáner específico con láseres de 532 nm y 635 nm compatible con todas las superficies usadas, incluidos los microarrays de nitrocelulosa, como el GenePix 4000B.



El láser de 635 nm se usa para la detección de Cy5, Alexa Fluor 647 o una señal con un espectro similar (canal rojo), mientras que el láser de 532 nm se utiliza para la detección de Cy3, Alexa Fluor 555 o una señal espectralmente similar (canal verde). En todos los casos, siempre es recomendable escanear los microarrays a dos potenciales de láser diferentes (100% y 10%) para un mejor análisis de los puntos de baja y alta reactividad, respectivamente.

Las imágenes obtenidas se cuantifican utilizando un software de adquisición de imágenes de microarrays, como el programa GenePix Pro 7 que permite la normalización intra-array usando la señal de ruido de fondo alrededor de cada punto (Fig. 4). Los datos obtenidos por el software para cada punto en el microarray son, posteriormente, normalizados inter-array y analizados utilizando los test estadísticos deseados para identificar los puntos (proteínas, o fagos reactivos) estadísticamente significativos en la patología. Para ello se pueden usar herramientas bioinformáticas específicas, como los programas gratuitos incluidos en la suite de Babelomics para la obtención de perfiles funcionales, Pomelo II o el MultiExperiment Viewer (55-57), los cuales permiten realizar tanto la normalización inter-array como el procesamiento de los datos obtenidos de los microarrays. Por otra parte, los microarrays de proteínas disponibles comercialmente también han desarrollado sus propias herramientas bioinformáticas que permiten realizar el análisis. Un ejemplo es el ProtoArray Prospector Software (58), que viene incluido con los microarrays de proteínas comerciales ProtoArray. Estas herramientas permiten identificar aquellas proteínas con diferencias estadísticamente significativas en cuanto a su abundancia relativa, en las condiciones del estudio que hayamos realizado.

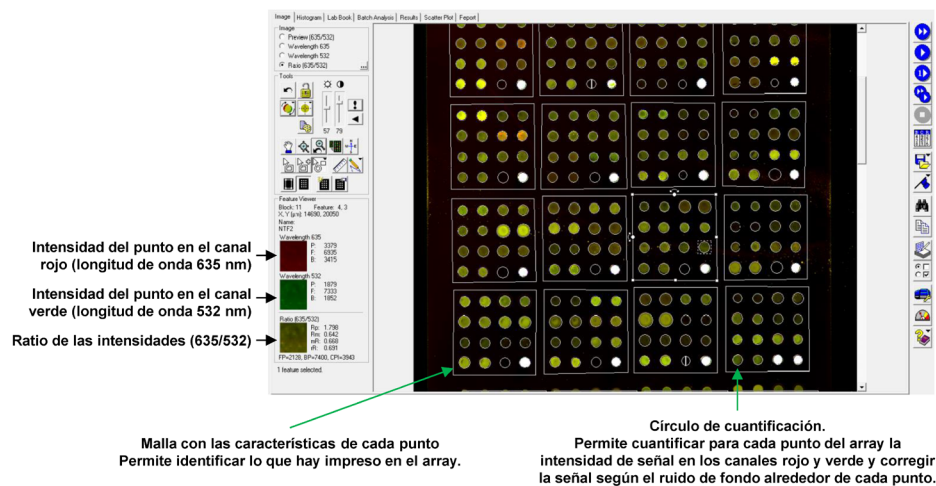


Figura 4. Cuantificación de microarrays de proteínas. Detalle de la imagen de un Microarray de Anticuerpos incubado con una muestra control marcada con Cy3 (verde) y muestra patológica marcada con Cy5 (rojo) utilizando el programa GenePix Pro 7. En la imagen se detallan 16 subarrays con la malla ajustada de un Microarray de anticuerpos comercial de la serie Panorama de la compañía Sigma.

3.4. Validación

Dado que en cualquier experimento de proteómica nos podemos encontrar con la identificación de falsos positivos, y los microarrays de proteínas no son una excepción, los resultados obtenidos a partir de microarrays de anticuerpos, proteínas y microarrays de fagos deben validarse mediante otras aproximaciones para garantizar

la precisión de los resultados. De hecho, un problema recurrente en relación con múltiples estudios obtenidos a través de análisis de microarrays de proteínas reside en la ausencia de validación del conjunto de datos. Así, las identificaciones erróneas y la ausencia de datos validados pueden suponer un problema para trasladar a la clínica los resultados obtenidos en investigación básica.

Por todo ello, es importante realizar una validación adicional y un análisis funcional de los datos obtenidos a partir de microarrays de proteínas mediante el uso de técnicas complementarias como dot-blot, WB o ELISA. Sin embargo, esta estrategia está limitada por la disponibilidad y el alto coste de los anticuerpos que reconocen a las proteínas de interés. Otra alternativa de validación consiste en el uso de espectrometría de masas para la validación de proteínas muy abundantes o moderadamente abundantes en fluidos biológicos o extractos de proteínas complejas (59). Los criterios de validación de los resultados, deberían tener en cuenta el (los) objetivo (s) del estudio, la naturaleza de la metodología y la variabilidad biológica del biomarcador a validar. Globalmente, estos ensayos de validación deberían dar como resultado la obtención y publicación de datos bioanalíticos validados de alta calidad.

4. Aplicaciones de los microarrays de proteínas.

Durante la última década, los arrays de proteínas se han empleado con éxito en múltiples ensayos para el descubrimiento de biomarcadores; aquí, se comentan algunos de estos estudios para ilustrar la utilidad de esta aproximación proteómica.

4.1. Biomarcadores específicos en oncología y neurodegeneración

Las primeras aproximaciones utilizadas para la identificación de biomarcadores utilizando microarrays de anticuerpos de alta densidad se publicó en la década de los 2000. El grupo de Brian Haab fue pionero en estos estudios, definiendo un panel de cinco proteínas séricas significativamente expresadas en el suero de pacientes de cáncer de próstata respecto a las muestras control (60). Este mismo grupo también identificó 84 proteínas con diferencias en la abundancia relativa entre pacientes con cáncer de pulmón y controles sanos (61). En un enfoque similar, Wittekind y colaboradores publicaron un estudio en el cual describían un conjunto de proteínas como candidatos a biomarcadores asociados a carcinoma hepatocelular (62).

Hoy en día, más allá de estos estudios iniciales con pocas decenas de anticuerpos, podemos encontrar microarrays de anticuerpos que contienen miles de anticuerpos diferentes que distinguen entre isoformas de la misma proteína y diferentes PTMs, permitiendo realizar múltiples estudios para la identificación de biomarcadores en varias patologías oncológicas; por ejemplo: Amonkar y colaboradores utilizando esta tecnología en cáncer de ovario determinaron 11 proteínas alteradas (63); Sreekumar y colaboradores identificaron un panel de proteínas como candidatos a biomarcadores en células de carcinoma de colon (64). En este mismo cáncer, se han podido determinar las rutas alteradas (35), y encontrar citoquinas, adipocinas y factores de crecimiento



asociados a la enfermedad (22, 36, 65). Más recientemente Díez y colaboradores identificaron un perfil de proteínas diferencialmente expresadas en leucemia linfocítica crónica de células B que están relacionadas con la respuesta a terapia y las alteraciones citogenéticas (66). Además, previamente, Abajo y colaboradores desarrollaron un array de anticuerpos para determinar el inmunofenotipo de 1100 muestras de leucemia y linfomas, llegando a identificar un panel de 82 clústeres de diferenciación estadísticamente significativo (67).

Más allá de los estudios oncológicos, en los que encontrar muestras de tejido y suero bien caracterizadas no es un problema, y en los que se han empleado estas aproximaciones en múltiples tipos de cáncer, es de destacar que recientemente se han utilizado para la caracterización de enfermedades neurodegenerativas, entre otras. Habitualmente, la caracterización de enfermedades neurodegenerativas ha sido un problema porque el diagnóstico definitivo se da habitualmente *post mortem*. Ello ha provocado que el acceso a muestras bien caracterizadas pareadas entre suero y tejido en este tipo de enfermedades haya sido difícil, recurriendo en algunos casos a la utilización de modelos animales (68). A pesar de ello, en los últimos años se ha empezado su caracterización con muestras plasmáticas y muestras de tejido de pacientes de enfermedades neurodegenerativas, encontrando tanto paneles de proteínas plasmáticas como proteínas tisulares alteradas en la enfermedad de Alzheimer (11, 34).

4.2. Diseño de Vacunas

En 2014, Cid y colaboradores publicaron el primer estudio sobre interacciones proteína patógeno-huésped. En este trabajo, se estudió la presencia de modificaciones postranscripcionales en proteínas efectoras, proteínas T3SS, de diferentes mutantes de *Salmonella typhimurium*, para infectar células HeLa *in vitro*. La colección de lisados que representaba todas las condiciones de infección se imprimieron y se utilizaron con varios anticuerpos validados, mostrando resultados comparativos entre los diferentes ensayos según la abundancia relativa de las proteínas o las modificaciones postranscripcionales (69). Otro estudio similar, realizado por Li y colaboradores permitió identificar 149 antígenos de *Yersinia pestis* potencialmente inmunogénicos (70), útiles para el diseño de vacunas. En 2009, Thanawastein y colaboradores diseñaron y desarrollaron un array de proteínas cuyo contenido estaba basado en activadores inmunes (EPSIA) que se empleó para la identificación de nuevas proteínas inmuno-estimuladoras frente a la bacteria de *Vibrio cholerae* (71).

En un estudio más reciente, Montor y colaboradores describieron, utilizando microarrays de proteínas NAPPA, antígenos de membrana de *Pseudomonas aeruginosa* que discriminaban entre pacientes infectados con *P. aeruginosa* e individuos sanos, y por ende proteínas potencialmente candidatas para vacunas. En este trabajo, se identificaron 12 proteínas, la mayoría de ellas relacionadas con la respuesta inmune adaptativa en pacientes infectados (46). Más recientemente, Manzano y colaboradores detectaron un conjunto de nuevas proteínas de *O. moubata* candidatas para vacunas basadas en una caracterización sistemática de más de 2000 interacciones huésped-patógeno evaluadas mediante microarrays de proteínas auto-ensamblado basados en una biblioteca de cDNAs que contenía más de 400 proteínas recombinantes de *O. moubata* (72).

4.3. Enfermedades auto-inmunes

El sistema inmune humoral genera anticuerpos (“auto-anticuerpos”) frente a auto-antígenos alterados en respuesta a muchos procesos patológicos. Se cree que los auto-anticuerpos se generan debido a la alteración de las proteínas por su sobre-expresión, mutación, modificaciones postranscripcionales alteradas, degradación alterada, o por su liberación del tejido dañado; alterando la antigeneidad de esas proteínas y por tanto su transformación en auto-antígenos (73). Todo ello conduce a su reconocimiento por el sistema inmune.

Los auto-anticuerpos específicos de patología tienen varios beneficios que los convierten en una fuente idónea de biomarcadores: 1.-se detectan antes de que aparezcan los síntomas clínicos; 2.-son fáciles de detectar incluso a niveles bajos una vez que se conoce su antígeno diana; 3.-son fáciles de analizar pues se obtienen a través de una prueba mínimamente invasiva (como es la sangre periférica); y 4.-están presentes en niveles más altos y con una vida media superior a sus antígenos diana, que están presentes transitoriamente en la sangre pues sufren una degradación intensiva.

Estas características de los auto-anticuerpos han provocado que se haya caracterizado con intensidad la respuesta humoral que se produce en situaciones patológicas mediante la utilización de microarrays de proteínas ensamblados y auto-ensamblados y microarrays de fagos (9, 10, 15-17, 22, 23, 74).

La primera vez que los microarrays de proteínas NAPPA se utilizaron para la detección serológica fue en 2007 por Anderson y colaboradores. Dichos autores estudiaron la presencia de anticuerpos contra antígenos tumorales en cáncer de mama. Los autores mostraron que los niveles de anticuerpos específicos de p53 eran significativamente más bajos en donantes sanos que en pacientes con cáncer de mama y que la respuesta al auto-antígeno p53 se detectaba en estadios II de la enfermedad [30]. En estudios posteriores a este trabajo, se analizó la respuesta humoral en cáncer de mama en microarrays NAPPA conteniendo aproximadamente 5000 antígenos asociados a cáncer (75, 76). Posteriormente, también analizaron la respuesta humoral en cáncer de ovario y la contribución de los auto-anticuerpos de p53 en diferentes tipos de cáncer (13, 75, 76). Con un flujo de trabajo similar, el grupo del Prof. Labaer desarrolló un NAPPA array piloto para evaluar los auto-anticuerpos presentes en la artritis idiopática juvenil (AIJ), que es una enfermedad caracterizada por inflamación crónica de las articulaciones en los niños (77). Recientemente, el laboratorio del Dr. Fuentes ha realizado una detección de auto-anticuerpos en osteoartritis y artritis reumatoide mediante el uso de microarrays NAPPA, validado con otras tecnologías de microarrays de proteínas (78).

Por otra parte, otra de las tecnologías más ampliamente utilizadas para caracterizar la respuesta humoral han sido los microarrays de proteínas comerciales –ProtoArray, HuProt- y microarrays de fagos. Dichas tecnologías se han utilizado con éxito para la caracterización de la respuesta humoral en cáncer colorrectal, cáncer de próstata, ovario, glioma,... (9, 10, 15-17, 22, 23, 74), entre otras enfermedades (12,79).



4.4. Biomarcadores funcionales como dianas farmacológicas

Entre los múltiples estudios realizados con microarrays de proteínas para análisis de Proteómica Funcional, los microarrays de proteína tipo NAPPa han permitido la caracterización funcional de múltiples interacciones en situaciones fisiológicas y/o patológicas que una vez trasladadas podrían constituir nuevas dianas farmacológicas.

Hace una década, el grupo del Prof. Labaer evaluó la utilidad de los microarrays NAPPa para la detección de interacciones proteína-proteína. En uno de los artículos iniciales de esta tecnología se verificó la utilidad de esta aproximación para la detección de interacciones conocidas, incluyendo Jun-Fos y p53-MDM2, mediante la utilización de un microarray con 647 genes únicos impresos por duplicado [33]. Más recientemente se ha utilizado esta tecnología para la caracterización de interacciones de factores de transcripción, y la detección de nuevas vías de señalización celular en *Arabidopsis thaliana* (49). Además, utilizando también microarrays NAPPa multiplexados que aumentan la densidad del array, se han identificado cuatro nuevos antígenos relacionados con la tuberculosis en cobayas vacunadas con *Bacillus Calmette-Guerin* (BCG) y validadas posteriormente mediante ELISA (45). En otro estudio reciente, Manzano y colaboradores publicaron un trabajo donde utilizaron NAPPa para estudiar las interacciones de proteínas. En este estudio, se identificaron nuevos interactores para P-selectina y fosfolipasa A2, validando otras múltiples interacciones (72).

5. Conclusiones y perspectivas.

En este capítulo de libro, hemos revisado los avances realizados en los últimos años en el campo de la proteómica basada en microarrays de proteínas; desde la impresión de dichos microarrays, su clasificación, hasta los usos y aplicaciones más recientes en el campo. Sin embargo, como describimos anteriormente, a pesar de todos los grandes avances realizados en microarrays de proteínas- tanto desde el punto de vista tecnológico/metodológico como en la ampliación del abanico de aplicaciones- aún sigue siendo un desafío su utilización para la caracterización del proteoma humano completo.

En los próximos años, se espera la obtención de microarrays de proteínas impresos con el proteoma completo, considerando un "open reading frame" por gen, de forma similar a como se han desarrollado los microarrays de DNA. Con el desarrollo de dichos microarrays, se conseguirá obtener información sobre las funciones de las proteínas y genes cuyas funciones se desconocen actualmente, y que según diferentes estudios puede suponer aproximadamente el 50% de todas ellas.

Finalmente, queremos destacar que aunque todavía es necesario desarrollar y optimizar algunos aspectos clave de los microarrays de proteínas, en un futuro cercano, la complementariedad que ofrecen los microarray de proteínas, junto con los ensayos proteómicos basados en espectrometría de masas proporcionarían los resultados complementarios necesarios para la definición concreta del proteoma humano, así como el análisis funcional del proteoma.

Agradecimientos:

Este trabajo se ha realizado con el apoyo de las siguientes fuentes de financiación: PI17CIIII/00045 del programa AES-ISCI a R.B. y la red temática ARADyAL (RD16/0006/0014) cofinanciadas por el Fondo Europeo de Desarrollo Regional -FEDER-. Asimismo agradecemos el apoyo del programa AES del Instituto de Salud Carlos III: FIS PI17/01930 y CB16/12/00400 y la Fundación Solórzano FS/38-2017 a M.F. La Unidad de Proteómica pertenece a ProteoRed, PRB3-ISCI, que está financiada por el programa PT17/0019/0023 del PE I+D+I 2017-2020 co-financiado por el ISCI y FEDER.

REFERENCIAS

1. Fasolo J, Im H, Snyder MP. Probing High-density Functional Protein Microarrays to Detect Protein-protein Interactions. *J Vis Exp*. 2015(102):e51872.
2. LaBaer J, Ramachandran N. Protein microarrays as tools for functional proteomics. *Curr Opin Chem Biol*. 2005;9(1):14-9.
3. Mattoon DR, Schweitzer B. Profiling protein interaction networks with functional protein microarrays. *Methods Mol Biol*. 2009;563:63-74.
4. Moore CD, Ajala OZ, Zhu H. Applications in high-content functional protein microarrays. *Curr Opin Chem Biol*. 2016;30:21-7.
5. Barderas R, Villar-Vázquez R, Casal JI. Colorectal Cancer Circulating Biomarkers. In: Preedy VR, Patel VB, editors. *Biomarkers in Cancer*. Dordrecht: Springer Netherlands; 2014. p. 1-21.
6. Barderas R, Babel I, Casal JI. Colorectal cancer proteomics, molecular characterization and biomarker discovery. *Proteomics Clin Appl*. 2010;4(2):159-78.
7. Ray S, Reddy PJ, Jain R, Gollapalli K, Moiyadi A, Srivastava S. Proteomic technologies for the identification of disease biomarkers in serum: advances and challenges ahead. *Proteomics*. 2011;11(11):2139-61.
8. Sutandy FX, Qian J, Chen CS, Zhu H. Overview of protein microarrays. *Curr Protoc Protein Sci*. 2013;Chapter 27:Unit 27 1.
9. 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. *Molecular & cellular proteomics : MCP*. 2009;8(10):2382-95.
10. Babel I, Barderas R, Diaz-Uriarte R, Moreno V, Suarez A, Fernandez-Acenero MJ, et al. Identification of MST1/STK4 and SULF1 proteins as autoantibody targets for the diagnosis of colorectal cancer by using phage microarrays. *Molecular & cellular proteomics : MCP*. 2011;10(3):M110 001784.
11. Garranzo-Asensio M, San Segundo-Acosta P, Martinez-Useros J, Montero-Calle A, Fernandez-Acenero MJ, Haggmark-Manberg A, et al. Identification of prefrontal cortex protein alterations in Alzheimer's disease. *Oncotarget*. 2018;9(13):10847-67.
12. San Segundo-Acosta P, Garranzo-Asensio M, Oeo-Santos C, Montero-Calle A, Quiralte J, Cuesta-Herranz J, et al. High-throughput screening of T7 phage display and protein microarrays as a methodological approach for the identification of IgE-reactive components. *J Immunol Methods*. 2018;456:44-53.
13. Katchman BA, Barderas R, Alam R, Chowell D, Field MS, Esserman LJ, et al. Proteomic mapping of p53 immunogenicity in pancreatic, ovarian, and breast cancers. *Proteomics Clin Appl*. 2016;10(7):720-31.
14. Villar-Vazquez R, Padilla G, Fernandez-Acenero MJ, Suarez A, Fuente E, Pastor C, et al. Development of a novel multiplex beads-based assay for autoantibody detection for colorectal cancer diagnosis. *Proteomics*. 2016;16(8):1280-90.
15. Chatterjee M, Ionan A, Draghici S, Tainsky MA. Epitomics: global profiling of immune response to disease using protein microarrays. *OMICS*. 2006;10(4):499-506.
16. Chatterjee M, Mohapatra S, Ionan A, Bawa G, Ali-Fehmi R, Wang X, et al. Diagnostic markers of ovarian cancer by high-throughput antigen cloning and detection on arrays. *Cancer Res*. 2006;66(2):1181-90.
17. Wang X, Yu J, Sreekumar A, Varambally S, Shen R, Giacherio D, et al. Autoantibody signatures in prostate cancer. *N Engl J Med*. 2005;353(12):1224-35.
18. Nagele E, Han M, Demarshall C, Belinka B, Nagele R. Diagnosis of Alzheimer's disease based on disease-specific autoantibody profiles in human sera. *PLoS One*. 2011;6(8):e23112.
19. Nagele EP, Han M, Acharya NK, DeMarshall C, Kosciuk MC, Nagele RG. Natural IgG autoantibodies are abundant and ubiquitous in human sera, and their number is influenced by age, gender, and disease. *PLoS One*. 2013;8(4):e60726.
20. DeMarshall C, Sarkar A, Nagele EP, Goldwaser E, Godsey G, Acharya NK, et al. Utility of autoantibodies as



biomarkers for diagnosis and staging of neurodegenerative diseases. *Int Rev Neurobiol*. 2015;122:1-51.

21. Han M, Nagele E, DeMarshall C, Acharya N, Nagele R. Diagnosis of Parkinson's disease based on disease-specific autoantibody profiles in human sera. *PLoS One*. 2012;7(2):e32383.
22. Barderas R, Babel I, Diaz-Uriarte R, Moreno V, Suarez A, Bonilla F, et al. An optimized predictor panel for colorectal cancer diagnosis based on the combination of tumor-associated antigens obtained from protein and phage microarrays. *J Proteomics*. 2012;75(15):4647-55.
23. Barderas R, Villar-Vazquez R, Fernandez-Acenero MJ, Babel I, Pelaez-Garcia A, Torres S, et al. Sporadic colon cancer murine models demonstrate the value of autoantibody detection for preclinical cancer diagnosis. *Sci Rep*. 2013;3:2938.
24. Cox E, Uzoma I, Guzzo C, Jeong JS, Matunis M, Blackshaw S, et al. Identification of SUMO E3 ligase-specific substrates using the HuProt human proteome microarray. *Methods Mol Biol*. 2015;1295:455-63.
25. Ayoglu B, Haggmark A, Khademi M, Olsson T, Uhlen M, Schwenk JM, et al. Autoantibody profiling in multiple sclerosis using arrays of human protein fragments. *Molecular & cellular proteomics : MCP*. 2013;12(9):2657-72.
26. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419.
27. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol*. 2010;28(12):1248-50.
28. Ahn SB, Khan A. Detection and quantitation of twenty-seven cytokines, chemokines and growth factors pre- and post-high abundance protein depletion in human plasma. *EuPA Open Proteomics*. 2014;3:78-84.
29. Stampolidis P, Ullrich A, Iacobelli S. LGALS3BP, lectin galactoside-binding soluble 3 binding protein, promotes oncogenic cellular events impeded by antibody intervention. *Oncogene*. 2015;34(1):39-52.
30. Yu XB, Wallstrom G, Magee DM, Qiu J, Mendoza DEA, Wang J, et al. Quantifying antibody binding on protein microarrays using microarray nonlinear calibration. *Biotechniques*. 2013;54(5):257+.
31. Hospach I, Joseph Y, Mai MK, Krasteva N, Nelles G. Fabrication of Homogeneous High-Density Antibody Microarrays for Cytokine Detection. *Microarrays (Basel)*. 2014;3(4):282-301.
32. Barderas R, Bartolome RA, Fernandez-Acenero MJ, Torres S, Casal JI. High expression of IL-13 receptor alpha2 in colorectal cancer is associated with invasion, liver metastasis, and poor prognosis. *Cancer Res*. 2012;72(11):2780-90.
33. Huang RP. Detection of multiple proteins in an antibody-based protein microarray system. *J Immunol Methods*. 2001;255(1-2):1-13.
34. Jaeger PA, Lucin KM, Britschgi M, Vardarajan B, Huang RP, Kirby ED, et al. Network-driven plasma proteomics expose molecular changes in the Alzheimer's brain. *Mol Neurodegener*. 2016;11:31.
35. Madoz-Gurpide J, Canamero M, Sanchez L, Solano J, Alfonso P, Casal JI. A proteomics analysis of cell signaling alterations in colorectal cancer. *Molecular & cellular proteomics : MCP*. 2007;6(12):2150-64.
36. Pena C, Garcia JM, Larriba MJ, Barderas R, Gomez I, Herrera M, et al. SNAI1 expression in colon cancer related with CDH1 and VDR downregulation in normal adjacent tissue. *Oncogene*. 2009;28(49):4375-85.
37. Babel I, Barderas R, Pelaez-Garcia A, Casal JI. Antibodies on demand: a fast method for the production of human scFvs with minimal amounts of antigen. *BMC Biotechnol*. 2011;11:61.
38. Delfani P, Dextrin Mellby L, Nordstrom M, Holmer A, Ohlsson M, Borrebaeck CA, et al. Technical Advances of the Recombinant Antibody Microarray Technology Platform for Clinical Immunoproteomics. *PLoS One*. 2016;11(7):e0159138.
39. Borrebaeck CA, Wingren C. Design of high-density antibody microarrays for disease proteomics: key technological issues. *J Proteomics*. 2009;72(6):928-35.
40. Valiokas R. Nanobiochips. *Cell Mol Life Sci*. 2012;69(3):347-56.
41. Matarraz S, Gonzalez-Gonzalez M, Jara M, Orfao A, Fuentes M. New technologies in cancer. Protein microarrays for biomarker discovery. *Clin Transl Oncol*. 2011;13(3):156-61.
42. Ekins R, Chu F. Immunoassay and other ligand assays: present status and future trends. *J Int Fed Clin Chem*. 1997;9(3):100-9.
43. Fuentes M, Diez P, Casado-Vela J. Nanotechnology in the Fabrication of Protein Microarrays. *Methods Mol Biol*. 2016;1368:197-208.
44. Ramachandran N, Raphael JV, Hainsworth E, Demirkan G, Fuentes MG, Rolfs A, et al. Next-generation high-density self-assembling functional protein arrays. *Nat Methods*. 2008;5(6):535-8.
45. Yu X, Song L, Petritis B, Bian X, Wang H, Vilorio J, et al. Multiplexed Nucleic Acid Programmable Protein Arrays. *Theranostics*. 2017;7(16):4057-70.
46. Montor WR, Huang J, Hu Y, Hainsworth E, Lynch S, Kronish JW, et al. Genome-wide study of *Pseudomonas aeruginosa* outer membrane protein immunogenicity using self-assembling protein microarrays. *Infect Immun*. 2009;77(11):4877-86.
47. Ramachandran N, Hainsworth E, Bhullar B, Eisenstein S, Rosen B, Lau AY, et al. Self-assembling protein microarrays. *Science*. 2004;305(5680):86-90.
48. Garranzo-Asensio M, Guzman-Aranguiz A, Poves C, Fernandez-Acenero MJ, Torrente-Rodriguez RM, Ruiz-Valdepenas Montiel V, et al. Toward Liquid Biopsy: Determination of the Humoral Immune Response in Cancer Patients Using HaloTag Fusion Protein-Modified Electrochemical Bioplatfroms. *Anal Chem*. 2016;88(24):12339-45.
49. Yazaki J, Galli M, Kim AY, Nito K, Aleman F, Chang KN, et al. Mapping transcription factor interactome networks using HaloTag protein arrays. *Proc Natl Acad Sci U S A*. 2016;113(29):E4238-47.
50. Gonzalez-Gonzalez M, Bartolome R, Jara-Acevedo R, Casado-Vela J, Dasilva N, Matarraz S, et al. Evaluation of homo- and hetero-functionally activated glass surfaces for optimized antibody arrays. *Anal Biochem*.

51. Driskell JD, Tripp RA. Emerging Technologies in Nanotechnology-Based Pathogen Detection. *Clinical Microbiology Newsletter*. 2009;31(18):137.
52. Fuentes M. Protein microarrays overview. *Recent Pat Biotechnol*. 2013;7(2):83.
53. Ellington AA, Kullo IJ, Bailey KR, Klee GG. Antibody-based protein multiplex platforms: technical and operational challenges. *Clin Chem*. 2010;56(2):186-93.
54. Barbulovic-Nad I, Lucente M, Sun Y, Zhang M, Wheeler AR, Bussmann M. Bio-microarray fabrication techniques--a review. *Crit Rev Biotechnol*. 2006;26(4):237-59.
55. Morrissey ER, Diaz-Uriarte R. Pomelo II: finding differentially expressed genes. *Nucleic Acids Res*. 2009;37(Web Server issue):W581-6.
56. Alonso R, Salavert F, Garcia-Garcia F, Carbonell-Caballero J, Bleda M, Garcia-Alonso L, et al. Babelomics 5.0: functional interpretation for new generations of genomic data. *Nucleic Acids Res*. 2015;43(W1):W117-21.
57. Howe EA, Sinha R, Schlauch D, Quackenbush J. RNA-Seq analysis in MeV. *Bioinformatics*. 2011;27(22):3209-10.
58. Sboner A, Karpikov A, Chen G, Smith M, Mattoon D, Freeman-Cook L, et al. Robust-linear-model normalization to reduce technical variability in functional protein microarrays. *J Proteome Res*. 2009;8(12):5451-64.
59. McIntosh M, Fitzgibbon M. Biomarker validation by targeted mass spectrometry. *Nat Biotechnol*. 2009;27(7):622-3.
60. Miller JC, Butler EB, Teh BS, Haab BB. The application of protein microarrays to serum diagnostics: prostate cancer as a test case. *Dis Markers*. 2001;17(4):225-34.
61. Gao WM, Kuick R, Orzechowski RP, Misek DE, Qiu J, Greenberg AK, et al. Distinctive serum protein profiles involving abundant proteins in lung cancer patients based upon antibody microarray analysis. *BMC Cancer*. 2005;5:110.
62. Tannapfel A, Anhalt K, Hausermann P, Sommerer F, Benicke M, Uhlmann D, et al. Identification of novel proteins associated with hepatocellular carcinomas using protein microarrays. *J Pathol*. 2003;201(2):238-49.
63. Amonkar SD, Bertenshaw GP, Chen TH, Bergstrom KJ, Zhao J, Seshiah P, et al. Development and preliminary evaluation of a multivariate index assay for ovarian cancer. *PLoS One*. 2009;4(2):e4599.
64. Sreekumar A, Nyati MK, Varambally S, Barrette TR, Ghosh D, Lawrence TS, et al. Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res*. 2001;61(20):7585-93.
65. Pelaez-Garcia A, Barderas R, Batlle R, Vinas-Castells R, Bartolome RA, Torres S, et al. A proteomic analysis reveals that Snail regulates the expression of the nuclear orphan receptor Nuclear Receptor Subfamily 2 Group F Member 6 (Nr2f6) and interleukin 17 (IL-17) to inhibit adipocyte differentiation. *Molecular & cellular proteomics : MCP*. 2015;14(2):303-15.
66. Diez P, Lorenzo S, Degano RM, Ibarrola N, Gonzalez-Gonzalez M, Nieto W, et al. Multipronged functional proteomics approaches for global identification of altered cell signalling pathways in B-cell chronic lymphocytic leukaemia. *Proteomics*. 2016;16(8):1193-203.
67. Belov L, Huang P, Barber N, Mulligan SP, Christopherson RI. Identification of repertoires of surface antigens on leukemias using an antibody microarray. *Proteomics*. 2003;3(11):2147-54.
68. Agoston DV, Gyorgy A, Eidelman O, Pollard HB. Proteomic biomarkers for blast neurotrauma: targeting cerebral edema, inflammation, and neuronal death cascades. *J Neurotrauma*. 2009;26(6):901-11.
69. Molero C, Rodriguez-Escudero I, Aleman A, Rotger R, Molina M, Cid VJ. Addressing the effects of Salmonella internalization in host cell signaling on a reverse-phase protein array. *Proteomics*. 2009;9(14):3652-65.
70. Li B, Jiang L, Song Q, Yang J, Chen Z, Guo Z, et al. Protein microarray for profiling antibody responses to Yersinia pestis live vaccine. *Infect Immun*. 2005;73(6):3734-9.
71. Thanawastien A, Montor WR, Labaer J, Mekalanos JJ, Yoon SS. Vibrio cholerae proteome-wide screen for immunostimulatory proteins identifies phosphatidylserine decarboxylase as a novel Toll-like receptor 4 agonist. *PLoS Pathog*. 2009;5(8):e1000556.
72. Manzano-Roman R, Diaz-Martin V, Gonzalez-Gonzalez M, Matarraz S, Alvarez-Prado AF, LaBaer J, et al. Self-assembled protein arrays from an Ornithodoros moubata salivary gland expression library. *J Proteome Res*. 2012;11(12):5972-82.
73. Anderson KS, LaBaer J. The sentinel within: exploiting the immune system for cancer biomarkers. *J Proteome Res*. 2005;4(4):1123-33.
74. Syed P, Gupta S, Choudhary S, Pandala NG, Atak A, Richharia A, et al. Autoantibody Profiling of Glioma Serum Samples to Identify Biomarkers Using Human Proteome Arrays. *Sci Rep*. 2015;5:13895.
75. Anderson KS, Cramer DW, Sibani S, Wallstrom G, Wong J, Park J, et al. Autoantibody signature for the serologic detection of ovarian cancer. *J Proteome Res*. 2015;14(1):578-86.
76. Anderson KS, Sibani S, Wallstrom G, Qiu J, Mendoza EA, Raphael J, et al. Protein microarray signature of autoantibody biomarkers for the early detection of breast cancer. *J Proteome Res*. 2011;10(1):85-96.
77. Gibson DS, Qiu J, Mendoza EA, Barker K, Rooney ME, LaBaer J. Circulating and synovial antibody profiling of juvenile arthritis patients by nucleic acid programmable protein arrays. *Arthritis Res Ther*. 2012;14(2):R77.
78. Henjes F, Lourido L, Ruiz-Romero C, Fernandez-Tajes J, Schwenk JM, Gonzalez-Gonzalez M, et al. Analysis of autoantibody profiles in osteoarthritis using comprehensive protein array concepts. *J Proteome Res*. 2014;13(11):5218-29.
79. San Segundo-Acosta P, Montero-Calle A, Fuentes M, Rabano A, Villalba M, Barderas R. Identification of Alzheimer's Disease Autoantibodies and Their Target Biomarkers by Phage Microarrays. *J Proteome Res*. 2019. doi: 10.1021/acs.jproteome.9b00258.



Proyecto Proteoma Humano (Human Proteome Project, HPP)

Fernando J. Corrales¹, Francisco Blanco² y Concha Gil³

¹Centro Nacional de Biotecnología (CNB)-CSIC. ProteoRed-ISCIII, CIBERehD. Madrid, Spain.

²INIBIC Hospital Universitario de A Coruña, ProteoRed-ISCIII, RIER-RED. A Coruña, Spain.

³Facultad de Farmacia, Universidad Complutense de Madrid, ProteoRed-ISCIII. Madrid, Spain.

1. Introducción histórica.

El conocimiento del proteoma humano supone un reto de una dimensión sin precedentes para la comunidad científica. Más allá de productos génicos que cabría esperar a partir de la traducción de los alrededor de 20.000 genes codificantes del genoma humano, procesos tales como el procesamiento del ARN o las modificaciones post-traduccionales (PTMs), entre otros, dan lugar a una gran diversidad de proteoformas (variantes de proteínas estables en las condiciones celulares), cuyo número se estima en cientos de miles (1,2). La identificación de todas las entidades peptídicas, su detección en los diferentes tipos celulares, la asignación de los cambios dinámicos de abundancia en respuesta a estímulos y la asignación de su implicación funcional, tanto en un entorno fisiológico como patológico son los objetivos últimos de la proteómica actual. La evolución de las técnicas de separación de proteínas y de espectrometría de masas, ha permitido diseñar estrategias para el estudio del proteoma con una gran precisión y sensibilidad, tales como las descritas en otros capítulos de este manual. Así, es posible hoy día realizar estudios prácticamente de proteomas completos (3) y de anotar miles de PTMs en un único experimento (4). A pesar de todo ello, el progreso en la investigación sobre el proteoma humano requiere de un continuo esfuerzo multidisciplinar que garantice el desarrollo de nuevos recursos tecnológicos y de nuevos métodos de análisis e integración de datos, la aplicación de los recursos generados en la investigación biomédica y clínica y, en último término, la identificación y cualificación de biomarcadores que contribuyan a establecer las bases de una medicina molecular diseñada a medida del paciente. En este contexto se fundó la HUPO en febrero de 2001 para coordinar en una iniciativa internacional a los diferentes proyectos ya operativos en varios países, promover la investigación en el ámbito de la proteómica y diseminar el conocimiento generado haciéndolo accesible a la comunidad científica en general (<http://www.HUPO.org>). Entre 2001 y 2002 se definieron las áreas estratégicas en las que centrar los objetivos: desarrollo tecnológico para llevar a cabo estudios más profundos y precisos, establecimiento de recursos informáticos para el procesado y análisis estadístico de los datos en formatos estandarizados y accesibles al resto de la comunidad científica, selección de tipos celulares y tejidos de



amplio interés para abordar su caracterización proteómica (cerebro, hígado, plasma/suero) y desarrollo de anticuerpos específicos para todas las proteínas codificadas por el genoma humano (5). Por entonces se comenzaba a plantear el abordaje del estudio sistemático del proteoma humano, incluso desde una perspectiva privada liderada por Celera (6). Desde el primer congreso de la HUPO celebrado en Versalles en noviembre de 2002, la actividad realizada se presenta y revisa en el congreso anual que se organiza de forma alternante en las 3 regiones geográficas que participan en HUPO, Europa/África, Asia/Oceanía y las Américas. Cabe destacar que este congreso fue organizado en España por miembros de la SEProt en 2014 (7). La actividad de las diferentes iniciativas (8) permitió establecer los primeros mapas del proteoma del plasma (9), del hígado (10), cerebro (11), así como comenzar a establecer estándares para facilitar la validación, comparación e intercambio de datos (12,13) y desarrollar colecciones masivas de anticuerpos de forma sistemática para establecer un atlas de expresión del proteoma humano(14,15). A pesar de que en los años siguientes se sumaron otras iniciativas centradas en el estudio del proteoma de tejidos específicos (riñón y orina) o enfermedades concretas (diabetes) (16), no fue hasta unos años más tarde, durante el Congreso HUPO celebrado en Amsterdam en 2008, cuando se comenzó a discutir la posibilidad de dotar de una estructura de proyecto único a la actividad que se desarrollaba en el seno de la HUPO y se discute el esquema del HPP (17,18), que se presentará oficialmente durante el congreso HUPO celebrado en Sidney en 2010 (19,20).

2. Objetivos y estructura del HPP.

Siguiendo con los principios inicialmente establecidos en la HUPO, el HPP tiene como misión la elaboración de un mapa completo de las proteínas humanas en su contexto biológico sobre el que construir modelos funcionales y determinar las variaciones en respuesta a situaciones tanto fisiológicas como patológicas que eventualmente serán la base para desarrollar nuevos métodos de diagnóstico, pronóstico y terapia. El proyecto descansa sobre tres pilares tecnológicos fundamentales: la espectrometría de masas (tanto *shotgun* como dirigida, Pilar MS), reactivos de afinidad (Pilar Ab), anticuerpos monoclonales y policlonales principalmente y recursos bioinformáticos (Pilar KB) (repositorios, bases de datos) que permitieran compartir la información generada en formatos estandarizados (19). Los objetivos que se plantearon en un inicio fueron:

1. La elaboración de un catálogo de ensayos robustos para detectar y cuantificar al menos un producto polipeptídico de cada gen codificante del genoma humano en base a las transiciones definidas para sus péptidos proteotípicos (específicos de una proteína y proteoforma concreta) en análisis de monitorización selectiva de iones (MRM/SRM, *Multiple/Selected Reaction Monitoring*).
2. Obtención de una librería de anticuerpos que permita detectar y cuantificar al menos un producto polipeptídico de cada gen codificante del genoma humano en los diferentes tipos celulares, tejidos y órganos, así como su localización subcelular, durante las etapas del desarrollo y edad adulta, tanto en situación basal como en condiciones patológicas.

3. Creación de una plataforma de conocimiento sobre proteínas humanas que integre la información estructural y funcional depositada en los recursos disponibles (UniProt/SwissProt, PRIDE, PeptideAtlas) y los resultados generados por las iniciativas del HPP bajo estrictos estándares de calidad previamente definidos.

Además de estos objetivos principalmente asociados a los pilares tecnológicos, se iniciaron los dos subproyectos principales del HPP: el centrado en el cromosoma (*Chromosome Centric* HPP, C-HPP) y el centrado en la biología y la enfermedad (*Biology and Disease driven* HPP, B/D-HPP). El C-HPP es un consorcio en el que 25 equipos de investigación internacionales adoptan un cromosoma específico para caracterizar las proteínas codificadas por los genes que lo integran (20), por ejemplo, en España se adoptó el cromosoma 16 (21). Aunque realmente esta distribución del trabajo no tiene una base biológica clara, sí permitió abordar de forma organizada la construcción del mapa del proteoma humano definiendo de forma sistemática las proteínas que lo integran. Por otra parte, las iniciativas ya existentes en HUPO centradas en tejidos o enfermedades concretas, se agruparon en el B/D-HPP con la idea de promover la investigación en sus áreas de conocimiento específicas aplicando la proteómica y en particular los recursos desarrollados en el HPP. A su vez, el HPP debía de constituirse como plataforma de interacción con la comunidad investigadora en general y con otros sectores como el biotecnológico, farmacéutico y clínico, para facilitar la transición del conocimiento generado hacia un beneficio neto para la sociedad.

3. Pilares tecnológicos.

Los tres pilares tecnológicos del HPP: MS, Ab y KB, tienen como misión aportar herramientas para el estudio del proteoma y métodos estandarizados para compartir la información generada. Como se detallará a continuación, el apartado más tecnológico del HPP integra tanto iniciativas que comenzaron en los primeros años de la HUPO, como nuevas colaboraciones que ya han generado valiosos recursos para la comunidad científica. El pilar MS ha sido desde el principio uno de los motores principales en el avance de la proteómica. La contribución de los desarrollos tecnológicos y la colaboración de muchas de las empresas protagonistas con el trabajo de la HUPO a través del IAB (*Industrial Advisory Board*), conjunto de empresas que participan en el consejo directivo de la HUPO, ha sido fundamental en el origen y diseño del HPP. Las primeras iniciativas de HUPO centradas en el estudio del proteoma de órganos y fluidos biológicos sirvieron de banco de pruebas para determinar las limitaciones de la metodología disponible e impulsar nuevos recursos analíticos más sensibles y precisos. Los análisis exploratorios del proteoma se basaban fundamentalmente en análisis dependientes de datos (DDA) y requirieron de la elaboración de una estrategia de estandarización que dotara de trazabilidad y reproducibilidad a los experimentos así como de la definición de estándares que facilitasen el intercambio de datos entre investigadores. Este ha sido el principal objetivo de la *Proteomics Standards Initiative* (PSI, <http://www.psidev.info>) desde su creación en el año 2002 (13) que ha definido y actualiza formatos estandarizados y guías en diferentes ámbitos, que incluyen: interacciones moleculares (MIMix, MIABA, MIAPAR, PSI-MI-XML, MITAB), espectrometría de masas e informática en proteómica (mzML, TraML, mzIdentML, mzQuantML, MIA-



PEs (*Minimum Information About a Proteomic Experiment*)) y separación de proteínas (MIAPEs, GelML, spML) (revisado en (22)). A su vez, se crea la *PRoteomics IDEntification Database* (PRIDE) como un repositorio público de datos de espectrometría de masas cuyo objetivo es representar todos los aspectos de un experimento de proteómica y hacerlos accesibles al usuario (<http://www.ebi.ac.uk/pride>) (23) en los formatos estandarizados desarrollados por la PSI. En la actualidad PRIDE contiene más de 10.000 proyectos y alrededor de 90.000 ensayos así como herramientas para la visualización y análisis de datos (PRIDE Inspector) y conversores a los formatos estándar más comunes (PRIDE Converter) (24). PRIDE es el principal repositorio de datos proteómicos basados en espectrometría de masas y se ha convertido en una referencia que proporciona a cualquier investigador la posibilidad de disponer de los datos depositados para desarrollar nuevas herramientas de análisis estadístico o reanalizarlos en el contexto de su propia investigación. PRIDE es uno de los miembros fundadores del ProteomeXchange (PX) consortium (<http://www.proteomexchange.org>) que se creó en 2011 con el objetivo de estandarizar el depósito de datos y la disseminación de los datos de proteómica a nivel mundial. Actualmente, además de los miembros fundadores (PRIDE and PeptideAtlas, incluido el recurso PASSEL) también se han unido al consorcio los siguientes repositorios: Massive, jPOST, iProX y Panorama (24). De forma análoga a otras tecnologías "ómicas", la deposición de los datos derivados de experimentos de MS/MS en repositorios públicos es un requisito exigido o recomendado por un número creciente de revistas científicas. Ante las limitaciones de los experimentos DDA (*shotgun*) para realizar cuantificaciones reproducibles en muestras de alta complejidad debido al elevado número de péptidos (selección del precursor más abundante, interferencias cruzadas...) se diseñaron estrategias dirigidas mediante monitorización selectiva de iones (*Selected Reaction Monitoring*, SRM). El desarrollo de ensayos SRM para los proteomas de *S cerevisiae* (25), *S pyogenes* (26) y (27) y sus aplicaciones (28), inspiró la puesta en marcha de una iniciativa cuyo objetivo fue el desarrollo de ensayos SRM para todo el proteoma humano. El resultado fue la generación de un conjunto de ensayos SRM específicos que permiten la cuantificación del 99,7% de las más de 20.000 proteínas humanas anotadas. Los resultados están recogidos en el SRMATlas (<http://www.srmatlas.org>) que contiene las coordenadas para identificar más de 166.000 péptidos proteotípicos y proporcionan múltiples ensayos independientes para cuantificar cualquier proteína humana, algunas variantes de procesamiento alternativo, mutaciones no sinónimas y modificaciones postraduccionales (29).

Para complementar la utilización de la espectrometría de masas en la caracterización del proteoma humano, se planteó el desarrollo de reactivos de afinidad a gran escala, esencialmente anticuerpos. Con este fin, el grupo de Mathias Uhlén propuso la generación de colecciones de anticuerpos para realizar estudios de proteómica de alta densidad a partir de fragmentos recombinantes de proteína (PrESTs, *Protein Epitopes Signature Tags*) a finales de los noventa. Esta estrategia se utiliza en 2002 para mapear las proteínas del cromosoma 21, como prueba de concepto (14) y más adelante da lugar al inicio del *Human Protein Atlas* (HPA, <http://www.proteinatlas.org>) en el que se describe la distribución específica de tejido y la localización subcelular de las proteínas humanas en situación normal y patológica (30). Desde entonces, se ha producido una continua evolución de esta iniciativa con una primera versión del atlas que contenía información de expresión de proteínas basada en aproximadamente 700 anticuerpos hasta la versión actual, la 18, con información sobre más de 18.000 proteínas y

26.000 anticuerpos. El HPA proporciona actualmente un valioso mapa de las proteínas humanas en células, tejidos y órganos integrando varias tecnologías, imagen basada en inmunoafinidad, espectrometría de masas y transcriptómica, en un formato abierto a cualquier usuario. Este recurso se subdivide en tres partes: 1: el Atlas de Tejidos o *Tissue Atlas* que integra datos de 76 tipos celulares pertenecientes a 44 tejidos (31), 2: el Atlas Celular o *Cell Atlas* (32) que determina la distribución subcelular de más de 12.000 proteínas en 33 orgánulos y estructuras celulares y 3: el Atlas de Patología o *Pathology Atlas* (33), que proporciona análisis de correlación de expresión génica y la evolución clínica de más de 8.000 pacientes de los tipos de cánceres más comunes y más de 5 millones de imágenes de inmunohistoquímica. El impacto científico del HPA puede deducirse de los miles de publicaciones a las que ha dado lugar, tanto en el entorno de la HUPO y del HPP como fuera de él.

Dentro del pilar KB, cabe destacar dos iniciativas: PeptideAtlas y neXtProt. PeptideAtlas (<http://www.peptideatlas.org>) fue diseñado para anotar genomas eucariotas con secuencias de péptidos derivadas de experimentos de espectrometría de masas (34). Los datos crudos generados por los espectrómetros de masas (se utilizan datos enviados directamente a PeptideAtlas, a PRIDE o a otros repositorios) se convierten a un formato estándar, mzXML y se realiza la asignación de secuencia en bases de datos y la validación automática utilizando la TPP (Trans Proteomic Pipeline; (35)). En la actualidad, tras el reprocesamiento de los archivos raw, el atlas humano contiene más de 220.000.000 de PSMs (*Peptide Spectrum Matches* o el número total de secuencias peptídicas identificadas para una proteína, incluyendo aquellas redundantemente identificadas) que corresponden a más de 1.400.000 péptidos únicos de 15.898 proteínas, identificadas de acuerdo a los criterios de aceptación establecidos por PeptideAtlas. Por su parte, NeXtProt (<http://nextprot.org>) nace con el objetivo de crear una plataforma de conocimiento en la red que contenga la información relativa a las proteínas humanas, tomando como punto de partida la información curada de UniProtKB/Swiss-Prot (36), que se integra con otras fuentes de información como PeptideAtlas, SRMATlas y HPA, entre otras (37). La última versión de neXtProt (03/09/2018) contiene información relativa a 20230 proteínas, 42241 proteoformas (splicing), más de 200.000 interacciones binarias, cerca de 200.000 modificaciones postraduccionales y más de 5 millones y medio de variantes. Además de PeptideAtlas y neXtProt, durante estos años se han desarrollado otras herramientas para recoger, ordenar y hacer disponible de forma sistemática la información generada por las diferentes iniciativas asociadas al HPP, entre las que cabe destacar, la *Global Proteome Machine* (GPM) (38), CAPER (39), GenomewidePDB (40) (41). Son, sin embargo, PeptideAtlas y neXtProt las iniciativas que están desempeñando un papel clave en el HPP ya que son las responsables de definir los umbrales de filtrado y certidumbre en relación a la detección de las proteínas del proteoma humano y sus variantes, curando la información que genera la comunidad científica (espectrometría de masas, inmunodetección, estructura, interacción, etc) de acuerdo a los criterios establecidos.

Uno de los objetivos fundamentales del HPP es completar la anotación del proteoma humano, habida cuenta de que existen proteínas para las que no disponemos de suficiente información experimental que avale su detección, las *missing proteins* (MPs). Tras las primeras clasificaciones (*gold, silver and bronze*), es en la versión de neXtProt de 2013 donde se establecen las diferentes categorías en relación al grado de evidencia de la existencia de las proteínas, que siguen vigentes en la actualidad: PE1,



observación a nivel de proteína, PE2, evidencia a nivel de transcrito, PE3, deducida a partir de datos de ortólogos, PE4 por predicción a partir del genoma y PE5, existen dudas a cerca de la existencia del gen codificante (42). Las MPs incluyen las categorías PE2-4, sin considerar PE5, cuya existencia es dudosa. Entre las razones que pueden explicar la dificultad de detectar las MPs se podrían considerar: los niveles bajos de expresión, expresión específica en algunos tipos celulares muy concretos (como por ejemplo los receptores olfatorios), de alguna etapa del desarrollo o en respuesta a determinados estímulos. En 2012 representaban alrededor del 32% del proteoma humano, cifra que ha ido modificándose en las sucesivas actualizaciones anuales de neXtProt hasta un 11% en la versión de 2018 (43), gracias a las contribuciones de las iniciativas del HPP (44) principalmente C-HPP, así como de trabajos de investigación de otros grupos. Para el desarrollo de este trabajo era esencial definir unas reglas básicas que asegurasen uniformidad de criterios y fiabilidad de la información anotada. Por ello, se elaboraron unas guías en el marco del HPP que indican el protocolo para enviar los datos a los repositorios de espectrometría de masas, como presentar el análisis del error experimental, y que información es necesaria para confirmar la observación de nuevas proteínas (45). Los resultados de espectrometría de masas han de filtrarse con un $FDR \leq 1\%$ (a nivel de PSM, péptido y proteína) y se considerarán identificaciones válidas aquellas que se basen en la asignación de al menos dos péptidos proteotípicos (permiten identificar inequívocamente una proteína) de nueve o más residuos. En relación a esto último, es relevante mencionar la herramienta de neXtProt *peptide uniqueness checker*, que permite de forma sencilla determinar si un péptido es proteotípico o no (46).

Ya en 2017, durante el XVI Congreso de la HUPO en Dublín, se propone la creación de un nuevo pilar dentro del esquema del HPP enfocado en patología (*Pathology Pillar*), de gran interés para muchas de las iniciativas del HPP. Sus objetivos principales son: 1: inventariar e incrementar las capacidades en este área dentro de HUPO, 2: iniciar un programa de formación, 3: elaborar guías de para la interpretación de los datos derivados del análisis histopatológico 4: establecer normas de buena práctica para biobancos, 5: construir librerías de perfiles proteómicos normales y patológicos conectados con repositorios de imágenes histológicas validadas, 6: construir un Atlas de Patología Humano (en colaboración con el HPA) y 7: iniciar colaboraciones con la industria.

4. HPP centrado en el cromosoma, C-HPP.

El C-HPP nace en Busan durante una reunión mantenida en 2011, con la misión de mapear y anotar todas las proteínas codificadas por cada uno de los genes del genoma humano. Inicialmente se organizó con una configuración en red en la que equipos internacionales adoptaban un cromosoma con la idea de caracterizar todos los productos proteicos codificados por sus genes (47). El principal objetivo que se aborda en todos los casos es la detección de la MPs para conseguir definir un mapa en el que al menos se describa un producto proteico para cada gen codificante (posteriormente se abordará la identificación de proteoformas). Para ello y aunque cada uno de estos consorcios internacionales adopta una estrategia diferente, en casi todos los casos

se adopta una estrategia que combina el análisis de experimentos de expresión génica masivos disponibles en repositorios públicos para determinar la expresión de los genes que codifican MPs en diferentes tejidos y tipos celulares, con experimentos de proteómica para verificar la expresión de las proteínas candidato (47-49). En España se creó el consorcio para el estudio del proteoma humano (SpHPP) que asumió la caracterización del cromosoma 16 (21) y en estos años, ha realizado aportaciones en relación a este cromosoma específico así como al desarrollo de recursos computacionales para definir los tejidos o líneas celulares en los que la probabilidad de detectar una determinada MP es máxima, entre los que cabe destacar testículo, esperma y cerebro (50,51). Asimismo, en paralelo a otros grupos, se configuraron nuevos sistemas de análisis que han permitido extraer información adicional a partir de los datos generados en experimentos de espectrometría de masas depositados en repositorios como PRIDE (52-54). Los progresos realizados se han publicado anualmente, desde 2013, en números temáticos del *Journal of Proteome Research*. En ellos, se describen las nuevas versiones de PeptideAtlas y neXtProt, los progresos en la identificación de MPs, nuevos métodos de análisis o flujos de trabajo, etc. En los últimos años, los objetivos del C-HPP se han diversificado para abarcar temas tan interesantes como el estudio de PTMs, variantes de splicing, SAPs (*Single Aminoacid Polymorphism*), smORFs (*small Open Reading Frame*) o la anotación funcional de proteínas PE1 (uPE1) (55). Todos estos aspectos tienen un claro trasfondo biológico lo que promueve el trabajo sinérgico con el B/D-HPP, algo lógico y ya previsto en los primeros esquemas del HPP (47).

5. HPP dirigido al estudio de la biología y enfermedad.

El objetivo general del B/D-HPP es explorar el impacto de las tecnologías comúnmente utilizadas en proteómica, como la espectrometría de masas, en áreas específicas de la biología, estableciendo colaboraciones con investigadores expertos. La complementariedad entre los dos subproyectos del HPP es evidente ya que el motor de esta iniciativa es la utilización de la descripción del proteoma elaborada por el C-HPP para descubrir las redes moleculares implicadas en la regulación fisiológica del organismo y las alteraciones que dan lugar a la progresión de enfermedades. Se trata de organizar el proteoma en los módulos funcionales que son la base de la complejidad de las células, tejidos y órganos del organismo (16). Con el propósito general de determinar si el acceso a técnicas de medición fiables y reproducibles facilitaría el abordaje de estudios más amplios del proteoma accesibles a investigadores no especialistas en proteómica el B/D-HPP plantea 4 objetivos específicos: 1: selección de las áreas biológicas/clínicas de interés, 2: generación de listas de proteínas de relevantes en cada una de las áreas definidas, 3: generación ensayos y reactivos para medir las proteínas diana y 4: divulgación del conocimiento adquirido y de los reactivos/métodos generados. En su inicio, el B/D-HPP integró a muchas de las iniciativas iniciales de la HUPO, mayoritariamente enfocadas en la descripción del proteoma de órganos y tejidos humanos, a las que se sumaron otras específicas de enfermedades como la diabetes o el cáncer. A día de hoy, el B/D-HPP cuenta con 19 iniciativas activas: cerebro, hígado,



plasma, riñón/orina, músculo esquelético, ojo, pediátrica, condiciones extremas, diabetes, cáncer, infecciosas, inmunopeptidómica, cardiovascular, mitocondria, organismos modelo, agregación de proteínas, nutrición y glicoproteómica, cuya descripción se puede consultar en <https://www.hupo.org/B/D-HPP>. Algunas de las iniciativas han generado listas de proteínas, disponibles en PeptideAtlas (<https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/proteinListSelector>), y se han desarrollado aplicaciones que permiten la consulta sistemática de bases de datos bibliográficas para extraer las proteínas (y metabolitos) más comúnmente asociados a una determinada situación biológica o enfermedad (56-59). Como se comentó anteriormente, SRMATlas y HPA proporcionan valiosos recursos complementarios para monitorizar la abundancia de las proteínas en muestras biológicas complejas.

6. Perspectivas.

El HPP se plantea como principal objetivo el impulsar el estudio del proteoma humano para conseguir su continua actualización incorporando el nuevo conocimiento generado a versiones curadas, accesibles a la comunidad científica. La anotación del "proteoma oscuro", en referencia a las MPs y a las proteínas sin anotación funcional (uPE1), es el principal reto y contará con la participación coordinada de las iniciativas anteriormente mencionadas (C-HPP, B/D-HPP y Pilares). Para ello se profundizará en la caracterización de tejidos/células específicos en los que estas proteínas puedan estar especialmente representadas, haciendo énfasis en condiciones concretas que puedan favorecer su expresión. En este sentido, los estudios de proteómica clínica, ya en curso en muchas de las iniciativas del B/D-HPP, juegan un papel esencial al ser fuente de tejidos patológicos, con programas de expresión específicamente adaptados. La caracterización funcional de estas proteínas y su implicación en enfermedades se realizará en el marco de los programas de investigación activos en cada uno de los laboratorios participantes.

Además de las modulaciones en la abundancia de las proteínas, su regulación mediante modificaciones postraduccionales es un aspecto íntimamente ligado al control de los procesos celulares, cuya alteración es parte de los mecanismos de progresión de muchas patologías. En este sentido, el HPP mantiene abiertas varias líneas de trabajo para la anotación curada de modificaciones en neXtProt y dos experimentos multicéntricos para analizar el alcance de diferentes procedimientos de análisis de fosfopéptidos (Pilar MS), por una parte y de glicoproteínas (Iniciativa B/D Glicoproteómica). La identificación de variantes de proteínas que incorporan sustituciones de aminoácidos, que son el resultado de procesamiento alternativo, de la traducción de marcos de lectura abierta pequeños (SMORF, de *small open reading frame*), e incluso péptidos derivados de RNAs no codificantes, es otro de los objetivos que ya se está abordando utilizando aproximaciones proteogenómicas.

Finalmente, además de hacer accesible el conocimiento generado, el HPP quiere seguir apoyando y promoviendo iniciativas que permitan el desarrollo de aplicaciones de interés clínico. La identificación de listas de proteínas relevantes (potenciales biomarcadores o dianas terapéuticas) en áreas concretas (*popular/priority proteins initiative*) y

la puesta a punto de métodos que permitan su cuantificación precisa y su desarrollo como herramientas aplicables en la rutina clínica, son objetivos prioritarios. Se requiere para ello de una colaboración interdisciplinar que asegure la orientación de los proyectos hacia las necesidades actuales de los pacientes. El recientemente creado Pilar de Patología va a desempeñar un papel decisivo en la creación de los puentes necesarios entre las diferentes iniciativas del HPP y así definir proyectos concretos claramente orientados a fines enmarcados en los objetivos de la medicina de precisión.

En conclusión, el HPP proporciona un marco ideal para desarrollar la actividad pluridisciplinar imprescindible para profundizar en el conocimiento del proteoma humano. Una de las prioridades es asegurar la disponibilidad de una versión curada en continua actualización del proteoma humano, fundamentada en los proyectos propios y en la incorporación de datos generados por la comunidad científica, previo filtro de acuerdo a las reglas de calidad ya establecidas. Este conocimiento y las herramientas tecnológicas que ya están en marcha, constituyen la base para promover la amplia utilización de la proteómica en investigación, algo que es ya una realidad, y dar los primeros pasos en la definición de fenotipos moleculares para la estratificación de pacientes.

REFERENCIAS

1. Smith LM, Kelleher NL. Proteoforms as the next proteomics currency. *Science*. 2018;359(6380):1106–7.
2. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, et al. How many human proteoforms are there? *Nat Chem Biol*. 2018;14(3):206–14.
3. Geiger T, Wehner A, Schaab C, Cox J, Mann M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics*. 2012;11(3):M111.014050.
4. Zagorac I, Fernandez-Gaitero S, Penning R, Post H, Bueno MJ, Mouron S, et al. In vivo phosphoproteomics reveals kinase activity profiles that predict treatment outcome in triple-negative breast cancer. *Nat Commun*; 2018; 9(1):3501.
5. Hanash S, Celis JE. The Human Proteome Organization: a mission to advance proteome knowledge. *Mol Cell Proteomics*. 2002;1(6):413–4.
6. Butler D. Celera in talks to launch private sector human proteome project. *Nature*; 2000 Feb 24;:815–6.
7. Gil C, Calvete JJ, Corrales FJ. The proteome quest to understand biology and disease (HUPO 2014). *J Proteomics*. 2015;127(Pt B):223–4.
8. Hanash S. HUPO initiatives relevant to clinical proteomics. *Mol Cell Proteomics*; 2004;3(4):298–301.
9. Omenn GS. Exploring the human plasma proteome. *Proteomics*. 2005; 5(13):3223–5.
10. He F. Human liver proteome project: plan, progress, and perspectives. *Mol Cell Proteomics*. 2005;4(12):1841–8.
11. Meyer HE, Klose J, Hamacher M. HBPP and the pursuit of standardisation. *Lancet Neurol*. 2003 Nov;2(11):657–8.
12. Orchard S, Zhu W, Julian RK, Hermjakob H, Apweiler R. Further advances in the development of a data interchange standard for proteomics data. *Proteomics*. 2003;3(10):2065–6.
13. Orchard S, Hermjakob H, Apweiler R. The proteomics standards initiative. *Proteomics*. 2003; 3(7):1374–6.
14. Agaton C, Galli J, Höidén-Guthenberg I, Janzon L, Hansson M, Asplund A, et al. Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. *Mol Cell Proteomics*. 2003; 2(6):405–14.
15. Nilsson P, Paavilainen L, Larsson K, Odling J, Sundberg M, Andersson A-C, et al. Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics*. 2005; 5(17):4327–37.
16. Aebersold R, Bader GD, Edwards AM, Van Eyk JE, Kussmann M, Qin J, et al. The biology/disease-driven human proteome project (B/D-HPP): enabling protein research for the life sciences community. *J Proteome Res*. 2013; 4;12(1):23–7.
17. Gavin A-C, Aebersold R, Heck AJR. Meeting report on the 7th World Congress of the Human Proteome Organization (HUPO) in Amsterdam: *Proteome Biology*. 2008. pp. 2288–91.
18. Baker MS. Building the “practical” human proteome project - the next big thing in basic and clinical proteomics. *Curr Opin Mol Ther*. 2009;11(6):600–2.
19. Legrain P, Aebersold R, Archakov A, Bairoch A, Bala K, Beretta L, et al. The human proteome project: Current state and future direction. *Mol Cell Proteomics*; 2011;mcp.O111.009993.



20. Hancock W, Omenn G, Legrain P, Paik Y-K. Proteomics, human proteome project, and chromosomes. *J Proteome Res.* 2011 7;10(1):210–0.
21. Segura V, Medina-Aunon JA, Guruceaga E, Gharbi SI, González-Tejedo C, Sánchez del Pino MM, et al. Spanish human proteome project: dissection of chromosome 16. *J Proteome Res.*; 2013; 4;12(1):112–22.
22. Deutsch EW, Orchard S, Binz P-A, Bittremieux W, Eisenacher M, Hermjakob H, et al. Proteomics Standards Initiative: Fifteen Years of Progress and Future Work. *J Proteome Res.*; 2017; 16(12):4288–98.
23. Jones P, Côté RG, Martens L, Quinn AF, Taylor CF, Derache W, et al. PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res.* 2006; 34(Database issue):D659–63.
24. Deutsch EW, Csordas A, Sun Z, Jarnuczak A, Perez-Riverol Y, Ternent T, et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* 2017; 45(D1):D1100–6.
25. Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell.* 2009; 138(4):795–806.
26. Karlsson C, Malmström L, Aebersold R, Malmström J. Proteome-wide selected reaction monitoring assays for the human pathogen *Streptococcus pyogenes*. *Nat Commun*; 2012;3(1):1301.
27. Schubert OT, Mouritsen J, Ludwig C, Röst HL, Rosenberger G, Arthur PK, et al. The Mtb proteome library: a resource of assays to quantify the complete proteome of *Mycobacterium tuberculosis*. *Cell Host Microbe.* 2013;13(5):602–12.
28. Ebhardt HA, Root A, Sander C, Aebersold R. Applications of targeted proteomics in systems biology and translational medicine. Lilley KS, Beynon RJ, Eyers CE, Hubbard SJ, editors. *Proteomics*. Wiley-Blackwell; 2015 Sep;15(18):3193–208.
29. Kusebauch U, Campbell DS, Deutsch EW, Chu CS, Spicer DA, Brusniak M-Y, et al. Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell.* 2016 Jul 28;166(3):766–78.
30. Uhlen M, Björling E, Agaton C, Szgyarto CA-K, Amini B, Andersen E, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics.* 2005 Dec;4(12):1920–32.
31. Uhlen M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015; 347(6220):1260419–9.
32. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. *Science.* 2017; 356(6340):eaal3321.
33. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. *Science.* 2017; 357(6352):eaan2507.
34. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, et al. The PeptideAtlas project. *Nucleic Acids Res.* 2006;34(Database issue):D655–8.
35. Keller A, Eng J, Zhang N, Li X-J, Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol.* 2005;1(1):2005.0017–E8.
36. Lane L, Argoud-Puy G, Britan A, Cusin I, Duek PD, Evalet O, et al. neXtProt: a knowledge platform for human proteins. *Nucleic Acids Res*; 2012; 40(Database issue):D76–83.
37. Gaudet P, Michel P-A, Zahn-Zabal M, Britan A, Cusin I, Domagalski M, et al. The neXtProt knowledgebase on human proteins: 2017 update. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D177–82.
38. Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res*; 2004;3(6):1234–42.
39. Guo F, Wang D, Liu Z, Lu L, Zhang W, Sun H, et al. CAPER: a chromosome-assembled human proteome browser. *J Proteome Res*; 2013;12(1):179–86.
40. Jeong S-K, Lee H-J, Na K, Cho J-Y, Lee MJ, Kwon J-Y, et al. GenomewidePDB, a proteomic database exploring the comprehensive protein parts list and transcriptome landscape in human chromosomes. *J Proteome Res*; 2013; 12(1):106–11.
41. Islam MT, Garg G, Hancock WS, Risk BA, Baker MS, Ranganathan S. Protannotator: A Semiautomated Pipeline for Chromosome-Wise Functional Annotation of the “Missing” Human Proteome. *J Proteome Res.* 2014 Jan 3;13(1):76–83.
42. Lane L, Bairoch A, Beavis RC, Deutsch EW, Gaudet P, Lundberg E, et al. Metrics for the Human Proteome Project 2013–2014 and Strategies for Finding Missing Proteins. *J Proteome Res.* 2014 Jan 3;13(1):15–20.
43. Omenn GS, Lane L, Overall CM, Corrales FJ, Schwenk JM, Paik Y-K, et al. Progress on Identifying and Characterizing the Human Proteome: 2018 Metrics from the HUPO Human Proteome Project. *J Proteome Res*; 2018; 17(12):4031–41.
44. Segura V, Garin-Muga A, Guruceaga E, Corrales FJ. Progress and pitfalls in finding the “missing proteins” from the human proteome map. *Expert Rev Proteomics.* 2017;14(1):9–14.
45. Deutsch EW, Overall CM, Van Eyk JE, Baker MS, Paik Y-K, Weintraub ST, et al. Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. *J Proteome Res.* 2016 Aug 24;acs.jproteome.6b00392–10.
46. Schaeffer M, Gateau A, Teixeira D, Michel P-A, Zahn-Zabal M, Lane L. The neXtProt peptide uniqueness checker: a tool for the proteomics community. Hancock J, editor. *Bioinformatics.* 2017;33(21):3471–2.
47. Paik Y-K, Omenn GS, Uhlen M, Hanash S, Marko-Varga G, Aebersold R, et al. Standard Guidelines for the Chromosome-Centric Human Proteome Project. *J Proteome Res.* 2012;11(4):2005–13.
48. Paik Y-K, Jeong S-K, Omenn GS, Uhlen M, Hanash S, Cho SY, et al. The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat Biotechnol.* 2012;30(3):221–3.
49. Paik Y-K, Omenn GS, Thongboonkerd V, Marko-Varga G, Hancock WS. Genome-wide proteomics, Chromosome-Centric Human Proteome Project (C-HPP), part II. *J Proteome Res*; 2014;13(1):1–4.

50. Segura V, Medina-Aunon JA, Mora MI, Martínez-Bartolomé S, Abian J, Aloria K, et al. Surfing transcriptomic landscapes. A step beyond the annotation of chromosome 16 proteome. *J Proteome Res*; 2014; 13(1):158–72.
51. Guruceaga E, Sanchez del Pino MM, Corrales FJ, Segura V. Prediction of a missing protein expression map in the context of the human proteome project. *J Proteome Res*. 2015;14(3):1350–60.
52. Carapito C, Lane L, Benama M, Opsomer A, Mouton-Barbosa E, Garrigues L, et al. Computational and Mass-Spectrometry-Based Workflow for the Discovery and Validation of Missing Human Proteins: Application to Chromosomes 2 and 14. *J Proteome Res*. 2015;14(9):3621–34.
53. Garin-Muga A, Odrizola L, Martínez-Val A, Del-Toro N, Martínez R, Molina M, et al. Detection of Missing Proteins Using the PRIDE Database as a Source of Mass Spectrometry Evidence. *J Proteome Res*. 2016; 15(11): 4101–4115
54. Guruceaga E, Garin-Muga A, Prieto G, Bejarano B, Marcilla M, Marín-Vicente C, et al. Enhanced Missing Proteins Detection in NCI60 Cell Lines Using an Integrative Search Engine Approach. *J Proteome Res*; 2017;16(12):4374–90.
55. Paik Y-K, Omenn GS, Hancock WS, Lane L, Overall CM. Advances in the Chromosome-Centric Human Proteome Project: looking to the future. *Expert Rev Proteomics*. 2017;14(12):1059–71.
56. Lam MPY, Venkatraman V, Xing Y, Lau E, Cao Q, Ng DCM, et al. Data-Driven Approach To Determine Popular Proteins for Targeted Proteomics Translation of Six Organ Systems. *J Proteome Res*. 2016;15(11):4126–34.
57. Lau E, Venkatraman V, Thomas CT, Wu JC, Van Eyk JE, Lam MPY. Identifying High-Priority Proteins Across the Human Diseasesome Using Semantic Similarity. *J Proteome Res*; 2018;17(12):4267–78.
58. Yu K-H, Levine DA, Zhang H, Chan DW, Zhang Z, Snyder M. Predicting Ovarian Cancer Patients' Clinical Response to Platinum-Based Chemotherapy by Their Tumor Proteomic Signatures. *J Proteome Res*. 2016;15(8):2455–65.
59. Yu K-H, Lee T-LM, Chen Y-J, Ré C, Kou SC, Chiang J-H, et al. A Cloud-Based Metabolite and Chemical Prioritization System for the Biology/Disease-Driven Human Proteome Project. *J Proteome Res*. 2018;17(12):4345–4357



Manual de Proteómica

Volumen II





SEProt

Sociedad Española
de Proteómica