

LA PROTEÓMICA, UN RETO CONSTANTE EN BIOMEDICINA

M^a Luisa Hernández Sánchez

Responsable Técnico Unidad de Proteómica UCM-PCM

Concha Gil García

Catedrática de la UCM. Responsable Científico Unidad de Proteómica UCM-PCM

PRESENTACIÓN

La Unidad de Proteómica es una Unidad de Desarrollo Tecnológico (UDT) de la Fundación Parque Científico de Madrid ubicada en la Facultad de Farmacia de la Universidad Complutense de Madrid (UCM). Esta Unidad se creó en el año 2001 como una ampliación del Centro de Asistencia a la Investigación de Secuenciación de DNA, dando origen al Centro de Genómica y Proteómica de la UCM. A finales de los años noventa nuestro grupo de investigación estaba participando en el Proyecto EUROFAN (Análisis funcional del genoma de la levadura). Uno de cuyos objetivos era la identificación de proteínas requeridas en la biogénesis de la pared celular. Dadas las características de nuestra muestra los procedimientos de identificación de proteínas disponibles (inmunodetección y secuenciación amino terminal) no resultaban adecuados.

Fue entonces cuando algunos grupos europeos estaban desarrollando otras estrategias para identificar proteínas basadas en la digestión enzimática o química de las proteínas y en el análisis de los péptidos mediante espectrometría de masas (MS). Utilizando dicha metodología en colaboración con el Dr. Blackstock (Glaxo-Wellcome) pudimos llevar a cabo nuestro objetivo, además de introducirnos en este apasionante campo de la proteómica. Por este motivo impulsamos la creación de la primera Unidad de Proteómica en una universidad española para dar apoyo a los investigadores en sus proyectos de Proteómica.

Para ello se precisó de equipamientos tecnológicos de alto coste que fueron financiados por el Ministerio de Ciencia y tecnología (Fondos FEDER y financiación para Parques Científicos) y de personal técnico cualificado (financiado por la UCM y el PCM). Este Centro ha contado desde el principio con un importante apoyo de la Cátedra de Genómica y Proteómica UCM-Merck-Sharp & Dohme, cuyo director es el Prof. César Nombela, y cuya misión es la de promocionar la docencia y la investigación en Genómica y Proteómica en la UCM.

Durante el año 2002, los servicios que desarrolla la Unidad de Proteómica de la UCM quedaron integrados dentro del Parque Científico de Madrid. En el año 2005 se crea el Instituto Nacional de Proteómica, ProteoRed, que es una plataforma de servicios proteómicos financiada por la Fundación Genoma España (2005-2010) y desde el año 2010 por el Instituto de Salud Carlos III. La red de Proteómica cuenta con seis nodos geográficos, uno de ellos con sede en Madrid, de la que es miembro la Unidad de Proteómica UCM-PCM. El objetivo principal de ProteoRed es coordinar y desarrollar Servicios ya existentes, ofreciendo esta plataforma a los grupos de investigación españoles y fomentando el desarrollo de nuevas aplicaciones tecnológicas y la estandarización y perfeccionamiento de las técnicas existentes en los 21 laboratorios que integran este proyecto.

Las principales funciones son: la implementación tecnológica, las recomendaciones sobre el manejo y almacenamiento de muestras, la promoción de experimentos multicentro, el soporte bioinformático, el establecimiento de los precios y la organización funcional, la promoción y difusión

de las técnicas de proteómica en general mediante cursos y seminarios, la internacionalización y conexión con plataformas europeas y, finalmente, la evaluación de las nuevas técnicas biomédicas.

La Unidad de Proteómica UCM-PCM tiene como objetivo principal fomentar la investigación científica y el desarrollo tecnológico en el área de la Proteómica, ofreciendo a los grupos de investigación, hospitales y empresas diferentes servicios de proteómica, asesoramiento en sus proyectos, así como seminarios y cursos de formación. El personal de la Unidad de Proteómica tiene como objetivos fundamentales: dar servicio de proteómica a la comunidad científica y desarrollar y poner a punto nuevas metodologías que se están desarrollando en dicho campo. Además participa en cursos de técnicas proteómicas organizados por diferentes instituciones. Por supuesto, el personal de la Unidad está implicado activamente en los diferentes grupos de trabajo de ProteoRed, además son miembros fundadores de la SeProt, una asociación de carácter multidisciplinar, cuyo objetivo es impulsar el desarrollo de la Proteómica en España y han formado parte del Comité organizador del IV Congreso de Proteómica (New Trends in Proteomics, Segovia, 2011). Además la directora científica forma parte del Comité directivo de la EuPA (European Proteomics Association), ha coordinado el Comité de Educación y actualmente coordina el Comité de Congresos y Comunicación.

¿QUÉ ES LA PROTEÓMICA?

El término “proteoma” fue utilizado por primera vez en 1995 por Mark Wilkins, un científico Australiano, para describir el conjunto de PROTEÍNAS expresadas por un genOMA, una célula o un tejido. Así, surgió la disciplina denominada Proteómica, que consiste en el estudio a gran escala del proteoma con el fin de obtener una visión global e integrada de los procesos celulares.

La proteómica proporciona un conjunto de herramientas muy poderosas para el estudio a gran escala de la función de las proteínas, de su estructura, localización, modificaciones postraduccionales e interacciones. Hubo dos factores decisivos para el desarrollo de la proteómica:

1. Por un lado la secuenciación de los genomas a gran escala (se conoce la secuencia de los genes pero no su función) y su anotación en las bases de datos de secuencias de proteínas.
2. Y por otro lado, el desarrollo de técnicas de separación y análisis de proteínas (Electroforesis Bidimensional y Espectrometría de Masas que más adelante se detallarán).

La secuenciación del genoma humano ha permitido conocer el número de genes que poseemos y que dicho número no es muy diferente al de otros organismos. La complejidad de los organismos parece radicar en las proteínas, ya que la información que está contenida en los genes se traduce en proteínas, que son las que llevan a cabo las funciones del organismo. A diferencia del genoma, que es más o menos constante, el proteoma es altamente dinámico, ya que en distintos tipos de células se expresan genes diferentes y además las proteínas varían dependiendo de las condiciones del entorno donde se encuentren, de determinadas señales bioquímicas, del estado fisiológico de las células, por lo que en cada momento y para cada tipo de célula, el proteoma será diferente.

Además, la complejidad del estudio del proteoma reside en que un gen puede dar lugar a diferentes formas proteicas (“splicing” o procesamiento alternativo). Además las proteínas van a sufrir diferentes modificaciones post-traduccionales, existiendo más de 200 (fosforilación, acetilación, glicosilación, desaminación, miristoilación...). Ambas fuentes de complejidad incrementan sustancialmente el número de proteínas diferentes que pueden existir: se estima que a partir de los 20.000 genes que codifican proteínas en el genoma humano, podrían generarse un número de proteínas diferentes que oscilaría entre 500.000-1.000.000.

Así la era post-genómica ha dado lugar al desarrollo de la Proteómica, que algunos investigadores la incluyen dentro de la genómica funcional junto con otras disciplinas “ómicas” como la transcriptómica y la metabolómica.

METODOLOGÍAS PROTEÓMICAS

La proteómica surgió como consecuencia de los avances en diferentes técnicas de separación de proteínas, principalmente de las basadas en gel y por la aplicación de la espectrometría de masas a la caracterización de las proteínas.

Vamos a comentar brevemente las técnicas proteómicas más usadas.

La tecnología más utilizada para la separación de proteínas es la electroforesis en geles de poliacrilamida. Para muchas aplicaciones proteómicas, la electroforesis en una dimensión es el método de elección. Las proteínas se separan de acuerdo a su masa y las proteínas son solubilizadas en dodecil sulfato sódico (SDS). Es una técnica sencilla, reproducible y permite la separación de proteínas de 10-300 kDa. Una de las aplicaciones más comunes de la 1-DE es la caracterización de proteínas después de realizar algún tipo de purificación previa.

Uno de los pilares de la proteómica fue el desarrollo de la electroforesis bidimensional 2D-PAGE, que permite separar hasta miles de proteínas en un sólo experimento, y está basada en una separación de las proteínas en función de la carga, seguida de una separación de las proteínas en función de su masa molecular. La separación de la primera dimensión se realiza mediante isoelectroenfoque, durante el cual las proteínas son separadas en un gradiente de pH hasta alcanzar una posición en la que su carga neta es cero, es decir, su punto isoeléctrico. En una segunda dimensión, las proteínas son separadas mediante electroforesis en presencia de SDS. La alta resolución de la técnica se debe a que las dos separaciones se basan en parámetros independientes. En cuanto a la detección de las proteínas, tradicionalmente se ha venido empleando la tinción con azul de Coomassie, o con plata, o tinciones fluorescentes para conseguir mayor sensibilidad.

La principal herramienta de la investigación proteómica es la Espectrometría de Masas (EM), una técnica analítica que mide la relación masa/carga de una molécula ionizada. A finales de los años ochenta, los científicos Tanaka y J. Fenn desarrollaron dos nuevos métodos de ionización suave de macromoléculas, MALDI, del inglés *Matrix Assisted Laser Desorption Ionization* y ESI del inglés *electrospray ionization*. El método de ionización por electrospray permite la ionización de moléculas a partir de una solución acuosa, al hacerla pasar por un capilar metálico bajo la aplicación de alto voltaje, mientras que la ionización por MALDI produce iones mediante pulsos de láser aplicados a muestras en estado sólido embebidas en matrices cristalizables.

Estas metodologías solucionaron la dificultad de generar iones a partir de macromoléculas polares y no volátiles, como las proteínas y péptidos, sin que se fragmentaran. Desde entonces, la EM es la técnica de elección para la identificación de proteínas, debido a su sensibilidad (pmol-fmol), exactitud (100-5ppm) y rapidez (minutos-segundos), variando de acuerdo con la complejidad del análisis y el equipo utilizado. Los espectrómetros de masas constan de tres componentes fundamentales: la fuente de ionización, el analizador y el detector. De la combinación de estos elementos dependen los diferentes tipos de espectrómetros de masas y sus aplicaciones.

El analizador de masas sirve para separar y analizar los iones según su masa/carga, que se han formado en la fuente de ionización. Hay diferentes tipos de analizadores, TOF (Tiempo de vuelo, en inglés, *Time of Flight*), cuadrupolo y trampa de iones, que van desde el más sencillo, compuesto por fuente de iones tipo MALDI y analizador tipo tiempo de vuelo, MALDI-TOF, hasta los espectrómetros de masas en *tandem*, donde se combinan varios analizadores en el espacio (TOF/TOF, triple cuadrupolo, cuadrupolo-TOF), o en el tiempo como la trampa de iones.

En la identificación de proteínas existen dos flujos de trabajo independientes:

- a) Basado en gel.
- b) En solución o libre de gel.

a) El más clásico, común y sencillo es el basado en gel, o procedente de la *escuela europea*. Consiste en recortar las manchas de proteínas directamente del gel bidimensional donde han sido aisladas, y se digieren con una proteasa, generalmente tripsina, que produce un conjunto de péptidos. Estos péptidos se analizan mediante espectrometría de masas tipo MALDI-TOF, y se obtiene el espectro de masas del conjunto de péptidos, también denominado “huella peptídica”. Las huellas peptídicas son características y específicas de cada una de las proteínas, y permiten identificarlas una a una las proteínas por comparación de la huella peptídica experimental con las huellas peptídicas teóricas que hay en las bases de datos de proteínas, mediante técnicas bioinformáticas, siempre y cuando la secuencia de la proteína esté depositada en una base de datos de acceso público.

Dado que el criterio más determinante para la identificación de una proteína es su secuencia aminoacídica, y debido a que a veces las proteínas no están anotadas en las bases de datos, en muchas ocasiones se recurre a la obtención de espectros de fragmentación (MS/MS). Consiste en seleccionar un ión peptídico en un primer analizador, se fragmenta mediante colisión y en un segundo analizador se separan las masas de los fragmentos derivados del péptido. De la misma manera, que la huella peptídica, los espectros MS/MS son también característicos de cada uno de los péptidos, y permiten su identificación inequívoca, buscando y comparando en las bases de datos. Un estudio en detalle de los espectros MS/MS nos permiten determinar sitios de modificaciones post-traduccionales, e incluso identificar proteínas de organismos cuyo genoma es aún desconocido, elucidando su secuencia de aminoácidos (secuenciación *de novo*).

b) El flujo de trabajo, libre de gel, o procedente de la *escuela americana*, se basa en el análisis de una mezcla de proteínas más o menos compleja, mediante la digestión con tripsina en solución del conjunto de proteínas. Los péptidos resultantes se separan mediante cromatografía de fase reversa según la hidrofobicidad y/o por cromatografía de intercambio iónico, según la carga de la molécula. Los péptidos así separados pueden ser analizados directamente mediante espectrometría de masas *en tandem, en tiempo real*.

En los estudios de proteómica cuantitativa, se pueden comparar la expresión de proteínas de varias muestras de tejido, cultivo, etc., para detectar variaciones en el nivel de expresión de cada una de las proteínas de las diferentes condiciones y su posterior identificación.

La aproximación de proteómica cuantitativa en gel se realiza mediante la tecnología DIGE (Difference Gel Electrophoresis) que se basa en el marcaje de tres muestras a comparar con tres fluorocromos diferentes y analizarlos en un mismo gel bidimensional para disminuir las variaciones técnicas.

En las aproximaciones de proteómica cuantitativa libre de gel, las muestras a comparar se pueden marcar diferencialmente con marcaje metabólico de los cultivos, SILAC (Stable Isotope Label Aminoacids), o mediante marcaje isobárico diferencial de los péptidos de la digestión de una mezcla de proteínas con un reactivo químico (iTRAQ).

El análisis y la interpretación de los datos obtenidos con las técnicas de proteómica descritas anteriormente, especialmente en los experimentos a gran escala, se necesitan herramientas bioinformáticas y han sido y son esenciales para el desarrollo y auge de la proteómica.

Como hemos comentado, no existe un diagrama de flujo único para el análisis proteómico, por lo que queda a criterio del investigador la elección de la aproximación y el conjunto de técnicas más adecuado para abordar sus estudios.

APROXIMACIONES DE LA PROTEÓMICA

La Proteómica, sin embargo, va más allá de la mera identificación de las proteínas. Con los experimentos de Proteómica se pueden abordar diferentes estudios proteómicos como se indican a continuación:

- A) Identificación de las proteínas que componen un proteoma de forma masiva para obtener una visión global e integrada del proteoma, así como la realización de mapas de referencia.
- B) Realización de estudios de proteómica cuantitativa. Consisten en comparar la abundancia relativa de las proteínas entre diferentes muestras. Por ejemplo, se pueden comparar proteomas expuestos a diversos agentes externos (tratado y sin tratar con una determinada droga), en determinadas condiciones fisiológicas (ej. expuestos a diferentes tipos de estrés), o patológicas (un tejido sano y uno enfermo), etc. De esta manera se pueden identificar proteínas de expresión diferencial y elucidar qué procesos biológicos están implicados.
- C) Estudios de la interacción de proteínas con otras proteínas o moléculas con el fin de determinar su función, ya que las proteínas no actúan de forma aislada sino formando complejos macromoleculares y redes funcionales.
- D) Estudio de las modificaciones postraduccionales que sufren las proteínas, como la fosforilación (fosfoproteómica), glicosilación (glicoproteómica), etc. Dichas modificaciones sirven para modular la actividad y/o localización de una proteína.
- E) Diseño de estudios de proteómica dirigida (“targeted proteomics”) para identificar y cuantificar una proteína de interés en muestras biológicas complejas.

APLICACIONES DE LA PROTEÓMICA

El desarrollo de la Proteómica ha despertado grandes expectativas en diferentes áreas, como en biotecnología, industria agroalimentaria y especialmente en biomedicina. La identificación de las proteínas que intervienen en las diferentes etapas de la enfermedad ayudará a comprender las bases moleculares de dichas patologías, y así estas proteínas identificadas y cuantificadas se podrán utilizar como biomarcadores de diagnóstico o pronóstico de la enfermedad y nuevas dianas terapéuticas.

Los biomarcadores son moléculas que sirven como indicadores del estado fisiológico o patológico de las células o tejidos objeto de estudio y cuyos requisitos fundamentales que deben cumplir son una elevada especificidad y sensibilidad.

El desarrollo de la proteómica ha abierto grandes expectativas para la identificación de biomarcadores, ya que la espectrometría de masas permite identificar proteínas en muy baja concentración y puede realizarse un análisis sistemático de cientos o miles de proteínas en una muestra clínica. Sin embargo, debido a la complejidad de estas muestras, que consisten en fluidos biológicos humanos (sangre, plasma y suero, líquido cefalorraquídeo, orina, saliva, cristalino, etc.) que son la principal fuente de biomarcadores y a la baja concentración de proteínas que pueden actuar como biomarcadores (ng/ml), se deben tener en cuenta algunas consideraciones en la puesta a punto de la preparación de la muestra.

Desde el punto de vista metodológico el problema a resolver consiste en que en los tejidos y especialmente en el suero, las proteínas poco abundantes, que son las más significativas, se encuentran enmascaradas por unas pocas (pocos cientos) proteínas muy abundantes, por lo que es necesario un primer prefraccionamiento, depleción de las proteínas más abundantes, cuantificación de las proteínas, con la finalidad de que los resultados sean comparables y reproducibles. Además, este tipo de muestra

en algunos casos es de difícil recolección y son muestras escasas y muy valiosas ya que proceden de pacientes o biobancos bien caracterizados.

En general, las enfermedades no se deben a alteraciones de una única proteína, sino de muchas proteínas y a la variación en su abundancia. Por estos motivos la proteómica aplicada a la biomedicina no sólo consiste en la identificación del mayor número posible de proteínas en una muestra, sino estudios de proteómica cuantitativa de las proteínas entre muestras clínicas diferentes. Estos estudios son complejos y necesitan de una posterior validación, ya que las enfermedades suelen tener una naturaleza multifactorial, y por tanto no se trataría de un único biomarcador, sino de un conjunto de marcadores.

Participación de la Unidad en diferentes Proyectos

A continuación se enumeran algunas áreas en las que la proteómica tiene grandes impactos y en los proyectos en los que participa la Unidad de Proteómica, dando apoyo tanto técnico como de asesoría científica a los investigadores:

- Estudio de nuevos biomarcadores de diagnóstico y pronóstico en diferentes enfermedades: cardiovasculares, artritis reumatoide, cáncer y obesidad.
- Estudio de nuevos biomarcadores de diagnóstico y pronóstico en enfermedades metabólicas y autoinmunes.
- Identificación de nuevos alérgenos y proteínas candidatas para el desarrollo de vacunas.
- Identificación de antígenos útiles para el diagnóstico y prevención de enfermedades infecciosas y búsqueda de dianas para el desarrollo de nuevos antifúngicos y otros fármacos.
- Estudios de proteómica aplicados a neurología y psiquiatría, autismo, y alzhéimer.
- Estudios diferenciales de familias y especies de animales.
- Mejora de plantas y alimentos.

La Unidad de Proteómica ha prestado servicio a más de 250 investigadores, tanto del ámbito público como privado, en las áreas anteriormente descritas.

FUTURO DE LA PROTEÓMICA

En la X Conferencia Anual de la Organización del Proteoma Humano (HUPO, en sus siglas en inglés), celebrada en Sydney en septiembre de 2010, se lanzó el Proyecto Proteoma Humano (HPP). El objetivo del proyecto es la caracterización de las proteínas codificadas por los 21.000 genes que constituyen el genoma humano. Este estudio generará un mapa de proteínas y ayudará a elucidar las funciones moleculares y biológicas y supondrá un avance en el diagnóstico y tratamiento de las enfermedades. Este avance científico, comparable en su momento a la investigación sobre el genoma, permitirá a los científicos descubrir y analizar cualquier proteína humana en cualquier muestra biológica.

Para llevar a cabo la organización del trabajo, la tarea se ha segmentado por cromosomas, de forma que cada grupo de países se centra en las proteínas codificadas por todos los genes agrupados en diferentes cromosomas. Los grupos españoles, incluida la Unidad de Proteómica UCM-PCM, estudiarán, junto con equipos suecos y noruegos, las proteínas del cromosoma 19, al que están ligadas enfermedades como el Alzheimer, el cáncer de ovario, de páncreas, de mama, de pulmón, los gliomas o la leucemia, entre otras múltiples patologías. Además todos los conocimientos adquiridos durante los últimos años se integrarán en el desarrollo del Proyecto HPP. El trabajo es de gran magnitud y será necesario un enorme desarrollo tecnológico para llevarlo a cabo.

Respecto al futuro, una vez que conozcamos todas las proteínas humanas, podremos definir cuáles de ellas se alteran y participan en el desarrollo de una enfermedad. Esto significa que, en unos

años, se podrán definir biomarcadores de utilidad para conocer quién padece una enfermedad (diagnóstico precoz), cómo será su pronóstico y si responderá o no a un tratamiento determinado para elegir el tratamiento más adecuado en cada caso. En este sentido, la proteómica representa un paso más en los estudios dirigidos a la medicina individualizada o personalizada, en la que se tienen en cuenta las respuestas de cada paciente ante distintas terapias.

ENLACES DE INTERÉS

Página de la Fundación Parque Científico de Madrid-Unidad de Proteómica:

<http://www.fpcm.es/srvAProteomica.htm>

Página de la Universidad Complutense de Madrid-Unidad de Proteómica:

<http://www.ucm.es/info/gyp/proteomica/>

Página de la Sociedad Española de Proteómica: <http://www.fpcm.es/srvAProteomica.htm>

Página de la Plataforma en Red de Proteómica (ProteoRed): <http://www.proteored.org/>

Página de la Asociación Europea de Proteómica (EuPA): <http://www.eupa.org/>

Página de acceso a la información del Proyecto del Proteoma Humano(HPP):

<http://www.hupo.org/research/hpp/>