

La metodología proteómica, una herramienta para la búsqueda de función

Concha Gil García

Departamento de Microbiología II, Facultad de Farmacia, Universidad Complutense de Madrid
E-mail: conchagil@farm.ucm.es.

Nuestro interés por la proteómica.

Antes de comenzar a hablar de "proteómica" y de su enorme potencial en el campo de la biología, voy a comentar brevemente cómo nuestro grupo entro en contacto con esta tecnología. Hace siete años, el grupo del Profesor César Nombela del que formo parte, estaba llevando a cabo un proyecto de investigación que requería la utilización de unas técnicas de análisis de proteínas muy novedosas, y que desafortunadamente en ese momento no disponía ningún laboratorio en España. Participábamos, junto con varios grupos europeos, en el Proyecto EUROFAN (Análisis funcional del genoma de la levadura *Saccharomyces cerevisiae*). El genoma de la levadura acababa de ser secuenciado y se disponía de una gran información; sin embargo, muchos genes eran de función desconocida. Al objeto de estudiar la función de estos genes se establecieron una serie de nodos integrados por diferentes grupos de investigación especializados. Nuestro grupo estaba en el nodo de pared celular y morfogénesis. El estudio de la pared celular presenta un enorme interés como diana de agentes antifúngicos, ya que dicha estructura está ausente en la célula animal. Nuestro objetivo era realizar un *screening* con aproximadamente mil mutantes de delección de *S. cerevisiae* al objeto de identificar genes implicados con la formación de la pared celular de la levadura. Para ello se propusieron diversos análisis fenotípicos relacionados con defectos en la pared celular con diferente grado de complejidad. Uno de los ensayos propuestos fue la identificación de proteínas secretadas por protoplastos en regeneración, al objeto de tener una visión global e integrada de todas la proteínas que son requeridas en la biogénesis de dicha estructura. De esta forma obtendríamos un mapa de referencia para después poder realizar estudios comparativos con mutantes de pared celular, es decir, para estudiar la expresión diferencial de las proteínas de pared. En primer lugar, obtuvimos los productos de secreción de protoplastos y las proteínas se separaron mediante electroforesis bidimensional (2D-PAGE) utilizando gradientes de pH inmovilizados (IPGs) que permiten obtener una gran reproducibilidad y gran capacidad de carga de proteínas en el gel.

Finalmente, procedimos a la caracterización de las proteínas mediante dos aproximaciones: inmunodetección y secuenciación amino-terminal o degradación de Edman. Ambos procedimientos son bastante limitados, para la inmunodetección se requiere disponer de anticuerpos y para la secuenciación de Edman es preciso obtener gran cantidad de proteína y transferirla a una membrana. Dadas las características de nuestra muestra, estos procedimientos 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). Mediante dicha estrategia se requería una menor cantidad de proteína (manchas proteicas teñidas con plata) y se podía digerir la proteína directamente en gel de poliacrilamida para la obtención de los péptidos. 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 (Pardo *et al.*, 2000).

Se puede decir que hubo 3 factores decisivos para el desarrollo de la proteómica:

- La secuenciación de genomas a gran escala y el desarrollo de bases de datos de proteínas.
- El desarrollo de técnicas de espectrometría de masas para analizar proteínas y péptidos.
- Los avances realizados en la separación de proteínas mediante 2D-PAGE con la introducción de los gradientes de pH inmovilizados (IPGs).

Esto fue el comienzo y gracias a la conjunción de todo ello se inició lo que hoy llamamos proteómica. Hace veinte años esto era impensable y los investigadores nos encontrábamos siempre con el cuello de botella de la identificación de las proteínas. Recuerdo varios proyectos, de enorme interés, que no pudieron desarrollarse hasta el final debido a esta limitación. Creo que algunas de estas ideas y proyectos se podrían retomar y sacarles muchísimo partido utilizando esta tecnología.

En esta pequeña revisión sobre tecnología proteómica pretendo solamente resaltar algunos aspectos de interés especialmente de la separación de proteínas mediante electroforesis y del

análisis de éstas mediante espectrometría de masas (*MS*) señalando las aproximaciones para la identificación y caracterización de las proteínas. También comentaré nuevas estrategias para el análisis del proteoma. Para mayor información se pueden consultar diversas revisiones que han aparecido durante los últimos años (Banks *et al.*, 1999; Blackstock y Man, 2001; Aebersold y Cravatt, 2002; Graves y Haystead, 2002).

Definición y tipos de proteómica

Los rápidos avances conseguidos en la tecnología del DNA están permitiendo la secuenciación sistemática de los genomas de diversos organismos. Actualmente se ha completado y publicado la secuencia de 140 genomas, que pertenecen a 16 arqueas, a 106 bacterias y a 18 eucariotas (en dos, sólo un cromosoma), y existen más de 717 proyectos de secuenciación en marcha (<http://wit.integratedgenomics.com/GOLD/>; última actualización 8 de Mayo de 2003). Recientemente hemos asistido a un hito histórico, la secuenciación del genoma humano completo.

Todos estos proyectos de secuenciación a gran escala están proporcionando una ingente cantidad de secuencias de DNA. Sin embargo, aún se desconoce la función biológica de la mayoría de las proteínas codificadas por los genes detectados. Así pues, el siguiente paso en la era post-genómica debe ser el estudio funcional de todos estos genes. La proteómica es uno de los campos que puede ayudar a establecer una conexión entre las secuencias genómicas y su comportamiento biológico, constituyendo una herramienta importante en el análisis funcional de genes de función desconocida.

El término "proteoma" fue usado por vez primera en 1995 para describir el conjunto de PROTEÍNAS de un genOMA, una célula o un tejido. De forma imperceptible, la palabra proteoma dio lugar a una nueva disciplina, la "proteómica". Pero, ¿qué es la proteómica? Esencialmente la proteómica es el estudio a gran escala de los productos génicos de un genoma mediante métodos bioquímicos, con el fin de obtener una visión global e integrada de los procesos celulares. El término proteómica se ha asociado tradicionalmente con la separación de un gran número de proteínas de una célula u organismo mediante 2D-PAGE. Según esto, la proteómica comenzó en los años setenta cuando se empezaron a construir bases de datos de proteínas utilizando la electroforesis bidimensional. Sin embargo, la identificación de las proteínas era difícil debido a la falta de métodos

analíticos rápidos y sensibles para la caracterización de proteínas. En los años noventa, la espectrometría de masas surge como un método analítico muy poderoso, ya que elimina la mayoría de las limitaciones del análisis de proteínas. Este desarrollo, junto con la disponibilidad de los genomas secuenciados marca el comienzo de una nueva era. Actualmente muchas áreas de estudio han sido agrupadas dentro de la proteómica. Se pueden incluir, entre otros, los estudios de interacciones de proteínas, de modificaciones pos-traduccionales, el análisis funcional de proteínas y estudios de localización.

Se puede hablar de dos tipos de proteómica: proteómica de expresión y proteómica del mapa celular.

La **proteómica de expresión** es el estudio cuantitativo de la expresión de proteínas entre muestras que difieren en alguna variable. En esta estrategia se compara la expresión del proteoma total o de subproteomas entre diferentes muestras. La información obtenida puede permitir la identificación de nuevas proteínas implicadas en transducción de señales, la identificación de proteínas específicas de una enfermedad y proteínas de interés en microbiología médica.

La **proteómica del mapa celular o estructural** es el estudio de la localización subcelular de las proteínas y de las interacciones proteína-proteína mediante la purificación de orgánulos o complejos y la posterior identificación de sus componentes mediante espectrometría de masas.

También se utiliza el término de **proteómica funcional** para referirse a diversas aproximaciones proteómicas que permiten el estudio y caracterización de un grupo de proteínas determinado proporcionando información importante sobre señalización, mecanismos de la enfermedad o interacciones proteína-fármaco.

Para poder desarrollar los ambiciosos objetivos de la proteómica se requiere la implicación de diversas disciplinas como biología molecular, bioquímica, microbiología y bioinformática. Siendo esta última de enorme importancia, ya que será necesario desarrollar ordenadores con una gran capacidad para organizar la gran cantidad de información que se generará en estas investigaciones. Además, la proteómica complementa otras aproximaciones genómicas funcionales, como los perfiles de expresión utilizando DNA-*arrays*, perfiles fenotípicos sistemáticos, genética sistemática y *arrays* basados en moléculas pequeñas.

Para caracterizar un proteoma de una célula es importante tener en cuenta que el proteoma es dinámico y que es reflejo del medio ambiente en el que es estudiado. Como respuesta a estímulos

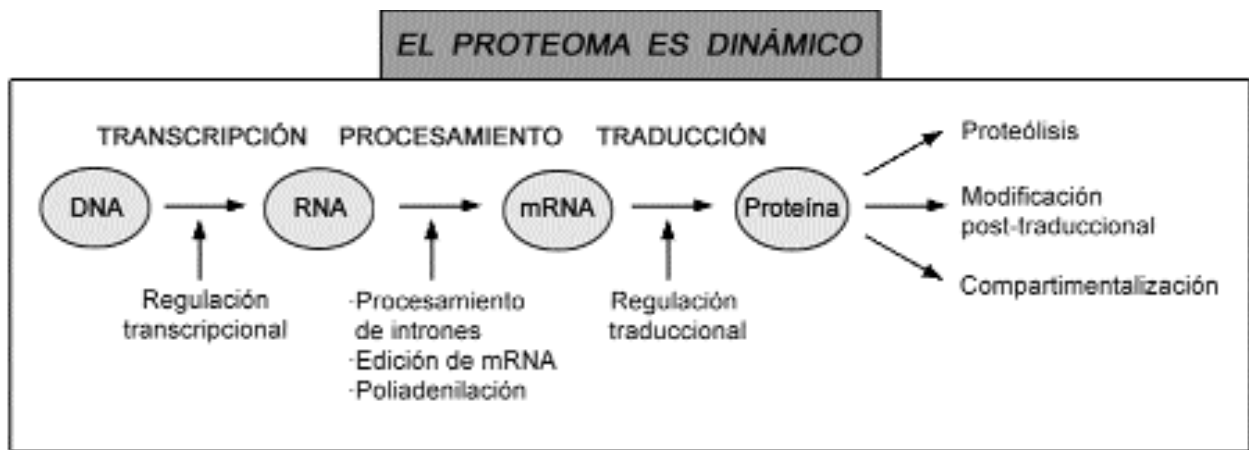


Figura 1. Mecanismos por los que un gen puede dar lugar a múltiples productos génicos.

externos e internos, las proteínas pueden ser modificadas post-traduccionalmente, translocadas, sintetizadas o degradadas (Figura 1). Obtener el proteoma de una célula es como tomar una fotografía de todas las proteínas en un momento determinado. Considerando todas las posibilidades se podría decir que un genoma podría dar lugar a un número infinito de proteomas.

Tecnología proteómica

Como ya he comentado, el desarrollo de la proteómica se debe en parte a los avances importantes realizados en la tecnología de proteínas. Sin embargo, la metodología disponible tiene sus limitaciones y actualmente todavía no es posible realizar muchos tipos de proteómica. Es necesario mejorar algunas de estas limitaciones y desarrollar nuevas tecnologías para conseguir el máximo partido.

1. Separación de proteínas. Electroforesis mono y bidimensional (2D-PAGE).

Limitaciones de la 2D-PAGE.

Alternativas a la electroforesis.

La tecnología más utilizada para la separación de proteínas es la electroforesis en geles de poliacrilamida. Fue introducida hace 32 años y hasta el momento es la técnica más eficaz para resolver mezclas complejas de proteínas.

Para muchas aplicaciones proteómicas, la electroforesis en una dimensión es el método de elección. Las proteínas se separan de acuerdo a su masa y como las proteínas son solubilizadas en dodecil sulfato sódico (SDS), no suele haber problemas de solubilización. 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 pro-

teínas después de realizar algún tipo de purificación previa.

La electroforesis bidimensional 2D-PAGE permite separar hasta miles de proteínas en un sólo experimento, y constituye actualmente el método más eficiente para la separación de mezclas muy complejas de proteínas. 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 isoelectroforesis, 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 isoelectrónico. 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. La innovación clave para la 2D-PAGE fue el desarrollo de geles con un gradiente de pH inmovilizado (IPG). El gradiente de pH inmovilizado elimina los problemas de inestabilidad del gradiente y baja capacidad de carga que iban asociados a los gradientes de pH preparados con anfólitos *carrier*. En los geles IPG, el gradiente es generado por las llamadas "inmovilinas" y está co-polimerizado con la matriz de acrilamida del gel. Este sistema ha permitido mejorar la resolución y reproducibilidad de los geles así como aumentar la cantidad de proteína que puede ser cargada. La reproducibilidad conseguida con los IPGs ha hecho posible la comparación de mapas entre distintos laboratorios, facilitando así el intercambio de información.

En cuanto a la detección de las proteínas, tradicionalmente se ha venido empleando el marcaje radioactivo o la tinción con azul de Coomassie, o con plata, para conseguir mayor sensibilidad. También se ha desarrollado un método de tinción de plata superficial compatible con la digestión

proteica y la espectrometría de masas, aunque el umbral de detección no es tan sensible como el conseguido con los protocolos de tinción de plata más utilizados. Así mismo, se han comenzado a utilizar tinciones y marcajes fluorescentes (Sypro-Ruby, Cy3, Cy5...) que presentan una sensibilidad comparable a la plata y también permiten el análisis posterior de las proteínas mediante espectrometría de masas. También se han desarrollado programas para comparar las imágenes de 2D-PAGE y facilitar la identificación y cuantificación de manchas de proteínas entre diferentes muestras. Un avance reciente de la 2D-PAGE es la técnica DIGE (*Difference gel electrophoresis*) que más adelante explicaré.

La aplicación principal de la 2D-PAGE es la proteómica de expresión. En esta aproximación, la expresión de proteínas de dos muestras se puede comparar de forma cualitativa y cuantitativa. La aparición o desaparición de manchas proporciona información sobre la expresión diferencial de proteínas y la intensidad de las manchas permite conocer los niveles de expresión. Para realizar estos estudios se pueden utilizar organismos completos, líneas celulares o fluidos biológicos. Se pueden comparar tejidos normales con tejidos enfermos, o células tratadas con drogas o diferentes estímulos. Un ejemplo de este tipo de estudio aparece en la figura de la portada, donde se muestra la expresión diferencial de proteínas de pared de formas levaduriformes y de hifas de *Cándida albicans*.

Otra aplicación importante es la proteómica del mapa celular, donde la 2D-PAGE se utiliza para hacer mapas de proteínas de microorganismos, orgánulos celulares y complejos de proteínas. También se puede utilizar para caracterizar proteínas en subproteomas que se han obtenido mediante alguna forma de purificación del proteoma.

La 2D-PAGE también presenta muchas limitaciones. Es una técnica muy laboriosa que requiere bastante tiempo, y difícil de automatizar. La 2D-PAGE está limitada por el número y el tipo de proteínas a resolver. Las proteínas muy grandes o hidrofóbicas no entran en el gel durante la primera dimensión y las proteínas muy ácidas o muy básicas no se resuelven bien. Algunos de estos problemas se pueden resolver mediante fraccionamiento, la utilización de determinadas condiciones de solubilización y la utilización de IPGs con diferentes rangos de pH. Otra limitación importante de la 2D-PAGE es la detección de proteínas poco abundantes, algunas de ellas se consideran muy importantes (proteínas regulatorias, proteínas implicadas en la transducción de señales,

receptores). En estos estudios es necesario realizar un fraccionamiento de la muestra para reducir la complejidad de los extractos. Debido a estas limitaciones, la mayor aplicación de esta técnica en el futuro será el análisis y caracterización de subproteomas y de complejos proteicos.

Las limitaciones de la 2D-PAGE han impulsado el desarrollo de otras metodologías. Una estrategia desarrollada consiste en digerir con tripsina una mezcla de proteínas para después purificar y analizar los péptidos mediante MS. Los péptidos se pueden purificar mediante cromatografía líquida, electroforesis capilar o mediante una combinación de técnicas como cromatografía de intercambio iónico y cromatografía de fase reversa. La ventaja de este procedimiento es que se dispone de una gran cantidad de proteínas. Sin embargo el análisis de los datos requiere una enorme cantidad de tiempo y ordenadores muy potentes. Además se puede perder mucho tiempo y esfuerzo en el análisis de proteínas que no tienen interés. Una de las

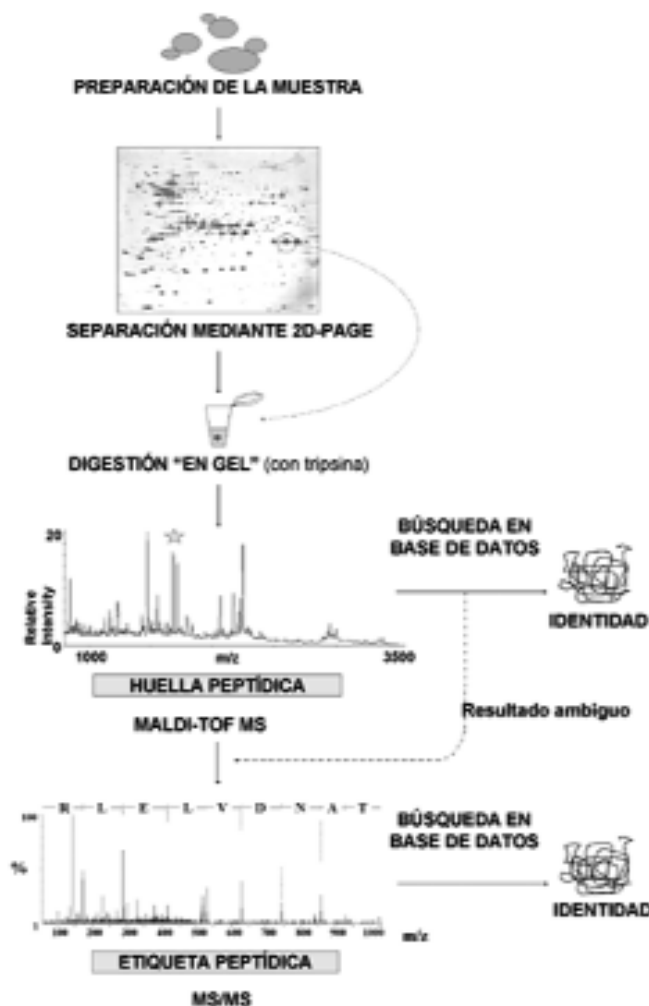


Figura 3. Estrategia para la identificación de proteínas mediante espectrometría de masas (Pitarch *et al.*, 2003).

técnicas más interesantes y que surge como alternativa a la 2D-PAGE es el ICAT (*Isotope-coded affinity tags*) que luego comentaré.

2. Identificación y caracterización de proteínas. La espectrometría de masas (MS), una técnica clave para la investigación proteómica.

En los proyectos proteómicos la identificación de proteínas es esencial. Es el primer paso para otros estudios que suponen en última instancia la caracterización funcional. Además, en el caso de los geles bidimensionales, la identificación de las manchas conduce a la creación de "mapas de referencia", que definen las proteínas expresadas por un organismo o tejido en unas condiciones determinadas.

Las proteínas pueden ser identificadas por diversos procedimientos, entre los que se incluyen la secuenciación del extremo N-terminal, detección con anticuerpos específicos, composición de aminoácidos, co-migración con proteínas conocidas, y sobre-expresión y delección de genes. Todos estos métodos generalmente son lentos, laboriosos, o caros, y por tanto no resultan apropiados para su utilización como estrategias a gran escala.

Debido a su rapidez y elevada sensibilidad, la espectrometría de masas se ha convertido en el método de elección para la identificación de proteínas a gran escala, el primer paso para el estudio del proteoma de distintos organismos. También permite la caracterización de modificaciones post-traduccionales que presentan relevancia fisiológica, tales como la glicosilación y la fosforilación. La estrategia general empleada para la identificación a gran escala de proteínas mediante espectrometría de masas se muestra en la Figura 2.

El análisis de las proteínas mediante espectrometría de masas ha sido posible gracias al desarrollo de varios métodos de ionización suave para convertir biomoléculas grandes, polares y no volátiles en iones en fase gaseosa.

Los espectrómetros de masa están formados al menos por una fuente de iones, un analizador de masas y un detector y mide la relación masa/carga (m/z) de los iones en fase gaseosa. Para analizar las proteínas mediante espectrometría de masas deben ser convertidas en péptidos, mediante proteólisis, generalmente con tripsina, ya que la masa de una proteína no es suficiente para identificarla.

Esta técnica tan robusta implica:

1.- La conversión de los péptidos en iones en fase gaseosa mediante técnicas de ionización suave, como ionización desorción con láser asistida con matriz (MALDI) a partir de una muestra en

Concha Gil García es Licenciada y Doctora en Farmacia por la Universidad Complutense de Madrid. Realizó su Tesis Doctoral sobre el dimorfismo de la levadura patógeno oportunista *Candida albicans* bajo la dirección del Dr. César Nombela, y una estancia postdoctoral en el laboratorio de la Dra. J. Douglas en Glasgow (UK). Desde 1989 es Profesora Titular en el Departamento de Microbiología II de la UCM. Actualmente es Secretaria Académica del Departamento, coordinadora del Programa de Doctorado de Microbiología y Parasitología, y responsable del Centro de Proteómica de la UCM. Este centro ofrece servicios de identificación de proteínas mediante huella peptídica, etiqueta de secuencia o secuenciación de novo, y en breve contará con las técnicas DIGE y ICAT para el análisis de la expresión diferencial de proteínas.



Su investigación se ha enfocado hacia el estudio del dimorfismo de *C. albicans* y hacia la búsqueda de dianas de nuevos antifúngicos. El grupo de investigación que dirige está centrado ahora en el estudio de la pared celular de levaduras (*C. albicans* y *Saccharomyces cerevisiae*), de la interacción *Candida*-hospedador y en el análisis funcional de genomas.

estado sólido, o la ionización mediante electrospray (ESI) de una muestra en solución.

2.- Separación de los iones según su m/z en un analizador de masas (por ejemplo un analizador tipo TOF (*Time Of Flight*), cuadrupolo, trampa iónica, etc.)

3.- Fragmentación opcional de los iones peptídicos seleccionados mediante descomposición metaestable (o técnica de PSD: *postsource decay*) o mediante disociación inducida por colisión (CID) llevada a cabo en un espectrómetro de masas en tandem combinando dos analizadores diferentes.

4.- Medida de las masas en un detector obteniendo un espectro de masas que refleja la abundancia de los iones frente a su valor m/z .

Aunque se han desarrollado diversas combinaciones de fuentes de ionización y de analizadores de masas, la fuente de MALDI se suele acoplar a un analizador tipo TOF mientras que la ionización mediante ESI normalmente se combina con un triple cuadrupolo, con una trampa iónica o con un híbrido cuadrupolo tiempo de vuelo (Q-TOF).

Recientemente se han desarrollado también fuentes de MALDI que se acopan a un analizador Q-TOF o a dos analizadores TOF en tándem (MALDI-TOF-TOF)

Para la identificación de proteínas se han desarrollado dos estrategias que se encuentran representadas esquemáticamente en la Figura 3:

- Identificación mediante huella peptídica (PMF: *peptide mass fingerprinting*) o mapeo peptídico utilizando un espectrómetro tipo MALDI-TOF.
- Identificación mediante fragmentación de péptidos obteniendo la secuencia total o parcial de los aminoácidos (etiqueta de secuencia) utilizando un espectrómetro de masas en tándem.

2.1. Huella peptídica. Herramienta para la identificación a gran escala de proteínas presentes en las bases de datos

El mapeo de péptidos es una técnica utilizada rutinariamente para identificar proteínas de forma rápida, normalmente a partir de geles de SDS-PAGE o 2D-PAGE y que se realiza normalmente en un espectrómetro de masas tipo MALDI-TOF. En esta aproximación la proteína se digiere con una enzima, normalmente tripsina (que rompe específicamente por lisinas y argininas, si no están unidas a una prolina). La muestra es incorporada a una placa metálica junto con una matriz (moléculas que absorben una pequeña cantidad de energía, como el ácido 2,5-dihidroxi-benzoico o el ácido ciano-4-hidroxicinámico) y al evaporarse se forman cristales. Posteriormente, la muestra es irradiada con láser para ionizar las moléculas. Los iones son acelerados por un campo eléctrico hacia un detector, el valor m/z de cada ión viene determinado por el tiempo de vuelo para llegar desde la fuente al detector.

La huella peptídica (PMF) de una determinada proteína es un conjunto de péptidos generados mediante la digestión de una proteasa específica. Estas masas peptídicas experimentales son comparadas con las masas peptídicas teóricas de proteínas presentes en bases de datos. Para ello se han desarrollado diversos algoritmos (PEPSEA, PROFOUND, MS-FIT, PEPTIDENT...) disponibles en Internet. Para la identificación correcta de la proteína se requiere que las masas de un gran número de péptidos coincidan con las masas teóricas de los péptidos y que cubran parte de la secuencia de la proteína de la base de datos.

La espectrometría de masas tipo MALDI-TOF se considera un pilar de la proteómica. El instrumento es robusto, relativamente caro y con posibilidades de ser automatizado. Como el tiempo de análisis de una muestra es muy corto se puede utilizar para análisis a gran escala. Sin embargo, este tipo de espectrometría tiene sus limitaciones

- La ionización de los péptidos es selectiva y no es cuantitativa. En un conjunto equimolecular de péptidos procedentes de la digestión de una

proteína, algunos péptidos pueden no ser detectados y en el resto de ellos puede haber una gran variación en la señal de intensidad.

- Si la cantidad de proteína en el gel es pequeña, el número de péptidos observados puede ser pequeño y por tanto la proteína no se puede identificar con seguridad.
- El MALDI-TOF *MS* tiene poca utilidad para analizar mezclas de proteínas. Manchas de proteínas muy claras de geles de 2D pueden contener varias proteínas

La principal ventaja del MALDI reside en la posibilidad de realizar análisis a gran escala de organismos que tengan el genoma completamente secuenciado. Cuando no se dispone de la secuencia completa del genoma y cuando se tiene poca cantidad de proteína el MALDI-TOF es insuficiente ya que se requiere más información.

2.2.- Secuencia peptídica o etiqueta de secuencia. Estrategia para la identificación de proteínas no anotadas en las bases de datos o para las identificaciones ambiguas mediante MALDI-TOF.

Los espectrómetros de masas en tándem *MS/MS* permiten también la determinación de la secuencia de aminoácidos. Se selecciona un ión por la masa en un primer espectrómetro y se fragmenta por colisión con un gas y los fragmentos se analizan en un segundo espectrómetro. Puede utilizarse con una fuente de ionización tipo MALDI o ESI.

En la ionización mediante *electrospray* los iones se forman a partir de una muestra en solución, que se vaporiza haciéndola pasar a través de una fina aguja a la que se aplica una alta diferencia de potencial. Las gotas cargadas pasan a una zona de potencial más bajo, y son desolvatadas, adquiriendo protones las moléculas de la muestra y dando lugar a iones con carga múltiple. El último avance de la espectrometría ESI-MS ha sido la introducción de microcapilares de borosilicato recubiertos de oro para la inyección de la muestra en el espectrómetro. A esta técnica se le ha denominado *nanoelectrospray* y permite analizar volúmenes de muestra de 1-2 μ l durante un periodo de tiempo de unos 30 minutos, optimizando mucho la señal y posibilitando la realización de espectrometría de masas en tándem en un equipo de tipo cuadrupolo, para determinar la secuencia aminoacídica de péptidos. Se llega a conseguir una sensibilidad del orden de femtomoles. Para ESI hay varias formas de incorporar la muestra al espectrómetro. El método más sencillo es cargar tubos microcapilares individuales con la muestra. Como se utiliza un tubo nuevo para cada muestra no

hay contaminación. En ESI los péptidos requieren cierta purificación que se puede realizar en el microcapilar. Tanto la necesidad de purificar la muestra como la de cargar de forma manual los capilares son dos problemas que hacen que esta técnica sea lenta y tediosa. Como alternativa, este tipo de espectrómetros suelen estar conectados *on-line* a un equipo de cromatografía líquida (LC) que de forma automática purifique e introduzca la muestra en el espectrómetro. Los espectrómetros de masas en tándem pueden ser de diferentes tipos:

- Triple cuadrupolo (acoplado a una fuente ESI). El primer cuadrupolo separa el ión de interés y éste es fragmentado en una CID cámara de colisión con un gas inerte. El tercer cuadrupolo mide la m/z de los productos de disociación.
- Híbrido cuadrupolo-TOF (Q-TOF). Combina un cuadrupolo, un hexapolo para la colisión y un analizador tipo TOF.
- Trampa de iones. Atrapan iones en un campo eléctrico 3D. A diferencia de los cuadrupolos, donde los iones son eliminados antes de comenzar el análisis, la principal ventaja de la trampa es su capacidad para conservar los iones y posteriormente eliminarlos selectivamente.
- Recientemente se ha desarrollado un TOF en tándem con fuente de MALDI, MALDI-TOF-TOF. Los iones fragmentados en una CID son reacelerados y analizados en un segundo analizador tipo TOF.

Cuando un ión se introduce en una CID, interacciona con el gas de colisión (normalmente nitrógeno o argón) y tiene lugar la fragmentación a lo largo del esqueleto peptídico. Como los iones se pueden fragmentar en múltiples sitios se ha creado un nomenclatura para indicar el tipo de iones generados. Si, después de la ruptura del enlace peptídico, la carga se mantiene en el extremo amino del ión, éste se denomina ión **b**, mientras que si la carga se mantiene en el extremo carboxilo, el ión se denomina **y**. Los espectros de fragmentación MS/MS proporcionan información de las diferencias de masas entre dos iones consecutivos del mismo tipo (por ejemplo, iones de la serie **y** o **b**.) correspondientes a la pérdida de un aminoácido en el péptido. Estos espectros proporcionan información sobre la identidad y posición del aminoácido en el péptido. Por tanto es posible adquirir la secuencia completa del péptido si se parte de una fragmentación de buena calidad. Otras veces sólo se obtiene una secuenciación parcial del péptido. La información sobre la secuencia parcial, masa del péptido, y la localización exacta de los aminoácidos puede permitir la

identificación de la proteína en las bases de datos. A esta aproximación se le denomina identificación por "etiqueta de secuencia". Existen diversas herramientas bioinformáticas que se pueden utilizar para la interpretación de la fragmentación de un péptido (PEPSEA, SEQUEST, PEPFRAG, MS-TAG y MASCOT).

El método de la secuenciación de novo es la estrategia de elección para aquellos organismos que no han sido secuenciados o su secuencia de DNA no ha sido anotada. El objetivo de esta aproximación es obtener secuencias completas de aminoácidos que se deduzcan *de novo* de los espectros MS/MS (bien mediante interpretación manual o con ayuda de un programa) y buscar homólogas frente a proteínas presentes en bases de datos. Esta estrategia permite encontrar proteínas homólogas en otras especies utilizando programas de búsqueda como BLAST o FASTA. Aunque la proteína de interés pueda no ser identificada o sea una proteína desconocida, la información de la secuencia puede ser utilizada para clonar el gen correspondiente.

Nuevos desarrollos de la metodología proteómica: Cromatografía multidimensional, ICAT, DIGE y arrays de proteínas.

Actualmente se están desarrollando tecnologías proteómicas sin llevar a cabo la separación de las proteínas en un gel, como son la cromatografía líquida bidimensional o la electroforesis capilar acoplada a MS , los *arrays* de proteínas o el etiquetado de péptidos con diferentes reactivos como en la técnica de ICAT (*isotope-coded affinity tags*). Estas técnicas son complementarias a las que se vienen utilizando. También se están desarrollando mejoras en la separación y análisis de las proteínas mediante electroforesis bidimensional.

Cromatografía multidimensional. Para no tener las limitaciones asociadas a los geles en la separación y caracterización de las proteínas se han llevado a cabo aproximaciones multidimensionales utilizando cromatografía líquida bidimensional (columna de intercambio iónico y columna de fase reversa), seguida de un análisis de los péptidos mediante MS (MudPIT: *multidimensional protein identification technology*). En dicha aproximación se requiere realizar la digestión de las proteínas antes de la separación primero por su carga y posteriormente por su hidrofobicidad. Este método permite la separación e identificación de cientos de proteínas en un solo experimento (Washburn *et al.*, 2001).

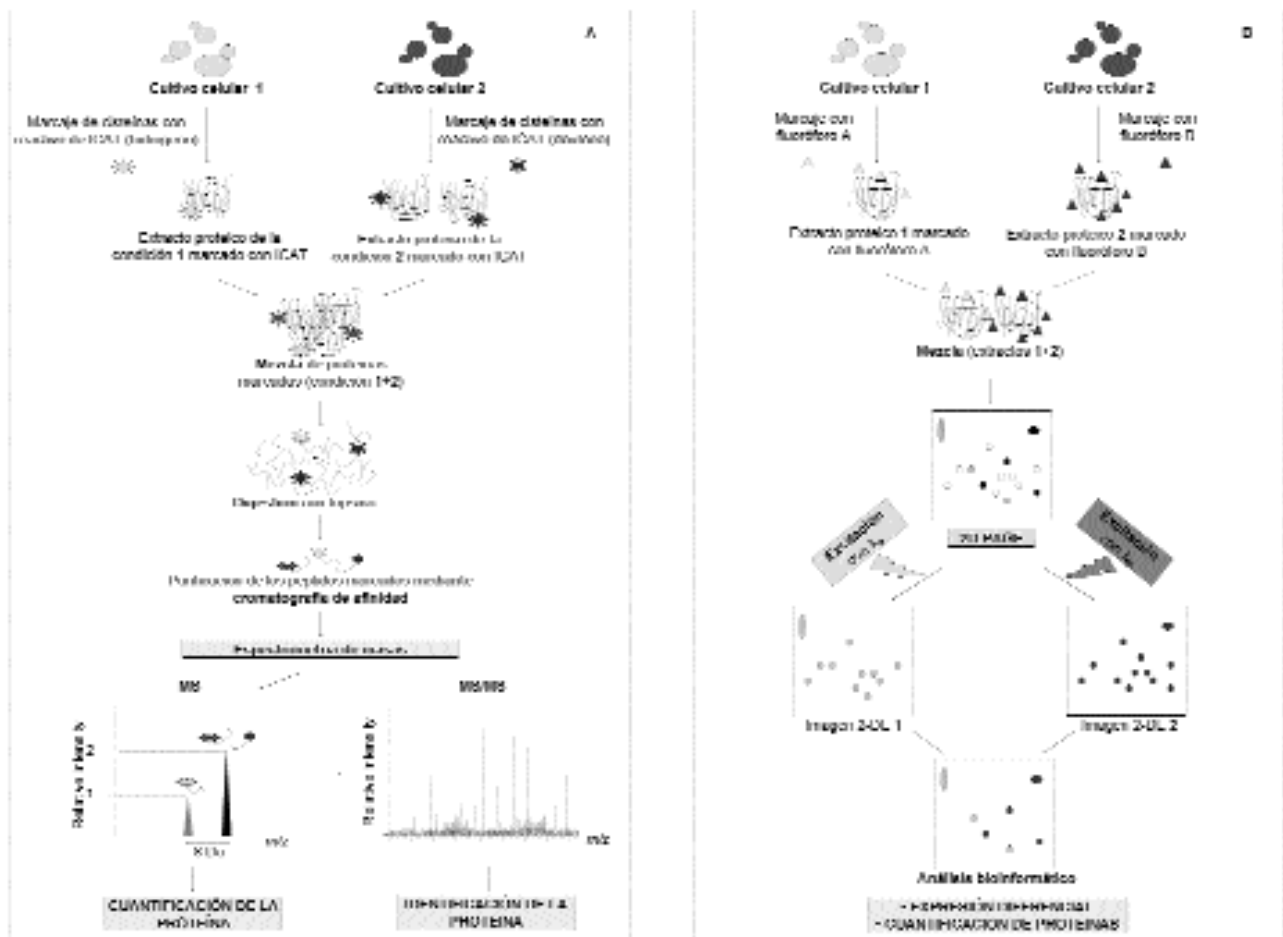


Figura 3. Dos nuevas aproximaciones para la detección y cuantificación diferencial de proteínas. **A.** ICAT *Isotope coded affinity tag*. **B.** DIGE (*Differential gel electrophoresis*). (Pitarch *et al.*, 2003).

Tecnologías proteómicas para cuantificar el nivel de expresión de las proteínas: ICAT y DIGE. La principal aplicación de la proteómica es el estudio del perfil de expresión de proteínas. Existen dos estrategias que mejoran el estudio de la expresión diferencial de proteínas entre diferentes muestras. Una de ellas es el ICAT, que permite determinar la cantidad de proteína relativa entre dos muestras (Gygi *et al.*, 1999). Las dos muestras proteicas son marcadas con el reactivo del ICAT ligero o pesado (según lleve hidrógeno o deuterio). Este reactivo se une a las cisteínas y contiene biotina para facilitar la purificación. Posteriormente, las dos muestras se mezclan y se digieren con tripsina. Los péptidos marcados con el reactivo del ICAT son separados en un columna de afinidad y analizados mediante MS. La intensidad relativa de los péptidos idénticos de cada muestra (se diferencian en una masa de 8 Dalton) muestran la abundancia de la proteína de la que proceden. La fragmentación del péptido mediante MS/MS conduce a la identificación de la proteína (Figura 3A).

Recientemente también se ha descrito una

aproximación basada en el marcaje de las proteínas con diferentes fluorocromos y la separación de las muestras mediante 2D-PAGE en un mismo gel. Dicha metodología, denominada DIGE (*Differential gel electrophoresis*), minimiza la variabilidad de los gels, disminuye el tiempo de análisis y permite una cuantificación del perfil de expresión muy precisa (Figura 3B).

Arrays de proteínas. Los *arrays* de proteínas se están desarrollando rápidamente para la caracterización de actividades y para detección de las interacciones proteína-proteína a gran escala. Al igual que los *arrays* de DNA, los *arrays* de proteínas serán esenciales para la investigación básica así como para la investigación más aplicada al descubrimiento de medicamentos y desarrollo de métodos de diagnóstico. En un trabajo pionero realizado por el grupo de Snyder (Zhu *et al.*, 2000), se desarrolló un *chip* con 6.000 proteínas de levadura para identificar nuevas proteínas que interaccionasen con calmodulina o con fosfolípidos. Las proteínas se obtuvieron por clonación de los ORFs correspondientes y cada proteína se expresó fusionada a GST (glutathion-S- transferasa) y a una

etiqueta de histidinas. Este trabajo tan importante mostró que es posible preparar *microarrays* con miles de proteínas y utilizarlos para estudiar interacciones. Sin embargo, aunque ya se han realizado avances importantes para la preparación de los *arrays*, todavía es necesario enfrentarse a varios retos tecnológicos que permitan hacer posible la utilización de esta herramienta a un gran número de investigadores.

Conclusiones

La proteómica proporciona un conjunto de herramientas muy poderosas para el estudio a gran escala de la función de los genes a nivel de proteína. Los estudios mediante espectrometría de masas de las proteínas separadas en geles está conduciendo al renacimiento de las aproximaciones bioquímicas para el estudio de la función de las proteínas. Actualmente existen dos estrategias bien definidas para el estudio del proteoma: una de ellas se basa en la utilización de la electroforesis bidimensional y la otra utiliza otros métodos de separación de proteínas, principalmente métodos cromatográficos. En ambos casos la identificación y caracterización de proteínas se realiza mediante espectrometría de masas. Actualmente existe un gran interés en la mejora y en el desarrollo de estrategias que faciliten el estudio de las proteínas y de los proteomas. En mi opinión la proteómica sólo esta en sus inicios. La proteómica en combinación con otras aproximaciones genómicas tiene un enorme potencial para comprender mejor como funcionan los sistemas biológicos complejos.

Bibliografía

- Banks RE, Dunn MJ, Hochstrasser DF, Sanchez J, Blackstock W, Pappin DJ, Selby PJ. 1999. Proteomics: new perspectives, new biomedical opportunities. *The Lancet* 356, 1749-1756.
- Blackstock W, Man M (Eds). 2001. A trends guide for proteomics. A supplement to *Trends in Biotechnology* Vol 19 (10).
- Aebersold R, Cravatt B (Eds). 2002. A trends guide for proteomics. A supplement to *Trends in Biotechnology* Vol 20 (12).
- Graves PR, Haystead TAJ. 2002. Molecular Biologist's guide to proteomics. *Microbiology and Molecular Biology Reviews* 66, 39-63.
- Gigy SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* 17, 994-999.
- Pardo M, Monteoliva L, Pla J, Sánchez M, Gil C, Nombela C. 1999. Two-dimensional analysis of proteins secreted by *Saccharomyces cerevisiae* protoplasts: a novel approach to study the cell wall. *Yeast* 15, 459-472.
- Pardo M, Ward M, Pitarch A, Sánchez M, Nombela C, Blackstock W, Gil C. 2000. A proteomic approach to study *Saccharomyces cerevisiae* cell wall biogenesis. *Electrophoresis* 21, 3396-3410.
- Pitarch A, Sánchez M, Nombela C, Gil C. 2002. Sequential fractionation and two-dimensional gel analysis unravels the complexity of the dimorphic fungus *Candida albicans* cell wall proteome. *Molecular and Cellular Proteomics* 1, 967-982.
- Pitarch A, Sánchez M, Nombela C, Gil C. 2003. Analysis of the *Candida albicans* proteome. I. Strategies and applications. *Journal of Chromatography B* 787, 101-128.
- Washburn MP, Wolters D, Yates III JR. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* 19, 242-247.
- Zhu, H. *et al.* 2001. Global analysis of protein activities using proteome arrays. *Science* 293, 2101-2105