
PROTEÓMICA

HERRAMIENTAS DE LA ERA POSGENÓMICA EN LA VIDA COTIDIANA

Joaquín Abián i Susanna Baqué (coords.)



Cicle de conferències celebrat
els mesos de febrer i març de 2002
a la Residència d'Investigadors
CSIC-Generalitat de Catalunya

PROTEÓMICA

«Publicacions de la Residència d'Investigadors»

El cicle de conferències «Proteómica. Herramientas de la era pos-genómica en la vida cotidiana» va comptar amb la col·laboració de la Fundació Catalana per a la Recerca i d'Applied Biosystems



PROTEÓMICA

HERRAMIENTAS DE LA ERA
POSGENÓMICA EN LA VIDA COTIDIANA

Joaquín Abián i Susanna Baqué (coords.)



Cicle de conferències celebrat
els mesos de febrer i març de 2002
a la Residència d'Investigadors
CSIC-Generalitat de Catalunya

RESIDÈNCIA D'INVESTIGADORS
CSIC-GENERALITAT DE CATALUNYA

Barcelona, 2003

**Consorci de la Residència d'Investigadors
CSIC-Generalitat de Catalunya**

President del CSIC: ROLF TARRACH SIEGEL
Conseller d'Universitats, Recerca i Societat de la Informació:
ANDREU MAS-COLELL

Consell de Govern

President del Consorci: DAVID SERRAT I CONGOST, fins al 30 d'abril de 2002; JOAQUIM CASAL I FÀBREGA (Director General de Recerca, Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya [DURSI]), a partir del 30 d'abril i fins al 25 d'octubre de 2002; ROLF TARRACH SIEGEL (President del CSIC), a partir del 25 d'octubre de 2002;
Director: FRANCESC FARRÉ RIUS
Director científicocultural: LLUÍS CALVO CALVO

Vocals:

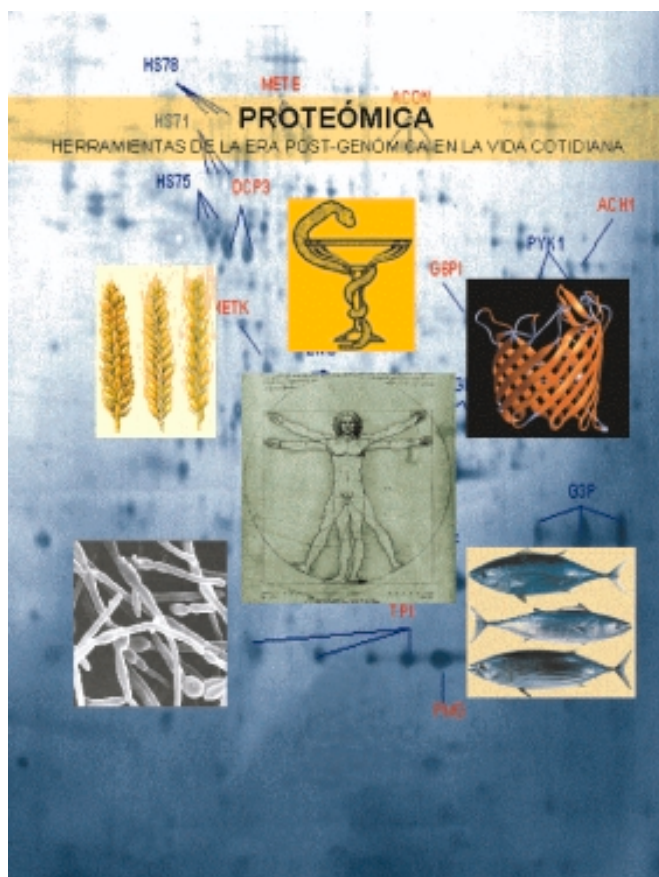
ROLF TARRACH SIEGEL (President del CSIC), fins al 25 d'octubre de 2002; JOAQUIM CASAL I FÀBREGA (Director General de Recerca del DURSI), a partir del 25 d'octubre de 2002;
JOSEP GRIFOLL GUASCH (Secretari General del DURSI)
LLUÍS CALVO CALVO (Coordinador Institucional del CSIC a Catalunya)

© DELS AUTORS

Primera edició: febrer de 2003

Impressió: Alta Fulla · Taller

D. L. B 7593-2003



Composición original de J. Abián y S. Baqué referida, con motivos alegóricos, a los temas tratados en el ciclo de conferencias.

Fondo: Imagen de gel bidimensional correspondiente al análisis de proteínas de *C. albicans*. Las diferentes manchas corresponden a proteínas distintas, algunas de ellas (las marcadas con su nombre) identificadas mediante espectrometría de masas (Imagen cedida por la Dra. Concha Gil, UCM, Madrid).

Imagen central: Estudio de las proporciones de Leonardo da Vinci, del tratado *De Architectura* de Vitrubio; lápiz y tinta, 34,3 × 24,5; Galleria dell'Accademia, Venecia.

Imágenes laterales (desde arriba y en dirección de las agujas del reloj):

- Copa de Higía (o Hygeia), diosa de la salud, una de las hijas de Asclepio (o Escolapio), dios griego de la medicina.
- Estructura tridimensional de la proteína de membrana Matrix Porin de *E. coli*. (Imagen obtenida de la base de datos Swiss3D).
- Diferentes especies de atún: yellowfin, *Thunnus albacares*; bonito, *Sarda sarda*; y skipjack, *Katsuwonus pelamis*. (Extraído de www.atuna.com).
- Hifas de *Candida albicans* vistas al microscopio electrónico de barrido (Imagen cedida por la Dra. Concha Gil, UCM, Madrid).
- Espigas de trigo, cebada y centeno. Estos cereales contienen gluten y suelen encontrarse contaminando otras harinas o alimentos supuestamente libres de estas proteínas (Imágenes cedidas por Enrique Méndez, CNB, CSIC, Madrid).

PROTEÓMICA: HERRAMIENTAS DE LA ERA POSGENÓMICA EN LA VIDA COTIDIANA

El anuncio de la finalización de la secuenciación del genoma humano ha sido un hecho histórico que ha trascendido las fronteras del entorno científico. Se ha abierto una etapa que, aunque bautizada como «era posgenómica», debe considerarse más el principio que el final de la revolución genómica. Gracias a la información aportada por los proyectos de secuenciación es hoy posible abordar de forma efectiva el estudio de la funcionalidad del genoma, y en especial el análisis de los productos codificados por los genes: las proteínas. La proteómica, ciencia que estudia las proteínas y sus interacciones en los distintos modelos vivos, es un área de investigación y desarrollo que ha evolucionado vertiginosamente en los últimos años. En febrero de 2001 se fundó la Organización para el proteoma humano (HUPO), cuyo objetivo es coordinar los esfuerzos para el estudio de este proteoma, y numerosas iniciativas tanto públicas como privadas están en marcha. Recientemente, tres investigadores cuya actividad ha sido vital para el desarrollo de las técnicas espectrométricas (electrospray y MALDI) y de resonancia magnética empleadas en proteómica, John Fenn, Koichi Tanaka y Kurt Wuthrich, han recibido el premio Nobel de química (2002).

En la actualidad se dispone de poderosas herramientas analíticas para llevar a cabo estos estudios de manera automatizada y al mismo tiempo de forma cada vez más exhaustiva, rápida y precisa. Se han iniciado diversos proyectos a gran escala para el estudio de partes del proteoma, de especial significa-

ción como el de las proteínas fosforiladas o el de las glicoproteínas. Estos subgrupos de proteínas constituyen lo que se denomina el fosfoproteoma y el glicoproteoma o glicoma, respectivamente. Otros proyectos se han centrado en la descripción del proteoma de tejidos específicos, incluyendo fluidos como el plasma o la orina, o en el estudio de los cambios sufridos en un proteoma como resultado de enfermedades como el cáncer o el parkinson. Este tipo de investigación se ha dirigido tanto a la caracterización de los mecanismos implicados en estos procesos como a la búsqueda de proteínas cuya presencia o ausencia pueda servir como marcador de la enfermedad y por tanto como medio de diagnóstico o como diana para la producción de nuevos medicamentos. Es por este motivo que el interés de las compañías privadas en el estudio del proteoma es tanto o más alto que el mostrado en el estudio del genoma humano.

El interés suscitado por estos nuevos conocimientos reside en el potencial de la proteómica como generadora de una nueva revolución biotecnológica en sectores como el desarrollo de fármacos, o la caracterización de marcadores de enfermedad de uso en el diagnóstico médico. Entre las diversas aplicaciones existen otras, quizás menos conocidas, pero de un mayor impacto directo en nuestra vida cotidiana, como el control de alimentos dietéticos, la clasificación de pescados para el consumo o la preparación de vacunas.

En este ciclo de conferencias, reconocidos expertos nos han acercado a esta nueva ciencia, cuya potencialidad y limitaciones fueron comentadas en detalle por el Dr. Juan Pablo Albar del Centro Nacional de Biotecnología (CSIC, Madrid). La complementariedad de los estudios genómicos y proteómicos así como la importancia de la bioinformática para la manipulación de las grandes cantidades de datos producidos en estos estudios fueron puestas de relieve por los Dres. Roderic Guigó (IMIM) y Gonzalo Firpo (Applied Biosystems).

Diversas conferencias ilustraron ejemplos concretos de cómo la proteómica juega un papel relevante en la práctica diaria. Así, el grupo de la Dra. Concha Gil, de la Universidad Complutense de Madrid, nos mostró cómo esta aproximación científica permite abordar el estudio de las infecciones fúngicas. *Candida albicans*, un hongo normalmente presente en ciertas zonas del organismo humano, puede, en determinadas condiciones, producir infecciones graves denominadas Candidiasis sistémicas, responsables del 4% de las infecciones hospitalarias. El estudio del perfil proteico de estos hongos y la caracterización de las proteínas diana de los anticuerpos que aparecen en los pacientes afectados por esta infección permiten identificar de forma muy efectiva marcadores proteicos de enfermedad y proponer posibles dianas terapéuticas para el control de este problema. En el ámbito de la alimentación, el Profesor J. M. Gallardo nos mostró cómo estas técnicas pueden utilizarse para la caracterización de especies marinas o para la determinación, mediante biomarcadores proteicos, de la frescura de los pescados destinados al consumo. Una implicación directa en la salud derivada del consumo de ciertos pescados son las respuestas alérgicas a algunas proteínas del pescado, como las parvalbúminas de determinadas especies de merluza y bacalao. El control de la composición de los alimentos es por tanto imprescindible en cuanto que estas alteraciones pueden afectar directamente a la salud. En este sentido, el Profesor Enrique Méndez, del CNB (CSIC, Madrid), nos mostró la importancia de dicho control en el caso de las personas afectadas por la enfermedad celíaca (1 de cada 200-300 nacidos). Esta enfermedad se produce por una intolerancia permanente al gluten, una familia de proteínas presentes en el trigo y otros cereales y usadas comúnmente como aditivos en muchos alimentos preparados o como contaminantes de otras harinas vegetales. La eliminación de cualquier presencia de gluten en la alimentación de la población celíaca es por tanto primordial para su

calidad de vida. El estudio de la composición química del gluten está permitiendo el desarrollo de métodos específicos y muy sensibles para la caracterización de su presencia en alimentos en principio destinados a la población celíaca.

Los organizadores hemos querido, con estas conferencias, mostrar algunos de los aspectos más prácticos de la investigación en Proteómica, una ciencia cuyas implicaciones científicas, pero también económicas y sociales, sólo estamos empujando a vislumbrar.

Finalmente, queremos agradecer a la Residencia de Investigadores CSIC-Generalitat de Catalunya, y en concreto a sus responsables, los Drs. Francesc Farré —Director— y Luis Calvo —Director Científico-Cultural—, el apoyo y el estímulo continuado que nos han proporcionado tanto para la organización del seminario como para la edición del ciclo de conferencias.

JOAQUÍN ABIÁN y SUSANNA BAQUÉ
Coordinadores

Febrero-Marzo 2002

IDENTIFICACIÓN Y CATALOGACIÓN DE PROTEÍNAS: PRIMEROS OBJETIVOS DE LA ERA POSGENÓMICA

JUAN A. LÓPEZ, L. EMILIO CAMAFEITA, ENRIQUE CALVO,
ANA BELOSO Y JUAN P. ALBAR

Laboratorio de Proteómica

Centro Nacional de Biotecnología, CSIC

I. Resumen

El análisis exhaustivo de la expresión génica en sistemas biológicos complejos ha requerido del desarrollo de distintas metodologías. Habitualmente ha supuesto el estudio bien de una porción del transcriptoma celular o bien de una porción del proteoma celular. Cada aproximación posee ventajas y desventajas tanto conceptuales como metodológicas.

La proteómica es una de las áreas tecnológicas de mayor relevancia científica de la era posgenómica. La enorme cantidad de información generada por los proyectos de secuenciación de genomas, y la necesidad de descifrar esta información han desplazado el foco de atención hacia el estudio directo de las proteínas, su estructura, su función, sus interacciones y sus modificaciones. La proteómica se está convirtiendo, pues, en una estrategia experimental clave para la ciencia del siglo XXI, y constituye un campo emergente de la biotecnología cuyo desarrollo reviste una importancia estratégica de primer orden.

II. Antecedentes

Bases de datos generadas por los proyectos de secuenciación de genomas

Los programas de secuenciación de genomas han definido el contenido informacional de una amplia variedad de organismos, desde el genoma pequeño de *Haemophilus influenzae* (1,8 Mb) al de la mosca *Drosophila melanogaster* (137 Mb) pasando por el de levaduras (*Saccharomyces cerevisiae*: 12 Mb), nematodos (*Caenorhabditis elegans*: 97 Mb), plantas (*Arabidopsis thaliana*: 115,4 Mb/25.598 orfs), más de 108 organismos han sido secuenciados y anotados, incluyendo el de nuestra especie *Homo sapiens* (3.285 Mb, 99% anotado); además cerca de 700 genomas virales, 250 de plásmidos, 90 de fagos y 200 de organelos están depositados en las bases de datos públicas. A todo esto habrá que sumar los 329 genomas de organismos procariotas y 219 de eucariotas que están en proceso de secuenciación (GOLD, 17-10-2002, Kyrpides 1999). Esta anotación de genes y los análisis a gran escala de mRNAs con técnicas como SAGE (Velculescu et al. 1997) están arrojando luz acerca de qué proteínas esperamos encontrar en un organismo, en qué nivel deben estar presentes y en qué momento ellas podrían ser expresadas.

La enorme información así generada, y en continuo crecimiento, supone el sustento sobre el que empieza a definirse el nuevo reto de la etapa posgenómica: la identificación y catalogación de los productos que expresan estos genes. No obstante, el ADN sólo define lo que podría suceder. Un gen puede dar lugar a más de una proteína o «especie proteica». El genoma no define las modificaciones postraduccionales de las proteínas, ni su actividad, ni su localización, ni las interacciones con otras proteínas y/o metabolitos.

Definición de términos

El **proteoma** puede definirse como «el conjunto de proteínas expresadas por un genoma» y la **proteómica** como «la genómica funcional al nivel de proteínas» (Wilkins et al. 1995). La **proteómica** es, pues, la ciencia que correlaciona las proteínas con sus genes: las células expresan varios miles de proteínas diferentes y cada una de éstas puede experimentar numerosas modificaciones postraduccionales dinámicas en respuesta a microambientes cambiantes, lo cual incrementa de forma significativa el número de especies proteicas diferentes presentes (Wilkins et al. 1999). Por tanto el **proteoma** es «una imagen dinámica de todas las proteínas expresadas por un organismo, célula o compartimento subcelular concreto en un momento dado y bajo determinadas condiciones», constituyendo el mapa de expresión proteica de un organismo, tejido o célula dado en condiciones concretas de tiempo y ambiente.

La ciencia de la **proteómica** no es nueva en sus bases experimentales, pero ha dado un salto cualitativo y cuantitativo sensible gracias al desarrollo convergente de diferentes áreas de investigación como son los proyectos de secuenciación de distintos genomas, la potencia computacional actual y las herramientas bioinformáticas, y finalmente, un refinamiento sin precedentes en las técnicas de análisis de proteínas, tanto en sus posibilidades como en su sensibilidad.

Básicamente, la **proteómica** pone en juego un conjunto de técnicas destinadas a *resolver* (geles de poliacrilamida tanto monodimensionales como bidimensionales de alta resolución, cromatografía líquida multidimensional, cromatografía de retención sobre «chips» de proteínas), *cuantificar* (escáner, phosphorimage, análisis de imágenes, ICAT) e *identificar* y *caracterizar* proteínas (espectrometría de masas, EM), secuenciación de proteínas por degradación de Edman, inmunoblot, etc.), así como *almacenar* y *analizar* los datos (bioinformática,

bancos de datos de geles 2D-E y proteínas, LIMS...), y *comunicar y entrecruzar* tanto información como resultados (web, publicaciones, etc.).

La **proteómica** puede aplicarse a tres tipos de estudios y como tales suelen referirse las áreas principales en las que ésta puede ser dividida:

a) *La identificación y microcaracterización sistemática de proteínas*, con el objetivo de abordar la identificación a gran escala de los componentes de un proteoma y de determinar sus modificaciones postraduccionales.

La identificación a gran escala de los componentes de un proteoma permite la catalogación de los mismos y la construcción de bases de datos, que suelen organizarse en base a la imagen del proteoma que resulta de su separación en geles de electroforesis bidimensional. Estas bases de datos tienen una enorme utilidad para los proyectos de investigación que necesiten un conocimiento detallado de esos proteomas.

b) La identificación de los componentes del proteoma que sufren alteraciones en sus niveles de expresión a consecuencia de alteraciones fisiopatológicas (*«Proteómica de Expresión Diferencial»*).

Estos proyectos pretenden identificar qué proteínas sufren alteraciones en sus niveles de expresión a consecuencia de cambios en su entorno, situaciones de estrés, administración de drogas, efectores o señales bioquímicas o su estado fisiológico o patológico.

El análisis exhaustivo de la expresión génica en sistemas biológicos complejos ha requerido del desarrollo de distintas metodologías. Habitualmente han supuesto el estudio bien de una porción del transcriptoma celular mediante el uso de «chips» o «arrays» de ADN, o bien de una porción del proteoma celular. Cada aproximación posee ventajas y desventajas tanto conceptuales como metodológicas. En general, estas dos

aproximaciones resultan complementarias. Así, los niveles de expresión de mRNA no siempre reflejan adecuadamente el nivel de expresión real de las proteínas correspondientes, por lo que la aproximación proteómica permite obtener resultados más concluyentes desde un punto de vista tanto fisiológico como farmacológico. Sin embargo, esta aproximación sólo permite analizar los niveles de expresión de proteínas que estén presentes, en un proteoma dado, a un nivel de sensibilidad por encima de un límite de detección particular, mientras que la tecnología de los chip de DNA permite concentrar el análisis en el nivel de expresión de determinados genes, con una sensibilidad muy superior.

c) La caracterización de las interacciones subcelulares existentes entre las proteínas y la determinación de los componentes de complejos macromoleculares (*«Proteómica de Mapa Celular»*).

Estos proyectos pretenden elucidar la función de las proteínas caracterizando las interacciones que tienen lugar en el interior de la célula, y responden a la noción cada vez más extendida de que las proteínas no actúan en forma aislada sino que tienden a formar grandes complejos. Mediante este tipo de proyectos, aplicados de forma sistemática, se pretende la construcción de un mapa físico de las interacciones existentes entre las proteínas celulares.

Herramientas bioinformáticas

La accesibilidad a las bases de datos que contienen las secuencias genómicas, junto con la disponibilidad de técnicas como la electroforesis en geles bidimensionales de alta resolución y la espectrometría de masas, adecuada para análisis de péptidos y proteínas (Roepstorff 1997), han hecho posible la identificación de cientos de proteínas de numerosos organis-

mos (Sánchez et al. 1997). Los avances en los métodos y tecnología proteómicos se han centrado principalmente en la identificación de proteínas a gran escala mediante el análisis de sus *huellas peptídicas*. Esto implica la digestión de una proteína dada con una endoproteasa de especificidad de rotura conocida, la medida de las masas de los péptidos resultantes por espectrometría de masas y la identificación de la proteína por ajuste de las masas peptídicas observadas frente a bases de datos de proteínas cuyas masas peptídicas son generadas teóricamente (Mann et al. 1993) La eficacia de esta técnica es tal que está convirtiéndose en el punto de encuentro para la identificación de cientos de proteínas (Shevchenko et al. 1996; Traini et al.1998).

Incremento de la complejidad: Modificaciones postraduccionales

Sin embargo, hasta el momento se ha prestado menos atención a cómo la información así generada puede ser utilizada en una caracterización proteica detallada. Los procesos de *modificación postraducciona*l, aunque resultan difíciles o imposibles de predecir a partir únicamente de las secuencias de proteínas, son cruciales en la estructura y función de muchas proteínas, en las interrelaciones génicas y en el control de rutas metabólicas. Si hablamos de la complejidad inducida por las modificaciones postraduccionales baste decir que mientras el *genoma* humano esta constituido por unos 30-40.000 genes (número provisional y todavía en discusión), el *proteoma* se especula que pueda estar conformado por 1,5-2.000.000 de proteínas. La diferencia es apreciable.

Se han aportado soluciones parciales para la caracterización de modificaciones postraduccionales mediante el uso de la espectrometría de masas y algunos softwares específicamente di-

señados para este fin, disponibles en la red publica de Internet (FindMod) (Wilkins et al. 1999). Las predicciones resultantes pueden ser comprobadas por espectrometría de masas mediante técnicas de fragmentación peptídica (p. ej. colisión por gas de iones atrapados, en ESI o por decaimiento posfuente, PSD, en MALDI-TOF o recientemente, por MALDI-TOF/TOF).

La identificación y caracterización de modificaciones co- y post-traduccionales puede considerarse un caso particular en el campo de la identificación de proteínas. Si bien algunas modificaciones co- y post-traduccionales pueden predecirse a partir de la secuencia de genes, la naturaleza, localización exacta y variabilidad (en respuesta a estímulos externos) de dichas modificaciones deben determinarse experimentalmente. Aunque para este fin se siguen empleando con profusión técnicas bioquímicas clásicas (anticuerpos monoclonales, lectinas, radio marcaje de fosfoproteínas, liberación química o enzimática de las modificaciones, HPLC, etc.), la EM muestra cada vez un mayor potencial para el análisis de modificaciones co- y post-traduccionales; en particular, la EM resulta especialmente valiosa cuando se conoce la secuencia de la proteína.

Aproximación proteómica al diagnóstico clínico

La genómica ha aumentado sustancialmente el conocimiento que se poseía sobre muchas enfermedades, abriendo nuevas avenidas terapéuticas. La **proteómica** es un área de investigación que conjuga factores genéticos y ambientales. *La composición proteica «proteoma» representa el estatus funcional de un compartimento biológico.* Una enfermedad se puede producir por un intercambio de un solo par de bases en un genoma. Sin embargo, la mayor parte de las enfermedades son más complejas y la elucidación de sus mecanismos patológicos más

complicada. Para comprender los mecanismos moleculares subyacentes al desarrollo de la enfermedad se precisan métodos adecuados a esta complejidad. Las proteínas, como macromoléculas efectoras, son los candidatos principales como moléculas diana en el desarrollo de enfermedades.

La estrategia proteómica antes descrita, se ha implantado con decisión para la detección de proteínas asociada a tumores: cáncer de colon, hepatocarcinomas (Jungblut et al. 1999), fibrosarcoma (Sinha et al. 1999), cáncer de vejiga (Celis et al. 1996, 1999), cáncer de mama (Giometti et al. 1997; Perou et al. 1999), cáncer de pulmón (Okuzawa et al. 1994), cáncer de riñón (Sarto et al. 1997), cáncer de ovario (Lawson et al. 1991), médula ósea (Hannash et al. 1982, 1993). Existen numerosos marcadores tumorales descritos (Carney, 1988). Sin embargo, sólo unos pocos se han revelado útiles en el diagnóstico y en el pronóstico clínico de supervivencia y recurrencia. Por ello, continua siendo de relevancia la descripción y caracterización de nuevos marcadores. Más recientemente se ha comenzado su aplicación a distintos y variados problemas biológicos, como son: *a*) el estudio de determinantes de patogenicidad y resistencia a antibióticos en microorganismos (Niimi et al. 1999; Cash et al. 1999); *b*) desórdenes y degeneración del sistema nervioso (Johnston-Wilson et al. 2001; Rohlf et al. 2000); *c*) patologías humanas organo-específicas (Cutler et al. 1999), entre otras.

III. Metodología

La estrategia experimental troncal (figura 1) es común para la mayoría de los proyectos proteómicos, básicamente supone:

1. Separación de proteínas por electroforesis bidimensional

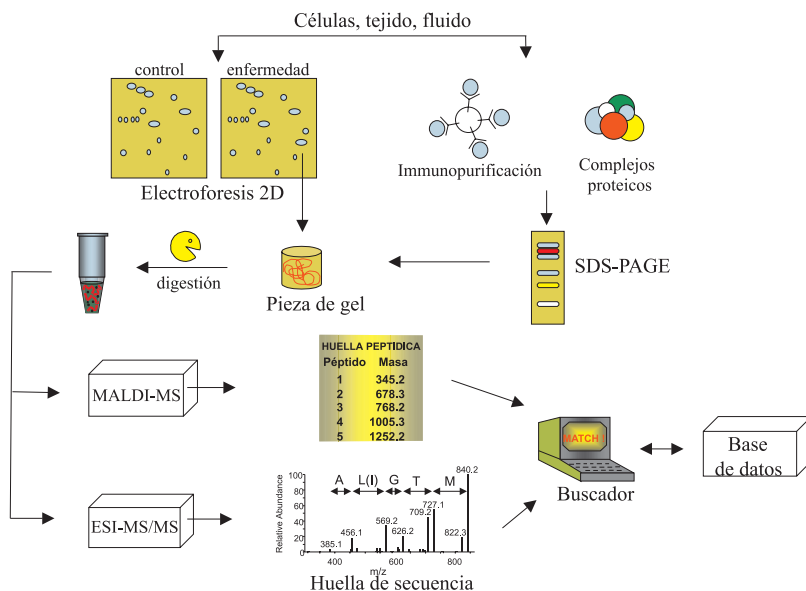


Figura 1: Estrategia analítica generalmente utilizada en análisis de proteómica.

El primer paso lo constituye la **electroforesis bidimensional** (2D-E) —centro neurálgico de la tecnología del proteoma—, método de elección en los grandes proyectos de proteómica para separar simultáneamente hasta miles de componentes individuales en mezclas complejas de proteínas (figura 2). La primera dimensión separa las proteínas en función de sus puntos isoeléctricos (electroenfoque), mientras que la segunda dimensión lo hace según su tamaño por electroforesis en gel de poliacrilamida en presencia de SDS (SDS-PAGE).

2. Obtención del patrón de expresión proteico y análisis de imágenes

Después de desarrollar la segunda dimensión de la electroforesis se visualizará el *patrón de expresión proteico* utilizando un amplio abanico de métodos de tinción (plata, azul de Coomassie, tinción inversa, fluorescencia, radioactividad) que con distintos grados de sensibilidad (por debajo del ng por mancha) permitirá *visualizar, escanear y analizar* las imágenes del patrón de expresión de proteínas. El análisis de imágenes (patrón de expresión) se llevará a cabo mediante el uso de programas informáticos (ImageMaster, Mellanie u otros) que facilitaran la comparación entre las distintas condiciones y con bases de datos de geles de 2-D existentes en el dominio público de la red de Internet.

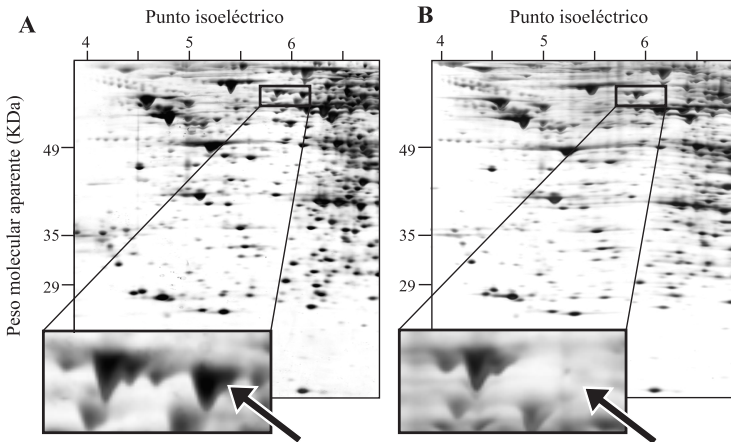


Figura 2: Patrón de expresión proteico obtenido mediante la separación de proteínas mediante electroforesis en gel bidimensional y tinción con plata. Las muestras corresponden a extractos de hígado de rata control (A) y de animal con una pancreatitis inducida (B). En la zona ampliada se muestra un ejemplo de expresión diferencial de una proteína descubierto mediante el análisis de imagen.

3. Identificación y caracterización de proteínas

La identificación de proteínas se lleva a cabo a partir del conocimiento de uno o varios atributos de las mismas que son ajustados frente a una o varias bases de datos.

3.1 La secuenciación N-terminal de proteínas

La secuenciación N-terminal de proteínas por degradación automática de Edman aún hoy, después de 30 años, constituye el método principal de identificación de proteínas gracias a su inambigüedad. Se pueden obtener datos sobre la secuencia de aminoácidos a partir de proteínas purificadas o por secuenciación interna de fragmentos proteolíticos, bien directamente o previa transferencia a membranas de PVDF. La secuencia N-terminal de una proteína puede ser utilizada para su identificación mediante búsquedas en bases de datos o para determinar si la secuencia representa una proteína identificada *de novo*.

3.2 Obtención de huellas peptídicas de proteínas (Peptide Mass Fingerprints)

Sin embargo, la estrategia de elección más importante en el proceso de identificación y caracterización para las proteínas elegidas de los geles de 2-D vendrá definida por la *espectrometría de masas (EM)*, que es una tecnología analítica esencial en el contexto de la proteómica debido a su alta capacidad de análisis, su sensibilidad y su precisión en la determinación de las masas moleculares de péptidos y proteínas, así como por la información sobre la composición molecular que se deriva de la masa molecular. Utilizaremos como primera y más importante opción para el análisis de mezclas complejas de digeridos proteicos la espectrometría MALDI-TOF, que utiliza una técnica de generación de iones por desorción/ionización median-

te láser inducida por matriz (Matrix-Assisted Laser Desorption/Ionisation, MALDI). Estos iones son acelerados en un campo eléctrico para a continuación penetrar en un tubo de recorrido o vuelo libre sin campo eléctrico alguno. Para un voltaje de aceleración dado, el tiempo de vuelo (Time-Of-Flight, TOF) que lleva a un ion alcanzar el detector es proporcional a su relación masa/carga, m/z .

La EM constituye una técnica relativamente nueva para el análisis de macromoléculas biológicas; por lo tanto, instrumentos, protocolos y aplicaciones se hallan en continua mejora y desarrollo.

3.3 Análisis de las masas moleculares de péptidos y proteínas (búsquedas en bases de datos)

Distintas características de las proteínas se utilizan para las búsquedas en las bases de datos: origen de la especie de la proteína, composición de aminoácidos, punto isoelectrico, masa molecular aparente en SDS-PAGE, etc. Sin embargo, la estrategia más utilizada para identificar una proteína dada es volcar en una base de datos de proteínas las masas provenientes de los digeridos proteicos (huellas peptídicas) obtenidas por espectrometría de masas con softwares específicos (ProFound, Mascot, p.ej.), donde se ajustan éstas con las que provienen de digerir teóricamente las distintas proteínas presentes en las distintas bases de datos.

Si el genoma del organismo o especie está secuenciado y su información es accesible a las bases de datos públicas esta búsqueda-rastreo dará un resultado que será una proteína candidata con un grado de fiabilidad determinado.

3.4 Obtención de espectros PSD o MS/MS

La comprobación de que la proteína candidata con mayor probabilidad sea inambiguamente la proteína aislada se realiza

por fragmentación de alguno de los iones moleculares obtenidos en el espectro de huellas peptídicas y que corresponda a una secuencia dada de la proteína candidata. Este proceso en espectrometría MALDI se denomina decaimiento posfuente (Post-Source Decay, PSD). El espectro de fragmentación así generado constituye la base para que programas bioinformáticos, que utilizan algoritmos de deconvolución adecuados, den información sobre la *secuencia peptídica* putativa y por tanto si la identificación proteica ha sido correcta. Obviamente, la accesibilidad a una instrumentación de espectrometría de masas que permita la obtención de espectros MS/MS reales (ESI ver más abajo, Q-TOF, o recientemente MALDI-TOF-TOF) facilitaría enormemente este aspecto de la comprobación-confirmación de la identificación además de resultar casi imprescindible en el análisis de las modificaciones postraduccionales.

3.5 Espectrometría de masas ESI-NSI

En determinadas circunstancias no podemos descartar el recurrir a la obtención de espectros de masas, tanto para huellas peptídicas como sobre todo para espectros de fragmentación, basados en otro tipo de fuente de ionización para generar los iones peptídicos: la ionización por electronebulización (Electrospray Ionisation, ESI). Esta técnica de ionización, que genera una amplia colección de iones metaestables (Gaskell, 1997) y por tanto más fácilmente fragmentables, exige una pureza extrema en la preparación de la muestra, requiriendo una purificación previa de ésta por cromatografía capilar de alta resolución (nano-HPLC); la confinación de los iones generados en una trampa iónica o cuadrupolo y su posterior fragmentación en una cámara de colisión, permite la obtención de información secuencial.

4. *Caracterización de las modificaciones co- y postraduccionales*

Las mutaciones puntuales y las modificaciones co- y postraduccionales pueden quedar de manifiesto en un análisis de masas por la diferencia que se observa entre la masa esperada y la obtenida experimentalmente. La huella peptídica, junto con la información parcial de secuencia, permite interpretar los desplazamientos de masa en términos de modificaciones en pequeños fragmentos de secuencia de la proteína. A pesar del enorme potencial que posee la EM en lo que respecta al estudio de modificaciones, sigue siendo necesario en algunos casos recurrir a técnicas bioquímicas clásicas para obtener una caracterización completa.

FindMod es un programa bioinformático, disponible desde la página ExpASy de Internet, diseñado para la identificación de modificaciones postraduccionales de proteínas. Analiza las masas del fingerprinting peptídico de proteínas conocidas para determinar la presencia de hasta 22 tipos de modificaciones postraduccionales de masas discretas como, entre otras, acetilación, amidación, biotinilación, C-manosilación, desamidación, tripalmitoilación, flavinización, farnesilación, formilación, geranil-geranilación, carboxilación, glicosilación, hidroxilación, lipoilación, metilación, miristoilación, palmitoilación, fosforilación y sulfatación. Si una diferencia de masa corresponde a una modificación conocida que aún no está anotada en SWISS-PROT, el programa aplica un conjunto de reglas «inteligentes» a la secuencia del péptido de interés y realiza predicciones acerca del aminoácido que más probablemente porte la modificación.

IV. Limitaciones

Hasta aquí se presenta una imagen más o menos ideal de la separación de proteínas por 2D-E y su identificación mediante degradación de Edman y técnicas de espectrometría de masas, fundamentalmente, MALDI-TOF. Obviamente, estas metodologías no están exentas de problemas, resultando difícil en algunos casos la interpretación de los resultados. Así, entre los principales problemas asociados a la 2D-E se encuentra la solubilización de las proteínas, tanto en lo referente al procedimiento inicial de extracción como al proceso de transferencia de proteínas desde las tiras de IPG a los geles de la segunda dimensión.

Por otro lado, aunque el uso de gradientes de pH inmovilizados ha mejorado notablemente la reproducibilidad de la primera dimensión, esta característica no está garantizada, necesitándose un alto número de geles para la consecución de réplicas en número estadísticamente relevante a fin de obtener geles promediados maestros. Otros problemas provienen de la sensibilidad de detección de proteínas y del poder de resolución de la técnica, principalmente en el rango básico de pH. El desarrollo de software adecuado para el análisis automático de imágenes ayudaría enormemente en los procesos de comparación de geles.

El bloqueo N-terminal de proteínas o péptidos, debido tanto a modificaciones naturales como a manipulaciones de la muestra, imposibilita en un buen número de casos la degradación de Edman. Aunque se han descrito métodos para el desbloqueo de muestras, no existen estrategias eficientes y aplicables de forma general. Más aún, la eficacia de la química de Edman no es del 100%, lo cual disminuye de facto la longitud de la secuencia que puede ser identificada satisfactoriamente e incrementa la cantidad teórica de proteína necesaria para obtener secuencia. Así, una banda de proteína electrotransferida

a membranas de PVDF y visualizada con Coomassie podría resultar insuficiente para secuenciación. Por otro lado, esta técnica consume reactivos caros. Todo esto hace hoy por hoy inusual la utilización extensiva de la degradación de Edman, de manera que la técnica se utiliza para la secuenciación de marcas terminales, cadenas polipeptídicas de corta longitud que se emplean como atributos para la identificación de proteínas.

Respecto a la espectrometría de masas, la masa experimental de proteínas aisladas en geles o membranas de PVDF puede ser diferente de la obtenida para proteínas puras como resultado de reacciones laterales como las que tienen lugar en Met o Trp (oxidación), Cys o grupos N-terminales libres (acrilamidación) o formación de iones aductos en presencia de altos niveles de sales, Coomassie o SDS.

En el fingerprinting peptídico, la interpretación del espectro de masas se complica porque no todos los fragmentos proteolíticos se extraen de igual manera, permaneciendo, tras la digestión, un número indeterminado de péptidos retenidos en los geles o membranas. Por tanto, la cobertura de péptidos que se ajustan a una proteína en las búsquedas o rastreos en bases de datos puede ser reducida, complicando el proceso de identificación; no obstante, una cobertura baja, por debajo del 30%, puede ser suficiente para la correcta identificación de una proteína. Además, buena parte de las posibles modificaciones de residuos particulares como cisteína, metionina, residuos N-terminales, etc. están previstas en las rutinas de búsquedas en las bases de datos. Habitualmente, los péptidos no ajustados se descartan sin análisis posteriores debido a la carencia de herramientas automáticas para analizarlos con seguridad; sin embargo, dichos péptidos podrían portar modificaciones como postraduccionales que de esta forma quedan generalmente sin caracterizar.

Por otro lado, el modo de ionización intrínseco al bombardeo por láser, puede producir fragmentaciones no deseadas en

la molécula polipeptídica (rotura de algunos enlaces lábiles) que hagan inviable el análisis fiable de determinados péptidos. De ser así, resulta recomendable el empleo de otros métodos de ionización basados en principios diferentes (p. ej., ESI).

La interpretación de los espectros PSD no es una tarea fácil. El espectro de fragmentación completo por MALDI-TOF se adquiere en varias etapas, ya que el reflector de iones analiza las masas iónicas dentro de un intervalo definido. Los iones fragmento producidos cubren una región de masas relativamente amplia, por lo que a fin de analizar los iones de menor masa el potencial del reflector debe ser disminuido progresivamente hasta que todos los iones producidos se adquieran con suficiente resolución másica.

Finalmente, los espectros PSD parciales se calibran, se solapan y se calculan las masas iónicas. Por otra parte, no todos los posibles iones pueden observarse claramente y algunos están ausentes; además, en algunos casos, los patrones de fragmentación son específicos de residuos, complicando más aún la interpretación de los espectros PSD. Por todo ello, ocasionalmente se recurre a técnicas que suministran información adicional sobre la secuencia peptídica, como el intercambio hidrógeno-deuterio o la acetilación N-terminal.

Por último, cabe señalar que la accesibilidad a cualquier banco de datos, ya sea genómico, de EST o de proteínas, no es todavía universal. Asimismo se han de desarrollar programas predictivos fiables para traducir secuencias genómicas en las secuencias proteicas que realmente se expresan.

V. Perspectivas (*«Proteómica de segunda generación»*)

La disponibilidad de tiras de IPG de rango de pH estrecho mejorará el poder de resolución de la 2D-E al expandir áreas específicas de pH. Será necesario avanzar en la capacidad de

focalización de las tiras de IPG en la zona básica si se requiere el análisis total de un proteoma. El uso de distintos fluorocromos para visualizar diferencialmente (mediante lectores de fluorescencia adecuados) proteínas provenientes de muestras —problema versus control en un mismo gel (DIGE)— supone un claro avance en el análisis comparativo de la expresión diferencial de proteínas. Por otro lado, para el análisis a gran escala de proteomas es preciso desarrollar una plataforma integrada que cubra el análisis de imágenes y la excisión y digestión de las manchas de proteína de los geles de 2D-E.

Sin embargo, las limitaciones intrínsecas derivadas de la técnica de separación (proteínas de peso molecular superior a 100 kDa, inferior a 10 kDa, de pI inferiores a 3,5 y superiores a 9, las expresadas a bajo nivel y sobre todo las hidrofóbicas) están propiciando la necesidad de incluir otras aproximaciones para complementar la ventana de análisis que ofrece la 2-D. La utilización de técnicas basadas en la cromatografía multidimensional (intercambio iónico, afinidad y fase reversa) en escalas capilar o nano, en conjunción con espectrometría MS/MS (Q-TOF Nano-ESI) suponen herramientas irrenunciables para garantizar una identificación en la escala baja de los fmoles-attomoles de proteínas. Así, estrategias de análisis de expresión diferencial de proteínas como la ICAT (Isotope Code Affinity Tag) (Gygi et al. 2000) que combina marcaje químico diferenciado isotópicamente de lisados celulares, cromatografía capilar multidimensional y espectrometría masas/masas se aplicaran si se tiene acceso a la instrumentación requerida: espectrómetro de masas Q-TOF, MALDI quadrupole Time-of-Flight o nano-electrospray (nESI).

Alternativamente, el fraccionamiento de muestras complejas de proteínas mediante cromatografía de retención sobre chips y posterior análisis de éstas por espectrometría de masas SELDI (Surface Enhanced Laser Desorption/Ionization-Time Of Flight) (Merchant y Weinberger 2000) supone una plata-

forma de análisis rápida, reproducible y robusta, extremadamente útil para el análisis comparativo de perfiles de expresión de proteínas en el rango bajo de los fmoles, que, en conjunción con las otras espectrometrías de masas, facilita enormemente la identificación de proteínas.

El proceso de degradación de Edman se estableció hace 30 años y la química involucrada no ha experimentado cambios fundamentales. Se han producido mejoras en cuanto al nivel de sensibilidad optimizando el tamaño de los componentes, los volúmenes de uso, la sensibilidad de los detectores, etc.

Hoy en día es posible secuenciar de forma fiable muestras en el rango de pocos picomoles de muestra. Se han desarrollado otros instrumentos auxiliares que ayudan en la secuenciación de fragmentos internos obtenidos por digestión enzimática, tales como el Microblotter de ABI, que posibilita la separación de péptidos por HPLC capilar, colectando los picos como microgotas sobre membranas de PVDF, porta-muestras para MALDI o en tubos para ser usados directamente por el secuenciador o el espectrómetro de masas, tanto MALDI como ESI.

A medida que avanzan los distintos proyectos de secuenciación de genomas se va generando una vasta cantidad de datos que necesitan ser recogidos, comprobados, analizados y clasificados. A este respecto, el Swiss Institute of Bioinformatics (SIB) y el European Bioinformatics Institute (EBI) están poniendo un gran énfasis en anotar, describir y distribuir entre la comunidad científica una gran cantidad de información relativa a secuencias de proteínas humanas tales como la anotación de proteínas humanas, polimorfismos humanos, modificaciones postraduccionales, ortólogos de proteínas humanas en mamíferos, etc. Esta iniciativa, conocida como Human Proteomics Initiative, demanda a la comunidad de usuarios una participación activa en el proyecto a todos los niveles.

VI. Importancia y aplicabilidad

En resumen, las distintas técnicas de la proteómica pretenden obtener información acerca de las proteínas celulares que puedan resultar afectadas como consecuencia de un estadio fisiopatológico alterado, y de si el ángulo de observación y su nivel de sensibilidad es el apropiado para este seguimiento.

La identificación y caracterización de modificaciones posttraduccionales en proteínas es un campo que se escapa a la información genética propiamente dicha, por lo que la evaluación de la aproximación proteómica a este área tiene una gran relevancia.

El análisis comparado de los patrones de expresión proteica de p. ej. tejidos tumorales es la base tanto para hallar potenciales marcadores asociados al desarrollo tumoral como para señalar futuras dianas terapéuticas útiles para la prevención del crecimiento de las células tumorales. Su estudio debe de conducir al establecimiento de métodos moleculares para clasificar el tipo tumoral y su invasividad, lo que, a su vez, puede tener importantes consecuencias para el diagnóstico y tratamiento del cáncer.

En conclusión, la electroforesis bidimensional, la cromatografía multidimensional, el análisis de imágenes, la química de Edman, las espectrometrías de masas MALDI-TOF, M-TOF-TOF, Q-TOF y ESI –Trampa iónica (en tandem con nano-HPLC), constituyen poderosas técnicas bioanalíticas que junto a las bases de datos generadas por los proyectos de secuenciación de genomas y las herramientas bioinformáticas son útiles para la separación, identificación, caracterización y catalogación de proteínas en el contexto de los objetivos de la proteómica. La reducida cantidad de proteína necesaria (rango de sub-pmoles) y los bajos requisitos de pureza para la identificación de las mismas hacen de la proteómica un campo de trabajo en continuo crecimiento.

BIOINFORMÁTICA: NUEVAS HERRAMIENTAS PARA GESTIONAR Y ANALIZAR LA INFORMACIÓN GENÓMICA Y PROTEÓMICA

RODERIC GUIGÓ

IMIM, UPF, CRG, Barcelona

GONÇAL FIRPO

Applied Biosystems, Barcelona

Resumen

Los continuos avances que las Ciencias de la vida han experimentado en la pasada década, no sólo han revolucionado muchos de los conceptos de lo que actualmente entendemos como genómica —para referirnos al estudio y medida de la información contenida en el ADN y su relación con la vida, la salud y las leyes de la herencia e identificación humana— y de la proteómica —para referirnos al estudio de las proteínas que poseen funciones relacionadas con las Ciencias de la salud, producidas como expresión de un determinado mensaje genético— sino que también han alterado y revolucionado las técnicas de estudio e investigación en estos campos.

Entre estas nuevas tecnologías cabe destacar: la automatización y robotización de los métodos de secuenciación del ADN, que permiten la obtención de la secuencia completa de los genomas de los organismos vivos o la secuencia parcial de aquellos de entre los genes codificados en este genoma, que se expresan en una determinada estirpe celular; la invención de los «ADN arrays», que nos permiten monitorizar la expresión

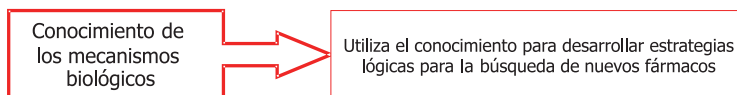
simultánea de miles de genes en condiciones distintas; el desarrollo de la PCR a tiempo real para la cuantificación precisa de la expresión génica y detección de SNPs; la mayor precisión de las técnicas de geles 2D y de espectroscopia de masas, que permiten la caracterización global de las proteínas en que estos genes son traducidos; técnicas como: la de ‘Doble Híbrido’ en levadura, que permiten inferir globalmente las interacciones entre estas proteínas, interacciones que sustentan los procesos de la vida; la automatización y robotización de las técnicas de rayos X y de resonancia magnética nuclear, que están acelerando substancialmente el descubrimiento de las estructuras tridimensionales de un gran número de estas proteínas.

La aplicación de estas tecnologías permite, por primera vez, la observación y la investigación del funcionamiento global de la célula viva. No se trata ya de obtener la secuencia de nucleótidos de un único gen, sino la secuencia completa del genoma. No se trata de determinar el grado de expresión de un único gen, sino la expresión global de todos los genes. No se trata de investigar la interacción entre dos proteínas, sino de determinar la red global de interacciones entre todas las proteínas que existen en la célula en un momento determinado. De esta manera, la biología, que, en cierto sentido, ha sido siempre una ciencia «analítica» en la que la realidad era disecionada en sus componentes más elementales para poder ser comprendida, ha pasado a ser una ciencia «sintética», en la que el reto consiste en integrar información global y heterogénea para convertirla en conocimiento sobre el funcionamiento de la célula. Por otra parte, y a raíz precisamente de su vocación de globalidad, la aplicación de estas técnicas genera una cantidad enorme de información, de una magnitud insólita en la historia de la biología. Y así, la biología, una disciplina en la que el mayor esfuerzo ha sido tradicionalmente dedicado a la obtención de los datos, ha pasado a ser, en pocos años, una disciplina en la que los datos se obtienen de modo prácticamente

automático y el esfuerzo debe dirigirse hacia su análisis. El caso de las matrices de ADN es paradigmático: hasta hace ciertamente muy poco, la investigación de la expresión de un gen bajo condiciones distintas era a menudo el trabajo de una tesis doctoral; representaba, en consecuencia, la labor de un investigador en el curso de tres o cuatro años. Ahora, unas pocas horas nos permiten disponer de resultados relativos al nivel de expresión simultáneo de decenas de miles de genes en condiciones distintas. La magnitud del cambio en la cantidad de información que genera la investigación en biología es abrumadora.

Este cambio se puede ejemplificar en el cambio de actitud que se está produciendo en investigación de nuevos fármacos.

Viejo Paradigma: Volumen pequeño de información rica en conocimiento;
Alto nivel de comprensión



Nuevo Paradigma: Gran volumen de datos con mucha información;
Bajo nivel de Conocimiento y comprensión



Ello lleva como consecuencia la necesidad de organizar correcta y eficazmente todos los datos obtenidos, sin tener demasiado claro cómo se utilizarán posteriormente. Esta organización se ha conseguido utilizando bases de datos relacionales capaces de manejar objetos y documentos. Y permite establecer libremente las relaciones y análisis estadísticos complejos entre las distintas fuentes de datos a posteriori.

En el Laboratorio de Investigación en Genómica o Pro-

teómica, tenemos actualmente herramientas que nos permiten una velocidad y productividad en el análisis de los datos muy superior a las que poseíamos hace poco tiempo. Se producen más y mejores datos, y más complejos. Se manejan secuencias de ADN, marcadores genéticos (microsatélites, SNPs), péptidos y proteínas... y se intentan correlacionar con los fenotipos y haplotipos de procedencia, analizando poblaciones enteras intentando determinar y obtener tratamientos válidos para disfunciones genéticamente transmisibles.

Es para responder a ese cambio abrumador que una nueva disciplina científica, la bioinformática, en la intersección entre biología y computación, ha emergido recientemente. Los objetivos de la bioinformática son, precisamente, la aplicación y el desarrollo de métodos para la obtención, el almacenamiento, el análisis y la integración de los datos que genera la investigación en biología.

La magnitud de la información que genera la investigación genómica es tal que probablemente supera la magnitud de la información que genera la investigación en otras disciplinas científicas. No en vano la vida es la forma más compleja de organización de la materia que conocemos. En estos momentos, los ordenadores no clasificados para uso civil más potentes del mundo —en Celera y en Oak Ridge National Laboratory, por ejemplo, con una capacidad de cálculo superior a los 2 Teraflops (billones de operaciones por segundo)— se encuentran ya dedicados a la investigación biológica, concretamente a la obtención y al análisis de las secuencias de nucleótidos de los genomas conocidos; IBM, por su parte, anuncia en un plazo de cuatro años un ordenador 500 veces más potente que «Deep Blue», el ordenador que en mayo de 1997 derrotó a Kasparov —acabando, de esta manera, con la hegemonía humana en el ajedrez—. Su nombre, «Blue Gene»; su objetivo, deducir, tras un año de cálculo, la conformación tridimensional de una pequeña proteína (de

entre las decenas de miles codificadas en nuestro genoma) a partir de su secuencia de aminoácidos.

Pero nos encontramos sólo en los inicios —balbuceantes— de la «era genómica». Tras el genoma de la especie humana, seguirá, por un lado, el genoma de otras especies y, por otro, el genoma de los individuos. El volumen de datos generados por estos proyectos será incomparable al volumen de datos que generan los proyectos genómicos hoy en día y que ya nos parece difícilmente tratable. Pero no se trata sólo de la información de secuencia; cada experimento con «DNA arrays» genera alrededor de unos de 60 Megabytes de información. Cientos, miles quizás, de dichos experimentos están siendo realizados estos días; centenares de miles, millones de ellos —cuando su uso como herramientas diagnósticas se generalice— se llevarán a cabo en un futuro no muy lejano.

Por otra parte, los conceptos tradicionalmente asociados a la investigación de nuevos fármacos, que pretenden asegurar la exactitud, calidad y reproductividad de los datos y conclusiones obtenidos, y que se expresan con claridad en las llamadas «*Buenas Prácticas del Laboratorio*» (o utilizando sus siglas inglesas, GLP, GMP, GCP... según se apliquen al entorno del laboratorio, de fabricación o clínico...) y en las nuevas normas como la *ISO17025* (que pretende asegurar que el instrumento de medida esté correctamente mantenido y calibrado y se conozca su precisión o, dicho de otra forma, su capacidad de medir correctamente) o la *21CFR parte 11* de FDA (que define como gestionar correctamente documentos electrónicos, controlando originales y copias, versiones y autores... y gestionar la «*firma electrónica*» como mecanismo de autenticación..), son cada vez mas de aplicación directa en los laboratorios de investigación genómica y proteómica, como mecanismo de garantía y aseguramiento de las conclusiones y decisiones tomadas.

La aplicación de estas normas se traduce, en primer lugar,

en un incremento enorme de la cantidad de datos generados: No basta con dar el dato. Hay que hacer constar también quién, cuándo, cómo, con qué instrumento y protocolo se ha obtenido. Y establecer su *trazabilidad* (de dónde proviene. Localización) y *auditabilidad* (¿se ha modificado? ¿Quién? ¿Cuándo? ¿Por qué?) como historial de cambios. Y mantener todos los «*datos originales*» obtenidos, no sólo los resultados finales o conclusiones, de forma que éstos se puedan reproducir.

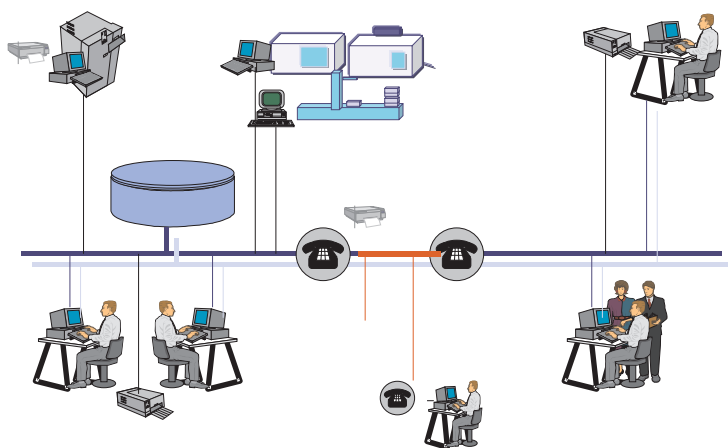
Uno de los aspectos prácticos que permiten un notable incremento en el control y flujo de información, es la conexión de los instrumentos de medida a la base de datos. No sólo ahorra esfuerzos y tiempo, también representa una drástica reducción de los errores de transcripción. Y permite, con el mismo coste, traspasar no los resultados, sino también toda aquella información asociada que permite aumentar el control, conocimiento y reproductividad de estos datos. Y en ello podemos incluir las imágenes gráficas producidas por la mayoría de técnicas: cromatogramas, espectros, imágenes de geles 2D, electroferogramas... Estas posibilidades modifican nuestra imagen del laboratorio, al centrarlo en la interconexión de instrumentos y personas, y convertirlo en un entorno distribuido de información completa y de calidad.

En este seminario se han presentado las soluciones que actualmente se pueden obtener y utilizar, tanto en el ámbito del control, gestión y organización de los datos en los laboratorios genómicos y/o proteómicos, como los avances y tendencias actuales en algoritmos y motores de búsqueda e interpretación de los datos, que intentan convertir la abrumadora cantidad de datos en información útil y manejable, relacionable con enfermedades o alteraciones de la salud y su posible corrección.

Las modernas herramientas de gestión del laboratorio proteómico son herramientas muy próximas al usuario final, gestionando la información en interfaces visuales, e interactuando directamente con las imágenes de geles 2D y los progra-

mas de gestión, normalización y análisis de dichas imágenes, visualizando los espectros de Masas o Masas-Masas relacionados con cada mancha o proteína, y que pueden automatizar la utilización de los motores de búsqueda e identificación de proteínas más populares.

Las modernas herramientas de gestión del laboratorio genómico o proteómico permiten que el investigador se centre en su trabajo y conclusiones, descargándolo de las tareas administrativas, de gestión y de documentación de cada fase de dicho trabajo, proporcionándole un nivel de confianza en su información, inimaginable hace unos años. Ello redundante, evidentemente, en un mejor aprovechamiento de las técnicas y algoritmos de interpretación y análisis y en una optimización general de los recursos tanto humanos como instrumentales y temporales dedicados a la investigación.



El volumen de información que genera la investigación genómica crece —y crecerá— a una velocidad vertiginosa. De hecho, mucho mayor que la velocidad que predice la famosa «ley de Moore», de acuerdo con la cual, la capacidad de los or-

denadores se duplica cada 18 meses —una tendencia que se inició a finales de los años cincuenta y que dura hasta nuestros días—. Parece ser que muy pocas actividades humanas crecen a un ritmo superior; una de ellas es la investigación genómica: a finales de los noventa, por ejemplo, el de tiempo necesario para que se duplicara el volumen de secuencias de nucleótidos almacenadas en las bases de datos públicas, era inferior a un año (a finales de 1999, estas bases de datos contenían alrededor 4.500 millones de nucleótidos. Un año más tarde, a finales del 2000, contenían alrededor de 11.000 millones). Este hecho tiene implicaciones trascendentales: la información genómica crece a una velocidad muy superior a la que crecen (es más, a la que pueden crecer) los recursos necesarios para analizarla. Y no se trata sólo, ni quizás principalmente, de recursos computacionales. Se trata, sobre todo, de recursos humanos.

Ante tal situación, uno de los retos de la Bioinformática es el desarrollo de métodos que permitan integrar los datos genómicos —de secuencia, de expresión, de estructura, de interacciones— para explicar el comportamiento global de la célula viva, que hagan necesarios menos recursos computacionales, pero también humanos.

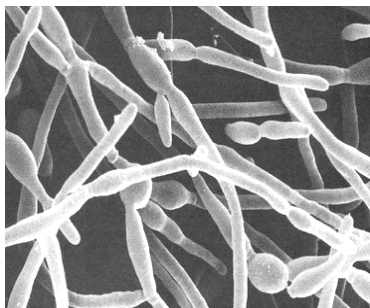
DIAGNÓSTICO, TRATAMIENTO Y PREVENCIÓN DE LAS INFECCIONES PRODUCIDAS POR EL HONGO PATÓGENO-OPORTUNISTA *CANDIDA ALBICANS*. ESTUDIO PROTEÓMICO.

CONCHA GIL

Facultad de Farmacia, Universidad Complutense de Madrid

Con la colaboración de Aida Pitarch, Lucía Monteoliva,
Mercedes Pardo, Gloria Molero, Rosalía Díez-Orejas
y César Nombela (UCM)

Las infecciones producidas por hongos constituyen actualmente un grave problema sanitario. Estas infecciones se han incrementado notablemente durante los últimos años debido a diversos factores como son los tratamientos con agentes inmunosupresores, con antibióticos de amplio espectro, utilización de catéteres y procedimientos quirúrgicos invasivos y una mayor supervivencia de los pacientes inmunocomprometidos. Entre los hongos patógenos más importantes (*Candida al-*



bicans, *Cryptococcus neoformans* y *Aspergillus fumigatus*) es *Candida albicans* el que presenta una mayor relevancia clínica.

Figura 1: Microscopía electrónica de filamentos de *C. albicans*.

Candida albicans es un hongo dimórfico, que forma parte de la microbiota de un gran número de personas (tracto gastrointestinal y vaginal) y en determinadas circunstancias puede producir infecciones que varían desde infecciones superficiales a generalizadas (invasivas o sistémicas). Estas últimas son las más graves y principalmente afectan a pacientes con deficiencias inmunológicas. La incidencia de las candidiasis invasivas se ha incrementado notablemente en las dos últimas décadas, constituyendo aproximadamente el 4% de las infecciones hospitalarias con un 30% de mortalidad.

El **diagnóstico** de las infecciones invasivas es difícil de realizar mediante criterios clínicos. Existen algunos ensayos serológicos para detectar antígenos o anticuerpos frente a este patógeno; sin embargo presentan poca sensibilidad y/o sensibilidad. Además, no son capaces de discriminar entre colonización e infección. Por tanto es necesario disponer de un ensayo sencillo, fiable y fácil de realizar para el diagnóstico de las infecciones invasivas producidas por *Candida*.

Otro problema importante que presentan las infecciones fúngicas es el de su **tratamiento**. Se dispone de un número muy reducido de agentes antifúngicos (antifúngicos poliénicos o azólicos) que o bien presentan efectos secundarios importantes o permiten la aparición de microorganismos resistentes. Es, por tanto, necesario descubrir nuevos agentes antifúngicos.

La tecnología proteómica constituye una herramienta muy importante para abordar toda la problemática anteriormente mencionada. La proteómica permite estudiar, identificar y caracterizar las proteínas expresadas por un genoma. Nuestro grupo de investigación ha desarrollado diversas estrategias proteómicas para identificar y estudiar proteínas que sean:

1. *Marcadores de diagnóstico*, que permitan desarrollar un método de diagnóstico sencillo, rápido y fiable de las candidiasis invasivas.

2. *Proteínas que induzcan anticuerpos protectores* para su posible utilización en inmunoprofilaxis o vacunación.
3. *Nuevas dianas para la búsqueda de nuevos agentes antifúngicos.*

Para identificar proteínas de utilidad en el diagnóstico se llevó a cabo la estrategia detallada en la figura 2.

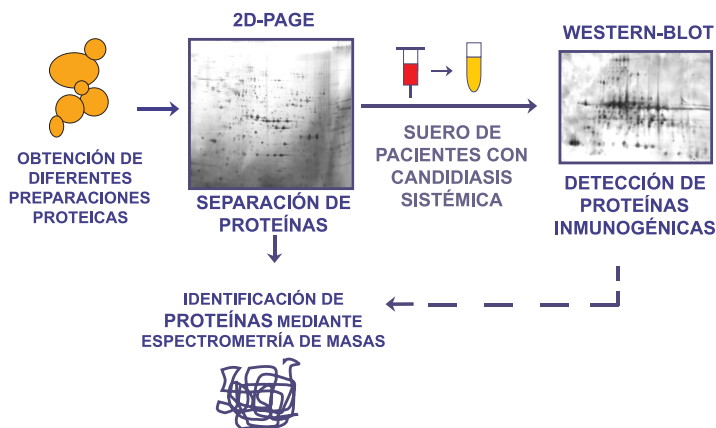


Figura 2: Estrategia para la identificación de proteínas antigénicas de *C. albicans*.

Se obtuvieron diferentes extractos proteicos de *C. albicans* (extractos celulares completos y diferentes fracciones de proteínas de la pared celular), y se separaron mediante electroforesis bidimensional (2D-PAGE). Dichas proteínas se transfirieron a membrana y mediante la técnica de «western-blotting» utilizando sueros de enfermos y sueros de ratón con candidiasis sistémica, se detectaron las proteínas inmunogénicas (Pitarch et al. 1999; Pitarch et al. 2001). Las proteínas inmunogénicas se identificaron mediante diferentes métodos: inmunodetección, secuenciación amino-terminal y principalmente mediante espectrometría de masas.

La identificación de las proteínas mediante espectrometría de masas se muestra en la figura 3. La identificación de las proteínas de interés se realizó mediante digestión con tripsina de la proteína. Los péptidos resultantes se analizaron en un espectrómetro de masas tipo MALDI-TOF, obteniéndose una huella peptídica de la proteína que se comparó con las huellas peptídicas de las proteínas presentes en las bases de datos. En caso de ambigüedad, los péptidos se analizaron mediante espectrometría de masas en tandem (MS/MS), obteniéndose la secuencia total o parcial del péptido y buscando en las bases de datos. Como el número de proteínas de *Candida albicans* presentes en bases de datos era bastante reducido (aproximadamente 400 proteínas en Swiss-Prot), algunas proteínas se identificaron por homología con las proteínas de *Saccharomyces cerevisiae*, una levadura muy relacionada y primer microorganismo del que se obtuvo la secuencia completa de su genoma en 1996

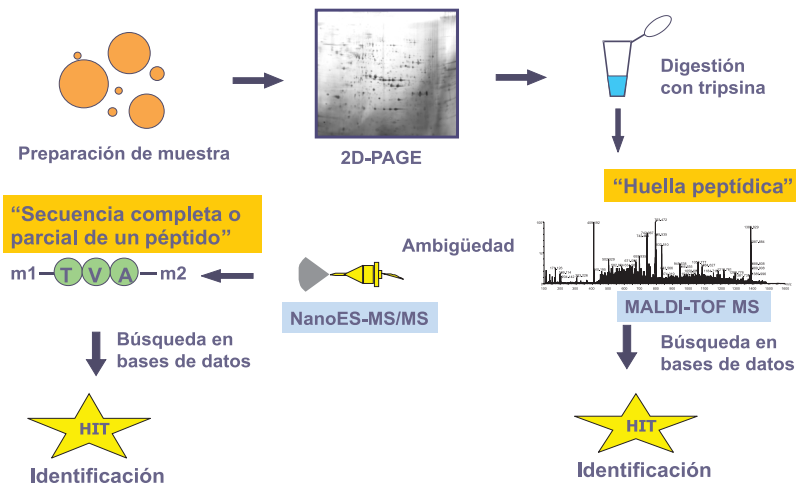


Figura 3: Procedimiento para la identificación de proteínas mediante espectrometría de masas

(Pardo et al. 2000b). Recientemente el genoma de *C. albicans* ha sido anotado.

Hasta el momento hemos identificado veintiuna proteínas inmunogénicas, doce de ellas son antígenos no conocidos hasta el momento. Estas proteínas inmunogénicas incluyen proteínas glucolíticas y del metabolismo, proteínas de choque térmico y proteínas de la pared celular. En la figura 4 se muestra un mapa de proteínas citoplasmáticas de *C. albicans* teñidas con plata donde están señaladas las proteínas inmunogénicas identificadas.

Algunas de estas proteínas se están expresando en la levadura *Pichia pastoris* y purificando para poder desarrollar un

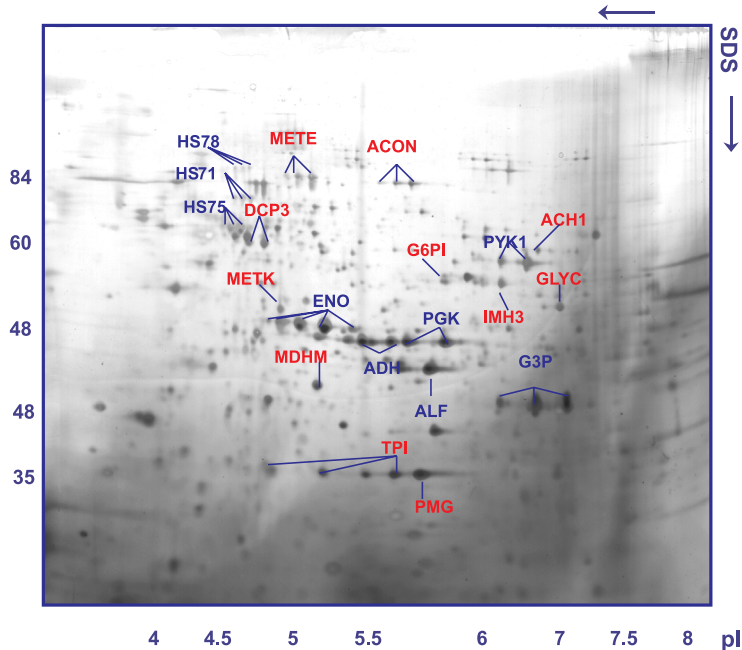


Figura 4: Mapa de proteínas de *C. albicans* separadas mediante 2D-PAGE y teñidas con plata. Las proteínas cuyo carácter antigénico se describe por primera vez están señaladas en rojo.

sistema de diagnóstico de la candidiasis sistémica. Dicho método de diagnóstico estaría basado en la realización de un western-blot o un ELISA (método inmunoenzimático) utilizando de 4-6 proteínas recombinantes de *C. albicans*.

Además estamos obteniendo anticuerpos policlonales frente a estas proteínas al objeto de estudiar su capacidad protectora frente a la candidiasis sistémica en un modelo de infección experimental en ratón. Dichos estudios pueden ayudar a estudiar el papel que juegan los anticuerpos en las infecciones sistémicas y podrían ser de utilidad en inmunoprevención y vacunación.

Otro aspecto importante que está desarrollando nuestro grupo de investigación, es la búsqueda de nuevas dianas de la célula fúngica susceptibles de ser inhibidas por **nuevos agentes antifúngicos**. Una de las estructuras que podría ser inhibida sin afectar a la célula animal es la pared celular (dicha estructura está ausente en las células animales). Se conoce con bastante detalle la composición química de la pared celular de las levaduras, pero aspectos como el ensamblaje de sus componentes resultan poco claros. La pared celular de las levaduras es una estructura compleja formada por polisacáridos (glucano y quitina) y manoproteínas, unidos entre si mediante enlaces covalentes. Las manoproteínas de la pared celular constituyen una especie de material de relleno de la red estructural. Dichas proteínas pueden estar débilmente unidas a la pared mediante enlaces no covalentes o mediante puentes disulfuro a otras proteínas, o unidas covalentemente al glucano y a la quitina (Molina et al. 2000). En la figura 5 se muestra un esquema de la estructura de la pared celular.

Debido a las dificultades de extracción de todas estas proteínas ideamos una estrategia para poder estudiarlas sin utilizar agentes químicos o enzimáticos. Este estudio se llevó a cabo en la levadura modelo *Saccharomyces cerevisiae*. Se realizó un tratamiento para eliminar la pared celular completamente y

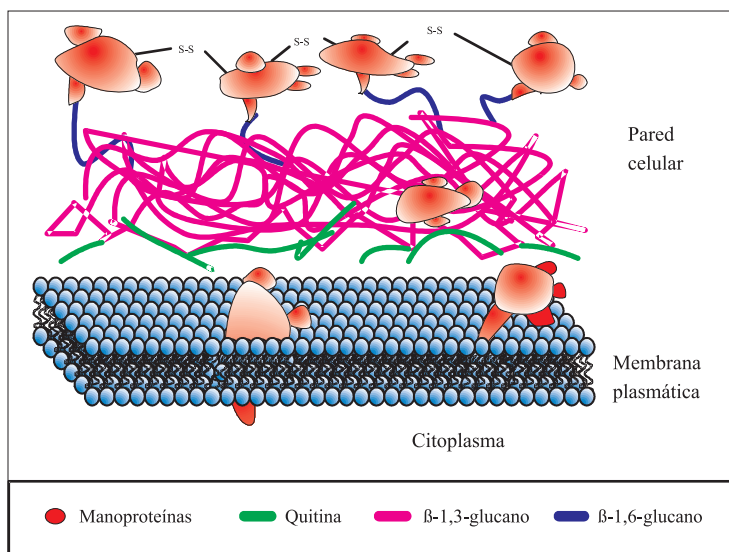


Figura 5: Estructura de la pared celular de *S. cerevisiae*.

obtener protoplastos. Dichos protoplastos se pusieron en condiciones de regeneración de la pared celular. En la figura 6 se muestra el aspecto que presentan los protoplastos. Como la célula está formando «de novo» la pared, va a haber una alta expresión de genes de pared y muchas proteínas van a pasar al

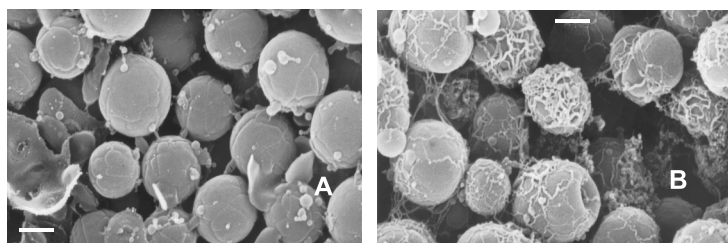


Figura 6: Microscopía electrónica de barrido de protoplastos de *S. cerevisiae* recién preparados (A) y tras 2 horas de incubación en medio de regeneración (B).

medio de cultivo y no van a quedar retenidas a los polisacáridos de la pared que se están formando. Realizamos un análisis proteómico de las proteínas secretadas al medio en condiciones de regeneración de la pared celular de los protoplastos (Pardo et al. 1999).

De esta forma pudimos identificar más de treinta y cinco proteínas relacionadas directa o indirectamente con la formación de la pared celular (Pardo et al. 2000a). Dichas proteínas se pueden clasificar: proteínas de pared celular o implicadas en su biosíntesis, enzimas glucolíticas o de la fermentación, proteínas de choque térmico y proteínas de función desconocida. Actualmente estamos realizando el análisis funcional de dos proteínas de función desconocida y que pertenecen a una misma familia tanto en *S. cerevisiae* como en *C. albicans*. Como último objetivo de este trabajo está la búsqueda de compuestos capaces de inhibir la función de estas proteínas y que se puedan utilizar en el tratamiento de las infecciones fúngicas.

Todas las proteínas identificadas en estos trabajos se encuentran en una base que hemos desarrollado y que está federada a la base de datos Europea SWISS- PROT, llamada **COM-PLUYEAST-2D-PAGE** <http://baggage.csc.ucm/2d/2d.html>

Para finalizar me gustaría señalar la enorme importancia que tiene la tecnología proteómica y las grandes expectativas que abre en el campo de la biomedicina. Sin embargo es muy importante tener en cuenta que su aprovechamiento depende del diseño de los experimentos con una clara hipótesis de trabajo, del conocimiento de las limitaciones de la tecnología y de la posibilidad de validar los descubrimientos utilizando aproximaciones experimentales alternativas.

Estas investigaciones están subvencionadas por la CICYT (SAF2000-0108), Comunidad de Madrid (Proyecto Estratégico) y Comunidad Europea (QLK3-CT-2000-01537).

PROTEÓMICA Y CALIDAD DE PRODUCTOS PESQUEROS

JOSÉ M. GALLARDO

Instituto de Investigaciones Marinas de Vigo (CSIC)

La producción pesquera mundial se sitúa alrededor de los 120 millones de toneladas anuales, de las que un 25% corresponde a la acuicultura. El consumo de productos pesqueros constituye en España una tradición, que nos coloca entre los primeros países del mundo, sólo detrás de Japón y Dinamarca. Una clásica dieta occidental contiene unos 25 gramos de pescado por día, frente a la dieta japonesa de 100 gramos diarios o los casi 80 gramos de la población española (30 Kg/persona/año), alejada así de la occidental, para parecerse más a lo que se denomina dieta típica mediterránea. De estos 30 Kg de consumo, unos 3,3 Kg/persona/año son de merluza y pescadilla frescas y 2,26 Kg/persona/año son de merluza y pescadilla congeladas, es decir, los merlúcidos son los principales productos pesqueros comercializados en España tanto en fresco como congelados.

En la sociedad actual la producción de alimentos en general y la pesquera en particular está alejada de los puntos de venta y consumo, por lo que teniendo en cuenta que los productos pesqueros son muy perecederos es necesario tomar medidas que eviten o ralenticen la alteración de los mismos, son necesarias operaciones como eviscerado, descabezado, pelado o fileteado y métodos de conservación encaminados a asegurar la calidad. Estas operaciones dan lugar a la desaparición de las características morfológicas, que hacen imposible la identificación de las especies marinas. A nivel mundial se

comercializan unas 7.000 especies de pescado de diversa procedencia, calidad y precio, por lo que tanto el consumidor como el importador y distribuidor comercial quieren conocer su identidad.

En los últimos años, las empresas procesadoras y comercializadoras de productos marinos han estado sometidas a una serie de factores de tipo socioeconómico como:

- a) La demanda, por parte de los consumidores, de productos alimentarios más frescos, sanos, más nutritivos y con una mayor información en el etiquetado de los mismos.
- b) El aumento de la competitividad de las empresas.
- c) Las nuevas normativas mundiales.

Hasta 1990, las metodologías empleadas en la autenticación de especies marinas se basaban en el análisis de proteínas, mediante técnicas electroforéticas, tales como: electroforesis nativa en geles de almidón, agarosa y poliacrilamida, isoelectroenfoque nativo (IEF), electroforesis en geles de poliacrilamida con SDS, isoelectroenfoque con urea, electroforesis capilar y técnicas inmunológicas.

En la última década del siglo pasado, la aplicación de las técnicas de biología molecular para la identificación de especies marinas ha supuesto un gran avance y los análisis de ADN comienzan a emplearse para identificar especies, especialmente en aquellos productos que han sufrido algún tratamiento térmico exhaustivo, como es el caso de las conservas de pescado, en las que se produce una desnaturalización de las proteínas.

El desarrollo reciente de la electroforesis bidimensional ha permitido la separación de mezclas complejas de proteínas, que con la implantación de la espectrometría de masas, ha supuesto un gran avance en el análisis y secuenciación de las proteínas. Fue en 1994 cuando se aceptó el término proteoma, que

fue definido en aquel momento como el conjunto de proteínas equivalentes a un genoma. Realmente un proteoma representa el patrón de proteínas de un organismo, una célula, un orgánulo o un fluido corporal, determinado en un cierto momento y en condiciones determinadas; y el conjunto de técnicas empleadas para el estudio del proteoma recibe el nombre de proteómica.

Aplicaciones de la proteómica

Identificación de especies de pescado

Las aplicaciones de las técnicas de proteómica en productos marinos son hasta el momento muy escasas. Los estudios existentes se centran en la secuenciación de proteínas por espectrometría de masas con fines fisiológicos, inmunológicos, con implicaciones biomédicas.

Desde 1997 el grupo de Química de Productos Marinos de Vigo (CSIC) en colaboración con el Laboratorio de Química de Proteínas del Centro de Biología Molecular «Severo Ochoa» están aplicando las técnicas de proteómica para la autenticación de especies marinas, lo que constituye una novedosa aplicación de las técnicas de proteómica. Al amparo de un Proyecto de Investigación financiado por la CICyT se están aplicando estas herramientas en la identificación de especies de merluza. El mapa de péptidos de una fracción de 17,5 kDa ha permitido diferenciar especies de merluza de las costas del Atlántico oriental de las del Atlántico occidental y Pacífico; y el mapa de péptidos de otra fracción correspondiente a otras proteínas, denominadas parvalbúminas, que son estables al calor, permite diferenciar la merluza austral del resto de especies. La resistencia al calor de las parvalbúminas y los diferentes isotipos existentes, dependiendo de la especie, motiva que las parval-

búminas sean apropiadas para investigar la autenticidad de especies en productos cocinados. El objetivo final de esta investigación es la obtención de anticuerpos, elaborados a partir de péptidos sintéticos, cuya composición se obtiene de la secuenciación de proteínas específicas mediante espectrometría de masas, que permitirían así la identificación de especies de merluza.

Frescura del pescado. Calidad post-mortem

La frescura es uno de los criterios más importantes para juzgar la calidad del pescado. La pérdida de frescura del pescado se debe a la combinación de procesos microbianos y bioquímicos. Las reacciones iniciales son hidrolíticas, catalizadas por enzimas endógenos, que dan lugar a metabolitos que permiten el posterior desarrollo microbiano; estos procesos proteolíticos afectan a la textura del pescado. Sin embargo, se ha prestado poca atención al seguimiento de los productos de la proteólisis y poco se sabe de los cambios de las proteínas durante el almacenamiento post-mortem. Estudios previos muy recientes han puesto de relieve que las tecnologías de proteómica pueden ser de gran utilidad para investigar las degradaciones de proteínas durante el almacenamiento refrigerado del pescado, con el objetivo de identificar proteínas o polipéptidos biomarcadores, que podrían emplearse en el control de la frescura y calidad del pescado.

Parásitos patógenos. Anisakiasis

La anisakiasis es una enfermedad causada por la ingesta de larvas de *Anisakis* simples, que es un nemátodo que infesta peces de agua salada y agua dulce. La vía de parasitación en humanos es la ingesta de pescados crudos o poco cocinados.

Las larvas de *Anisakis* nadan en el medio marino hasta que

son ingeridas por pequeños crustáceos como los eufásidos, que son los hospedadores intermedios, o bien a través de un hospedador de transporte como son los copépodos, de los que se alimentan estos crustáceos. Los eufásidos son a su vez ingeridos por los peces teleósteos y cefalópodos, como: merluza, palometa, besugo, caballa, abadejo, jurel, pota y pulpo.

Una vez ingeridas por el hombre, las larvas de *Anisakis*, para invadir la mucosa intestinal, utilizan un diente de penetración que está situado en la región cefálica y que ejerce una acción mecánica, y sus potentes enzimas proteolíticas capaces de degradar la matriz extracelular. Cuando las larvas de *Anisakis* invaden la pared gástrica o intestinal, se producen náuseas, vómitos y dolor epigástrico, y se conoce como la forma de anisakiasis intestinal, y también pueden producirse alergias en la forma gastroalérgica.

Existen numerosos estudios para caracterizar los antígenos implicados en estas reacciones. Cuando estos antígenos se someten a electroforesis de poliacrilamida, se observan alrededor de 20-30 bandas con pesos moleculares que oscilan entre los 12-120 kDa., habiéndose identificado y caracterizado proteínas implicadas en las reacciones alérgicas en humanos.

La electroforesis bidimensional combinada con western blotting con sueros de enfermos es útil para la detección de antígenos de *Anisakis*, que inducen la formación de anticuerpos en pacientes con anisakiasis. Con la electroforesis bidimensional se pueden separar mezclas de proteínas complejas, según su carga y peso molecular, y enfrentarlas a sueros de enfermos con anisakiasis para detectar los antígenos, cuya identificación se puede realizar mediante técnicas de proteómica.

Alergenicidad

Está demostrado que el consumo de pescado tiene efectos positivos sobre la salud, sin embargo se debe mencionar que hay una toxicidad natural, ya sea porque contienen sustancias potencialmente tóxicas, o porque se pueden producir en la conservación de los productos marinos o en su transformación, que pueden dar lugar a efectos no deseables en personas que presenten problemas farmacológicos o metabólicos.

En este sentido, son dignas de mencionar las alergias producidas, en personas sensibles, por las parvalbúminas que están presentes en los gádidos como merluza y bacalao. Se ha comentado con anterioridad que las parvalbúminas son buenos marcadores para la identificación de especies de merluza y de productos cocinados derivados, pero además las parvalbúminas se han investigado por su carácter alergénico, contrastado recientemente en la parvalbúmina beta de la merluza europea y en bacalao del Atlántico. Las técnicas de proteómica permiten estudiar la estructura inmunológica de estas proteínas alergénicas, por la modificación de ciertos residuos de aminoácidos y el examen de la reactividad de los derivados modificados y de la reactividad inmunológica de los péptidos obtenidos por digestión trípica selectiva.

Aunque la mayoría de los trabajos y aplicaciones de la proteómica están relacionados con la biomedicina, las aplicaciones en otros campos, como la calidad de los productos marinos, constituyen un reto actual de esta tecnología.

ENFERMEDAD CELÍACA, GLUTEN Y ESPECTROMETRÍA DE MASAS MALDI-TOF: NUEVA HERRAMIENTA PARA EL CONTROL DEL TRATAMIENTO DE LA POBLACIÓN CELÍACA

ENRIQUE MÉNDEZ

Unidad de Gluten, Centro Nacional de Biotecnología, Madrid

Con la colaboración de Alberto Hernando e Israel Valdés

La Enfermedad Celíaca es una intolerancia permanente al gluten de trigo, de cebada, de centeno y de avena. El único tratamiento posible para los pacientes celíacos es el seguimiento de por vida de una dieta estrictamente libre de gluten. De aquí la inquietud plenamente justificada de los celíacos por saber con exactitud la presencia o no de gluten en los alimentos, ya que de una dieta sin gluten depende su salud. De ahí la necesidad de desarrollar métodos fiables para la medición de gluten en alimentos. Convencionalmente se han venido empleando métodos inmunológicos tipo ELISAS o Western Blot, basados en anticuerpos, para medir el gluten de los alimentos. Debido a la gran complejidad de los alimentos, con frecuencia se originan falsos positivos en los ELISA, de aquí se deriva la necesidad de sistemas no inmunológicos alternativos y complementarios para confirmar la fiabilidad de los métodos inmunológicos.

Nuestro grupo ha sido pionero en la utilización de la espectrometría de masas MALDI-TOF (EM MALDI-TOF) como herramienta no inmunológica para la detección de gluten en

alimentos dietéticos para celíacos (Méndez et al. 1995; idem 2000; Camafeita et al. 1997; idem 1997b, idem 1998). A diferencia de ELISA, la EM MALDI-TOF detecta gluten de cebada y avena, además de trigo y de centeno.

El análisis consiste en una extracción de los componentes tóxicos del gluten de los alimentos con etanol al 60% seguido de un análisis directo por EM MALDI-TOF. La técnica permite analizar automáticamente los alimentos en tan solo unos minutos y detectar gluten con una sensibilidad de 5-10 partes por millón —semejante a los ELISA—. Desde hace tres años esta nueva herramienta se ha utilizado para la confirmación de gluten en más de 3.000 alimentos dietéticos para celíacos.

La observación directa de estos patrones característicos de trigo, cebada, centeno o avena en alimentos confirma la presencia de contaminaciones de gluten de estos cereales.

Los espectros de masas de los componentes tóxicos de los cuatro tipos de gluten, presentan unos perfiles muy distintos con rangos de masas fácilmente diferenciables: de 30-45 kDa en el trigo, 20-30 kDa en la avena, dos señales de masas de 32 y 39 kDa en el centeno, mientras que el gluten de cebada presenta señales de masas entre 30-65 kDa dependiendo de la variedad. En base a estos perfiles, la técnica de espectrometría de masas MALDI-TOF se puede utilizar para la identificación directa de gluten de trigo en alimentos elaborados con otros cereales como el maíz, arroz, o almidones, etc., en los que el trigo está como contaminante. Utilizando estándares de trigo, la EM MALDI-TOF permite además semicuantificar el gluten de trigo de los alimentos, aunque esta técnica se utiliza rutinariamente más para confirmación del gluten que como sistema de cuantificación. De la misma manera que el trigo, el gluten de cebada, de centeno y de avena se pueden identificar en alimentos elaborados con otros alimentos. El límite de detección de la EM MALDI-TOF es del orden de 5-10 partes por millón (ppm), muy cercana a la detección de los ELISA. En contraste

con los ELISA, la EM MALDI-TOF da información adicional de modificaciones y grado de hidrólisis de las proteínas del gluten como consecuencia del tratamiento por calor o enzimas proteolíticas a los que los alimentos son sometidos durante el proceso de elaboración.

Uno de los problemas de la analítica del gluten es que la mayoría de los alimentos dietéticos para celíacos son procesados a altas temperaturas (240 °C). Durante los procesos térmicos las subunidades α , β y γ del gluten que contiene puentes disulfuros se insolubilizan total o parcialmente y no se pueden extraer, mientras que las fracciones ω que no presentan puentes disulfuros sufren menos modificaciones.

En estos alimentos la confirmación de la presencia de gluten por la técnica de Western Blot se hace prácticamente imposible debido a la dificultad en la interpretación del perfil resultante de proteínas con esta técnica inmunológica.

Por el contrario, la técnica de EM-MT permite en tan solo unos minutos confirmar sin ninguna ambigüedad la presencia de gluten en estos alimentos procesados por calor por la simple observación del perfil de masas de las diferentes subunidades α , β , γ y ω . Los espectros de masas de harinas de maíz tostado que habitualmente consumen pacientes celíacos en la Comunidad Canaria muestran perfiles de masas con diferencias de concentración de las subunidades α , β , γ y ω desde espectros con ausencia total de subunidades α , β y γ a señales de masas exclusivamente de fracciones ω más o menos complejas (figura 1).

Otra de las limitaciones de la técnica de EM MALDI-TOF es la imposibilidad de detectar gluten en alimentos elaborados con cereales de maíz o arroz «no tóxicos para celíacos». Estos alimentos que utiliza el colectivo celíaco para su dieta libre de gluten están frecuentemente contaminados con gluten. La dificultad radica en que durante los procesos de extracción alcohólica de estos alimentos, el gluten (presente a niveles de trazas) se co-extrae junto con prolaminas de maíz o arroz que se

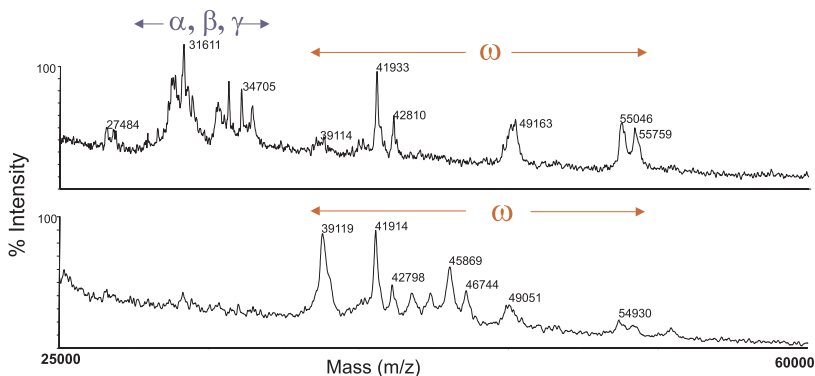


Figura 1: Dos perfiles MALDI-TOF correspondientes a alimentos procesados por calor conteniendo contaminaciones variables de gluten.

encuentran a altas concentraciones. Consecuentemente, los espectros de masas de los extractos alcohólicos de estos alimentos presentan predominantemente señales de masas muy intensas de prolaminas de maíz o arroz, mientras que las de gluten en general son indetectables.

Esta limitación de la técnica de EM MALDI-TOF la hemos resuelto desarrollando un procedimiento muy sencillo consistente en un tratamiento ácido del extracto alcohólico, que sorprendentemente produce un cambio muy selectivo de solubilidad de las prolaminas de arroz o de maíz. En condiciones ácidas los componentes del gluten permanecen solubles, mientras que las prolaminas de maíz o arroz son insolubles.

Por consiguiente hemos propuesto una nueva estrategia para la preparación de la muestra y posterior análisis por EM MALDI-TOF de gluten de alimentos elaborados con maíz o arroz, consistente en: 1) una extracción convencional con 60 % etanol en la cual el gluten y las prolaminas de arroz o maíz se co-extraen conjuntamente; 2) un tratamiento de pocos minu-

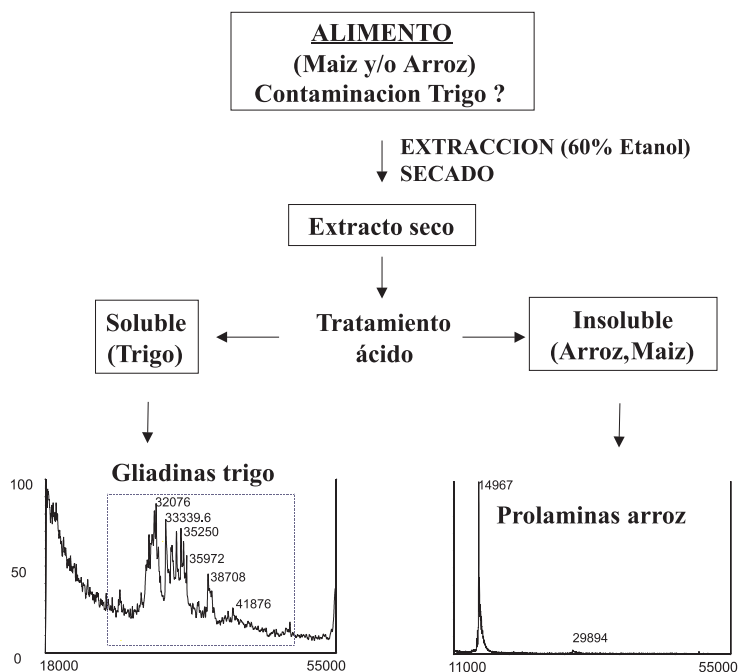


Figura 2: Caracterización mediante extracción ácida y análisis MALDI-TOF de la presencia de contaminaciones de harina de trigo en alimentos basados en maíz o arroz.

tos de una alícuota seca del extracto alcohólico en medio ácido; 3) centrifugación en un tubo eppendorf y separación del sobrenadante —donde selectivamente se encuentran los componentes del gluten—; y 4) análisis directo del sobrenadante ácido por EM-MT (figura 2). La eficiencia de este procedimiento ha sido demostrada en 20-30 alimentos elaborados con maíz y arroz con distintos grados de contaminación de gluten.

Probablemente la razón por la que el medio ácido produce cambios selectivos de solubilidad en algunos tipos de prolaminas (proteínas con un alto grado de prolínas y de glutaminas) puede ser debido a cambios conformacionales y/o

desaminaciones de las múltiples glutaminas presentes en todos los múltiples componentes proteicos de las prolaminas. Esta última parece ser la mas probable. Si la desaminación tiene lugar, la conversión de glutamina (Q, $-\text{CONH}_2$, 16 Da) a ácido glutámico (E $-\text{COOH}$, 17 Da) originaría un incremento de masa de 1Da/Q modificada. Por lo tanto la masa molecular de cadena polipeptídica desaminada por la acción del ácido experimentalmente un incremento de masas dependiendo del número de Q convertidas en E. Datos preliminares han permitido confirmar incrementos de masas entre 6 y 37 Da en las prolaminas de maíz y de arroz tratadas con ácido y que son insolubles en ese medio.

BIBLIOGRAFÍA DE LAS PONENCIAS

- Camafeita E, Alfonso P, Acevedo B y Méndez E J. *Mass Spectrom* (1997) 32, 444-449.
- Camafeita E, Alfonso P, Mothes T y Méndez E J. *Mass Spectrom* (1997b) 32, 940-947.
- Camafeita E y Méndez E J. *Mass Spectrom* (1998) 33, 1023-1028.
- Camafeita E, Solís J, Alfonso P, López JA, Sorell L y Méndez E J. *Chrom* (1998b) A823, 299-306.
- Cash P et al. *A proteomic analysis of erythromycin resistance in Streptococcus pneumoniae*. *Electrophoresis* (1999) 11, 2259-68.
- Carney W. *Human tumor antigens and specific tumor therapy*. *Immunol Today* (1988) 9, 363-364.
- Celis JE et al. *Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas*. *Cancer Res* (1996) 56, 4782-4790.
- Celis JE et al. *A comprehensive protein resource for the study of bladder cancer*: <http://biobase.dkcg-bin/celis>. *Electrophoresis* (1999) 20, 300-309.
- Cutler P et al. *An integrated proteomic approach to studying glomerular nephrotoxicity*. *Electrophoresis* (1999) 18, 3647-58.
- Gaskell SJ. *Electrospray: Principles and practice*. *Mass Spectrom* (1997) 32, 677-688.
- Gygi SP, Rist B, Aebersold R. *Measuring gene expression by*

- quantitative proteome analysis*. Current Opinion in Biotechnology (2000) 11, 396-401.
- Giometti CS et al. *A two-dimensional electrophoresis data base of human breast epithelial cell proteins*. Electrophoresis (1997) 18, 573-581.
- Hannash SM et al. *Two-dimensional gel electrophoresis of cell proteins in childhood leukemia, with silver staining: a preliminary report*. Clin Chem (1982) 28, 1026-1030.
- Hannash SM et al. *Data base analysis of protein expression patterns during T-cell ontogeny and activation*. Proc Natl Acad Sci USA (1993) 90, 3314-3318.
- Johnston-Wilson NL et al. *Emerging technologies for large-scale screening of human tissues and fluids in the study of severe psychiatric disease*. Int J Neuropsychopharmacol (2001) 1, 83-92.
- Jungblut PR et al. *Proteomics in human disease: Cancer, heart and infectious diseases*. Electrophoresis (1999); 20:2100-2110.
- Kyrpides N. *Genomes Online Database (GOLD): A monitor of complete and ongoing genome projects world wide*. Bioinformatics (1999); 15:773-774.
- Lawson SR et al. *Quantitative protein changes in metastatic versus primary epithelial ovarian carcinoma*. Gynecol Oncol (1991); 41:22-27.
- Lin X et al. *Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana*. Nature (1999) 402, 761-8.
- Luo L et al. *Gene expression profiles of laser-captured adjacent neuronal subtypes*. Nat Medicine (1999) 5, 117-122.
- Mann M, Hojrup P y Roepstorff P. *Use of mass spectrometric molecular weight information to identify proteins in sequence databases*. Biol Mass Spectrom (1993) 22, 338-45.
- Mann M, Talbo G. *Developments in matrix-assisted laser desorption/ionization peptide mass spectrometry*. Curr Opin Biotechnol (1996) 7, 11-19.

- Méndez E, Camafeita E, S.-Sebastián J, Valle Y, Solís J, Mayer-Posner FJ, Suckau D, Marfisi C y Soriano F. *Rapid Commun. Mass Spectrom.* (1995) 123-128.
- Méndez E, Valdés I y Camafeita E. *Methods in Molecular Biology* (2000) 146. 355-367, New Mass Spectrometric Applications. Edited by: J.R. Chapman, Humana Press Inc. Totowa, N.Y.
- Merchant M y Weinberger RS. *Recent advancements in surface-enhanced laser desorption ionisation-time of flight mass spectrometry.* *Electrophoresis* (2000); 21:1164-77.
- Molina M, Gil C, Pla J, Arroyo J y Nombela C. *Protein localisation approaches for understanding yeast cell wall biogenesis.* *Microscope Research and Technique* (2000) 51, 601-612.
- Nelson PS et al. *Comprehensive analyses of prostate gene expression: Convergence of expressed sequence tag databases, transcript profiling and proteomics.* *Electrophoresis* (2000) 21, 1823-1831.
- Niimi M et al *Candida albicans pathogenicity: a proteomic perspective.* *Electrophoresis* (1999) 11, 2299-308.
- Okuzawa K et al. *Characterization of gene expression in clinical lung cancer materials by two-dimensional polyacrylamide gel electrophoresis.* *Electrophoresis* (1994);15:382-390.
- Pardo M, Monteoliva L, Pla J, Sánchez M, Gil C y Nombela C. *Two-dimensional analysis of proteins secreted by Saccharomyces cerevisiae regenerating protoplasts: A novel approach to study the cell wall.* *Yeast* (1999) 15, 459-472.
- Pardo M, Ward M, Bains S, Molina M, Blackstock W, Gil C y Nombela C. *A proteomic approach for the study of Saccharomyces cerevisiae cell wall biogenesis.* *Electrophoresis* (2000a) 21, 3396-3410.
- Pardo M, Ward M, Pitarch A, Sánchez M, Nombela C, Blackstock W y Gil C. *Cross-species identification of novel Candida albicans immunogenic proteins by combination of 2D-*

- PAGE and mass spectrometry. *Electrophoresis* (2000b) 21, 2651-2659.
- Perou CM et al. *Distinctive gene expression patterns in human mammary epithelial and breast cancers*. *PNAS* (1999) 96, 9212-9217.
- Pitarch A, Pardo M, Jiménez A, Pla J, Gil C, Sánchez M y Nombela C. *Two-dimensional gel electrophoresis as analytical tool for identifying Candida albicans immunogenic proteins*. *Electrophoresis* (1999) 20, 1001-1010.
- Pitarch A, Díez-Orejas R, Molero G, Pardo M, Sánchez M, Gil C y Nombela C. *Analysis of the serologic response to systemic Candida albicans infection in a murine model*. *Proteomics* (2001) 1, 550-559.
- Roepstorff P. *Mass spectrometry in protein studies from genome to function*. *Curr Opin Biotechnol* (1997) 8, 6-13.
- Rohlf C. *Proteomics in molecular medicine: applications in central nervous systems disorders*. *Electrophoresis* (2000) 6, 1227-34.
- Sánchez JC et al. *Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients*. *Electrophoresis* (1997) 18, 324-7.
- Shevchenko A et al. *Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels*. *Proc Natl Acad Sci U S A* (1996) 93, 14440-5.
- Sarto C et al. *Renal cell carcinoma and normal kidney protein expression*. *Electrophoresis* (1997) 18, 599-604.
- Sinha P et al. *Search for novel proteins involved in the development of chemoresistance in colorectal cancer and fibrosarcoma cells in vitro using two-dimensional electrophoresis, mass spectrometry and microsequencing*. *Electrophoresis* (1999) 20, 2961-2969.
- Traini M et al. *Towards an automated approach for protein identification in proteome projects*. *Electrophoresis* (1998) 19, 1941-9.

- Velculescu VE et al. *Characterization of the yeast transcriptome*. Cell (1997) 88, 243-251.
- Wilkins MR et al. *Progress in proteome projects: why all proteins expressed by a genome should be identified and how to do it*. Biotechnol Genet Eng Rev (1995) 13, 19-50.
- Wilkins MR et al. *High-throughput mass spectrometric discovery of protein post-translational modifications*. J Mol Biol (1999) 289, 645-57.

Otras lecturas introductorias recomendadas

- Carol Ezzel. *La proteómica en el horizonte*. Investigación y Ciencia (2002) 309 (junio), 46-53
- Thomas Maeder. *Glucómica*. Investigación y Ciencia (2002) 312 (septiembre), 6-12

Páginas web de interés

Human Genome Program: www.ornl.gov/hgmis
ExPASy (Expert Protein Analysis System): www.expasy.org
HUPO: www.hupo.org

AUTORES

Juan Pablo Albar
Laboratorio de Proteómica
Centro Nacional de Biotecnología, CSIC
Cantoblanco, E-28049 Madrid
Tel: (34) 91 585 46 96
Fax: (34) 91 585 45 06
E-mail: jpalbar@cnb.uam.es

Roderic Guigó
Grup de Recerca en Bioinformàtica mèdica
IMIM
Doctor Aiguader, 80
E-08003 Barcelona
Tel: (34) 93 225 75 67
Fax: (34) 93 221 32 37
E-mail: rguigo@imim.es

Gonçal Firpo
División Bioinformática
Applied Biosystems
Parc Científic de Barcelona
Josep Samitier, 1-5
E-08028 Barcelona
Tel: (34) 93 447 02 15
Fax: (34) 93 447 02 17
E-mail: Gonzalo_Firpo@eur.appliedbiosystems.com

Concha Gil
Departamento Microbiología II
Facultad de Farmacia
Universidad Complutense de Madrid
Plaza Ramón y Cajal, s/n
E-28040 Madrid
Tel: (34) 91 394 17 48

Fax: (34) 91 394 17 45
E-mail: conchagil@farm.ucm.es

José M. Gallardo
Instituto Investigaciones Marinas-CSIC
Eduardo Cabello, 6
E-36208 Vigo
Tel: (34) 986-23 19 30
E-mail: gallardo@iim.csic.es

Enrique Méndez
Laboratorio de Proteínas
Centro Nacional de Biotecnología, CSIC
Cantoblanco, E-28049 Madrid
Tel: (34) 91 585 48 42
Fax: (34) 91 585 45 06
E-mail: emendez@cnb.uam.es

Joaquín Abián
Unidad de Espectrometría de Masas Estructural y Biológica
Institut d'Investigacions Biomèdiques de Barcelona
CSIC-IDIBAPS
Rosselló 161, 6ª planta
E-08036 Barcelona
Tel: (34) 93 363 83 12
Fax: (34) 93 363 83 01
E-mail: jambam@iibb.csic.es

Susanna Baqué
Applied Biosystems
Parc Científic de Barcelona
Josep Samitier, 1-5
E-08028 Barcelona
Tel: (34) 93 447 02 15
Fax: (34) 93 447 02 17
E-mail: Susanna_Baque@eur.appliedbiosystems.com

Sumario

Proteómica: herramientas de la era posgenómica en la vida cotidiana. Joaquín Abián y Susanna Baqué	7
Identificación y catalogación de proteínas: primeros objetivos de la era posgenómica. Juan A. López, L. Emilio Camafeita, Enrique Calvo, Ana Beloso y Juan P. Albar	11
Bioinformática: nuevas herramientas para gestionar y analizar la información genómica y proteómica. Roderic Guigó y Gonçal Firpo	31
Diagnóstico, tratamiento y prevención de las infecciones producidas por el hongo patógeno-oportunista <i>candida albicans</i> . Estudio proteómico. Concha Gil (con la colaboración de Aida Pitarch, Lucía Monteoliva, Mercedes Pardo, Gloria Molero, Rosalía Díez-Orejas y César Nombela)	39
Proteómica y calidad de productos pesqueros. José M. Gallardo	47
Enfermedad celíaca, gluten y espectrometría de masas MALDI-TOF: nueva herramienta para el control del tratamiento de la población celíaca. Enrique Méndez (con la colaboración de Alberto Hernando e Israel Valdés)	53
Bibliografía de las ponencias	59

