

La Proteómica, otra cara de la genómica

Tobías Mojica Ph.D., Oscar Sánchez y Leonardo Bobadilla

Instituto de Genética, Departamento de Biología, Facultad de Ciencias, Departamento de Ingeniería de Sistemas, Facultad de Ingeniería, Universidad Nacional de Colombia, Bogotá.

Recibido: 07-07-03; Aceptado: 24-09-03

Las proteínas son lo que uno podría llamar los arquitectos de la vida, pues son cruciales en los procesos celulares de todos los seres vivos. Las proteínas están implicadas en la catálisis de las reacciones químicas celulares, el transporte de moléculas, la transducción de señales, la segregación del material genético, la producción y el manejo de la energía. El programa celular vital necesita del trabajo coordinado de muchos tipos diferentes de proteínas (1). La mayor parte del peso seco de una célula está constituida por proteínas. Parece una tautología, pero tendremos que entender las proteínas antes de que podamos entender la célula.

La síntesis de todas las proteínas celulares está codificada por los genomas. En la actualidad, tenemos disponibles las secuencias de más de 150 genomas celulares, incluyendo varios de organismos multicelulares (<http://www.tigr.org>).

La proteómica es la nueva etapa en la investigación biológica que emana naturalmente de la genómica y que incluye la caracterización de la expresión de las proteínas codificadas por un genoma y el establecimiento de sus propiedades funcionales y estructurales (2). En cierta manera, la proteómica es la continuación de lo que se ha llamado la geonómica o ciencia de los genomas que se ha encargado de establecer la secuencia de muchos genomas. Los genomas definen el contenido informacional de los organismos y, por lo tanto, definen la tipología del organismo. La secuencia del genoma no le dice a uno de qué manera funciona el organismo; para tener respuestas a la pregunta de cómo funciona el organismo es necesario estudiar las proteínas, es decir, es necesario acercarnos a la proteómica. Para dar una idea de la complejidad del problema de la proteómica, consideremos que el genoma de levaduras tiene 6.225 proteínas calculadas por bioinformática y por otros medios. Por

anotación bioinformática, se calcula que el 17% de esas proteínas están involucradas en el metabolismo general de la célula de levadura, el 30% funciona en organización celular y biogénesis de organelos y membranas y el 10% en transporte molecular (3).

Los tópicos grandes de la proteómica son los siguientes:

- a. Perspectiva de la empresa, incluyendo la perspectiva de la relación entre genoma y proteoma.
- b. Fuente y manejo de las proteínas, incluyendo el organismo donante y la extracción y almacenamiento de la muestra, etc.
- c. Separación de las proteínas, incluyendo, como en el caso de las proteínas subcelulares, la purificación.
- d. Identificación de las proteínas que se realizan actualmente por espectrometría de masas.
- e. Función de las proteínas, lo cual incluye, además de la función propiamente dicha, la localización celular, interacciones proteína-proteína, determinación de la estructura terciaria, en alto grado alcanzado por cristalografía, la relación entre estructura y función y modificaciones postraduccionales.

f. Aplicaciones que pueden estar en el área del diagnóstico de las enfermedades humanas y en el descubrimiento de nuevas drogas.

g. Aplicaciones en informática que generan muchas anotaciones basadas en la homología de las secuencias, la construcción de bases de datos, generación de algoritmos para el análisis y, por último, la estandarización.

h. Otras áreas de importancia humana que incluyen la colaboración internacional, consideraciones éticas, legales, papel de la sociedad en la construcción y uso del conocimiento nuevo que resulte.

Gran parte de la empresa de investigación en biología contemporánea es llevada a cabo por países avanzados, con economías y personal adecuado. La proteómica no es la excepción. Los Institutos Nacionales de Salud de los Estados Unidos financian una actividad investigativa llamada la Iniciativa de la Estructura de Proteínas, llevada a cabo en 9 centros universitarios distribuidos por toda la unión. Este es un esfuerzo coordinado a 10 años para producir estructuras completas de 10.000 proteínas. La actividad está en sus comienzos y se considera un blanco importante y difícil llegar a la cristalización de 200 proteínas por año. Laboratorios de industrias que esperan ganar dinero con el desarrollo en conocimiento también han proliferado recientemente. El problema radica en que hay casi tantas proteínas en la Tierra como estrellas en el cielo. La iniciativa de la estructura de proteínas ha llevado al desarrollo de nuevas tecnologías que incluyen estudios estructurales detallados por nanotecnologías y procesos rápidos de cristalización. Esta es claramente una muestra de la enorme dimensión de la tarea, que contrasta con el rezago tecnológico de nuestros países.

Podemos entonces preguntarnos ¿qué es un proteoma? Se trata del complemento proteínico expresado por un genoma o por un tejido. Los proteomas son dinámicos y cambian con el tiempo, con el estado del desarrollo y con las condiciones intra y

extracelulares. La proteómica es el estudio de los proteomas; separa, identifica y caracteriza proteínas a gran escala, define niveles de proteínas celularmente, investiga complejos de proteínas, elucida funciones, caminos metabólicos e interrelaciones. La proteómica tiene que analizar un número muy grande de proteínas (4) (en organismos eucarióticos es usualmente un número mayor que el número de genes presentes en el genoma); también se encarga de la caracterización funcional de tales proteínas y de sus relaciones estructurales. En esencia, podemos decir que la genómica produce las estructuras primarias de las proteínas, mientras que la proteómica se encargará de producir las estructuras secundaria, terciaria y cuaternaria.

La información fluye en forma lineal del DNA genómico al RNAm por dos conjuntos de procesos contiguos llamados transcripción y procesamiento y sigue fluyendo linealmente del RNAm a la proteína, mediante un conjunto de procesos llamados traducción (5). El resultado de ese flujo de información lineal es la estructura primaria de la proteína y de todas las proteínas de un genoma. A partir de la estructura primaria de la proteína, ocurren tres hechos en tres dimensiones: todas las proteínas se pliegan en estructura secundaria y terciaria, algunas sufren modificaciones postraduccionales y algunas adquieren estructura cuaternaria. Hasta hoy, el pensamiento genético ha sido lineal. La proteómica empieza donde empiezan los eventos que tenemos que entender en tres dimensiones. Ni la estructura primaria del gene ni la del RNAm ni de la proteína misma dan indicaciones claras de la naturaleza del producto proteico ni de su función; por lo tanto, la primera tarea de la proteómica consiste en identificar las proteínas producidas por un genoma. Es decir, principalmente problemas tecnológicos frente a cómo se pueden separar y visualizar las proteínas en un proteoma, cómo se puede utilizar esta información para estudiar complejos proteicos y caminos metabólicos, cómo se

pueden identificar proteínas separadas y cómo se pueden caracterizar en detalle tales proteínas.

Las principales tecnologías en desarrollo son las siguientes: tecnología reproducible de electroforesis en geles en 2D; tecnología de tinción y análisis de los geles; tecnología para la identificación de las proteínas, tal como la espectrometría de masas (2); tecnologías de manejo electrónico, tales como bases de datos (de genomas y de proteínas); y, también, algoritmos de búsqueda y de recuperación de datos. El reto principal de la proteómica es la automatización y la integración de las tecnologías mencionadas. Este es un reto cuya respuesta saldrá, principalmente, de la bioinformática.

La electroforesis es la técnica central para analizar un número grande de proteínas al mismo tiempo (3). La técnica electroforética es relativamente antigua, fue inventada por Tiselius hace más de 50 años y, en 1956, Smithies y Poulick describieron el primer gel en dos dimensiones, mientras que en 1975, Patrick O'Farrell optimizó el proceso de separación en 2D, técnica que se utiliza hoy en día. En la primera dimensión se efectúa el isoelectroenfoco utilizando gradientes inmovilizados de pH, es decir, las proteínas se separan por su carga. En la segunda dimensión, la separación de proteínas se hace en presencia de SDS y las proteínas se separan por su masa. Se pueden separar miles de proteínas en un solo gel. La correcta preparación de la muestra es frecuentemente el paso más importante para un resultado exitoso.

Las manchas en el gel son extraídas y cada una es tratada con tripsina para producir un patrón característico de péptidos, y éstos se identifican por espectrometría de masas, una técnica que mide la masa (en este caso, la masa de los péptidos producidos por digestión con tripsina) de las moléculas con gran precisión. Los espectrómetros de masa tienen dos partes: una fuente de iones (como por ejemplo, MALDI o ESI, ver el siguiente párrafo) y un aparato medidor (como por ejemplo, cuadrupolos, trampas

iónicas y TOF). Las combinaciones de esta técnica que más se utilizan son las siguientes:

- MALDI-TOF-MS (Matrix Assisted Laser Desorption Ionization - Time Of Flight Mass Spectrometry) o Espectrometría de masas, basada en la desorción ionización asistida por una matriz sólida de rápida ocurrencia. Esta identificación de proteínas ocurre cuando éstas están inmovilizadas en una membrana o en una matriz de un gel. La técnica es de baja eficiencia.

- MALDI o ESI-MS (ElectroSpray Ionization Mass Spectrometry) o Espectrometría de masas por ionización en electroatomizado. Esta técnica analiza proteínas inmovilizadas en membranas o eluidas y digeridas en los geles.

Los datos que salen de la técnica de espectrometría de masas son fingerprints o “huellas digitales” de péptidos separados por masa. La identificación de las proteínas se hace por aprendizaje de máquina, refiriendo los patrones particulares de espectrometría de masas a una base de datos que sirve para entrenamiento.

Un objetivo más o menos inmediato de la proteómica es la aplicación de las tecnologías y datos a la investigación en clínica humana. Otras áreas de menor desarrollo en el conocimiento incluyen los estudios ambientales, investigación en agricultura y en veterinaria, las cuales podrán usar los desarrollos de la proteómica solo de mediano a largo plazo. Un objetivo claro de la proteómica clínica es el descubrimiento de drogas clínicamente importantes, es decir, blancos potenciales para desarrollo farmacológico, y mayor distancia entre los ricos y los pobres.

En la actualidad, la computación es un componente esencial de la investigación en biología. La masa de los datos producidos es tan grande que no se puede ni siquiera “mirar” sin la ayuda de un computador. Quizás la bioinformática nunca haya enfrentado un problema de magnitud tan grande. Necesitamos bases de datos

integradas, eficientes para entrar y sacar datos; así mismo, necesitamos métodos computacionales avanzados tales como reconocimiento de patrones y otros enfoques de aprendizaje de máquina, para analizar y entender los datos.

La computación suministra herramientas poderosas para detectar relaciones sutiles entre los datos y sugiere hipótesis y rutas para la validación experimental. Los genetistas antiguos hacían genética impulsada por hipótesis, mientras los nuevos plantean hipótesis basados en los datos del computador, lo cual añade así una dimensión nueva a la intuición humana capacitándonos para obtener entendimientos adicionales. Quienes puedan integrar la computación a sus tareas investigativas tendrán mucho éxito. Para aprovecharnos de la riqueza de los datos necesitamos dos tipos de desarrollos computacionales:

1. Infraestructura de las bases de datos que permitan almacenamiento y recuperación eficientes, que sea biológicamente intuitiva y con una interfase que permita la comunicación entre bases de datos y científicos de diferentes disciplinas.

2. Sistemas inteligentes, agentes y software para discernir las relaciones entre los datos, de tal manera que se puedan formular hipótesis junto con experimentos de validación. Debemos esperar que estos desarrollos en bioinformática ayuden a responder preguntas fundamentales de toda la biología, incluyendo preguntas que aún no se han formulado. Podemos incluir en este punto la necesidad de entrenar la nueva generación de científicos multidisciplinarios, en lo cual la Universidad Colegio Mayor de Cundinamarca puede hacer una contribución importante.

Conclusión

La definición más amplia de proteómica puede ser la más útil; de hecho cualquier definición de proteómica debe llegarle al alma a los biólogos y los científicos de la computación. La proteómica representa la línea de conocimiento que quiere establecer las identidades, las cantidades, las estructuras, las funciones bioquímicas y celulares de todas las proteínas de un organismo, órgano u organelo, y las formas como las anteriores propiedades varían en el espacio y el tiempo y con el estado fisiológico y anímico. Una tarea de mucho tiempo que pueden dejar más atrasados que antes.

Referencias

1. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. *Molecular biology of the cell*. New York: Garland Publishing; 1994.
2. Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts and new words. *Electrophoresis* 1998;19:1853-61.
3. Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. *Trends in Biotech* 1999;17:121-7.
4. Kenyon GL, DeMarini^b DM, Fuchs E, Galas DJ, Kirsch^e JF, Leyh TS, et al. Defining the mandate of proteomics in the post-genomics era: workshop report. *Molecular & Cellular Proteomics* 2002;1:763-80. Disponible en: URL: <http://www.nap.edu/catalog/10209.html>
5. Lodish H, Berk A, Zipursky S, Matsudaira P, Baltimore D, Darnell J. *Molecular cell biology*. New York: W.H. Freeman and Company; 2000.
6. Wilkins M. editor. *New frontiers in functional genomics*. New York: Springer Verlag; 1997.