

4. **Proteómica, poderosa pero poco explorada ómica en la búsqueda y desarrollo de nuevos fármacos**

LAUREANO SIMÓN BUELA y
ANTONIO MARTÍNEZ MARTÍNEZ

RESUMEN

La Proteómica es una disciplina comparativa que pretende identificar y caracterizar aquellas proteínas que diferencian, a nivel de expresión o de modificación química o estructural, la célula o el tejido enfermo del sano. Esas proteínas diferenciales serán los marcadores con los que desarrollar métodos de predicción, diagnóstico, pronóstico o respuesta a fármacos de la enfermedad en estudio; esas proteínas diferenciales también podrán ser candidatas a dianas terapéuticas contra las que desarrollar nuevas generaciones de fármacos con los que tratar la enfermedad de una manera más eficaz y menos tóxica para el individuo.

Pese a todas sus limitaciones técnicas, la Proteómica permite la identificación de nuevas dianas terapéuticas contra las que dirigir nuevas generaciones de fármacos; la optimización no invasiva de los ensayos clínicos de esos fármacos, seleccionando aquellos pacientes con una mayor probabilidad de responder mejor al fármaco y ajustando, idealmente de una manera personalizada, las dosis de administración; y la optimización no invasiva del tratamiento con el fármaco de pacientes crónicos que lo necesiten, a través de la monitorización de marcadores de respuesta en períodos largos de tiempo.

1. **GENÓMICA Y PROTEÓMICA**

El conocimiento de la secuencia completa del genoma humano ha sido un gran avance de la Ciencia y de la Humanidad para la que todos los científicos

trabajamos, o deberíamos trabajar, pero no es la panacea que se ha presentado como la solución a todas las complicaciones fisiológicas que comprometen el deseado estado de buena salud.

Dentro de unos años las generaciones futuras de brillantes cerebros a los que tendremos que ilusionar y atraer al mundo científico, podrán confirmar la sabiduría popular clásica que no duda en calificar a la ignorancia como siempre osada. Creemos que tenemos la solución a nuestros problemas en nuestras cabezas y bases de datos, pero desconocemos el significado de la mayor parte de ese conocimiento, la secuencia del genoma humano; y en nuestra ignorancia, como no sabemos qué sentido tiene la existencia de enormes regiones de esa secuencia completa, la despreciamos denominándola *DNA basura*, y centramos nuestros esfuerzos en esas regiones más pequeñas y discretas, que codifican las proteínas responsables de la ejecución de la estructura y las funciones del cuerpo humano (1, 2); sin reparar en muchas ocasiones en que esas proteínas posiblemente no sean más que obreras dirigidas por entramados moleculares mucho más complejos de interacciones proteína-proteína, proteína-DNA, proteína-RNA, DNA-DNA, RNA-RNA,, interacciones en las que esa basura de DNA, que nos estorba en nuestra simplicidad, posiblemente juegue un papel básico y determinante (3, 4).

Pero, dejémosles a las futuras generaciones algo con lo que trabajar y apasionarse, y limítenos a lo más obvio, lo más tangible, lo más fácilmente asimilable por nuestros cerebros y por los cerebros de nuestros colegas que evaluarán la calidad científica de nuestros trabajos y de nuestras solicitudes de financiación. Y si simplicidad es lo que buscamos, nada mejor que el universo bidimensional del DNA codificante y su RNA transcrito, sólo incordiado por intrones estructurales y enojosos procesos de splicing alternativo; cierto es que incluso podemos dedicar una carrera entera, por no decir institutos enteros, a disfrutar con la búsqueda de nuevas mutaciones, variaciones alélicas, aberraciones cromosómicas, incluso enrevesadas metilaciones del DNA, o complejos mecanismos de regulación transcripcional, en *cis* o en *trans*, de modificaciones post-transcripcionales, de degradación del RNA, hasta podemos estudiar la razón de ser de las ribozimas. Y si queremos ser todavía más prácticos y aplicados, nos zambulliremos en el mundo obrero de las ejecutoras de funciones y estructuras: Las proteínas, y podremos dedicar carreras, institutos, incluso empresas enteras, al estudio de la regulación de su expresión, de las modificaciones co-traduccionales, de su ensamblaje, del mantenimiento de su conformación (tridimensional!), de su compartimentalización, de su degradación y de las modificaciones post-traduccionales que ya han conseguido convencernos de su importancia cla-

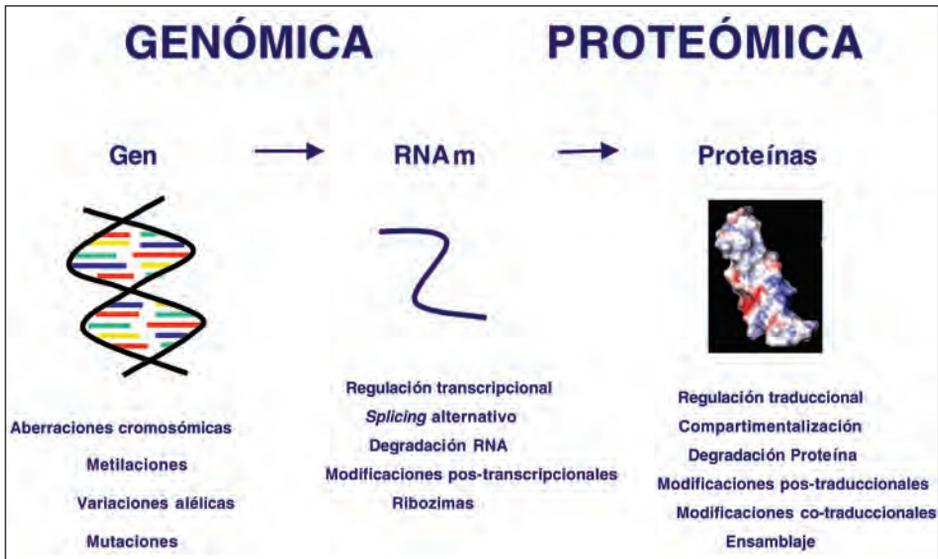


FIGURA 1. *Genómica vs Proteómica I*

ve para el mantenimiento de una estructura y función acordes con un estado saludable del individuo que constituyen. (Fig. 1).

El término *Proteoma* se refiere al contenido total de proteínas que se traducen del Genoma de un organismo. A diferencia del Genoma, que se puede considerar una estructura bastante constante, con excepciones como los procesos de replicación y reparación del DNA, el Proteoma de una célula es muy dinámico y varía continuamente en respuesta a estímulos o a condiciones anormales de stress o enfermedad.

Para analizar los Proteomas de células o tejidos se han desarrollado un conjunto de nuevas técnicas que se engloban en el término *Proteómica*. La Proteómica facilita el análisis cualitativo y cuantitativo de la expresión de las proteínas que existen dentro de una célula. La Genómica analiza el potencial de expresión génica y proteica de un organismo, determinado por su Genoma, pero es la Proteómica la que especifica qué proteínas se expresan en los distintos tipos de células de ese organismo, y en las distintas situaciones fisiológicas en las que se encuentran esas células, cual es el nivel de expresión de esas proteínas y qué modificaciones post-traduccionales afectan a su estructura y función (Fig. 2). Es muy representativo de esta diferencia y de la importancia de las modificaciones post-traduccionales, entre otras, que el número de genes que existen en el Genoma humano, alrededor de 30.000, podría originar más de

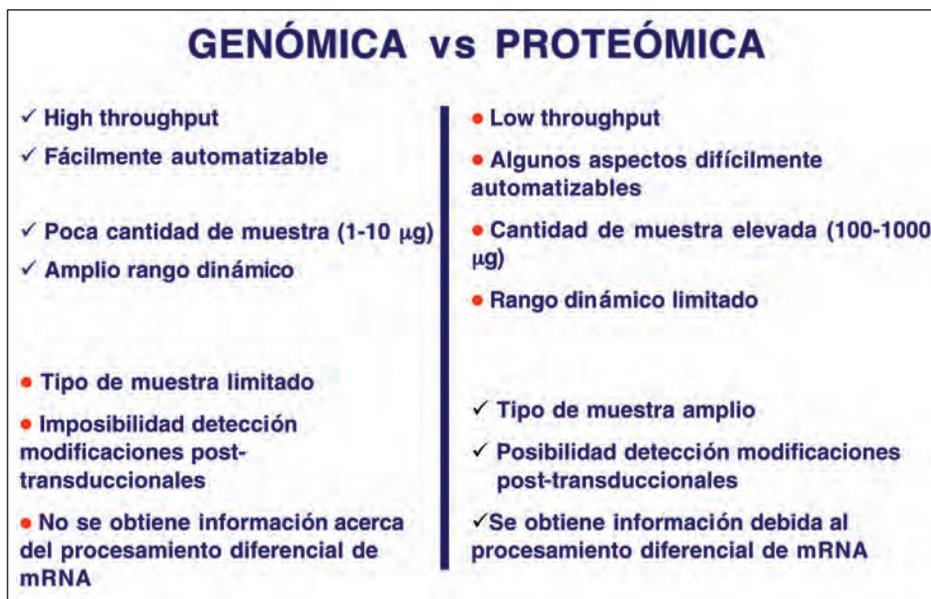


FIGURA 2. *Genómica vs Proteómica II*

1.000.000 de proteínas diferentes (5); aunque también es muy posible que de la misma manera que se estimaba hace menos de 5 años que el número de genes humanos podía llegar a 100.000, se compruebe dentro de otros 5 años que el número posible de proteínas humanas diferentes no llega ni a un tercio de las cifras que ahora se estiman (Fig. 3).

Un análisis reduccionista podría considerar que la Proteómica es más útil para analizar los cambios funcionales que ocurren durante un proceso patológico, que la Genómica, que analiza el mantenimiento de la estructura del DNA y los cambios a nivel del mRNA, que no tienen que traducirse necesariamente en cambios en la concentración de las proteínas funcionales, y que no puede analizar las modificaciones post-traduccionales, que juegan un papel muy importante en los procesos patológicos (Fig. 4). Sin embargo, no es necesario ser un gurú de la Biología para aceptar las ventajas de una aproximación analítica que integre herramientas genómicas y herramientas proteómicas, para predecir la ocurrencia y progresión del proceso patológico en base a la estructura del DNA del individuo, analizar la expresión génica en base a los niveles de mRNA, y analizar la expresión proteica en base a los niveles de proteínas y a las modificaciones post-traduccionales que esas proteínas experimenten (6).

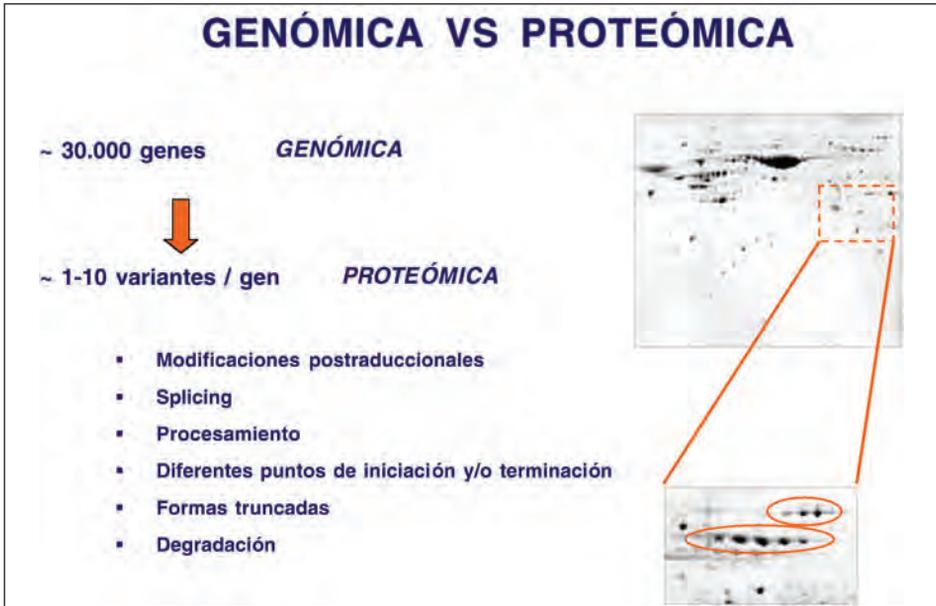


FIGURA 3. *Genómica vs Proteómica III*

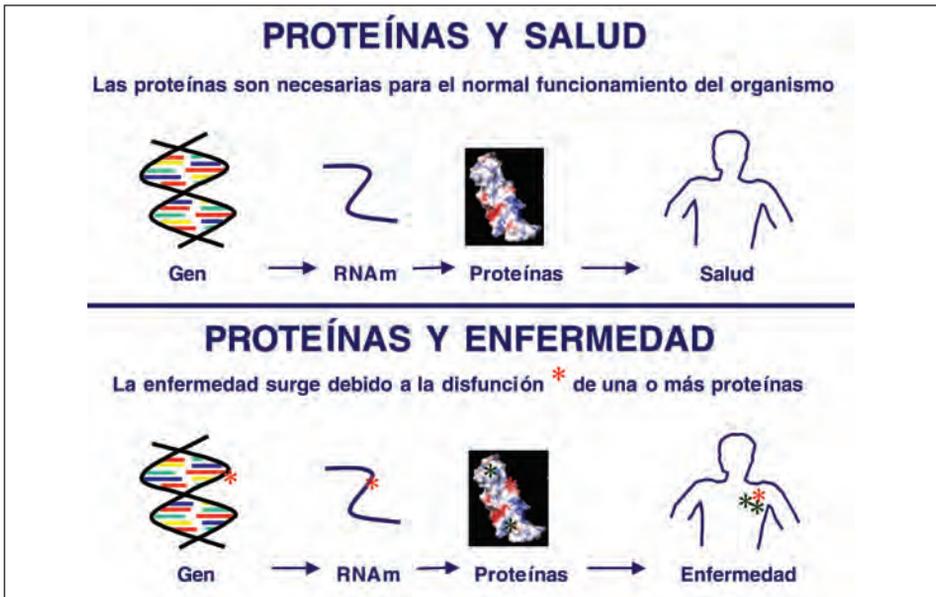


FIGURA 4. *Disfunciones proteicas y enfermedad.*

2. TECNOLOGÍA

La Proteómica es una disciplina comparativa. En el estudio de una enfermedad, la Proteómica pretende identificar y caracterizar aquellas proteínas que diferencian, a nivel de expresión o de modificación química o estructural, la célula o el tejido enfermo del sano. Esas proteínas diferenciales serán los marcadores con los que desarrollar métodos de predicción, diagnóstico, pronóstico o respuesta a fármacos de la enfermedad en estudio; esas proteínas diferenciales también podrán ser candidatas a dianas terapéuticas contra las que desarrollar nuevas generaciones de fármacos con los que tratar la enfermedad de una manera más eficaz y menos tóxica para el individuo (Fig. 5).

La tecnología proteómica de comienzos del siglo XXI combina potentes, modernos y sofisticados equipos de espectrometría de masas para la identificación de las proteínas, con métodos de preparación y separación previa de las proteínas a analizar que muy poco han evolucionado desde mediados del siglo XX. Grandes bioquímicos revolucionaron el análisis proteico hace ya más de 50 años, pero los métodos que ellos desarrollaron estaban encaminados al estudio de una o unas pocas proteínas, en ningún momento osaban siquiera plantearse el estudio global del proteoma completo de una célula en una determinada condición fisiológica. Los métodos que a continuación describimos brevemente y desde nuestra condición de meros usuarios, permiten el análisis de una porción de ese Proteoma, pero tienen grandes limitaciones para acometer debidamente la identificación y caracterización de las proteínas minoritarias, que pueden jugar un importante papel en muchos procesos fisiológicos y principalmente en las alteraciones de estos procesos, que causan las enfermedades.

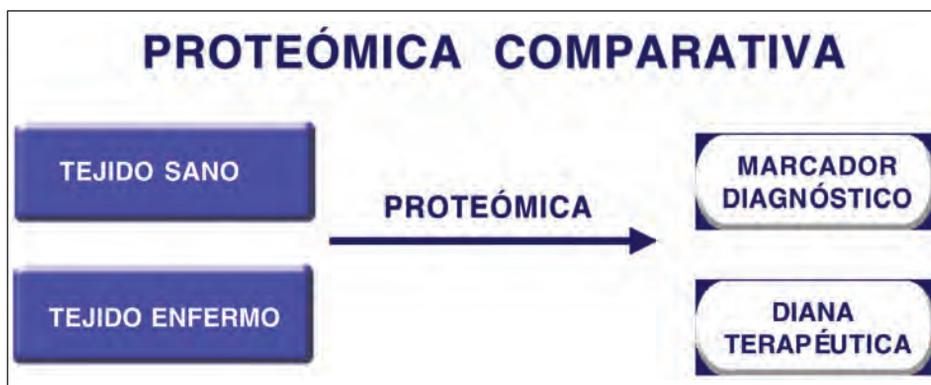


FIGURA 5. *Proteómica comparativa.*

Las técnicas proteómicas abarcan desde la extracción, preparación y purificación de las proteínas que queremos analizar en un extracto de múltiples proteínas procedente de un tejido determinado, de una manera que permita aplicar correctamente los métodos de análisis elegidos, hasta la manipulación, identificación y caracterización final de esas proteínas individuales.

2.1. Extracción, purificación y preparación de las proteínas

Los protocolos de extracción, purificación y preparación de las proteínas pueden ser determinantes en el resultado del método completo de identificación proteica; la sensibilidad general del protocolo global depende más de la estrategia de purificación que de la sensibilidad de los sofisticados métodos de identificación que se utilizan en pasos posteriores.

El protocolo habitual de purificación de proteínas consiste en un lisado celular, extracción de la mezcla de proteínas presentes en ese lisado, separación de las proteínas individuales por su carga (punto isoeléctrico o pI) y por su masa mediante gel-electroforesis bidimensional, tinción en el gel de las proteínas separadas, escisión del gel y posterior digestión (generalmente tripsinización) de las proteínas escindidas, análisis por espectrometría de masas de los péptidos resultantes e identificación de esos péptidos comparando los resultados de la espectrometría de masas con bases de datos de masas y espectros de proteínas previamente identificadas (7, 8) (Fig. 6-10).

Los métodos clásicos de purificación de proteínas tienen una serie de limitaciones: No pueden trabajar con proteínas muy ácidas ($\text{pH} < 3$) o muy básicas ($\text{pH} > 10$), ni con proteínas hidrofóbicas o liposolubles; el bajo rango dinámico de la técnica provoca que proteínas minoritarias queden ocultas tras proteínas *housekeeping* mayoritarias; el tamaño del gel impide estéricamente la detección de más de 2.000-3.000 proteínas, un número muy lejano de las 30.000 proteínas que podrían, de una manera aproximada, estar presentes en el extracto original.

Sobre esta base tecnológica clásica, se han desarrollado en los últimos años una serie de variaciones que aumentan considerablemente la resolución de los geles bidimensionales, principalmente reduciendo la complejidad de la muestra a analizar (9, 10):

- Aumento de la resolución de la primera dimensión, disminuyendo el rango de pH, desde el 3-10 habitual a 4-7 ó 5-8 (Fig. 10).

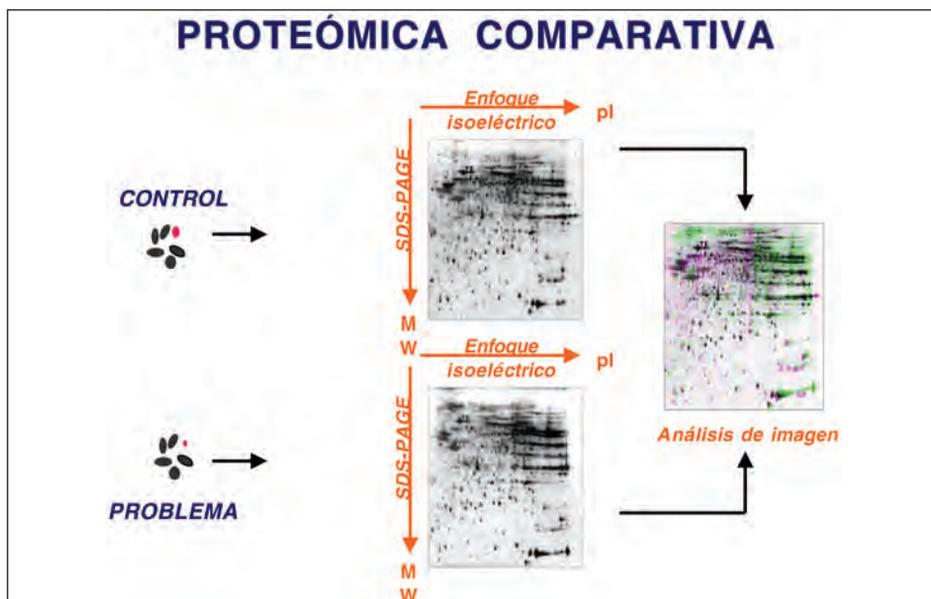


FIGURA 6. Separación de proteínas: Electroforesis bidimensional.

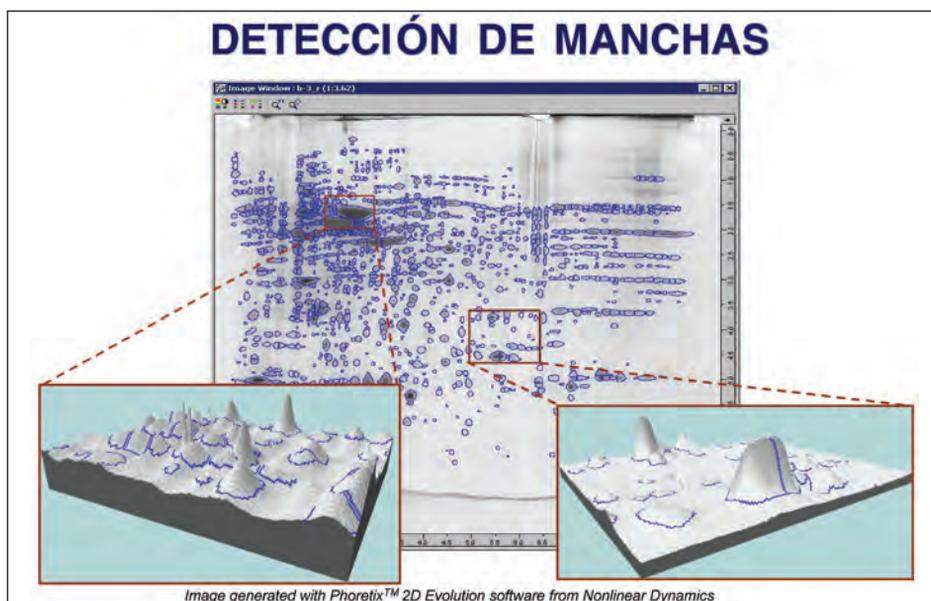


FIGURA 7. Detección de proteínas separadas (manchas) con programas de análisis de imágenes.

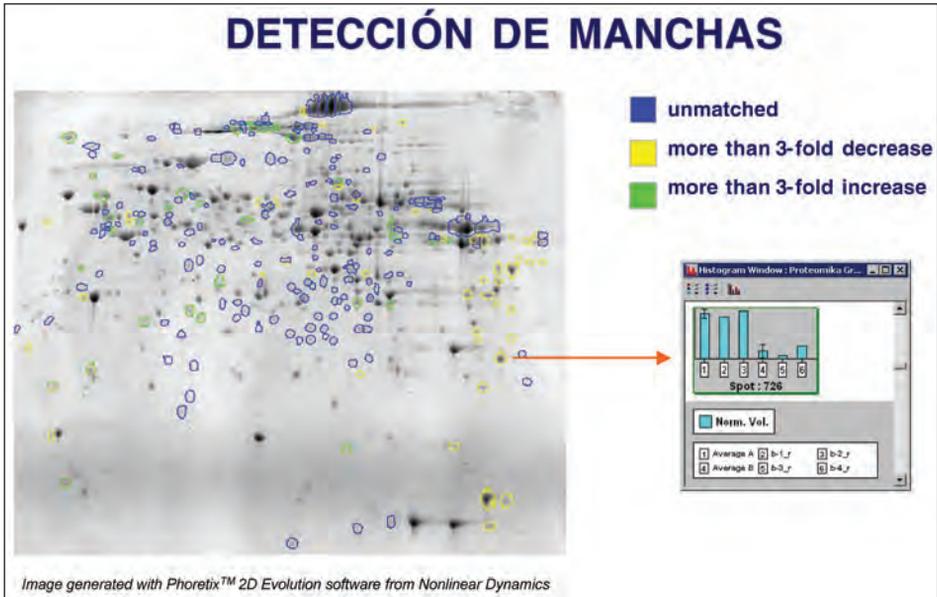


FIGURA 8. Identificación de proteínas (manchas) diferenciales sano vs enfermo con programas de análisis de imágenes.

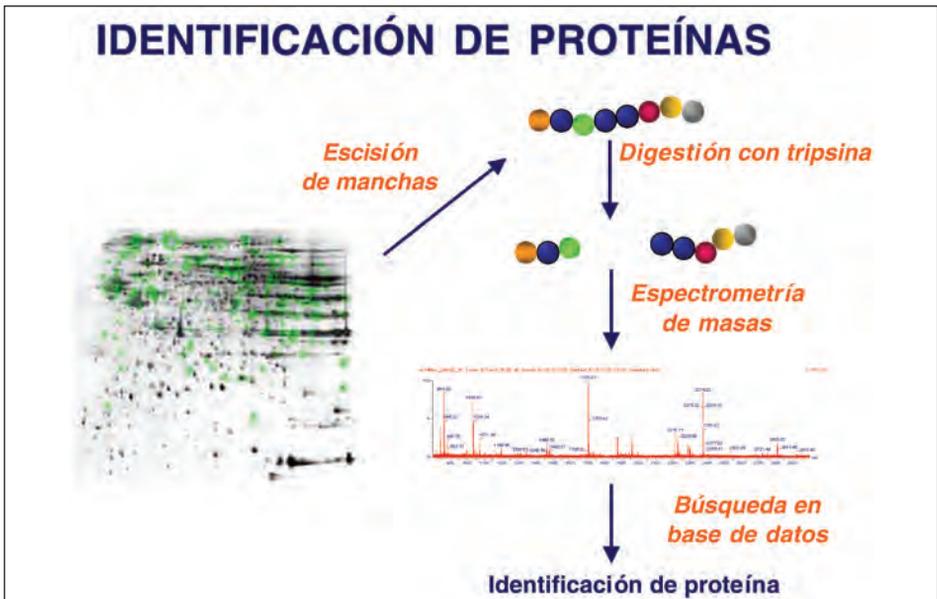


FIGURA 9. Escisión de proteínas (manchas) diferenciales sano vs enfermo, tripsinización e identificación por espectrometría de masas y comparación con bases de datos.

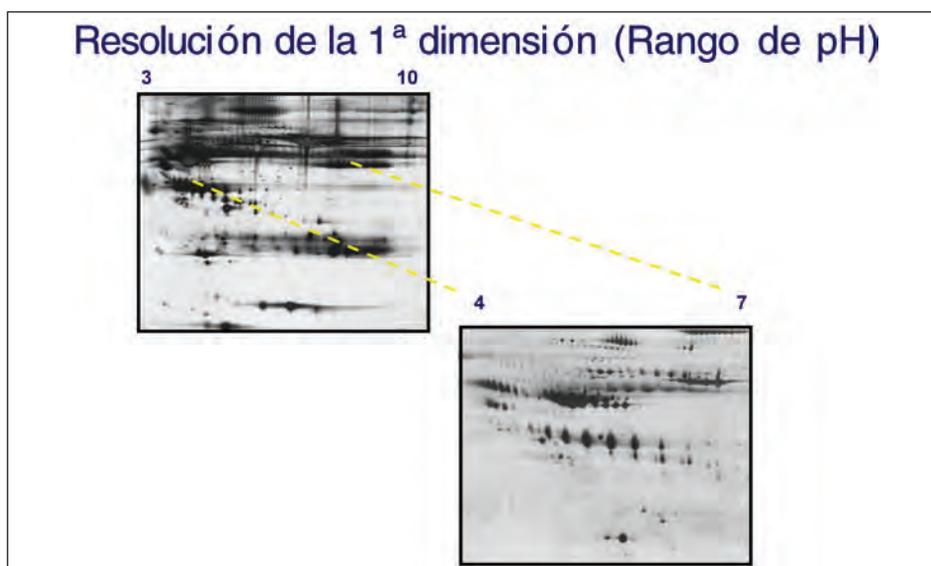


FIGURA 10. *Aumento de resolución de la electroforesis bidimensional I: Resolución de la primera dimensión: Punto Isoeléctrico.*

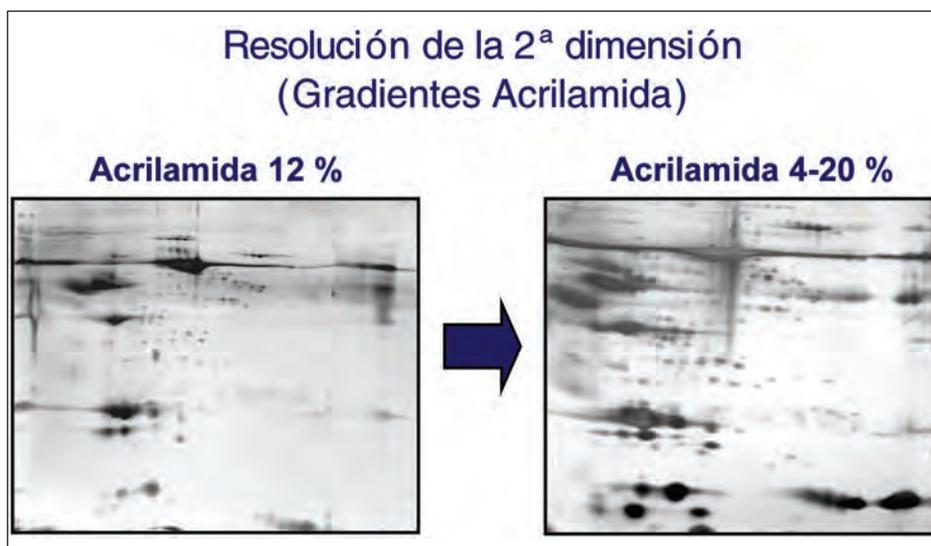


FIGURA 11. *Aumento de resolución de la electroforesis bidimensional II: Resolución de la segunda dimensión: Peso Molecular.*

- Aumento de la resolución de la segunda dimensión, trabajando con gradientes de poliacrilamida (Fig. 11).



FIGURA 12. Aumento de resolución de la electroforesis bidimensional III: Eliminación de proteínas mayoritarias.

- Eliminación de proteínas mayoritarias. La eliminación de proteínas mayoritarias como la albúmina, IgGs, transferrina, etc, ha permitido el análisis del proteoma del suero y la identificación de proteínas minoritarias como marcadores no invasivos con un elevado interés clínico (Fig. 12).
- Fraccionamiento subcelular. La extracción selectiva de las proteínas de un determinado compartimento celular ha facilitado análisis de gran importancia farmacéutica, como la identificación de potenciales dianas terapéuticas en fracciones de membrana plasmática (Fig. 13).

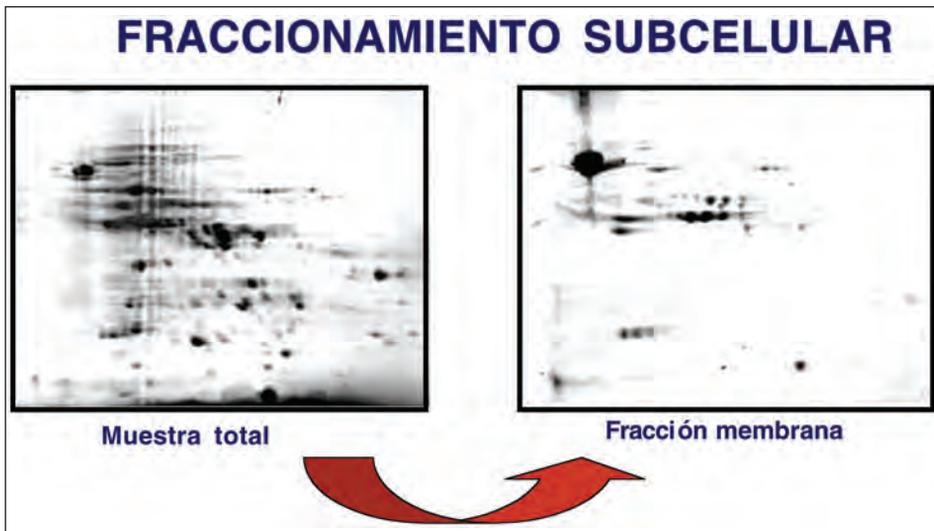


FIGURA 13. Aumento de resolución de la electroforesis bidimensional IV: Fraccionamiento subcelular

Estas mejoras de la tecnología clásica han permitido que los geles bidimensionales, pese a sus limitaciones, y a la aparición de técnicas alternativas, sigan siendo una herramienta fundamental del análisis proteómico.

Entre estas técnicas alternativas, que eliminan el uso de los geles bidimensionales, o lo enriquecen con otras tecnologías complementarias, se pueden mencionar:

- Tecnología Multidimensional de Identificación de Proteínas: Las proteínas son separadas por cromatografía líquida (LC) bidimensional, que combina columnas de intercambio iónico con columnas capilares de fase reversa (11, 12). Esta tecnología, que culmina en la identificación de las proteínas por Espectrometría de masas en tandem, es básicamente cualitativa, no cuantitativa y se ha demostrado útil para el análisis de proteínas de membrana, pero tiene grandes problemas con las proteínas mayoritarias, ya que la mayoría de los protocolos de eliminación de proteínas mayoritarias, imprescindibles para trabajar con fluidos como el suero, son incompatibles con las columnas de LC. En nuestros laboratorios de Proteomika, SL, hemos paliado este problema añadiendo una tercera dimensión antes de la espectrometría de masas, lo que nos ha permitido aplicar esta tecnología al estudio del Plasma humano en el proyecto Proteoma del Plasma de la HUPO (Human Proteome Organization), en la que participamos activamente.
- Marcaje por afinidad codificado por isótopos (ICAT): Dos muestras proteicas diferentes, por ejemplo tejido sano y tejido enfermo, se marcan con reactivos ligeramente diferentes: Un marcador contiene un isótopo de deuterio o C-13 que aumenta su peso molecular en 8 unidades respecto al segundo marcador. Las muestras se digieren entonces y se analizan por LC microcapilar y espectrometría de masas por Electrospray. Los cambios en la expresión proteica entre la muestra sana y enferma, se determinan midiendo la diferencia en las intensidades de los picos ligeros y pesados. ICAT genera una lista de proteínas que están presentes en 1 o en las 2 muestras y la abundancia relativa de cada proteína en las muestras comparadas. Esta tecnología tiene un alto poder cuantitativo, pero una importante limitación, que es que sólo se marcan los péptidos que contienen residuos de Cisteína.
- Combinación LC - Gel bidimensional: El extracto de proteína total se fracciona en una columna de intercambio iónico, antes de separar por electroforesis bidimensional las proteínas presentes en cada fracción.

2.2. Espectrometría de masas

La espectrometría de masas (MS) es una técnica que se remonta a los comienzos del siglo XX. Esta técnica, más venerable que antigua, debe su éxito y prevalencia a que mide con gran sensibilidad una propiedad intrínseca de las moléculas: su masa (13). Pese a la fecha de su origen, la MS no se aplicó en Biología hasta la década de los 80; este retraso se debió a que los espectrómetros de masas requieren moléculas gaseosas cargadas para su análisis, y las moléculas biológicas, pese a su polaridad, no se ionizan y convierten a estado gaseoso con facilidad. Dos técnicas de ionización han evitado este problema y originado el éxito de la MS en Biología: Electrospray (ES) (14) y Matriz-asistida por láser desorción/ionización (MALDI) (15); el desarrollo de algoritmos en la década de los 90 facilitó la correlación de los datos originados por los espectrómetros de masas con las bases de datos de secuencias proteicas, facilitando la automatización de la interpretación de los resultados y la identificación de la proteína analizada (16, 17). MALDI se utiliza principalmente para el análisis de péptidos, mientras que Electrospray puede analizar proteínas enteras y se ha popularizado su uso en combinación con la Cromatografía Líquida de Alta resolución (HPLC). Tres principios básicos se aplican actualmente para lograr la separación de masas, separación en base al tiempo de vuelo (Time Of Flight, TOF), que se acopla al MALDI (MALDI-TOF), y separación en base a campos eléctricos cuadrupólicos (Quadrupole) o separación en base a una eyección selectiva de iones desde un campo atrayente tridimensional (Ion trap), que se acoplan a ES. Para secuenciar el péptido a analizar, dos pasos de espectrometría de masas se realizan en tandem (MS/MS). A principios del siglo XXI la espectrometría de masas ha avanzado vertiginosamente, se ha combinado con éxito el triple Quadrupole con un analizador TOF, que reemplaza la tercera sección del Quadrupole (Q TOF), con gran éxito como equipo secuenciador de fragmentos peptídicos escindidos de geles bidimensionales, pese a su lentitud; combinaciones de MALDI con dos secciones de TOF son las preferidas en la actualidad por muchos laboratorios, incluido el nuestro, por su rapidez, que permite un análisis muy sensible, de múltiples muestras en espacios cortos de tiempo (18).

3. APLICACIONES DE LA PROTEÓMICA EN EL DESARROLLO DE FÁRMACOS

Pese a todas sus limitaciones técnicas, la Proteómica permite la identificación de nuevas dianas terapéuticas contra las que dirigir nuevas generaciones de fármacos; la optimización no invasiva de los ensayos clínicos de esos fármacos,

seleccionando aquellos pacientes con una mayor probabilidad de responder mejor al fármaco y ajustando, idealmente de una manera personalizada, las dosis de administración; y la optimización no invasiva del tratamiento con el fármaco de pacientes crónicos que lo necesiten, a través de la monitorización de marcadores de respuesta en períodos largos tiempo.

3.1. Identificación y validación de dianas terapéuticas

Los fármacos que están actualmente en el mercado se dirigen contra un número de dianas que no supera 500. Si la ciencia del desarrollo farmacéutico se acercase más a la ingeniería que a la biología, podríamos afirmar que identificando la asociación a las enfermedades de los 30.000 genes estimados en el genoma humano y del millón de proteínas resultantes tras todas las posibles modificaciones, podríamos encontrar millares de nuevas dianas terapéuticas (19). Números como 5.000 han aparecido en artículos, imaginamos que de opinión, en revistas científicas del prestigio de Nature o Science; números que también imaginamos resultantes de magníficas tardes de *brain storming* en el *pub* que suele estar cercano a los grandes institutos de investigación ingleses y americanos. No nos resulta difícil creer, y es que de eso estamos hablando en este párrafo, de Fe, más que de empirismo y experimentación, que haya más de 5.000 potenciales dianas identificables tras sofisticados experimentos de Genómica y Proteómica, pero si actuamos como científicos rigurosos, esas dianas potenciales habrán de ser validadas molecularmente con diferentes técnicas que coincidan en su resultado y validadas funcionalmente en modelos *in vitro* e *in vivo* que demuestren no sólo su asociación a la enfermedad en estudio, sino que también demuestren que la modificación de esas dianas juega un papel causal de la enfermedad y que la actuación sobre esas dianas, inhibiéndolas, expresándolas o modificándolas, en un sentido que no necesariamente será opuesto al relacionado con la enfermedad, tiene un efecto curativo, o al menos paliativo, de esa enfermedad. Siendo muy optimistas, esas 5.000 potenciales dianas del *pub* se reducirán a unos pocos cientos, que en este maravilloso ejercicio de opinión que nos ha permitido la editora de este interesante y necesario libro, reduciremos a las 150 dianas de la tasca; que los medios de la Ciencia española, no permiten mayores alegrías. Pero aún siendo sólo 150 dianas, es todo un filón a explorar, un filón suficiente no para salvar a la humanidad y que vivamos sanos y felices durante tantos años como nuevas dianas estimamos encontrarnos, pero sí suficiente para provocar un gran avance de la Farmacia en los próximos 10 años.

Debemos ser realistas, no disponemos de las herramientas para curar todas las enfermedades diagnosticadas, o no sabemos utilizarlas apropiadamente en proyectos en los que tanta importancia debería tener el médico facultativo, como el investigador académico y como el investigador industrial que canalice y optimice el conocimiento generado en la Academia, para resolver los problemas clínicos, que el médico debe plantear y supervisar en todo momento, para que no se pierda la perspectiva del fin que se persigue, que es combatir la enfermedad; pero sí hemos desarrollado una nueva generación de herramientas de estudio genómico y proteómico que están multiplicando la eficacia y la velocidad del desarrollo farmacéutico, y sus efectos masivos se comprobarán en los próximos años con la aparición de nuevas generaciones de fármacos y con la optimización y personalización de esos tratamientos nuevos y de los ya existentes, lo que multiplicará su eficacia.

3.1.1. *Identificación de nuevas dianas terapéuticas*

La Proteómica es, junto a la Genómica, la estrella de esas nuevas herramientas. Las técnicas y equipos descritos en el apartado anterior permiten un estudio comparativo de la expresión de las proteínas en el tejido sano y el tejido enfermo, para identificar aquellas proteínas que están asociadas a la enfermedad. La Proteómica ofrece una ventaja frente a la Genómica, al permitirnos encontrar las proteínas que hayan sufrido modificaciones post-taduccionales, glicosilaciones, fosforilaciones, ubiquitinaciones, ..., un elevado número de cambios que hoy sabemos experimentalmente que juegan un papel determinante, no sólo en el desarrollo de la enfermedad, sino también en su causa.

La causalidad es la pieza clave en la identificación de la diana terapéutica; encontraremos cientos, incluso miles de cambios en la expresión y la estructura de las proteínas de la célula al comparar el tejido sano con el enfermo, pero la vasta mayoría de esos cambios serán consecuencia de la enfermedad, no la causa. La validación funcional de las potenciales dianas identificadas será determinante para evaluar su papel en la enfermedad, pero ya en la etapa de identificación hemos de cribar en base al conocimiento publicado en la literatura científica y en la multitud de bases de datos, que cobran más y más importancia; bases de datos en las que encontraremos si esa potencial diana ha sido identificada previamente en la enfermedad que estamos estudiando, o en otra enfermedad molecularmente cercana, bases de datos en las que también posiblemente encontraremos la localización celular de esa diana, la actividad y función, o funciones, que se han descrito para esa diana, o que podría tener en base a los do-

minios que presenta, su localización en cascadas moleculares y bioquímicas, la localización cromosómica del gen que codifica esa proteína y otros muchos parámetros que ocasionan que el investigador proteómico dedique más tiempo a la consulta de bases de datos que a la consulta de la literatura científica (20, 21).

Con esta información, el investigador seleccionará preferente, que no exclusivamente, las proteínas de membrana, si quiere que la molécula terapéutica a desarrollar contra ella sea un anticuerpo, o una molécula guiada por ese anticuerpo, una de las aproximaciones terapéuticas preferidas por las empresas biotecnológicas, que no disponen de las grandes colecciones de moléculas de tamaño menor que el anticuerpo y por tanto capaces de atravesar la membrana plasmática y acceder a proteínas intracelulares; el investigador seleccionará también proteínas que estén en los primeros pasos de las cascadas bioquímicas, en las que de nuevo estarán bien situadas la mayoría de las proteínas de membrana; el investigador también seleccionará proteínas con actividad conocida experimentalmente, o inferida en base a sus dominios, que encaje en un modelo de causalidad, proteínas como las quinasas son siempre bien evaluadas, incluso hay grandes empresas farmacéuticas que han centrado en ellas su esfuerzo de desarrollo de nuevas moléculas (22, 23).

Hemos de ser conscientes de que la causa primera de una enfermedad será la conjunción de varios factores, como una estructura genómica que predisponga a padecer la enfermedad, por aumentar la susceptibilidad al efecto de factores ambientales que faciliten el desencadenamiento de la enfermedad, bien modificando la estructura genómica, mutando un gen concreto, bien provocando directamente la modificación post-traducciona de una proteína, que ocasione una cascada bioquímica que produzca la enfermedad, o bien induciendo la expresión de una proteína que será la encargada de modificar post-traducciona otra proteína, o bien induciendo la expresión de una proteína que será la encargada de inducir la expresión de otra proteína que será la encargada de modificar post-traducciona otra proteína,, y así podemos complicar la cascada tanto como queramos, sin alejarnos de la realidad. Pero también hemos de ser conscientes de que la diana terapéutica tampoco tiene que ser necesariamente la causa primera de la enfermedad. En una cascada de fosforilaciones en la que participan quinasas, quinasasquinasas, que son las quinasas que fosforilan a otras quinasas, quinasasquinasasquinasas, quinasasquinasasquinasasquinasas,, es posible que una de esas quinasasquinasas sea una buena diana terapéutica, incluso es posible que sea mejor diana que la quinasa original, por muchos factores como pueden ser la facilidad de interacción con moléculas de síntesis sencilla.

La conclusión es que debemos cribar con criterios científicos, no sólo bioquímicos sino también farmacéuticos, la lista original de potenciales dianas terapéuticas que el análisis proteómico identifique asociadas a la enfermedad.

3.1.2. *Validación molecular de las potenciales dianas terapéuticas*

Las proteínas que el estudio proteómico ha identificado asociadas a la enfermedad y que la evaluación científica ha seleccionado como potenciales dianas terapéuticas, han de ser validadas por otras técnicas moleculares que confirmen esa asociación. Entre estas técnicas destacan las inmunoquímicas, que permiten con el uso de anticuerpos específicos de esa proteína confirmar los niveles de expresión de la diana, por ELISA, Western blot, ..., su localización celular por Inmunohistoquímica y las modificaciones post-traduccionales en geles unidimensionales y/o bidimensionales con anticuerpos específicos de la diana y con anticuerpos que detecten la modificación post-taduccional.

3.1.3. *Validación funcional de las potenciales dianas terapéuticas*

Las potenciales dianas terapéuticas, una vez cribadas con criterios de conocimiento bioquímico y farmacéutico, y validadas molecularmente con diferentes técnicas que coincidan en su resultado, han de ser validadas funcionalmente en modelos *in vitro* e *in vivo*, que demuestren no sólo su función causal de la enfermedad en estudio, sino que también demuestren que la actuación farmacéutica sobre esas dianas, inhibiéndolas, expresándolas o modificándolas, en un sentido que no necesariamente será opuesto al relacionado con la enfermedad, tiene un efecto curativo, o al menos paliativo, de esa enfermedad.

En un primer nivel de validación funcional, las potenciales dianas se validan *in vitro* en cultivos de líneas celulares del tejido que se conoce afectado por la enfermedad. La utilidad de estos modelos celulares depende de la existencia de fenotipos analizables; estos modelos son muy útiles para enfermedades como el cáncer, en el que el fenotipo de proliferación celular es fácilmente analizable y cuantificable en líneas celulares inmortalizadas del tejido afectado, en respuesta a la inducción o represión de la expresión de la proteína a validar. En nuestros laboratorios de Proteomika, SL, hemos validado *in vitro* potenciales dianas como el Receptor 3 del factor de crecimiento de fibroblastos (FGFR3), una proteína de membrana que habíamos identificado inducida en carcinomas transicionales de vejiga; el tratamiento de líneas celulares de vejiga que expre-

saban FGFR3 con anticuerpos monoclonales específicos de FGFR3 y también con fragmentos de anticuerpos humanos específicos de FGFR3, inhibía la proliferación de esas células (24); estos modelos son por tanto útiles no sólo para validar la diana, sino también como modelos para el screening de moléculas con actividad antitumoral, como los fragmentos de anticuerpos humanos específicos de FGFR3. También hemos validado una proteína, posiblemente supresora de tumores, que estaba reprimida en oncocitomas de riñón, comprobando que al inducir la expresión de esa proteína en líneas celulares de riñón, inhibíamos la proliferación celular, postulando de esa manera la proteína en cuestión como una molécula con actividad antitumoral *per se*.

La validación funcional *in vivo* en modelos animales es el siguiente paso, más cercano a la realidad humana de la enfermedad. El modelo ya clásico de validación es la generación de animales transgénicos que expresan, o sobreexpresan, en el tejido diana de la enfermedad, la proteína que hemos identificado como inducida en el estudio proteómico de la enfermedad, o la generación de animales transgénicos *Knock-Outs*, en los que está reprimida la expresión de la proteína que hemos identificado con una expresión reprimida en el estudio proteómico de la enfermedad. Si estos animales presentan un fenotipo asociado a la enfermedad, desarrollo de tumores en cáncer, comportamientos psicóticos si las dianas las hemos identificado estudiando muestras procedentes de individuos con trastornos psiquiátricos, complicaciones cardiovasculares, complicaciones neurodegenerativas, etc, no sólo habremos validado la importancia funcional del papel que esa diana juega en la enfermedad, sino que también habremos generado un magnífico modelo animal para el *screening* de moléculas terapéuticas con las que combatir esa enfermedad. El modelo animal más utilizado para estos experimentos de validación es el ratón (25), pero existen otros modelos como la mosca de la fruta o el pez cebra, que aún siendo modelos mucho más alejados del hombre y por tanto generadores de innumerables resultados falsos positivos y negativos al extrapolarlos al hombre, son modelos que permiten un análisis rápido de múltiples dianas. En la tasca consideramos que estos modelos son interesantes para complementar, pero no para sustituir a modelos mamíferos, entre los que el ratón es el que presenta mayores ventajas.

3.2. Identificación de nuevas moléculas con actividad terapéutica.

En el apartado anterior hemos comprobado con ejemplos reales que estudios proteómicos realizados en nuestra empresa, Proteomika, SL, rendían no sólo la identificación y validación de dianas terapéuticas, sino que también pos-

tulaban la utilidad terapéutica de las proteínas identificadas como reprimidas en la enfermedad en estudio, o de otras proteínas, los anticuerpos humanos o humanizados desarrollados contra proteínas inducidas en la enfermedad. Estas aproximaciones son las favoritas de las empresas biotecnológicas, muy alejadas por motivos técnicos y económicos del desarrollo químico de fármacos a través del *screening* de colecciones de compuestos, o de aproximaciones más sofisticadas de diseño y síntesis química de nuevos compuestos; pero debemos ser conscientes de que las proteínas como moléculas terapéuticas tienen enormes limitaciones, por su fácil degradabilidad, casi ineludible en el tracto digestivo, imposibilitando su administración oral, y por su gran tamaño, que impide su accesibilidad a las dianas intracelulares, o su paso a través de la barrera hematoencefálica.

Aún así, son ya varios los fármacos de base proteica aprobados para su uso terapéutico, desde clásicos como la insulina, hasta resultantes de un desarrollo farmacogenómico, como la Herceptina, que es un anticuerpo humanizado específico del receptor Her-2 que está inducido en el 30% de los carcinomas de mama, y que ha demostrado una gran utilidad para combatir esos tumores en los que está inducido.

Un escenario de lucha contra enfermedades como el cáncer, con combinaciones de fármacos, que bien podrían ser de naturaleza proteica, desarrollados contra varias dianas identificadas con herramientas genómicas y proteómicas, está mucho más cercano en el tiempo de lo que podemos imaginar; incluso en nuestra modesta y conservadora tasca lo consideramos ya como una realidad, tan sólo eludible por criterios económicos. La combinación de fármacos supondrá una necesaria alianza entre distintas empresas productoras de los distintos fármacos a combinar, muy difícil de conseguir si no es a través de la continuación de macroprocesos de compras y fusiones, y supondrá también un incremento del coste del tratamiento, ya muy alto cuando de fármacos individuales se trata, y más alto todavía si esos fármacos son biológicos. Quizás hemos de ir concienciándonos en los países avanzados, en los que tenemos la suerte de poder hacerlo, de que un importante porcentaje de nuestros ingresos personales deberá dedicarse al cuidado de nuestra salud. ¿Por qué nos vamos a rasgar las vestiduras por dedicar un porcentaje de nuestros ingresos a nuestra salud, cuando no pestañeamos al dedicarle más del 30% a la compra o alquiler de las 4 paredes entre las que vivimos?. Mayores impuestos para mantener la estructura de bienestar social, de la que tan orgullosos podemos estar en Europa, serán pronto ineludibles, como ineludible debería ser una dura regulación inmobiliaria que evite un descabellado, por ilógico e injusto, gasto de los ingresos personales. Si

la contabilidad macroeconómica pudo evitar el impacto del Euro en el IPC, también podrá evitar el impacto del pinchazo de la burbuja inmobiliaria en los índices de crecimiento, pinchazo que es necesario para una correcta adecuación del gasto farmacéutico, público y privado (*in tasca dixit*).

3.3 Ensayos proteómicos en la fase preclínica del desarrollo farmacéutico

La aplicación de las herramientas proteómicas en los ensayos preclínicos, *in vivo* e *in vitro*, de los fármacos biológicos o de los convencionales, está todavía menos implantada que la utilización de las herramientas genómicas. La causa principal es que la caracterización de un parámetro determinado que nos aporte información sobre la actividad, eficacia, toxicidad, mecanismo de acción, ..., de un fármaco determinado, ha de cubrir el mayor número de variables posible, y eso no es posible con una disciplina incipiente como es la Proteómica. No existe el conocimiento necesario de los perfiles proteómicos asociados a distintos mecanismos de actividad o toxicidad farmacológica, necesarios para poder interpretar los resultados de un experimento proteómico en el que queramos evaluar la toxicidad o la actividad de ese fármaco; conocimiento que sí existe ya en Genómica, con el desarrollo de bases de datos generados en miles de experimentos en los que se ha analizado con herramientas genómicas *in vitro* e *in vivo* los efectos de la administración de cientos de fármacos a modelos celulares y a animales modelo.

El análisis simultáneo de todos los posibles cambios de expresión, modificaciones post-traduccionales, ..., de las proteínas de una célula o de un tejido concreto, diana de la enfermedad y del fármaco en estudio, no es factible hoy en día, porque no se conocen muchas de esas modificaciones y su asociación a enfermedades y porque no existen herramientas de análisis masivo sensible, específico y reproducible, ni siquiera de todos los posibles cambios y asociaciones a enfermedades ya conocidos. El conocimiento de todas las modificaciones y de su asociación a las distintas enfermedades será cuestión de tiempo y de fuertes inversiones en los institutos de investigación académicos e industriales. El desarrollo de herramientas equivalentes a los microarrays de DNA debería ser una prioridad de cualquier plan nacional de ciencia y tecnología, y no lo es, ni aquí ni en USA; los microarrays permiten el análisis simultáneo de casi todos los posibles cambios de expresión génica (los 30.000 genes conocidos son analizados en un solo experimento), y de un importante número de cambios estructurales (millones de polimorfismos analizados en un único experimento); el desarrollo de una herramienta

equivalente que identifique las modificaciones post-traduccionales quizás esté todavía lejano, pero el desarrollo de una herramienta equivalente que mida los cambios de expresión de todas las proteínas codificadas por esos 30.000 genes es en este momento un abordaje más técnico que científico, perfectamente factible (26, 27), únicamente dependiente de los recursos económicos necesarios, nunca superiores a 90 millones de euros, 3.000 euros por anticuerpo específico para cada una de las proteínas codificadas por los 30.000 genes humanos conocidos, si queremos o tenemos que comenzar de cero; esta cantidad se reduciría drásticamente si implicamos a las empresas que tienen ya importantes colecciones de anticuerpos. Pero en el peor de los casos, 90 millones de euros, que son muchos euros en términos absolutos, son una cantidad insignificante si los repartimos en 3 años y los comparamos con los presupuestos nacionales, o europeos, que se diluyen en miles de proyectos de cuando menos cuestionable trascendencia. Una vez más tendremos que esperar a que lo desarrolle y nos lo venda a precio de oro una empresa americana, a la que hayan dotado, entre su gobierno y los inversores bursátiles, con 100 millones de euros, por si surgen complicaciones inesperadas; no nos creamos que el gobierno americano no subvenciona a las empresas biotecnológicas con recursos entre 10 y 100 veces superiores a los que las empresas europeas pueden acceder desde las 4-5 administraciones distintas con competencias para dárselos, y mientras el NasDaq recupera y multiplica la fuerza que tuvo a finales del siglo XX, sigamos observando complacientes la muerte de los nuevos mercados de las bolsas europeas, contentos con el escarmiento a esos chavales que se querían comer el mundo en unos meses; los nuevos mercados son los únicos capaces de financiar empresas con los recursos necesarios para competir con las empresas americanas de hoy y con las asiáticas de un futuro muy cercano.

En definitiva, la Proteómica no es en este momento la tecnología de elección para la caracterización de una molécula farmacéutica durante los ensayos preclínicos. La Proteómica en su estado actual es una poderosa disciplina de búsqueda, de identificación, pero no de caracterización.

3.4. Ensayos proteómicos en las fases clínicas del desarrollo farmacéutico

Al contrario que en las fases preclínicas, la Proteómica es una tecnología que puede facilitar mucho la optimización de los ensayos clínicos, no ayudando a la caracterización del fármaco, sino identificando biomarcadores no invasivos de diagnóstico y pronóstico de la respuesta del paciente a la administración del fármaco (28).

La Genómica, que en otros momentos del desarrollo farmacéutico es tanto o más útil que la Proteómica, pierde el protagonismo en los ensayos clínicos, porque la medida de los parámetros a analizar ha de ser no invasiva, para no infringir un daño innecesario e inviable al paciente. La imposibilidad de recoger a través de técnicas quirúrgicas muestras del tejido diana del fármaco, obliga a la recolección no invasiva de fluidos corporales en los que se buscarán y analizarán los biomarcadores que permitan seleccionar los pacientes que responderán con un mejor coeficiente eficacia / toxicidad al tratamiento con el fármaco, y que también ayudarán al ajuste de la posología, idealmente personalizado en función del biomarcador. En esos fluidos corporales, suero, orina, jugo gástrico, ..., no existen células procedentes del tejido diana, salvo excepciones como células descamadas de tejidos del tracto urinario en la orina, y por tanto no es posible realizar un análisis genómico, ya que no hay DNA o RNA que analizar; pero sí hay proteínas que se han secretado desde los tejidos diana, que pueden ser identificadas y analizadas con técnicas proteómicas (29, 30).

La comparación, durante las fases I y II, del perfil proteómico del suero, u otro fluido, entre individuos respondedores y no respondedores a un fármaco puede rendir:

- La identificación de proteínas biomarcadoras antes de la administración del fármaco, que diferencian aquellos individuos que posteriormente han respondido bien al fármaco, de aquellos en los que el fármaco ha provocado una mayor toxicidad y de aquellos en los que el fármaco ha presentado unos niveles de actividad ineficaces para combatir la enfermedad.

Estos biomarcadores se analizarán en los candidatos a participar en las fases III y IV posteriores, antes del inicio del ensayo, para seleccionar los pacientes que participarán en los ensayos y para ajustar, idealmente de una manera personalizada, las dosis de administración.

- La identificación de proteínas biomarcadoras tras la administración del fármaco, que diferencian aquellos individuos que han respondido bien al fármaco, de aquellos en los que el fármaco ha provocado una mayor toxicidad y de aquellos en los que el fármaco ha presentado unos niveles de actividad ineficaces para combatir la enfermedad.

Estos biomarcadores se analizarán en los individuos que participan en las fases III y IV del ensayo, para suspender el tratamiento o para ajustar la posología, si es posible.

En las fases III y IV, no sólo se utilizarán los biomarcadores identificados en las fases I y II, sino que se validará el uso de esos biomarcadores en la pos-

terior administración del fármaco ya registrado y comercializado y se buscarán nuevos biomarcadores que faciliten no sólo la selección de pacientes a tratar y la determinación de la posología, sino también la monitorización del fármaco en administraciones prolongadas a enfermos crónicos.

Nuestra empresa, Proteomika, SL, ha comparado durante los dos últimos años los perfiles proteicos antes y después del tratamiento con fármacos contra la Hipercolesterolemia Familiar y la Enfermedad de Gaucher. La asociación de los perfiles proteómicos del suero de los individuos tratados, a la respuesta al tratamiento de esos individuos, ha permitido la identificación de biomarcadores que están en fase de validación clínica.

3.5. Ensayos proteómicos en la administración de fármacos registrados

Los biomarcadores identificados en los ensayos clínicos permitirán desarrollar sencillos métodos no invasivos de detección de esos biomarcadores en fluidos corporales, con tecnologías inmunoquímicas de fácil aplicación en la rutina hospitalaria de los laboratorios de análisis clínicos.

- Biomarcadores a analizar antes del comienzo del tratamiento, que faciliten la selección de pacientes a tratar.
- Biomarcadores a analizar un período corto de tiempo tras el comienzo del tratamiento, que faciliten la determinación de la posología, o que indiquen la suspensión del tratamiento.
- Biomarcadores a analizar periódicamente tras el comienzo del tratamiento, que faciliten la monitorización del fármaco en administraciones prolongadas a enfermos crónicos.

El análisis proteómico comparativo en fluidos corporales, con el objetivo de desarrollar métodos de diagnóstico, pronóstico y monitorización de la respuesta a fármacos, es actualmente la principal aplicación de la Proteómica. Muchos grupos, institutos y empresas como la nuestra centran sus recursos humanos y materiales en la optimización de los protocolos para analizar los proteomas de esos fluidos. El proteoma del suero y plasma es el principal objeto de la Organización Proteoma Humano (HUPO), de la que estamos orgullosos de formar parte en Proteomika, SL, y que es el proyecto científico internacional más ambicioso tras el Proyecto Genoma Humano, que coordinó los esfuerzos públicos internacionales en la secuenciación del Genoma Humano. Tendrá que aparecer una empresa americana con 350 millones de dólares de presupuesto para iden-

tificar el proteoma humano, para que las administraciones públicas reparen en la existencia de HUPO; tiempo al tiempo, la Ciencia no es posible sin toneladas de paciencia, pero mientras, millones de seres humanos se mueren todos los días por diagnósticos o tratamientos erróneos o inexistentes.

REFERENCIAS

- (1) Pandey, A. & Mann, M. Proteomics to study genes and genomes (2000). *Nature* 405, 837-846.
- (2) Patterson, S.D. & Aebersold, R.H. Proteomics: the first decade and beyond (2002). *Nature Genetics* 33, supplement 311-323.
- (3) Blagoev, B. *et al.* A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling (2003). *Nat. Biotechnol.* 21, 315-318.
- (4) Bader, G. D., Betel, D. & Hogue, C. W. BIND: The biomolecular interaction network database. (2003). *Nucleic Acids Res.* 31, 248-250.
- (5) Mann, M. & Jensen. Proteomic analysis of post-translational modifications (2003). *Nature Biotechnology* 21, 255-261.
- (6) Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network (2001). *Science* 292, 929-934.
- (7) O'Farrell, P.H. High resolution two-dimensional gel electrophoresis of proteins (1975). *J. Biol. Chem.* 250, 4007-4021.
- (8) Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y. & Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology (2000). *Proc. Natl. Acad. Sci. USA* 97, 9390-9395.
- (9) Rabilloud, T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains (2002). *Proteomics* 2, 3-10.
- (10) Henzel, W.J. *et al.* Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases (1993). *Proc. Natl. Acad. Sci. USA* 90, 5011-5015.
- (11) Gygi, S.P., Rist, B., Griffin, T.J., Eng, J. & Aebersold, R. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags (2002). *J. Proteome Res.* 1, 47-54.
- (12) Washburn, M.P., Ulaszek, R., Deciu, C., Schieltz, D.M. & Yates, J.R. III. Analysis of quantitative proteomic data generated via multidimensional protein identification technology (2002). *Anal. Chem.* 74, 1650-1657.

- (13) Glish, G.L. & Vachet, R.W. The basics of mass spectrometry in the twenty-first century (2003). *Nature Reviews Drug Discovery* 2, 140-150.
- (14) Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules (1989). *Science* 246, 64-71.
- (15) Caprioli, R.M., Farmer, T.B. & Gile, J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS (1997). *Anal. Chem.* 69, 4751-4760.
- (16) Eng, J.K., McCormack, A.L. & Yates, J.R.I. An approach to correlate MS/MS data to amino acid sequences in a protein database (1994). *J. Am. Soc. Mass Spectrom.* 5, 976-989.
- (17) Sadygov, R.G., Cociorva, D. & Yates J.R. III. Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book (2004). *Nature Methods* 1, 195 - 202.
- (18) Medzihradszky, K.F. *et al.* The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer (2000). *Anal. Chem.* 72, 552-558.
- (19) Kramer, R. & Cohen D. Functional Genomics to New Drug Targets (2004). *Nature Reviews Drug Discovery* 3, 965-972.
- (20) Zeeberg, B. R. *et al.* GoMiner: a resource for biological interpretation of genomic and proteomic data. (2003). *Genome Biol.* 4, R28.
- (21) Boguski, M.S. & McIntosh, M.W. Biomedical informatics for proteomics (2003). *Nature* 422, 233-237.
- (22) Johnson, S.A. & Hunter, T. Kinomics: methods for deciphering the kinome (2005). *Nature Methods* 2, 17-25.
- (23) Koike, A., Kobayashi, Y. & Takagi, T. Kinase pathway database: An integrated protein-kinase and NLP-Based protein-interaction resources (2003). *Genome Res.* 13, 1231-1243.
- (24) Gomez-Roman, J.J., Saenz, P., Cuevas Gonzalez, J., Escuredo, K., Santa Cruz, S., Junquera, C., Simon, L., Martinez, A., Gutierrez Banos, J.L., Lopez-Brea, M., Esparza, C. & Val-Bernal, J.F.. Fibroblast growth factor receptor 3 is overexpressed in urinary tract carcinomas and modulates the neoplastic cell growth (2005). *Clin Cancer Res.* 15:459-65.
- (25) Rangarajan, A. & Weinberg, R. A. Comparative biology of mouse versus human cells: modelling human cancer in mice (2003). *Nature Rev. Cancer* 3, 952-359.
- (26) Zhu, H. *et al.* Global analysis of protein activities using proteome chips (2001). *Science* 293, 2101-2105.

- (27) Mitchell, P. A. Perspective on protein microarrays (2002). *Nat. Biotechnol.* 20, 225-229.
- (28) Petricoin, E.F., Zoon, K.C., Kohn, E.C., Barrett, J.C. & Liotta, L.A.. Clinical proteomics: translating benchside promise into bedside reality (2002). *Nature Reviews Drug Discovery* 1, 683-695.
- (29) Petricoin, E.F. *et al.* Use of proteomic patterns in serum to identify ovarian cancer (2002). *Lancet* 359, 572-577.
- (30) Adam, B.L. *et al.* Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men (2002). *Cancer Res.* 62, 3609-3614.