

# EL PROTEOMA, UNA HERRAMIENTA PARA EL ANÁLISIS GLOBAL DEL ESTADO CELULAR DE ORGANISMOS MICROBIANOS.<sup>1</sup>

Sergio Manuel Encarnación Guevara

Las técnicas de proteoma (electroforesis de alta resolución en dos dimensiones y la caracterización de proteínas) son usadas en gran escala en la investigación microbiana para analizar la síntesis global de proteínas como un indicador de la expresión genética de la bacteria. Los rápidos progresos en proteomas microbianos se deben a la gran disponibilidad de secuencias de genomas que varios grupos de investigación han reportado.

Además de proveer un entendimiento básico de la expresión genética en microorganismos, el proteoma juega también un papel fundamental en importantes áreas de la microbiología tanto médica como agrícola. Notables progresos han sido realizados con la utilización de esta metodología, en campos como son la epidemiología, la taxonomía de patógenos microbianos y bacterias de interés agrícola, además en la identificación de nuevos mecanismos de patogenicidad y en el análisis de la resistencia bacteriana a los antibióticos. En cada una de estas áreas, el proteoma a provisto de nueva información que complementa las investigaciones basadas en los genomas bacterianos. Este capítulo describe los progresos más recientes en el campo de la investigación realizada con proteomas

además de algunos de los más importantes retos técnicos existentes para la aplicación de esta técnica en la microbiología.

En microbiología médica es aún un reto para el futuro la caracterización de proteomas de patógenos o simbioses bacterianos creciendo o desarrollando sus actividades fisiológicas en su huésped o compañero natural sin embargo en microbiología agrícola algunos grupos de investigación como el nuestro ya estamos abordando este tipo de estudios.

Iniciaremos este capítulo con un análisis de las técnicas y metodologías más usadas en el proteoma, sus principios, para posteriormente entrar de lleno en lo que es esta importante técnica de biología moderna, seguido de un análisis de los conceptos más importantes que nos ayuden a entender este manuscrito, para posteriormente pasar al análisis de proteomas en bacterias y profundizar lo más posible en sus posibilidades, utilización actual y perspectivas.

## Introducción

La Biología moderna empezó con la meta de definir las propiedades del material genético y la manera de como esta información es utilizada. En la actualidad se sabe que el material genético esta constituido de un tipo particular de moléculas (ácidos nucleicos), los

---

<sup>1</sup>Centro de Investigación Sobre Fijación de Nitrógeno, Universidad Nacional Autónoma de México

cuales son diferentes en estructura de los otros tipos de moléculas (proteínas, lípidos, azúcares etc.). El gene es la unidad genética de información y el paradigma básico consiste en que los genes codifican proteínas, las cuales son responsables en su momento de la síntesis de otro tipo de estructuras, incluidos los ácidos nucleicos. La secuencia de un gene, especificará la secuencia de la proteína y por lo tanto su estructura molecular. Pero las características individuales de estructura no determinaran exclusivamente la función de la molécula, la localización y el estado fisiológico del organismo son importantes. La producción de una proteína funcional ocurre después de una serie de interacciones. La secuencia de un gene especificará la secuencia de la proteína. La secuencia de la proteína especificará la estructura de la molécula. La estructura molecular determinará la ubicación de la proteína en la célula. Es importante notar que el orden de eventos puede diferir en diferentes grupos de proteínas; muchas proteínas alcanzan estructura antes de localización otras obtienen su localización antes que su estructura. En cada estado metabólico un aparato específico interpreta la información proporcionada a un nivel jerárquico para proceder posteriormente al siguiente nivel. Este nivel jerárquico puede ser resumido en algunos pasos siendo el primero la perpetuación del DNA o de sus genes por medio de replicación, posteriormente la transcripción en RNA (RNAm) y la traducción de este para producir la proteína, lo anterior utilizando el código genético, el paso jerárquico siguiente es el armado de la proteína mediante "chaperonas", lo cual proporcionará la estructura que definirá si esta proteína será localizada en el citosol o asociada a membranas, definiendo esto su papel estructural o actividad catalítica. Con lo anterior podemos concluir que el DNA contiene la

Información pero las proteínas proveen la manera de ejecutarla.

Para muchos microorganismos, los proyectos de investigación se encuentran ya en la era post-genómica. La secuencia de DNA de estos microorganismos está siendo analizada con varias herramientas computacionales para extraer la mayor cantidad de información teórica. Esta estática descripción teórica de la célula es actualmente acompañada por dos metodologías muy dinámicas conocidas como proteomas y transcriptomas. El transcriptoma se refiere al perfil de transcritos de la célula (RNAm) (132). El proteoma se refiere al producto de la traducción de estos RNAm y las diferentes variantes o modificaciones generadas por la célula después de este proceso. Antes de 1975 las proteínas celulares eran estudiadas por medio de métodos bioquímicos como entidades individuales o como una mezcla de las proteínas de la célula (por ejemplo la concentración total de la proteína de la célula, o su tasa de síntesis) todo por medio de métodos químicos o radio químicos. Desde que O'Farrel y Klose en 1975 (99) demostraron que era posible separar una muestra compleja de proteínas en geles de acrilamida de doble dimensión, basados en su punto isoeléctrico y peso molecular (2-ED) esta metodología se ha mantenido como la forma más viable de obtener resultados confiables, aunado a esto, modernos formatos de geles, más grandes (ejemplo 23 x 28 cm) son altamente reproducibles y la tinción por Coomassie o tinción por plata permiten la cuantificación, además tinciones fluorescentes o mareaje radiactivo permiten la detección de miles de proteínas en un gel. Este método es comúnmente usado para estudiar la arquitectura proteica de la célula, conociendo la cantidad y localización subcelular de proteínas individuales (38), pero también ha sido usado en muchos estudios para examinar los cambios en la expresión de proteínas en respuesta a estímulos (97), o para identificar proteínas regulatorias (92). En todas esas aplicaciones,

la electroforesis en geles de doble dimensión es usada para localizar y aislar proteínas de interés. Esas aplicaciones han hecho y continuaran contribuyendo con aportaciones importantes para entender como funcionan las proteínas en la célula. Cada grupo de proteínas expresadas en un momento metabólico específico de la célula podría ser llamada la firma celular, relacionándola con el estado de trabajo de la célula en un proceso metabólico particular.

El estado fisiológico de la célula en una circunstancia particular es obtenido de la unión de las firmas metabólicas observadas. El progreso en este campo se basa en la construcción de una tabla de correspondencia, ligando las firmas a las funciones celulares.

La combinación del proteoma con genética, biología molecular, bioquímica y biofísica de proteínas han producido una técnica de gran alcance, dando por resultado 2 novedosos métodos para analizar mezclas complejas de la proteína. Las tecnologías de proteoma que emergen prometen aumentar la velocidad de procesamiento de identificación de las mezclas complejas de proteínas y conocer sus niveles de expresión.

Como mencionamos en el párrafo anterior, el análisis tipo proteoma es el resultado de la combinación de una serie de técnicas que han sido reportadas y perfeccionadas durante los últimos años, técnicas como la electroforesis en geles de doble dimensión, de tinción altamente sensibles, de caracterización de proteínas mediante secuencia de la porción amino o carboxilo terminal y la utilización de la espectrometría de masas dirigida a la identificación de proteínas presentes en concentraciones pico molares, son algunas metodologías actualizadas para su uso en el análisis de expresión global del genoma mediante las proteínas que sintetiza la célula, mejor conocido como proteoma.

Pero, ¿que es la electroforesis de proteínas? La electroforesis de proteínas

en geles de poliacrilamida es una técnica indispensable en el análisis y estudio de las proteínas. Cada tipo de separación electroforética puede ser llevada a cabo en una gran variedad de tamaños de geles, desde minigeles que no son más grandes que un sello postal hasta geles gigantes más grandes que esta página. La electroforesis en geles de poliacrilamida puede ser de dos tipos, de una dimensión y dos dimensiones, esta última será la que ocupe nuestra atención. La electroforesis en geles de doble dimensión involucra la combinación ortogonal de dos diferentes métodos electroforéticos. La técnica más común de electroforesis en geles de doble dimensión combina dos métodos individuales de alta resolución, consistentes en el isoelectroenfoco seguido por la electroforesis con Dodecil Sulfato de Sodio (SDS) (Fig. 1). Cada uno de esos métodos es individualmente capaz de separar alrededor de 100 proteínas pero cuando son combinados, más de 1500 proteínas pueden ser separadas en un gel de doble dimensión. Por lo tanto este método puede ser usado para analizar y comparar mezclas complejas de proteínas, incluyendo extractos totales de células o tejidos.

Después de la separación electroforética, las proteínas son detectadas usualmente con métodos comunes de tinción como la tinción con azul de Coomassie o uno más sensible que es la tinción con plata. Esos métodos de tinción fijan la proteína en la matriz del gel para prevenir la difusión de las manchas proteicas.

Un método de análisis muy usado después del aislamiento de proteínas en geles de doble dimensión es el que involucra la transferencia de estas, en un campo eléctrico a una membrana o soporte inerte como pueden ser las membranas de difluorido de polivinilideno (PVDF), esta técnica suple la electro-elusión de las proteínas y es muy utilizada ya que las membranas de PVDF son fáciles de manejar y químicamente inertes, por lo tanto son compatibles con subsecuentes métodos de análisis que incluyen análisis de aminoácidos,

de modificación química in situ, digestión proteica in situ y análisis de secuencia AMerimal, además la estabilidad y resistencia química de las membranas de PVDF provee un medio óptimo para la combinación de múltiples métodos de detección como son métodos inmunológicos con el uso de anticuerpos.

La combinación de la tecnología del proteoma con otras técnicas, ha dado lugar al descubrimiento acelerado de la información funcional de las proteínas.

Aunque esta técnica es de gran alcance, tiene sus limitaciones. Además de ser necesarios geles laboriosos, en este tipo de geles raramente se identifican las proteínas hidrofóbicas y las de baja abundancia. Recientemente nuevos métodos de análisis de proteomas se están desarrollando para poder resolver estas clases de proteínas y acelerar la velocidad del análisis. Finalmente, se están desarrollando métodos cuantitativos permitiendo conocer la abundancia y síntesis relativa de estas proteínas. Este capítulo presenta y discute una amplia variedad de aplicaciones del proteoma a la microbiología. Además, se discuten metodologías utilizadas actualmente en proteoma así como métodos de la cuantificación que pueden revolucionar el estudio de proteomas.

### **¿Por qué es necesario trabajar con el proteoma?**

Con la generación constante de extensas cantidades de secuencias en las bases de datos de ADN, se ha observado que tener simplemente las secuencias completas de genomas no es suficiente para aclarar la función biológica. Una célula es normalmente dependiente para su supervivencia de una multiplicidad de caminos metabólicos y reguladores. No existe una relación lineal terminante entre los genes y el complejo proteico o el "proteoma" de una célula. El proteoma es complementario al genoma porque

se centra en los productos del gene, que son los agentes activos en las células. La existencia de una fase de lectura abierta (ORF) o las secuencias genómicas, no implican necesariamente la existencia de un gene funcional. Además a pesar de los avances en bioinformática, sigue siendo difícil predecir genes de datos geonómicos (29,32,70). Una ventaja en la constante secuenciación de genomas de microorganismos será que la secuenciación de genomas de organismos relacionados facilitará la predicción del gene con base a una comparación genómica, sin embargo la tasa de éxito para la predicción correcta de la estructura primaria es aún baja (211Q3). Fundamentalmente se observan problemas de predicción de fases de lectura en genes pequeños (que se puedan perder completamente) o en genes con poco o nada de homología a secuencias conocidas o previamente reportados. Un estudio reciente concluyó que la tasa de error en la anotación o definición de genes del genoma de *Mycoplasma genitalium* constituido por 340 genes es de al menos el 8%. Si tales tasas de error se extrapolan al genoma humano, el resultado y las consecuencias pueden ser fácilmente calculadas. Por lo tanto, la verificación del producto de un gene por el método de proteoma es un importante primer paso en la "anotación" de los genomas. Por otro lado las modificaciones de las proteínas que no son evidentes a partir de la secuencia de ADN (ácido desoxiribonucleico), tal como isoformas y modificaciones post-traduccionales, se pueden determinar solamente por metodologías proteómicas. Además, si pretendemos estudiar la regulación en la expresión del genoma será necesario correlacionar el nivel de la expresión de la proteína directamente con los niveles de RNAm aunque en varios casos estos pueden o no correlacionar con la presencia de la proteína(42,48).

### **Combinación de proteomas con modernas técnicas biológicas.**

Como mencionamos en párrafos anteriores, el potencial real de los proteomas surge cuando el proteoma se combina con genética, biología molecular, bioquímica y/o biofísica de proteínas. Combinando estudios de proteoma con genética y biología de levaduras, Lee et al. (74) descifraron los papeles que juegan las proteínas Yap1 y Skn7 en la respuesta al estrés oxidativo, por otro lado Alms et al. (3) identificaron fosfostratos del complejo fosforilado de la proteína Reg1, un complejo involucrado en la regulación de las vías metabólicas que son reprimidas por glucosa. Estudios adicionales que utilizan la tecnología del proteoma, purificación de proteínas y genética, identificaron las proteínas implicadas en ensamblaje de la actinia (44) además de un nuevo componente del nucléolo mitocondrial requerido en la reparación del ADN (83) además ha servido en la clasificación de los componentes del complejo de la histona acetiltransferasa de levadura (31). Rigaut et al. (110) desarrollaron un método que han aplicado para caracterizar las pequeñas partículas del ácido ribonucleico de la levadura, complejo implicado en empalmar el RNAm (19,34) y la proteína Dbp5 de la caja DEAD (116). Combinando el proteoma y biología estructural, Houry et al. (57) identificaron in vivo 52 de los substratos de la chaperonina GroEL, una proteína de E coli implicada en el plegamiento de polipéptidos, además determinaron los motivos comunes en 24 de los substratos de esta proteína. Sin el proteoma, los complejos estudios emprendidos por Houry et al no hubieran sido posibles.

### **Identificación y análisis de proteínas**

Uno de los pasos cruciales en proteomas es la obtención y manejo de la muestra proteica. Debemos considerar que existen genomas microbianos de cerca de 10 000 genes, o líneas celulares potencialmente capaces de regular la expresión de un

número incluso más alto de genes. Además, el rango dinámico de la abundancia de proteínas en muestras biológicas en este último ejemplo puede ser tan alto como 10<sup>6</sup>. Si consideramos que en los mejores geles de dos dimensiones de una mezcla cruda de proteínas solo se puede resolver rutinariamente 1,000 proteínas, es obvio que solamente se pueden visualizar mediante electroforesis en geles, las proteínas más abundantes. La solución ideal para reducir la complejidad y las diferencias en abundancia es utilizar estrategias de purificación de proteínas a partir del complejo proteico a analizar. Por ejemplo, existen receptores de abundancia media, que se encuentran en cerca de 1000 copias por la célula, o menos de dos picomoles, (100 ng) en un litro de cultivo celular. Estas proteínas no serían visualizadas en los extractos de celulares si no se enriquece la muestra previamente mediante purificación utilizando una columna de afinidad con anticuerpos.

Este hecho debe ser tomado en cuenta por los que trabajamos con proteomas, si consideramos que este puede ser el caso de algunas moléculas reguladoras. Después de obtener la fracción de proteínas, el método de opción inmediato para este tipo de análisis es la electroforesis en geles de dos dimensiones.

### **Identificación espectrométrica de proteínas**

Uno de los descubrimientos más significativos para la investigación en proteomas, ha sido la identificación por medio de espectrometría de masas, de proteínas separadas mediante electroforesis en geles de doble dimensión, lo cual ha ampliado el análisis de estas más allá de su simple visualización. La espectrometría de masas esencialmente ha substituido la técnica clásica de secuenciación de la porción N-terminal de la proteína, por medio de degradación Edman, sustituyéndola incluso en la química tradicional de proteínas, esto es porque es mucho más sensible, puede ocuparse de mezclas de proteínas y

tiene un rendimiento de procesamiento muy alto, lo anterior esta basado en la cantidad de proteínas que pueden analizarse, en ocasiones grupos de 100 proteínas en pocos minutos, algunas de ellas en concentraciones pico molares. Esta metodología analiza las proteínas separadas en geles de doble dimensión las cuales son digeridas por una proteasa secuencia-específica como la tripsina o quimiotripsina y los masas de los péptidos son analizadas y determinadas en un espectrómetro de masas. La razón de analizar los péptidos preferentemente a las proteínas es porque las proteínas separadas en el gel son difíciles de extraer del gel que esta constituido de acrilamida, además de que el conocimiento de la masa total de la proteína generalmente no es suficiente para la identificación de estas en las bases de datos. En contraste, los péptidos se eluyen fácilmente del gel de acrilamida y un conjunto pequeño de péptidos de una proteína proporciona suficiente información para su identificación. La progresión de pasos típicamente implicados en un análisis mediante espectrometría de masas de una proteína son ilustrados por el ejemplo que muestra el análisis de una molécula (Fig. 2b y 3). Un protocolo detallado que describe métodos y las estrategias para la identificación mediante espectrometría de masas se puede encontrar en la referencia (107).

Existen dos estrategias principales para la identificación espectrométrica de las proteínas. Una es el "mapeo de la masa del péptido", sugerido inicialmente por Henzel y colaboradores (52), en esta técnica se obtiene el espectro total de la mezcla de los péptidos obtenidos de la digestión secuencia-específica de la proteína, lo que da lugar a una huella digital de la masa de los péptidos de la proteína que esta siendo estudiada. Este espectro total es obtenido por un método de espectrometría de masas relativamente simple utilizando una fuente de ionización y

la desorción de muestras asistido por un láser (MALDI-TOF, por sus siglas en inglés) lo que resulta en una medición del "tiempo de vuelo" de los péptidos que constituyen la mezcla de péptidos (Fig. 3 y 2b). Recientes avances en la automatización del procedimiento de identificación de proteínas mediante MALDI han hecho que cientos de proteínas puedan ser cortadas de los geles, digeridas enzimáticamente, su espectro de masas obtenido analizado automáticamente e identificado en las bases de datos de proteínas (12,62). En un procedimiento de dos etapas para la identificación rápida e inequívoca de la proteína, la huella digital de la proteína obtenida mediante MALDI sería el primer paso (119,120). El segundo paso será la secuencia de péptidos individuales. En este método, los péptidos son ionizados directamente por ionización en electroaspersión a partir de la fase líquida. Los iones de los péptidos se rocían en un espectrómetro de masas en serie que tenga la capacidad de separar los péptidos de una mezcla, además de aislar una especie al mismo tiempo y de disociarla en los fragmentos amino o carboxilo-terminal (Fig. 1c). El método de espectrometría de masas en serie es técnicamente más complejo y menos escalable que la "huella digital" de la proteína obtenida por MALDI. Su ventaja principal es que la información obtenida de la secuencia de aminoácidos de la porción amino o carboxilo terminal derivada del análisis de varios péptidos es mucho más específica para la identificación de una proteína que una lista de las masas de los péptidos que la constituyen. Los datos de la secuencia no solo se pueden utilizar para identificar proteínas en bases de datos de secuencias de aminoácidos sino también en bases de datos de secuencias de nucleótidos y bases de datos de genomas.

### **Análisis de modificaciones postraduccionales de las proteínas**

Una de las ventajas de las investigaciones tipo proteoma es la capacidad única que

ofrece este tipo de estudios de analizar las modificaciones post-traduccionales de las proteínas. La fosforilación, la glicosilación y la sulfatación así como muchas otras modificaciones son extremadamente importantes para la función de las proteínas pues pueden determinar la actividad, estabilidad, localización y el recambio de las mismas. Estas modificaciones no son evidentes en datos de secuencia de genomas, o en estudios de expresión de RNAm. La espectrometría de masas es el método de proteoma de elección para determinar las modificaciones postraduccionales de las proteínas, sin embargo, esta tarea es mucho más difícil que la misma identificación de las proteínas. Mientras que para la identificación de las proteínas en las bases de datos, es necesaria en ocasiones la secuencia de uno o dos péptidos, para conocer la naturaleza y la localización de modificaciones postraduccionales, se necesitan analizar detalladamente todos los péptidos que no tienen la masa molecular prevista. Debido a esto y otras razones, es necesaria una mayor cantidad de material para estudiar las modificaciones postraduccionales que el utilizado para la identificación de la proteína. Un avance continuo se está teniendo en este campo, especialmente en el caso de estudios de fosforilación. Las fosforilaciones se pueden identificar por estrategias genéricas, ya que los fosfopéptidos son 80 daltones más pesados que sus contrapartes sin modificar, dando lugar a un fragmento específico (P03-, masa 79), los cuales se unen en una microcolumnas de resinas metálicas para posteriormente ser reconocidos por anticuerpos específicos, adicionalmente los grupos fosfato pueden ser removidos por fosfatasas y analizados por medio de MALDI (13, 23, 93, 96, 144). Como ejemplo, figura 2d.

### **Bases de datos de proteínas y proteomas**

Varias parciales bases de datos de proteomas se encuentran disponibles actualmente, incluyendo mapas de varios organismos, tales como *Escherichia coli*, *H. influenzae*, *Saccharomyces cerevisiae*, *Bacillus subtilis*, siendo las dos primeras bases de datos las que agrupan más proteínas identificadas, en el caso de la de *H.influenzae* cerca de 500 proteínas han sido identificadas. Muchas de estas bases de datos de proteomas son disponibles de manera electrónica, por ejemplo en el servidor ExPASy, el cual es accesible vía World Wide Web (<http://www.expasy.ch>).

Quizás el análisis tipo proteoma más detallado se ha realizado en *Saccharomyces cerevisiae*. La actualización más reciente de su proteoma muestra 279 productos proteicos diversos identificados de 401 proteínas resueltas mediante geles de doble dimensión (106). Debido a la cantidad de información publicada sobre *S. cerevisiae*, especialmente sobre la secuencia del genoma, existe una gran cantidad de bases de datos que tienen clasificadas todas las proteínas conocidas de levadura. Esto ha hecho posible un catálogo de estas en la base de datos del proteoma de este organismo, incluso la clasificación es con base a su papel funcional y a la localización subcelular (24,84). Por lo anterior un análisis tipo proteoma de *S. cerevisiae* puede correlacionarse fácilmente con estos grupos de clasificación e identificar la función de la proteína y su papel en el metabolismo.

Utilizando la información genómica y proteómica disponible para *S.cerevisiae*, algunos grupos de investigación han analizado la posible correlación entre el mRNA y la presencia de la proteína. Con base a un análisis logarítmico, Fitcher et al. (42) descubrieron que la concentración de 148 proteínas correlaciona con la abundancia en RNAm. Sabemos que no en todos los casos se ha encontrado este tipo de correlación, por lo tanto estudios en otros organismos que correlacionen el RNAm y los niveles de expresión de las proteínas darán

información valiosa de la organización y la expresión de sus genomas proporcionando en el futuro al conocimiento de las reglas comunes que regulan la expresión de las proteínas.

### **El Proteoma en la Internet**

Dada la cantidad de información generada por los diversos grupos de investigación que trabajan con proteomas en diversos organismos, existe un importante esfuerzo para hacer esta información fácilmente disponible en la Internet. Las recientes actualizaciones de muchos proteomas microbianos, incluyendo los de la cianobacteria *Synechocystis* (115), *S. cerevisiae*, (106), *M. tuberculosis* y *M. bovis* (85), están disponibles en sus respectivas ventanas de Internet donde se pueden consultar como bases de datos interactivas. Cordwell et al. (22) recientemente publicaron una aproximación que permite integrar toda la información generada por El Centro Australiano de Análisis de Proteomas, con proteomas obtenidos de *E. coli*, *Pseudomonas auhginosa* y *Staphylococcus aureus*. Lo anterior se debe a que no toda la información generada por proyectos tipo proteoma puede ser presentada en una publicación y debe ser accesible de alguna otra forma a los grupos de investigación interesados, por lo cual la Internet es la herramienta ideal para este propósito.

### **Definiciones**

En 1994, casi 20 años después de que la metodología de los geles de doble dimensión fue introducida, el término proteoma fue utilizado por primera vez. Actualmente han sido sugeridas muchas derivaciones de la palabra y todas son usadas para describir el análisis de proteomas. El proteoma de un organismo es el complejo de proteínas que este puede producir. Los proteomas pueden ser

determinados sin ambigüedad por los métodos electroforéticos o descritos de manera teórica como el predicho por Blattner (14) para *E.coli* MG1655 el cual consiste de 4405 proteínas o el construido por nuestro grupo para el plásmido simbiótico de *Rhizobium* NGR234. Estos proteomas pueden ser observados en las figuras 4 y 5. Estas figuras muestran el rango de peso molecular y punto isoeléctrico teórico de las proteínas de un organismo y puede ayudar a decidir una estrategia adecuada para separar óptimamente las proteínas, estos proteomas también pueden revelar otras características interesantes. Por ejemplo los proteomas parecen tener una distribución bimodal, probablemente relacionada con el pI de las proteínas y el pH intracelular de la célula. Proteoma, puede ser definido como el análisis en gran escala de proteínas, lo cual contribuirá enormemente a nuestra comprensión de la función de los genes en la era post-genómica. Proteoma se puede dividir en tres áreas principales: (i) micro-caracterización de las proteínas y de sus modificaciones postranscripcionales; (ii) proteoma de visualización diferencial para la comparación de los niveles de síntesis de las proteínas con la aplicación potencial en una amplia gama de enfermedades microbianas; y (iii) estudios de las interacciones proteína-proteína usando técnicas tales como espectrometría de masas. Como es a menudo difícil predecir la función de una proteína basada en la homología a otras proteínas o aún a su estructura tridimensional, la determinación de componentes de una proteína compleja o de una estructura celular es central en el análisis funcional. Este aspecto en estudios tipo proteoma es quizás el área de trabajo más prometedora. Después de la revolución en la biología molecular, el proteoma, en los años por venir jugará un importante papel en la comprensión de la bioquímica de proteínas y de vías metabólicas.

Proteoma es un término usado para definir las proteínas de células, tejidos y organismos

incluyendo las interacciones de estas con el medio ambiente. El perfil de expresión de proteínas es el catálogo cuantitativo de proteínas sintetizadas por la célula bajo una circunstancia determinada. Por ejemplo, cada gel de doble dimensión o grupo de geles de doble dimensión obtenidos de una muestra revelan mucho de un perfil particular de proteínas. Obtener un perfil de expresión de proteínas completo no es técnicamente posible en este momento debido a dificultades asociadas con la solubilidad de proteínas, pl extremos y baja abundancia de algunas de ellas.

Fenotipo proteico es un carácter o estado de una proteína en especial en una condición y tiempo definido. Este término fue introducido por primera vez por J.Klose (67).

La información en el fenotipo proteico incluye: la cantidad, la tasa de síntesis, la tasa de degradación, las modificaciones postraduccionales y la actividad de la proteína. Por ejemplo, fenotipo proteico de la fosfatasa alcalina en *E.coli* incluye la información acerca de esta proteína para cada condición de crecimiento de la célula. Los fenotipos proteicos pueden incluir resultados de muchas fuentes. La más completa colección de fenotipos proteicos fue compilada por Garreis et al, (56) para *Saccharomyces cerevisiae*.

Regulón es un grupo de proteínas cuya síntesis es regulada por la misma proteína reguladora (91) El análisis genético ha sido utilizado clásicamente para encontrar y estudiar estos regulones, pero los proteomas son de mucha más utilidad para identificar los miembros de estos regulones. Por ejemplo, comparando el perfil de expresión de proteínas de una cepa silvestre y una mutante, pueden identificarse proteínas miembros de un mismo regulón.

Estimulón es un grupo de proteínas para las cuales se puede determinar la

cantidad y tasa de síntesis de estas en respuesta a un estímulo, varios estímulos han sido descritos (91). Un estimulón puede consistir de varios regulones.

Firma proteómica es el grupo de proteínas en las cuales la alteración en su expresión es característica de una respuesta a una condición definida o cambio genético. Las firmas con frecuencia relacionan vías metabólicas específicas y funciones (Fig. 8).

### **El uso de perfiles de expresión de proteínas**

La correlación entre la expresión de proteínas y la fisiología de la célula conduce a la obtención de perfiles de proteínas. En la tabla 1a, tabla 1b y figura 9, se puede consultar un listado y un diagrama de flujo que ilustra una serie de procesos necesarios para la obtención de un perfil de expresión de proteínas. El seguimiento de estos pasos producirá la información necesaria para la obtención del fenotipo proteico, estímulos, regulones y firmas proteómicas. Una parte de la información obtenida por medio de proteoma estará basada en características tales como cambios en la cantidad y tasa de síntesis de cada proteína. Por ejemplo, una proteína podría ser considerada en base a cambios cualitativos (ausencia o presencia) o en cambios cuantitativos. Otro tipo de características a estudiar son los cambios de relación. La tasa de abundancia entre dos proteínas en un perfil de expresión individual o entre dos subunidades de una sola proteína y puede ser tomado en cuenta como un cambio en la regulación de la función.

Fenotipos proteicos, estímulos, regulones y firmas proteómicas pueden y han sido encontrados utilizando simples métodos de análisis de imágenes y también con la utilización de métodos computarizados de análisis de imágenes como el pdQuest y el Melanie3. La profundidad del análisis en microbiología corresponderá a la profundidad y metas del proyecto. Un análisis comprensible de cada proteína que sea inducida o reprimida en 1.5 veces en una

condición comparada con otra, puede ser excesivo si la meta fue solo realizar un reconocimiento o búsqueda de marcadores proteicos. Por otro lado este tipo de análisis sería justificado si la meta fuese la identificación de todas las proteínas controladas por una cierta proteína reguladora. En algunos casos la meta es el definir fenotipos adicionales de una proteína en particular. En otros casos será necesario identificar estímulos y regulones que puedan ayudar a revelar la firma o firmas celulares que provean la información necesaria para entender la reacción celular a un estímulo.

Estímulos son directamente identificados usando los perfiles de expresión de proteínas. Comparando perfiles de expresión de una condición control con una condición de interés revelará el estímulo de la condición de interés. En una serie de estudios cualitativos realizados en *Vibrio sp.*, varios estímulos fueron identificados y revelaron la potencial función de la célula (97).

Los regulones son usualmente identificados de dos maneras: (i) usando cepas en las que es mutado el gene que codifica para una proteína reguladora de interés, y (ii) examinando la respuesta a múltiples condiciones de prueba y analizando las proteínas que son comunes en la mayoría de las respuestas obtenidas. El regulón del choque por alta temperatura (heat shock), fue el primer regulón identificado usando perfiles de expresión de proteínas (92). En este caso se observó el regulón al utilizar una cepa mutante comparando su perfil de proteínas con el perfil de proteínas de la cepa padre.

En un estudio reciente, miembros potenciales del regulón Pho en *E.coli* fueron identificados cuantitativamente por comparación de tres estímulos, la privación temprana de fosfatos, la privación tardía de fosfatos y condiciones de crecimiento con restricción de fósforo (130). La intercepción de esos

tres estímulos, el cual contiene cerca de 150 proteínas, incluye miembros del regulón Pho. En 8 de los genes que codifican para estas proteínas se identificó arriba, una secuencia promotor similar a la secuencia consenso que enlaza el regulador PhoB. La mayoría de estas 150 proteínas aún no han sido identificadas.

Firma Proteómica, es un nuevo término presente en una reciente publicación (128). Pero a pesar de que el término es nuevo, ya han sido previamente identificadas firmas proteómicas en bacterias durante los últimos veinte años. Muchos de los esfuerzos realizados y reportados en la base de datos EC02DBASE tienen como fin generar los datos para identificar firmas proteómicas relacionadas con la mayoría de estados fisiológicos de la célula. La meta fundamental es usar esta firma proteómica para diagnosticar el estado fisiológico de la célula cuando el blanco de un estímulo de una función o mutante es desconocido.

Ejemplos de fenotipos proteicos y firmas proteómicas

En los últimos 20 años se han llevado a cabo numerosos estudios buscando perfiles de expresión de proteínas en diferentes bacterias los cuales han definido los fenotipos de varias proteínas como son RpIL, MetE y PhoE.

### **La electroforesis en geles de doble dimensión y el proteoma microbiano**

Como ya mencionamos reiteradamente la meta básica de un estudio de proteoma es identificar en un organismo las proteínas implicadas en un proceso determinado. Generalmente, un organismo se cultiva bajo diversas condiciones experimentales y las proteínas presentes en cada muestra se analizan en 2-ED (Fig.6). Después de la comparación de los geles en programas de computación diseñados para ello, las proteínas de interés se seleccionan, proporcionando datos acerca de cambios en

la expresión de las proteínas (Fig.6). Usando esta metodología, se han identificado grupos funcionales de proteínas en *Bradyrhizobium japonicum*, cultivado bajo condiciones anaerobias y aeróbicas (26) (Fig.7) y sujetas a presiones ambientales como el estrés por calor (88). Además, Guerrero et al. (46) determinaron los cambios en los patrones o modelos de expresión de la proteínas de *Sinorhizobium meliloti* expresados durante la fase de crecimiento exponencial temprana y tardía, Vasseur et al. (131) investigaron las alteraciones en la expresión de las proteínas de *Pseudomonas fragi* cultivada bajo una variedad de condiciones, incluyendo el choque osmótico, choque por pH, y la combinación de tratamientos. Estos estudios han contribuido exitosamente en el conocimiento de nuevos descubrimientos en la biología de microorganismos, una revisión de las aplicaciones de esta metodología en varios organismos microbianos se ha publicado recientemente (128).

También recientemente, fue publicado el mapa de dos dimensiones del proteoma de *H. influenzae*. El cual contiene 502 proteínas identificadas (73), las cuales fueron obtenidas de fracciones solubles de la célula, proteínas de membrana (estas últimas fueron fracciones proteicas obtenidas con un sistema que utiliza dos detergentes con la finalidad de resolver las proteínas hidrofóbicas) y proteínas presentes en baja concentración, las cuales para su observación y posterior identificación fueron enriquecidas parcialmente mediante varios pasos cromatográficos, de esta manera en este sobresaliente trabajo se presentan 4 grupos de proteínas, no habiéndose reportado previamente algo similar en proteomas de microorganismos. Las proteínas más abundantes fueron factores de elongación y proteínas de membrana. Por secuencia amino terminal se determinó que en aproximadamente un 70% de las proteínas existió un péptido señal, el cual

ya no existía en la proteína caracterizada, por otro lado se estableció que aproximadamente el 15% de las proteínas identificadas eran proteínas truncadas. Esta base de datos debe ser considerada una base muy sólida para posteriores estudios de proteoma en *H.influenzae*.

La respuesta a la deprivación de fosfato en *Bacillus subtilis* fue analizada usando extractos celulares y sobrenadantes mediante electroforesis en geles de poliacrilamida de 2-dimensiones (8). Muchas de las proteínas inducidas durante la deprivación de fósforo se encuentran bajo el control de un factor sigma. Para poder definir cuales son las proteínas reguladas por el regulón Pho, el cual esta constituido por el sistema de dos componentes formado por proteínas reguladoras llamadas PhoP y PhoR, se analizaron patrones de proteínas de la cepa silvestre comparándolos con las mutantes en los genes *s/g6*, el cual codifica para el factor sigma B y la mutante en *phoR* el cual regula para uno de los miembros del sistema de dos componentes. Mediante MALDI-TOF fueron identificadas dos fosfatasa alcalinas (PhoA y PhoB), una fosfodiesterasa alcalina (PhoD), una glicerofosforil diester fosfodiesterasa (GlpQ) y una lipoproteína YdhF las cuales fueron inducidas fuertemente por las proteínas PhoPR. Por otro lado en la fracción citoplásmica, las proteínas PstB1, PstB2 y TuaD también fueron inducidas por este par de genes reguladores solo que de estas ya se había previamente publicado este tipo de regulación sobre ellas. Estudios transcripcionales de los genes *glpQ* y *ydhF* corroboraron lo ya observado mediante proteoma, la fuerte dependencia en la regulación de estos genes por PhoPR. También mediante este tipo de estudios en las mutantes *s/g6* y *phoR* se pudo observar algunas proteínas que se inducen durante la deprivación de fosfato, lo cual sugiere que estas son reguladas de manera independiente de PhoPR. A pesar de la valiosa información presentada por este grupo mediante un análisis tipo proteoma del estimulen de fosfato, no todas las proteínas

que lo componen pudieron ser analizadas ya que algunas de ellas se encuentran integradas en la membrana y para ello tendrán que realizar estudios tipo proteoma con detergentes que las solubilizen.

### Unión entre el proteoma y el genoma

Un requerimiento clave para el proteoma es la habilidad para ligar las proteínas individuales resueltas mediante 2-ED a sus correspondientes secuencias génicas. En paralelo con estudios de proteomas, especialmente en eucariontes, han sido descritos algunos que pueden ser usados para identificar estas proteínas. Las primeras aplicaciones de 2-ED examinaron la síntesis de proteínas y expresión del genoma de *Escherichia coli*, identificando las proteínas presentes en lisados celulares totales en base a su comigración con proteínas purificadas (90,105). Anticuerpos monoclonales y policlonales pueden ser usados para identificar proteínas de interés junto con técnicas inmunológicas. Sin embargo estas metodologías son limitadas por la disponibilidad de proteínas purificadas o anticuerpos específicos. Métodos más generalizados de identificación, están basados en la determinación de la composición de aminoácidos (140) o en una secuencia parcial de aminoácidos (76,125) o perfil de las masas de los péptidos que constituyen la proteína (35,77). Esos datos pueden ser usados para buscar en una apropiada base de datos de secuencias de proteínas o de ácidos nucleicos similitud u homología con un gene o proteína relacionado. En estos casos, los grupos de investigación que trabajan con microorganismos a los cuales se les ha determinado su secuencia genómica, sin duda alguna tienen una enorme ventaja.

En proteomas de organismos con genomas totalmente secuenciados el uso de métodos de identificación únicos es

suficiente, sin embargo en organismos poco caracterizados genéticamente, es necesaria la combinación de dos o más técnicas. Por ejemplo en el caso del proteoma de *Spiroplasma melliferun*, un miembro de la clase Mollicutes, usando 2-ED teñidos con plata, fueron detectadas 456 proteínas de las cuales 156 fueron seleccionadas para una posterior caracterización usando una combinación de un análisis tipo composición de aminoácidos, análisis del patrón de péptidos con MALDI-TOF y secuencia amino terminal por degradación de Edman. Usando una combinación de esos métodos, 98 proteínas fueron identificadas por homología con secuencias génicas ya existentes en un grupo diverso de bacterias, treinta de estas proteínas no están presentes en *Mycoplasma genitalium* y *M. pneumoniae* organismos muy relacionados filogenéticamente a los cuales su genoma ya ha sido secuenciado (40,54). Un análisis similar empleando una combinación de metodologías como composición de aminoácidos y espectrometría de masas (MALDI) fue usado para caracterizar el proteoma de *Ochrobacterum anthropi* (136). Sin embargo solo 45 de las proteínas resueltas fueron identificadas. Mediante una estrategia similar nuestro grupo en *Rhizobium etii* se encuentra en la tarea de identificar proteínas presentes en el metabolismo celular en vida libre y en simbiosis con *Phaseolus vulgaris*, la estrategia en este momento se encuentra en la combinación entre espectrometría de masas (MALDI-TOF) y secuencia amino terminal mediante degradación de Edman, lo cual nos ha permitido identificar cerca de 90 proteínas en total.

Desde la primera secuencia genómica completa de un organismo vivo (*Haemphilus influenzae*) publicada en 1995 (33), ha habido un rápido incremento en el número de genomas bacterianos secuenciados. Hasta el momento el "Megpie Genome Project List" (<http://www-fp.mcs.anl.gov/gaasterland/genomes.html>) presenta 25 secuencias de genomas

completamente secuenciados en procariontes además de 48 en proceso de terminación. A esto se adiciona a un número extenso de secuencias de genes individuales que constantemente están siendo publicados en las bases de datos. Estas secuencias genómicas son la mejor fuente para la identificación de proteínas en proyectos de proteomas bacterianos. No obstante lo anteriormente comentado aunado a que la secuencia genómica completa de al menos nueve patógenos humanos y una bacteria de interés agrícola (*Sinorhizobium meliloti*) serán pronto disponibles, existen grupos de investigación interesados en trabajar con microorganismos en los cuales una limitada base de datos de secuencias genómicas han sido publicadas. En este caso algunos grupos utilizan la búsqueda cruzada con secuencias de genomas totalmente secuenciados como en el caso de *Listeria monocytogenes* los cuales analizaron fundamentalmente sus proteínas en comparación con la secuencia genómica de *Bacillus subtilis* logrando identificar de un extracto celular bacteriano, mediante espectrometría de masas (MALDI) 13 de las más abundantes proteínas inducidas durante su cultivo en un pH (W7). En resumen los estudios en proteomas bacterianos solo podrán continuar si a la par se continúa la caracterización genómica completa de microorganismos.

### **Análisis de variabilidad genética**

Las técnicas de biología molecular han sido extensivamente utilizadas para investigar aspectos epidemiológicos y taxonómicos de microorganismos. Las proteínas también pueden ser utilizadas como marcadores moleculares basando el análisis en diferencias en sus movilidades electroforéticas. La movilidad electroforética de proteínas específicas (ejemplo, enzimas metabólicas (60,H7) y proteínas de membrana externa (78) o

proteínas de extractos celulares completos) son analizadas en geles de una dimensión (1-ED). Pero el análisis de proteínas bacterianas mediante 2-ED ha sido recientemente documentado. Estudios comparativos de extractos celulares analizados en 2-ED de *Neisseha sp* (60, 61, 66), *Treponema pallidum* (4, 114), *Campylobacter sp.*, (30), *Haemophilus sp.*, (1\_7) *Mycoplasma sp.*, (100. 135, 138) y *Plasmodium falciparum* (10, H) han sido utilizados para diferenciar aislados de interés clínico. De hecho, las cepas bacterianas pueden algunas veces mostrar un mayor grado de variabilidad por 2-ED que la observada mediante hibridización ADN (112). Jackson et al., (60) compararon mediante 2-ED proteínas de múltiples aislados de *Neissería gonorrhoea*, *N. meningitidis* y *Brahhamella catarrhalis*. Comparaciones cuantitativas y cualitativas de aproximadamente 200 proteínas resueltas por 2-ED mostraron diferencias en los perfiles de 2-ED colaborando en el establecimiento de su ya reconocida clasificación taxonómica (60). De manera similar, seis cepas de *Mycoplasma*, que representan 4 diferentes especies, mostraron grados equivalentes de similitud por 2-ED al por hibridización de su ADN (5). Un estudio extensivo realizado en *Listona sp.*, clasifico 29 aislados que conformaron 6 diferentes especies, este análisis fue realizado en base a las proteínas celulares resueltas por 2-ED (45).

### **Estudios de patogenicidad bacteriana**

Una meta importante en la microbiología médica es lograr el conocimiento de como los organismos patógenos interactúan con sus huéspedes para producir los procesos clínicos. Durante un largo período de tiempo varios grupos de investigación han trabajado en la identificación de nuevos determinantes de virulencia que puedan servir como futuros objetivos para la síntesis de vacunas o para el desarrollo de fármacos. Algunas técnicas de biología molecular permiten ahora a los investigadores analizar este tipo de eventos de manera detallada y como discutiremos

más adelante, los proteomas juegan un papel importante en estos desarrollos.

### **Expresión de determinantes de virulencia en microorganismos**

La habilidad de los microorganismos patógenos para causar enfermedades en hospederos susceptibles es determinada por múltiples factores que actúan individualmente o en grupo durante diferentes grados de la infección. Por ejemplo, los dos más importantes factores de virulencia, en *Helicobacter pylori* *vacA* y *cagA*, han sido identificados y mostrado que se encuentran involucrados en distintos aspectos de la patogénesis de esta bacteria (9). El *cagA* es actualmente un marcador de una isla de patogenidad que contiene 29 genes (124). El proteoma tiene el potencial de proporcionar una importante información que colabore en el control de la expresión de esos genes en aislados de *H.pylori* *cagA+*. La capacidad del proteoma para analizar la síntesis global de proteínas y por extensión la expresión genética, convierte a esta técnica en una poderosa herramienta de estudio para comparar proteínas sintetizadas por cepas bacterianas virulentas y avirulentas. La electroforesis en geles de doble dimensión puede ser catalogada como un popular método para la identificación de determinantes de virulencia al nivel de síntesis de proteínas. Estudios tempranos comparando cepas virulentas y no virulentas de *Mycoplasma pneumoniae* identificaron tres nuevas proteínas expresadas solo por cepas virulentas y que se no se encuentran en las cepas no virulentas (49). Un trabajo similar fue realizado en *B. abortus* para comparar una cepa virulenta con una cepa avirulenta que había sido utilizada como una vacuna así como unos aislados deficientes en la síntesis de lipopolisacáridos derivados de ambas cepas.

En promedio 935 proteínas fueron resueltas por 2-ED en cada una de las cuatro cepas, un número menor a las 2129 proteínas potenciales predichas para el genoma de esta bacteria que tiene un tamaño de 3.3 X10<sup>6</sup> pares de bases. Las cepas virulentas y avirulentas mostraron una homología a nivel de AND del 98.4 y 99.3%. La cantidad de ADN equivalente a la diferencia tiene un potencial de capacidad codificadora de 34 proteínas. La comparación en 2-D de los perfiles de proteínas de la cepa virulenta y la avirulenta mostró un total de 86 diferencias cualitativas y 6 diferencias cuantitativas, lo cual se acerca bastante a lo esperado si consideramos las posibles entidades electroforéticas productos de las modificaciones que pueden sufrir las proteínas, estos resultados les permitirá a los autores, previa identificación de estas proteínas, caracterizar las moléculas responsables de la virulencia en esta bacteria.

### **Selección de cepas virulentas y avirulentas**

Una dificultad en la interpretación de datos obtenidos en estudios comparativos como los mencionados en párrafos anteriores es la selección de cepas virulentas y avirulentas usadas como la base de la comparación. En muchos casos las cepas virulentas y las avirulentas son genéticamente distintas y las diferencias en el proteoma con frecuencias no son ligadas a la virulencia. Maharias et al (79) describieron una estrategia alternativa para minimizar este problema. Una hibridación genómica sustractiva fue usada para localizar diferencias genéticas entre la cepa vacuna BGC de *M. Bovis* y aislados virulentos de *M.bovis* y *M. tuberculosis*. Tres regiones fueron deletadas en la vacuna BCG comparada con las cepas virulentas. Un fragmento de 9.5 kb (designado RD1) ausente en seis BCG subcepas fue conservado en la cepas virulentas de *M. tuberculosis* y *M. Boris* así como en 62 de 62 aislados clínicos. Análisis de secuencia del

fragmento RD1 revelaron la presencia de al menos 8 genes. Análisis por 2-ED de las proteínas expresadas por las cepas virulentas de *M. bovis* y *M. tuberculosis* mostraron un indistinguible perfil de proteínas. En contraste, la cepa BCG expresó al menos 10 proteínas adicionales y niveles inducidos en la expresión de otras. La región RD1 fue introducida en el genoma de la cepa BCG para generar la cepa BCG::RD1 (Fig. 10).

Cuando se analizaron por 2-ED, el perfil de proteínas de la cepa BCG::RD1 fue indistinguible de la cepa virulenta de *M. bovis*, sugiriendo que la porción de ADN catalogada como RD1 causa una supresión específica en la síntesis de proteínas en la cepa virulenta de Mycobacteria. Algunas proteínas de bajo peso molecular fueron identificadas en la cepa BCG::RD1 lo cual es consistente con el tamaño de los genes pequeños codificados en la secuencia del fragmento RD1.

### **Patogénesis de *Haemophilus influenzae***

Para examinar la patogénesis de *Haemophilus influenzae* ha sido utilizada una estrategia genética. *H. influenzae* tipo B es un patógeno humano común que coloniza la parte superior del tracto respiratorio pero puede invadir el torrente circulatorio, las bacterias son expuestas a una decreciente concentración de oxígeno y se ha propuesto que la enzima superóxido dismutasa (SOD) juega un importante papel en la supervivencia durante la transición (28). *H. influenzae* tipo B posee una única superóxido dismutasa (SOD) funcional que es codificada por el gene *sodA*. Una mutación en el gene *sodA* muestra una reducida capacidad colonizadora del tracto respiratorio comparado con la cepa padre (28) (Fig. 11). Sin embargo, esta mutante aún causa bacteriemia indicando que la bacteriemia

podría proceder aún en ausencia de colonización. Esto sugirió que la SOD no es por mucho un factor de virulencia pero contribuye enormemente a la supervivencia bacteriana durante la colonización. Las proteínas sintetizadas por la cepa padre y la mutante fueron comparadas por medio de 2-ED. Adicionado a la ya esperada pérdida del producto del gene *sodA* dos proteínas mostraron cuantitativas diferencias en su expresión

(28). Ahora se sabe por análisis de MALDI, que esas dos manchas proteicas son codificadas por el ORF H10572. A pesar de la información obtenida hasta ahora es aún difícil saber como puede intervenir el producto de este ORF en la patogenicidad de *H. influenzae*, este párrafo solo pretende mostrar como un problema de análisis o estudio de determinantes antigénicos puede ser analizado de esta manera, dejando al descubierto la

interligada red de control genético presente durante la patogenicidad.

### **Expresión genética in vivo**

Durante la infección bacteriana in vivo, los patógenos modifican su expresión genética en un contexto espacial y temporal, esto en respuesta a la exposición a diferentes estímulos ambientales. Varios métodos han sido desarrollados para examinar la expresión genética de la bacteria durante la infección in vivo. Muchas de esas estrategias experimentales usan la tecnología de ADN recombinante para identificar y caracterizar promotores que son específicamente activados durante el crecimiento bacteriano con el huésped (50, 51, 80, 127, 134). El proteoma puede jugar un papel complementario a esas técnicas en la definición de expresión genética in vivo. Deiwick y Hansel (27) desarrollaron una estrategia experimental combinando las tecnologías de ADN recombinante y proteoma para identificar y caracterizar genes asociados a la virulencia en *S. typhimurium*, dentro de un modelo de salmonelosis en

ratones. Identificando un grupo de genes que afectan la virulencia, los cuales fueron posteriormente localizados en el grupo de genes de patogenicidad conocido como isla 2 (SP12). Esos datos fueron analizados mediante el proteoma estudiando la síntesis global de proteínas en respuesta a señales ambientales así como a mutaciones en genes reguladores presentes en la isla SP12. Este tipo de experimentos ilustra el potencial de la integración de proteoma con genoma para investigar la compleja y multifactorial regulación de los determinantes de la patogénesis microbiana.

El cultivar a las bacterias en presencia de células eucarióticas puede ser suficiente para inducir la síntesis de proteínas no expresadas por la bacteria cuando es cultivada sola en un medio definido. Durante este tipo de cocultivos de *Campylobacter jejuni* y la línea celular de epitelio INT407, la bacteria sintetizó al menos 14 nuevas proteínas nuevas. Estos cambios en la síntesis de proteínas fueron analizados mediante 2-ED combinando el mareaje radiactivo e inmunotransferencia (69, TV). La inducción de estas 14 proteínas también se observa en las bacterias previo cultivo con medio condicionado, es decir medio en el cual las células epiteliales fueron previamente cultivadas (68). Esto sugiere que la síntesis de novo de esas proteínas es requerida como parte de un programa de internalización de la bacteria en las células epiteliales.

Bacterias intracelulares cultivadas en fagosomas celulares al ser expuestas a una variedad de condiciones adversas, como acidez extrema, oxígeno y nutrientes en bajas concentraciones inducen una serie de respuestas que han sido monitoriadas mediante proteomas (71). Durante la infección de macrófagos, *Legionella pneumophila*, *Brucella abortus* y *S. typhimurium* permanecen asociadas con fagosomas y pueden

interferir con su maduración. Una observación consistente en esas bacterias patógenas intracelulares es que la síntesis de ciertas proteínas es inducida o reprimida durante el crecimiento intracelular comparadas con el crecimiento bacteriano en medios de cultivo artificiales. Además, un número de proteínas bacterianas inducidas durante su crecimiento intracelular son también alteradas en su biosíntesis bajo condiciones de estrés in vitro. La síntesis de proteínas bacterianas conocidas como proteínas de "heat shock" como son GroEL y DnaK son inducidas durante la infección de *B. abortus* a macrófagos de bovino (109). Las mismas dos proteínas son parte del espectro de proteínas inducidas por *S. typhimurium* durante la infección a macrófagos (16). Treinta y una de las (71) proteínas bacterianas inducidas durante la infección de *L. pneumophila* a macrófagos son también inducidas en condición de estrés in vitro. Estas incluyen a GroEL y GroES (1). Una proteína catalogada como proteína global de estrés llamada GspA, es expresada por *L. pneumophila* en respuesta a todas las condiciones de estrés probadas así como durante la infección, sugiriendo que esta bacteria en este medio ambiente particular puede estar expuesta a múltiples y simultáneas condiciones de estrés (2, 71).

### **La respuesta del hospedero analizada con proteomas**

Una gran cantidad de progresos han sido logrados con el uso de proteomas para investigar la respuesta inmune del huésped a la infección bacteriana, 2-ED combinados con inmunotransferencia son comúnmente usados como una herramienta para investigar la respuesta inmune celular y humoral contra patógenos microbianos (87, 143). La combinación de estas técnicas con anticuerpos policlonales han sido utilizadas en un gran rango de bacterias, por ejemplo en *Borrelia burgdorferi*, *Streptococcus pyogenes* (75) y *Brucella ovis* (123). La identificación de las principales proteínas

inmunogénicas es el camino lógico para definir el rango de anticuerpos específicos presentes en el suero humano durante los procesos infecciosos. Sánchez-Campillo et al. (113) utilizaron esta metodología en el proceso inflamatorio genital ocasionado por *Chlamydia trachomatis* detectando mediante anticuerpos una gran cantidad de proteínas inmunogénicas entre las cuales se encontraron proteínas de membrana además de GroEl y DnaK. Adicionalmente 4 nuevas proteínas que reaccionan con el suero fueron identificadas. La frecuencia en la cual esos antígenos fueron reconocidos por el suero, varió entre pacientes, sin embargo todos reconocieron una proteína de membrana conocida como OMP2. Estos resultados pueden colaborar posteriormente en la síntesis de vacunas contra patógenos bacterianos.

La respuesta inmunológica celular mediada por células T es el componente más importante en la infección producida por *M. tuberculosis*. La identificación de las proteínas bacterianas que inducen la respuesta de células T-dependiente será un resultado importante en el futuro, el cual colaborará en el desarrollo de nuevas y más eficientes vacunas. Algunos grupos de investigación han demostrado que durante el crecimiento en medios de cultivo específicos *M. tuberculosis* produce proteínas que pueden inducir la respuesta protectora mediada por células T (6, 72, 101, 111). Sonnenberg y Belisic (121) detectaron mediante 2-ED 205 proteínas en un proteoma de *M. tuberculosis*, 34 de ellas fueron identificadas con base en su reacción con anticuerpos monoclonales específicos y con el auxilio de la secuencia de aminoácidos. Un miembro de este grupo proteico previamente reconocido como un antígeno humoral importante (72), fue identificado como KatG la cual es una importante proteína en la defensa de la célula al estrés oxidativo.

### **Patogénesis y resistencia a antibióticos**

La aparición de la resistencia a antibióticos en un número alarmante de patógenos humanos ha hecho necesaria su investigación. El estudio mediante proteomas ha facilitado el conocimiento de los cambios fisiológicos en un organismo que conducen a la patogenicidad, a la resistencia a antibióticos y/o a una inmuno-respuesta. Mediante proteomas se han realizado recientemente estudios sobre la patogenicidad de *Candida albicans* (94) e investigaciones adicionales han identificado en sueros humanos varias proteínas inmunogénicas (108). Otros grupos han identificado las proteínas inmunoreactivas de *Chlamydia trachomatis* (113) y alteraciones en los patrones de proteínas expresadas por *Streptococcus pneumoniae* resultado de la resistencia a eritromicina (18) (Fig.12). En análisis comparativos, los proteomas de diversas cepas de microorganismos se resuelven y se comparan. Cuando se aplicaron este tipo de comparaciones a cepas de *Mycobacterium tuberculosis* y *Mycobacterium bovis* BCG, fueron identificadas proteínas que difieren en intensidad y posición, proporcionando información importante acerca de las diferencias en patogenicidad de ambas cepas de *Mycobacterium* (65). La meta de estos estudios con proteomas es identificar las proteínas implicadas en los procesos infecciosos, colaborando esperanzadoramente en el desarrollo de tratamientos médicos.

*Helicobacter pylori* es un importante patógeno gástrico del hombre ya que infecta a la mitad de la población mundial, este agente causa desordenes gastrointestinales, como gastritis crónica, úlcera péptica y cáncer estomacal. Después de la secuencia completa del genoma de dos cepas, 26695 y J99, un grupo de investigación se dio a la tarea de analizar la parte funcional de la información genética, el proteoma (64). Las proteínas expresadas por cepas de *H. pylori* 26695, J99 y SS1 fueron resueltas en geles de doble dimensión de alta resolución. Más de 1800 especies proteicas fueron obtenidas de extractos celulares. Mediante MALDI se

identificaron 152 proteínas, incluyendo 9 de ellas que fueron caracterizadas como factores de virulencia y 28 antígenos. Las cepas investigadas presentaron pocas proteínas con la misma movilidad electroforética, esto se debe a que modificaciones en su secuencia de DNA repercuten en un intercambio de aminoácidos en la secuencia de las proteínas y por consecuencia en la movilidad electroforética en el gel. Un estudio preliminar realizado con anticuerpos producidos por pacientes humanos, contra una membrana a la cual se habían electrotransferido las proteínas presentes en un gel, mostró que este tipo de estrategias puede ser ideal para identificar proteínas con valor diagnóstico o terapéutico. El patrón de geles de doble dimensión con las proteínas observadas así como la base de datos obtenida se pueden consultar en la siguiente dirección de internet (<http://www.mpiib-berlin.mpg.de/2D-ED/>). El conocimiento de las bases de datos generadas de estudios tipo proteoma son un instrumento efectivo para la identificación de factores patógenos, antígenos de potencial valor diagnóstico o curativo, lo cual se puede observar en el siguiente párrafo.

*H. pylori*, muestra un único y complejo patrón de componentes moleculares que se unen al complejo de carbohidratos presentes en la mucosa del huésped y otros tejidos.

Estas moléculas de adhesión son conocidas como adhesinas. Estas adhesinas y otras proteínas asociadas a membranas son importantes objetivos en el desarrollo de vacunas.

La identificación y caracterización de proteínas expresadas por la superficie celular de esta bacteria, es un pre-requisito para el desarrollo de vacunas diseñadas para interferir en la colonización bacteriana de tejidos del huésped. Este grupo de investigación desarrolló una interesante metodología para resolver mediante 2-ED el complejo proteico expresado por la

bacteria (95). Este grupo logró identificar 40 proteínas de una muestra de *H. pylori*, de este grupo un tercio son proteínas asociadas a membrana, entre ellas se encuentra una proteína membranal de baja abundancia (Leb) la cual ha sido identificada como una adhesina. El uso de la metodología del proteoma es catalogado como un método de identificación de proteínas expresadas durante la virulencia, así como proteínas blanco de vacunas en este y otros microorganismos patógenos.

*Mycoplasma genitalium* es el más pequeño miembro de la clase Mollicutes, este microorganismo tiene un genoma de solo 580 kilobases, el cual es potencialmente capaz de contener 480 genes, por lo anterior ha sido tomado como un excelente modelo para: i) conocer la unidad mínima metabólica necesaria por una célula viva ii) conocer mediante estudios tipo proteoma su patrón de proteínas. Por lo anterior Wasinger y colaboradores

(136) realizaron un trabajo de investigación que presenta 73% de las proteínas potencialmente factibles a ser expresadas por esta bacteria, de las cuales lograron analizar el 33%. En resumen, mediante el uso de 2-ED y una combinación de 4 diferentes tipos de geles de doble dimensión, con diversas ventanas de pI, lograron resolver 427 proteínas presentes en el crecimiento exponencial de la bacteria, de este grupo 201 se presentan en abundancia suficiente para ser observadas mediante tinción con plata y ser identificadas mediante espectrometría de masas, lo cual produjo un total de 158 proteínas identificadas. Observándose una disminución del 42% en la expresión de proteínas durante la fase estacionaria del crecimiento, siendo las proteínas más abundantes DnaK, GroEL, ADN-girasa mientras que las proteínas ribosomales disminuyeron en su abundancia en esta fase del cultivo. De esta manera se concluye que la tecnología del proteoma puede proveer una perspectiva en la bioquímica y actividad metabólica de este organismo.

## Predicciones de la funcionalidad del genoma mediante proteomas

El DNA cromosomal contempla dos principales funciones en la célula: sirve como templado para replicación y templado para transcripción. Existen una gran cantidad de actividades (tales como reparación del DNA y segregación del cromosoma) que muestran que el cromosoma es propiamente mantenido y cuidado para cumplir con sus funciones básicas (133). El fenotipo proteico de una proteína celular, llamada RecA, incluye un estado de inducción en todas las condiciones que son conocidas por comprometer la estructura o función del cromosoma (129). En la figura 13 mostramos la inducción de la proteína RecA en cultivos de *Escherichia coli* tratados con seis diferentes químicos.

Mientras esta proteína puede ser usada como un factor inicial de pérdida de funciones cromosómicas, el examen minucioso de esas seis condiciones, revela firmas celulares específicas. Veintidós proteínas fueron encontradas (por examen visual) que pueden servir como indicadores para separar esos seis agentes en cuatro grupos. Un análisis más comprensible de esas imágenes se encuentra en este momento en desarrollo (VanBogelen et al, comunicación personal) y podrán definir las firmas celulares específicas producidas por la presencia de cada uno de los agentes. Como algunos de los blancos de los agentes químicos son exclusivos de organismos microbianos esas firmas celulares pueden ayudar a seleccionar compuestos que puedan ser de utilidad con fines terapéuticos.

A pesar del valor y de la enorme información que la electroforesis en geles de doble dimensión ha generado en el campo del proteoma, existen algunas deficiencias en esta tecnología. A pesar de que hasta 1400 proteínas pueden resolverse en un gel, en cada estudio,

hasta el momento son muchos más las proteínas que se resuelven que las que se han logrado realmente identificar. Por ejemplo, en un estudio que correlaciona la abundancia de RNAm y la expresión de proteínas en el *S. cerevisiae*, Futcher et al. (42) resolvieron 1400 proteínas, pero solamente lograron identificar 148 proteínas. En un genoma de aproximadamente 6000 proteínas, sus resultados solo proporcionaron información del 2.5% del potencial proteoma de la levadura. El análisis de las 1400 proteínas es altamente consumidor en tiempo, porque cada proteína se debe extraer, digerir y analizar individualmente. Como ya mencionamos una porción significativa del proteoma, especialmente de proteínas de baja abundancia y asociadas a membrana, son raramente observadas en este tipo de estudio. Actualmente algunos grupos de investigación realizan intentos por resolver estas deficiencias, en el caso de las proteínas de baja abundancia, se trabaja mediante estudios cromatográficos previos al tratamiento de la muestra en geles de doble dimensión, esto es con la finalidad de enriquecer las proteínas de baja abundancia que no han sido observadas en geles convencionales.

Fountoulakis et al. (36) mediante cromatografía en columna de un lisado celular de *E.coli* identificaron 269 proteínas de 800 obtenidas del eluido de la columna de cromatografía.

En éstas proteínas, se incluyen varias proteínas de baja abundancia que no habían sido detectadas mediante electroforesis en geles de doble dimensión (36).

En estudios adicionales, Fountoulakis et al. (37) usando cromatografía de interacciones hidrofóbicas enriquecieron una muestra de proteínas de baja abundancia obtenida de un extracto celular de *Haemophilus influenzae*. En este caso, solamente las proteínas que se unen a la matriz de la columna fueron enriquecidas, y esto incluyó a proteínas de alta y baja abundancia. Aunque el enriquecimiento

mencionado es prometedor, es necesaria una metodología más general que identifique una mayor cantidad de proteínas de baja abundancia.

Otro de los problemas que más nos preocupan a los que trabajamos con proteomas, es la baja solubilidad de las proteínas unidas a la membrana celular las cuales al no solubilizarse completamente previenen su resolución en geles de doble dimensión, los recientes avances en solubilización de proteínas de membranas han sido publicados (53). Estos métodos son prometedores e incluyen la extracción de proteínas con solventes orgánicos seguido por la solubilización con detergentes, lo anterior previo al corrimiento electroforético en geles de doble dimensión (86) por otro lado es importante mencionar que actualmente se están sintetizando nuevos detergentes con una mayor eficiencia en la solubilización de proteínas unidas a membrana (20). A pesar de este tipo de esfuerzos hasta la fecha muy pocas proteínas de membrana han sido identificadas, siendo esta una área que necesita un fuerte desarrollo adicional.

### **Nuevos métodos de separación de proteínas**

Las dificultades para detectar e identificar proteínas de baja abundancia y asociadas a membrana han hecho necesario el desarrollo de nuevas tecnologías dirigidas a la resolución de este tipo de proteínas. Estas metodologías están demostrando ser más reproducibles y similares en rapidez de procesamiento que la electroforesis en geles de doble dimensión. El isoelectroenfoque capilar (81) acoplado a la espectrometría de masas (MS) han sido poco utilizados en proteomas, para la separación de mezclas complejas de proteínas (118), sin embargo han demostrado ser potencialmente útiles para una rápida identificación de hasta 1000 proteínas (63). Cuando se han

aplicado estas metodologías a pequeñas mezclas de proteínas, como extractos de ribosomas, se detectaron e identificaron entre el 80 y 90% de las proteínas presentes en el ribosoma de *S. cerevisiae* (126).

### **Proteomas cuantitativos**

Dos nuevos métodos dirigidos a la investigación de proteomas cuantitativos han sido desarrollados en el último año (82). Pása-Tolic et al. (104) y Oda et al (98) han desarrollado métodos para marcar metabólicamente proteínas de células cultivadas en presencia de  $^{15}\text{N}$  o  $^{13}\text{C}$  como la fuente única del nitrógeno. El producto ofrece muestras de proteínas idénticas solo que unas pueden ser catalogadas como proteínas "pesadas" o

"ligeras" dependiendo del tipo de isótopo incorporado, esto permite la cuantificación relativa de una proteína entre dos muestras al ser estas analizadas por MS. Oda et al.

(98) cultivaron dos cepas de levadura, una mutada en el gene que produce el producto catalogado como Cln2 y la cepa silvestre, ambas fueron cultivadas bajo idénticas condiciones (excepto que la única fuente del nitrógeno para una de ellas es  $^{15}\text{N}$ ). Después de reunir los dos extractos celulares estos fueron fraccionados por cromatografía líquida de alta presión y resueltos por electroforesis de geles de doble dimensión-SDS. Las proteínas obtenidas del gel fueron extraídas y digeridas e identificadas por MS. Las diferencias en masas de proteínas idénticas permitieron a Oda et al. determinar la abundancia relativa de 12 proteínas en cada muestra (98). Aunque esta metodología es de gran alcance, la limitación que presenta es que el mareaje metabólico está limitado a grupos de bacterias que puedan ser cultivadas bajo tales condiciones.

Gygi et al. (47) han desarrollado un nuevo método para la cuantificación de proteínas presentes en proteomas, en este método el residuo del cisteína de las proteínas fue isotópicamente marcado después del

crecimiento de la célula (Fig.2). Después de cultivar la levadura en etanol o galactosa (como fuente de carbono), las proteínas de cada condición de crecimiento fueron extraídas y la modificación en los residuos de la cisterna fueron analizados y clasificados como "pesados" o "ligeros"(Fig.2b). Gygi et al. (47) determinaron entonces los cambios en proteínas específicas, resultando del crecimiento en diversas fuentes de carbono. Estrategias de mapeo metabólico como las anteriores se pueden aplicar a cualquier sistema ya que las proteínas son modificadas posteriormente a su síntesis durante el crecimiento de celular.

## **Conclusiones**

Las ventajas de metodologías de proteomas en organismos procariontes y eucariontes han sido mencionadas de manera detallada en la literatura (55, 58, 59, 141). Las ventajas incluyen la capacidad de esta técnica para hacer análisis global del genoma analizando los cambios en su perfil de proteínas (7, 128) así como las ventajas que ofrece el análisis de modificaciones postranscripcionales de las proteínas, que no pueden ser observadas con el conocimiento de la simple secuencia nucleotídica (139). La identificación post-electroforética de las proteínas ha sido parcialmente resuelta con el desarrollo de técnicas de mapeo peptídico mediante MALDI-TOF y claro, con el desarrollo y culminación de algunos proyectos genómicos.

El Proteoma de esta manera ha colaborado la investigación en la nueva biología de microorganismos. La combinación del proteoma con otras herramientas disponibles permite al biólogo moderno un acercamiento al conocimiento integral de la célula. Sin embargo, todavía existe la necesidad de un desarrollo tecnológico adicional además es

importante considerar que desde mi juicio y la tecnología actual está subutilizada. Por ejemplo, los actuales métodos de investigación que permiten conocer las modificaciones post-traduccionales de las proteínas apenas se encuentran en desarrollo (139).

Actualmente se han realizado dos análisis con proteomas dirigidos a estudios sobre fosforilación de proteínas de fibroblastos de ratón (43, 122), pero esta es una área poco explorada, especialmente en el proteoma microbiano. Debido a las deficiencias en la electroforesis en geles de doble dimensión, avances tecnológicos en electroforesis capilar aunados al MS, se están perfeccionando para aumentar la capacidad de procesamiento de un análisis tipo proteoma. Finalmente, con el surgimiento del proteoma cuantitativo, serán posibles las investigaciones que analicen las alteraciones celulares a nivel de proteínas.

En el futuro, cuando el investigador que trabaje con proteomas pueda determinar semicuantitativamente las 2000 proteínas que sean resueltas de una muestra celular, el proteoma llegará a ser tan importante para la biología como el análisis cuantitativo de la expresión de los genes.