Ingeniería Genética II

Unidad X

Análisis de Proteínas

La complejidad del organismo no es explicable por el número de sus genes

Caption: Comparison of the number of genes in different organisms. Illustration: Dominic Li
Image for use with the Wellcome Trust genome announcement only, no archiving, no sales, no web use. For further clearance contact j.whiting@wellcome.ac.uk or Tel. +44 (0)20 7611 8347.
©Wellcome Trust 2001

“Cellular reality is more elaborate than the dreams of even the nucleus itself”

Anderson et al,
Electrophoresis 19, 1853 (1998)

Nada es lo que parece…

- Paradoja de los pocos genes
- Humano: 98% homología con chimpancé
- Empalme alternativo en ≈ 40% de los genes
- Más de 350 modificaciones químicas (proteina tau: 96 variantes)
- Abundancia de proteína no correlaciona con abundancia de mARN

Las nuevas preguntas…

- **Qué proteínas se expresan , en qué cantidad y en qué condiciones**

- **Dónde se ubican**

- **Cuál es su tiempo de vida**

- **Cuál es su función**

- **Cuál es su relación con otras (redes de interacción, ¨interactomas¨)**

↓

Biología de sistemas

Los métodos:

Bioquímica tradicional:

* Acumular material de partida (gramos)

•Múltiples etapas de purificación

* La identificación requiere de la proteína pura

Proteómica:

* Material de partida mínimo (microgramos)

* Resolución simultánea de miles de especies en pocas etapas o identificación sin separación

* Identificación de miles de especies en un experimento

Avalancha de datos → BIOINFORMATICA

La complejidad analítica en las muestras de proteínas

La complejidad analítica de las proteínas se debe a:

- **Número de especies**
- **Distribución subcelular de variantes**
- **Diversidad de:**

  Tamaño y forma

  Carga eléctrica

  Hidrofobicidad

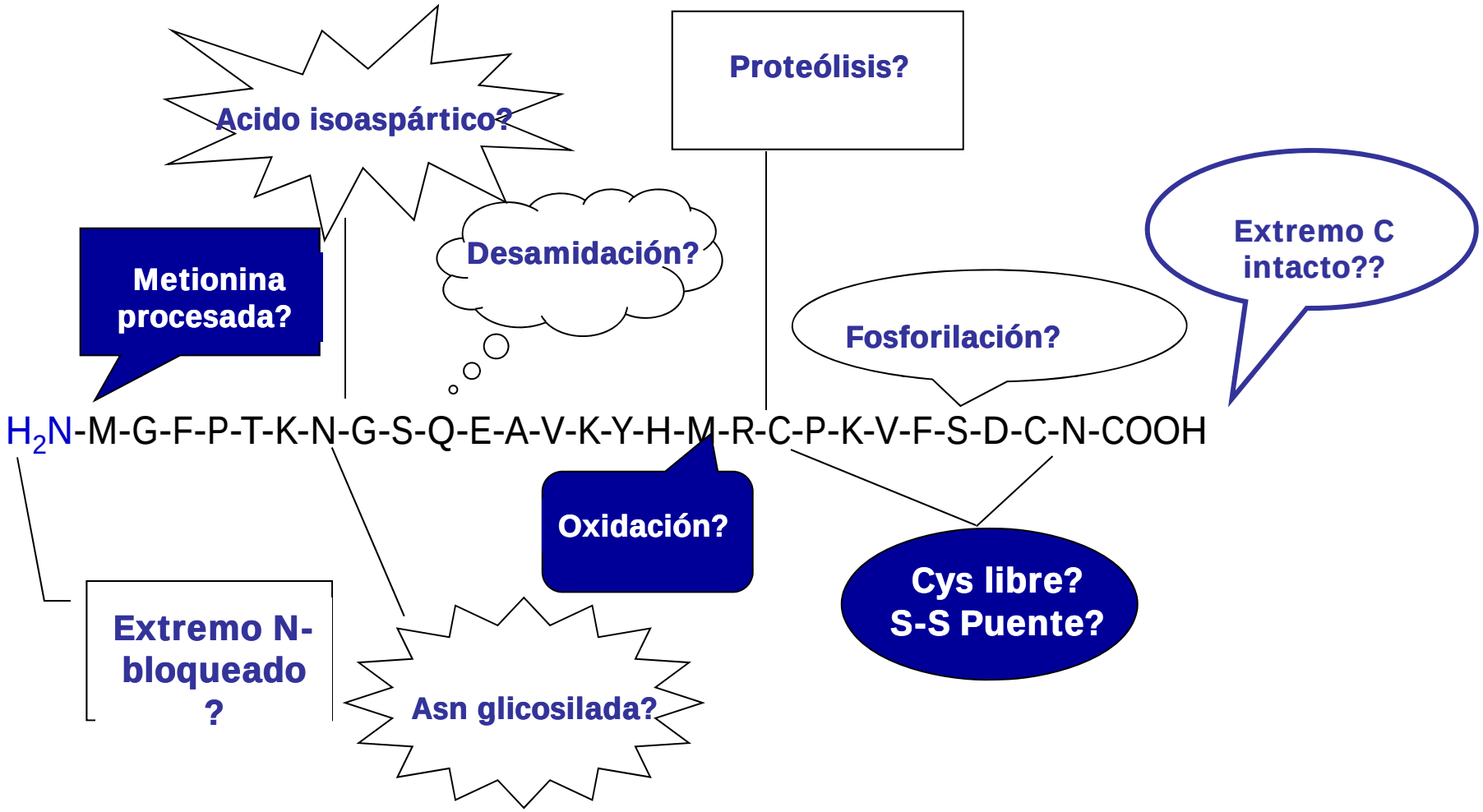
  Abundancia

La complejidad del proteoma:

Empalme alternativo: un gen produce varios transcriptos y varias proteínas (secuencias diferentes.

Modificaciones post traducción: Una proteína puede presentar muchas variantes, cada una con modificaciones químicas distintas.

Más de 400 modificaciones diferentes (www.DeltaMass.org)



Si: se expresa 1/3 del genoma humano, y hay promedio de 10 modificaciones por proteina, → 100,000 variantes / célula

La complejidad de un organismo está relacionada con la complejidad de su regulacion genetica y su proteoma

La expresión de proteínas es dinámica:

cantidad expresada,

tiempo de vida de la molecula,

distribución en la célula,

tipo y número de de modificaciones...

Número de especies:

Genoma humano: aprox. 23,000 genes

Si 1 / 3 del genoma se expresa (8,000 genes) y hay 10 modificaciones / proteína, → 80,000 especies / célula.

Ej.: Hsp27: identificadas 34 especies

Proteína tau: 96 especies

PERO: en tejidos 5% de las especies constituyen 80% de la masa de proteínas: problema del rango de expresión

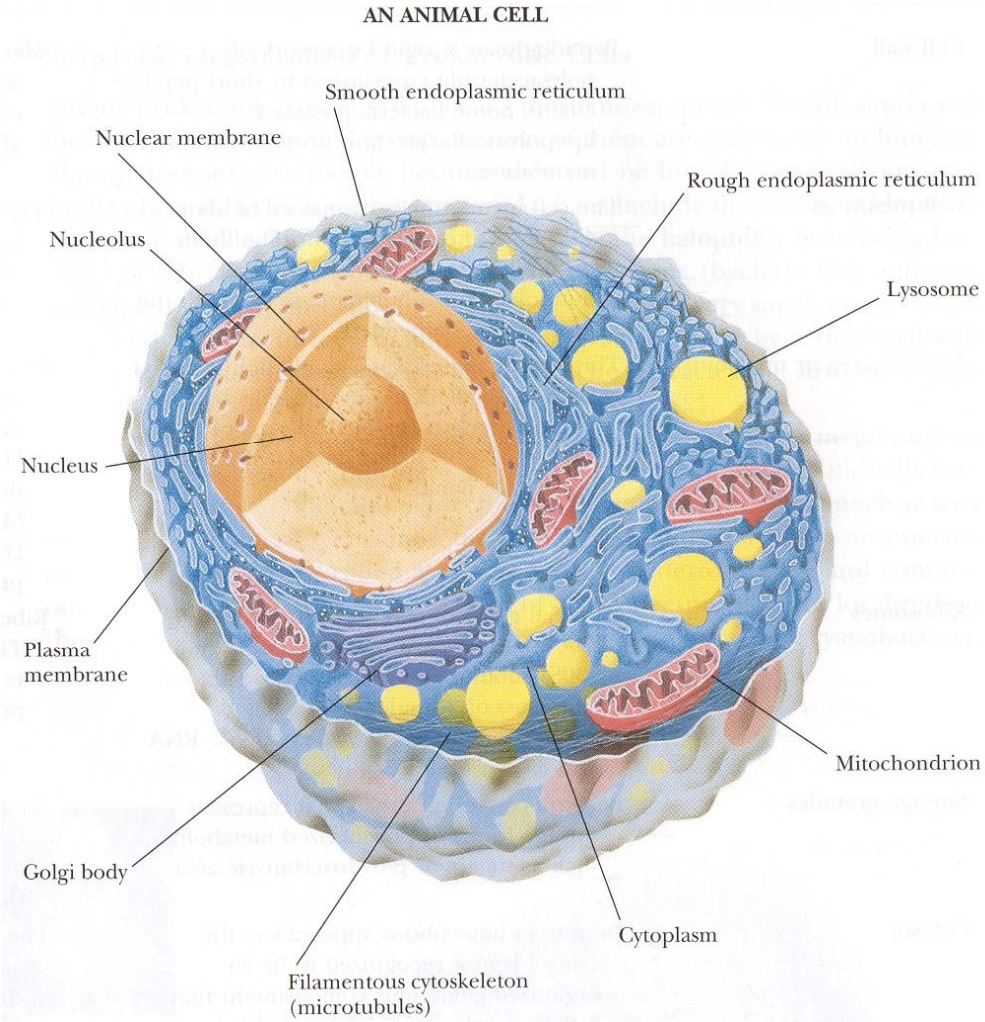
El rango dinámico de expresión

No. copias por célula

Alta: 10,000,000

Media: 100,000

Baja: 1,000



Actualmente es posible identificar proteinas expresadas en menos de 50 copias por célula

El reto del rango dinámico:

Por ej: Detectar y cuantificar 1 analito en suero a 10 pg/mL en presencia de albumina (10^{10} pg/ml)

Abundancia no es sinónimo de importancia

Subproteomas: una forma de simplificar el estudio de los proteomas

Una posibilidad es abordarlo a través de *subproteomas*, subconjuntos de proteínas definidos por alguna propiedad común, por ejemplo,

- **la ubicación subcelular (ej. subproteoma nuclear, proteoma secretado)**

- **o el tipo de modificación presente (ej. fosfoproteoma, glicoproteoma).**

La base metodológica es el fraccionamiento

Proteómica descriptiva y Proteómica comparativa.

Las preguntas que son objeto de estudio mediante herramientas proteómicas.

Investigación basada en hipótesis e investigación generadora de hipótesis.

¨Estudiar un proteoma¨ puede entenderse en un sentido muy amplio:

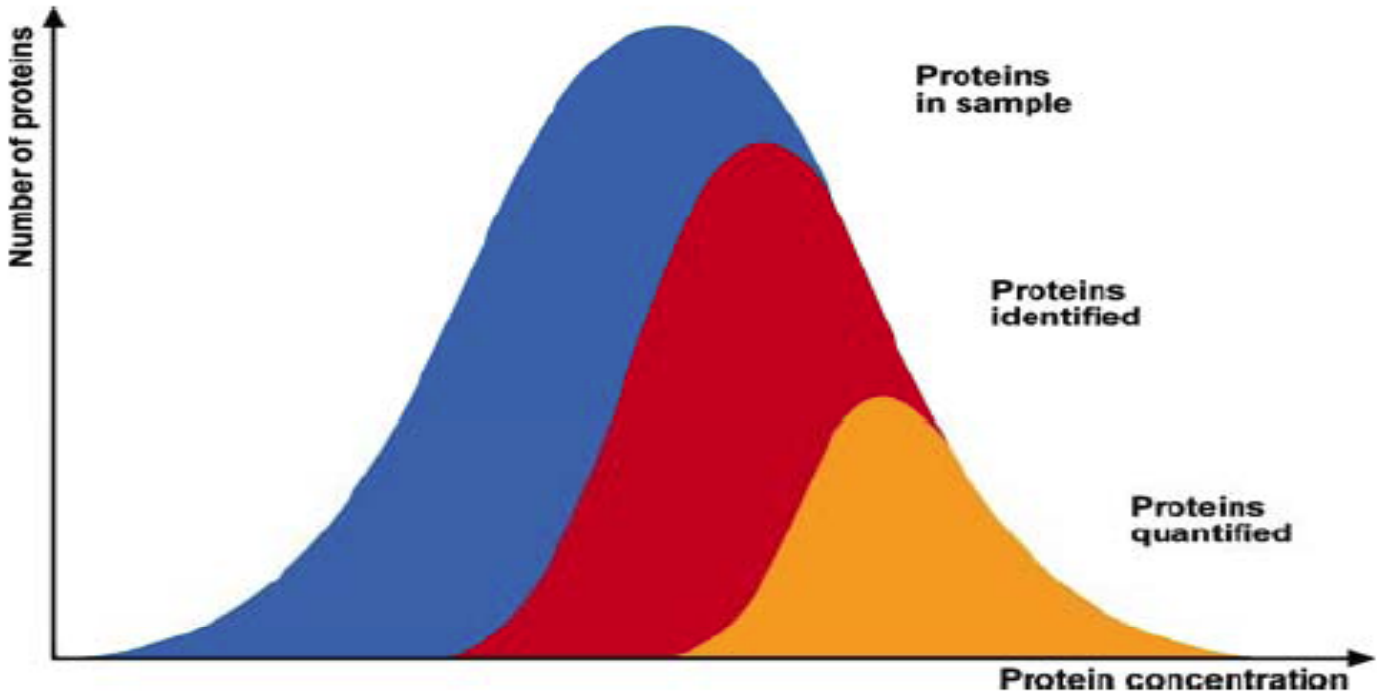
- **identificar ¨todas¨ las especies (*Proteómica descriptiva*)**
- **su tiempo de vida media (*Proteómica temporal*)**
- **su migración intra e intercelular**
- **su asociación con otras proteínas y las funciones de estas asociaciones (*Proteómica funcional*)**
- **los cambios cuantitativos en respuesta a una perturbación (*Proteómica cuantitativa*)**

Proteómica descriptiva **se propone caracterizar ¨todas¨ las proteínas que se expresan en un organismo o en una subestructura celular, de modo semejante a cómo la genómica se propone identificar ¨todos¨ los genes.**

Proteómica comparativa **estudia dos o más estados de un sistema (una célula, un tejido…): uno en condiciones ¨normales¨ y otro en condiciones ¨perturbadas¨ (por ejemplo, después de tratamiento con un medicamento). Evalúa los cambios en la expresión de proteínas a nivel cuantitativo.**

La proteómica comparativa es siempre cuantitativa.

•Normalmente en un experimento de proteómica se pueden identificar entre 500-3000 proteínas:



Esto representa menos de:

- 50 % proteoma de organismos unicelulares.
- 10 % proteoma de organismos superiores.

Mezcla compleja de proteínas

Convertir en mezcla de péptidos

↓

Separar péptidos

↓

Identificar péptidos y de ahí las proteínas

Separar proteínas

↓

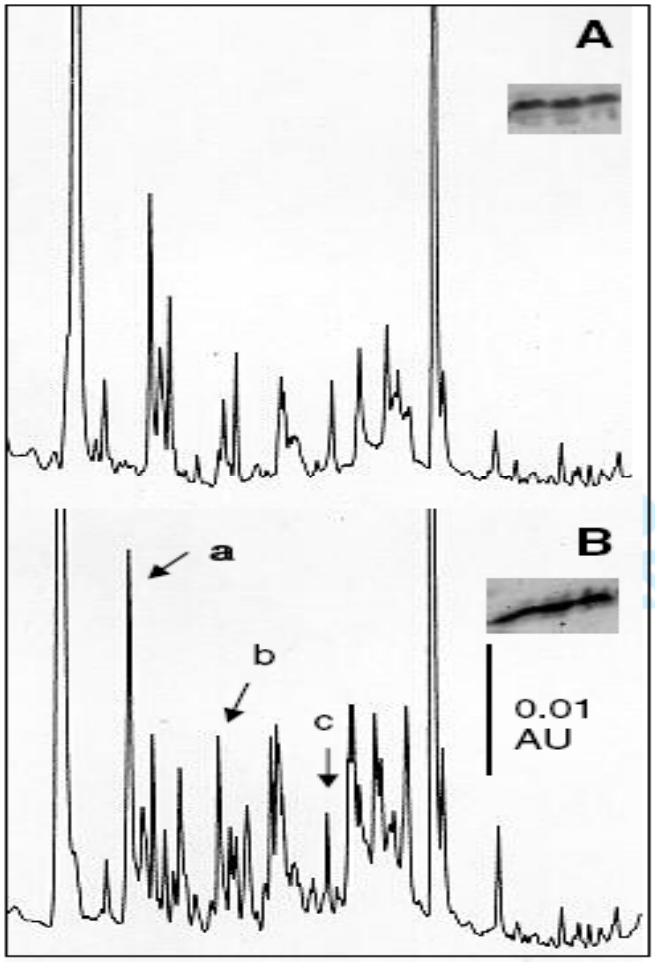
Convertir en mezcla de péptidos

↓

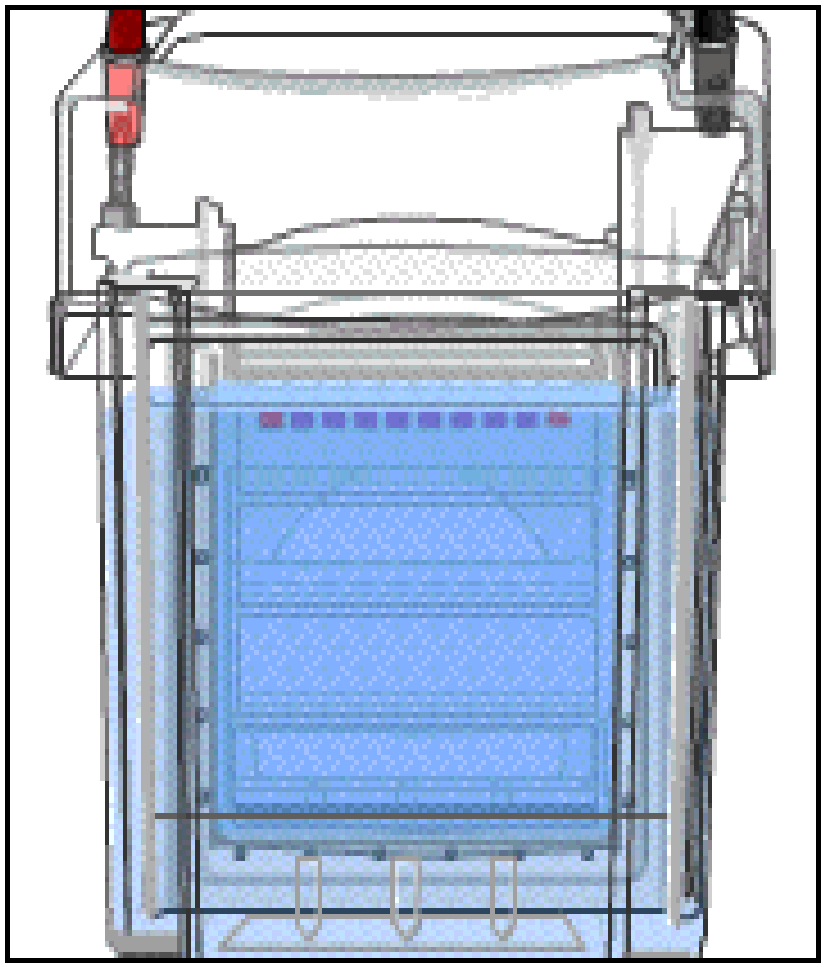
Identificar péptidos y de ahí las proteínas

Principales métodos de separación en Proteómica

Cromatografía



Electroforesis



Técnicas de uso en proteómica

Propiedad	Cromatografía	Electroforesis
Carga neta	Cromatoenfoque, intercambio iónico	Focalización isoeléctrica, Free-flow, OFF gel
Tamaño	Exclusión molecular	SDS-PAGE
Tamaño + carga	No	PAGE nativo
Hidrofobicidad	Interacción hidrofóbica, fase reversa, interacción hidrofílica	No
Afinidad	Ag/Ab, enzima/sustrato, ligando/receptor	No

Fraccionamientos basados en primera separación por punto isoeléctrico:

- **Electroforesis bidimensional**
- **Electroforesis OFF Gel**

Electroforesis bidimensional:

Primera dimensión: separación por p*I*

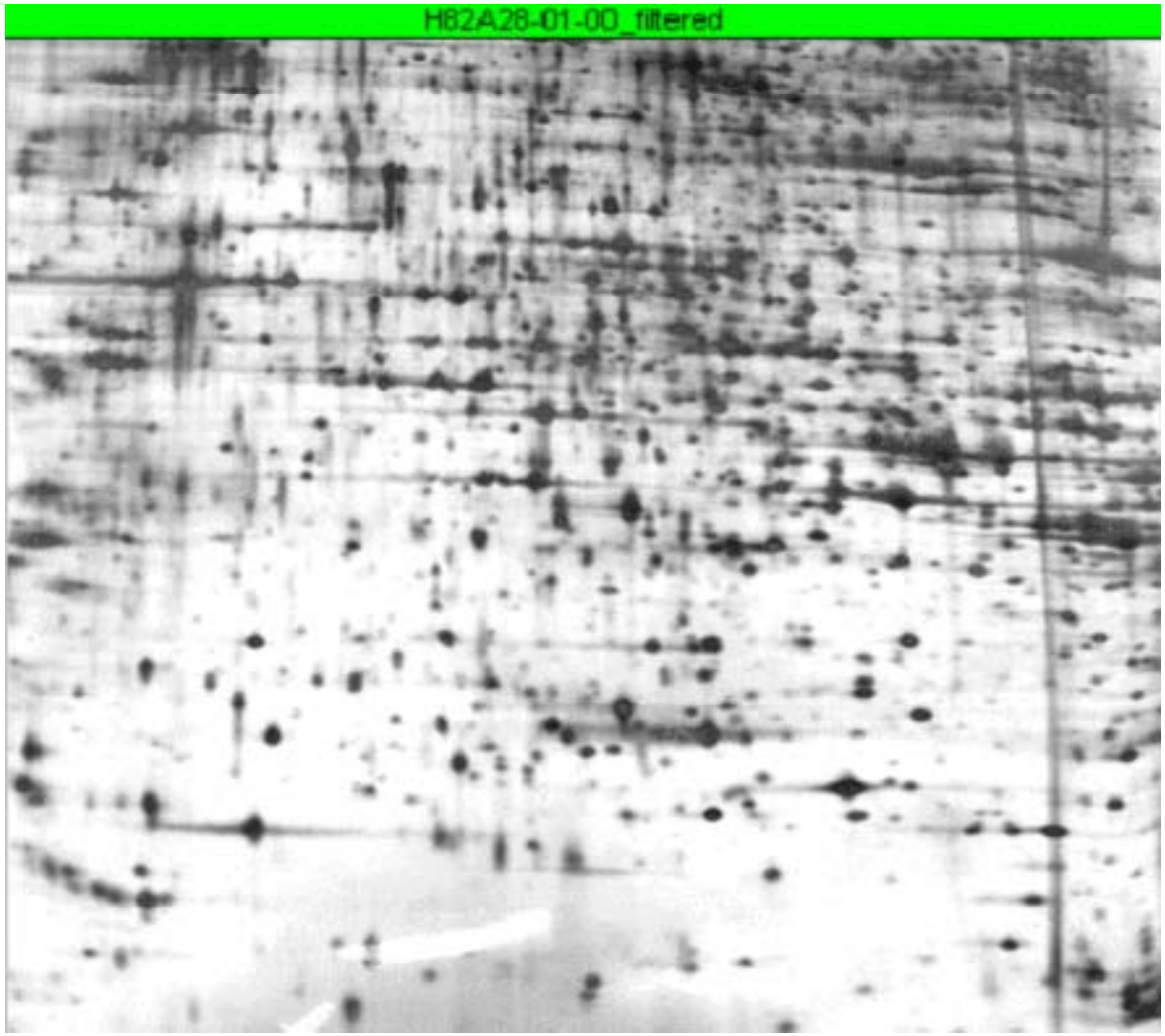
Focalización en anfolinas o en inmobilinas

Segunda dimensión: separación por masas

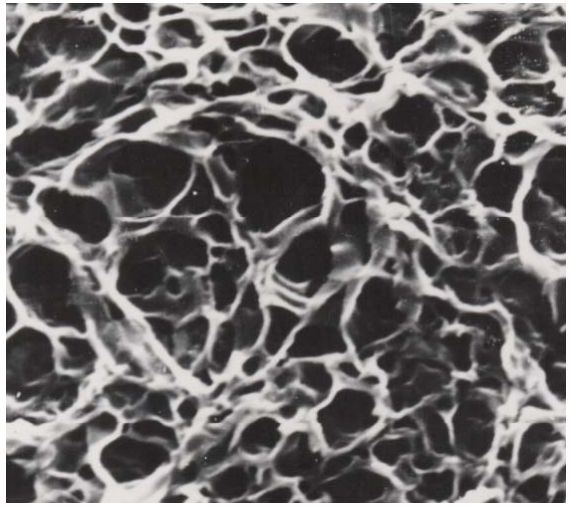
Electroforesis con SDS (Laemmli)

Separación de mezclas complejas de proteínas por electroforesis bidimensional

pI=3.0 pI=10.0



120 kDa



10 kDa

La primera etapa: separación por carga

Separación de péptidos y proteínas por focalización

@ Los aminoácidos ácidos y básicos:

suministran carga eléctrica dependiendo del pH donde se encuentran.

@ pI:

pH donde las cargas negativas aportadas por los residuos ácidos se iguala las cargas positivas aportadas por los residuos básicos (carga neta cero).

Separación de péptidos y proteínas por focalización

@ En un gel con un gradiente de pH:

la proteína migra bajo la acción de un campo eléctrico hasta alcanzar la región de gel cuyo pH coincide con su punto isoeléctrico.

Entonces, queda eléctricamente neutra por lo que el campo eléctrico deja de tener efecto sobre ella.

Dos formas de construir un gradiente de pH en un gel

Anfolitos portadores
(carrier ampholites)

Incorporar a la muestra una mezcla de miles de moléculas orgánicas cargadas, anfolitos, que cubran todo el rango de pI y con alta capacidad tampón.

Ellas se ordenan según su carga, co-migran con la muestra hasta su zona de pI y ahí quedan fijas mientras se mantenga el campo eléctrico.

IPG: Immobilized pH Gradient

Copolimerizar bajo un campo eléctrico una mezcla de miles de moléculas orgánicas cargadas, immobilinas, con la solución de gel.

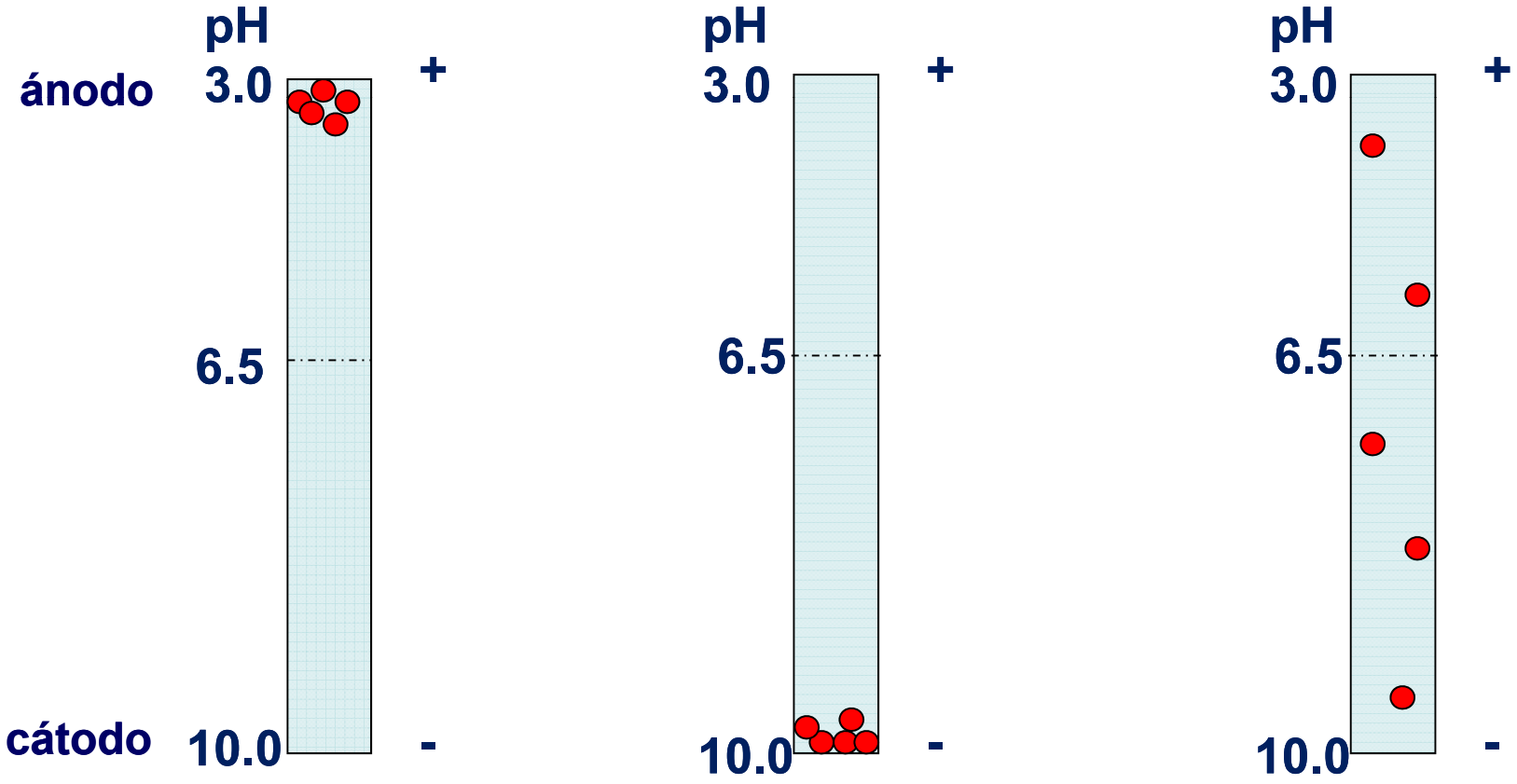
Las moléculas cargadas quedan inmóviles por enlaces covalentes.

Ejemplo: Proteína con pI 6,5

Aplicación anódica. Cuando la proteína está a un pH por debajo de su pI, su carga es positiva

Aplicación catódica. Cuando la proteína está a un pH superior a su pI, su carga es negativa

Aplicación por rehidratacion en todo el gel.

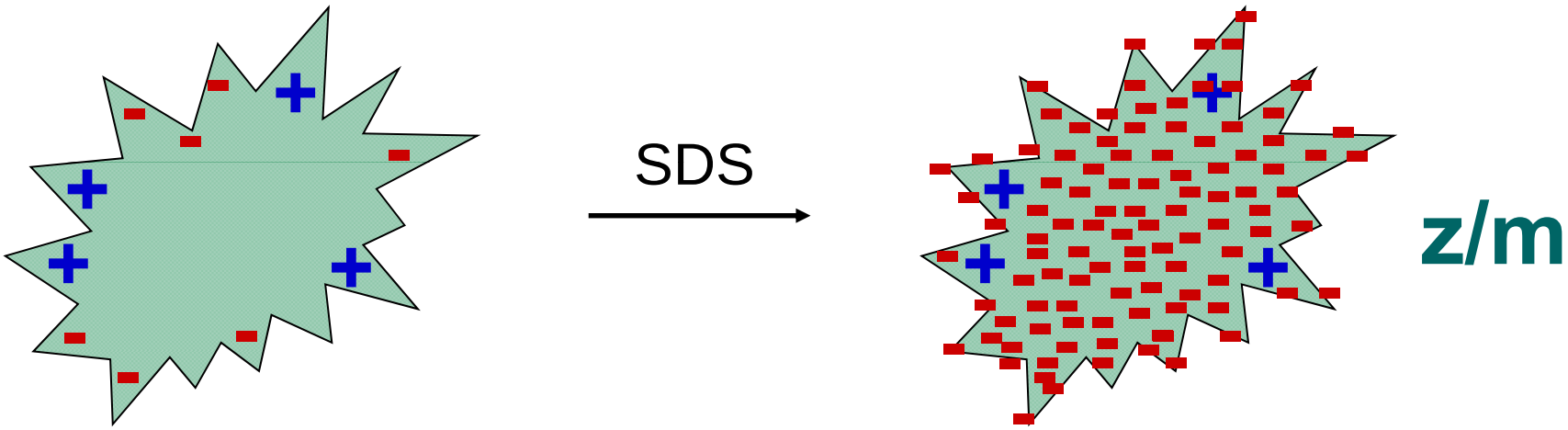


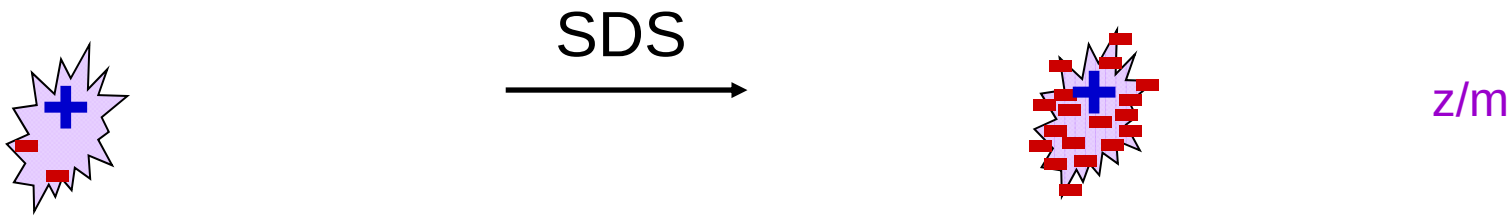
La segunda etapa: separación por tamaño

SDS: Detergente aniónico

CH_3-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-SO_4^-

dodecil sulfato de sodio





z/m = z/m

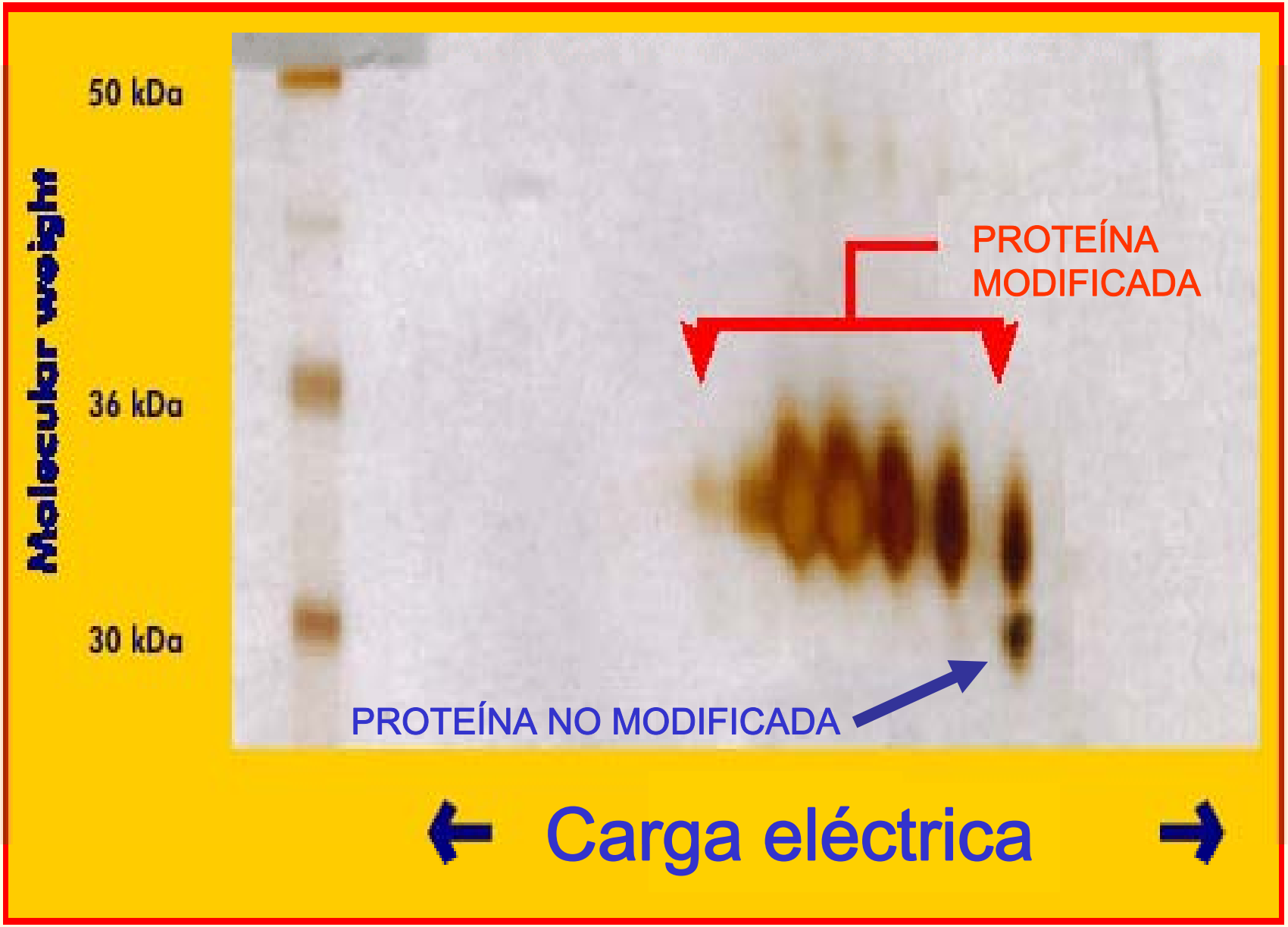
SDS:

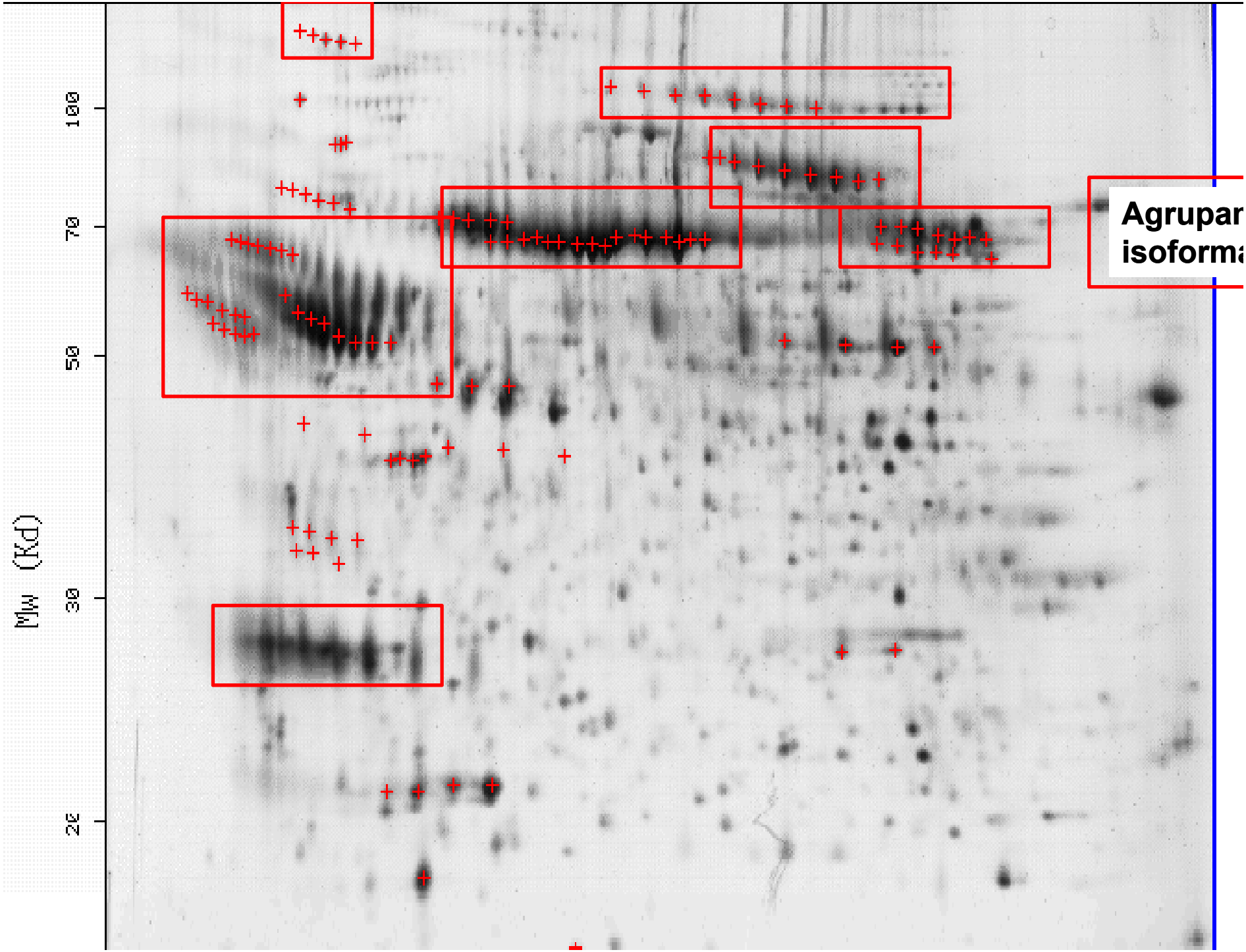
@ homogeniza la densidad de carga (z/m) de todas las proteínas

@ el principio de separación de la técnica de SDS-PAGE es el tamaño de las moléculas.

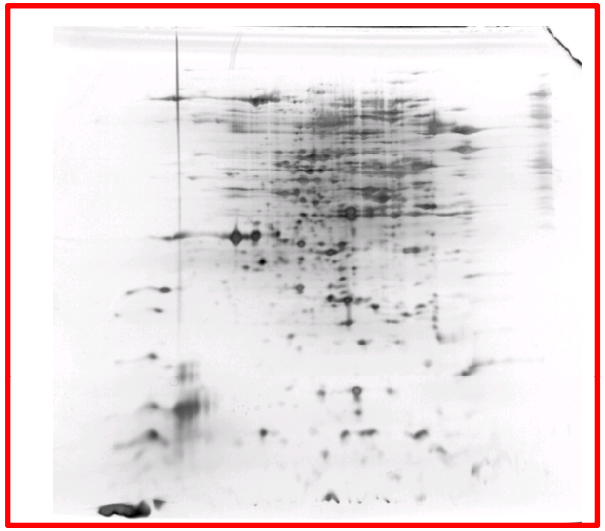
@ permite estimar la masa de una proteína (con referencia a marcadores de masa conocida)

La electroforesis bidimensional sigue siendo el único método que permite resolver isoformas


Molecular weight
50 kDa
36 kDa
30 kDa
PROTEÍNA
MODIFICADA
PROTEÍNA NO MODIFICADA
Carga eléctrica


Mw (Kd)
100
70
50
30
20
Agrupar
isoforma

El número de especies resueltas aumenta con el tamaño del gel



859 spots 7x7cm

Proteínas totales de *E. Coli* 18x20cm 1804 spots

La co-migración de proteínas: varias detectadas en una mancha

Mancha No. 35: seis proteínas identificadas

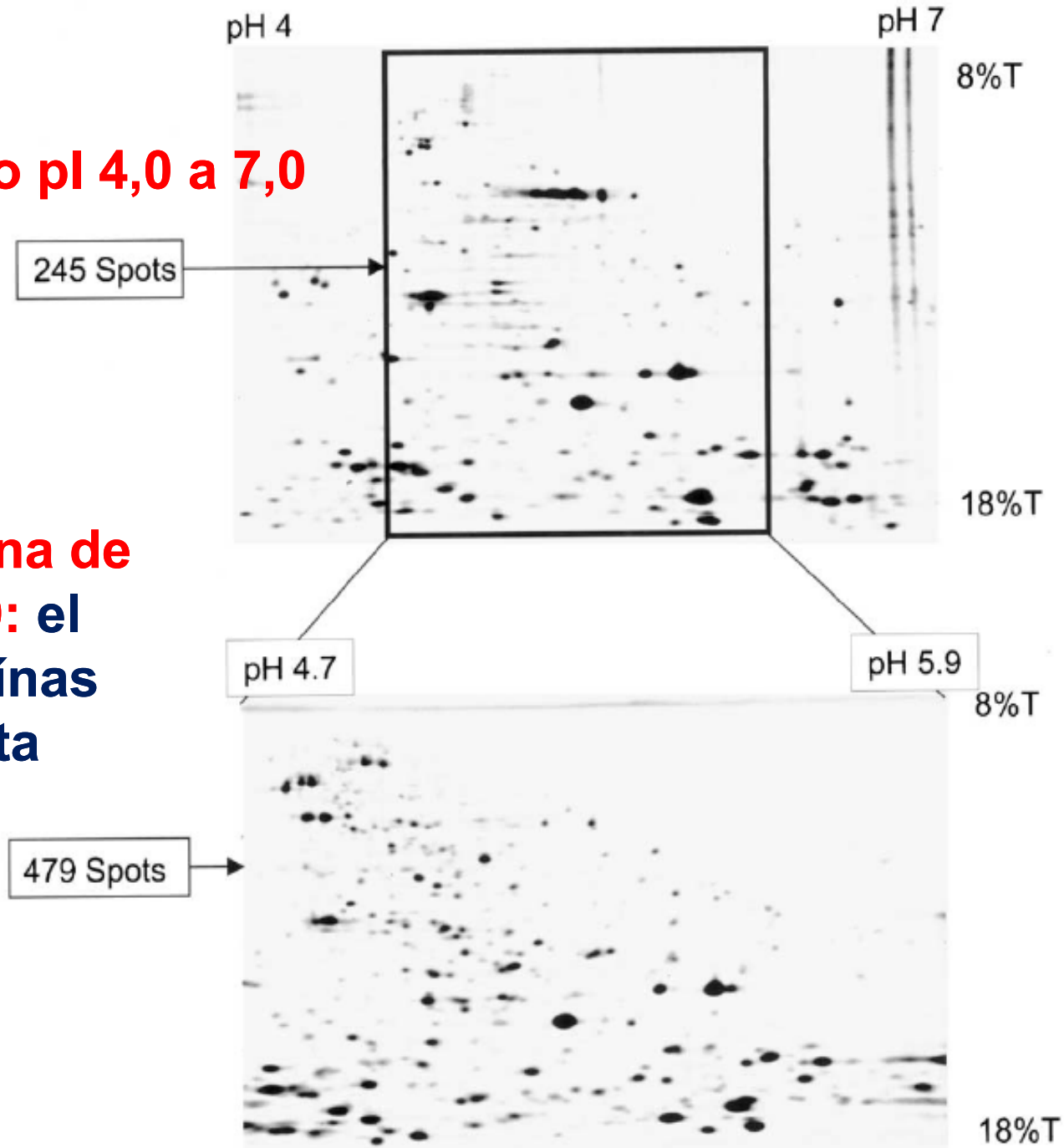
Gene name	Peptide sequences identified†	pI	MW (kD)
GDH1	(K) VIELGGTVVSLSDSK (K) FIAEGSNMGSTPEAIAVFETAR (R) EIGYLFGAYR	5.50	49.6
RPT3	(K) VLPIVSVPER (R) ENAPSIIFIDEVDSIATK	5.32	47.9
SUB2	(R) DVQEIFR (K) LTLHGLQQYYIK (R) INLAINYDLTNEADQYLHR	5.36	50.3
TIF3	(K) NKDTAPHIVVATPGR (R) FLQNPLEIFVDDEAK (R) GSNFQGDGREDAPDLDWGAAR (R) ADLVAVLK (K) ITIPIETANANTIPLSELAHAK (R) EREEVDIDWTAAR (R) EREEPDIDWSAAR	5.17	48.5
VMA1	(K) VGHDNLVGEVIR	5.09	67.7
YFR044C	(R) YPSLSIHGVEGAFSAQGAK (K) LVYGVDPDFTR (K) FISEQLSQSGFHDIK (R) TELIHDGAYWVSDPFNAQFTAAK (K) ILIDGIDEMVAPLTEK	5.54	52.9

La co-migración se reduce si hay subfraccionamiento

Separación rango pI 4,0 a 7,0

Geles Zoom:

Ampliemos la zona de rango pI 4,7 a 5,9: el número de proteínas resueltas aumenta



Detección de proteínas

@ Colorantes (Coomassie Blue)

@ Metales

Tincion positiva (Ag)

Tincion negativa (Zn-ImH)

@ Reactivos fluorescentes

Sypro

DIGE

@ Marcaje radiactivo

Propiedades de las técnicas de detección:

- @ Sensibilidad (lo mínimo que detecta)
- @ Rango dinámico (lo mínimo y lo máximo)
- @ Rango de linealidad (zona de respuesta lineal)
- @ Compatibilidad con análisis posteriores

La selección del método depende del propósito del análisis

Sensibilidad de la detección

Tinción de Coomassie R-250:	**100-200 ng**
Tinción con plata para espectrometría de masas:	**100 ng**
Tinción inversa:	**1-10 ng**
Tinción con plata con glutaraldehído:	**1 ng**
Tinción con colorantes fluorescentes (Sypro):	**1 ng-1 pg**
Fluorografía:	**1 pg**
Immunodetección en Western:	**1 pg**

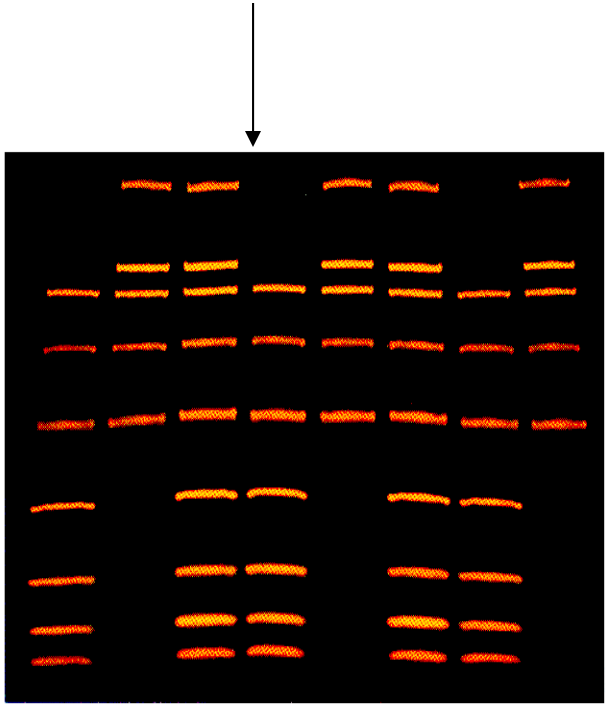
Rango dinámico de los métodos de detección

Coomassie blue, Ag+ (+ laser scanner)	100
Marcaje fluorescente (+ cámara CCD)	> 10,000
Marcaje isotópico (+ RX)	1,000
Marcaje isotópico (+ Phosphorimage)	>100,000

Ejemplos de técnicas de deteccion de proteínas en geles

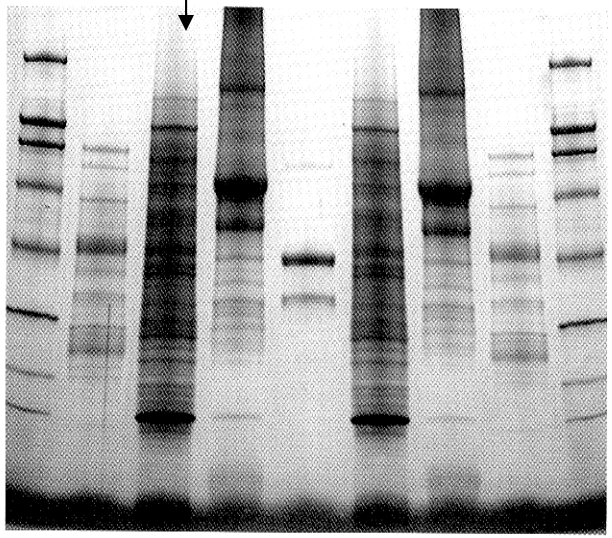
Fluorescente

Sypro Orange



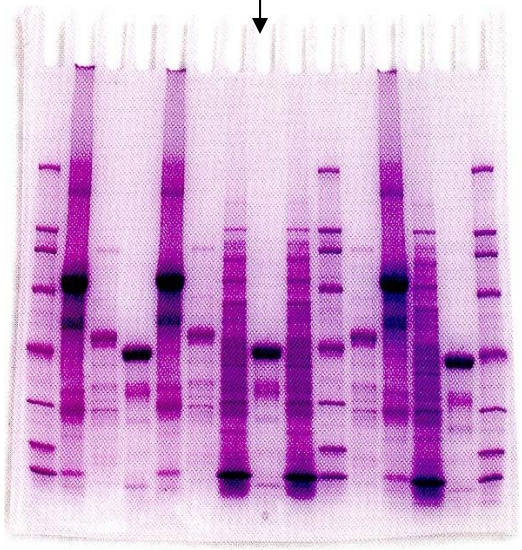
Negativa

Zinc-Imidazol

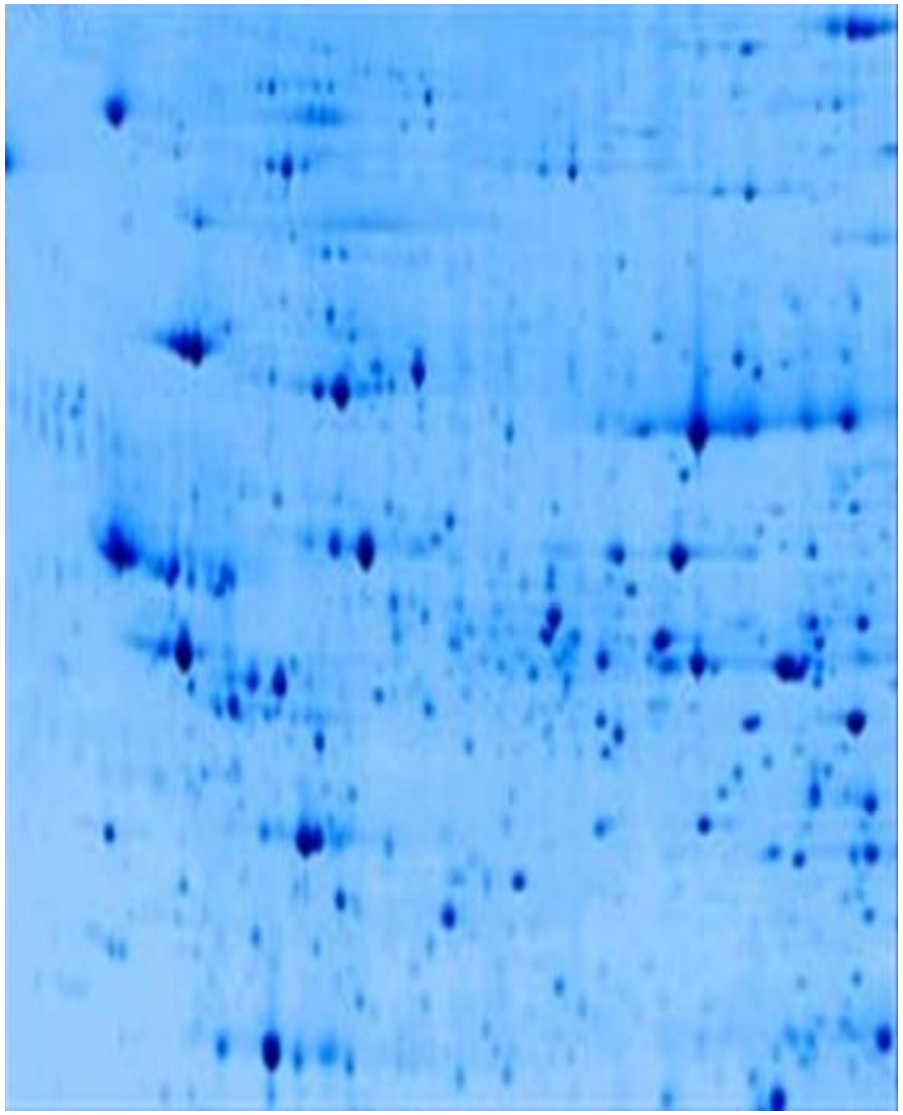


Colorante

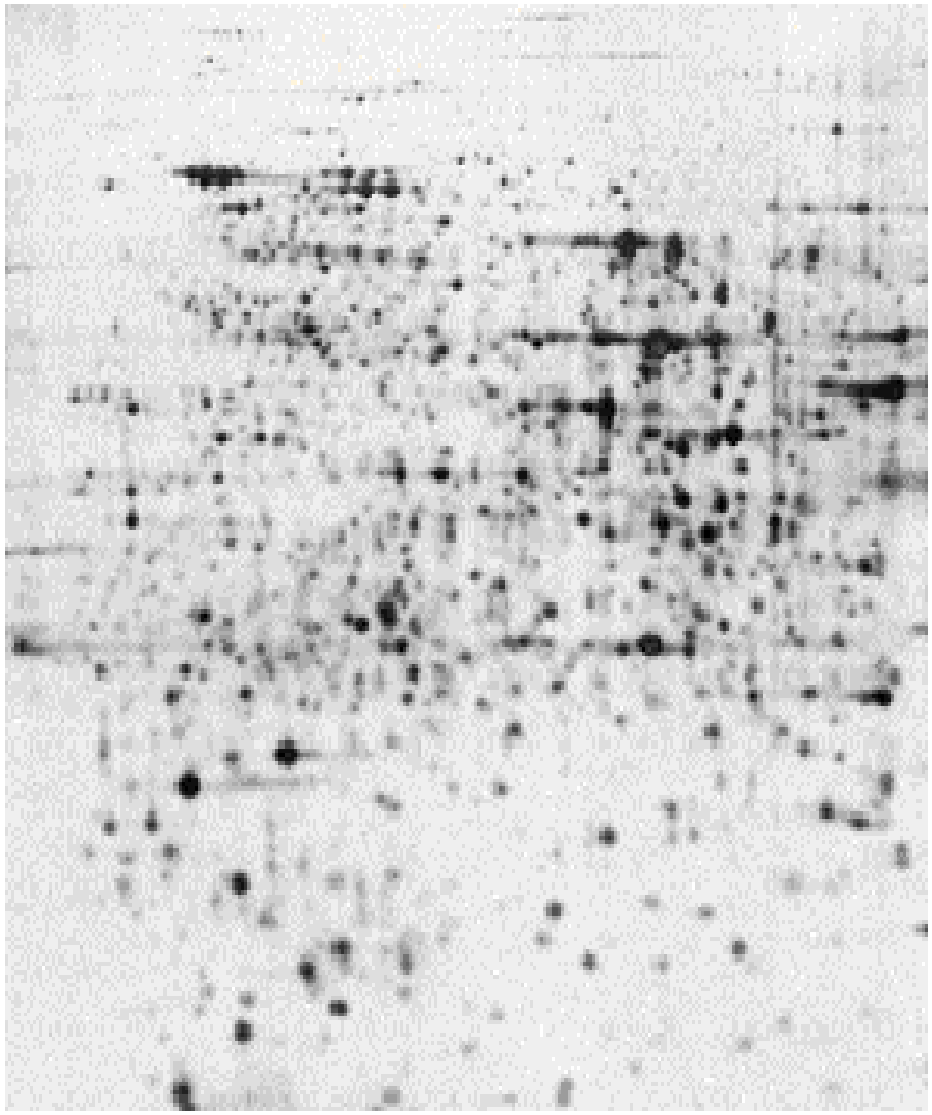
Coomassie Blue R250



Tinción: Coomassie G-250



Tinción: Nitrato de plata



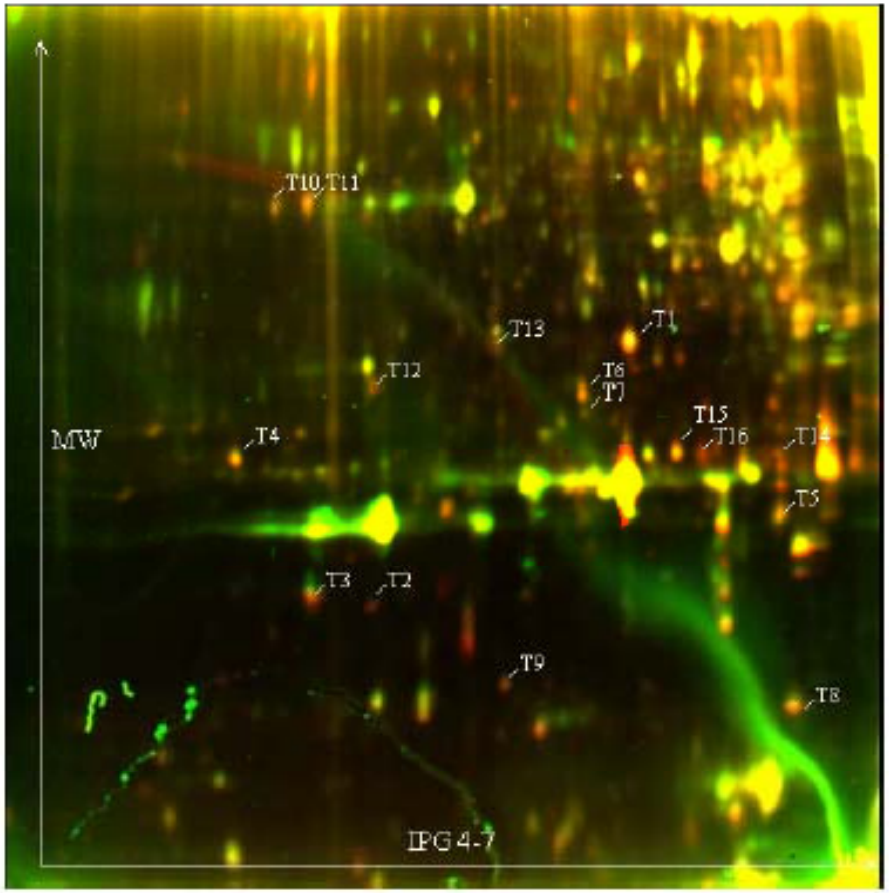
Differential Gel Electrophoresis (DIGE)

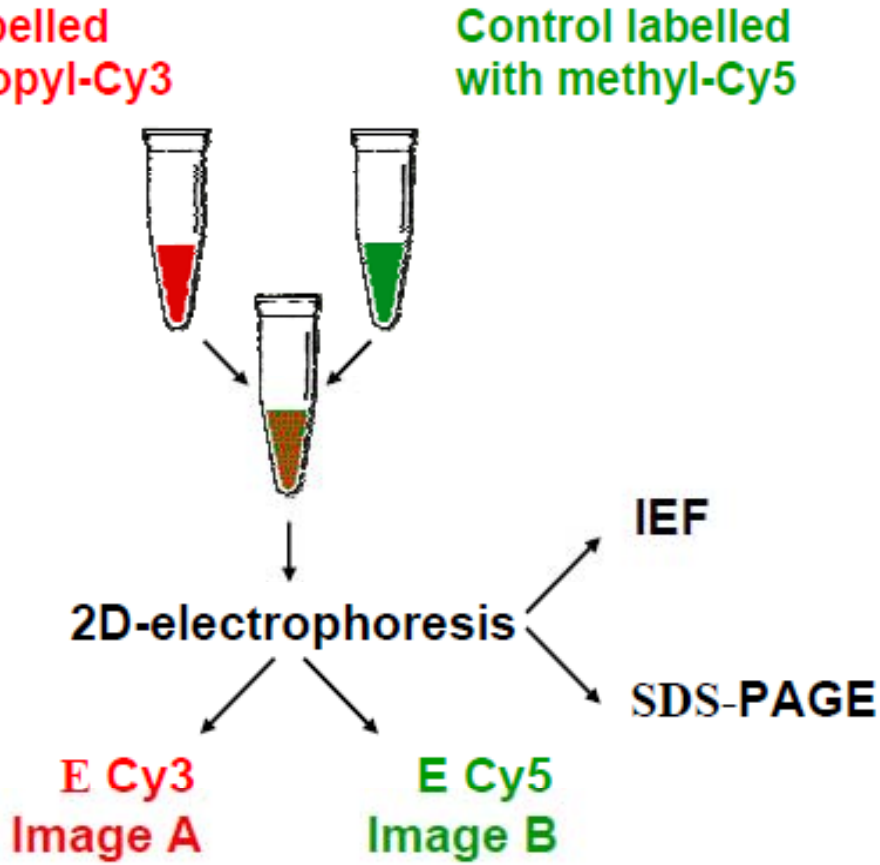
a)¨colorear¨ las muestras antes de separar: un ¨color¨ diferente para cada condición

b)Mezclar las muestras que se comparan, separar en un solo gel,

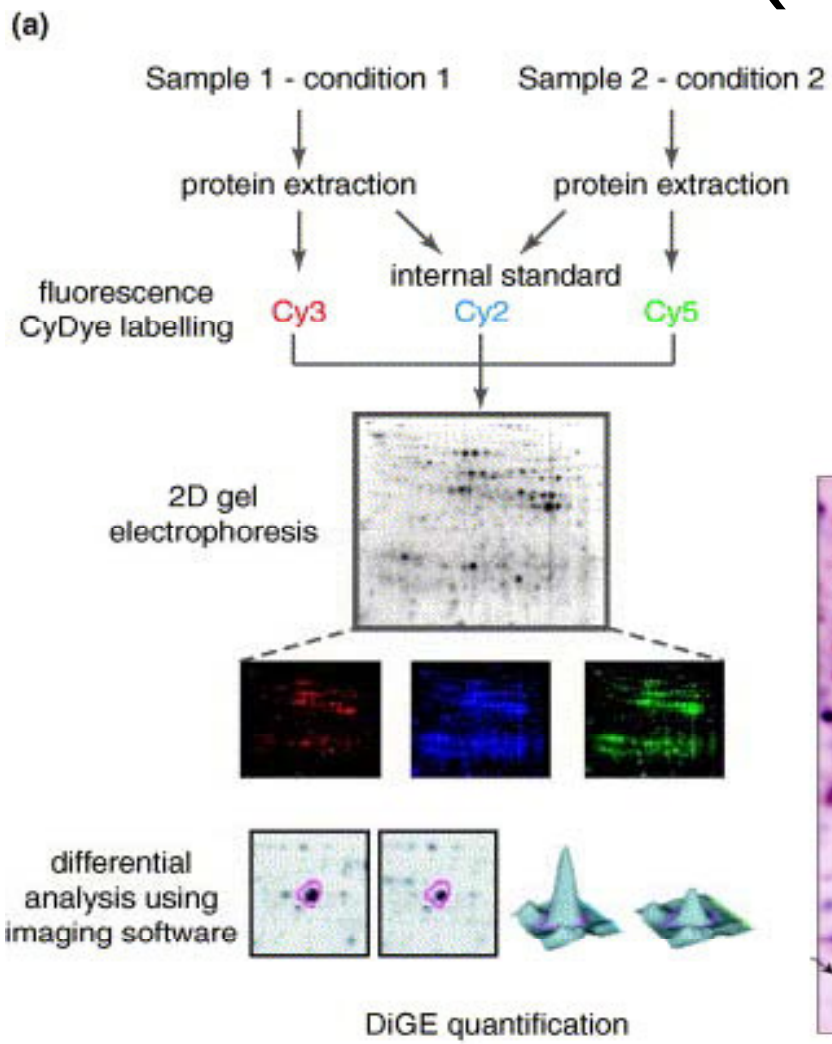
c)¨ver¨ la separación de cada muestra sin interferencia de la otra muestra presente

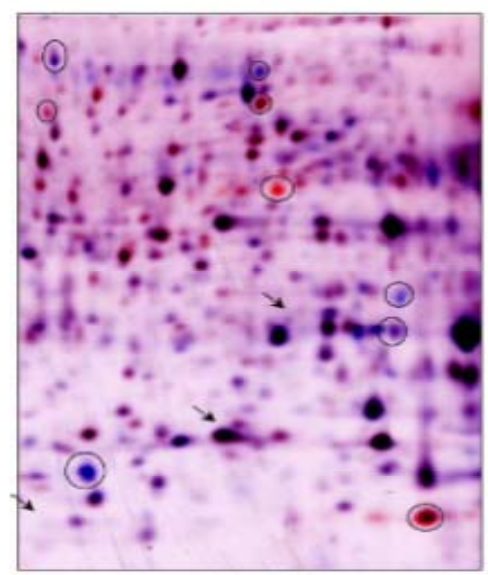
Differential gel electrophoresis (2D-DIGE)



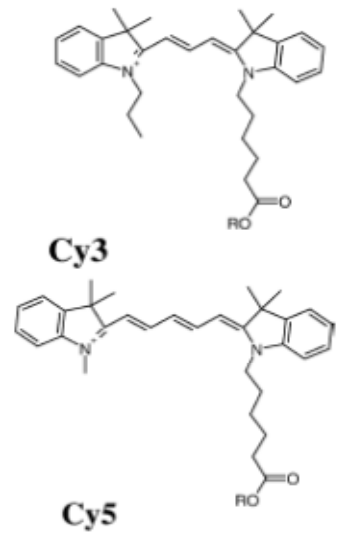


Differential gel electrophoresis (DIGE)





Cy Dye Image



- **Cy3: excitación – 540 nm**
  emisión – 590 nm
- **Cy5: excitación – 620 nm**
  emisión – 680 nm

Ventajas

- ✓ Hasta tres muestras en un único gel, elimina variación gel a gel.
- ✓ Mayor rango dinámico.

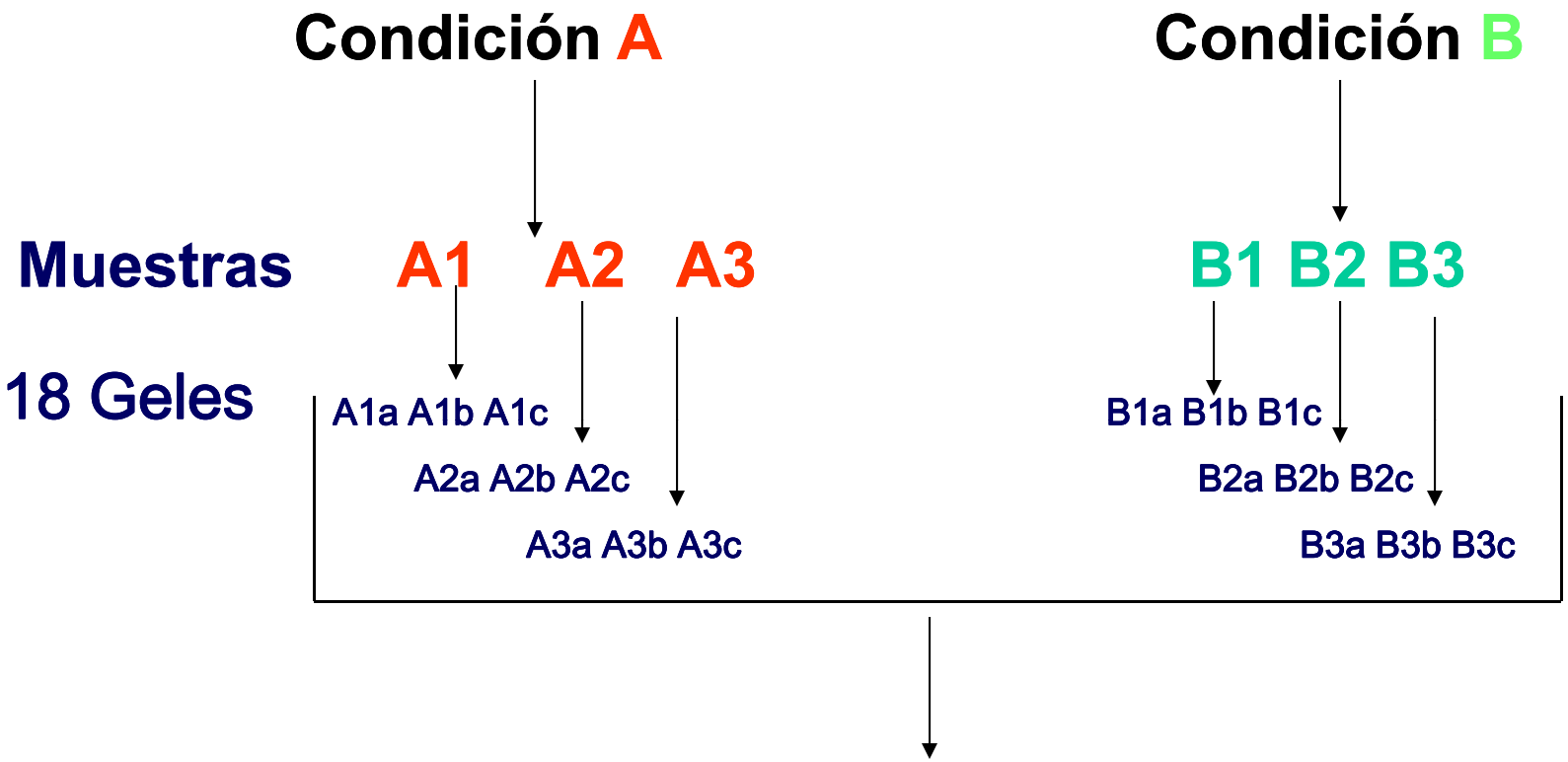
Desventajas

- ✓ Solo marca pocos residuos de Lys: algunas proteínas no se detectan
- ✓ Los fluoróforos aumentan 500 Da (un corrimiento respecto a su posición)
- ✓ Requiere observación ¨electrónica¨ (imagen ¨invisible¨). Precio elevado.

Proteómica comparativa basada en 2DE:

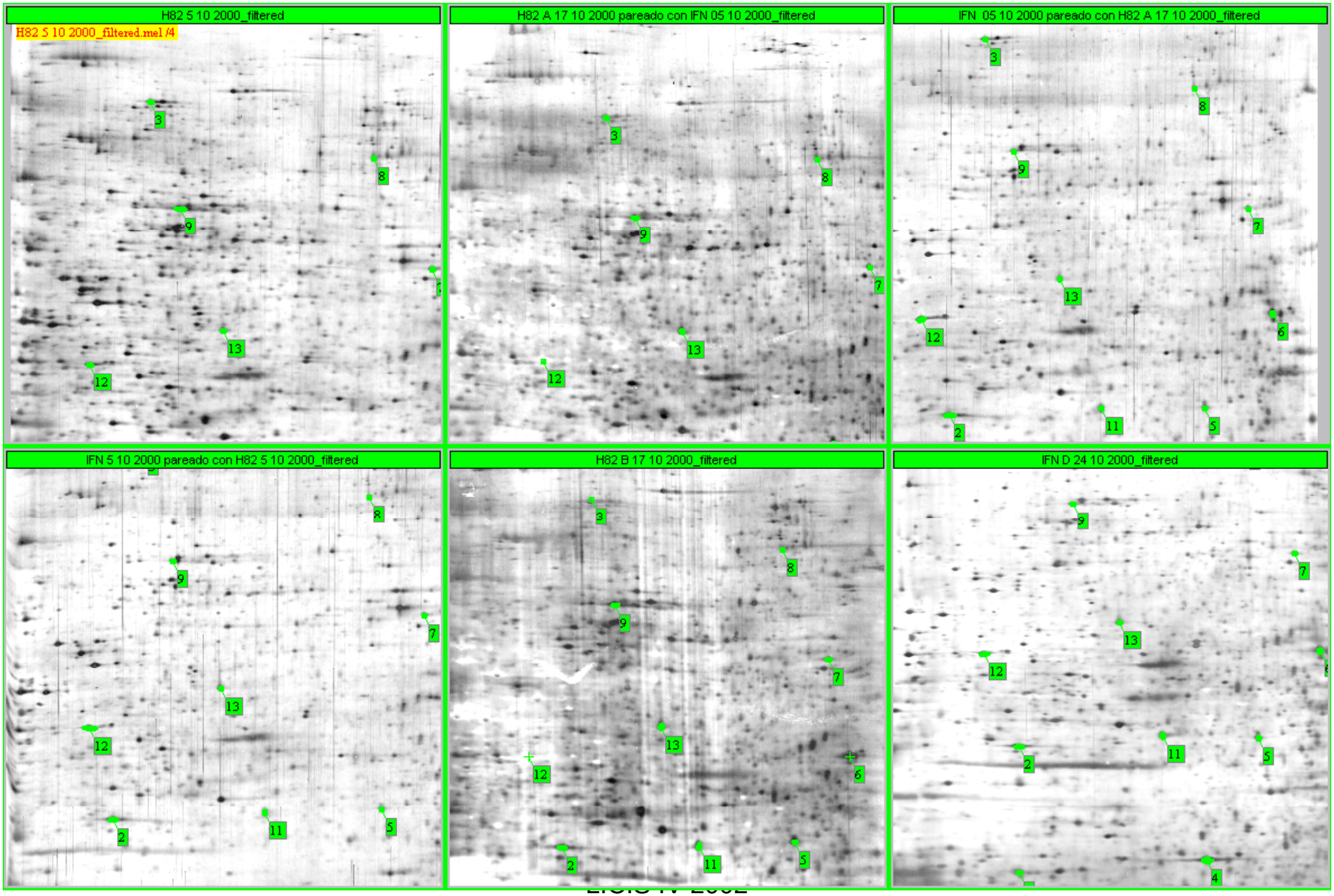
- es un **método relativo**: se estiman los factores de cambio
- la cuantificación ocurre a **nivel de proteínas**
- se basa en el **análisis comparativo de las imágenes** de geles

PROTEOMICA COMPARATIVA

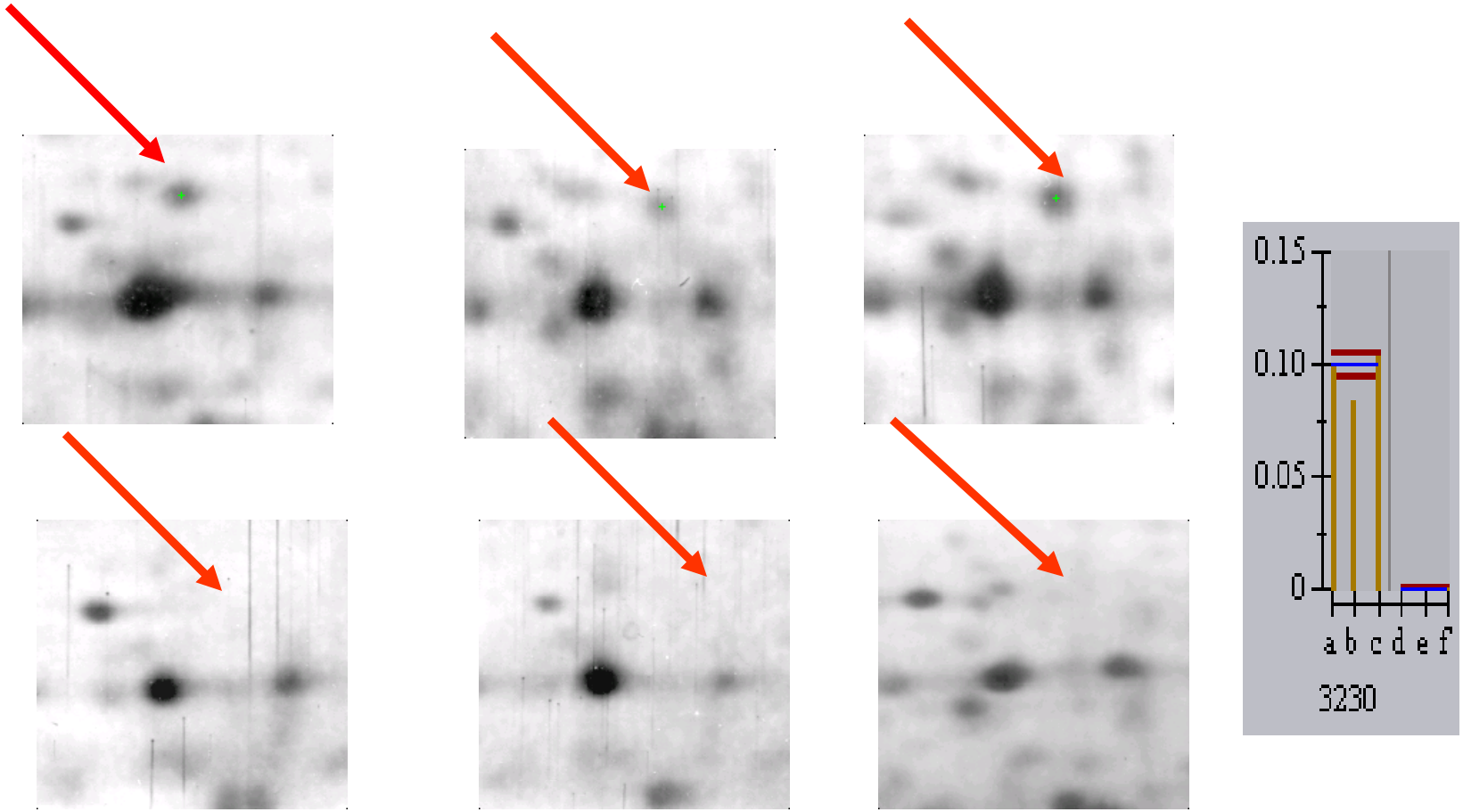


Análisis de Imágenes

Identificación de cambios estadísticamente significativos


H82 5 10 2000_filtered
H82 5 10 2000_filtered.mel /4
H82 A 17 10 2000 pareado con IFN 05 10 2000_filtered
IFN 05 10 2000 pareado con H82 A 17 10 2000_filtered
IFN 5 10 2000 pareado con H82 5 10 2000_filtered
H82 B 17 10 2000_filtered
IFN D 24 10 2000_filtered

Proteómica comparativa, cambios en spot 3230



Cuidado!: incorrecto decir que ¨no se expresa¨ :
Solo puedo afirmar que se reduce su expresión porque no la detecto.

Una forma de ¨ver¨ más proteínas en 2DE:

@ Fraccionamiento en subproteomas

@ Hacer varios ¨subproyectos¨ : proteínas
básicas (pI 7,5 a 13)
ácidas-neutras (pI 3 a 7,5)
de baja masa (2,000-10000 Da)
de alta masa (mas de 150 kDa)

@ ¨zoom¨ : Seleccionar un rango estrecho de pI (ej., pI 4,5 a 4,5, en 18 cm de longitud)

Principales ventajas de la electroforesis bidimensional:

@ Alto poder resolutivo

@ Ideal para separar isoformas

@ Permite identificar productos de degradación

@ Imagen "física" de la complejidad de la muestra

@ Fácil evaluación de la reproducibilidad de los resultados (incluso cuantitativos)

@ Detección fluorescente tiene amplio rango dinámico (pero muy costosa)

Principales desventajas de la electroforesis bidimensional:

@ No funciona bien con proteínas hidrofóbicas, básicas y de muy alta talla

@ Número de proteínas identificadas inferior a métodos basados en separación de péptidos

@ Manipulación intensiva, difícil automatización.

Fraccionamiento basado en una primera separación por electroforesis OFF Gel

Electroforesis OFF Gel:

Es una separación por focalización isoeléctrica sobre una tira de inmobilinas cubierta de una solución. Las proteínas o los péptidos se separan según su pI y se mantienen todo el tiempo en la solución que cubre la tira.

Fraccionamiento por electroforesis OFF Gel:

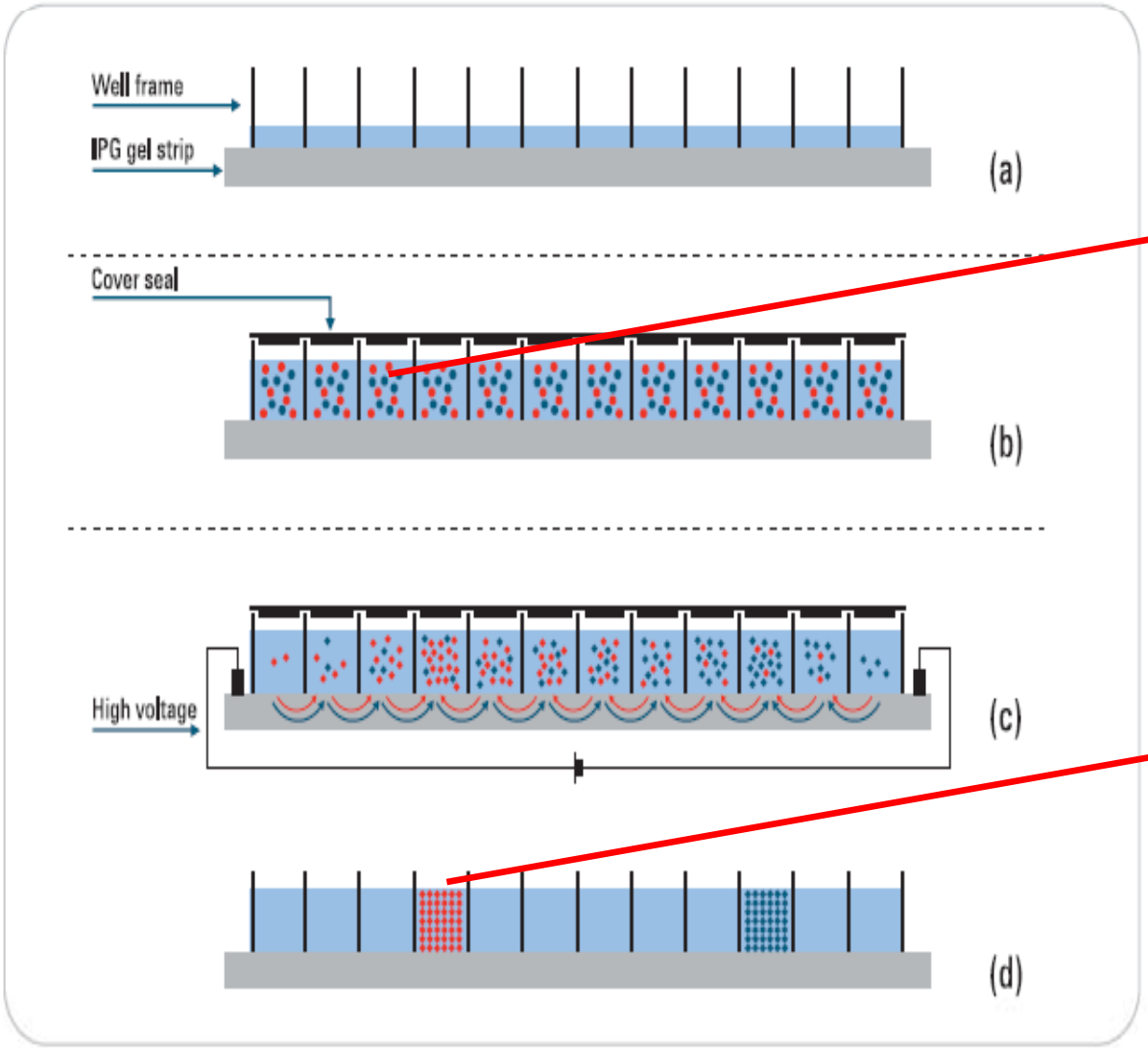
Separación OFF Gel según pI

↓

Moléculas en solución

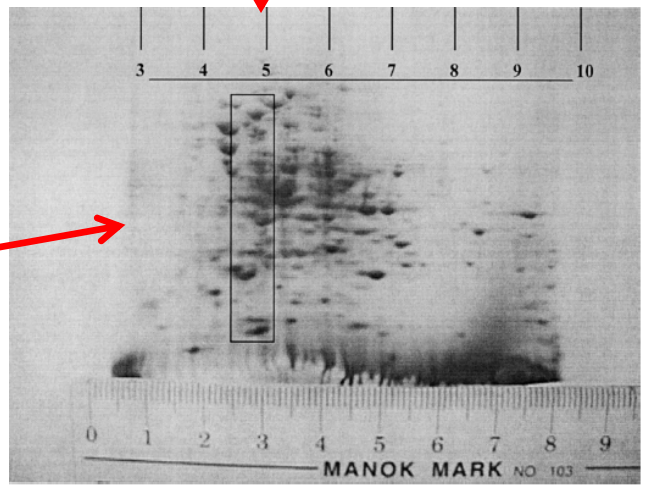
↓

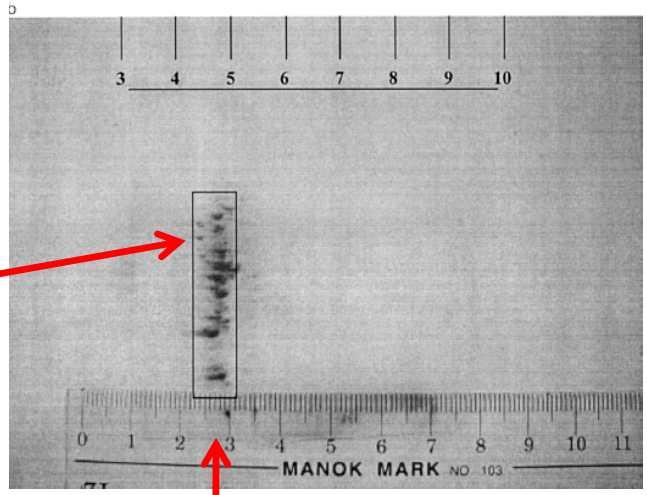
HPLC LC MS-MS
(separación por fase reversa
e identificación en línea)



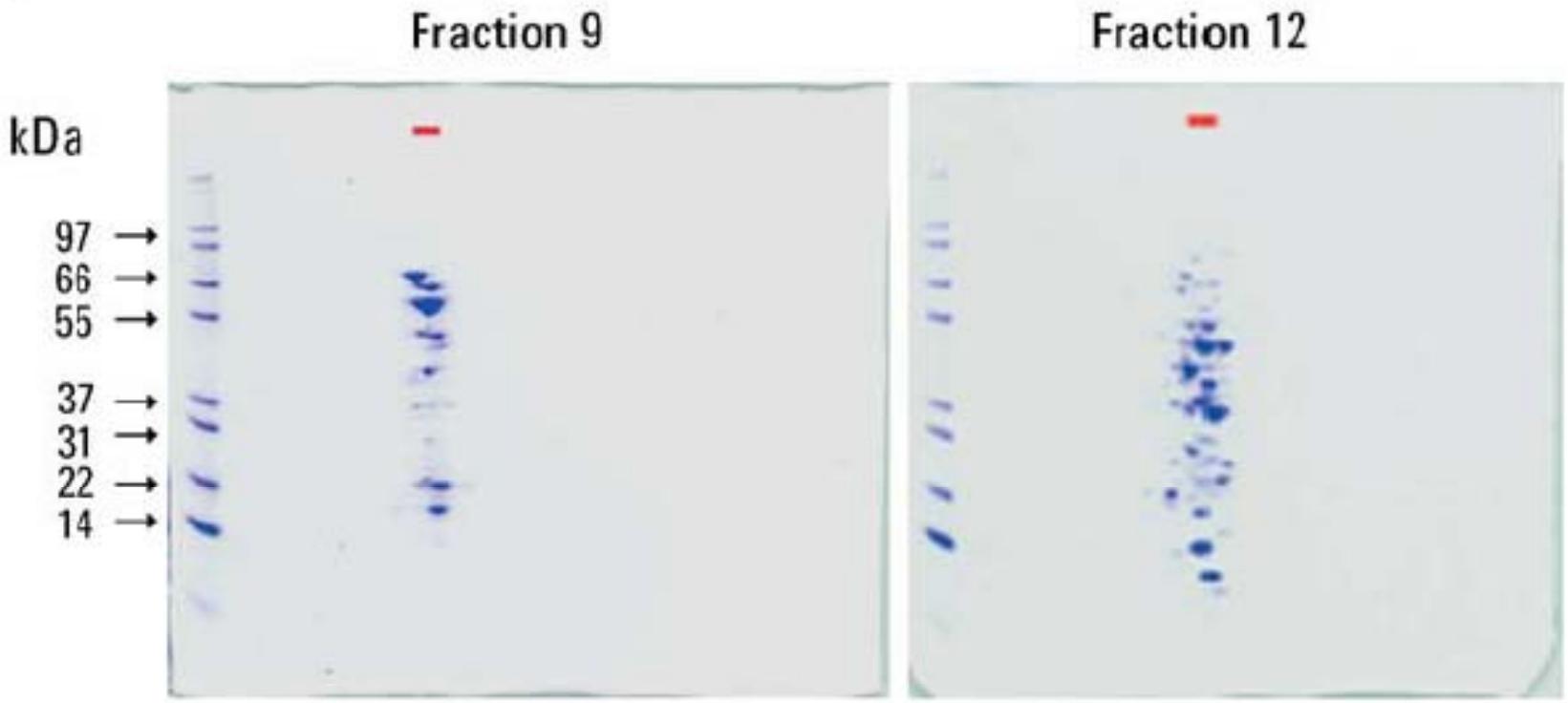
Fractionation principle

Muestra de partida





Una de las fracciones OFF Gel



Fractionation of E.coli into 24 fractions (pH 4-7). Analysis of fractions 9 and 12 by 2D electrophoresis. The theoretical resolution of 0.13 pH units is indicated by the red bars.

Características de OFF GEL

@ Alto poder resolutivo

@ Ideal para separar isoformas

@ Mayor recobrado de proteínas que 2DE

@ Perfectamente adaptable a separación de péptidos

@ Se acopla muy bien con HPLC en línea con MS

@ Mucho más simple y reproducible que 2DE

@ Requiere preparación de muestra muy cuidadosa (como 2DE) . Muy sensible a presencia de sales

¨High throughput proteomics ¨

Alude a experimentos que generan mucha información en breve tiempo y con relativamente poca intervención del analista.

Las técnicas basadas en electroforesis bidimensional no se consideran de este tipo.

Las técnicas de alto flujo requieren notable automatización, y generalmente de cierta autonomía instrumental en la toma de decisiones durante la adquisición de información,

- **Estudiar las proteinas expresadas en un tejido u organismo en condiciones normales (¨el proteoma normal¨)**

- **Identificar y cuantificar cambios en expresión de proteínas como efecto de un problema biológico (ej. Cáncer), o por acción de agente externo (ej., efectos de un medicamento)**

- **Identificar redes de interacción entre proteínas y entre proteínas y otras moléculas (¨interactomas¨)**

- **Identificar biomarcadores en tejidos o fluidos (ej., sangre, orina, LCR)**

Para cada problema biológico hay varios diseños experimentales posibles y diferentes.

Qué tienen todos en común?:

la conversión de proteínas en mezclas de sus péptidos

La identificación de las proteínas mediante espectrometría de masas de péptidos

El uso extensivo de bases de datos para identificación

El análisis post-proteómico (bioinformático)

Importancia de la existencia de las bases de datos internacionales

Las secuencias y masas experimentales se comparan con las obtenidas *in silico* para todas las proteínas en la base a partir de genomas secuenciados completos o de genes secuenciados en genomas incompletos o bien, a partir de secuencias de proteínas.

La identificación de proteínas en organismos con genoma desconocido es posible en muchos casos por la alta homología con secuencias de genes ya reportados.

¨Computational proteomics¨

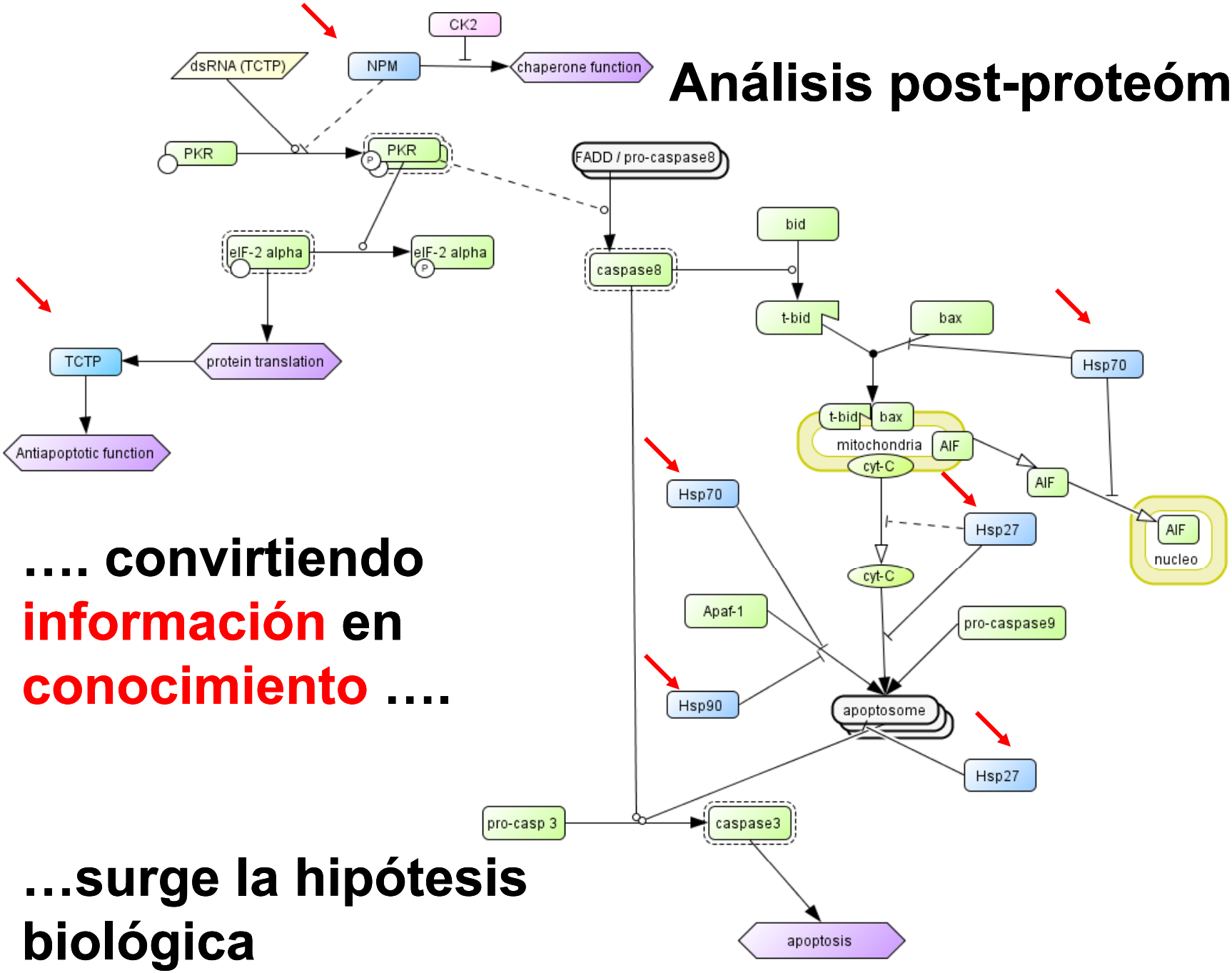
Proteómica computacional. Alude al desarrollo y al uso intensivo de programas para la interrogación de bases de datos y la creación de nuevas bases derivadas. Por ejemplo.

- **la identificación en las bases de datos de péptidos que son exclusivos de una proteína y que por tanto, su identificación equivale a la identificación de una proteína, denominados ¨péptidos proteotípicos¨,**

- **la creación de herramientas de validación de datos (por ejemplo ¨tarjet decoy¨)**

- **la creación de bases de datos de identificaciones basadas en la conjunción de múltiples parámetros físicos de los péptidos, calculados teóricamente, por ejemplo, los puntos isoeléctricos y los tiempos de retención cromatográficos.**

Análisis post-proteómico



…. convirtiendo **información** en **conocimiento** ….

…surge la hipótesis biológica