

# Módulo 1

## Técnicas de separación y Espectrometría de Masas en Ciencias de la Vida

Wojciech Luczaj<sup>1</sup>, Agnieszka Gegotek<sup>1</sup>, Katarzyna Bielawska<sup>1</sup>,  
Elzbieta Skrzydlewska<sup>1</sup>

*<sup>1</sup>Departamento de Química Analítica, Universidad Médica de Białystok, 15-089 Białystok, Polonia*

e-materials del  
Proyecto de Química Analítica Avanzada en Ciencias de la Vida [AACLifeSci]

Este proyecto ha sido financiado con ayuda de  
la Comisión Europea

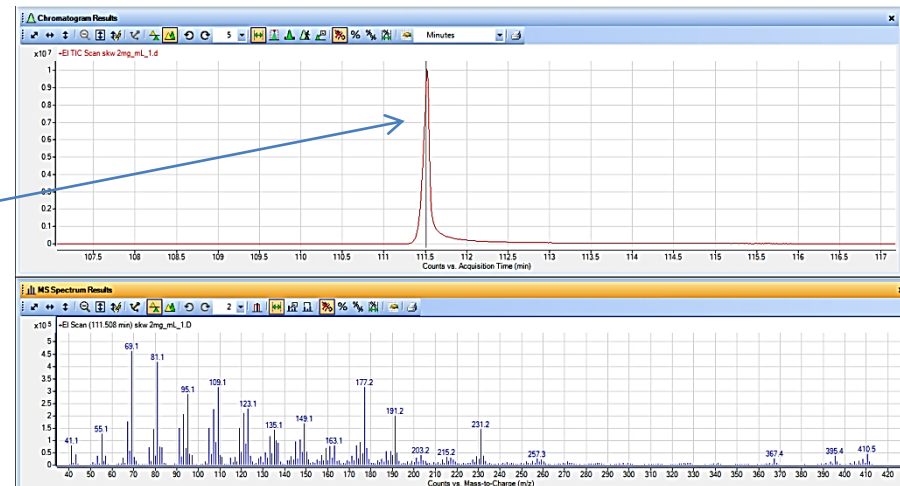
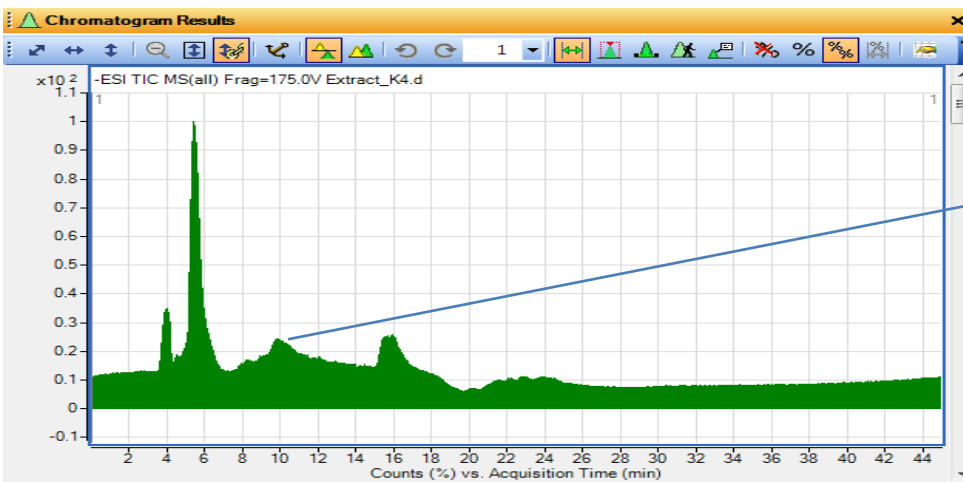


# Espectrometría de masas

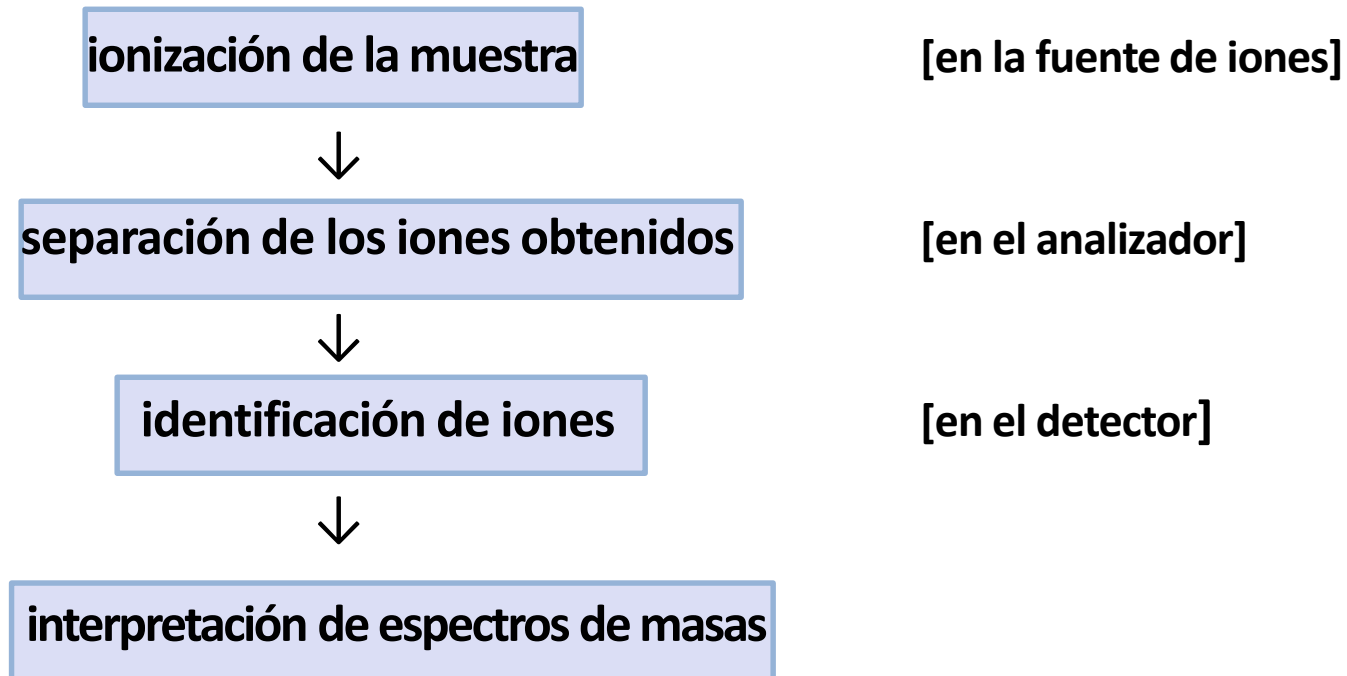
Técnica analítica que permite obtener información sobre:

- ✓ el peso molecular
- ✓ la estructura química
- ✓ cantidades de compuestos examinados

Técnica analítica que permite separar e identificar iones en función de su relación masa/carga ( $m/z$ )

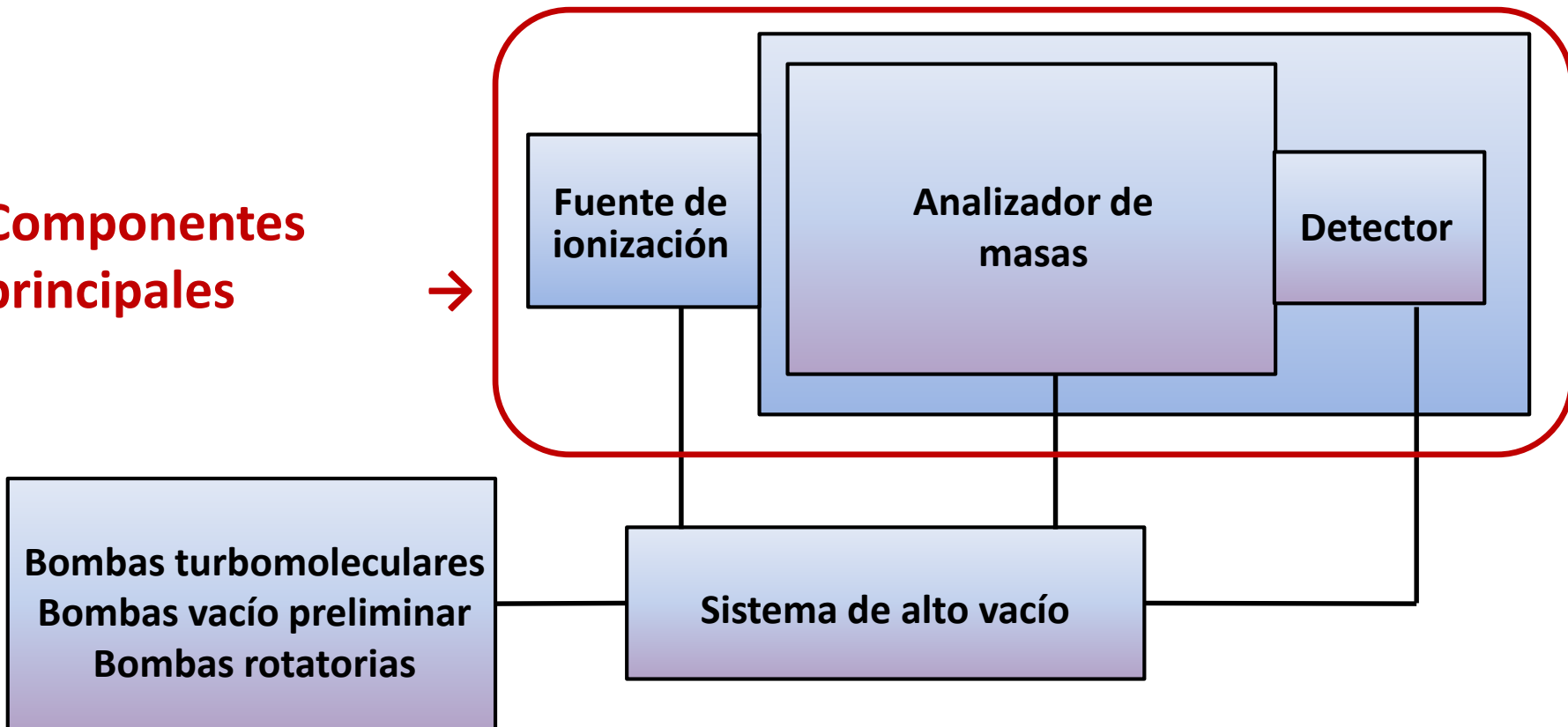


## Etapas del proceso analítico basado en la espectrometría de masas



# Espectrómetro de masas

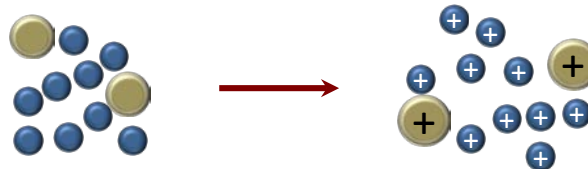
**Componentes principales**



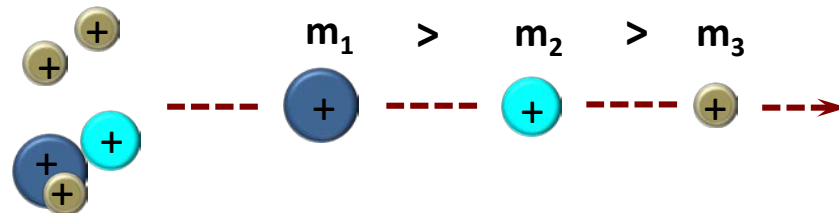


# Flujos de trabajo de la MS

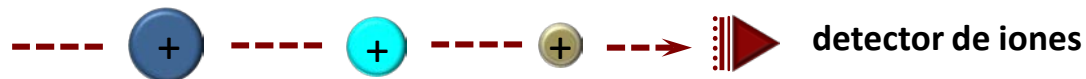
## 1. Ionización del analito



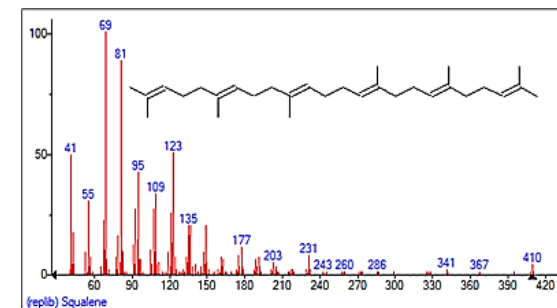
## 2. Separación de las moléculas ionizadas según su $m/z$



## 3. Detección de los iones



## 4. Análisis e interpretación del espectro de masas



## Métodos de ionización

- ionización por electroionización (impacto de electrones) [EI]
- ionización por electrospray [ESI]
- ionización por desorción láser asistida por matriz [MALDI]
- ionización química a presión atmosférica [APCI]

**ionización dura** – productos: **ion molecular** + **elemento fragmentado**

✓ EI

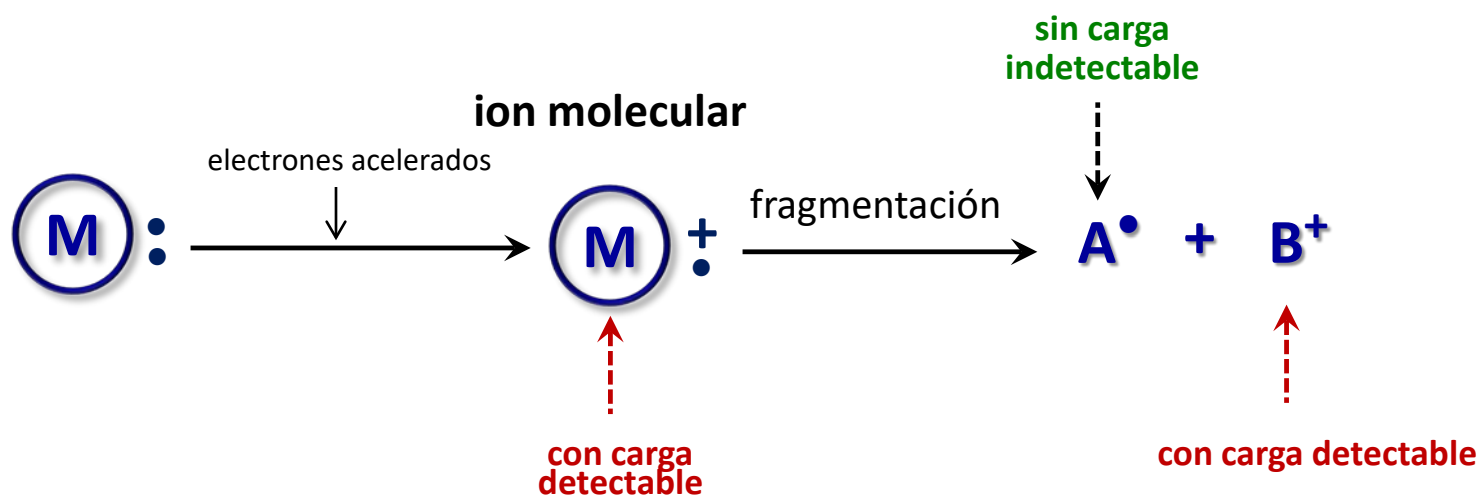
**ionización suave** – el producto principal - **ion molecular**

✓ ESI

✓ MALDI

✓ APCI

## Electroionización (EI)



*Mecanismo de ionización con electrones y fragmentación subsiguiente*

# Electroionización (EI)

## ➤ Analitos habituales:

- relativamente pequeños,
- no polares,
- volátiles,
- termoestables

## ➤ Rango de masas:

- <1 kDa

## ➤ Introducción de la muestra:

- GC o líquido/sólido

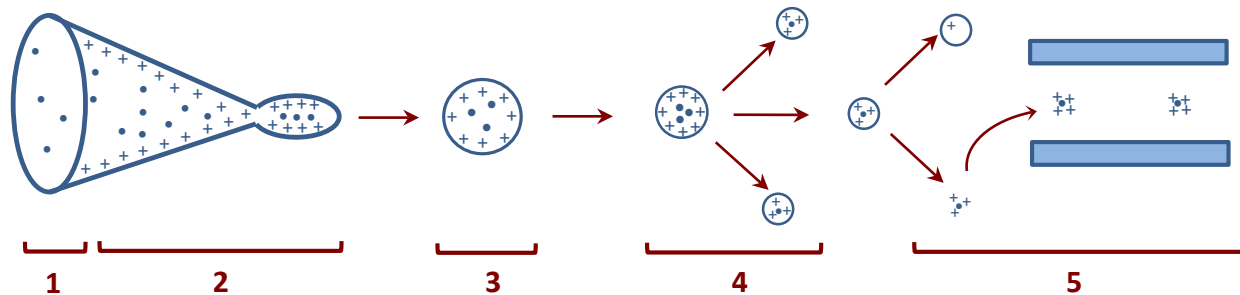
## ➤ Ventajas:

- analitos no polares,
- no hay supresión de iones,
- fácilmente acoplables mediante GC,
- librerías de espectros

## ➤ Inconvenientes:

- análisis
  - ✓ compuestos volátiles,
  - ✓ compuesto térmicamente estables,
  - ✓ compuestos de bajo peso molecular,
- ionización dura

## Ionización por electrospray (ESI)



- 1- producción de iones,
- 2- formación de un spray de gotitas cargadas,
- 3- desolvatación,
- 4- «fisión de Coulomb»,
- 5 - generación de iones en estado gaseoso

***El mecanismo de la ionización por electrospray***

## Ionización por electrospray (ESI)

### ➤ Analitos habituales:

- compuestos polares

*p. ej.* péptidos, proteínas, azúcares, nucleótidos

### ➤ Rango de masas:

- <200 kDa

### ➤ Introducción de la muestra:

- LC o disolución

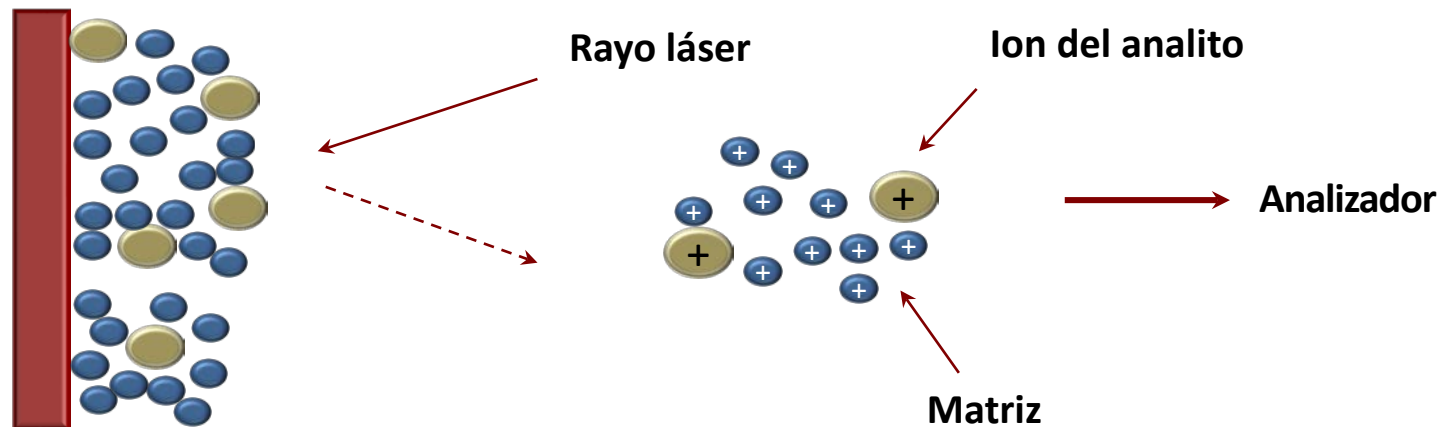
### ➤ Ventajas:

- compuestos termolábiles
- compuestos de alto PM
- iones multicargados
- sensibilidad
- fácil interacción con LC
- método de ionización suave

### ➤ Inconvenientes:

- analitos ionizables
- sensible a sales
- supresión de iones

## Ionización por desorción láser asistida por matriz (MALDI)



*El mecanismo de la ionización por desorción láser asistida por matriz*

## Ionización por desorción láser asistida por matriz (MALDI)

### ➤ Analitos habituales:

- compuestos polares  
*p. ej.* péptidos, proteínas, azúcares, nucleótidos

### ➤ Rango de masas:

- <500 kDa

### ➤ Introducción de la muestra:

- muestra mezclada con una matriz sólida

### ➤ Ventajas:

- compuestos termolábiles,
- compuestos de alto PM,
- sensibilidad,
- menos sensible a sales,
- método de ionización suave

### ➤ Inconvenientes:

- una amplia variedad de matrices,
- dificultades en el análisis cuantitativo,
- supresión de iones



## Ionización química a presión atmosférica (APCI)

### ➤ Analitos habituales:

- compuestos polares

*p. ej.* péptidos, proteínas, azúcares, nucleótidos

### ✓ Rango de masas:

- <1 kDa

### ✓ Introducción de la muestra:

- LC o disolución

### ➤ Ventajas:

- compuestos termoestables,
- sensibilidad,
- permite velocidades de flujo mayores,
- fácil interacción con LC,
- técnica de ionización suave

### ➤ Inconvenientes:

- necesita solubilidad en disolventes polares,
- sensible a sales,
- supresión de iones

## Analizadores de masas

Separan iones en función de su relación masa/carga ( $m/z$ )

- funcionan en alto vacío
- características fundamentales:
  - resolución
  - exactitud de masa
  - sensibilidad
  - rango dinámico

## Resolución

capacidad para diferenciar entre señales estrechamente relacionadas

$$R = \Delta m/m$$

donde el poder de resolución se define como:

$$m_1/(m_2 - m_1)$$

donde  $m_1$  es el ion más ligero y  $(m_2 - m_1)$  es la diferencia entre dos iones consecutivos

## Exactitud de masa

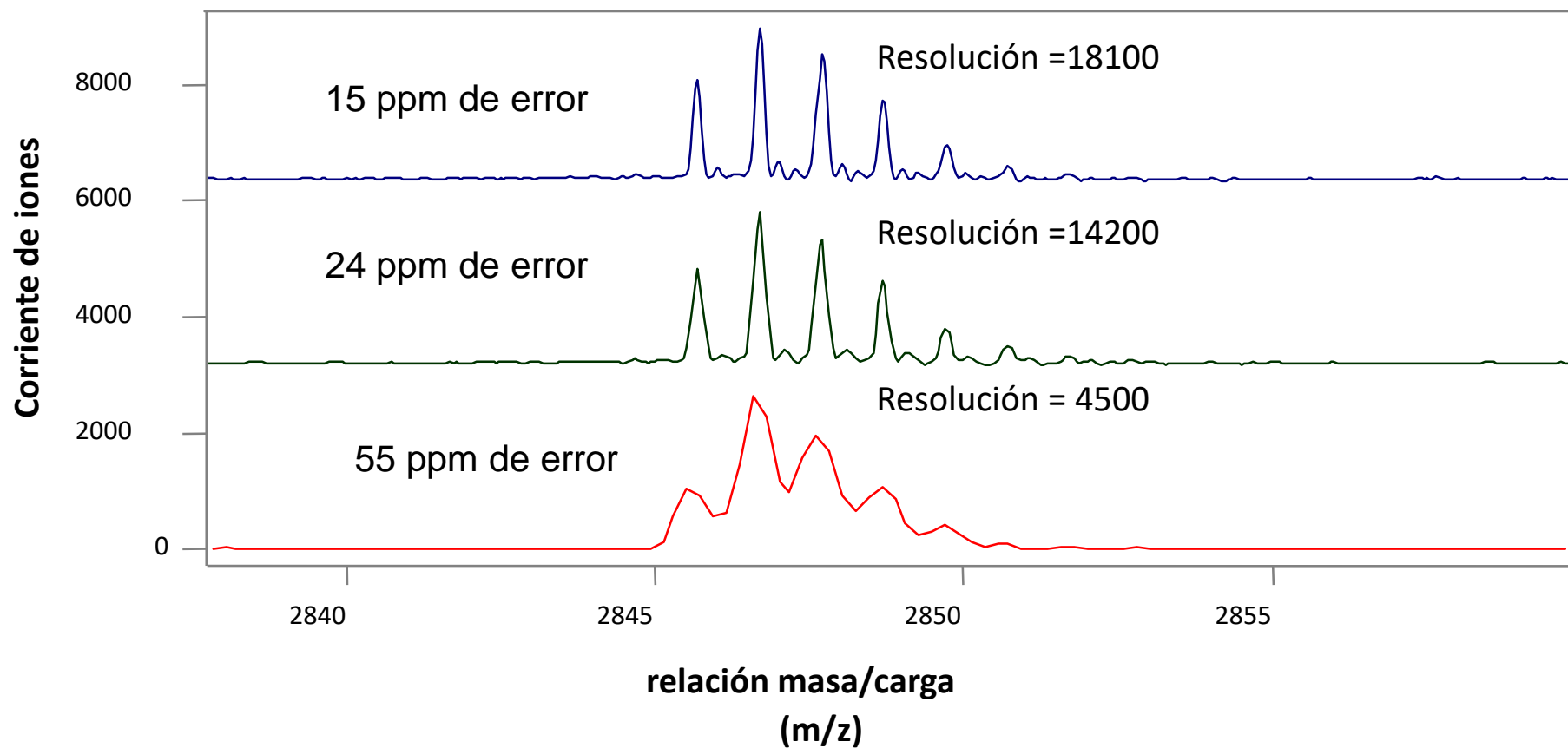
la proximidad de la masa experimental (masa exacta)  
y el valor verdadero (masa exacta)

$$\frac{(masa\ exacta\ monoisotópica - masa\ medida\ exacta)}{masa\ exacta\ monoisotópica \times 10^6}$$

$$masa\ exacta\ monoisotópica \times 10^6$$

- se calcula en [ppm]

**A mayor resolución  
mayor exactitud de masa**



## Sensibilidad

la respuesta del detector en relación con la concentración de un analito que llega al detector  
determina el límite de detección (LOD)

## Rango dinámico

el rango sobre el que la señal de iones es directamente proporcional a la concentración de analitos

- esencial para medidas exactas (análisis de cuantificación)

# Analizadores de masas

analizadores de masas

## Analizadores de masas más utilizados



Cuadropolo (Q)



Trampa de iones (IT)



Tiempo de vuelo (TOF)

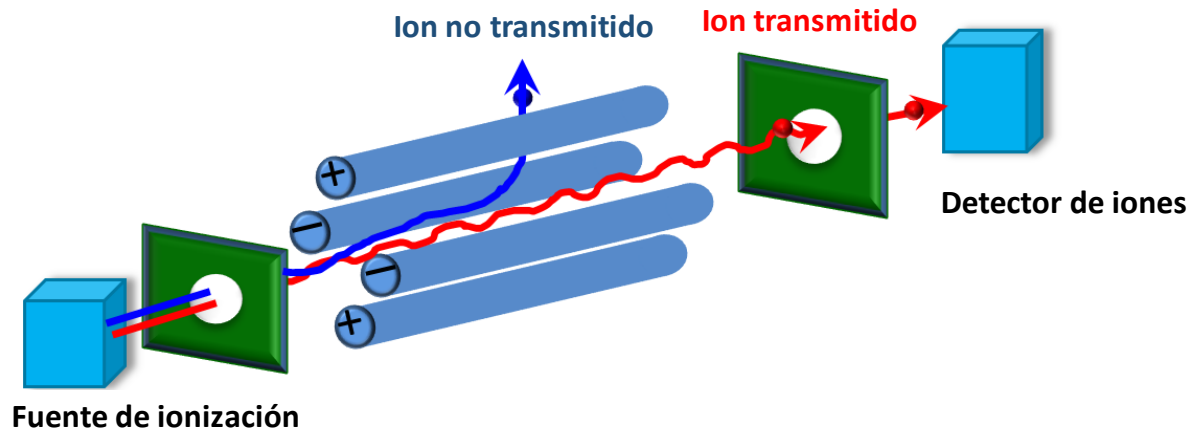


Orbitrap

## Analizadores de masas

### Cuadropolo (Q):

- se compone de cuatro barras paralelas
- utiliza una combinación de voltajes RF y DC para funcionar como filtro de masas
- dispone de modalidades de transmisión de iones variables:
  - ✓ escaneo de iones (SCAN),
  - ✓ monitorización de iones seleccionado (SIM)
- **baja resolución**
- **máxima sensibilidad (análisis cuantitativo)**



*Esquema de un analizador de masas de cuadrupolo*

## Analizadores de masas

### Cuadropolo (Q):

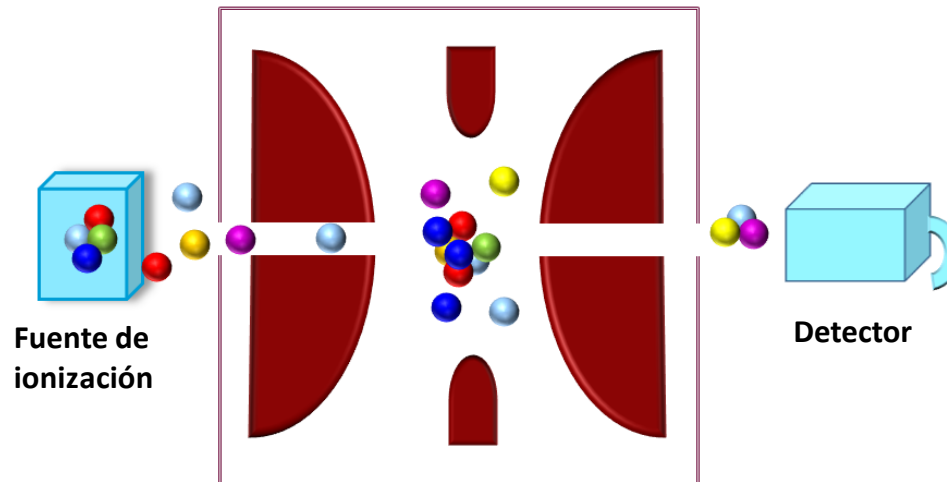
Caracterización	Cuadropolo
Velocidad de adquisición (Hz)	<b>2-10</b>
Exactitud de masa (ppm)	<b>baja</b>
Rango de masas (m/z)	<b>&lt;3000</b>
Resolución	<b>unitaria</b>



# Analizadores de masas

## Trampa de iones (IT):

- trampas de iones que utilizan campos cuadrupolares
- dos tipos:
  - ✓ trampa de iones 2D (trampa de iones lineal)
  - ✓ trampa de iones 3D (trampa de iones cuadrupolo)
- **baja resolución**
- **alta velocidad de escaneo**



*Esquema de un analizador de masas de trampa de iones 3D*

## Analizadores de masas

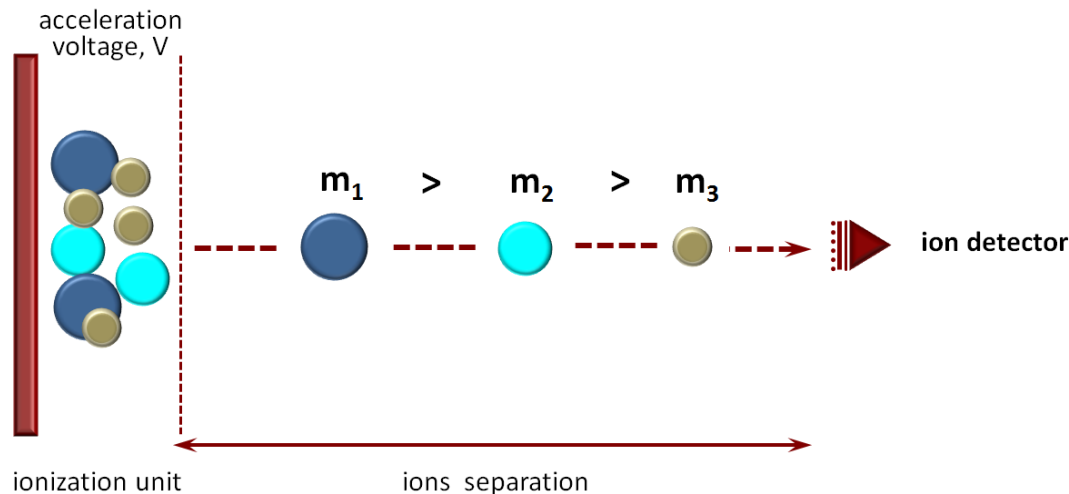
### Trampa de iones (IT)

Caracterización	Trampa de iones
Velocidad de adquisición (Hz)	<b>2-10</b>
Exactitud de masa (ppm)	<b>baja</b>
Rango de masas (m/z)	<b>&lt;6000</b>
Resolución	<b>unitaria</b>

# Analizadores de masas

## Tiempo de vuelo (TOF)

- los iones se forman en pulsos
- mide el tiempo que los iones tardan en alcanzar el detector
- los iones pequeños alcanzan el detector antes que los grandes
- **alta resolución**
- **alta exactitud de masa**
- **alta sensibilidad**



*El mecanismo de separación de iones en el analizador de masas de TOF*

## Analizadores de masas

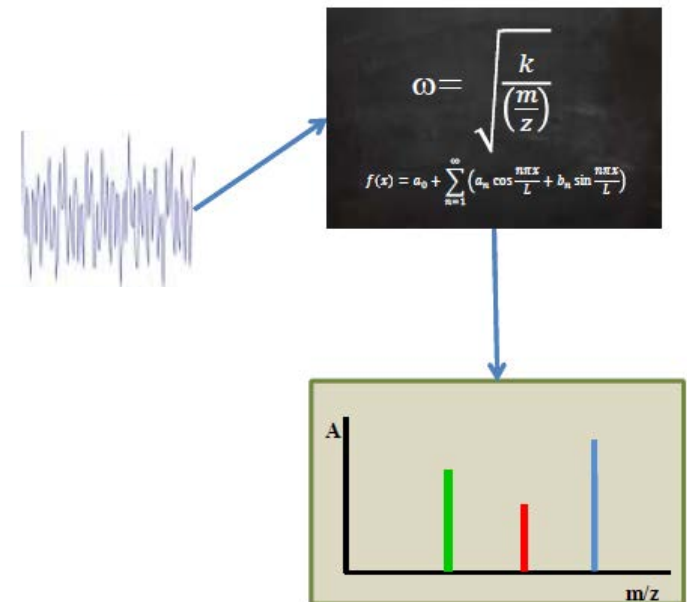
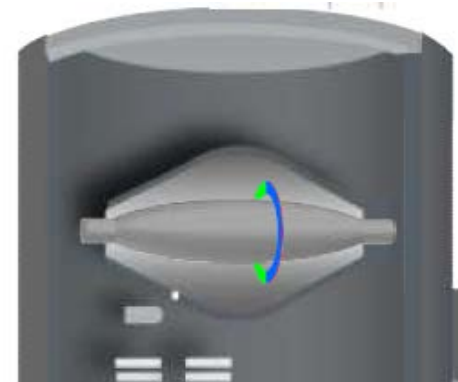
### Tiempo de vuelo (TOF)

Caracterización	TOF
Velocidad de adquisición (Hz)	<b>10-100</b>
Exactitud de masa (ppm)	<b>1-10 ppm</b>
Rango de masas (m/z)	<b>&lt;100.000 ilimitado</b>
Resolución	<b>&lt;50.000</b>

## Analizadores de masas

### Orbitrap:

- consiste en un electrodo con forma de barril
- los valores  $m/z$  se calculan rápidamente mediante la transformada de Fourier a partir de las frecuencias de oscilación de los iones atrapados
- **alta resolución**
- **alta exactitud de masa**
- **alta sensibilidad**



## Analizadores de masas

### Orbitrap

Caracterización	Orbitrap
Velocidad de adquisición (Hz)	<b>1-18</b>
Exactitud de masa (ppm)	<b>1-5 ppm</b>
Rango de masas (m/z)	<b>&lt;6.000</b>
Resolución	<b>&lt;500.000</b>

## Comparación de los distintos analizadores de MS

Analizador	Q	IT	TOF	Orbitrap
<b>Ventajas</b>	fácil interacción con varias técnicas de ionización, mayor rango dinámico, bajo coste	fácil interacción con varias técnicas de ionización, MS <sup>n</sup> bajo coste	escaneo rápido, alto rango de masas, alta exactitud de masa,	alta exactitud de masa, cambio rápido de polaridad
<b>Inconvenientes</b>	baja resolución, baja exactitud de masa, bajo rango de masas, baja velocidad de escaneo, MS/MS requiere múltiples analizadores	baja resolución, baja exactitud de masa, bajo rango de masas, baja velocidad de escaneo	rango dinámico menor que Q, alto coste	velocidad de escaneo menor que QTOF, rango dinámico menor que Q, alto coste

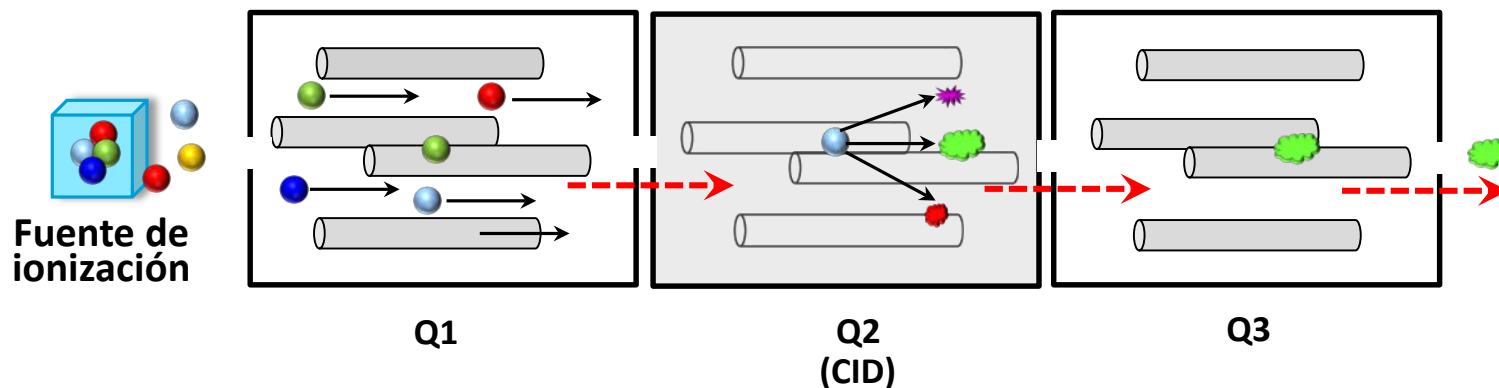
## Espectrometría de masas en tándem (MS/MS)

### Espectrómetros de masas en tándem

Triple cuadrupolo	[QqQ]
Cuadrupolo tiempo de vuelo	[QqTOF]
Trampa de iones	[IT]
Cuadrupolo Orbitrap	[QqOrbitrap]



## Espectrometría de masas en tándem



*La espectrometría de masas en tándem basada en el funcionamiento de un triple cuadrupolo [QqQ]*

### Análisis MS/MS:

- monitorización de reacción seleccionada o de reacción múltiple (SRM/MRM)
- escaneo del ion producto
- escaneo del ion precursor
- escaneo de pérdida neutra

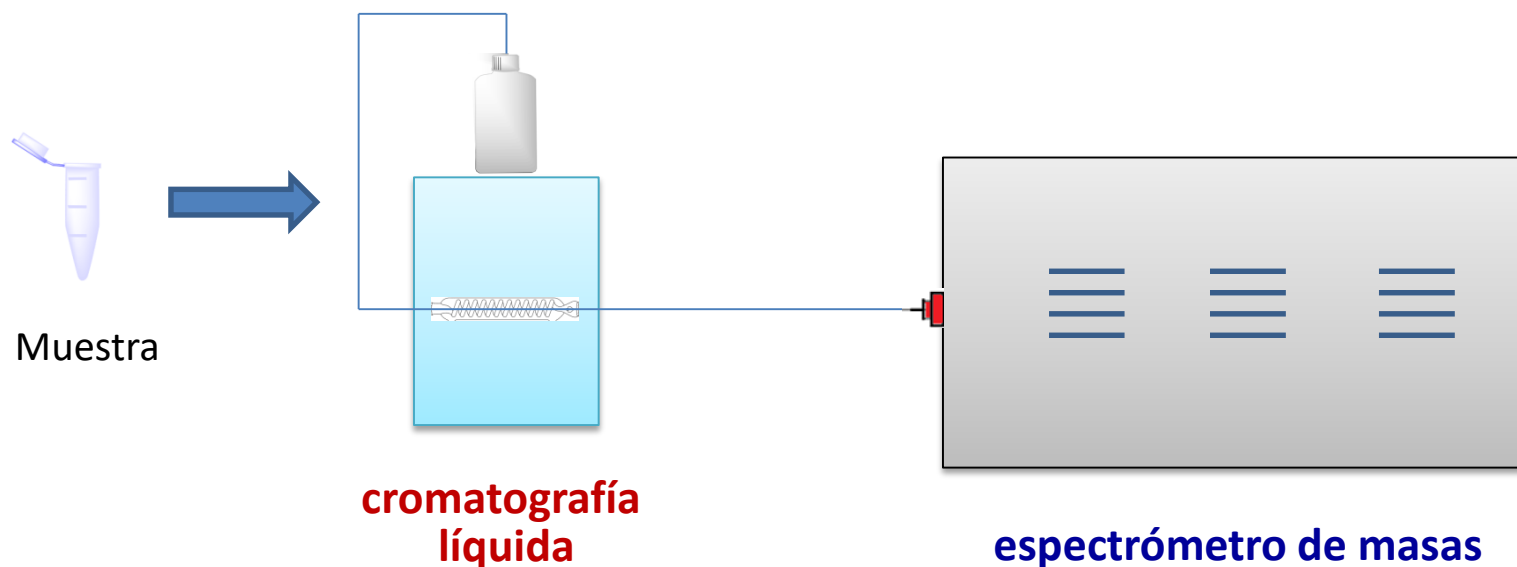
# Técnicas de separación acopladas a la espectrometría de masas

**La espectrometría de masas se combina normalmente con**

- cromatografía líquida [LC]
- cromatografía de gases [GC]
- electroforesis capilar [CE]

## Introducción a la LC-MS

Combinación de la capacidad de separación física de la cromatografía líquida (LC) y la capacidad de análisis de masas de la espectrometría de masas (MS)



La LC permite la separación de muchos compuestos en función de su tiempo de retención ( $t_R$ )

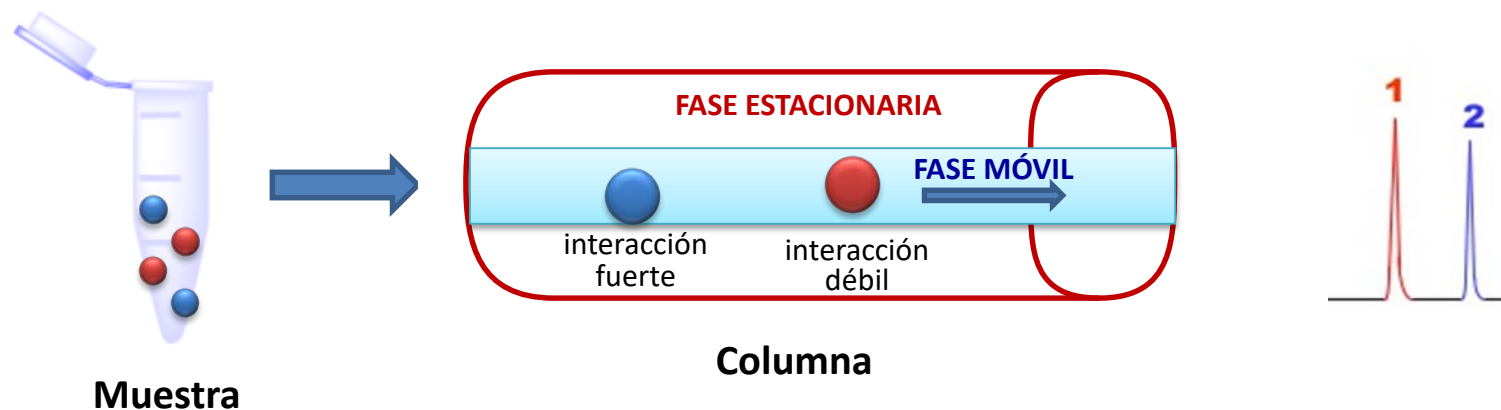
La LC-MS permite diferenciar muchos compuestos de  $t_R$  similar, pero con distinto  $m/z$  o patrón de fragmentación

# Introducción a la cromatografía líquida [LC]

LC - cromatografía en la que la fase móvil es un líquido («eluyente»)

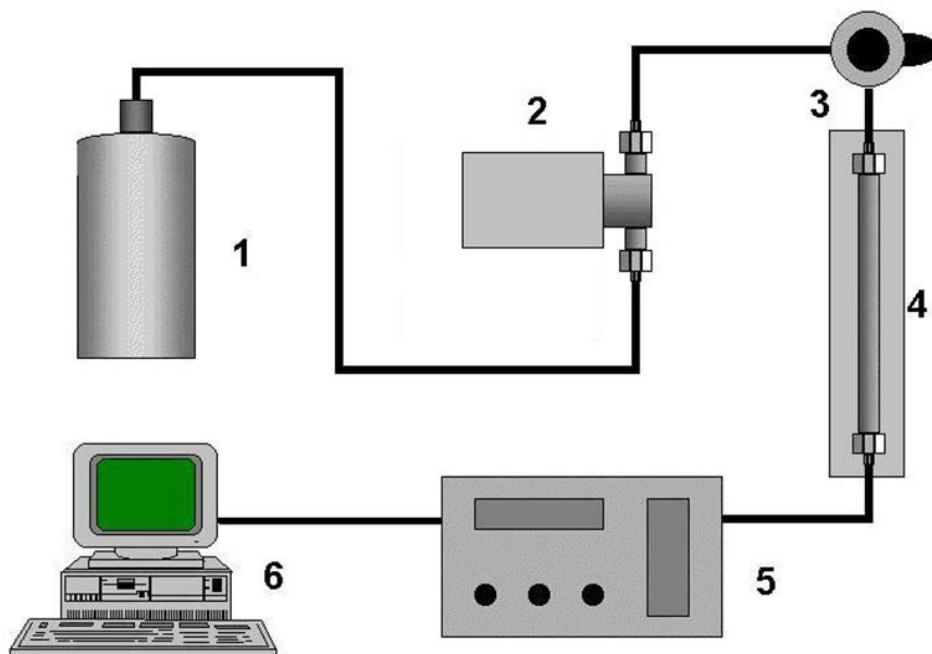
## Mecanismo de separación

Debido a la distinta interacción entre fase estacionaria/móvil y la polaridad de los compuestos de la muestra, sus moléculas se mueven a distinta velocidad y se eluyen de la columna en momentos distintos.



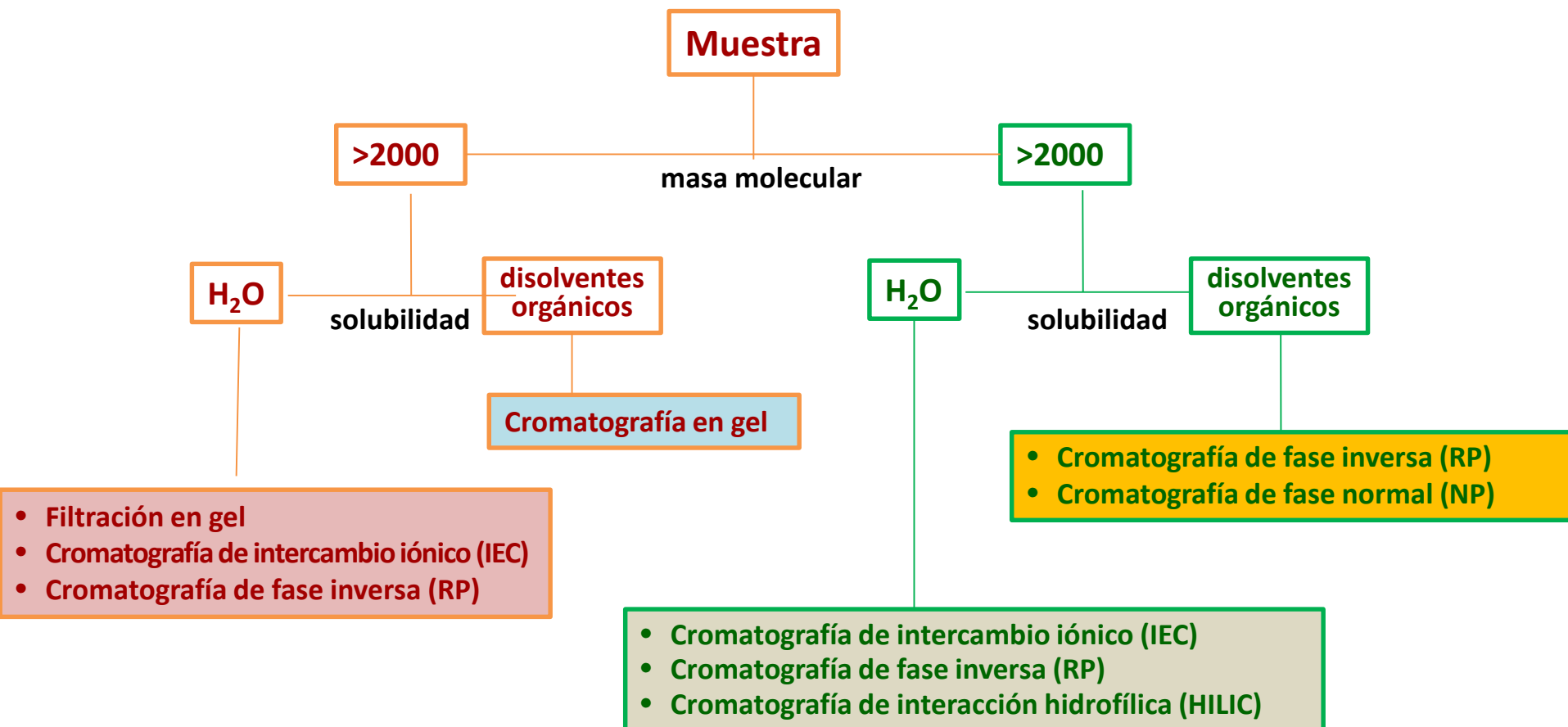
## Cromatografía líquida (LC)

Un sistema cromatográfico habitual contiene los siguientes componentes principales:



- 1 – contenedor de fase móvil,
- 2--bomba,
- 3-- inyector,
- 4 – columna cromatográfica,
- 5 - detector,
- 6 – ordenador

# Selección de la técnica



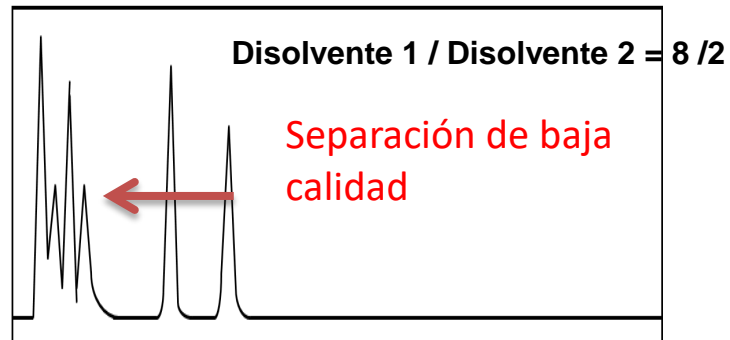
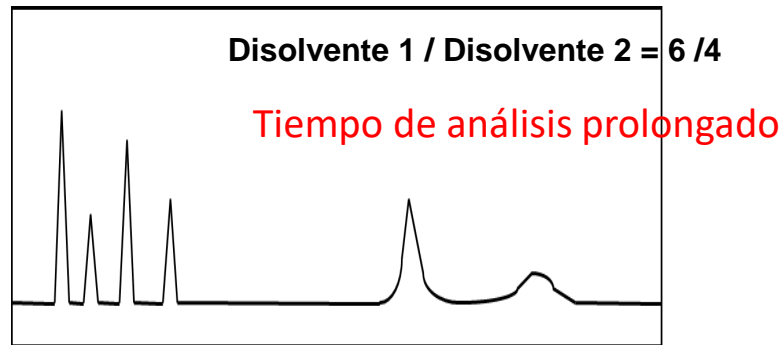
# Bomba de LC

La tarea principal de la bomba es proporcionar un flujo estable, que varía dependiendo de la interfaz que se utilice en la LC-MS

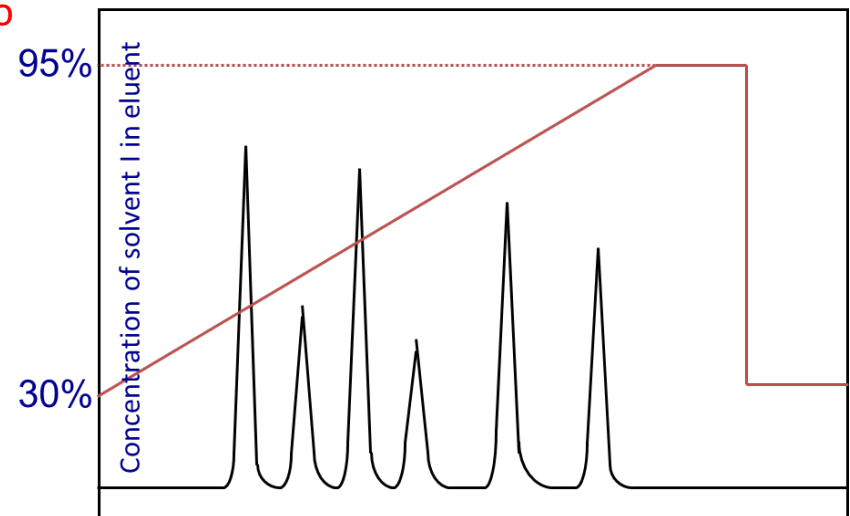
y  
los parámetros de la columna cromatográfica.

**La bomba proporciona fase móvil al sistema cromatográfico en:**

- **modo isocrático**  
(composición eluyente constante)



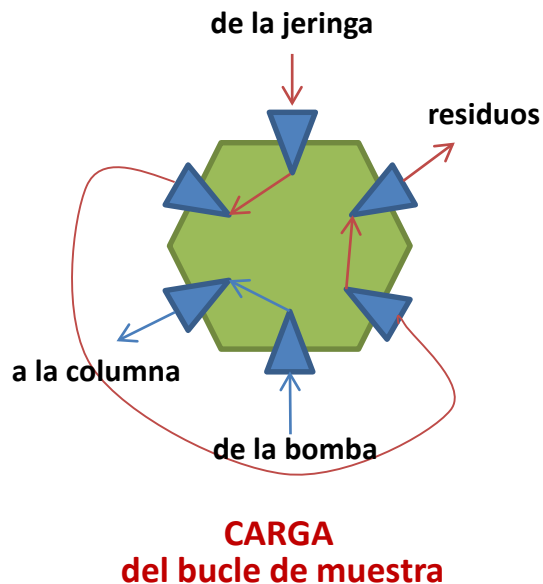
- **modo gradiente**  
(composición eluyente variable)



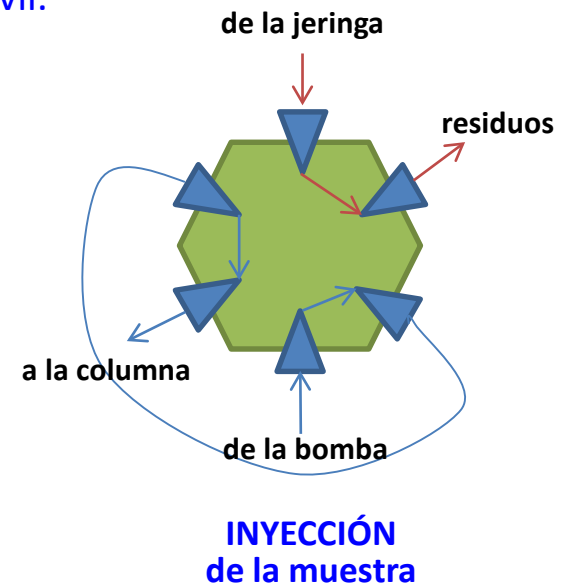
# Inyector de LC

En la LC se utiliza casi exclusivamente un inyector conocido como el «inyector de bucle» (o inyector de válvula de seis puertos).

1. Se introduce la muestra con una microjeringa en una fase móvil que llena un bucle de un volumen nominal.



2. Cuando el bucle esté lleno, se bombea la fase móvil a través de la válvula a la columna para mantener la columna en equilibrio con la fase móvil.



El inyector debe operar con

- gran reproducibilidad,
- exactitud,
- evitando la presencia de burbujas de aire y pulsos.



## Columna de LC

El formato de una columna de LC incluye la información:

- la longitud de la columna,
- el diámetro de la columna
- el tamaño de partícula de la fase estacionaria

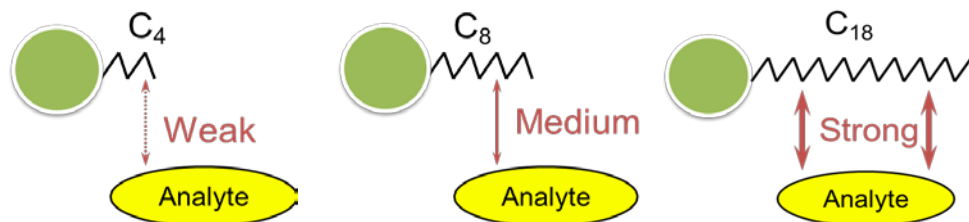
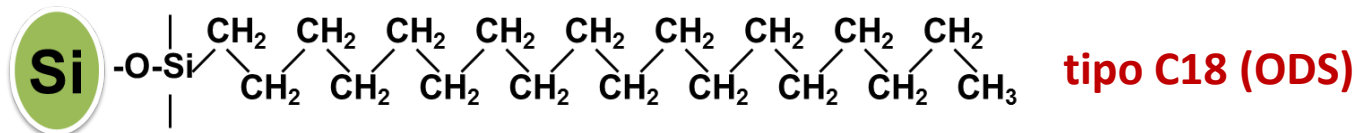
	Tamaño de la columna [mm]	Diámetro de la columna [mm]	Diámetro de la partícula [um]	Velocidad de flujo óptima [μl/min]	Presión [bares]
<b>Columna calibre nano</b>	50-1000	0,05-0,1	1-3	0,3	<300
<b>Columna Capilar</b>	50-1000	0,3	1-5	5	<500
<b>Columna calibre micro</b>	50-1000	0,5-1	1-5	10-50	<800
<b>Columna de calibre estrecho (pequeño)</b>	50-250	2,1	2-5	400	<1200
<b>Columna de calibre normal</b>	30-250	4,6	2-5	1000	<400

# Columna de LC



## Materiales de empaque de la columna LC

- **tipo C18 (ODS)**
- tipo C8 (octil)
- tipo C4 (butil)
- tipo fenil
- tipo TMS
- tipo de ciano



*interacción entre el analito y la columna*

## Fases móviles y estacionarias

La naturaleza del analito/de los compuestos que van a ser separados determina la selección de **la fase estacionaria** y **la fase móvil**.

Modo LC	Fase móvil	Fase estacionaria	Tipo de compuestos separados
NP	orgánicos: diclorometano, acetato de etilo	sílice, amino, ciano, diol	compuestos orgánicos no solubles en agua
RP	agua/disolvente orgánico con o sin aditivos	C18, C8, C4, ciano, amino	neutros, ácidos débiles, bases débiles
HILIC	acetonitrilo con agua, aditivos iónicos	sílice polar, pura	compuestos polares
IEC	soluciones tampón acuosas	anión o catión, resina de intercambio	iónicos, iónicos inorgánicos

# Cromatografía de fase normal (NP)

La separación cromatográfica en NP es el resultado de las interacciones de compuestos separados con **fase estacionaria polar** y **fase móvil no polar**

## Fase estacionaria utilizada en NP:

- gel de sílice:  $-\text{Si}-\text{OH}$
- tipo de ciano:  $-\text{Si}-\text{CH}_2\text{CH}_2\text{CH}_2 \text{CN}$
- tipo de amino:  $-\text{Si}-\text{CH}_2\text{CH}_2\text{CH}_2 \text{NH}_2$
- tipo de diol:  $-\text{Si}-\text{CH}_2\text{CH}_2\text{CH}_2\text{OCH}(\text{OH})-\text{CH}_2 \text{OH}$

## Fase móvil usada en NP:

- hidrocarburos
- diclorometano
- acetato de etilo
- otro disolvente inmiscible con agua



# Cromatografía de fase inversa (RP)

**La cromatografía RP es el método más común de los utilizados en HPLC**

La **separación cromatográfica en RP** es el resultado de las interacciones de compuestos separados con **fase estacionaria no polar** y **fase móvil polar**

## Fase estacionaria utilizada en RP:

(hidrocarburos de cadena larga enlazados covalentemente a la superficie de la sílice)

- C18
- C8
- C4
- ciano
- amino

## La fase móvil en RP:

agua o solución tampón, y disolventes orgánicos, entre los cuales, los más utilizados son:

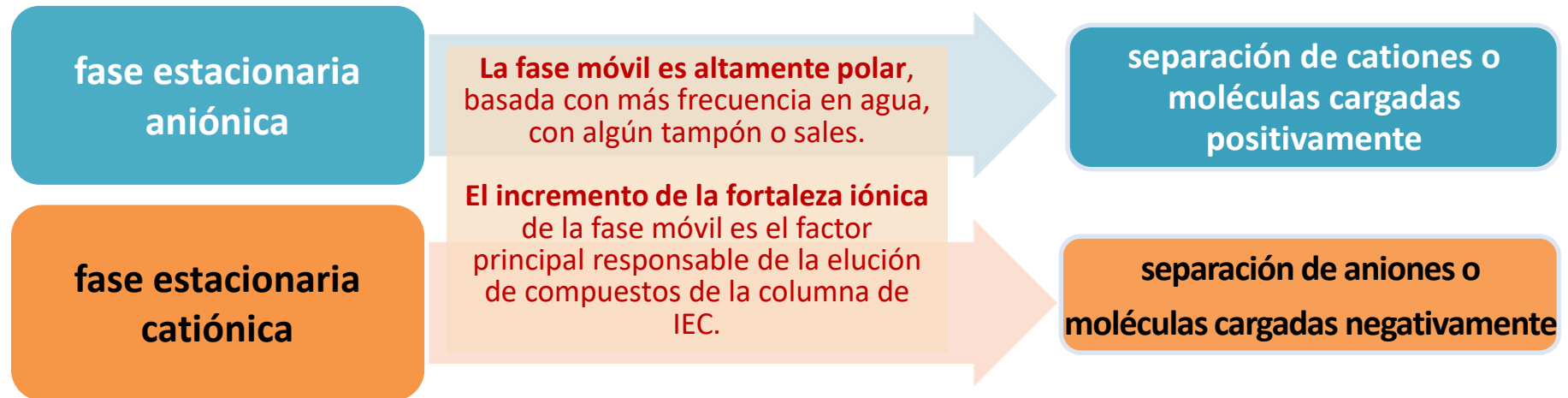
- metanol
- acetonitrilo

**Tanto la NP-LC como la RP-LC se utilizan con distintas finalidades en análisis de muestras biológicas:**

- La NP-LC se aplica para separar clases individuales de lípidos **basándose en los grupos de cabeza polar**,
- La RP-LC se usa para la separación de especies de lípidos **basándose en sus distintas hidrofobicidades** (cadenas de acilos grasos)
- La cromatografía RP en gradiente también resulta útil para la determinación de perfiles de metabolitos en los estudios metabólicos

# Cromatografía de intercambio iónico (IEC)

La separación cromatográfica en la fase de intercambio iónico es resultado de las interacciones de **compuestos iónicos e ionizables** con **grupos funcionales iónicos de la fase estacionaria**, normalmente con cargas opuestas a las de los analitos



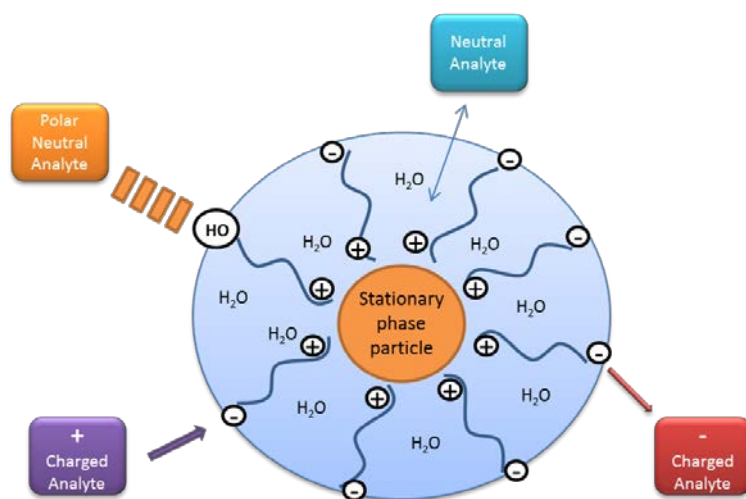
La IEC resulta útil para las separaciones de **biomoléculas grandes y pequeñas**, como aminoácidos, ácidos carboxílicos o aminos.

Debido a un fenómeno de supresión iónica, resulta relativamente difícil acoplar directamente la IEC, que utiliza una alta fortaleza iónica en la fase móvil, a una espectrometría de masas.

# Cromatografía de interacción hidrofílica (HILIC)

## HILIC para la separación de compuestos polares

El mecanismo de separación HILIC se basa en un sistema de extracción líquido/líquido con una formación de capas de agua sobre la superficie de la fase estacionaria polar y la fase móvil orgánica



Esquema de interacciones entre diferentes tipos de analitos polares y una fase estacionaria en modo HILIC

### fases estacionarias

- hidrofílicas
- Normalmente las columnas de HILIC contienen superficies polares de sílice o se pueden derivatizar a fases ligadas a aminos o amidas

### fase móvil

- sistema disolvente normal para RP, el más utilizado el acetonitrilo, con una pequeña cantidad de agua
- Con frecuencia se añaden a la fase móvil acetato o formiato de amonio para incrementar la polaridad y la fortaleza iónica

El modo HILIC se puede adaptar fácilmente a la MS

El uso de disolventes orgánicos aumenta la sensibilidad de la MS debido a una disminución de la supresión iónica

# Detectores

## Detectores de LC:

- UV,
- de fluorescencia,
- electroquímicos,
- de conductividad,
- de índice de refracción,
- **detectores de MS**

PARÁMETROS DE LOS DETECTORES

selectividad

- capacidad de determinar un analito de interés sin interferencias derivadas de la matriz, los disolventes u otras sustancias presentes en el sistema

límite de detección

- la concentración más pequeña de un analito que resulta suficiente para ser detectada con una probabilidad fija

sensibilidad

- la respuesta del detector que esté relacionada con la concentración de un analito que llega al detector.

**Espectrómetros de masa - detectores ideales tanto para análisis cualitativos como cuantitativos**



# Introducción a la GC-MS

La **cromatografía de gases (GC)** es una técnica de separación capaz de separar mezclas altamente complejas basadas principalmente en diferencias de **punto de ebullición/presión de vapor y de polaridad**



En **GC**:

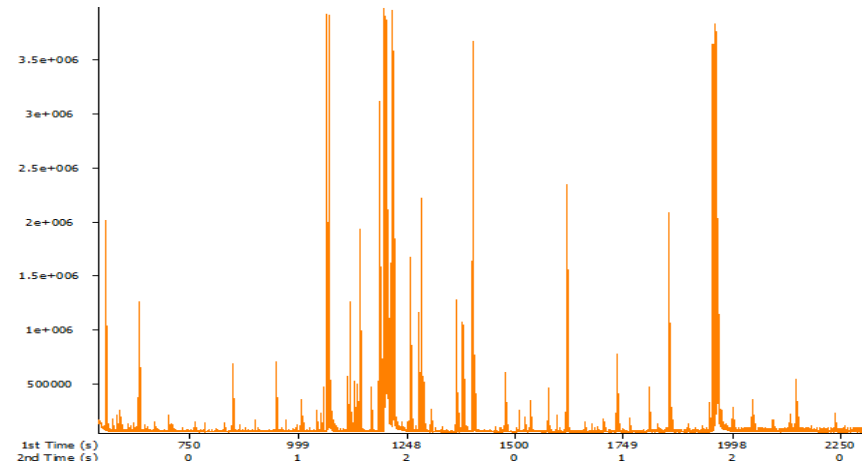
- la **fase móvil** es un gas (Ar, He, N<sub>2</sub> o H<sub>2</sub>)
- la **fase estacionaria** puede ser:
  - un **sólido (adsorbente)**- Cromatografía Gas Sólido (GSC) o bien
  - un **líquido polimérico inmovilizado** - Cromatografía Gas Líquido (GLC)

## La separación en GC depende de

La transferencia de una sustancia (como vapor) empleando un gas portador (fase móvil) a través de una columna.

En GC, el índice y el grado en el que se produce la partición de los compuestos depende de:

- la **afinidad química del analito para la fase estacionaria**
- la **presión de vapor del analito**— que depende de la temperatura de la columna



# Estrategias para la determinación por GC

Recogida de  
muestras

gaseosas  
líquidas  
sólidas

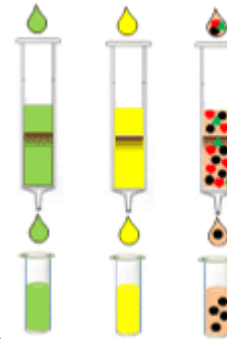
**Pretratamiento**

secado  
filtración  
homogeneización  
...



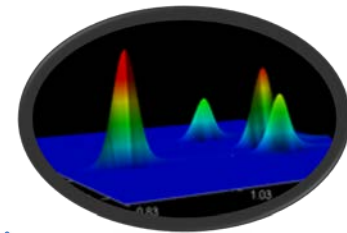
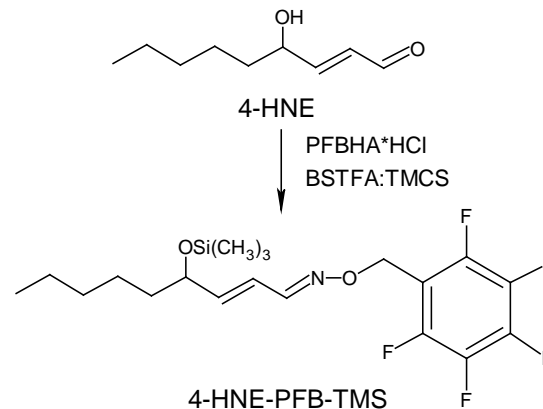
**Preparación  
de la muestra**

SPE  
SPME  
...



**Derivatización**

sililación  
alquilación  
...



**Análisis  
GC, GCxGC**

FID  
MS  
...

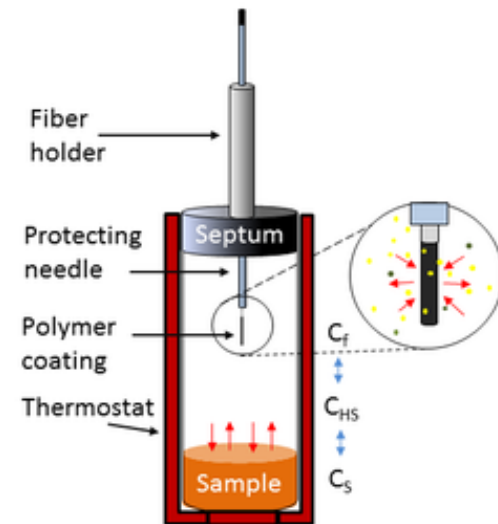
## Recogida de muestras

La muestra a analizar por **GC** puede ser:

- un gas
- un líquido
- moléculas absorbidas en una superficie después de la microextracción en fase sólida (SPME)

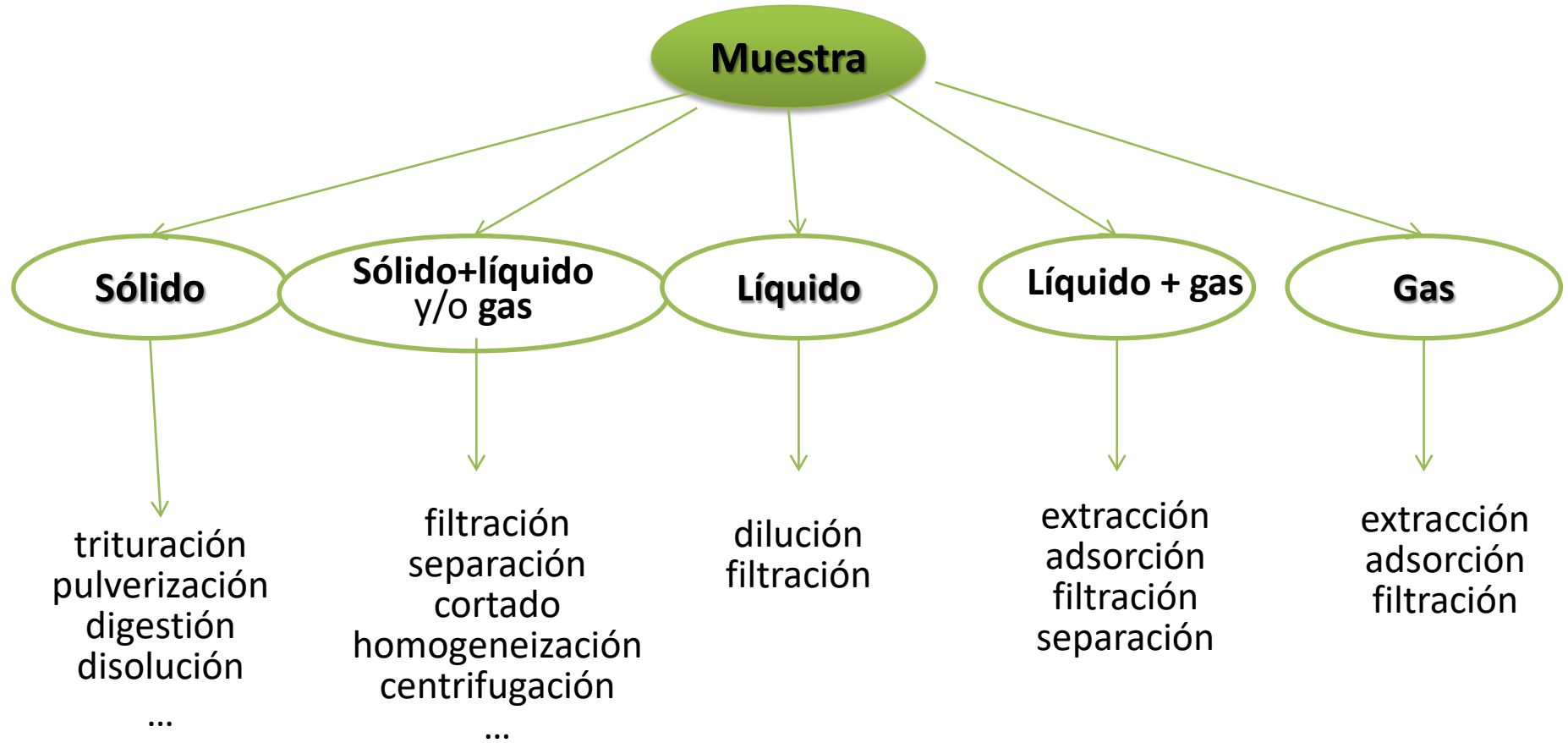
**La estabilización de la composición y las propiedades de las muestras** se obtiene normalmente añadiendo los compuestos apropiados, que:

- inhiben la actividad biológica de organismos presentes en las muestras
- eliminan la incidencia de absorción de componentes de la muestra en paredes vasculares
- eliminan la volatilización, descomposición térmica, reacciones químicas, etc.



Esquema de la SPME

# Pretratamiento de la muestra



# Preparación de muestras



Equipo de SPE

## Técnicas de extracción de muestras líquidas

Extracción líquido-gas

Análisis de espacio de cabeza estático (SHSA)

Análisis de espacio de cabeza dinámico (SHSA)

Purga y tampa (PT)

Extracción líquido-líquido

Extracción clásica líquido-líquido (LLE)

Extracción continua líquido-líquido (LLCE)

Microextracción (ME)

Microextracción en gota (SDME)

Microextracción líquido-líquido asistida por membrana (SM-LLME)

Extracción gas-líquido

Extracción a la fase líquida

Extracción utilizando las membranas semipermeables (SPMD)

Extracción líquido-fase sólida

Extracción de fase sólida (SPE)

Microextracción de fase sólida (SPME)

Extracción con disco rotatorio (RDE)

Microextracción en fase sólida con membrana (M-SPME)

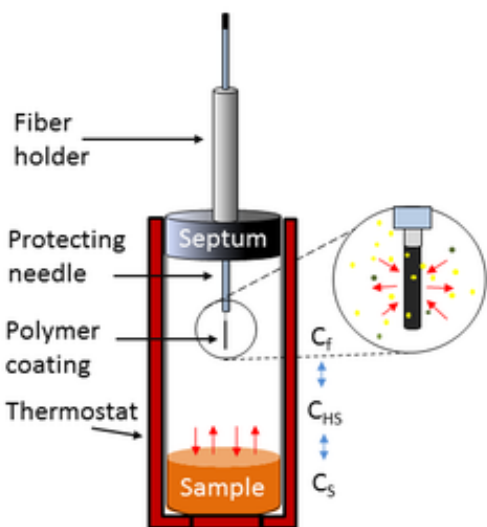
Extracción en fase sólida de impresión molecular (MI-SPE)

Extracción en fase sólida dispersiva (dSPE)

Extracción de membrana polimérica (PME)

## Técnicas de extracción de muestras gaseosas

Extracción gas-sólido



Esquema de SPME

# Preparación de muestras



Técnicas de extracción de muestras sólidas

Extracción sólido-gas

Análisis de espacio de cabeza estático (SHSA)

Análisis de espacio de cabeza dinámico (SHSA)

Desorción térmica (TD)

Extracción con agitación

Extracción en una corriente de disolvente

Extracción en Soxhlet

Homogeneización de la muestra

Extracción en varias etapas (secuencial)

Saponificación

Extracción asistida por ultrasonido (USE)

Ayudada por métodos clásicos

Extracción asistida por microondas (MAE)

Extracción acelerada de disolventes (ASE)

Extracción de alta presión (HPE)

Extracción de disolventes asistida por microondas (MASE)

Extracción de fluidos supercríticos (SFE)

Métodos clásicos

Técnicas modernas de extracción

Sistema de extracción por microondas



Equipo desorción térmica

Métodos de extracción de muestras sólidas

# Derivatización

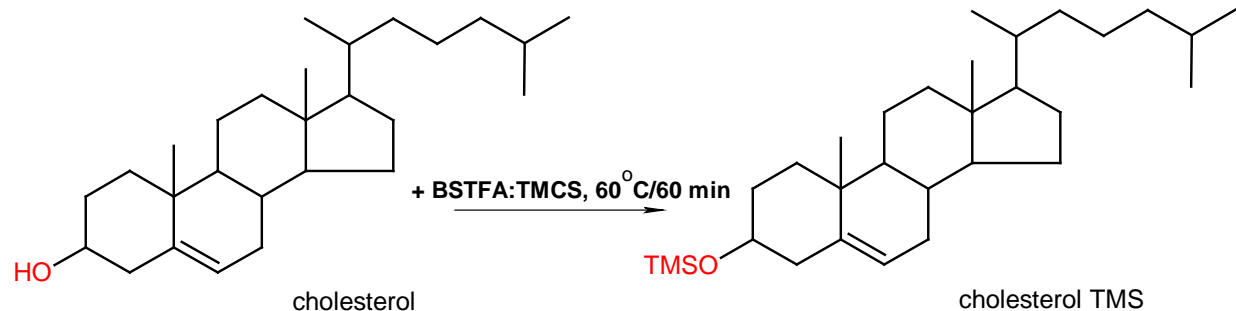
El análisis directo de mezclas de compuestos en GC es complicado debido a:

- la volatilidad demasiado baja de los compuestos analizados
- la estabilidad térmica demasiado baja de los compuestos analizados
- las interacciones entre los compuestos
- las interacciones entre los compuestos y la fase estacionaria de la columna de GC
- una sensibilidad o especificidad del ensayo demasiado bajas

Por tanto, la forma principal de preparar una muestra para el análisis es convertir los analitos en productos derivatizados

La derivatización de analitos se realiza mediante

- sililación
- alquilación
- acetilación



Reacción de sililación

# Derivatización

Procedimiento	Grupo funcional - Tipo de compuesto	Derivado	Reactivo
<b>Sililación</b>	-OH - alcoholes, fenoles -CO -cetonas, esteroides -COOH - aminoácidos , ácidos grasos, esteroides -(CH <sub>2</sub> OH) <sub>n</sub> -azúcares -NH, -NH <sub>2</sub> -aminas, urea -CONH, -CONH <sub>2</sub> - imidias, proteínas	Trimetilsilil éteres  Trimetilsilil amidas	<b>Bistrimetilsililfluoroacetamida (BSTFA)</b> N- metil-N-t-butildimetilsilil- trifluoroacetamida (MTBSTFA) N-metiltrimetilsililtrifluoroacetamida (MSTFA) <b>Trimetilsililimidazol (TMSI)</b> Reactivos halo-metilsililos
<b>Alquilación</b>	-OH - alcoholes, fenoles -CO -aldehídos -COOH - aminoácidos, ácidos grasos -NH, -NH <sub>2</sub> -aminas, aminoazúcares -CONH -amidias -SH -mercaptanos	Ésteres metílicos (DMF) Trifluoroacetatos (TFAA) Ésteres metílicos (BF <sub>3</sub> -metanol) Éteres de pentafluorobencilo (PFBBr) Metilamidias (TMAH) Ésteres metílicos (DMF)	Bromuro de Bencilo Trifluoruro de boro (BF <sub>3</sub> ) en metanol o butanol Dimetilformamida (DMF) Pentafluorobencil-hidroxilamina clorhidrato (PFBHA) Hidróxido de tetrabutilamonio (TBH) Anhídrido trifluoroacético (TFAA)
<b>Acilación</b>	-OH - alcoholes, fenoles -(CH <sub>2</sub> OH) <sub>n</sub> -azúcares -NH, -NH <sub>2</sub> -aminas -CONH -amidias -SH -mercaptanos	Pentafluoropropiónico (PFPA) Trifluoroacetamidias (TFAI) Trifluoroacetamidias (MBTFA) Trifluoroacetamidias (TFAA) Trimetilsilil éteres (MBTFA)	Anhídrido heptafluorobutírico (HFBA) N-Metil-bis(trifluoroacetamida) (MBTFA) Cloruro de pentafluorobenzoílo (PFBCI) Pentafluoropropanol (PFPOH) Anhídrido trifluoroacético (TFAA)



# Cromatógrafo de gases

## El cromatógrafo de gases se compone de

- **inyector**

se inyecta la muestra en el puerto de inyección caliente, donde se volatiliza y transporta hasta la columna a través del gas portador inerte pureza  $\geq 99,9995\%$  (Ar, He, N<sub>2</sub> o H<sub>2</sub>)

- **Columna de GC**

la muestra se separa dentro de la columna

- **detector**

responde a alguna propiedad fisicoquímica del analito y genera una señal electrónica que mide la cantidad de analitos presentes

**cromatograma** - resultado de la separación cromatográfica

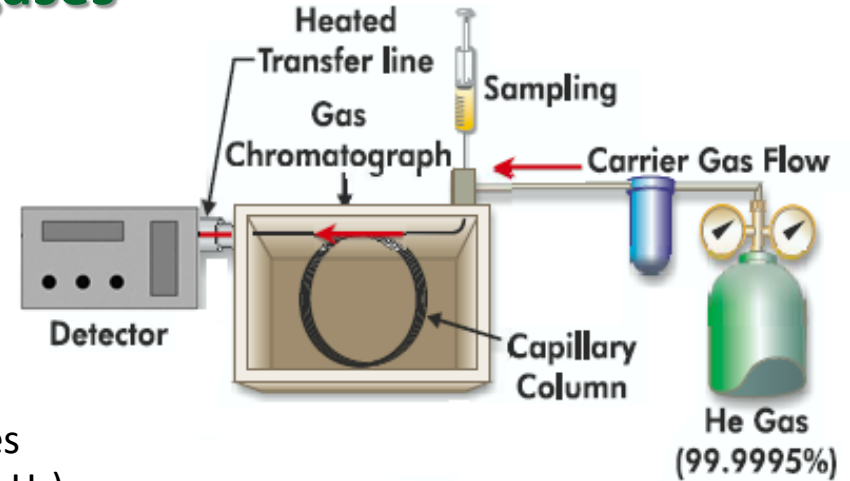
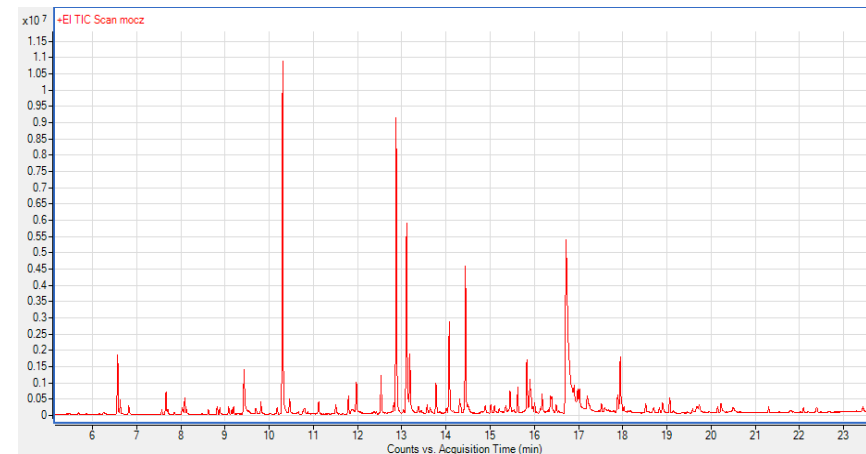


Figura. Diagrama del equipo GC

(modificado de *de.leco-europe.com*)



Cromatograma de GC de componentes del plasma humano

# Inyector de GC

## Etapas de la inyección

1. Se inyecta la muestra en el puerto de inyección caliente
2. En el puerto de inyección, la muestra se volatiliza
3. El gas portador atrapa la muestra volatilizada en la corriente portadora que se introduce en la columna de GC

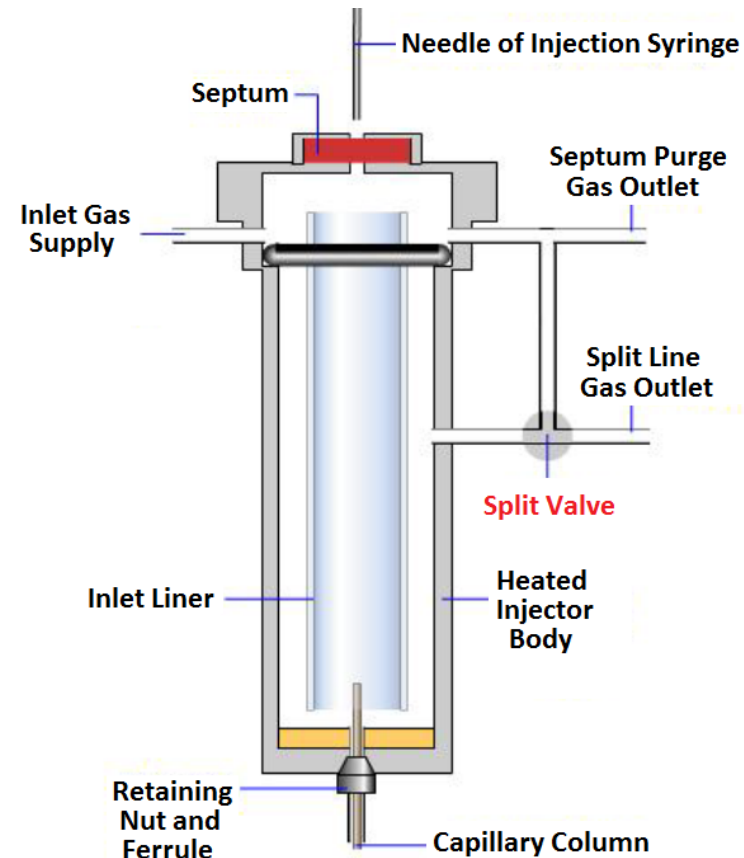
## Tipos de inyectoros GC: **Split/Splitless (SSI)**

### **Inyección split**

solo una pequeña parte de la muestra vaporizada se aplica a la columna (normalmente entre 1/20 y 1/500)  
 - se utiliza principalmente para análisis que no sean de trazas de muestras volátiles

### **Inyección Splitless**

la totalidad de la muestra alcanza la columna  
 - se utiliza principalmente para análisis de trazas y de ultra trazas



## Columna de GC

**Columna empaquetada** (analíticas, microempaquetadas) llenas de partículas sólidas

- adsorbentes de carbono
- sílice
- alúmina
- tamices moleculares
- polímeros sintéticos porosos



Los adsorbentes son fases estacionarias menos habituales debido a:

- la baja reproducibilidad de los resultados
- mayor tiempo de retención
- aparición de «colas» con eficiencias de separación mucho más bajas

**Columna capilar** (capilar, microcapilar)

▪ tubos capilares tubulares abiertos integrados con un líquido

- siliconas
- escualeno
- polietilenglicol

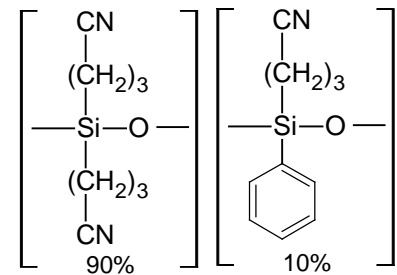
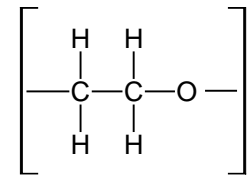
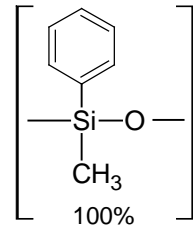
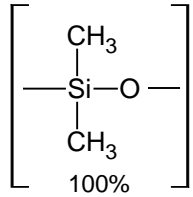
▪ la fase estacionaria líquida debería ser:

- químicamente inerte
- capaz de disolver componentes separados
- altamente selectivas para los componentes de la mezcla
- poco volátiles
- y poseer estabilidad térmica bajo las condiciones de operación de la columna
- especialmente aptas para la separación de componentes gaseosos con una alta eficacia de separación

▪ se utilizan con frecuencia



# Fases de columna



Incremento de la polaridad

Metilpolisiloxano

Metilpolisiloxano + 5% fenil

Metilpolisiloxano + 50% fenil

Metilpolisiloxano + 7% cianopropil + 7% fenil

Metilpolisiloxano + 25% cianopropil + 25% fenil

Polietilenglicol (PEG)

Metilpolisiloxano + 70% cianopropil

Polisiloxano +  $\geq 90\%$  cianopropil

Incremento de la selectividad



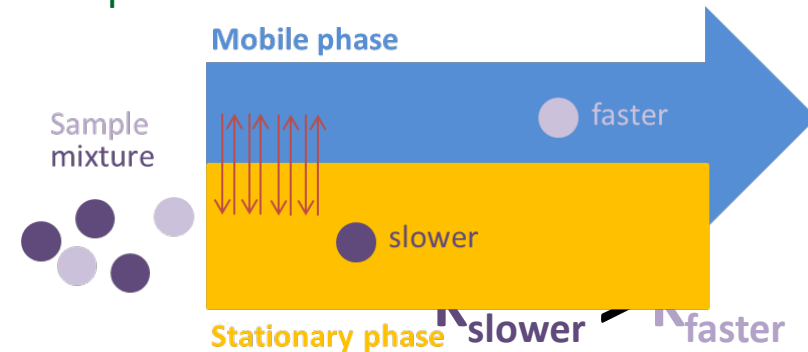
## Mecanismo de separación

El índice y el grado en el que se produce la separación de los compuestos en la columna de GC se corresponde con una función de la distribución de dichos compuestos en dos fases (**móvil y estacionaria**) en equilibrio.

La separación es el resultado de distintas velocidades de migración provocadas por distintos valores del **coeficiente de reparto (Ks)**

La **ecuación de Nernst** puede expresar el **coeficiente de reparto**:

$$K_S = C_L / C_G$$



$C_L$  - la concentración de las sustancias en la fase estacionaria

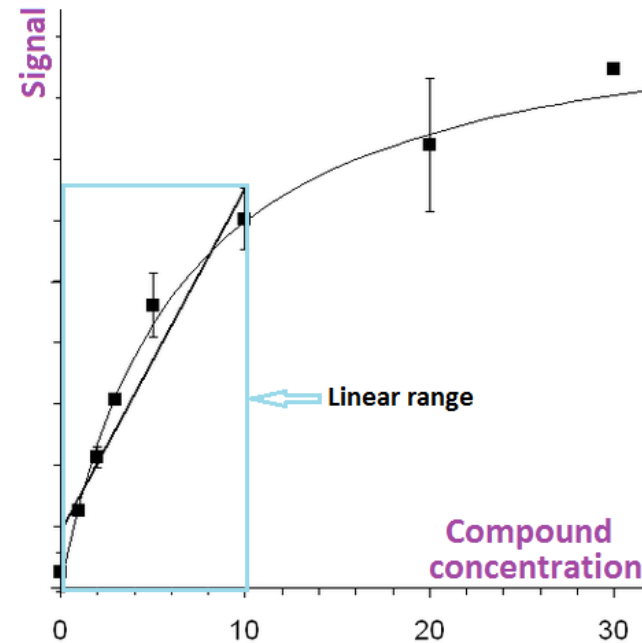
$C_G$  - la concentración de las sustancias en la fase móvil

Cuanto mayor sea la afinidad para el material de la fase estacionaria, mayor será el valor de  $K_s$  y mayor será el valor del tiempo de retención ( $t_R$ ).

# Detectores de GC

## Características del detector de GC ideal

- alta sensibilidad
- buena estabilidad y reproducibilidad
- amplio rango de respuesta lineal a solutos que se extiende a lo largo de varios órdenes de magnitud (propósitos de calibración)
- amplio rango de temperatura
- tiempo de respuesta corto con independencia de la velocidad de flujo
- alta fiabilidad y facilidad de uso
- similitud en respuesta a todos los solutos
- el detector debería ser no destructivo

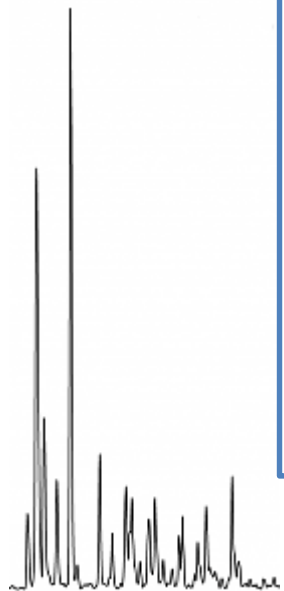
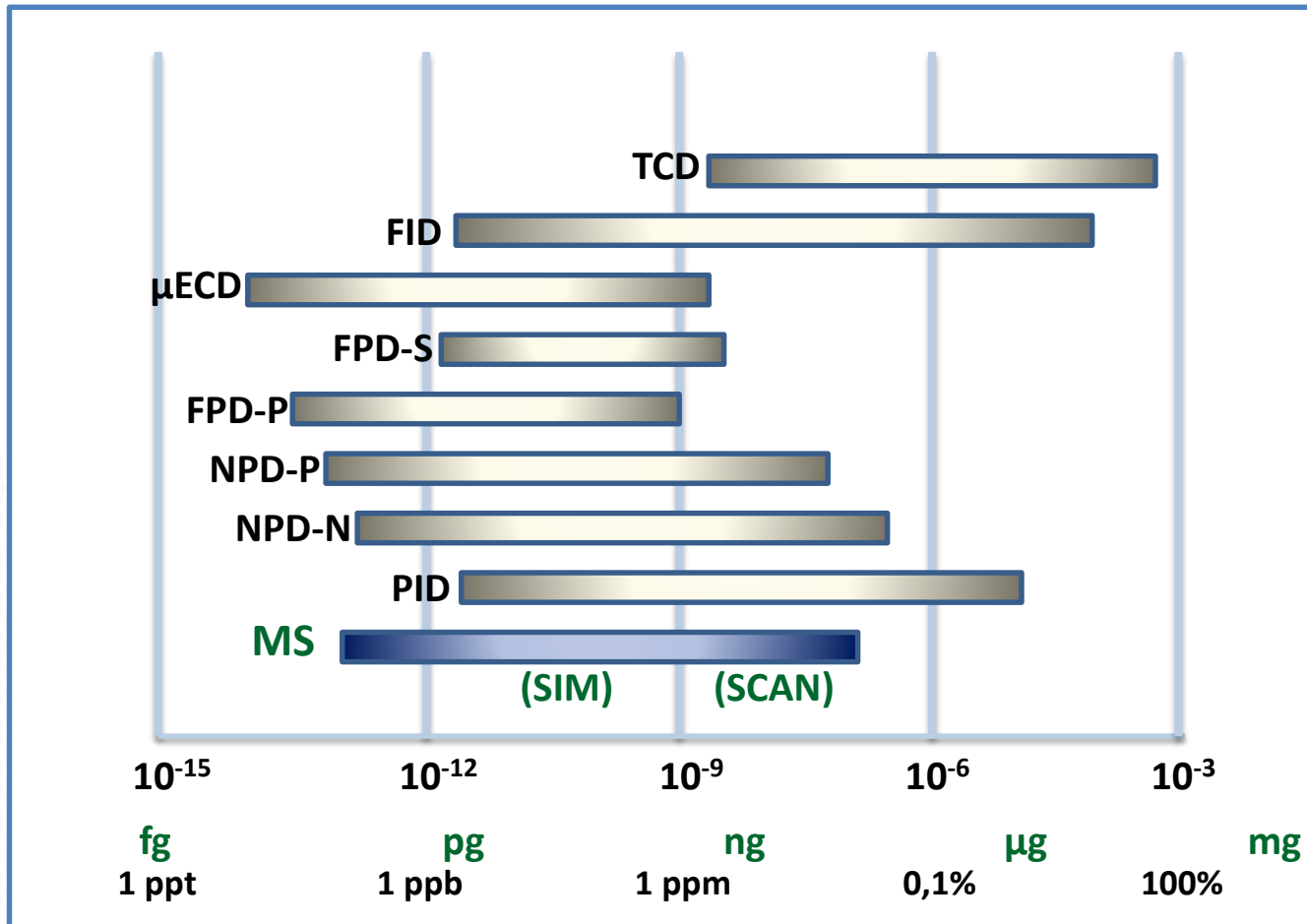


# Detectores de GC

Tipo	Tipo de respuesta	LOD [g/seg]	Rango lineal	Comentarios
Conductividad térmica (TCD)	universal	$10^{-5}$ - $10^{-6}$	$10^3$ - $10^4$	mide cambios en la conducción del calor
Ionización en llama (FID)	universal	$10^{-12}$	$10^6$ - $10^7$	mide corrientes iónicas de la pirólisis
Captura electrónica (ECD)	selectiva	$10^{-14}$	$10^2$ - $10^3$	detector de compuestos que contienen átomos con altas afinidades de electrones
Fotométrico de llama (FPD)	selectiva	$10^{-13}$	$10^2$	detector de compuestos que contienen S, P
Nitrógeno-fósforo (NPD)	selectiva	$10^{-8}$ - $10^{-14}$	$10^5$ - $10^7$	selectiva para compuestos que contienen N, P
Fotoionización (PID)	selectiva	$10^{-8}$ - $10^{-12}$	$10^5$	selectividad debido a la identificación de gas en la lámpara
Infrarrojo con transformada de Fourier (FTIR)	selectiva	$10^{-10}$	depende del grupo funcional	moléculas polares
Espectrómetro de masas (MS)	universal	$10^{-12}$	depende del tipo de analizador de MS, del modo de operación y del compuesto	detector no destructivo, una de las herramientas más exactas y eficientes para analizar muestras orgánicas <b>los detectores más potentes para GC</b>

# Detectores de GC

Comparación de la sensibilidad y el rango dinámico de detectores de GC



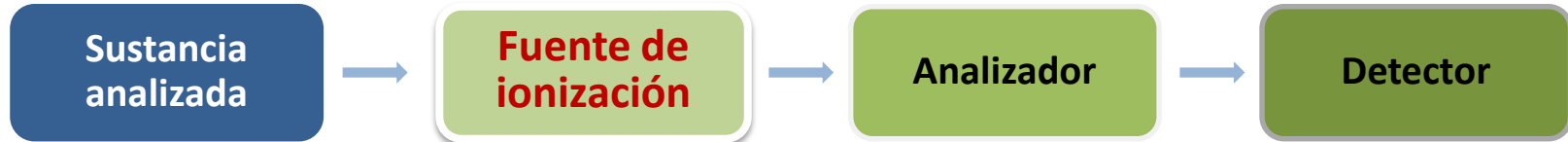


# Cromatografía de gases - espectrometría de masas (GC-MS)

## GC-MS

- herramienta versátil para:
  - separar
  - identificar
  - cuantificar sustancias desconocidas
- es la técnica más efectiva para el análisis de compuestos orgánicos volátiles de matrices complejas en una amplia variedad de concentraciones (desde ppb a ppm)
- se caracteriza por una alta selectividad y sensibilidad y por presentar una **amplia gama de aplicaciones**:
  - aplicaciones médicas y farmacéuticas
  - análisis biológico
  - aplicaciones forenses y criminológicas
  - monitorización medioambiental
  - seguridad y detección de agentes químicos bélicos
  - análisis de alimentos/sabor/fragancia
  - aplicaciones químicas/industriales
  - investigación geoquímica
  - análisis petroquímico

# Instrumentación GC-MS



## Técnicas de ionización

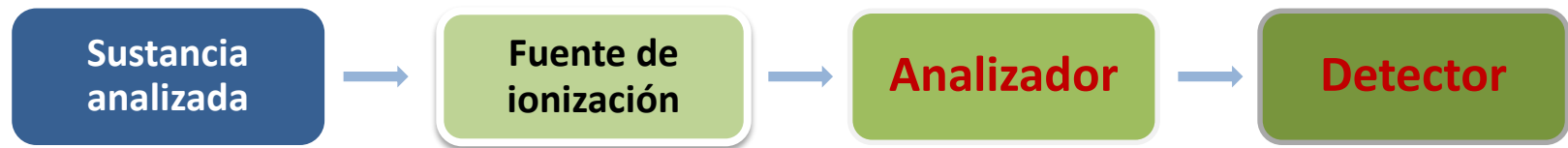
**EI - electro-ionización** - la técnica de ionización más empleada en GC-MS, ionización dura

**CI - ionización química**- relativamente suave (denominada «ionización suave»)

**otras técnicas de ionización utilizadas en GC-MS:**

- fotoionización (PI)
- ionización de campo (FI)
- desorción de campo (DI)
- desorción láser (LD)
- bombardeo con átomos acelerados (FAB)
- desorción por plasma (PD)
- espectrometría de masas de iones secundarios (SIMS)
- ionización por desorción láser asistida por matriz (MALDI)

# Instrumentación GC-MS



**Analizador**- separar iones cargados de acuerdo con su relación  $m/z$

- **Q** - cuadrupolo (simple Q, tripleQ)
- **IT** – trampas de iones (lineales, esféricas)
- **TOF** – «tiempo de vuelo»

▪ **analizadores híbridos de MS:**

- QqQ
- Q/IT
- Q/TOF
- Orbitrap

▪ **otros analizadores utilizados en GC-MS:**

- sector magnético (B)
- sector eléctrico (E)
- sector eléctrico y magnético
- resonancia ciclotrónica de iones (ICR)
- espectrometría de masas de resonancia ión-ciclotrón con transformada de Fourier (FT-ICR MS)

**Detector** - registro de datos

- Colector de Faraday
- multiplicador de electrones
- placas microcanal
- fotomultiplicador

# GC-MS

## Ventajas

- Análisis rápido
- Alta eficiencia que conlleva una alta resolución
- Detectores sensibles (ppb)
- Gran exactitud cuantitativa (<1% RSD)
- No destructiva - (acoplada a MS)
- Requiere muestras pequeñas (<1 mL)
- Técnicas sólidas y fiables
- Respaldada por una bibliografía extensa y múltiples aplicaciones

## Inconvenientes

- Limitada a muestras volátiles o se requiere derivatización
- No es adecuada para muestras térmicamente lábiles que se degraden a temperaturas elevadas - se requiere derivatización
- No es apta para la cromatografía preparativa
- Requiere un detector de MS para la elucidación estructural de analitos (caracterización)
- La mayoría de los detectores que no son de MS son destructivos
- Capacidad de picos muy limitada en el análisis de muestras muy complejas – **se requiere separación por GCxGC**

# Cromatografía de gases bidimensional (GC×GC; 2D GC)

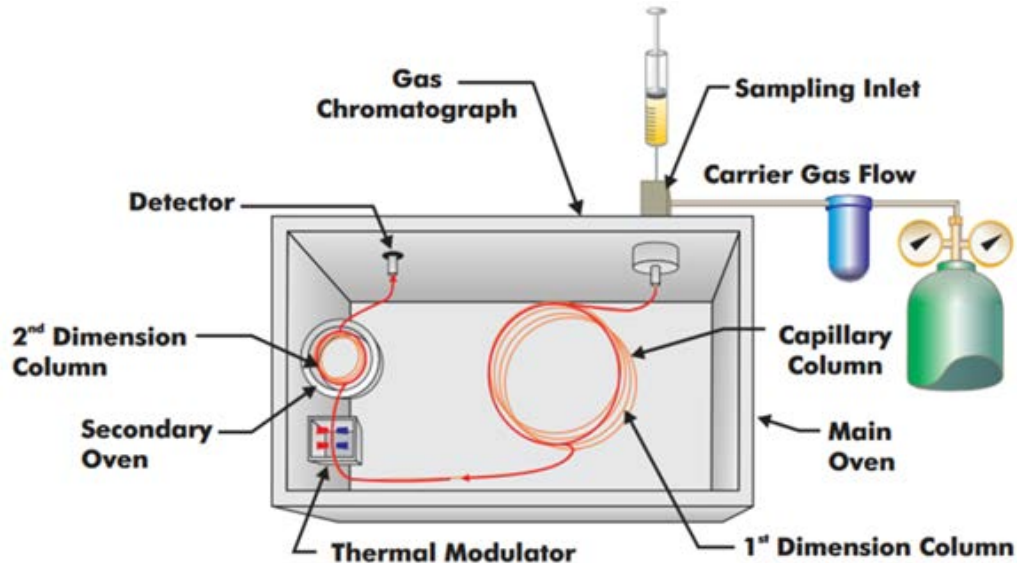
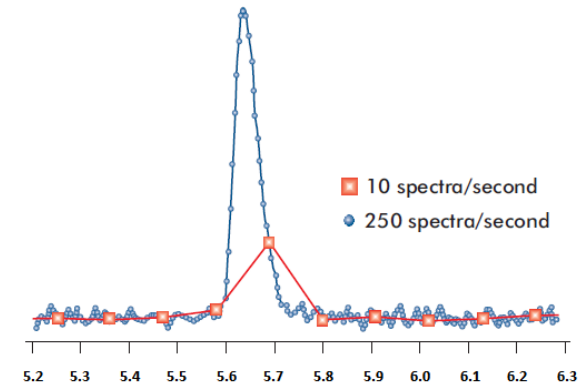


Diagrama del sistema GC×GC (*de.leco-europe.com*).

## Detectores GC×GC

- FID (máx. 300 Hz)
- TOFMS (máx. 500 Hz)
- $\mu$ -ECD (50-100 Hz)

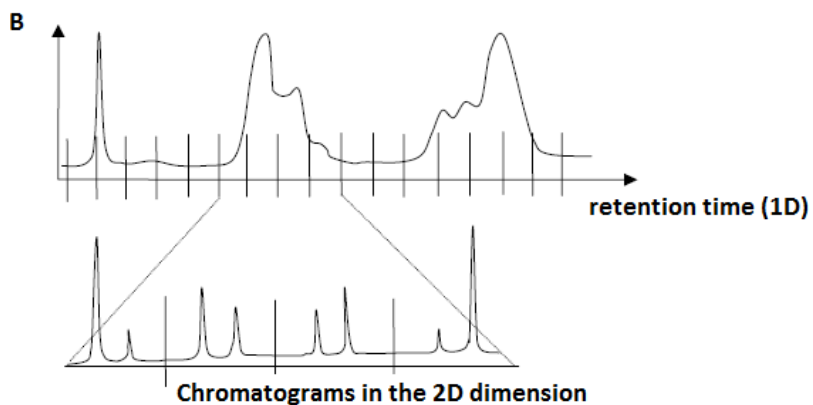
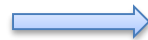
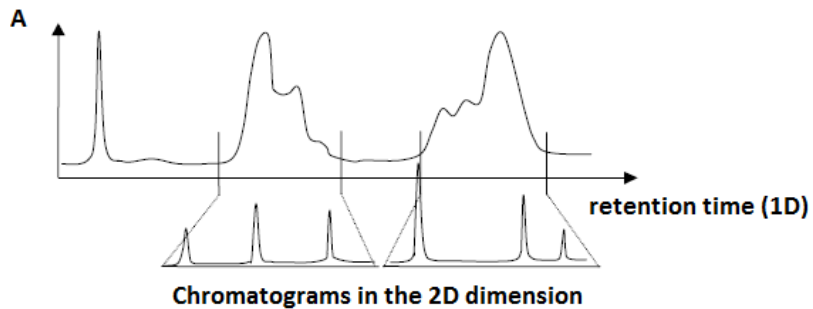


Velocidad de adquisición

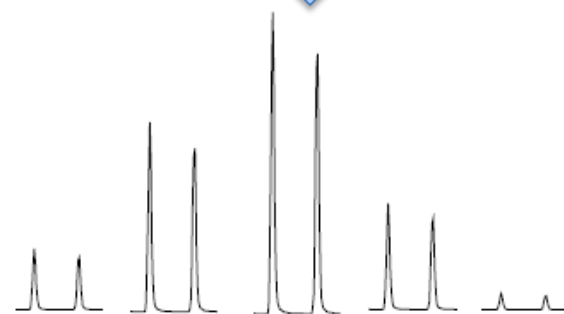
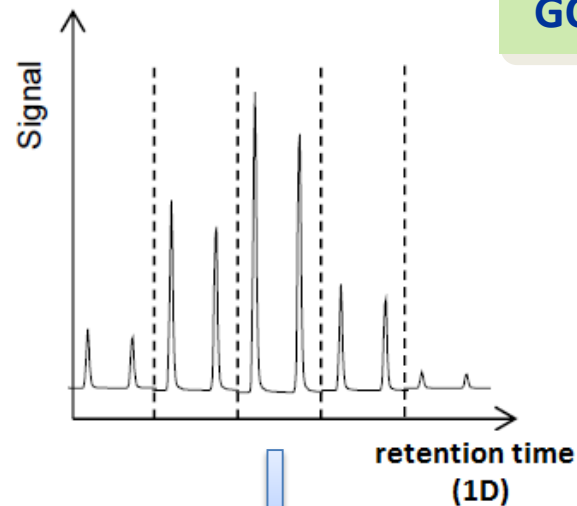
## Columnas GC×GC

Parámetros	Columna 1D	Columna 2D
longitud	15-30 m	0,5-2m
diámetro interior	0,25 mm	0,1 mm
grosor de la película de la fase estacionaria	0,25-1 $\mu$ m	0,1-0,25 $\mu$ m
tipo de fase estacionaria	no polar 100% polidimetilsiloxano o 5% fenil/95% dimetilsiloxano	polar 50% fenil/50% dimetilsiloxano o polietilenglicol (Carbowax)

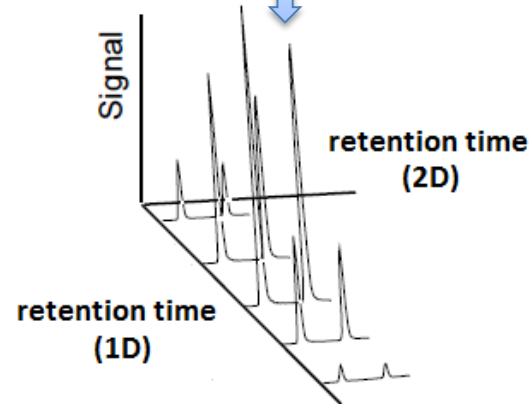
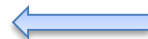
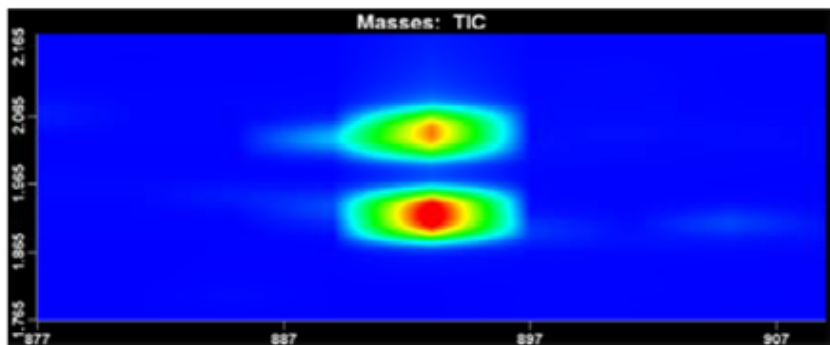
## GC×GC



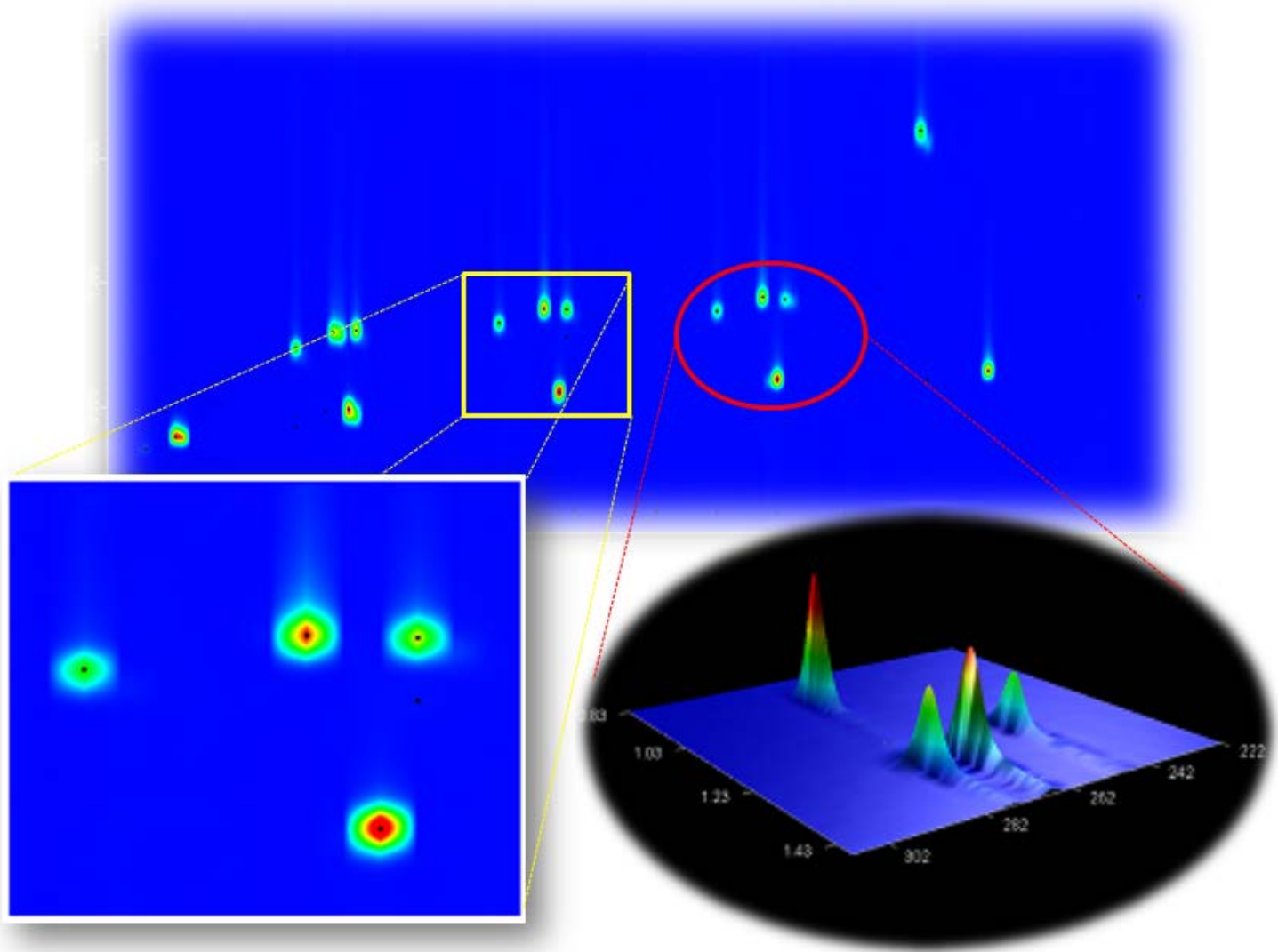
- (A) two fractions leaving I column are directed in turn to the II column  
 (B) comprehensive two-dimensional GC



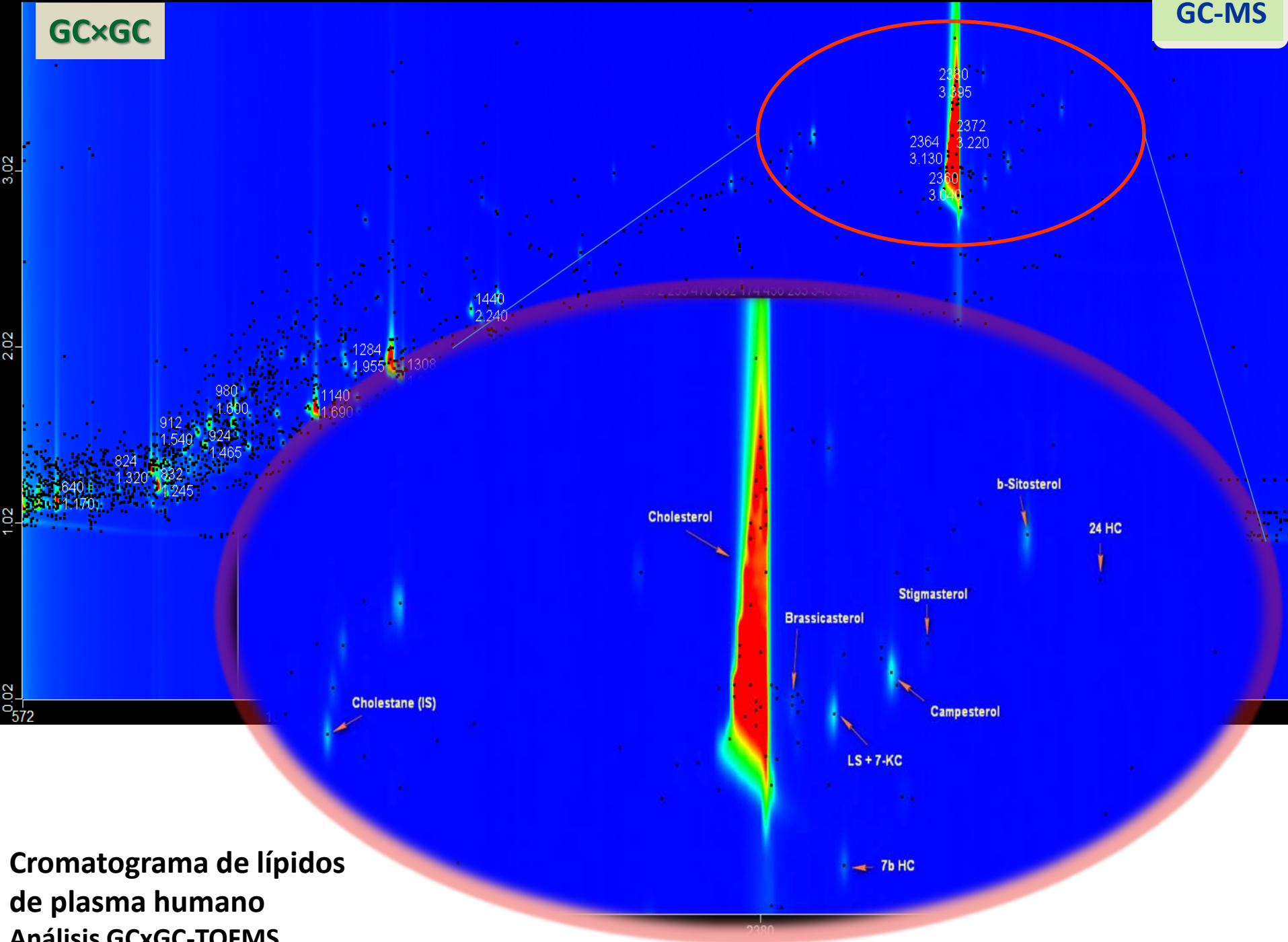
## Contour plot



GC×GC



Cromatograma de la mezcla de compuestos volátiles (alcanos, alcoholes, aldehídos, cetonas, aminas) analizados en GC 2D [Pegasus 4D; Leco].



**Cromatograma de lípidos  
de plasma humano  
Análisis GCxGC-TOFMS**



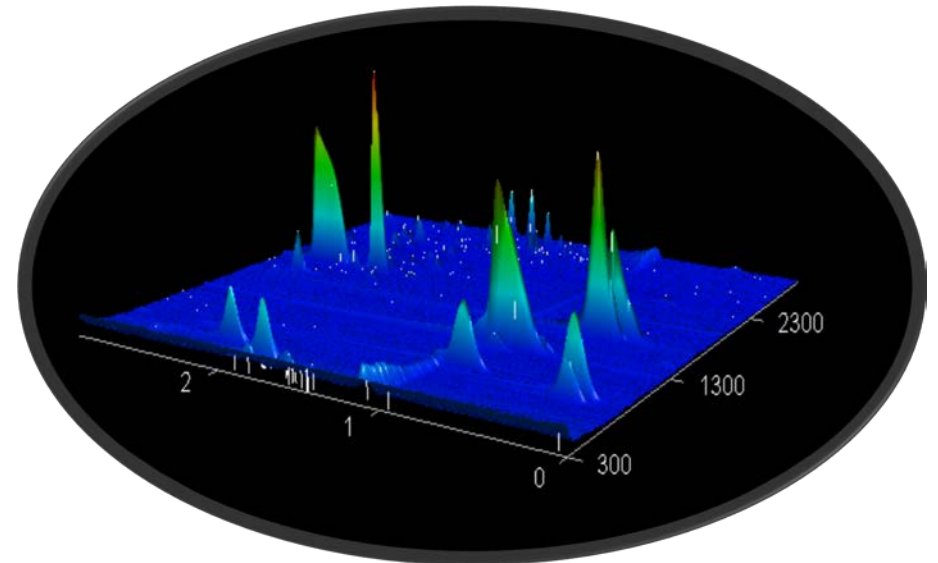
# GC×GC

## Ventajas

- Capacidad de picos incrementada - posibilidad de separación de mezclas complejas
- Separación total de todos los componentes
- Límites de detección (relación señal/ruido) mejorada (en comparación con la GC 1D)
- Se obtienen cromatogramas estructurados
- El cromatograma de GC 2D contiene mucha más información que el GC 1D - identificación más sencilla y fiable de sustancias desconocidas
- La naturaleza ordenada de los cromatogramas hace que el análisis de grupos sea mucho más sencillo

## Inconvenientes

- Se requieren detectores con una alta velocidad de adquisición (TOF,  $\mu$ ECD, FID)
- Amplia capacidad de datos registrados -el procesamiento de datos puede resultar muy largo
- Equipo muy costoso



AACLifeSci



Erasmus+

e-materials de

**Proyecto de Química Analítica Avanzada en Ciencias de la Vida [AACLifeSci]**

Este proyecto ha sido financiado con ayuda de la Comisión Europea.

La presente publicación recoge únicamente las opiniones de los autores, por lo que la Comisión no se hace responsable de cualquier uso que se haga de la información contenida en ella.

# Electroforesis Capilar- Espectrometría de Masas (CE-MS)

Coral Barbas  
Antonia García



# Electroforesis Capilar

- Separación basada en la movilidad electroforética
- Instrumentación simple
- Principales aplicaciones en el bioanálisis
  - Secuenciación de ADN
  - Análisis de fragmentos de ADN
- Múltiples modalidades para mejorar la selectividad con metabolitos neutros
  - MEKC
  - CEC



# Ventajas e inconvenientes de la CE

## Ventajas

- Ofrece nueva selectividad, una alternativa a la HPLC
- Selectividad sencilla y predecible
- Alta eficacia de separación (de  $10^5$  a  $10^6$  platos teóricos)
- Tamaños de muestras pequeños (1-10  $\mu$ l)
- Separaciones rápidas (de 1 a 45 min)
- Se puede automatizar
- Cuantificación (lineal)
- Distintas «modalidades» (a discutir)

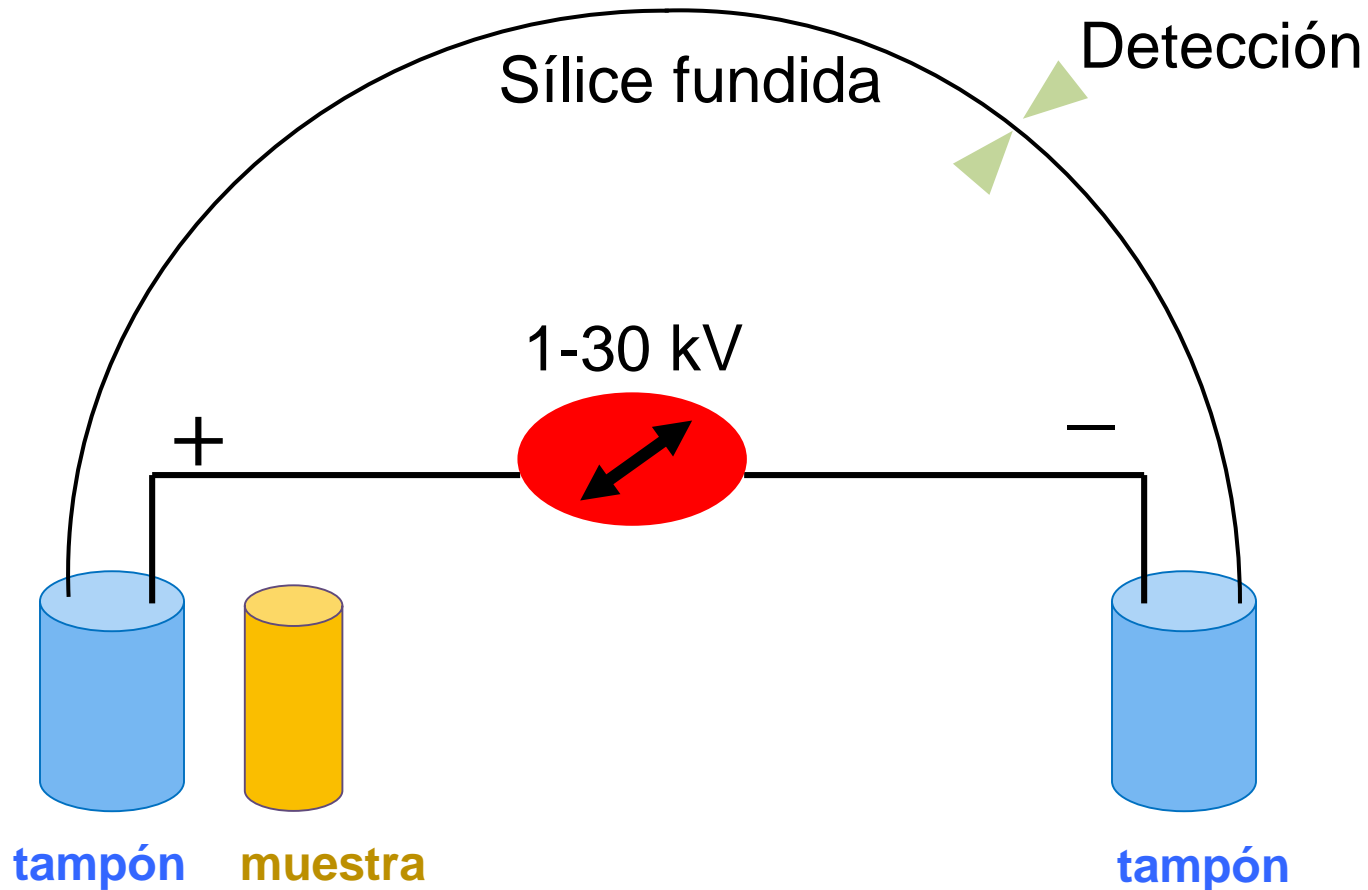
## Inconvenientes

- No se pueden realizar separaciones a escala preparativa
- Problemas con bajas concentraciones y grandes volúmenes
- Compuestos *pegajosos* («sticky»)
- Especies que son difíciles de disolver
- Problemas de reproducibilidad

# Aplicaciones de la CZE

- Gran variedad de aplicaciones
  - Pequeñas moléculas
  - Macromoléculas (proteínas, péptidos)
- Limitaciones
  - Debe tener distintas cargas
  - Muestra con baja fortaleza iónica
- Ventajas
  - Simple
  - Análisis directo de sistemas complejos

# Electroforesis Capilar (CE)

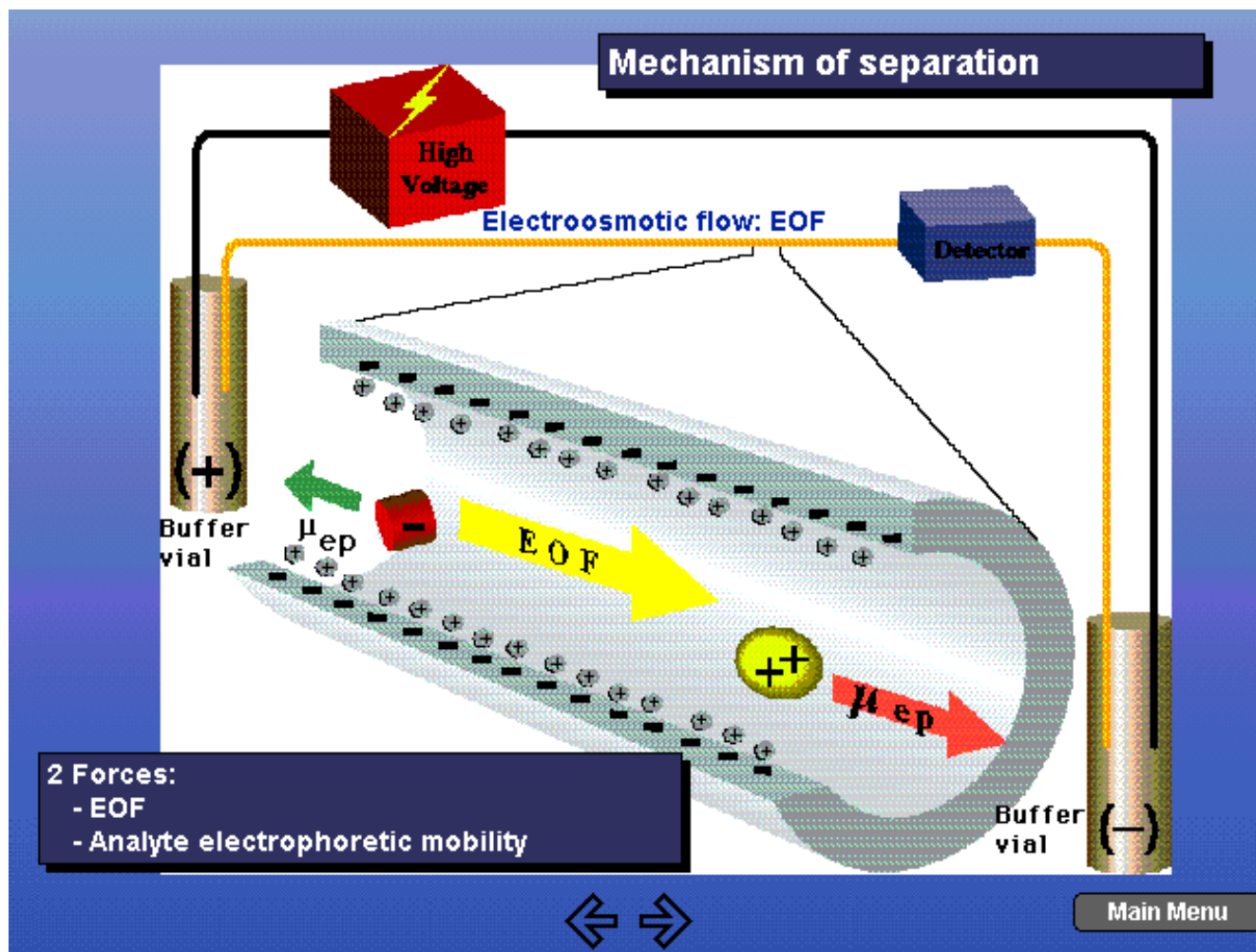


# Modalidades de CE

- Electroforesis Capilar Zonal (CZE)
  - Modalidad básica que utiliza canales abiertos
- Cromatografía Electrocinética Micelar (MEKC)
  - Separa compuestos con micelas
- Electroforesis Capilar en Gel (GCE)
  - Exclusión de tamaño que utiliza geles de filtración
- Electro cromatografía Capilar (CEC)
  - Híbrido entre CE y HPLC
- Isoelectroenfoque Capilar (CIEF)
- CE enantiomérica

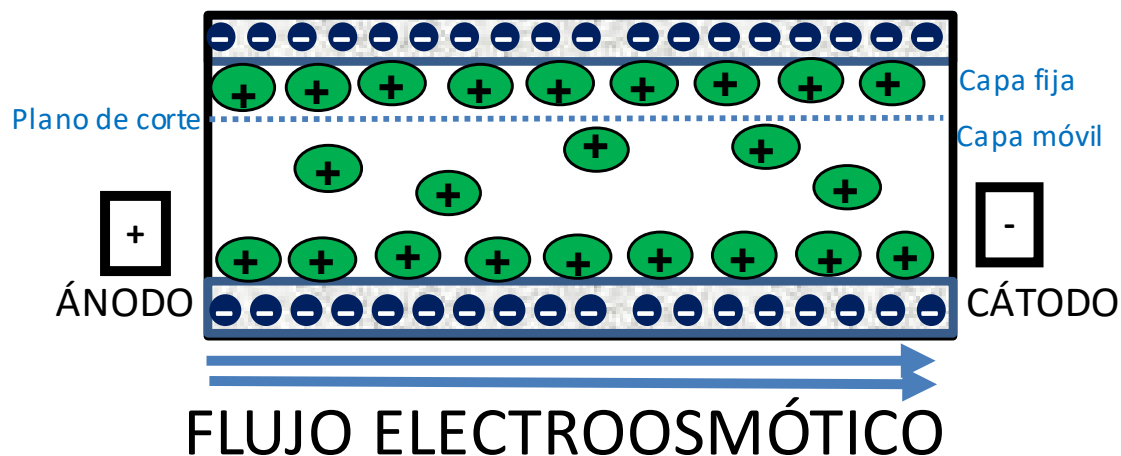


# Electroforesis Capilar Zonal (CZE)



# Flujo electroosmótico

- Mecanismo de flujo capilar basado en el potencial y el pH aplicados
- Proporciona un flujo de solución global («bulk») dentro de los capilares donde se utilizan tampones con concentración moderada o baja
- Gran dependencia de la fortaleza iónica de la solución y la química de superficies



# Movilidad Electroforética

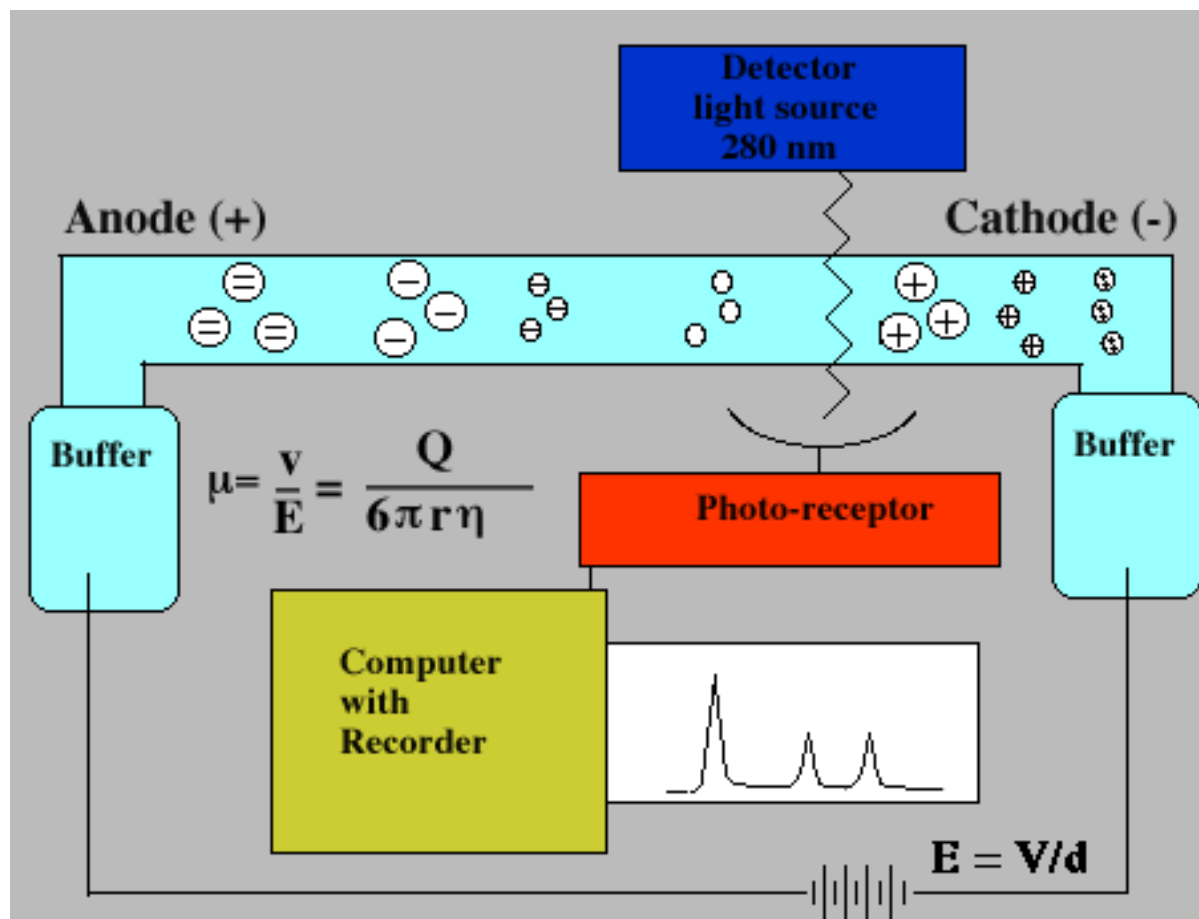
$$\mu = \frac{q}{6\pi\eta r}$$

$\mu$  = movilidad electroforética

$Q$  = carga en la partícula

$\eta$  = viscosidad de la solución

$r$  = Radio de Stokes de la partícula

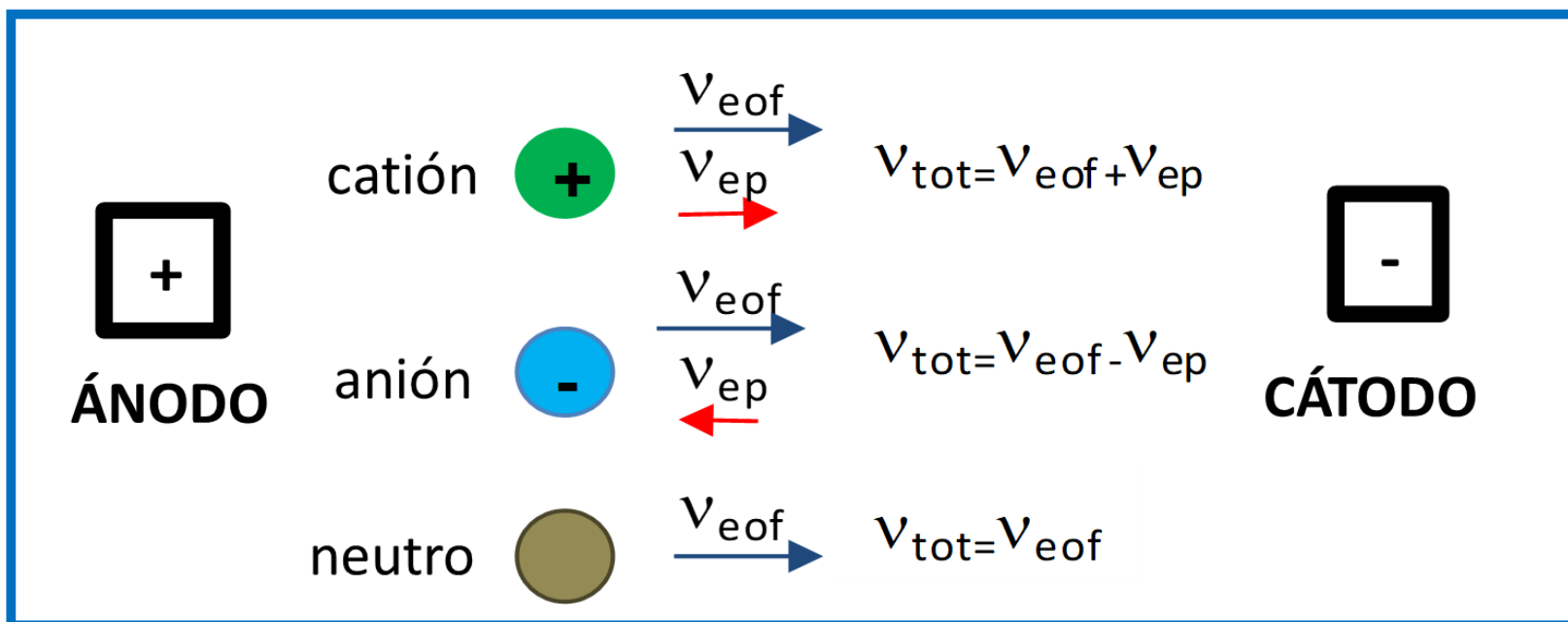


# Electroforesis y Electroosmosis

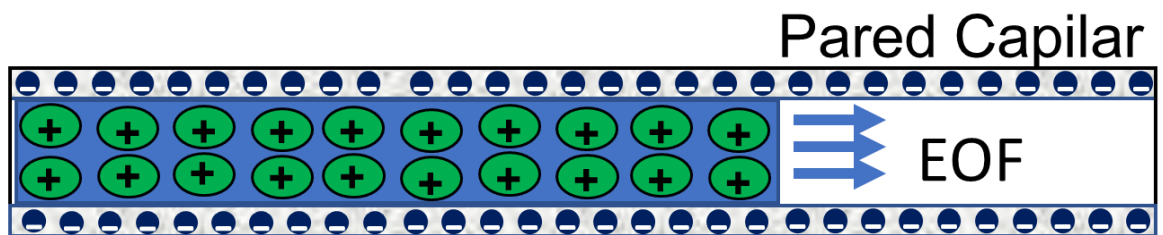
Representación gráfica del efecto combinado en un capilar cuando la electroosmosis (EO) es más rápida que la electroforesis (EP) (lo más habitual):

$$v = (\mu_{ep} + \mu_{eo})E = (\mu_{ep} + \mu_{eo})\frac{V}{L}$$

Imagen de R. N. Zare, Stanford

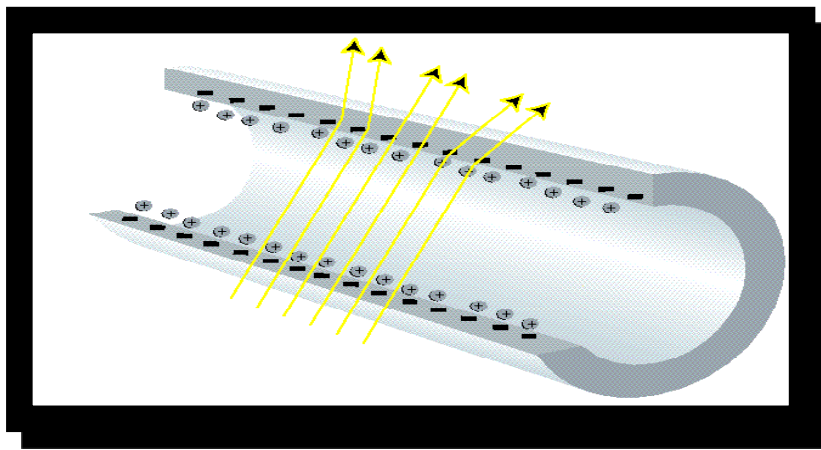


# Diagramas de flujo en CE y HPLC



# Opciones de Detección

Detector	Ventajas	Características	Limite de detección (M)
Absorbancia UV/vis	<ul style="list-style-type: none"><li>-Posibilidad de detección directa e indirecta</li><li>-Muy común</li></ul>	<ul style="list-style-type: none"><li>-Universal</li></ul>	<ul style="list-style-type: none"><li>-<math>10^{-3}</math></li><li>-<math>10^{-6}</math> para compuestos aromáticos</li></ul>
LIF (Fluorescencia inducida por láser)	<ul style="list-style-type: none"><li>-Muy sensible y selectivo</li><li>-Analitos fluorescentes o derivados</li></ul>	<ul style="list-style-type: none"><li>-Selectivo</li></ul>	<ul style="list-style-type: none"><li>-<math>10^{-6}</math> - <math>10^{-9}</math></li></ul>
Espectrometría de masas MS	<ul style="list-style-type: none"><li>-Información cualitativa y cuantitativa</li><li>-Muy sensible y selectivo</li></ul>	<ul style="list-style-type: none"><li>-Universal</li><li>-Selectivo</li></ul>	<ul style="list-style-type: none"><li>-<math>\approx 10^{-5}</math> (depende del tipo de MS y del tipo de metabolito)</li></ul>

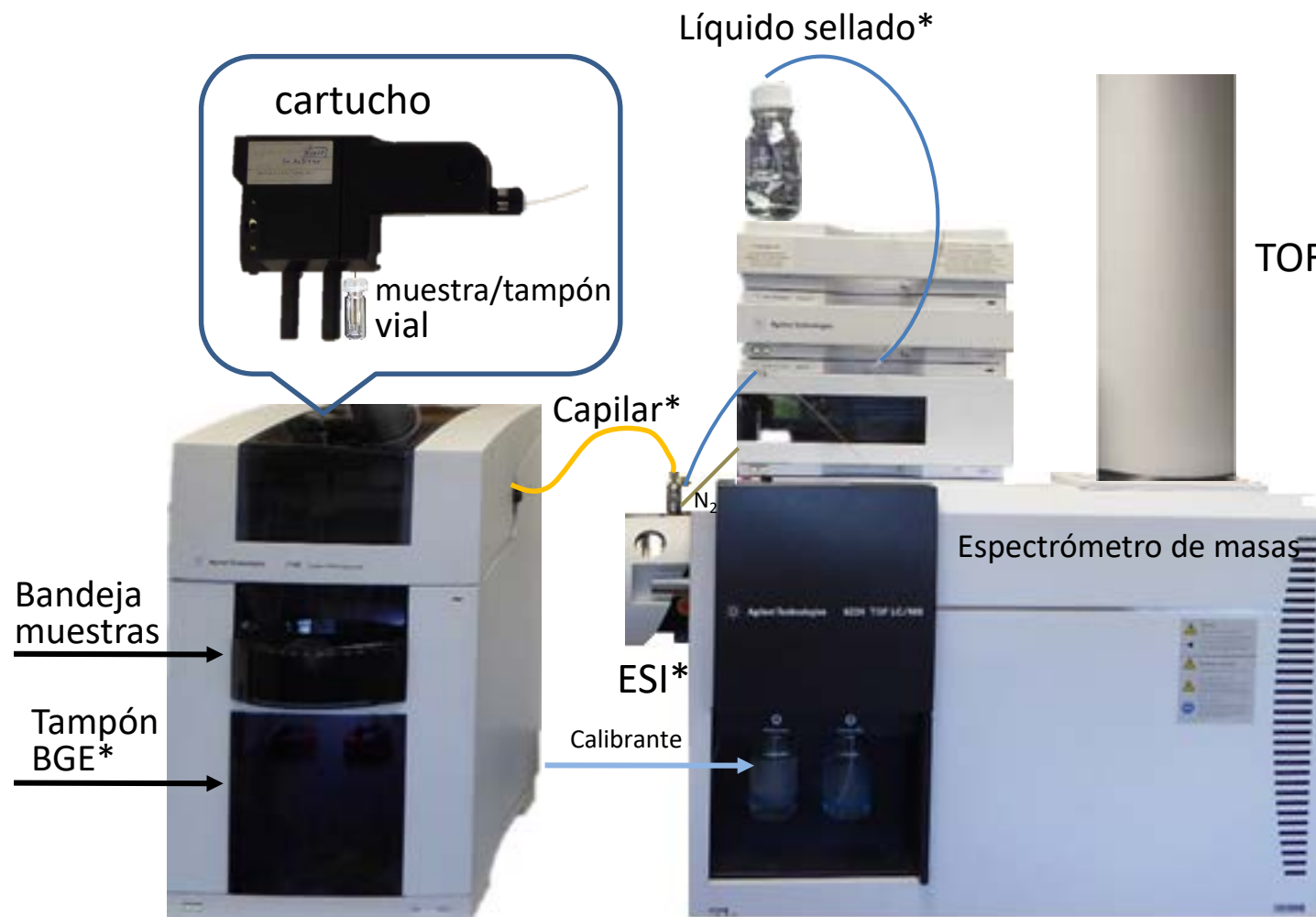


# Optimizar las Separaciones en CE

## PARÁMETROS

- pH
  - Primer parámetro que controlar
  - Afecta al EOF y a la movilidad (carga)
- Disolvente orgánico
  - Solvatación del analito
- Agente interactuante
  - Pares iónicos, solvatación, etc.
- Condiciones no acuosas
  - Solvatación y carga
- Temperatura
  - Solvatación, equilibrios químicos

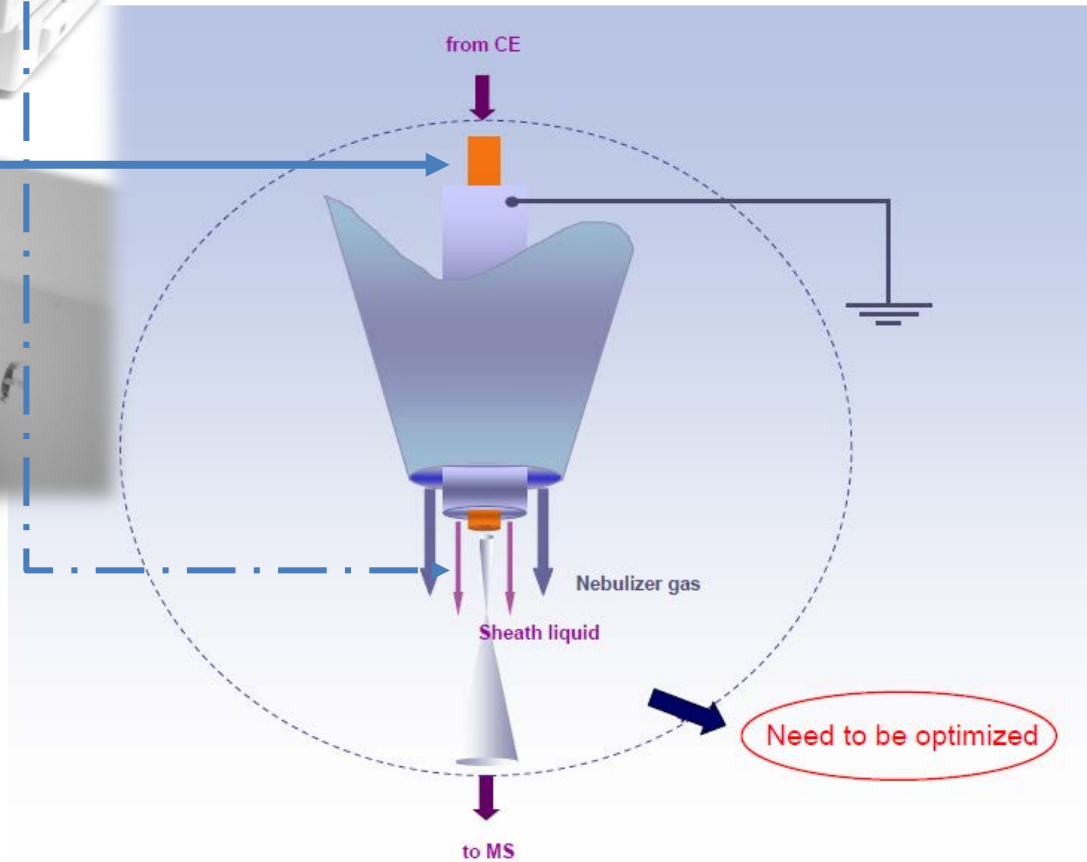
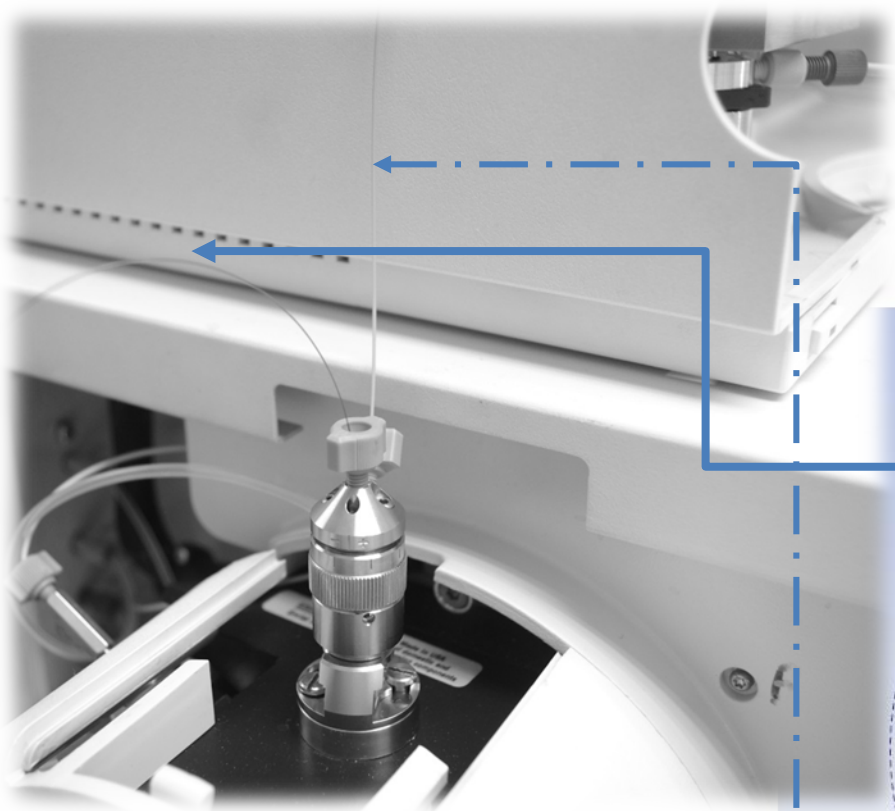
# INSTRUMENTACIÓN EN CE-MS



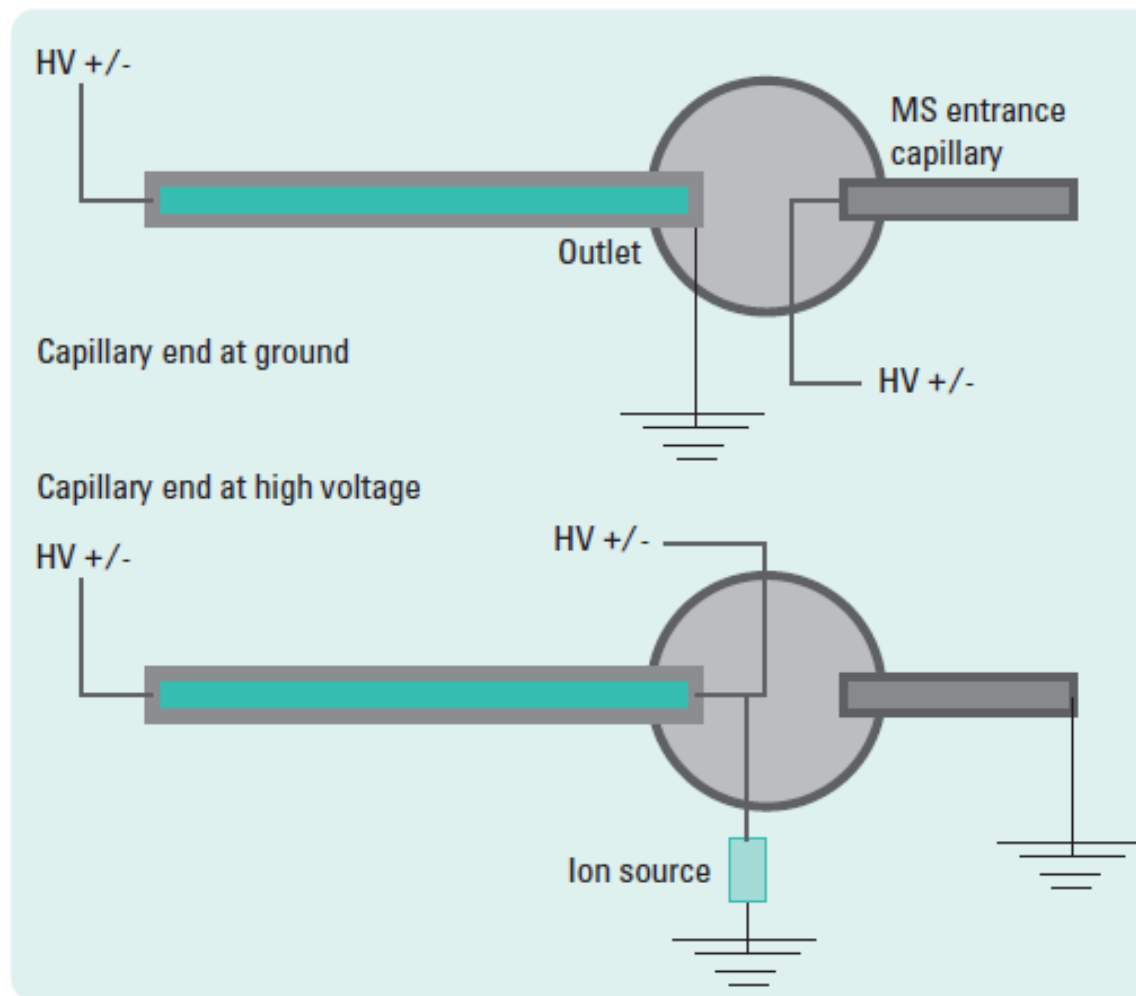
\* SE REQUIERE OPTIMIZACIÓN



# CE-MS CON FUENTE DE ESI



# CE-MS: Interfaz eléctrica



*imagen proporcionada por Agilent Technologies*

---

Este proyecto ha sido financiado con ayuda de la Comisión Europea.

La presente publicación recoge únicamente las opiniones de los autores, por lo que la Comisión no se hace responsable de cualquier uso que se haga de la información contenida en ella.



# METABOLÓMICA

Coral Barbas

Danuta Dudzik

M<sup>a</sup> Fernanda Rey-Stolle

Francisco J. Rupérez

Antonia García



# ÍNDICE

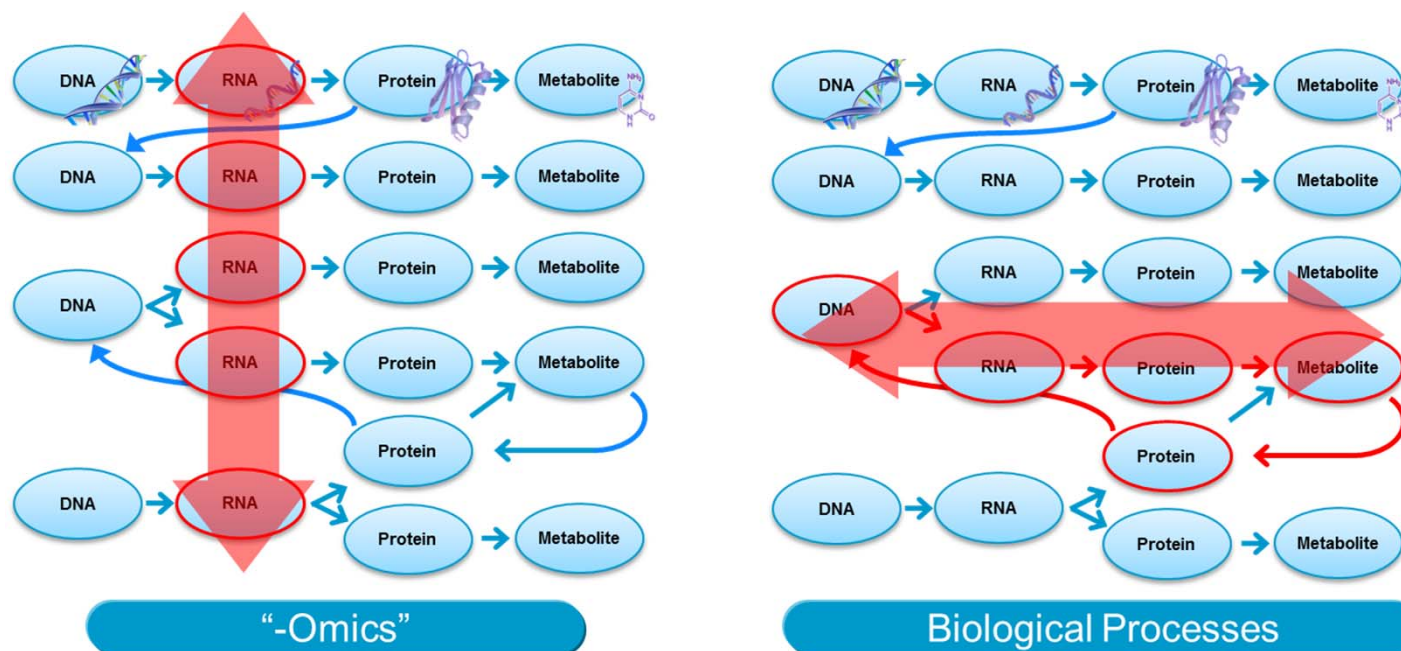
1. Introducción a la metabolómica.
2. Enfoques analíticos en la metabolómica.
  - Flujo de trabajo del estudio metabolómico
  - Procedimiento de Control de Calidad y Garantía de Calidad en la metabolómica
3. Procesamiento de datos e identificación de metabolitos
  - Flujo del procesamiento de datos
  - Tratamiento de datos en la metabolómica no dirigida
  - Identificación de metabolitos
  - Análisis estadístico
4. Análisis de datos
  - De la identificación de datos a las rutas metabolómicas
  - Validación de biomarcadores
5. Sesiones prácticas
  - Metabolómica dirigida y no dirigida
  - Metabolómica con herramientas gratuitas en línea



# Metabolómica

La investigación «ómica», nuevo campo emergente (que incluye la genómica, proteómica y metabolómica), se ocupa de la caracterización **completa** de los pequeños **metabolitos** moleculares presentes en sistemas biológicos.

## Ómica y Biología de Sistemas



## Anexo: Definición de Metabonómica

- Medida de la respuesta metabólica dinámica y multiparamétrica de un sistema vivo a estímulos patofisiológicos o modificaciones genéticas (Nicholson, 1999)
  - medida cuantitativa de la respuesta metabólica «total» en relación con el tiempo a estímulos patofisiológicos (nutricionales, xenobióticos, quirúrgicos o tóxicos)
- MetaboLómica - el cuadro / MetaboNómica - la película
- Hoy en día, todo es Metabolómica



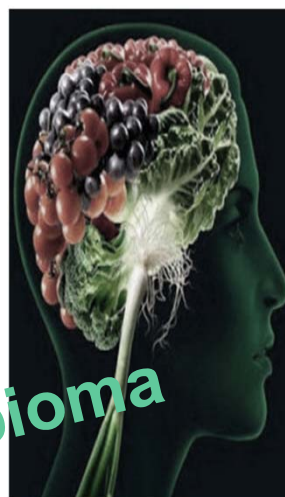
# Definición de Metaboloma

- «... el conjunto completo de metabolitos/sustancias intermedias con bajo peso molecular, que dependen de un contexto y que varían en función de la fisiología, el estado de desarrollo o patológico de la célula, tejido, órgano u organismo...» (Oliver 2002)
- **Origen:** Endometaboloma, Microbioma, Xenobioma, Nutribioma...
- **Naturaleza:** Glicoma, lipidoma, esfingolipidoma, peptidoma...
- Metaboloma ↔ Fenotipo

## Hospedador



Microbioma



Nutribioma

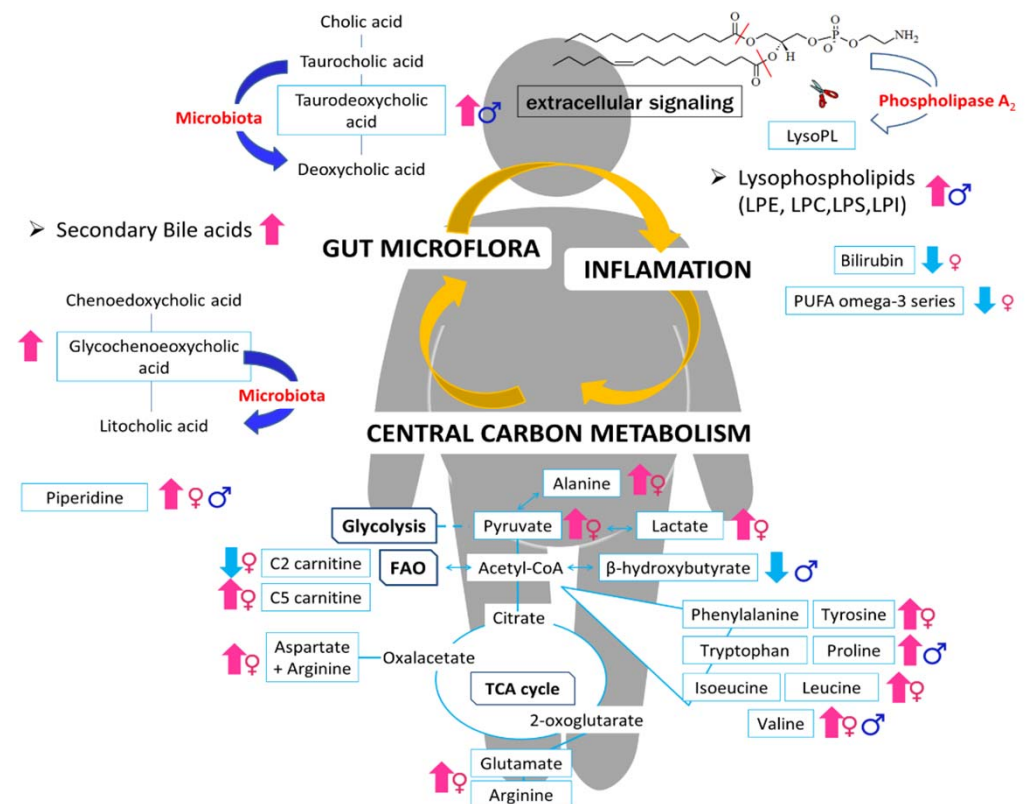


Xenobioma



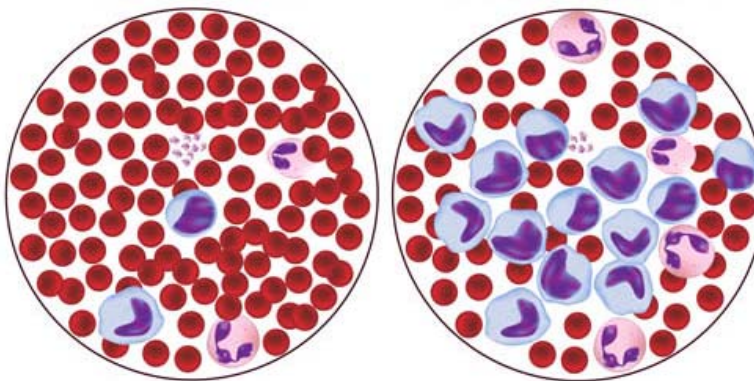
# Lo que puede aportar la metabolómica (I)

- **Perspectiva global** del estatus metabólico y de los eventos bioquímicos globales asociados a un sistema celular o biológico.
  - Situaciones patológicas sin mecanismo conocido, como la relación entre la obesidad y la resistencia a la insulina



# Lo que puede aportar la metabolómica (II)

- Identificación (propuesta) de nuevos **biomarcadores**, importantes en el proceso de descubrimiento de nuevos fármacos o como herramientas de diagnóstico in vitro.
  - Por ejemplo, nuevos biomarcadores diagnósticos para la agresividad en la leucemia linfática crónica



Utility of validated metabolites as biomarkers of aggressive state of CLL

Metabolite	AUC	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Acetylcarnitine	0.695	43.2	93.0	86.1	62.1
Butyrylcarnitine	0.548	10.8	98.0	84.4	52.4
Hexanoylcarnitine	0.690	27.0	96.0	87.1	56.8
Octanoylcarnitine	0.651	29.7	95.0	85.6	57.5
Decanoylcarnitine	0.662	27.0	94.0	81.8	56.3
Palmitoylcarnitine	0.719	40.5	94.0	87.1	61.2
Dodecanamide	0.497	8.1	100.0	100.0	52.1
Hexadecanamide	0.516	5.4	100.0	100.0	51.4
Oleamide	0.600	18.9	96.0	82.5	54.2
Linoleamide	0.672	16.2	98.0	89.0	53.9
Acylcarnitines <sup>a</sup>	0.743	32.4	95.0	86.6	58.4
FAA <sup>b</sup>	0.662	13.9	96.0	77.6	52.7
Acylcarnitines and FAA	0.750	54.0	89.0	83.1	65.9

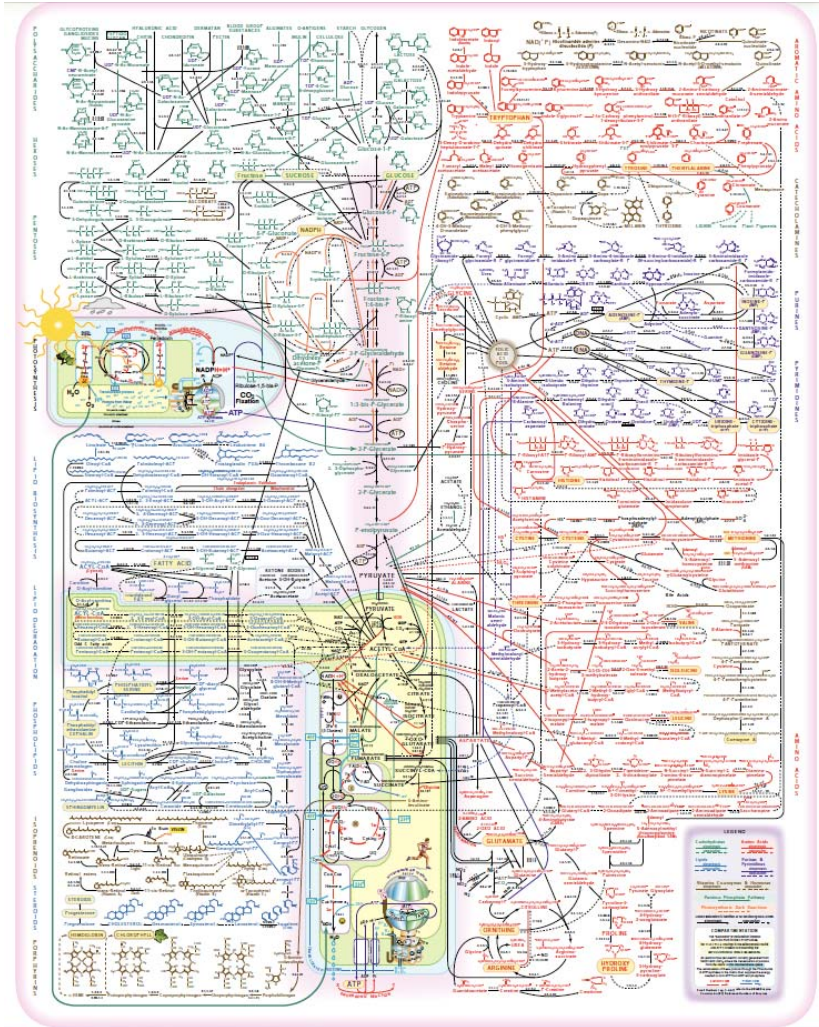
## La metabolómica sirve para...

- buscar diferencias metabólicas entre grupos de muestras (caso vs. control; pre-tratamiento vs. post-tratamiento; una condición vs. otra)
- identificar compuestos que sean significativos y proponer los mecanismos
- averiguar información sobre el fenotipo
- observar los efectos de un tratamiento
- encontrar nuevos objetivos farmacológicos

## La metabolómica NO es...

- un método para revelar el destino de un metabolito o un fármaco
- un método de cuantificación
- el uso de un simple kit para cuantificar un grupo de metabolitos (se requiere NMR, MS...)
- posible sin una comparación simultánea de muestras

# Definición de Metabolismo

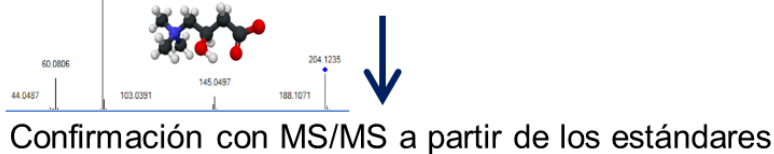
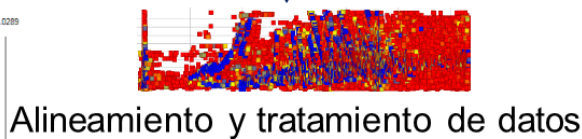
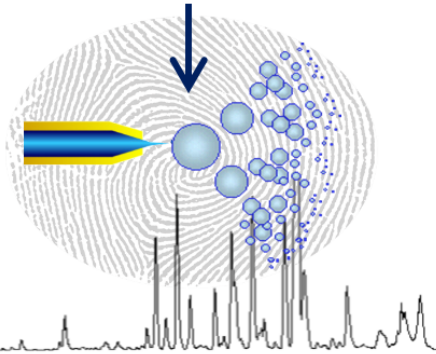


El grupo completo de procesos (bio)químicos dentro de un orgánulo, célula, tejido, órgano u organismo, esenciales para la vida



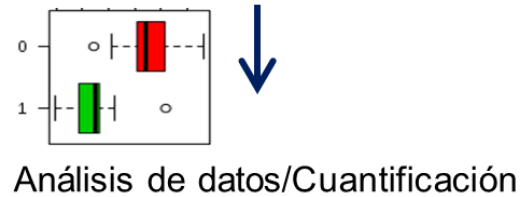
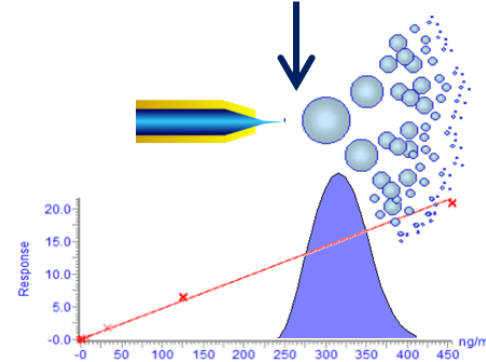
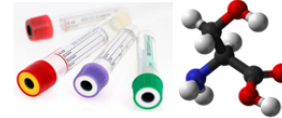
## Estrategias metabolómicas

### No dirigida



**Perfil metabólico global**

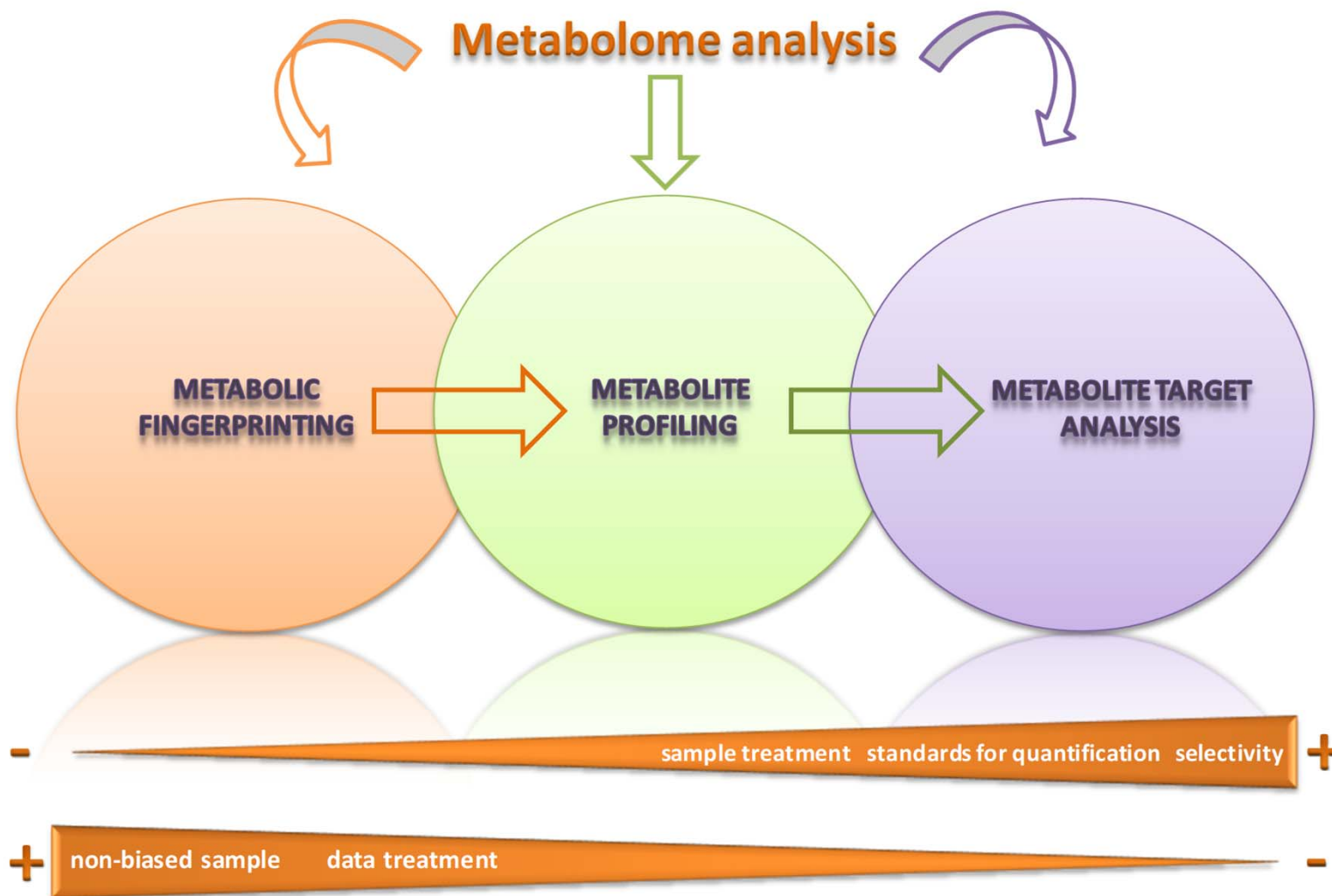
### Dirigida



Comparación de muestra y estándares

**Cuantificación de metabolitos específicos**

# Tres maneras de hacer metabolómica



# FLUJO DE TRABAJO EN METABOLÓMICA



# TÉCNICAS ANALÍTICAS

- GC/MS: Compuestos polares pequeños
  - Principalmente solubles en agua (algunos hidrofóbicos)
  - Tratamiento de muestras: Derivatización
  - Fragmentación reproducible - bases de datos
- NMR
  - Soluble en agua
  - Prácticamente sin tratamiento de muestra
  - LOD alto
- LC/MS
  - metabolitos de tamaño pequeño a grande (<1500 Da), y de polaridad media a nula
- CE/MS: Compuestos polares pequeños-medianos
  - Aminoácidos, acilcarnitinas, poliaminas, etc.
  - Sin derivatización





# Plataformas analíticas basadas en la MS en la metabolómica

## Técnica Analítica

## Aplicación

## Ventajas

## Desventajas

GC-MS

Separación, identificación, y cuantificación de metabolitos poco polares, volátiles y térmicamente estables

Elevada resolución, posibilidad de librerías de metabolitos de amplio espectro para identificación.

Imposibilidad de analizar metabolitos termoestables. Se requiere la derivatización de metabolitos no volátiles de elevado peso molecular

LC-MS

Separación, identificación, y cuantificación de numerosos grupos de metabolitos, dependiendo del tipo de columna y fase móvil

Alta sensibilidad, alta capacidad de carga de muestra, no necesaria derivatización, posibilidad de analizar compuestos termolábiles.

Disponibilidad limitada de librerías comerciales, restricción de eluyentes en LC, efecto matriz, limitado potencial de identificación a menos que se utilice como detección MS-MS

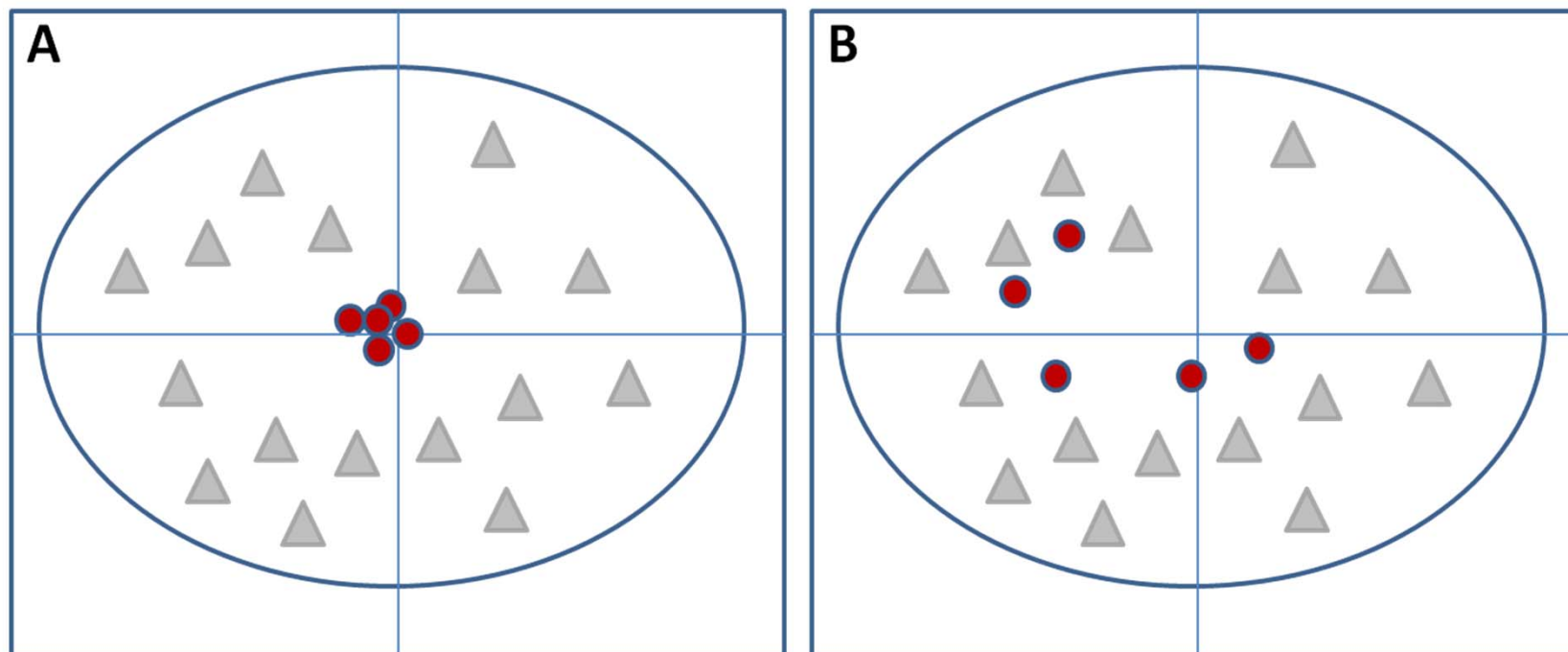
CE-MS

Separación, identificación, y cuantificación de metabolitos polares e ionizados, volúmenes de muestra muy pequeño.

Elevada resolución y análisis rápido, útil para muestras biológicas complejas, incluso con muy poco volumen

Disponibilidad limitada de librerías comerciales. Incompatibilidad de BGE, límites de detección. Limitado potencial de identificación a menos que se utilice como detección MS-MS

# Procedimiento de Control de Calidad y Garantía de Calidad en la metabolómica

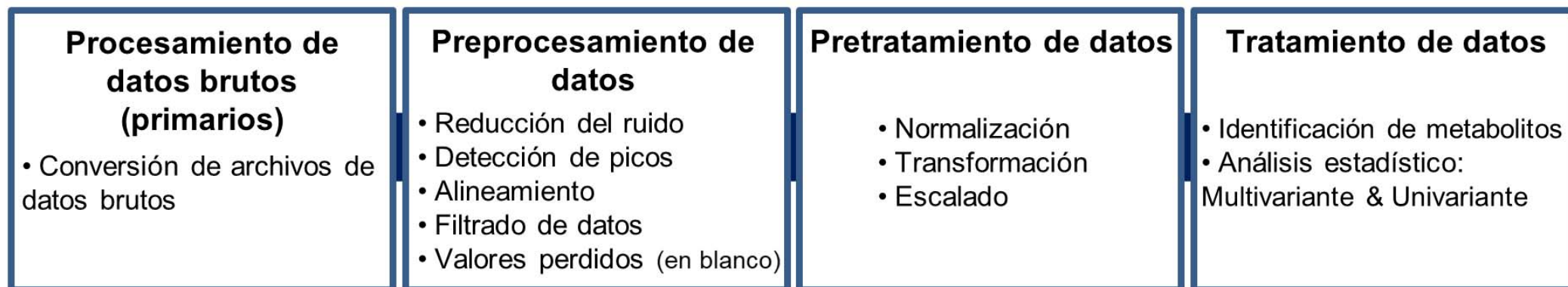


A: QC (puntos rojos) agrupadas

B: QC dispersas

# TRATAMIENTO DE DATOS EN METABOLÓMICA:

## Procesamiento de señales

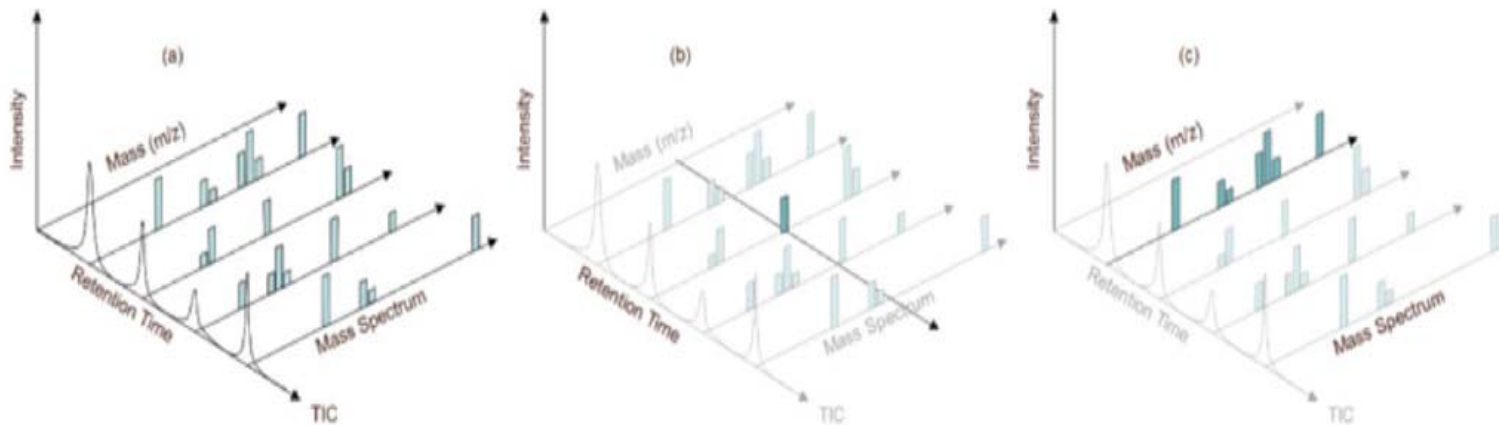


**FLUJO DEL PROCESAMIENTO DE DATOS**



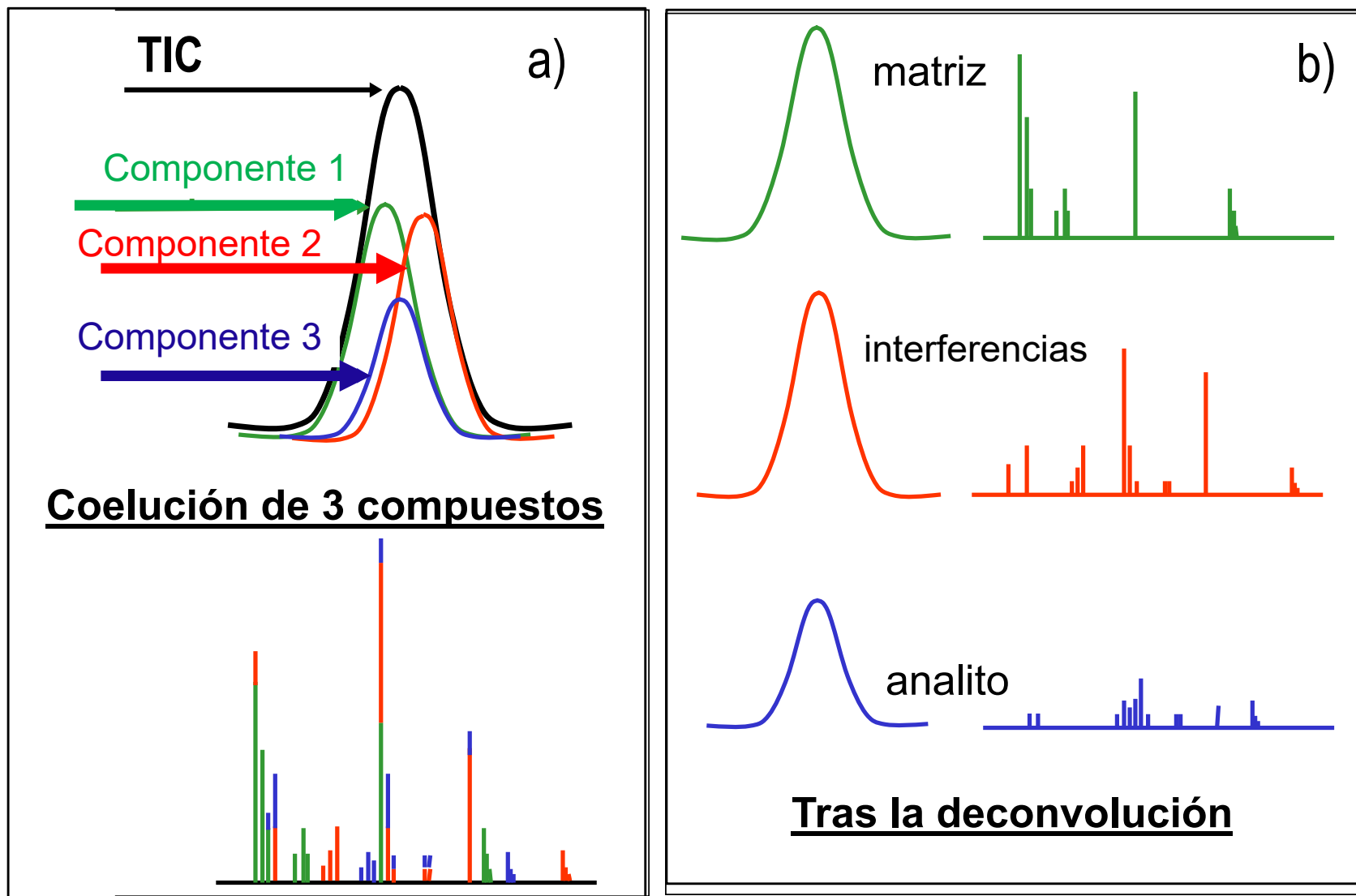
# TÉCNICA ANALÍTICA: GC-MS

- Cromatografía de gases acoplada a la espectrometría de masas
- Criterio de referencia
  - Altamente sensible y reproducible
  - Información: Calidad y Cantidad
  - Librerías de espectros para propósitos de identificación
  - El 10-20% de los compuestos conocidos se pueden analizar mediante GC
    - Alta relevancia metabólica



(a) Datos 3D de GC/MS; (b) Cromatograma de ion extraído para el ion seleccionado  
(c) Un único punto de datos en el tiempo proporciona un único espectro de masas  
*extraído de Chromatography Today*

# Deconvolución



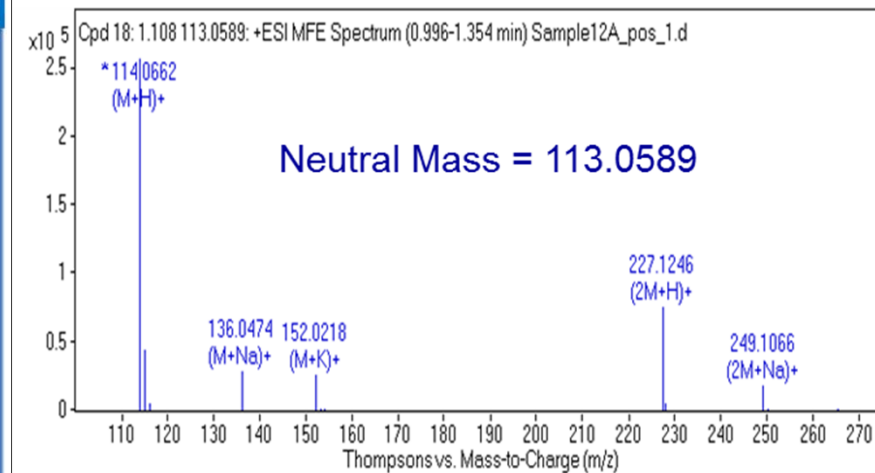
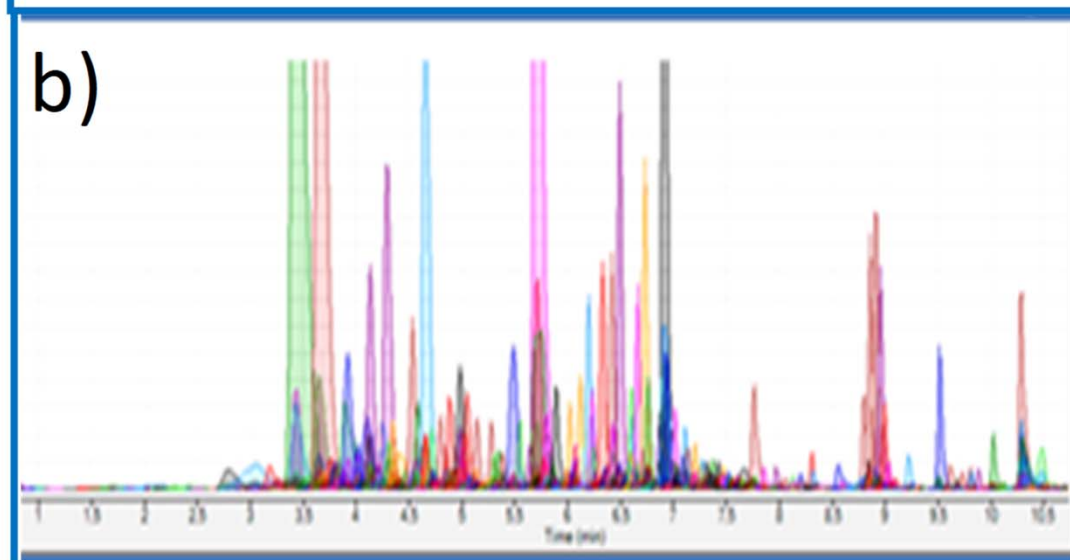
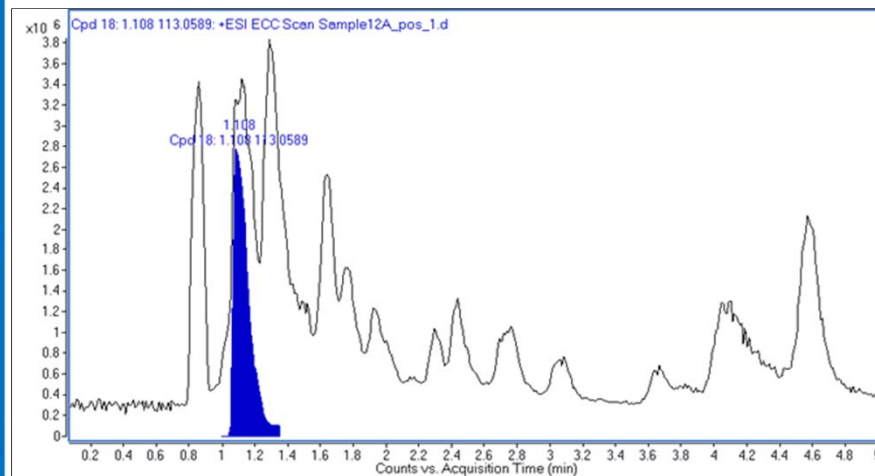
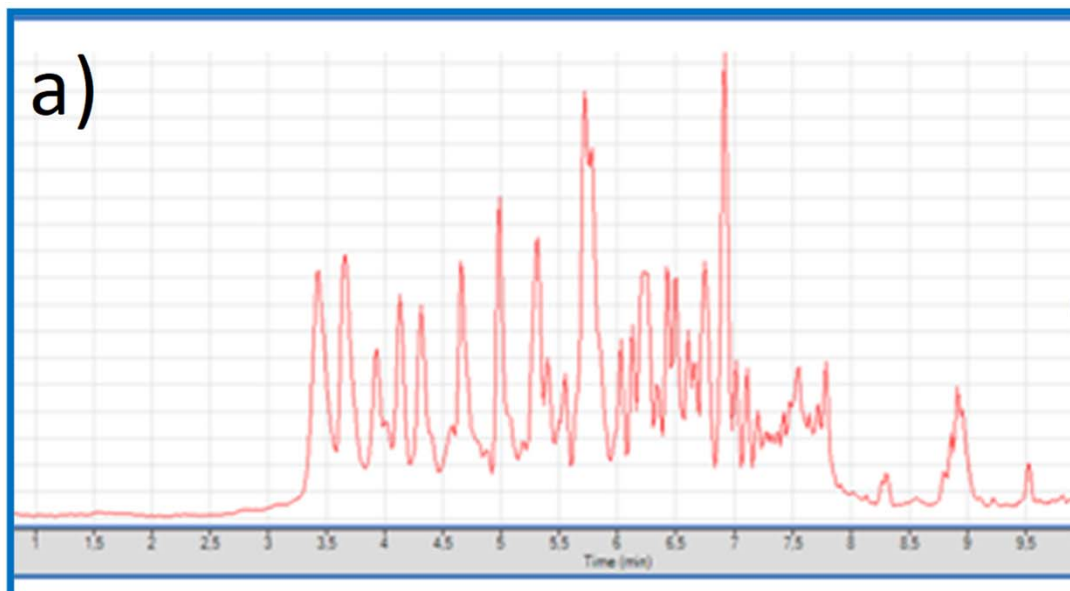
a) Antes y b) Después del proceso de deconvolución

extraído de <https://www.agilent.com/cs/library/Support/Documents/f05017.pdf>

# Deconvolución en LC-ESI-MS y CE-ESI-MS

- Métodos basados en picos
- A la hora de determinar si los distintos iones son del mismo “posible compuesto” (*feature*) el *Molecular Feature Extractor* (Agilent) considera la exactitud de las medidas de masas para iones relacionados en un grupo por su estado de carga, distribución isotópica, y las posibles relaciones químicas.
- También puede considerar iones vinculados como aductos: aductos de protón, sodio, potasio y amonio en la ionización positiva o pérdida de un protón, aductos con formiato, etc., en el modo de ionización negativa.

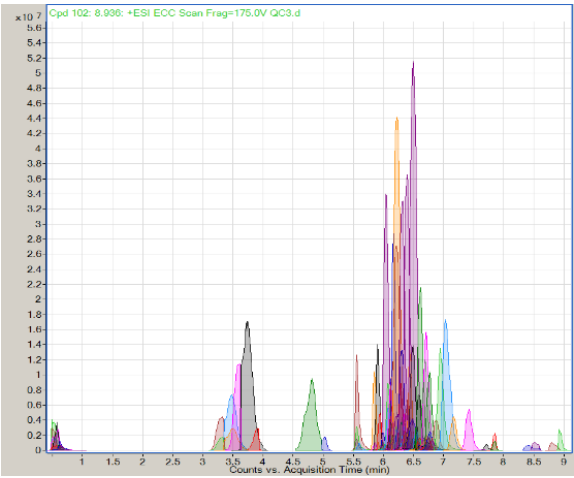
# Después de la deconvolución



a) Cromatograma de Iones Totales

b) Cromatogramas de cada uno de los compuestos obtenidos después de la deconvolución

# ¿Cromatograma o lista de características?



Automatically Show Columns

Show/Hide	Saturated	RT	m/z	Mass	Polarity	Ions	Height	Area	Vol	Quality Score
<input checked="" type="checkbox"/>	S	0.514	280.0923	279.085	Positive	5	2740215		14234373	100
<input checked="" type="checkbox"/>	S	0.517	203.0526	202.0454	Positive	6	4045612		20286412	80
<input checked="" type="checkbox"/>		0.526	140.0682	139.0609	Positive	3	1530642		7519633	100
<input checked="" type="checkbox"/>		0.529	136.0482	135.041	Positive	2	1187260		6016909	100
<input checked="" type="checkbox"/>	S	0.57	162.1126	161.1053	Positive	7	2758926		19465836	100
<input checked="" type="checkbox"/>	S	0.57	304.2998	303.2926	Positive	4	3021606		14360408	100
<input checked="" type="checkbox"/>		0.58	114.0664	113.0591	Positive	4	549599		6985240	80
<input checked="" type="checkbox"/>		0.614	175.1192	174.1119	Positive	3	760396		3860452	91.6
<input checked="" type="checkbox"/>		0.625	156.0768	155.0696	Positive	7	1055485		6005360	100
<input checked="" type="checkbox"/>		0.646	170.0927	169.0854	Positive	4	604901		3355886	100
<input checked="" type="checkbox"/>	S	3.29	520.3389	519.3316	Positive	7	3243984		36560884	80.2
<input checked="" type="checkbox"/>		3.298	544.3388	543.3315	Positive	4	1339590		12508261	87
<input checked="" type="checkbox"/>	S	3.483	520.3389	519.3316	Positive	11	5382675		86088232	85.2
<input checked="" type="checkbox"/>	S	3.5	544.3389	543.3316	Positive	7	2216628		36501788	100
<input checked="" type="checkbox"/>	S	3.573	496.3389	495.3316	Positive	6	8281163		80624664	86.1
<input checked="" type="checkbox"/>	S	3.746	496.3388	495.3316	Positive	13	11742118		200075408	80
<input checked="" type="checkbox"/>	S	3.89	522.3545	521.3472	Positive	8	2204035		17763280	87
<input checked="" type="checkbox"/>	S	4.801	524.3702	523.3629	Positive	12	6862920		111854136	100
<input checked="" type="checkbox"/>		5.027	524.3702	523.3629	Positive	6	1308318		8748089	87
<input checked="" type="checkbox"/>		5.559	163.0393	162.032	Positive	4	1349239		5379156	100
<input checked="" type="checkbox"/>	S	5.561	391.2839	390.2767	Positive	15	7367381		51796644	100
<input checked="" type="checkbox"/>	S	5.561	149.0235	148.0162	Positive	4	2915086		12886771	100
<input checked="" type="checkbox"/>		5.597	338.3418	337.3345	Positive	9	787867		4058574	100
<input checked="" type="checkbox"/>	S	5.849	675.5425	674.5353	Positive	7	6821314		39496860	100
<input checked="" type="checkbox"/>		5.857	627.5339	626.5266	Positive	3	712763		3673685	100
<input checked="" type="checkbox"/>	S	5.904	701.5583	700.551	Positive	8	8615328		59468044	100
<input checked="" type="checkbox"/>	S	5.941	689.558	688.5507	Positive	8	3178377		18867332	100
<input checked="" type="checkbox"/>		5.991	754.5372	753.5299	Positive	4	737180		4127601	100
<input checked="" type="checkbox"/>		5.995	730.5360	729.5297	Positive	7	1506513		8484420	100

Chromatogram Results | MS Spectrum



- **Alineamiento**

- Los desplazamientos de los picos se observan a través del eje del RT
- Dos grupos:
  - se alinean los datos antes de la detección de picos
  - métodos de alineamiento basados en el pico: se alinean en todas las muestras los picos espectrales detectados
  - programas:
    - MetaboAnalyst ([metaboanalyst.ca](http://metaboanalyst.ca))
    - mzmine and mzmine2 (<http://mzmine.sourceforge.net/>)
    - metAlign
    - BinBase ([fiehnlab.ucdavis.edu](http://fiehnlab.ucdavis.edu))
    - xcms and xcms2 (Scripps)
    - metaXCMS (Scripps)
    - XCMS Online (Scripps)

- **Valores en blanco**

- Problemas en nuevos análisis
- Distintas estrategias
  - Sustituir por la mitad del mínimo, por media/mediana, el k vecino-más-cercano (KNN), el PCA probabilístico (PPCA), el método PCA bayesiano (BPCA), o la Descomposición en Valores Singulares (SVD), etc.

- **Filtración**

- Variables de tamaños muy pequeños - detectados usando la media o la mediana
- Variables que son prácticamente constantes - detectadas usando la desviación estándar (SD)
- Variables que muestran baja repetibilidad - medidas usando la muestra QC

# Pretratamiento de datos

- Normalización

- Normalización **específica de muestra** (i.e. peso, volumen)
- Normalización por **suma o mediana**
- Normalización por **muestra** de referencia
- Normalización mediante muestra **combinada** a partir de un grupo de control
- Normalización por **característica** de referencia
- Normalización por **cuantiles**

- Transformación de datos

- Transformación **logarítmica**
- Transformación de **raíz cúbica**

- Escalado de datos

- Centrado a la **media**
- **Autoescalado** (centrado a la media y dividido por la desviación estándar de cada variable)
- **Escalado Pareto** (centrado a la media y dividido por la raíz cuadrada de la desviación estándar de cada variable)
- **Escalado por rango** (centrado a la media y dividido por el rango de cada variable)

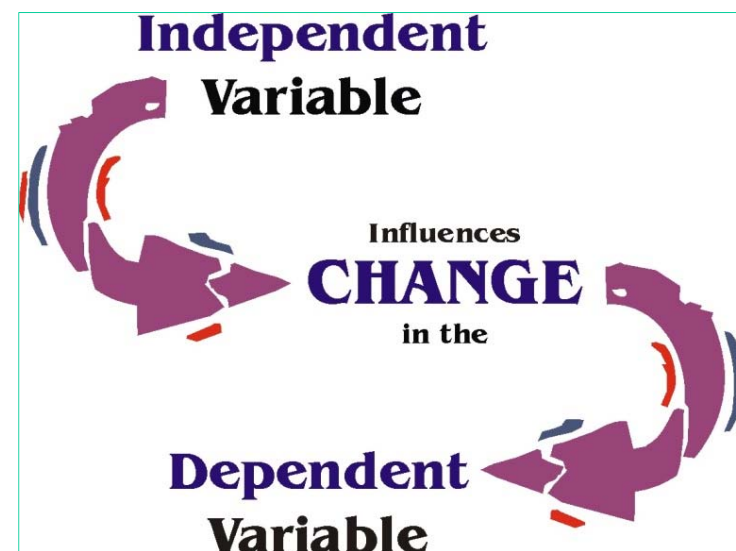
# Estadística para Metabolómica

## Objetivos:

- detectar diferencias entre grupos de muestras a nivel químico
- categorizar compuestos de acuerdo a su importancia relativa para la diferenciación de muestras

## VARIABLES

- **variable dependiente:** representa el resultado o el efecto, o se evalúa para comprobar si existe un efecto, p. ej.: abundancia de metabolito
- **variable independiente:** representa las entradas o las causas, o se evalúan para determinar si son las causas, p. ej.: condiciones de tratamiento dentro del experimento

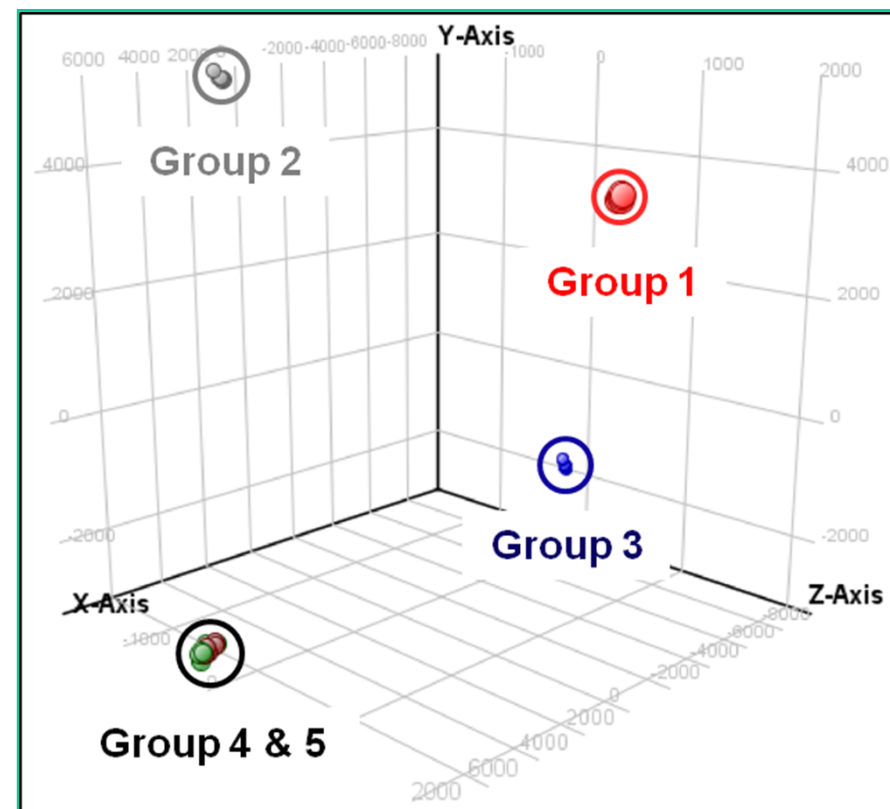


## TIPOS

- **Análisis univariante UVA:**
  - Distribución normal: Prueba *t* de Student, ANOVA
  - Distribución no normal: prueba de la U de Mann-Whitney, Kruskal-Wallis
- **Análisis multivariante MVA:** PCA, PLS-DA, OPLSDA

# PCA

- utilizado como herramienta en análisis exploratorio de datos
- cada punto representa gráficamente cada muestra medida
- el algoritmo no tiene conocimiento de las asociaciones en grupo de las muestras - análisis *no supervisado*
- el primer compuesto principal explica la mayor parte de la variación
- las cargas de compuestos indican el impacto del compuesto en el análisis
- cada punto es la suma de las cargas de compuestos de una muestra
- la densidad de la agrupación refleja la variación de las muestras



# Predicción de clases

un algoritmo que utiliza datos previos para predecir los resultados de observaciones futuras

- el algoritmo tiene conocimiento de las asociaciones en grupo de las muestras - análisis *supervisado*
- algoritmos habituales
  - **Análisis discriminante por mínimos cuadrados parciales (PLS-DA)**
  - Máquinas de soporte vectorial
  - Árbol de decisión
  - Bayesiano ingenuo
  - Red neuronal

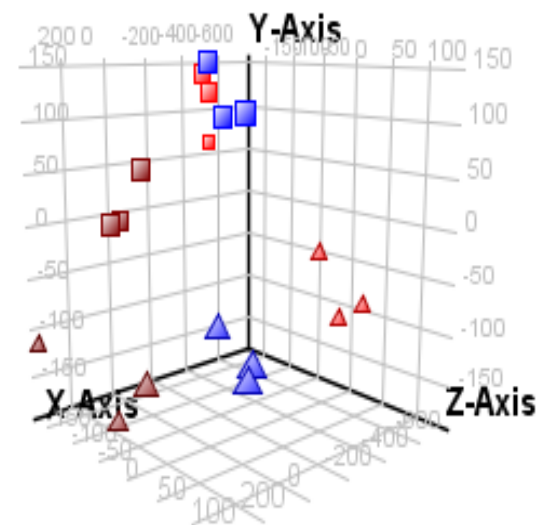
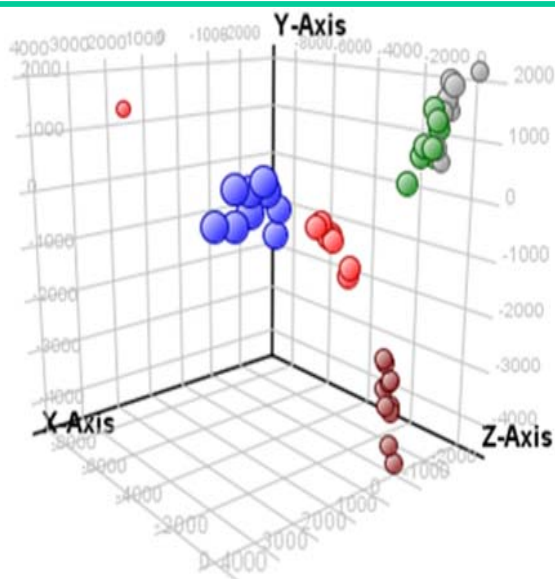


# Predicción de clases: PLS-DA

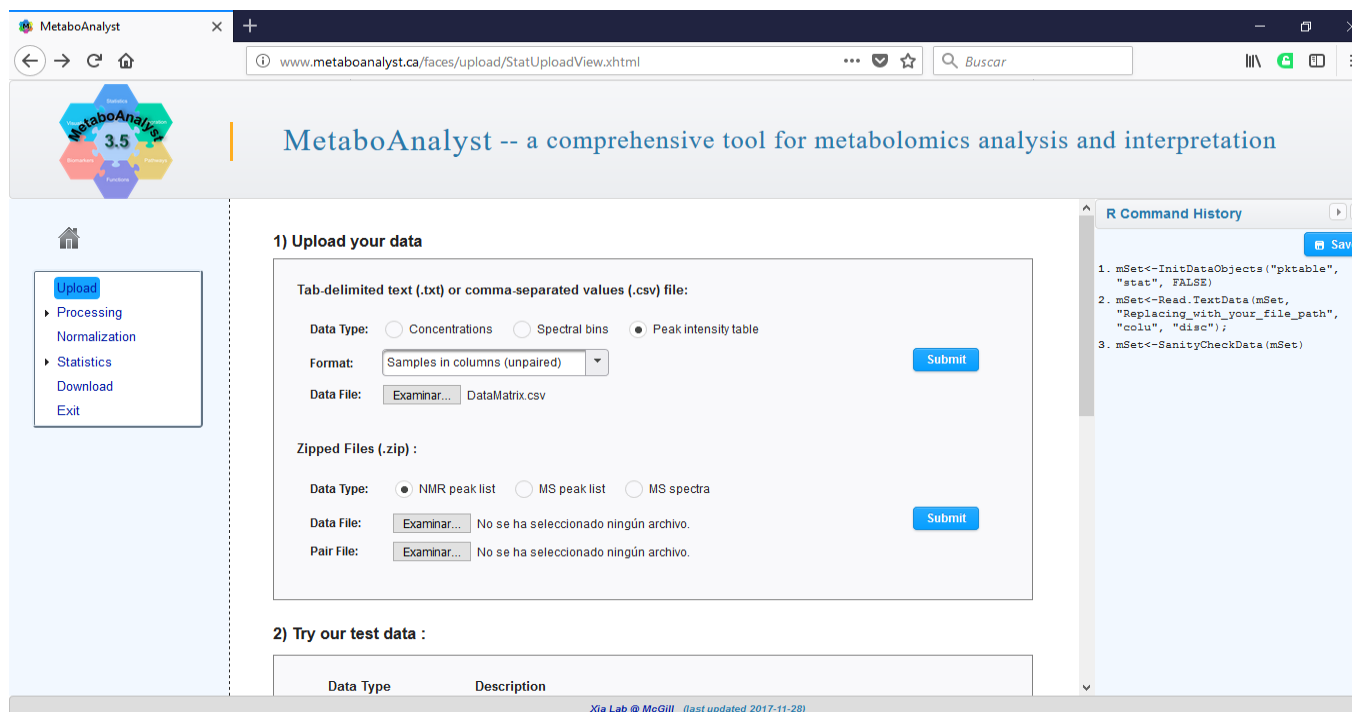
## Análisis Discriminante por Mínimos Cuadrados Parciales Análisis Discriminante por Proyección en Estructuras Latentes

un método estadístico que guarda cierta relación con el análisis de componentes principales (PCA), pero siendo un análisis *supervisado*

- o crea un modelo de regresión lineal proyectando las variables predichas y observables en un nuevo espacio
- o muy adecuado para cuando existen más indicadores (compuestos) que observaciones (muestras)
- o cada compuesto posee un *t-score* que representa su impacto en la predicción
- o se asigna un valor de confianza de la predicción al ejecutar el modelo



# Análisis estadístico univariante y multivariante



The screenshot displays the MetaboAnalyst 3.5 web interface. The browser address bar shows the URL `www.metaboanalyst.ca/faces/upload/StatUploadView.xhtml`. The page title is "MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation".

The interface is divided into several sections:

- Left Sidebar:** Contains navigation options: Upload, Processing, Normalization, Statistics, Download, and Exit.
- Main Content Area:**
  - 1) Upload your data:** This section is for uploading user data. It includes:
    - Tab-delimited text (.txt) or comma-separated values (.csv) file:** Options for Data Type (Concentrations, Spectral bins, Peak intensity table), Format (Samples in columns (unpaired)), and a Data File upload button.
    - Zipped Files (.zip):** Options for Data Type (NMR peak list, MS peak list, MS spectra), Data File, and Pair File upload buttons.
  - 2) Try our test data:** A section for testing the tool with pre-existing data, featuring a table with columns for Data Type and Description.
- Right Panel:** Titled "R Command History", it shows the R code used for the current session:

```
1. mSet<-InitDataObjects("pktable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_your_file_path", "colu", "disc");
3. mSet<-SanityCheckData(mSet)
```

The footer of the page reads "Xia Lab @ McGill (last updated 2017-11-28)".



# Predicción de clases: Validación del modelo

se evalúa la exactitud de la regla de predicción que se crea y proporciona una indicación de modelos de sobreajuste:

## dejando uno fuera (*leave one out*)

- se utilizan todas las muestras del conjunto de entrenamiento excepto una para construir la regla de predicción
- utilizando esta regla, se predice la clase de muestra que se ha dejado fuera
- la muestra se devuelve al conjunto de entrenamiento y se deja fuera una muestra diferente, volviendo a elaborar la regla de predicción con las muestras que quedan
- se repite este proceso hasta que se haya predicho cada muestra del conjunto de entrenamiento exactamente una vez
- se recuenta entonces el número de predicciones correctas e incorrectas para determinar el porcentaje de éxito

## $N$ - fold

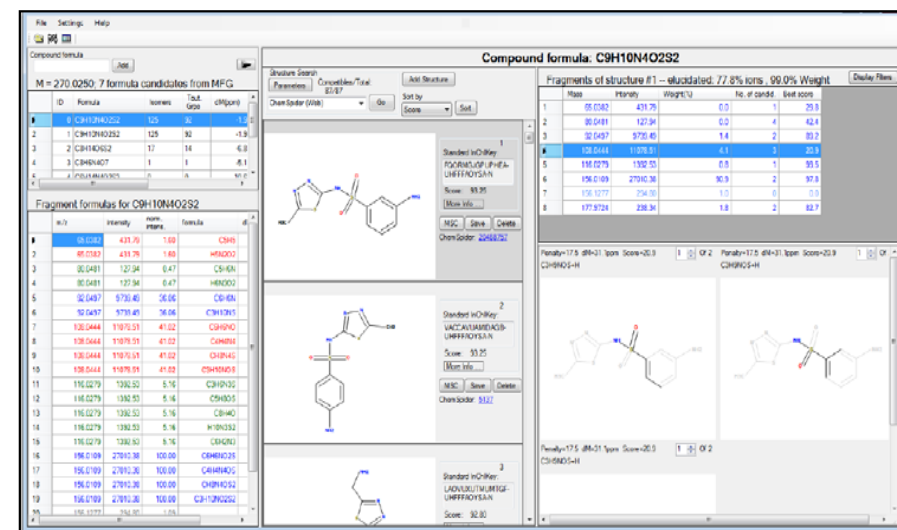
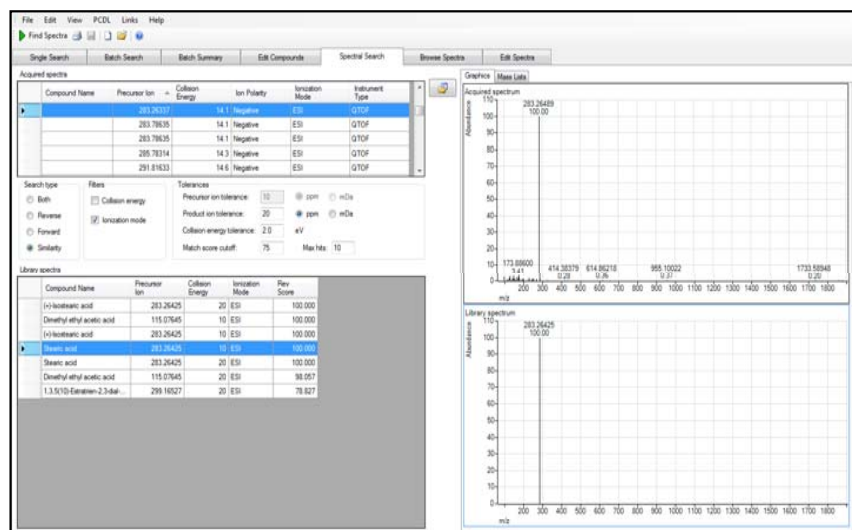
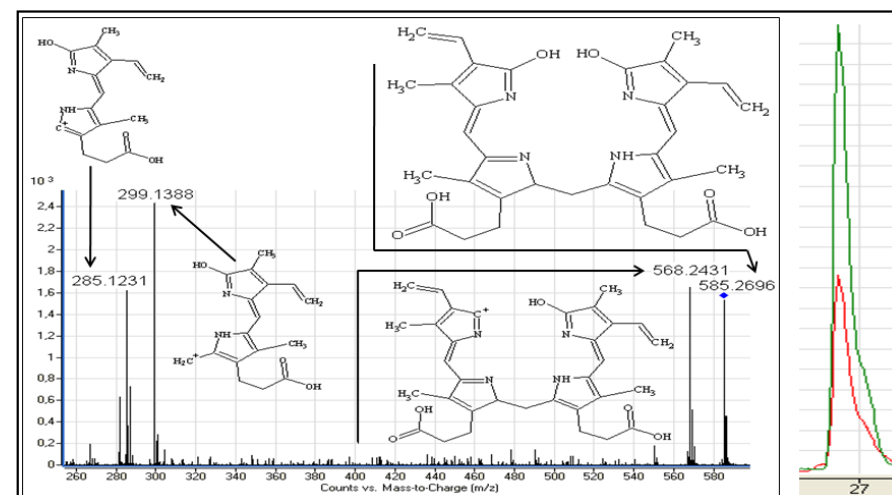
1. las muestras del conjunto de entrenamiento se dividen aleatoriamente en  $N$  subconjuntos iguales, manteniendo una frecuencia de clases relativa
2.  $N-1$  subconjuntos se combinan entonces para el entrenamiento y el conjunto sobrante se utiliza para el ensayo
3. se repite paso a paso dejando un grupo fuera cada vez
4. se repite el paso 1, 2, 3  $M$  veces
5. cada muestra se predice  $M$  veces y se comunica la clase que haya sido predicha más veces a lo largo de esas  $M$  veces en los resultados de validación



# Identificación

1. Búsqueda en base de datos con medida de masa exacta
2. Búsqueda en base de datos comparando el **patrón isotópico**
3. Búsqueda en base de datos comparando el patrón isotópico y el **tiempo de retención**
4. Búsqueda en librerías de MS/MS
5. Búsqueda en librerías de MS/MS y **tiempo de retención**

Confianza



# CLASIFICACIONES DE BASES DE DATOS

- **Basadas datos espectrales**
  - Principalmente pequeñas moléculas y no solo metabolitos
  - NMR
  - MS o MS/MS
- **Basadas en información de compuestos**
  - Nombre de compuesto, estructuras, propiedades físicas, identificación
- **Basadas en base de datos de rutas metabólicas**
  - Rutas de metabolitos, xenobióticos, proteínas y de señalización
- **Base de datos metabolómica completa**
  - Una combinación de las anteriores

# Lista de Bases de Datos en 2018

Nombre	URL	Nombre	URL
	<a href="http://aralip.plantbiology.msu.edu/pathways/pathways">http://aralip.plantbiology.msu.edu/pathways/pathways</a>		
ARALIP	<a href="http://aralip.plantbiology.msu.edu/pathways/pathways">ays</a>	KEGG	<a href="http://prime.psc.riken.jp/?action=metabolites_index">http://prime.psc.riken.jp/?action=metabolites_index</a>
AtIPD	<a href="http://www.atipd.ethz.ch/">http://www.atipd.ethz.ch/</a>	KEGG Glycan	<a href="http://www.genome.jp/kegg/glycan/">http://www.genome.jp/kegg/glycan/</a>
BiGG	<a href="http://bigg.ucsd.edu/">http://bigg.ucsd.edu/</a>	KNASAcK	<a href="http://prime.psc.riken.jp/?action=metabolites_index">http://prime.psc.riken.jp/?action=metabolites_index</a>
BioCyc	<a href="http://biocyc.org/">http://biocyc.org/</a>	LipidMaps	<a href="http://www.lipidmaps.org/">http://www.lipidmaps.org/</a>
BioNumbers	<a href="http://bionumbers.hms.harvard.edu/">http://bionumbers.hms.harvard.edu/</a>	MarkerDB	<a href="http://www.markerdb.ca/users/sign_in">http://www.markerdb.ca/users/sign_in</a>
BML-NMR	<a href="http://www.bml-nmr.org/">http://www.bml-nmr.org/</a>	MassBank	<a href="http://www.massbank.jp/">http://www.massbank.jp/</a>
BioMagResBank	<a href="http://www.bmrwisc.edu/metabolomics/">http://www.bmrwisc.edu/metabolomics/</a>	MetaboAnalyst	<a href="http://www.metaboanalyst.ca/MetaboAnalyst/">http://www.metaboanalyst.ca/MetaboAnalyst/</a>
BMDB	<a href="http://www.cowmetdb.ca/cgi-bin/browse.cgi">http://www.cowmetdb.ca/cgi-bin/browse.cgi</a>	MetaboLights	<a href="http://www.ebi.ac.uk/metabolights/index">http://www.ebi.ac.uk/metabolights/index</a>
ChEBI	<a href="http://www.ebi.ac.uk/chebi/">http://www.ebi.ac.uk/chebi/</a>	MetaCrop	<a href="http://metacrop.ipk-gatersleben.de/apex/f?p=269:111:">http://metacrop.ipk-gatersleben.de/apex/f?p=269:111:</a>
ChEMBL	<a href="https://www.ebi.ac.uk/chembl/about#">https://www.ebi.ac.uk/chembl/about#</a>	MetaCyc	<a href="http://metacyc.org/">http://metacyc.org/</a>
ChEBI	<a href="http://www.ebi.ac.uk/chebi/">http://www.ebi.ac.uk/chebi/</a>	METAGENE	<a href="http://www.metagene.de/program/a.prg">http://www.metagene.de/program/a.prg</a>
ChemMine	<a href="http://chemminedb.ucr.edu/">http://chemminedb.ucr.edu/</a>	METLIN	<a href="https://metlin.scripps.edu/index.php">https://metlin.scripps.edu/index.php</a>
ChemSpider	<a href="http://www.chemspider.com/">http://www.chemspider.com/</a>	MMCD	<a href="http://mmcd.nmrfa.wisc.edu/">http://mmcd.nmrfa.wisc.edu/</a>
CCD	<a href="http://ccd.chemnetbase.com/intro/index.jsp#about">http://ccd.chemnetbase.com/intro/index.jsp#about</a>	mzCloud	<a href="https://mzcloud.org/">https://mzcloud.org/</a>
CSF Metabolome Database	<a href="http://www.csfmetabolome.ca/">http://www.csfmetabolome.ca/</a>	OMIM	<a href="http://www.ncbi.nlm.nih.gov/omim/">http://www.ncbi.nlm.nih.gov/omim/</a>
CyberCell Database	<a href="http://ccdb.wishartlab.com/CCDB/">http://ccdb.wishartlab.com/CCDB/</a>	OMMBID	<a href="http://ommbid.mhmedical.com/">http://ommbid.mhmedical.com/</a>
DrugBank	<a href="http://www.drugbank.ca/">http://www.drugbank.ca/</a>	Oryzabase	<a href="http://www.shigen.nig.ac.jp/rice/oryzabase/">http://www.shigen.nig.ac.jp/rice/oryzabase/</a>
ECMDB	<a href="http://www.ecmdb.ca/">http://www.ecmdb.ca/</a>	PepBank	<a href="http://pepbank.mgh.harvard.edu/">http://pepbank.mgh.harvard.edu/</a>
ExPaSy Pathways	<a href="http://web.expasy.org/pathways/">http://web.expasy.org/pathways/</a>	PharmGKB	<a href="http://www.pharmgkb.org/">http://www.pharmgkb.org/</a>
	<a href="http://fiehnlab.ucdavis.edu/Metabolite-Library-2007/">http://fiehnlab.ucdavis.edu/Metabolite-Library-2007/</a>		
Fiehn GC-MS Database	<a href="http://fiehnlab.ucdavis.edu/Metabolite-Library-2007/">2007/</a>	PMN	<a href="http://www.plantcyc.org/">http://www.plantcyc.org/</a>
FoodB	<a href="http://www.foodb.ca">http://www.foodb.ca</a>	PubChem	<a href="http://pubchem.ncbi.nlm.nih.gov/">http://pubchem.ncbi.nlm.nih.gov/</a>
GMDB	<a href="http://gmd.mpimp-golm.mpg.de/">http://gmd.mpimp-golm.mpg.de/</a>	Reactome	<a href="http://www.reactome.org/">http://www.reactome.org/</a>
HMDB	<a href="http://metabolomics.pharm.uconn.edu/iimdb/">http://metabolomics.pharm.uconn.edu/iimdb/</a>	RiceCyc	<a href="http://pathway.gemstone.org/gramene/ricecyc.shtml">http://pathway.gemstone.org/gramene/ricecyc.shtml</a>
HumanCyc	<a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a>	Serum Metabolome Database	<a href="http://www.serummetabolome.ca/">http://www.serummetabolome.ca/</a>
IIDMB	<a href="http://www.genome.jp/kegg/glycan/">http://www.genome.jp/kegg/glycan/</a>	SetupX & BinBase	<a href="http://fiehnlab.ucdavis.edu/projects/binbase_setupx">http://fiehnlab.ucdavis.edu/projects/binbase_setupx</a>



# CEU MASS MEDIATOR

Seleccionar archivo Ningún archivo seleccionado

**Experimental Masses (\*):** enter significant input masses

**Retention Times:** enter significant retention times

**Composite Spectra:** enter significant composite spectra

---

Seleccionar archivo Ningún archivo seleccionado

**All Experimental Masses:** enter all input masses

**All Retention Times:** enter all retention times

**All Composite Spectra:** enter all composite spectra

**Chemical Alphabet (\*):**  
 All  
 CHNOPS  
 CHNOPS + Cl

**Modifiers (\*):**  
 None  
 NH3  
 HCOO  
 CH3COO  
 HCOONH3  
 CH3COONH3

**Databases (\*):**  
 All except MINE  
 All (Including In Silico Compounds)  
 Kegg  
 HMDB  
 LipidMaps  
 Metlin  
 MINE (Only In Silico Compounds)

**Metabolites (\*):**  
 All except peptides  
 Only lipids  
 All including peptides

**Input Masses Mode (\*):**  
 Neutral Masses  
 m/z Masses

**Ionization Mode (\*):**  
 Neutral  
 Positive Mode  
 Negative Mode

calculation of new m/z from neutral mass based on selected adducts

**Adducts (\*):**  
 All  
 M+H  
 M+2H  
 M+Na  
 M+K  
 M+NH4

- Dedicado a la anotación de metabolitos.
- Realiza búsquedas en compuestos unificados de distintas fuentes.
- Aplica conocimiento basado en la información de entrada proporcionada por el usuario.
- Ayuda a identificar lípidos oxidados.
- <http://ceumass.eps.uspceu.es/mediator>

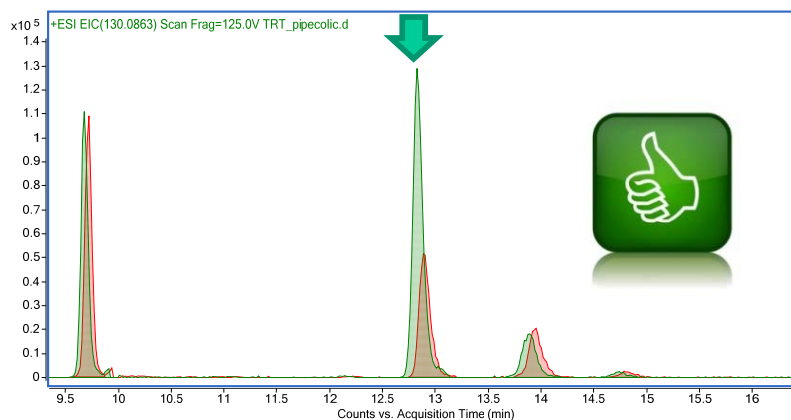
# CEU



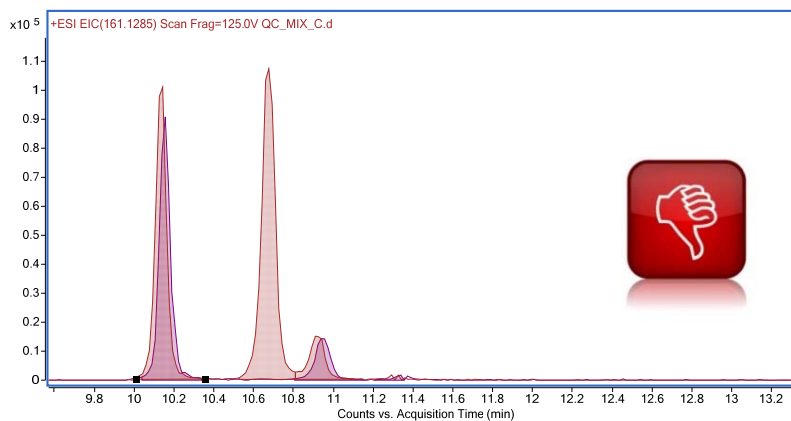
# MASS MEDIATOR

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
LIST OF COMPOUNDS	Experimental mass	Identifier Adduct	PPM Err	Molecular Weigh Name	Formula	CAS	Kegg	HMDB	LipidMaps	Metlin	PubChem	InChIKey	Pathways	
3	399.3367	17732 M-H	5	399.3349 L-palmitoylcarnitine	C23H45NO4				LME A07070079			XOMRRQKXKHMVMOC-NRFFANRHFSA-N		
4	399.3367	17751 M-H	5	399.3349 O-palmitoylcarnitine	C23H45NO4				LME A07070098			XOMRRQKXKHMVMOC-UHFFFAOYSA-N		
5	399.3367	17657 M-H	5	399.3349 Palmitoylcarnitine	C23H45NO4	2364-67-2	C02390	HMDB0000222	LME A07070004	36667	11953816	XOMRRQKXKHMVMOC-OAQYLSRUS	Fatty acid Metabolism Fatty a	
6	399.3367	0 M-Na	0	0 No compounds found for experimental mass 399.3367 and adduct: I										
7	399.3367	13442 M-NH4	5	382.3083 methyl 9-butylperoxy-10,12-octadecadienoate	C23H42O4				LME A01040036	74461		XVZNFQXKXKMPCH-ZIAPIVEJSA-N		
8	399.3367	13443 M-NH4	5	382.3083 methyl 13-butylperoxy-9,11-octadecadienoate	C23H42O4				LME A01040037	74462		GUVYTNFTPALBOJ-AEPWDTSSA-N		
9	399.3367	95769 M-NH4	5	382.3083 Lepidumterpenyl ester	C23H42O4			HMDB0036865		31868	11734320	PASMASQJCDKBJK-UHFFFAOYSA-N		
10	399.3367	53861 M-NH4	5	382.3083 MG(0.0/20.2(11Z,14Z)/0.0)	C23H42O4			HMDB0011544		62328	53480964	PMJSUEZTCFTBMD-HZJYTRNSA-N		
11	399.3367	52646 M-NH4	5	382.3083 MG(20.2(11Z,14Z)/0.0/0.0)	C23H42O4			HMDB0011574		62356	53480983	QRBGFYBOCBYOSN-KDTZJXJHSA-N		
12	399.3367	80631 M-NH4	5	382.3083 Pteronene B	C23H42O4			HMDB0035955		31145		NLXNQLZUQMHHEB-ISLRYVAYSAN		
13	399.3367	0 M-H-H2O	0	0 No compounds found for experimental mass 399.3367 and adduct: I										
14	421.3169	16968 M-H	6	421.3192 Gamma-linolyl carnitine	C25H43NO4			HMDB0006318	LME A07010893	58389	53477819	YDPMHMLMYSAQK-BAHSRKMSSA-N		
15	421.3169	96332 M-H	6	421.3192 Alpha-linolyl carnitine	C25H43NO4			HMDB0006319		58390	53477821	DFVGGGHKDAHYIU-UHMZJMFSA-N		
16	421.3169	126612 M-H	9	421.3205 AGELASINE	C26H39N5					43731				
17	421.3169	138401 M-H	9	421.313 Latanoprost ethyl amide-d4	C25H35D4NO4					36571				
18	421.3169	17732 M-Na	0	399.3349 L-palmitoylcarnitine	C23H45NO4				LME A07070079			XOMRRQKXKHMVMOC-NRFFANRHFSA-N		
19	421.3169	17751 M-Na	0	399.3349 O-palmitoylcarnitine	C23H45NO4				LME A07070098			XOMRRQKXKHMVMOC-UHFFFAOYSA-N		
20	421.3169	17657 M-Na	0	399.3349 Palmitoylcarnitine	C23H45NO4	2364-67-2	C02390	HMDB0000222	LME A07070004	36667	11953816	XOMRRQKXKHMVMOC-OAQYLSRUS	Fatty acid Metabolism Fatty a	
21	421.3169	1294 M-NH4	6	404.2927 alpha,25-dihydroxy-21-nor-20-oxavitamin D3 / alpha,25-dihydroxy-21	C25H40O4				LMST03020029	41970		ACMQZQKQVZ2TNY-OQGZSIESA-N		
22	421.3169	1295 M-NH4	6	404.2927 alpha,25-dihydroxy-24-nor-22-oxavitamin D3 / alpha,25-dihydroxy-2	C25H40O4				LMST03020030	41971		UMRLCGLUMINBCT-NPMMXCSOSA-N		
23	421.3169	2650 M-NH4	6	404.2927 7b-Hydroxy-3-oxo-5b-eholan-24-oate	C25H40O4				LMST04070028	57339		XHRLTYUHWGHDJ-JFRJXPMSA-N		
24	421.3169	117990 M-NH4	6	404.2927 Androstane-3,17-diol dipropionate,5alpha-Androstane-3alpha,17beta	C25H40O4	4350-14-5	C14624			70213	134572	FWAYUWSHYLKUEY-LFYVYRBCSA-N		
25	421.3169	86607 M-NH4	6	404.2927 11-Carboxy-gamma-chromanol	C25H40O4			HMDB0012517			53481453	ITULCNMOMXAH-YUULODDPSA-N		
26	421.3169	89679 M-NH4	6	404.2927 MG(0.0/22.5(7Z,10Z,13Z,16Z,19Z)/0.0)	C25H40O4			HMDB0011556		62339	53480971	LRBJLYDLUADN-JLNKQISASA-N		
27	421.3169	50261 M-NH4	6	404.2927 MG(22.5(4Z,7Z,10Z,13Z,16Z)/0.0/0.0)	C25H40O4			HMDB0011585		62367	53480993	HDIQCISTZKHUDD-AJWITYRPSA-N		
28	421.3169	105373 M-NH4	6	404.2927 MG(0.0/22.5(4Z,7Z,10Z,13Z,16Z)/0.0)	C25H40O4			HMDB0011555		62338	53480970	NPZVSBAEZLZYQU-WMPRHZDHSAN		
29	421.3169	56824 M-NH4	6	404.2927 MG(22.5(7Z,10Z,13Z,16Z,19Z)/0.0/0.0)	C25H40O4			HMDB0011586		62368	53480994	IDSLCYURGAOTDA-YAWQMQZOFSA-N		
30	421.3169	17695 M-H-H2O	5	439.3298 3-hydroxylinoleoylcarnitine	C25H45NO5				LME A07070042			WGYXCASYXUFNSI-UTJQPWESSAN		
31	315.2424	17712 M-H	5	315.241 Decanoylcarnitine	C17H33NO4			HMDB0000651	LME A07070059		10245190	LZOSYCMHQXPBFU-UHFFFAOYSA-N		
32	315.2424	126591 M-H	5	315.241 L-Hexanoylcarnitine n-butyl ester	C17H33NO4					3549				
33	315.2424	17659 M-H	5	315.241 O-decanoyl-R-carnitine	C17H33NO4	3992-45-8	C03239	HMDB0062631	LME A07070006	36669	11953821	LZOSYCMHQXPBFU-OAHLLOKOSA-N		
34	315.2424	0 M-Na	0	0 No compounds found for experimental mass 315.2424 and adduct: I										
35	315.2424	14277 M-NH4	5	298.2144 8E-Heptadecenedioic acid	C17H30O4				LME A01170053	74925		VDTSYDDUGXHLDY-OVQJBTEDSA-N		
36	315.2424	11819 M-NH4	5	298.2144 Plakortcic acid	C17H30O4			C17158		71590	10402441	ZCLJFHUIADAYRQ-CMDGGGBGSA-N		
37	315.2424	0 M-H-H2O	0	0 No compounds found for experimental mass 315.2424 and adduct: I										
38	337.2234	0 M-H	0	0 No compounds found for experimental mass 337.2234 and adduct: I										
39	337.2234	17712 M-Na	2	315.241 Decanoylcarnitine	C17H33NO4			HMDB0000651	LME A07070059		10245190	LZOSYCMHQXPBFU-UHFFFAOYSA-N		
40	337.2234	126591 M-Na	2	315.241 L-Hexanoylcarnitine n-butyl ester	C17H33NO4					3549				
41	337.2234	17659 M-Na	2	315.241 O-decanoyl-R-carnitine	C17H33NO4	3992-45-8	C03239	HMDB0062631	LME A07070006	36669	11953821	LZOSYCMHQXPBFU-OAHLLOKOSA-N		
42	337.2234	1060 M-NH4	6	320.1988 testolic acid	C19H28O4			C01618	LMST02020081	57817	439534	KMUJXIPRPXRP-PT-DZBHQSCQSA-N		
43	337.2234	116962 M-NH4	6	320.1988 10beta-Hydroxy-6beta-isobutyryl furanocremophilane	C19H28O4			C09685		67884	442377	WVBNALQGLGJMUUJ-BIGGFVSDSA-N		
44	337.2234	130809 M-NH4	6	320.1988 (+)-alpha-CMBHC	C19H28O4	7083-09-2				44834				
45	337.2234	53386 M-NH4	6	320.1988 [8]-Gingerdione	C19H28O4	77334-06-6		HMDB003276		33864	14440537	QDSRAFNZQKMPZ-UHFFFAOYSA-N		
46	337.2234	47665 M-NH4	6	320.1988 5-Carboxy-alpha-chromanol	C19H28O4			HMDB0012738			53481524	QWPNLVBFAZJBMIMEVZPHKSA-N		

# Confirmación mediante Adición estándar



Ácido pipercolico



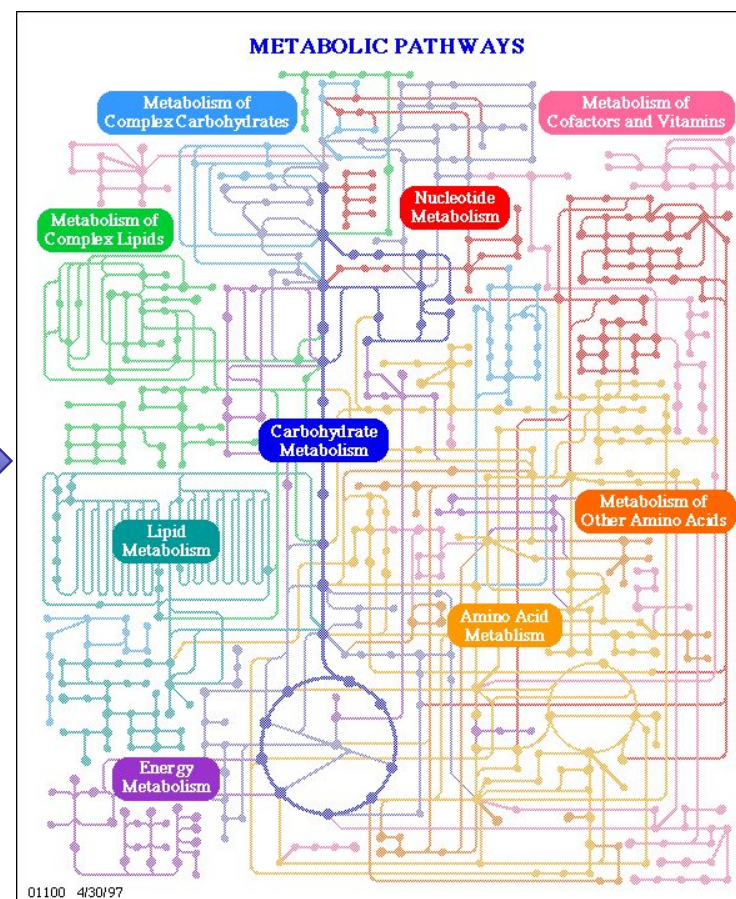
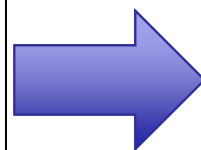
Metil-lisina



# De Listas a Rutas

## metabolómica

Compound	Retention Time (min)	Conc. in Urine (µM)	Compound	Retention Time (min)	Conc. in Urine (µM)
Dns-o-phospho -L-serine	0.92	<D.L.*	Dns-Ile	6.35	25
Dns-o-phospho -L-tyrosine	0.95	<D.L.	Dns-3-aminosalicylic acid	6.44	0.5
Dns-adenosine monophosphate	0.99	<D.L.	Dns-pipecolic acid	6.50	0.5
Dns-o-phosphoethanolamine	1.06	16	Dns-Leu	6.54	54
Dns-glucosamine	1.06	22	Dns-cystathionine	6.54	0.3
Dns-o-phospho -L-threonine	1.09	<D.L.	Dns-Leu-Pro	6.60	0.4
Dns-6-dimet hylamine purine	1.20	<D.L.	Dns-5-hydroxylysine	6.65	1.6
Dns-3-methyl -histidine	1.22	80	Dns-Cystine	6.73	160
Dns-taurine	1.25	834	Dns-N-norleucine	6.81	0.1
Dns-carnosine	1.34	28	Dns-5-hydroxydopamine	7.17	<D.L.
Dns-Arg	1.53	36	Dns-dimethylamine	7.33	293
Dns-Asn	1.55	133	Dns-5-HIAA	7.46	18
Dns-hypotaurine	1.58	10	Dns-umbelliferone	7.47	1.9
Dns-homocarnosine	1.61	3.9	Dns-2,3 -diam inopropionic acid	7.63	<D.L.
Dns-guanidine	1.62	<D.L.	Dns-L-ornithine	7.70	15
Dns-Gln	1.72	633	Dns-4-acetyamidophenol	7.73	51
Dns-allantoin	1.83	3.8	Dns-procaine	7.73	8.9
Dns-L-citrulline	1.87	2.9	Dns-homocystine	7.76	3.3
Dns-1 (or 3 -)-methylhistamine	1.94	1.9	Dns-acetaminophen	7.97	82
Dns-adenosine	2.06	2.6	Dns-Phe-Phe	8.03	0.4
Dns-methylguanidine	2.20	<D.L.	Dns-5-methyo xysalicylic acid	8.04	2.1
Dns-Ser	2.24	511	Dns-Lys	8.16	184
Dns-aspartic acid amide	2.44	26	Dns-aniline	8.17	<D.L.
Dns-4-hydroxy -proline	2.56	2.3	Dns-leu-Phe	8.22	0.3
Dns-Glu	2.57	21	Dns-His	8.35	1550
Dns-Asp	2.60	90	Dns-4-thialysine	8.37	<D.L.
Dns-Thr	3.03	157	Dns-benzylamine	8.38	<D.L.
Dns-ephedrine	3.05	<D.L.	Dns-1-ephedrine	8.50	0.6
Dns-ethanolamine	3.11	471	Dns-tryptamine	8.63	0.4
Dns-aminoadipic acid	3.17	70	Dns-pyridoxamine	8.94	<D.L.
Dns-Gly	3.43	2510	Dns-2-methyl -benzylamine	9.24	<D.L.
Dns-Ala	3.88	593	Dns-5-hydroxytrptophan	9.25	0.12
Dns-aminolevulinic acid	3.97	30	Dns-1,3-diaminopropane	9.44	0.23
Dns-r-amino -butyric acid	3.98	4.6	Dns-putrescine	9.60	0.5
Dns-p-amino -hippuric acid	3.98	2.9	Dns-1,2-diaminopropane	9.66	0.1
Dns-5-hydro xymethyluril	4.58	1.9	Dns-tyrosinamide	9.79	29
Dns-tryptophanamide	4.70	5.5	Dns-dopamine	10.08	140
Dns-isoguanine	4.75	<D.L.	Dns-cadaverine	10.08	0.08
Dns-5-aminopentanoic acid	4.79	1.6	Dns-histamine	10.19	0.4
Dns-sarcosine	4.81	7.2	Dns-3-methoxy -tyr amine	10.19	9.2
Dns-3-amino -isobutyrate	4.81	85	Dns-Tyr	10.28	321
Dns-2-aminobutyric acid	4.91	17	Dns-cysteamine	10.44	<D.L.



# Bases de datos de rutas metabólicas

- Excelente fuente de datos biológicos que sirve para relacionar metabolitos con genes, proteínas y enfermedades, señalando eventos y procesos
- Ofrece distintas herramientas que permiten la visualización y el mapeo de genes/metabolitos
- Normalmente cubre múltiples especies
- KEGG ([www.genome.jp/kegg/](http://www.genome.jp/kegg/)), BioCyc/MetaCyc (<https://biocyc.org/>), SMPDB ([www.smpdb.ca](http://www.smpdb.ca)), Reactome ([www.reactome.org](http://www.reactome.org)), WikiPathways (<http://www.wikipathways.org>)...
- «*De forma estricta, se podría afirmar que no existen las rutas metabólicas... son solo redes.*» (WikiPathways.org)



# KEGG – Enciclopedia de Genes y Genomas de Kioto

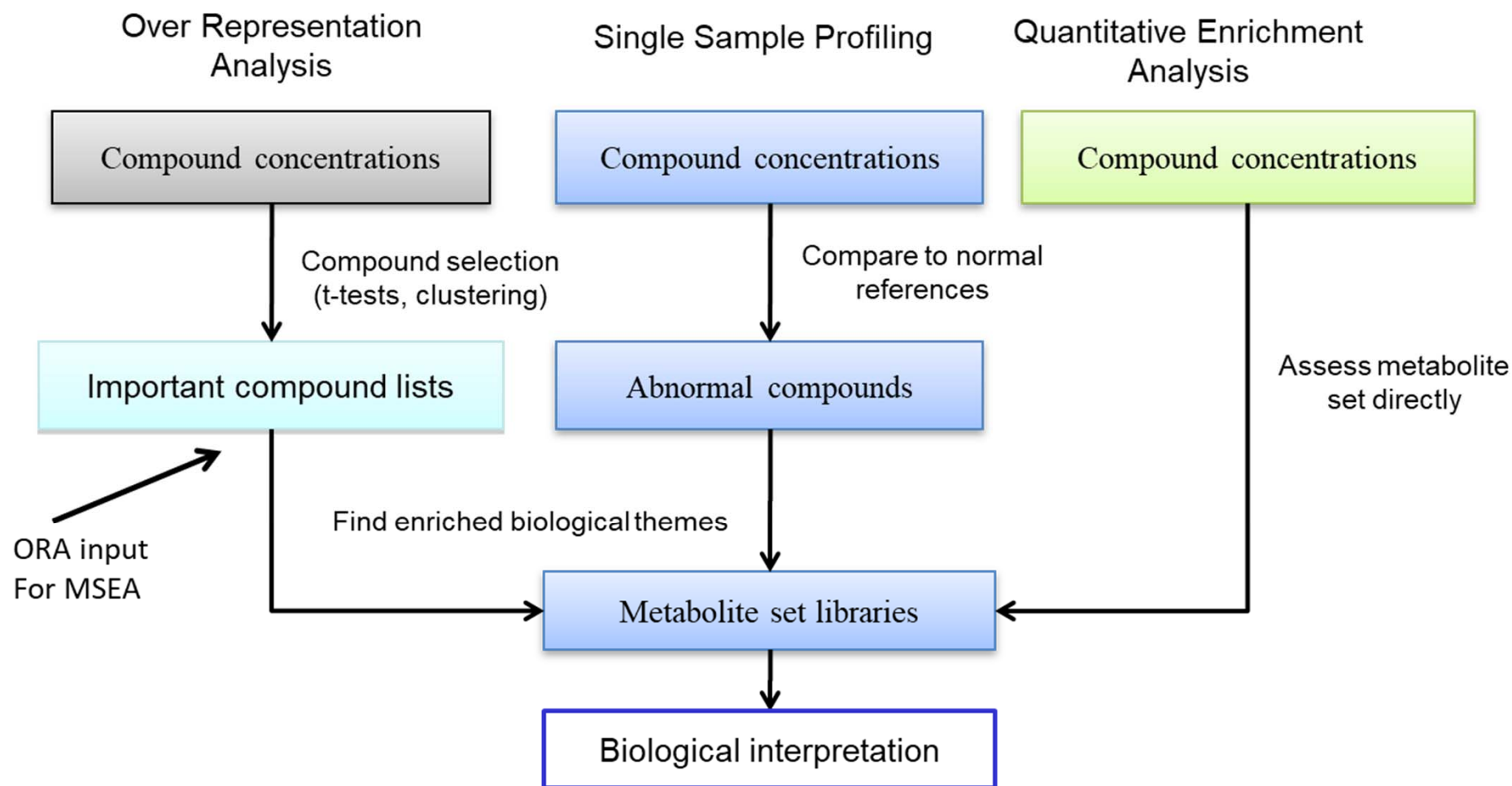
The image shows a screenshot of a web browser displaying the KEGG database. The left window shows the 'Pentose phosphate pathway - Reference pathway' map, with a box highlighting the 'PENTOSE PHOSPHATE PATHWAY' section. The right window shows the 'DBGET Result: COMPOUND C00198' details.

**COMPOUND: C00198**

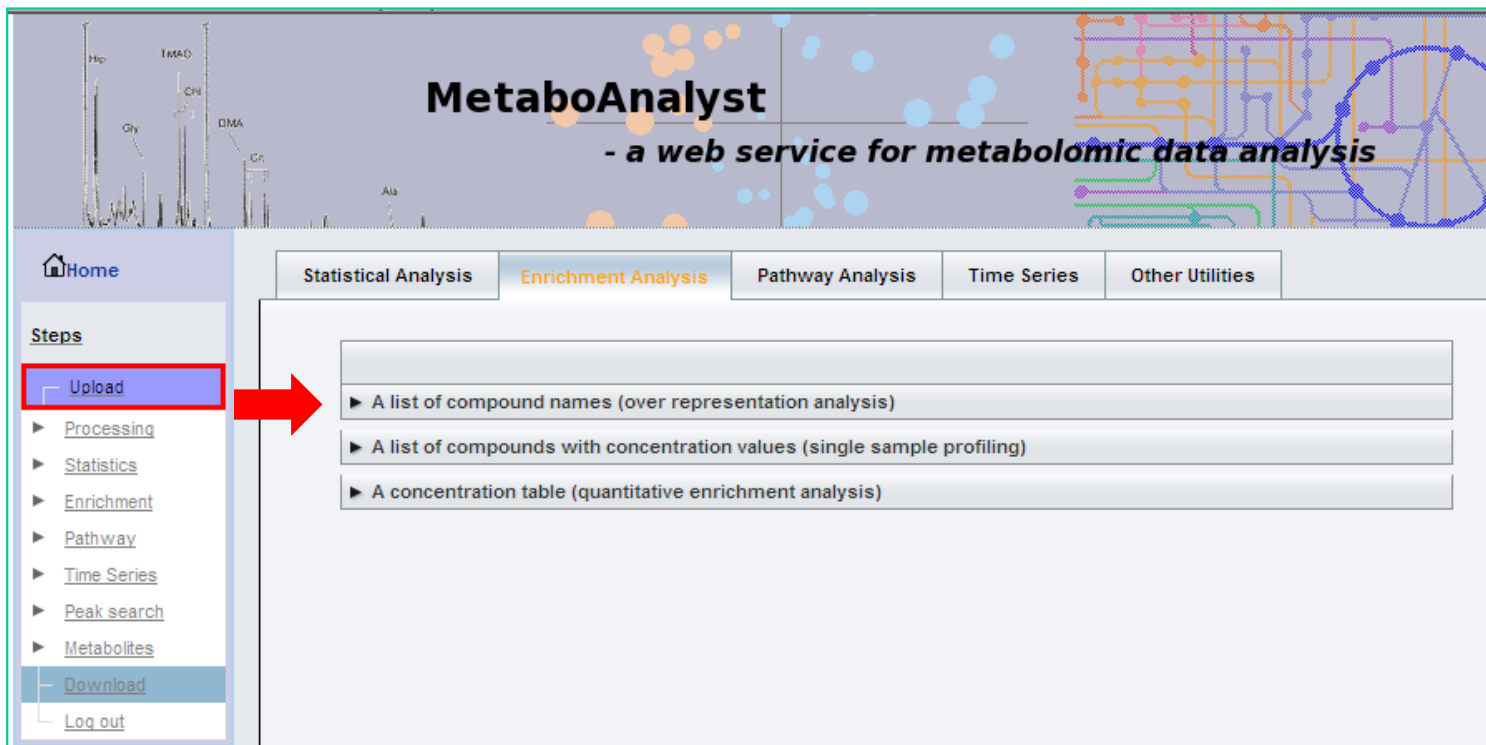
Entry	C00198	Compound												
Name	D-glucono-1,5-lactone; Gluconic lactone; Gluconic acid lactone; 1,5-gluconolactone; delta-Gluconolactone; D-gluconolactone; Glucarolactone; D-Aldonolactone; D-threo-Aldono-1,5-lactone													
Formula	C6H10O6													
Mass	178.0478													
Structure	<p>C00198</p> <p>Mol file KCF file DB search</p>													
Reaction	R00300 R00305 R00874 R01519 R01520 R01521 R01522 R02847 R03861 R06620													
Pathway	PATH: map00030 Pentose phosphate pathway													
Enzyme	<table border="1"> <tr> <td>1.1.1.47</td> <td>1.1.1.118</td> <td>1.1.1.121</td> <td>1.1.3.4</td> </tr> <tr> <td>1.1.3.5</td> <td>1.1.5.2</td> <td>1.1.99.10</td> <td>1.1.99.28</td> </tr> <tr> <td>2.2.1.122</td> <td>2.2.1.173</td> <td>2.2.1.17</td> <td></td> </tr> </table>		1.1.1.47	1.1.1.118	1.1.1.121	1.1.3.4	1.1.3.5	1.1.5.2	1.1.99.10	1.1.99.28	2.2.1.122	2.2.1.173	2.2.1.17	
1.1.1.47	1.1.1.118	1.1.1.121	1.1.3.4											
1.1.3.5	1.1.5.2	1.1.99.10	1.1.99.28											
2.2.1.122	2.2.1.173	2.2.1.17												

<http://www.genome.jp/kegg/>

# El enfoque MSEA (*Metabolite Set Enrichment Analysis*)



# Empezar con una lista de compuestos



**MetaboAnalyst**  
- a web service for metabolomic data analysis

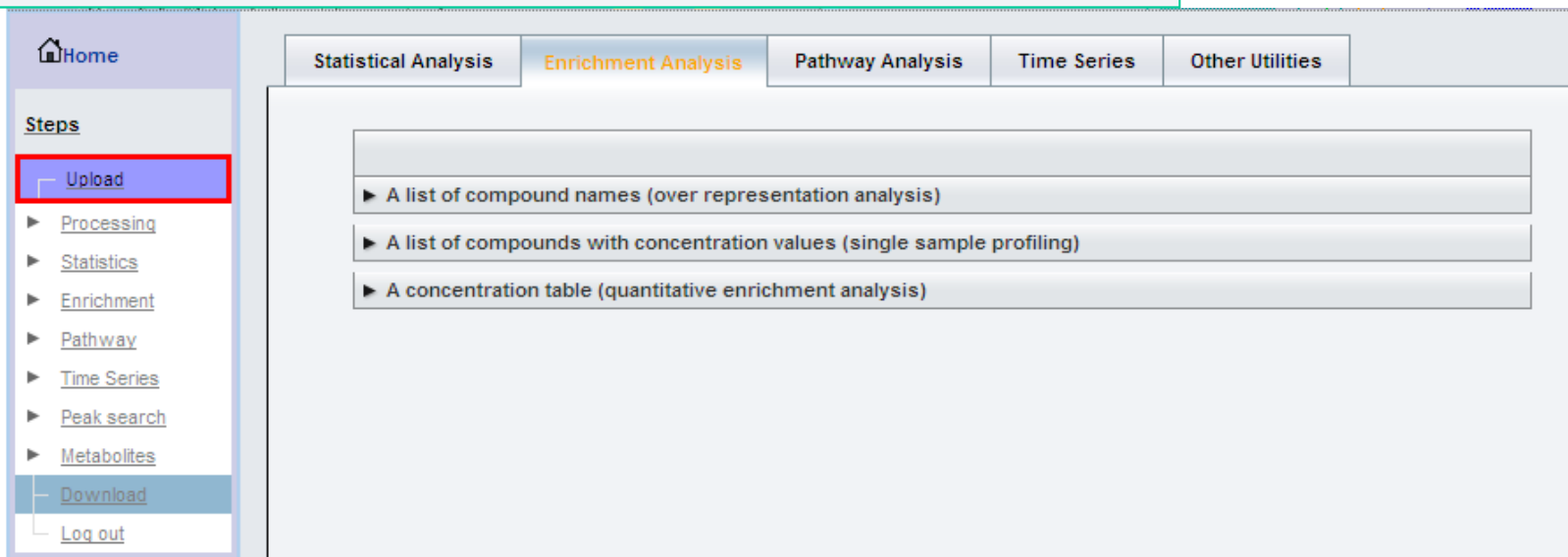
Home

Steps

- Upload
- Processing
- Statistics
- Enrichment
- Pathway
- Time Series
- Peak search
- Metabolites
- Download
- Log out

Statistical Analysis Enrichment Analysis Pathway Analysis Time Series Other Utilities

- A list of compound names (over representation analysis)
- A list of compounds with concentration values (single sample profiling)
- A concentration table (quantitative enrichment analysis)



Home

Steps

- Upload
- Processing
- Statistics
- Enrichment
- Pathway
- Time Series
- Peak search
- Metabolites
- Download
- Log out

Statistical Analysis Enrichment Analysis Pathway Analysis Time Series Other Utilities

- A list of compound names (over representation analysis)
- A list of compounds with concentration values (single sample profiling)
- A concentration table (quantitative enrichment analysis)

# Comparación con la concentración










Home


Steps

- Upload
- Processing
  - Pre-process
  - Name check
  - Conc. check**
  - Data check
  - Missing value
  - Data filter
  - Data editor
  - Color picker
  - Normalization
- Statistics
- Enrichment
- Pathway
- Time Series
- Peak search
- Metabolites
- Download
- Log out

## Comparison with Reference Concentration

Note: *reference concentrations* are in the form of **mean(min - max)** format. In cases where the ranges were not reported in the original literature, the min and max were calculated using the 95% confidence intervals. In the *Comparison* column, **H, M, L** means **higher, medium (within range), lower** compared to the reference concentrations. Click the **Image Icon** link to see a graphical summary for the comparisons.

Compound	Concentration	Reference Concentrations	Comparison	Detail	Include
L-Isoleucine	0.34	1.579 (0.789 - 2.368); 0.94 (0.27 - 1.61); 3.75 (1 - 6.5); 3 (1.5 - 4.5); 1.8 (0.8 - 2.8)	M		<input type="checkbox"/>
Fumaric acid	0.47	10.4 (2.8 - 53.7); 0.5 (0.1 - 1.7); 1 (0 - 2); 0.95 (0.02 - 1.88); 0.8 (0.1 - 1.7); 10.7 (0.1 - 28.2); 4.8 (0 - 35.2); 5 (1 - 33.5)	M		<input type="checkbox"/>
Acetone	0.58	4.2 (0.98 - 15.3); 0.92 (0.2 - 2.8); 320 (103 - 1290); 20 (2 - 180); 15.3 (2 - 120)	M		<input type="checkbox"/>
Succinic acid	9.4	14.4 (9.5 - 19.3); 3.8 (1.25 - 6.7); 12.6 (0.47 - 24.73); 14.48 (11.28 - 17.68); 9.9 (4.9 - 14.9); 39 (37 - 41); 197.2 (29.4 - 486.2); 185.4 (6 - 342.6); 7.7 (1.9 - 20); 11.6 (4 - 27.3); 8.25 (0.5 - 16)	M		<input type="checkbox"/>
1-Methylhistidine	9.6	2.3 (0 - 7.4); 33.6 (0 - 70); 28.1 (0 - 59.9); 30 (0 - 73); 45.5 (3.9 - 87.1); 1.3 (0 - 4.06); 4.6 (1.9 - 7.3); 46.1 (0 - 99.6); 15.9 (0 - 35.4)	M		<input type="checkbox"/>
L-Asparagine	19.62	35 (16.4 - 57.2); 9.211 (3.289 - 15.1); 0.96 (0.31 - 1.61); 10 (4.6 - 16.32)	M		<input type="checkbox"/>
3-Methylhistidine	9.7	42.76 (19.92 - 65.6); 15.1 (3.9 - 26.3); 12.5 (8.3 - 16.7)	M		<input type="checkbox"/>
L-Threonine	93.19	36.2 (10.82 - 61.58); 12.7 (4.934 - 20.4); 1 (0.16 - 2.4); 4.9 (2.4 - 7.4); 16 (7 - 25); 18 (8.4 - 27.6)	H		<input checked="" type="checkbox"/>
Creatine	720	46 (9 - 135); 113 (0 - 654); 26 (5 - 95); 167 (124 - 210); 212 (0 - 5000); 450 (0 - 10000)	M		<input type="checkbox"/>



# Análisis de Enriquecimiento Cuantitativo

## (Quantitative enrichment analysis)

Home

Statistical Analysis **Enrichment Analysis** Pathway Analysis Time Series Other Utilities

Steps

- Upload
- Processing
- Statistics
- Enrichment
- Pathway
- Time Series
- Peak search
- Metabolites
- Download
- Log out

A list of compound names (over representation analysis)

A list of compounds with concentration values (single sample profiling)

A concentration table (quantitative enrichment analysis)

**Upload your concentration data (.csv)**

[Format](#)

Browse...

Compound Label Type:

Phenotype Label:

Submit

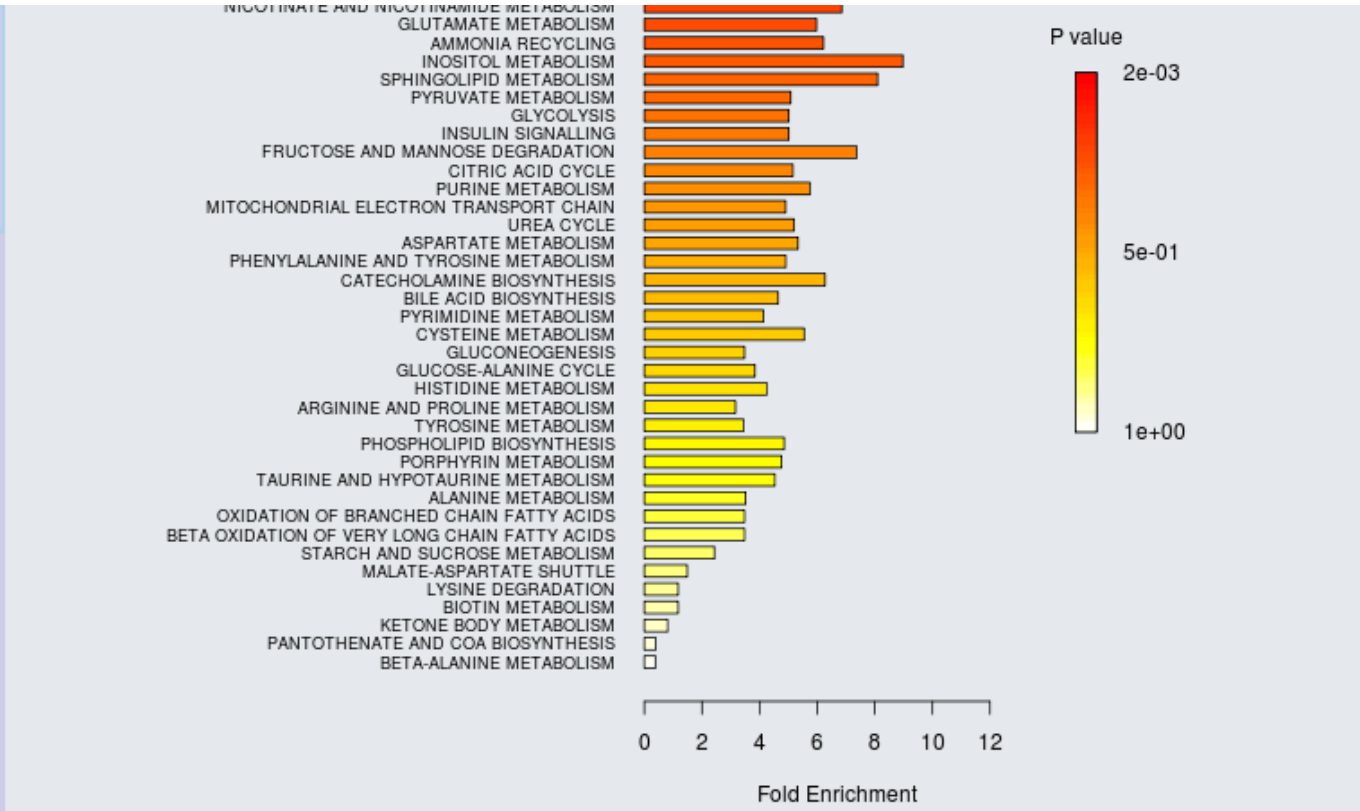
Try our test data:





Data	Compound	Phenotype	Description
<input checked="" type="radio"/> <a href="#">Data 1</a>	Common name	Discrete	Urinary metabolite concentrations from 77 cancer patients measured by 1H NMR. Phenotype: <b>N</b> - cachexic; <b>Y</b> - control
<input type="radio"/> <a href="#">Data 2</a>	PubChem CID	Continuous	Urinary metabolite concentrations from 97 cancer patients measured by 1H NMR. Phenotype: <b>muscle gain</b> (percentage within 100 days, negative values indicate muscle loss)

Submit

# RESULTADO

- ▶ [Time Series](#)
- ▶ [Peak search](#)
- ▶ [Metabolites](#)
- └─ [Download](#)
- └─ [Log out](#)



Metabolite Set	Total	Hit	Statistic	Expected	P Value	Holm P	FDR	Details
<b>TRYPTOPHAN METABOLISM</b>	34	2	15.088	1.3158	5.3712E-5	0.0024707	0.0020529	
<b>PROANOATE METABOLISM</b>	18	1	17.695	1.3158	1.3942E-4	0.0062741	0.0020529	
<b>BETAINE METABOLISM</b>	10	2	14.311	1.3158	1.4515E-4	0.0063865	0.0020529	
<b>METHIONINE METABOLISM</b>	24	4	11.386	1.3158	1.7852E-4	0.0076762	0.0020529	



# Metaboanalyst

## Metabolic Pathway Analysis (MetPA)

- Objetivo: extender y potenciar el MSEA para rutas metabólicas mediante
  - La consideración de las **estructuras de ruta metabólicas**
  - La visualización dinámica de las rutas metabólicas
- Actualmente admite aprox. 1500 rutas que comprenden 17 organismos (basadas en KEGG)

# SESIÓN PRÁCTICA. VISUALS\_1



## Workflow4metabolomics



Main menu

- Home
- Events
- History
- Introduction
  - The Galaxy environment
  - The LC-MS workflow
  - The GC-MS workflow
  - The NMR workflow
  - References
- HowTo
- Download
  - Datasets
- Referenced Workflows and Histories
- How to contribute?
- Developer resources
  - Source code
  - Virtual environments
- People
- Publications

### Workflow4Metabolomics 3.0

Welcome to the collaborative portal dedicated to metabolomics data processing, analysis and annotation for Metabolomics community.

*"We are happy to announce the next **Workflow4Experimenters (W4E) international course 2018: Using Galaxy and the Workflow4metabolomics infrastructure to analyse metabolomics data.** Please save the date: **8-12 October 2018** at Pasteur Institute, Paris - France. More news in December!"*

<b>STEP 1</b> Request an account 	<b>STEP 2</b> Connect to Galaxy 	<b>STEP 3</b> Upload your data 	<b>STEP 4</b> Make reproducible science 
--	---	--	---

About MSConvert

Filter:

Transfer Vendor:

MS Levels:

---

Parameters

peakPicking	true 1-
-------------	---------

---

Output Directory:

Options

Output format:  Extension:

Binary encoding precision:  64-bit  32-bit

Write index:  Use zlib compression:

TPP compatibility:  Package in gzip:

Use numpress linear compression:

Use numpress short logged float compression:

Use numpress short positive integer compression:

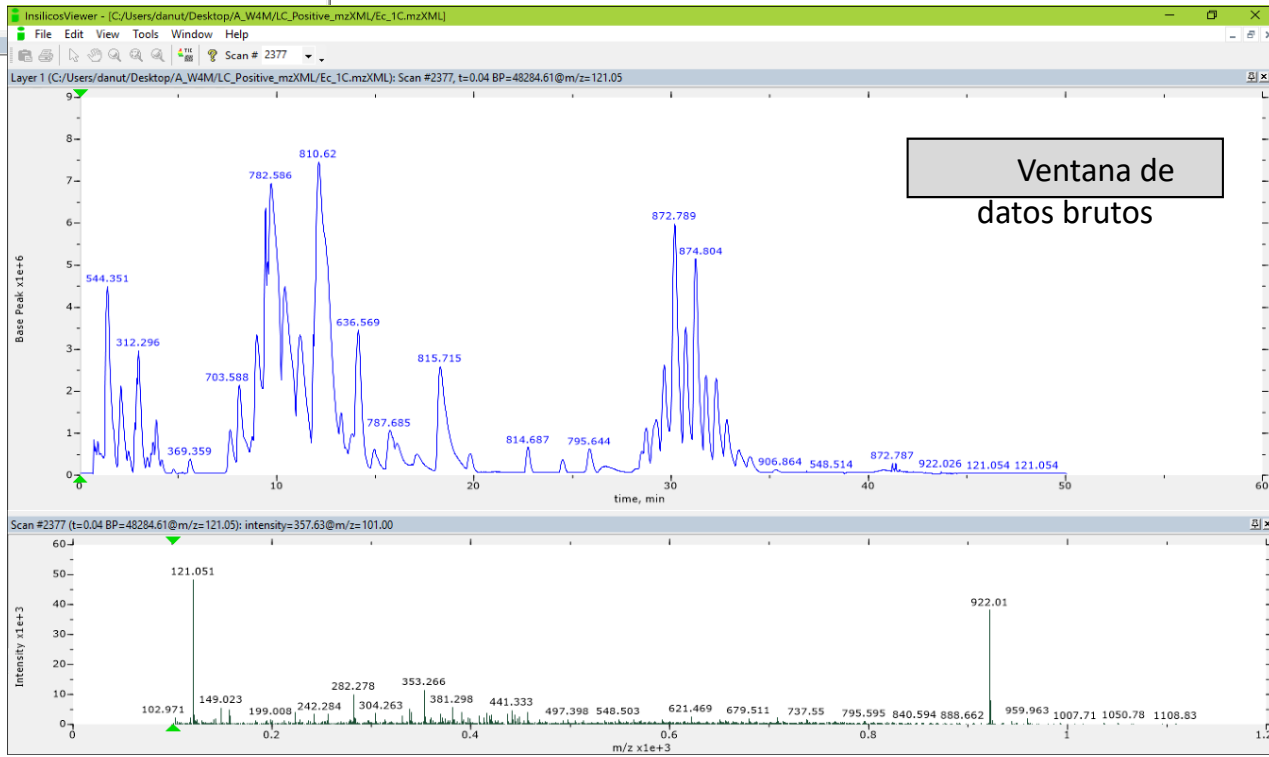


# SESIÓN PRÁCTICA. VISUALS\_2

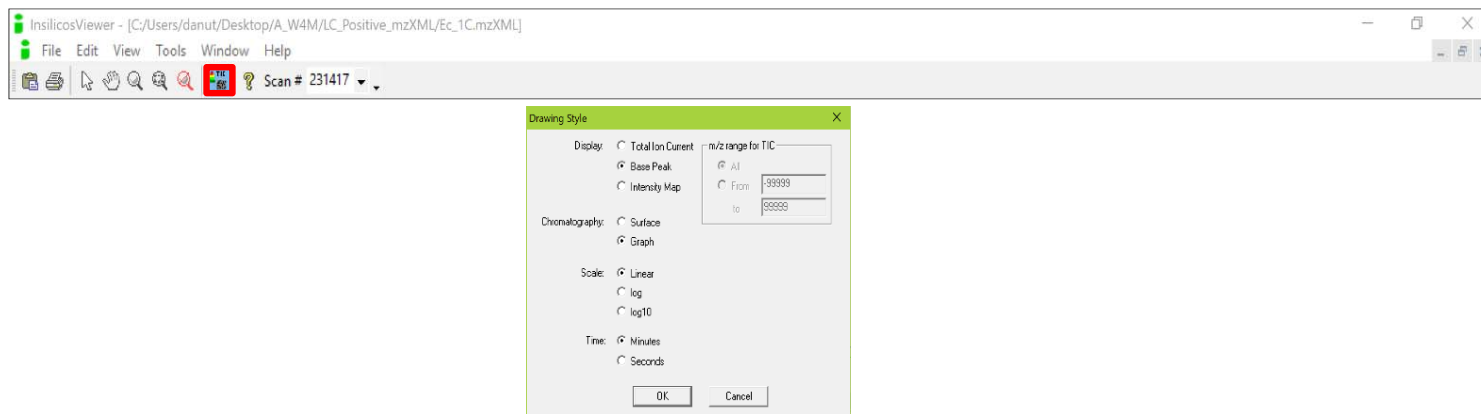
The screenshot shows the InsilicosViewer application window. The menu bar includes File, Edit, View, Tools, Window, and Help. The 'File' menu is open, showing options like New, Open..., Close, Print..., Print Preview, Print Setup..., and Exit. The main area displays the 'licos Life Science Software' logo. A file list window is open, showing a directory named 'LC\_Positive\_mzXML' with a table of files:

Nazwa	Data modyfikacji	Typ
Ec_1C	19.4.17 13:47	mass spec
Ec_2C	19.4.17 13:49	mass spec
Ec_3C	19.4.17 13:49	mass spec
Ec_4C	19.4.17 13:50	mass spec
Ec_5C	19.4.17 13:50	mass spec
Ec_6C	19.4.17 13:50	mass spec
Ec_7C	19.4.17 13:50	mass spec
Ec_8C	19.4.17 13:51	mass spec
Ec_9HC	19.4.17 13:51	mass spec
Ec_10HC	19.4.17 13:43	mass spec
Ec_11HC	19.4.17 13:43	mass spec
Ec_12HC	19.4.17 13:44	mass spec
Ec_13HC	19.4.17 13:44	mass spec
Ec_14HC	19.4.17 13:45	mass spec
Ec_15HC	19.4.17 13:45	mass spec

Below the list, there are fields for 'Nazwa pliku:' (set to 'Ec\_1C') and 'Pliki typu:' (set to 'MS Data Files (\*.mzXML;\*.mzData;\*.mzML;\*.RA)'), along with 'Otwórz' and 'Anuluj' buttons.



# SESIÓN PRÁCTICA. VISUALS\_3



Main menu

- Home
- Events
- History
- Introduction
  - The Galaxy environment
  - The LC-MS workflow
  - The GC-MS workflow
  - The NMR workflow
  - References
- HowTo
- Download
  - Datasets
  - Referenced WorkFlows and Histories
  - How to contribute?
- Developer resources
  - Source code
  - Virtual environments
- People
- Publications

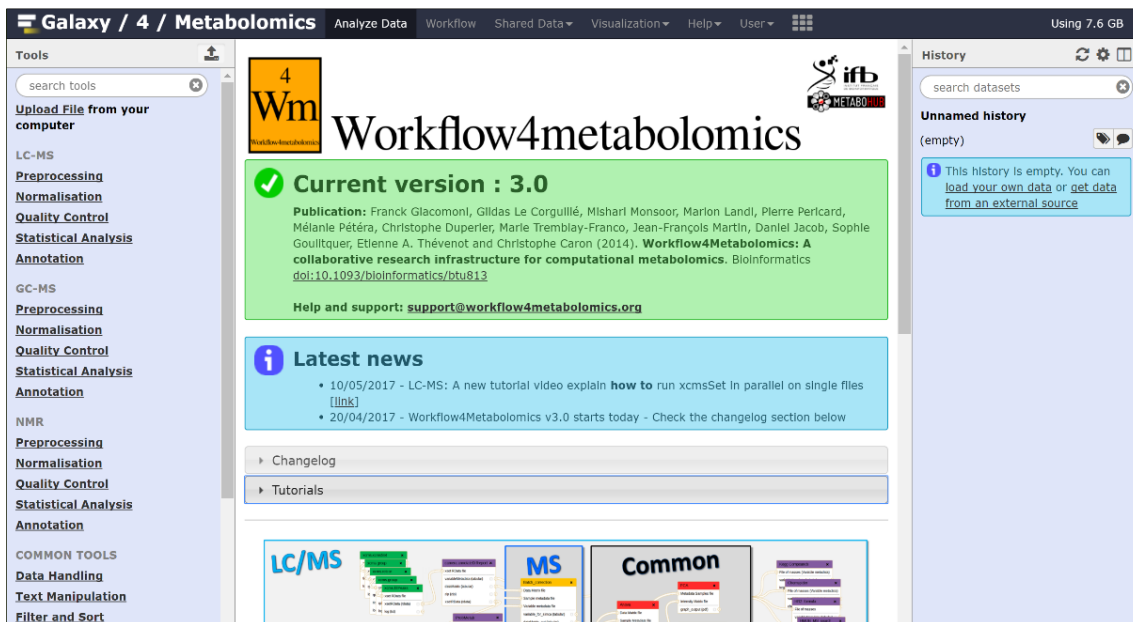
## Workflow4Metabolomics 3.0

Welcome to the collaborative portal dedicated to metabolomics data processing, analysis and annotation for Metabolomics community.

*"We are happy to announce the next **Workflow4Experimenters (W4E) international course 2018: Using Galaxy and the Workflow4metabolomics infrastructure to analyse metabolomics data.** Please save the date: **8-12 October 2018** at Pasteur Institute, Paris - France. More news in December!"*

<p><b>STEP 1</b> Request an account</p> 	<p><b>STEP 2</b> Connect to Galaxy</p> 	<p><b>STEP 3</b> Upload your data</p> 	<p><b>STEP 4</b> Make reproducible science</p> 
---	--	---	--

# SESIÓN PRÁCTICA. VISUALS\_4



**Galaxy / 4 / Metabolomics** Analyze Data Workflow Shared Data Visualization Help User Using 7.6 GB

**Tools**

search tools

Upload File from your computer

LC-MS

Preprocessing  
Normalisation  
Quality Control  
Statistical Analysis  
Annotation

GC-MS

Preprocessing  
Normalisation  
Quality Control  
Statistical Analysis  
Annotation

NMR

Preprocessing  
Normalisation  
Quality Control  
Statistical Analysis  
Annotation

COMMON TOOLS

Data Handling  
Text Manipulation  
Filter and Sort

**Workflow4metabolomics**

**Current version : 3.0**

**Publication:** Franck Glacomoni, Gildas Le Corguillé, Mishari Monsoor, Marion Landi, Pierre Pericard, Mélanie Pétéra, Christophe Duperter, Marie Tremblay-Franco, Jean-François Martin, Daniel Jacob, Sophie Goulltquer, Etienne A. Thévenot and Christophe Caron (2014). **Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics**. *Bioinformatics* doi:10.1093/bioinformatics/btu813

**Help and support:** [support@workflow4metabolomics.org](mailto:support@workflow4metabolomics.org)

**Latest news**

- 10/05/2017 - LC-MS: A new tutorial video explain **how to** run xcmsSet in parallel on single files [\[link\]](#)
- 20/04/2017 - Workflow4Metabolomics v3.0 starts today - Check the changelog section below

Changelog

Tutorials

LC/MS

MS

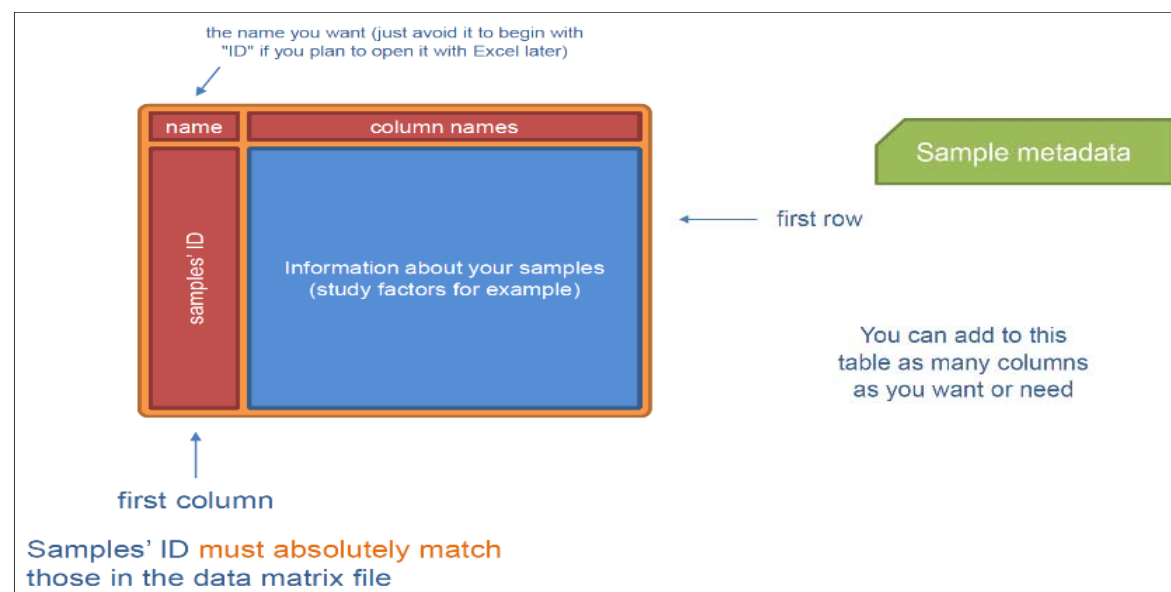
Common

History

search datasets

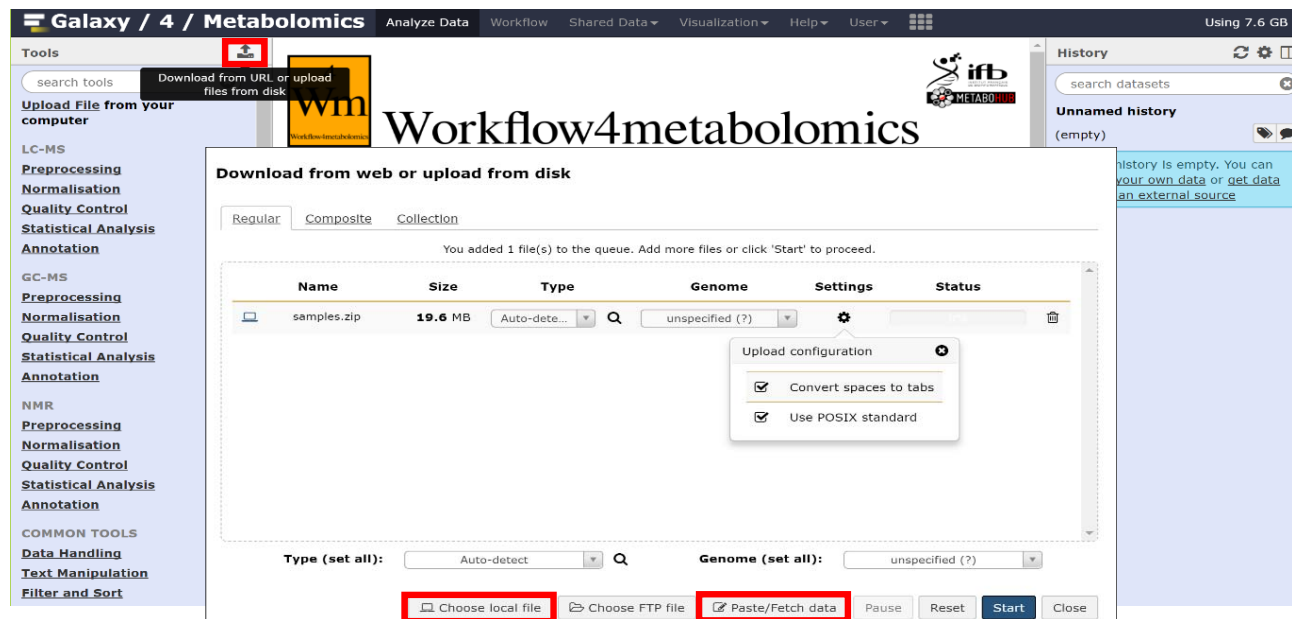
Unnamed history (empty)

This history is empty. You can [load your own data](#) or [get data from an external source](#)



# SESIÓN PRÁCTICA. VISUALS\_5

sampleName	class	polarity	sampleType	batch	injectionOrder	diet
QC	one	positive	pool	B1	1	NA
C1	one	positive	sample	B1	7	C
HC3	one	positive	sample	B1	10	HC
BL	one	positive	blank	B1	12	NA
...	...	...	...	...	...	...



Galaxy / 4 / Metabolomics

Tools

Download from URL or upload files from disk

Upload File from your computer

LC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

GC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

NMR

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

COMMON TOOLS

Data Handling

Text Manipulation

Filter and Sort

Workflow4metabolomics

Download from web or upload from disk

Regular Composite Collection

You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
samples.zip	19.6 MB	Auto-dete...	unspecified (?)		

Upload configuration

- Convert spaces to tabs
- Use POSIX standard

Type (set all): Auto-detect

Genome (set all): unspecified (?)

Choose local file Choose FTP file Paste/Fetch data Pause Reset Start Close

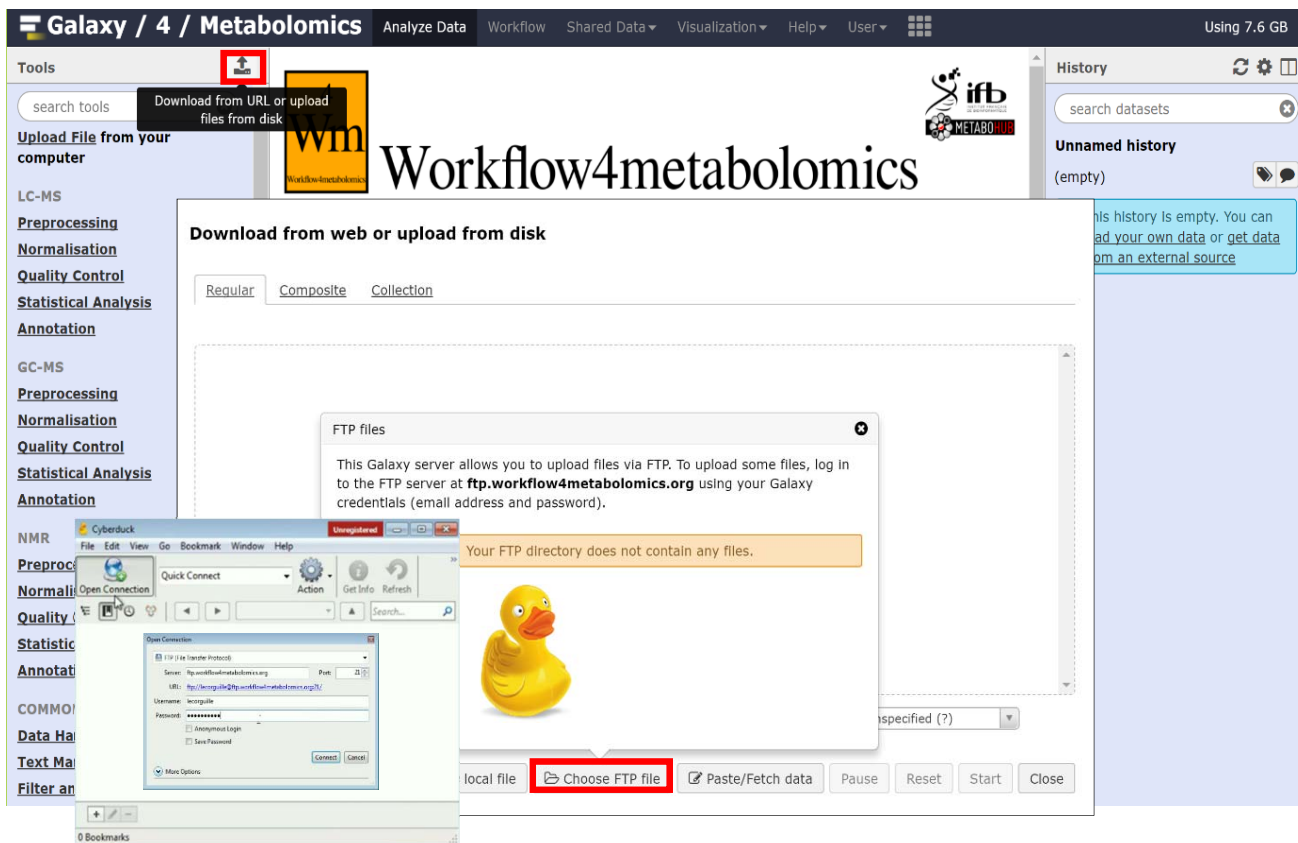
History

search datasets

Unnamed history (empty)

history is empty. You can your own data or get data an external source

# SESIÓN PRÁCTICA. VISUALS\_6



Galaxy / 4 / Metabolomics Analyze Data Workflow Shared Data Visualization Help User Using 7.6 GB

Tools Download from URL or upload files from disk

Upload File from your computer

LC-MS

Preprocessing  
Normalisation  
Quality Control  
Statistical Analysis  
Annotation

GC-MS

Preprocessing  
Normalisation  
Quality Control  
Statistical Analysis  
Annotation

NMR

Preprocessing  
Normalisation  
Quality Control  
Statistical Analysis  
Annotation

COMMON

Data Handling  
Text Manipulation  
Filtering

History

search datasets

Unnamed history (empty)

Workflow4metabolomics

Download from web or upload from disk

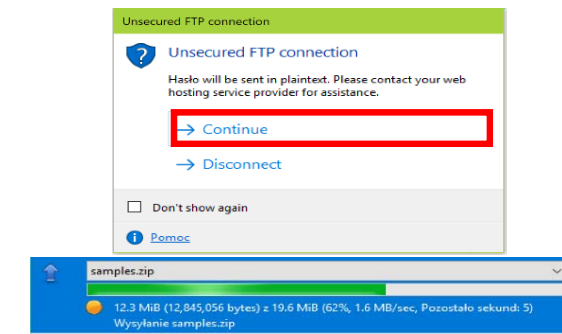
Regular Composite Collection

FTP files

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at **ftp.workflow4metabolomics.org** using your Galaxy credentials (email address and password).

Your FTP directory does not contain any files.

local file Choose FTP file Paste/Fetch data Pause Reset Start Close



Unsecured FTP connection

Unsecured FTP connection

Haslo will be sent in plaintext. Please contact your web hosting service provider for assistance.

Continue

Disconnect

Don't show again

Pomoc

samples.zip

12.3 MiB (12,845,056 bytes) z 19.6 MiB (62%, 1.6 MB/sec, Pozostalo sekund: 5)

Wysyłanie samples.zip

# SESIÓN PRÁCTICA. VISUALS\_7

**Download from web or upload from disk**

Regular Composite Collection

You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
samples.zip	19.6 MB				

FTP files

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at **ftp.workflow4metabolomics.org** using your Galaxy credentials (email address and password).

Available files: 1 files 19.6 MB

Name	Size	Created
samples.zip	19.6 MB	11/22/2017 05:09:03 PM

Type (set all): unspecified (?)

Choose local file Choose FTP file Paste/Fetch data Pause Reset **Start** Close

**Download from web or upload from disk**

Regular Composite Collection

Name	Size	Type	Genome	Settings	Status
LC_Positive_mzXML.zip	3.6 GB	Auto-dete...	unspecified (?)		100% ✓

Type (set all): Auto-detect Genome (set all): unspecified (?)

Choose local file Choose FTP file Paste/Fetch data Pause Reset Start **Close**

# SESIÓN PRÁCTICA. VISUALS\_8

**Galaxy / 4 / Metabolomics** Analyze Data Workflow Shared Data Visualization Help User Using 12.0 GB

**Workflow4metabolomics**

**Current version : 3.0**

**Publication:** Franck Giacconi, Gildas Le Corguillé, Mishari Monsoor, Marlon Landi, Pierre Pericard, Mélanie Pétéra, Christophe Duperier, Marie Tremblay-Franco, Jean-François Martin, Daniel Jacob, Sophie Goultiquier, Etienne A. Thévenot and Christophe Caron (2014). **Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics - Bioinformatics**. doi:10.1093/bioinformatics/btu813

**Help and support:** [support@workflow4metabolomics.org](mailto:support@workflow4metabolomics.org)

**Latest news**

- 10/05/2017 - LC-MS: A new tutorial video explain [how to run xcms](#) (link)
- 20/04/2017 - Workflow4Metabolomics v3.0 starts today - Check

**Changelog**

- 3.0.0 - 20/04/2017
  - LC-MS
    - Preprocessing
      - UPGRADE - xcms.\* (2.1.0): upgrade the 1.44.0 to 1.46.0
      - NEW** - xcms.\* (2.1.0): The W4M tools v take as input a single file. It will allow to several files and merge them afterward

**Upload File**

**Dataset Information**

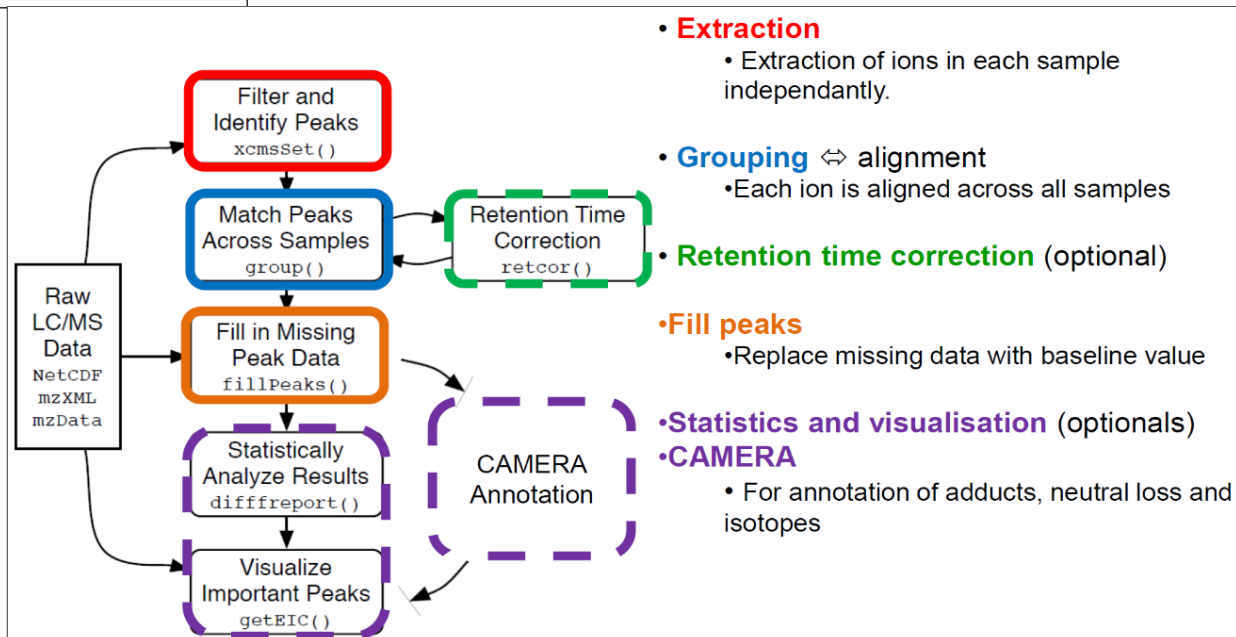
Number:	1
Name:	LC_Positive_mzXML.dp
Created:	Fri 24 Nov 2017 12:19:10 AM (UTC)
Filename:	0 bytes
DSkey:	?
Format:	txt

**Job Information**

Galaxy tool ID:	uploaded1
Galaxy Tool Version:	1.1.4
Tool Version:	
Tool Standard Output:	stdout
Tool Standard Error:	stderr
Tool Exit Code:	0
History Content API ID:	8d399c9b131f13b23
Job API ID:	c97c514f0105c63e
History API ID:	4e4095c119339f69
UUID:	9194cc24-ac90-410a-810c-410c79b5534

**Tool Parameters**

Input Parameter	Value	Note for rerun
file format:	auto	
async datasets:	None	
Specify Files for Dataset (auto):	1 uploaded datasets	
Genome:	unspecified (?)	
File Format:	auto	





# SESIÓN PRÁCTICA. VISUALS\_9

Galaxy / 4 / Metabolomics Analyze Data Workflow Shared Data Visualization Help User Using 15.6 GB

Tools

search tools

Upload File from your computer

LC-MS

Preprocessing

xcms.xcmsSet Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis

xcms.xcmsSet Merger Merge xcms.xcmsSet xset in one to be used by group

xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.

xcms.retcor Retention Time Correction using retcor function from xcms R package

xcms.fillPeaks Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas

xcms.summary Create a

1 job has been successfully added to the queue - resulting in the following datasets:

- 1: LC\_Positive\_mzXML.L.zip
- 2: LC\_Positive\_mzXML.xset.RData
- 3: LC\_Positive\_mzXML.sampleMetadata.tsv
- 4: LC\_Positive\_mzXML.xset.TICs\_raw.pdf
- 5: LC\_Positive\_mzXML.xset.BPCs\_raw.pdf
- 6: LC\_Positive\_mzXML.xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

Unnamed history

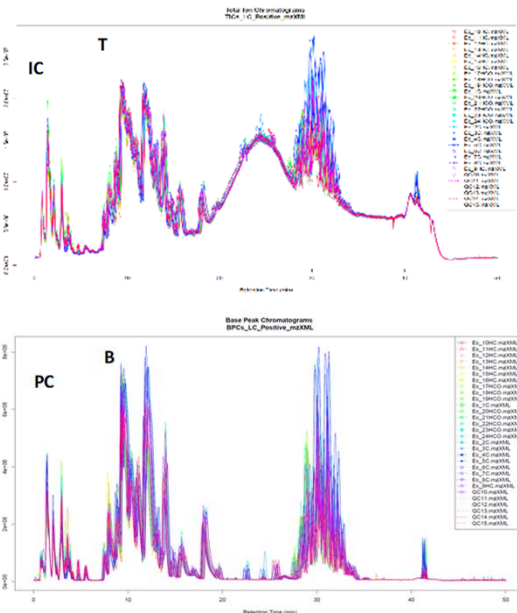
6 shown

3.61 GB

- 6: LC\_Positive\_mzXML.xset.log.txt
- 5: LC\_Positive\_mzXML.xset.BPCs\_raw.pdf
- 4: LC\_Positive\_mzXML.xset.TICs\_raw.pdf
- 3: LC\_Positive\_mzXML.sampleMetadata.tsv
- 2: LC\_Positive\_mzXML.xset.RData
- 1: LC\_Positive\_mzXML.L.zip

sample

1	Metadata	3
sampleMetadata	class	polarity
Ec_10HC	LC_Positive_mzXML	positive
Ec_11HC	LC_Positive_mzXML	positive
Ec_12HC	LC_Positive_mzXML	positive
Ec_13HC	LC_Positive_mzXML	positive
Ec_14HC	LC_Positive_mzXML	positive
Ec_15HC	LC_Positive_mzXML	positive
Ec_16HC	LC_Positive_mzXML	positive
Ec_17HCO	LC_Positive_mzXML	positive
Ec_18HCO	LC_Positive_mzXML	positive
Ec_19HCO	LC_Positive_mzXML	positive
Ec_1C	LC_Positive_mzXML	positive
Ec_20HCO	LC_Positive_mzXML	positive
Ec_21HCO	LC_Positive_mzXML	positive
Ec_22HCO	LC_Positive_mzXML	positive
Ec_23HCO	LC_Positive_mzXML	positive
Ec_24HCO	LC_Positive_mzXML	positive
Ec_2C	LC_Positive_mzXML	positive
Ec_3C	LC_Positive_mzXML	positive
Ec_4C	LC_Positive_mzXML	positive
Ec_5C	LC_Positive_mzXML	positive
Ec_6C	LC_Positive_mzXML	positive
Ec_7C	LC_Positive_mzXML	positive
Ec_8C	LC_Positive_mzXML	positive
Ec_9HC	LC_Positive_mzXML	positive
QC10	LC_Positive_mzXML	positive
QC11	LC_Positive_mzXML	positive
QC12	LC_Positive_mzXML	positive
QC13	LC_Positive_mzXML	positive
QC14	LC_Positive_mzXML	positive
QC15	LC_Positive_mzXML	positive



info

Data processing

PACKAGE INFO

```
parallel 3.2.5
BioGenerics 0.38.1
Biobase 2.30.0
Prop 0.32.10
mR 2.4.1
aces 1.60.0
time 0.6.2
batch 1.1.4
```

ARGUMENTS INFO

```
zipfile /usr/local/project/edu/galaxy/metabolomics/galaxy-dist/database/files/000/439/dataset_439795.dat
xfunction xcmsSet
xsetDataOutput /usr/local/project/edu/galaxy/metabolomics/galaxy-dist/database/files/000/440/dataset_440044.dat
sampleMetadata /usr/local/project/edu/galaxy/metabolomics/galaxy-dist/database/files/000/440/dataset_440044.dat
file /usr/local/project/edu/galaxy/metabolomics/galaxy-dist/database/files/000/440/dataset_440044.dat
file /usr/local/project/edu/galaxy/metabolomics/galaxy-dist/database/files/000/440/dataset_440044.dat
method centroidave
ppm 25
offset -136.640
width 0.0015
subsearch 10
integrate 1
noise 0
prefilter <(), 1000
```

INFIL PROCESSING INFO

ARGUMENTS PROCESSING INFO

```
files_root_directory LC_Positive_mzXML
Compute m/z checksum...
Checking file structure...
Checking files filenames compatibility with mcs...
```

- 6: LC\_Positive\_mzXML.xset.log.txt
- 5: LC\_Positive\_mzXML.xset.BPCs\_raw.pdf
- 4: LC\_Positive\_mzXML.xset.TICs\_raw.pdf
- 3: LC\_Positive\_mzXML.sampleMetadata.tsv
- 2: LC\_Positive\_mzXML.xset.RData
- 1: LC\_Positive\_mzXML.L.zip



# SESIÓN PRÁCTICA. VISUALS\_10

`xcms.group` Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.

		pool1B1			pool1B2			pool1B3		
		mz	rt	int	mz	rt	int	mz	rt	int
Listas de picos independientes		196.0905	66.6	7810936	196.0910	66.7	11733921	196.0902	66.6	7933325
		158.1180	67.4	71736	342.0310	69.0	74594	158.1173	67.4	82969
		342.0308	67.6	202268	267.0581	65.5	260877	342.0308	21.3	2581
		267.0581	65.5	282039	283.0318	65.2	424631	283.0320	65.3	357448
Iones del grupo por m/z		196.0905	66.6	7810936	196.0910	66.7	11733921	196.0902	66.6	7933325
		158.1180	67.4	71736	342.0310	69.0	74594	158.1173	67.4	82969
		342.0308	67.6	202268	267.0581	65.5	260877	342.0308	21.3	2581
		267.0581	65.5	282039	283.0318	65.2	424631	283.0320	65.3	357448
Iones del grupo por RT		196.0905	66.6	7810936	196.0910	66.7	11733921	196.0902	66.6	7933325
		158.1180	67.4	71736				158.1173	67.4	82969
		342.0308	67.6	202268	342.0310	69.0	74594	342.0308	21.3	2581
		267.0581	65.5	282039	267.0581	65.5	260877			
		<b>Matriz</b>								
		mz	rt	pool1B1	pool1B2	pool1B3				
		196.0905	66.6	7810936	11733921	7933325				
		158.1176	67.4	71736		82969				
		342.0308	21.3			2581				
		342.0309	68.3	202268	74594					
		267.0581	65.5	282039	260877					
		283.0319	65.2		424631	357448				

Parameter : num + label	Format
Or : RData file	rdata.xcms.raw
Or : RData file	rdata.xcms.retcor

# SESIÓN PRÁCTICA. VISUALS\_11

**xcms.group** Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time. (Galaxy Version 2.1.0) Versions Options

**xset RData file**

No rdata.xcms.raw, rdata.xcms.group, rdata.xcms.retcor or rdata dataset available. ▼  
output file from another function xcms (xcmsSet, retcor etc.)

**Method to use for grouping**

density ▼  
[method] See the help section below

**Bandwidth**

30  
[bw] bandwidth (standard deviation or half width at half maximum) of gaussian smoothing kernel to apply to the peak density chromatogram

**Minimum fraction of samples necessary**

0.5  
[minfrac] in at least one of the sample groups for it to be a valid group

**Width of overlapping m/z slices**

0.01  
[mzwid] to use for creating peak density chromatograms and grouping peaks across samples

**Advanced options**

show ▼

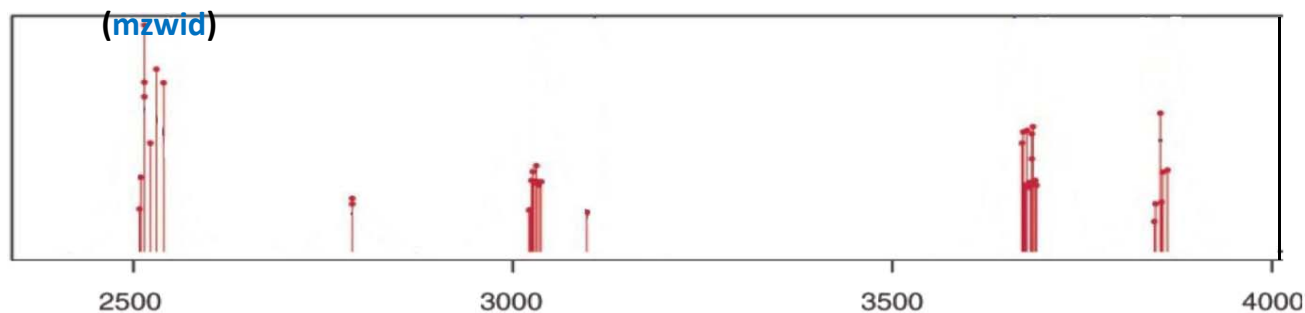
**Maximum number of groups to identify in a single m/z slice**

50  
[max]

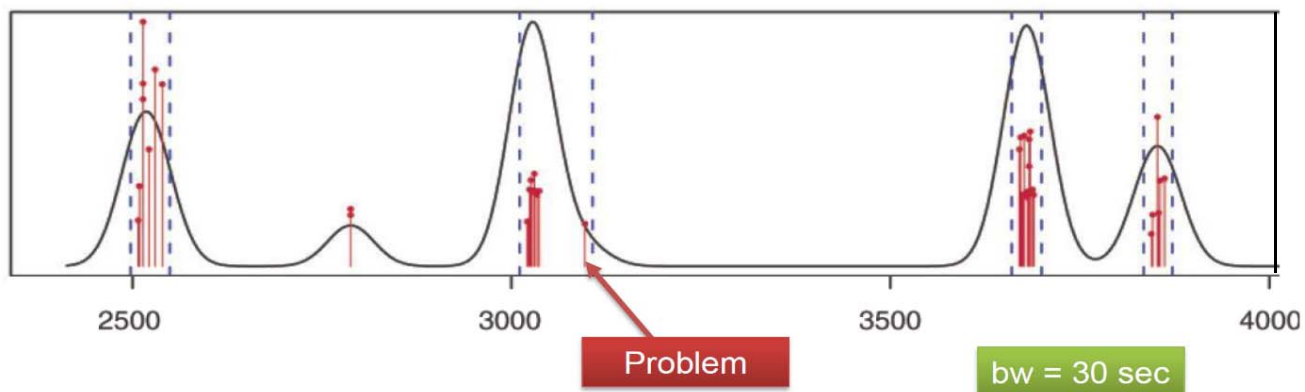
**Get a Peak List**

# SESIÓN PRÁCTICA. VISUALS\_12

Agrupación de picos en *bin* de masa: 337.975 – 338.225 m/z

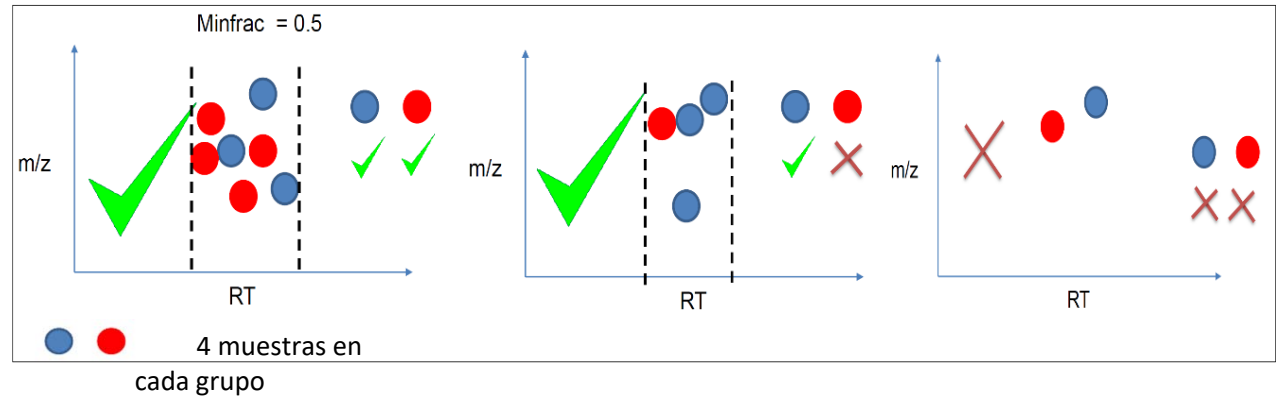
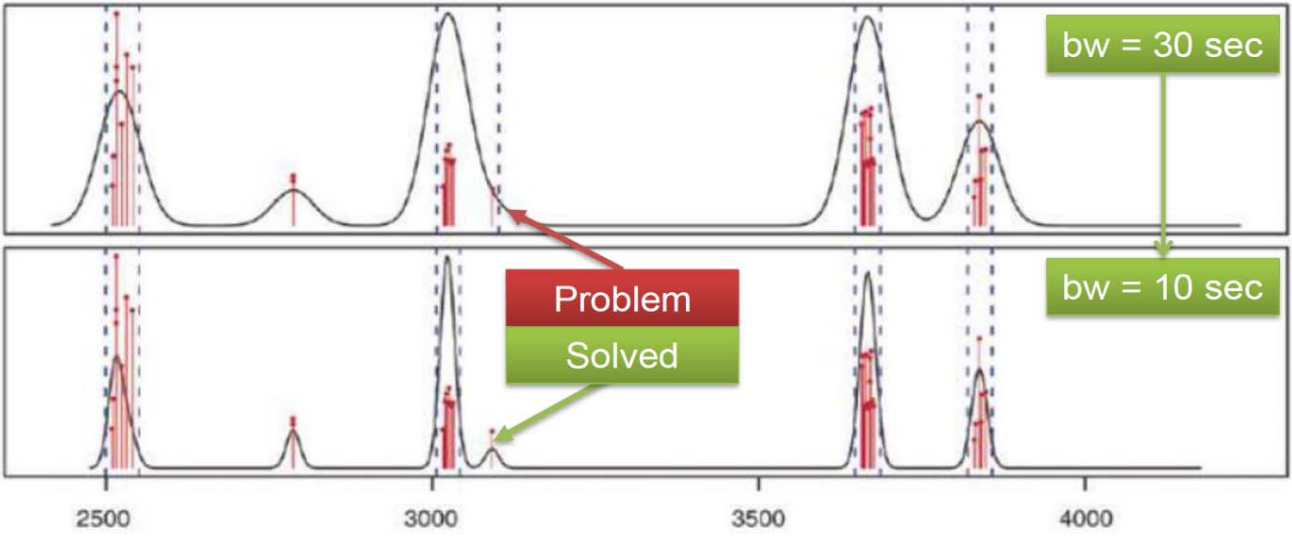


Agrupación de picos en *bin* de masa: 337.975 – 338.225 m/z (mzwid)



# SESIÓN PRÁCTICA. VISUALS\_13

Agrupación de picos en bin de masa: 337.975 – 338.225 m/z (mzwid)



# SESIÓN PRÁCTICA. VISUALS\_14

**Galaxy / 4 / Metabolomics** Analyze Data Workflow Shared Data Visualization Help User Using 15.7 GB

**Tools**

search tools

**Upload File from your computer**

LC-MS

**Preprocessing**

- xcms.xcmsSet** Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis
- xcms.xcmsMerger** Merge xcms.xcmsSet xset in one to be used by group
- xcms.group** Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.
- xcms.retcor** Retention Time Correction using retcor function from xcms R package
- xcms.fillPeaks** Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas
- xcms.summary** Create a summary of XCMS analysis

**History**

search datasets

Unnamed history  
10 shown  
3.62 GB

1 job has been successfully added to the queue - resulting in the following datasets:

- 7: LC\_Positive\_mzXML.xset.group.RData
- 8: LC\_Positive\_mzXML.xset.group.Rplots.pdf
- 9: LC\_Positive\_mzXML.xset.group.variableMetadata.tsv
- 10: LC\_Positive\_mzXML.xset.group.dataMatrix.tsv
- 11: xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from "running" to "finished" if completed successfully or "error" if problems were encountered.

**Group.variableMetadata.**

Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
MS100001	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000	1000000000

**Group.dataMatrix.ts**

Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
MS100001	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

**Group.Rpl**

ots.pdf

Shows the different samples (dot) grouped together in a peak density chromatogram defined by bw and mzwid value

mzwid define the intervals of m/z

204 - 204.01

bw define the width of the gaussian curve

Two distinct m/z merge as one group. mzwid and bw too large

50.44 - 50.66

Density

429.37 - 429.38

Density

Density

N = 30 Bandwidth = 15

N = 77 Bandwidth = 15



N = 437 Handwidth = 10

# SESIÓN PRÁCTICA. VISUALS\_15




[xcms.retcor](#) Retention Time Correction using retcor function from xcms R package

Parameter : num + label	Format
1 : RData file	rdata.xcms.group

**xcms.retcor Retention Time Correction using retcor function from xcms R package (Galaxy Version 2.1.0)**

 Versions
 
 Options

**xset RData file**

▼

output file from another function xcms (xcmsSet, retcor etc.)

**Method to use for retention time correction**

▼

[method] See the help section below

**Smooth method**

▼

[smooth] either 'loess' for non-linear alignment or 'linear' for linear alignment

**Number of extra peaks to allow in retention time correction correction groups**

[extra]


**Number of missing samples to allow in retention time correction groups**


Number of admitted missing well behaved peak in a group.

[missing]

**Advanced options**

▼

**Resubmit your raw dataset or your zip file** 

 Execute

# SESIÓN PRÁCTICA. VISUALS\_16

## Advanced options

show ▼

### Degree of smoothing for local polynomial regression fitting

0.2

[span]

### Family

gaussian ▼

[family] if gaussian fitting is by least-squares with no outlier removal, and if symmetric a re descending M estimator is used with Tukey's biweight function, allowing outlier removal

### plottype

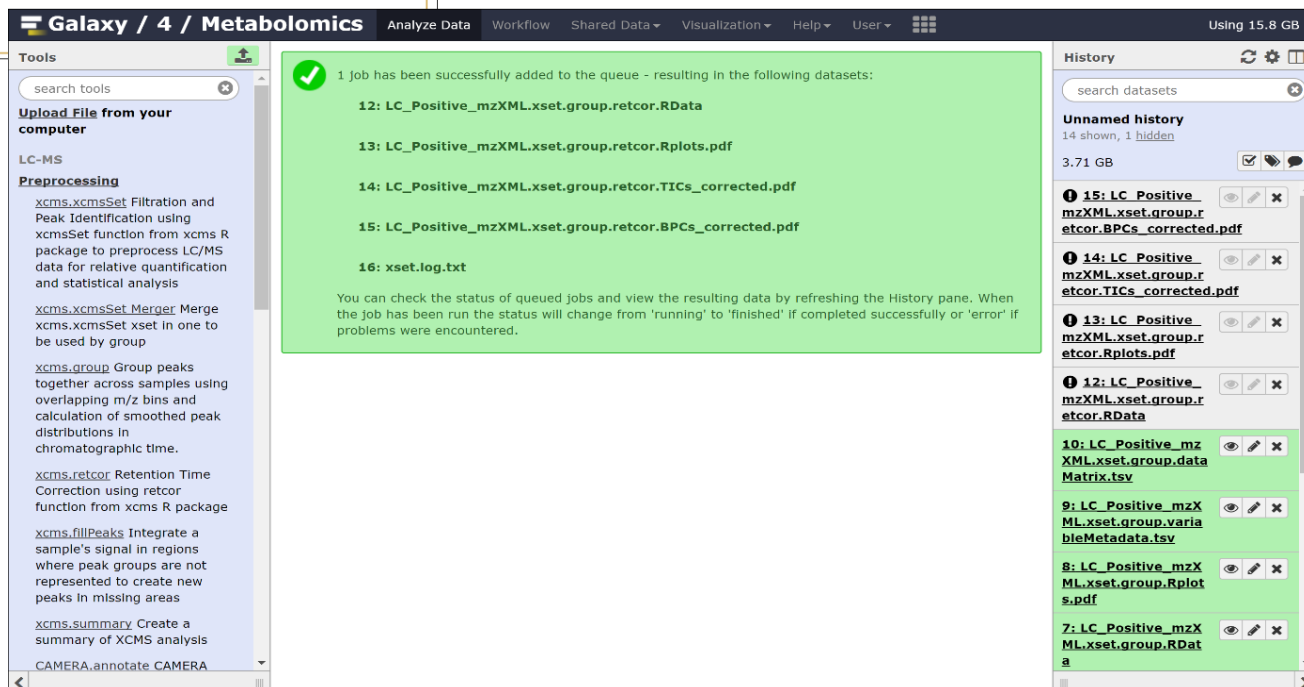
deviation ▼

**Plot to visualize the result of the retention time correction.**

[plottype] if deviation plot retention time deviation points and regression fit, and if mdevden also plot peak overall peak density and retention time correction peak density

**Resubmit your raw dataset or your zip file** 

 **Execute**



**Galaxy / 4 / Metabolomics** Analyze Data Workflow Shared Data Visualization Help User Using 15.8 GB

**Tools**

search tools

**Upload File from your computer**

LC-MS

**Preprocessing**

- [xcms.xcmsSet](#) Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis
- [xcms.xcmsSet Merger](#) Merge xcms.xcmsSet xset in one to be used by group
- [xcms.group](#) Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.
- [xcms.retcor](#) Retention Time Correction using retcor function from xcms R package
- [xcms.fillPeaks](#) Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas
- [xcms.summary](#) Create a summary of XCMS analysis

CAMERA.annotate CAMERA

**1 job has been successfully added to the queue - resulting in the following datasets:**

- 12: LC\_Positive\_mzXML.xset.group.retcor.RData
- 13: LC\_Positive\_mzXML.xset.group.retcor.Rplots.pdf
- 14: LC\_Positive\_mzXML.xset.group.retcor.TICs\_corrected.pdf
- 15: LC\_Positive\_mzXML.xset.group.retcor.BPCs\_corrected.pdf
- 16: xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

**History**

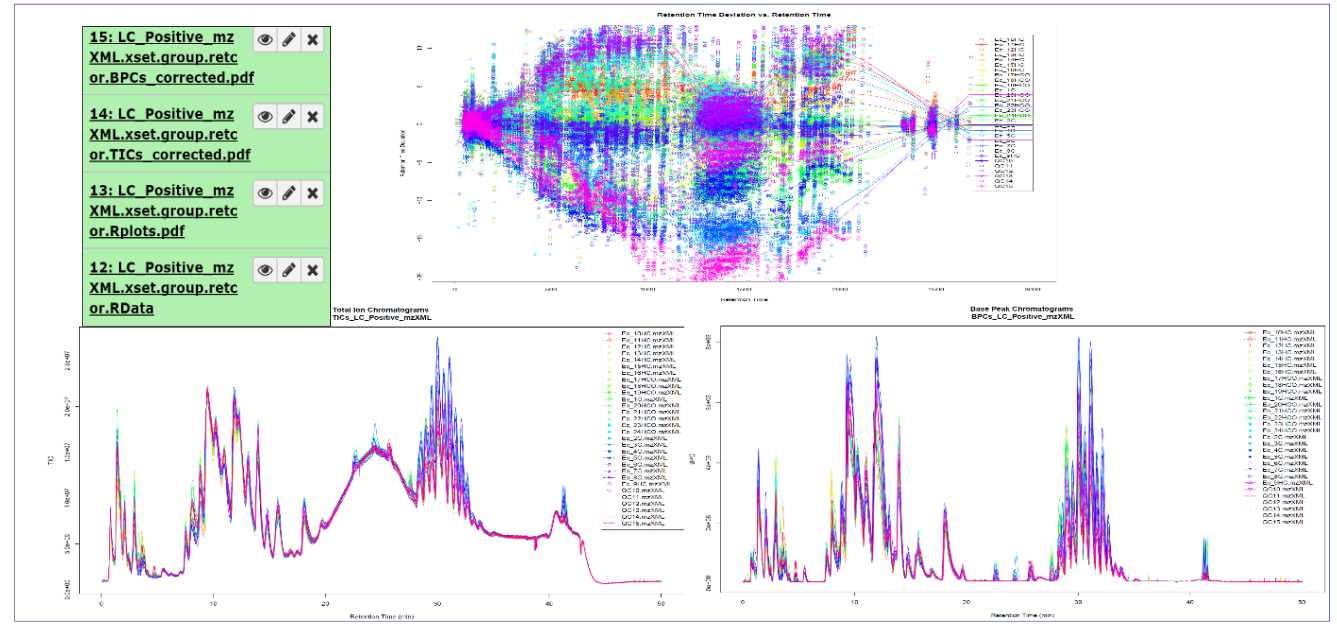
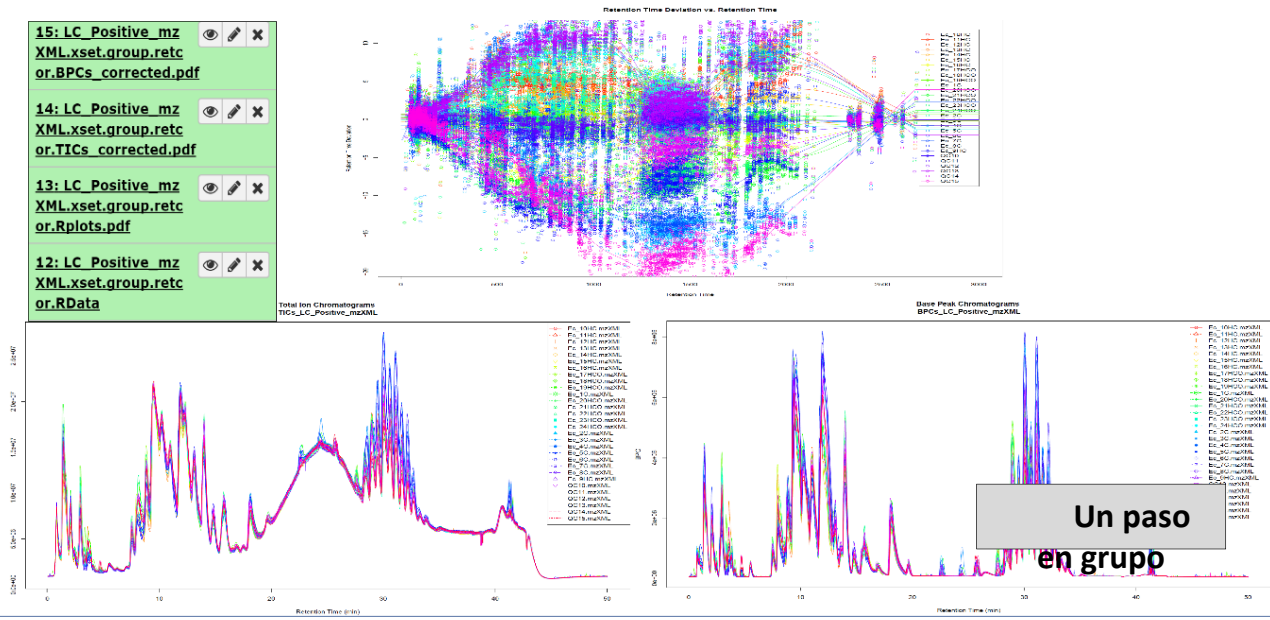
search datasets

Unnamed history  
14 shown, 1 hidden  
3.71 GB

- 15: LC\_Positive\_mzXML.xset.group.retcor.BPCs\_corrected.pdf
- 14: LC\_Positive\_mzXML.xset.group.retcor.TICs\_corrected.pdf
- 13: LC\_Positive\_mzXML.xset.group.retcor.Rplots.pdf
- 12: LC\_Positive\_mzXML.xset.group.retcor.RData
- 10: LC\_Positive\_mzXML.xset.group.data.Matrix.tsv
- 9: LC\_Positive\_mzXML.xset.group.variateMetadata.tsv
- 8: LC\_Positive\_mzXML.xset.group.Rplots.pdf
- 7: LC\_Positive\_mzXML.xset.group.RData



# SESIÓN PRÁCTICA. VISUALS\_17





# SESIÓN PRÁCTICA. VISUALS\_18



**Galaxy / 4 / Metabolomics** Analyze Data Workflow Shared Data Visualization Help User Using 15.8 GB

Tools

1 job has been successfully added to the queue - resulting in the following datasets:

- 17: LC\_Positive\_mzXML.xset.group.retcor.group.RData
- 18: LC\_Positive\_mzXML.xset.group.retcor.group.Rplots.pdf
- 19: LC\_Positive\_mzXML.xset.group.retcor.group.variableMetadata.tsv
- 20: LC\_Positive\_mzXML.xset.group.retcor.group.dataMatrix.tsv
- 21: xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

Unnamed history  
18 shown, 2 hidden  
3.73 GB

- 20: LC\_Positive\_mzXML.xset.group.retcor.group.dataMatrix.tsv
- 19: LC\_Positive\_mzXML.xset.group.retcor.group.variableMetadata.tsv
- 18: LC\_Positive\_mzXML.xset.group.retcor.group.Rplots.pdf
- 17: LC\_Positive\_mzXML.xset.group.retcor.group.RData
- 15: LC\_Positive\_mzXML.xset.group.retcor.BPCs\_corrected.pdf
- 14: LC\_Positive\_mzXML.xset.group.retcor.TICs\_corrected.pdf
- 13: LC\_Positive\_mzXML.xset.group.retcor.Rplots.pdf
- 12: LC\_Positive\_mzXML.xset.group.retcor.RData

Un paso en grupo

**xcms.fillPeaks** Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas

Parameter : num + label	Format
1 : RData file	rdata.xcms.group

**xcms.fillPeaks** Integrate a sample's signal in regions where peak groups are not represented to create new peaks in missing areas (Galaxy Version 2.1.0) Versions Options

**xset RData file**

output file from another xcms function (group)

**Filling method**

[method] See the help section below

**Get a Peak List**

**Resubmit your raw dataset or your zip file**

# SESIÓN PRÁCTICA. VISUALS\_19

Galaxy / 4 / Metabolomics

Tools

- search tools
- xcms.xcmsSet
- xcms.xcmsSet Merger
- xcms.group
- xcms.retcor
- xcms.fillPeaks
- xcms.summary
- CAMERA.annotate
- CAMERA.combineXAnno

History

24: LC\_Positive\_mzXML.xset.group.retcor.group.fillPeaks.dataMatrix.tsv

23: LC\_Positive\_mzXML.xset.group.retcor.group.fillpeaks.variableMetadata.tsv

22: LC\_Positive\_mzXML.xset.group.retcor.group.fillPeaks.RData

25: xset.log.txt

1 job has been successfully added to the queue - resulting in the following datasets:

22: LC\_Positive\_mzXML.xset.group.retcor.group.fillPeaks.RData

23: LC\_Positive\_mzXML.xset.group.retcor.group.fillpeaks.variableMetadata.tsv

24: LC\_Positive\_mzXML.xset.group.retcor.group.fillpeaks.dataMatrix.tsv

25: xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from "running" to "finished" if completed successfully or "error" if problems were encountered.

variableMeta					dataM				
data.tsv					atrix.tsv				
1	2	3	4	5	1	2	3	4	5
name	namecustom	mz	mzmin	mzmax	name	Ec_10HC	Ec_11HC	Ec_12HC	Ec_13HC
M103T2673	M102.9703T44.55	102.970311351949	102.970195579109	102.970399312385	M103T2673	1061832.30493052	1144403.44135983	1020539.16801412	1121248.09306137
M104T52	M104.10770.86	104.106964995157	104.105982748003	104.107508430139	M104T52	116131.202463165	113145.345969103	221247.835289123	85812.9626298084
M107T56	M107.0706T0.94	107.070606586921	107.070137215825	107.071251918772	M107T56	11028.4704300842	24224.0223756465	40340.7417124155	20019.0940904541
M109T176	M109.10172.93	109.100995864082	109.100845900078	109.101142306016	M109T176	201788.89433418	208794.490735228	95350.587358747	100622.014506292
M10T66_1	M110.0712T1_1_1	110.071232504779	110.071135326801	110.071413890512	M10T66_1	447052.480245756	509069.134419706	529036.944020849	419073.389406295
M10T66_2	M110.145T1.11_2	110.145035383832	110.144834345671	110.145230783207	M10T66_2	29398.852561569	33127.152696311	36290.5541728252	27314.146364311
M11T179	M111.116T1.31	111.116790950832	111.116666786018	111.116966139657	M11T179	154633.650364231	107591.113126188	162187.205989563	153506.517209983
M11T180	M111.1165T3	111.116486144107	111.116213639299	111.11662410968	M11T180	54067.0514414524	65254.58342024	49274.2740727852	60560.6364823781
M112T51	M112.0521T0.86	112.052056215129	112.051444952237	112.052702344466	M112T51	82566.9537246094	100534.324090609	98626.7065472707	98909.7776453247
M112T79	M112.1201T1.31	112.120663107016	112.119644951814	112.120279047431	M112T79	12748.4660072896	20138.4564137684	13863.0038125153	12884.1150847926
M113T48	M112.8959T0.81	112.895923556653	112.895275075808	112.896851780014	M113T48	66262.9480098419	62663.7677114182	88327.424822492	70933.4877627869
M113T59	M112.9995T0.98	112.999451218223	112.999065891781	112.999691444871	M113T59	96435.8076669617	85020.2573272705	84614.7288627994	90950.2532185818
M115T48	M114.8933T0.81	114.893323974783	114.892556681008	114.894135034299	M115T48	32698.9970340881	32496.02112052232	41719.2826053286	31671.5001469726
M116T54	M116.0706T0.9	116.070621610785	116.07008824786	116.07086028302	M116T54	172069.77323927	199833.290424495	259939.762536967	189934.579256049
M117T171	M117.0697T2.85	117.069716380854	117.069232623419	117.070682716424	M117T171	25315.1286311528	28061.6590161876	17398.6140722212	19771.6361491645
M118T2331_1	M118.0862T38.85_1	118.086184577653	118.086051978809	118.08632326958	M118T2331_1	425563.804303394	338087.459169224	394919.65065011	266230.851963383
M118T53_1	M118.0860T70.88_1	118.086740230705	118.086065712653	118.08718457984	M118T53_1	14370174.3398173	11354958.4483364	13709306.7272189	8072194.34844452
M118T112	M118.0862T1.87	118.086184011752	118.086107212736	118.086292372487	M118T112	2716813.17062041	2023073.28526089	2918253.74628112	1559191.99272233
M118T2331_2	M118.1624T38.85_2	118.162424196361	118.162275890008	118.162654424063	M118T2331_2	29155.2018136445	20060.6112223546	25748.7617057155	14566.8909274575
M118T53_2	M118.1625T0.88_2	118.162541997427	118.161930061116	118.16295469244	M118T53_2	1645565.67086	1214220.15293176	1501034.72970629	818662.841329422
M118T52_1	M118.188T0.87_1	118.187994739206	118.187269133652	118.188054468827	M118T52_1	146040.357902741	115410.21893396	147278.242704987	73784.7286913757
M118T52_2	M118.2124T0.87_2	118.212381136854	118.211443391281	118.212900282301	M118T52_2	143442.314667542	111096.156310303	143350.57062082	69977.2087654114
M118T52_3	M118.236T0.87_3	118.236519662066	118.2358711863	118.237417400209	M118T52_3	1627977.36458696	78505.770380249	103179.631372188	48305.8035364075
M119T53_1	M119.0895T0.88_1	119.089482441589	119.08882743951	119.08995212236	M119T53_1	798878.512076424	636208.391813786	754206.198042717	451126.926595215
M119T2330	M119.0885T38.84	119.088498058298	119.088215220181	119.08889805151	M119T2330	24954.8740942383	23399.8112346703	23575.9488785678	13338.1884562498
M119T53_2	M119.1656T0.88_2	119.165632425698	119.164957479794	119.166194657233	M119T53_2	54103.2992341919	49257.9664361572	60770.7195596075	35086.4712632904
M121T173	M121.1002T2.88	121.1001599399284	121.100016017078	121.100255809592	M121T173	240164.240526719	3207661.5872076	331021.874998269	251582.171688811
M121T1739	M121.1001T28.98	121.100136272057	121.09952084354	121.100214241705	M121T1739	2425847.45517966	174789.838189487	288696.654292461	249404.370447165
M121T176	M121.1241T2.94	121.124074506687	121.123181163719	121.12682345796	M121T176	352694.554213724	130687.160363821	426307.452419805	24478.4578494873

# SESIÓN PRÁCTICA. VISUALS\_20

**24: LC\_Positive\_mz**

**XML.xset.group.ret**

**or.group.fillpeaks.dataMatrix.t**

**v**

6,929 lines

format: **tabular**, database: ?

---

**PACKAGE INFO**

parallel 3.2.5

BiocGenerics 0.16.1

Biobase 2.30.0

Rcpp 0.12.10

mzR 2.4.1

xcms 1.46.0

snow 0.4.2

batch 1.1.4

---

**ARGUMENTS INFO**

xfunction fillPeaks

Image

/work/project/w4m/galaxy4metab

dist/database/files/000/440/datas

### Matriz de datos exportada

name	Ec_10HC	Ec_11HC	Ec_12HC	Ec_13HC	Ec_14HC	Ec_15HC	Ec_16HC	Ec_17HCO	Ec_18HCO	Ec_19HCO	Ec_1C	Ec_20HCO	Ec_21HCO	Ec_22HCO	Ec_23HCO	Ec_24HCO	Ec_2C	Ec_3C	Ec_4	
1	M1037267	1061832	1144403	1202539	1212248	947508.3	1045868	978855.6	1175380	1011507	1118730	1150975	975248.5	1031258	457858.8	878824.4	1019229	1088809	1136409	10
2	M104752	116131.2	113145.3	221247.8	85812.96	146116.3	140097.1	133505.4	86630.77	95829.4	96932.66	43766.73	109815.5	124091.2	102362.2	122175.4	132149.1	47053.63	59721.62	463
3	M107756	11028.47	24224.02	40340.74	20019.09	26727.98	19011.27	35323.02	20886.77	18390.19	23702.54	23869.28	23943.1	24901.63	19243.22	16770.16	24718.96	20585.67	28613.4	201
4	M109176	201788.9	208794.5	95350.59	100622	194415.7	127442.4	113288.2	196363.7	389619.8	217344	237685.2	218852.2	272707.9	151016.4	133146.2	114873.2	148750.3	218011.8	133
5	M110766	44705.25	509699.1	529036.9	419073.4	412756.3	529367.8	391197.7	538527.8	543039.3	532286.6	672981.9	460273.1	529586.6	399481.3	534636.6	511456.8	636662.1	564149.4	562
6	M110766	29398.85	33212.15	36290.55	27214.15	27749.53	34127.79	25516.78	34531.99	31252.68	35446.52	45575.19	31008.11	37144.59	21991.53	36946.01	33620.55	41071.43	35151.01	385
7	M111179	154633.7	107591.1	162187.2	153509.5	169885.5	112960.5	157391.6	105751.8	86714.63	98921.62	115968.6	238308.7	153576.9	92006.39	126529	159932.3	111906.3	99650.9	104
8	M111180	54067.05	65254.58	49274.27	60560.64	56842.71	79019.7	65522.75	61210.14	110854.6	67253.51	64479.75	63502.71	84569.38	92845.65	83259	62958.21	69265.39	5379.49	573
9	M112151	82566.95	100534.3	98626.71	89809.78	99830.09	107412.1	93364.82	94608.51	89708.21	82677.17	77575.35	102612.2	108832.4	89941.34	93667.31	103061.6	96405.49	77078.09	911
10	M112179	12748.47	20138.46	13869	12884.12	12731.65	21747.13	13215.56	20259.91	17220.5	21101.15	18760.8	23401.64	12511.27	16114.86	10182.54	12618.25	18439.15	14202.86	195
11	M113148	66262.95	62663.77	88327.42	70933.49	68463.19	81837.4	88823.26	52497.03	67809.3	34205.02	49157.28	86037.92	62005.08	42987.11	69887.25	61477.21	38803.76	41662.85	281
12	M113159	96435.81	85020.26	84614.73	96950.25	83132.07	89945.89	87088.7	109169.2	90344.09	101318.5	152500.6	105461	92616.84	94935.86	70281.61	79582.32	152045.6	162440.9	134
13	M115148	32091	32496.02	41719.28	31671.5	32969.86	44350.62	44357.66	25932.56	15777.95	11776.8	26325.72	48005.22	33828.57	28310.83	34330.24	31391.11	18951.08	21805.35	273
14	M116754	172069.8	199833.3	259939.8	189934.6	195105.2	239906.9	205441.7	180374.6	176592.2	200264.9	169972.1	204549.7	175611.2	165365.3	202068.1	213644.5	22074.8	192031.8	166
15	M117171	25315.13	28091.66	17398.61	19771.64	14305.55	30913.3	17978.14	25773.44	31245.05	25222.35	38144.71	22604.85	30235.47	22155.06	31155.07	28308.49	33011.4	38659.66	275
16	M118233	425563.8	338087.5	394919.7	266230.9	526366.9	363375.3	248248	280576.4	207221.9	288967.7	251278.5	288688.1	372121.4	323816.9	291506.3	368688	226753.5	305659.1	291
17	M118753	143701.74	1135495.8	13709307	8072194	15334177	11814324	8434178	7379334	7208947	7454897	3643826	12433611	11406345	9192595	10015179	13592387	7736001	7611000	68
18	M1187112	2716813	2023073	2918254	1559192	2833037	2127166	1465745	1365296	1137802	1287075	713326.2	2185308	184210	162428	1540899	2958062	9288389	1066894	982
19	M1187233	23155.2	20060.61	25748.76	14566.89	34234.07	23916.89	16655	17578.42	10762.96	17576.66	15281.99	19255.18	24177.14	18135.73	17026.41	21657.19	15197.02	17589.02	165
20	M118753	1645566	1214220	1501035	818662.8	1754264	1246622	887988.6	739222	719811.1	745133.8	435892.3	1356507	1263127	1115128	932822.4	1522090	782386.9	751276.9	711
21	M118752	146404.0	115410.2	147278.2	73784.73	180394.3	117377.9	84299.53	66606.05	61898.48	73686.84	25228.96	1533264	124390.7	106251.7	89433.6	152055.6	781939.4	50443.41	665
22	M118752	143442.3	111096.2	143350.6	69977.21	177728.3	112887.7	80537	63235.3	59029.35	69600.07	22661.69	1353264	121611.4	100314.3	84359.51	150006.2	775875.7	49053.04	627
23	M118752	1627977	78505.77	103179.6	48305.8	131062.6	80797.53	54655.8	44279.75	39887.74	45365.3	23545.19	1344630	87993.92	72424.75	59866.54	108794.5	778445.6	34642.17	436
24	M119753	798878.5	636208.4	754206.2	451126.9	833842.5	660059.2	477399.7	417104.4	403071.1	418959.8	210077.3	689006	634051.1	536035.7	499624.4	738711.3	349614.6	260218.8	396
25	M1197233	24954.87	23399.81	23575.95	13338.19	33246.27	26156.69	15766.12	25487.23	19042.66	18934.83	13620.29	20611.4	23869.51	22634.32	22294.85	22462.74	12240.95	19630.98	197
26	M119753	54103	49257.97	60770.72	35086.47	66681.28	52155.84	37296.66	35473.82	31866.72	32966.6	13828.12	58007.47	49774.35	45230.78	39611.76	59209.22	32914.16	21121.31	314
27	M121173	240164.2	3207662	331021.9	251582.1	277004.1	193025.7	438040.5	150957.3	230271.7	275420.5	293919	293830.4	231079.1	308697.2	220404.6	290114.2	414509.1	332537.5	2
28	M121173	2425847	174789.8	288869.6	240404.4	3378697	166736.2	3713308	2691870	165942.6	2600780	83681.79	282210.2	1992916	288429.7	322657.4	323585.8	2463562	2777134	30
29	M121176	352694.6	130687.2	426307.5	24478.46	418560.6	293284.8	413021.9	39924.64	42985.45	43437.39	56638.84	33514.14	57733.0	42693.03	70068.71	412904.2	53443.06	49367.68	706
30	M122768	140002.4	116594.1	158186.9	155712.5	94234.01	113308.7	126416.3	198397.4	115663.7	153246.4	129375.5	199662.6	117730.1	148161.3	116020.3	122180.6	113448.4	134329.7	883
31	M123746	144741.3	130602.7	171793	137241.2	165415.4	166985.3	177487.4	133245.3	128352.5	94373.31	149283.6	146908	158039.3	128054.9	142776.9	151719.7	126448	130698.8	131
32	M123732	88356.7	81693.14	76197.37	75632.53	86450.62	99250.24	65960.49	162082.1	76124.61	81249.24	106183.5	147584.7	82086.18	78140.29	71682.06	88380.97	144482.6	131978.6	107
33	M123753	1135725	1283104	867423.9	921317.9	1410529	1177023	900965.6	1433746	1260881	1421248	1462581	1320889	1370182	1275632	1148370	1337942	1696713	1797970	15
34	M123752	117756	66355.3	56642.84	998293.2	75267.71	79438.89	60165.52	1323879	78001.79	83213.13	1705030	1184399	94035.13	78453.68	72901.28	86045.16	1623042	62782.46	105
35	M125767	179644.3	111714.3	51851.2	96462.24	160915.7	110145	109939.3	187179.6	47349.44	133821.7	189536.4	127374.3	118933.8	205647.3	130404.3	92409.4	179574.8	207092.8	250

[xcms.summary](#) Create a summary of XCMS analysis

**xcms.summary** Create a summary of XCMS analysis (Galaxy Version 1.0.3) Versions Options

**xset RData file**

output file from another function xcms (xcmsSet, group, retcor, fillpeaks etc.)

**Execute**

---

Este proyecto ha sido financiado con ayuda de la Comisión Europea.

La presente publicación recoge únicamente las opiniones de los autores, por lo que la Comisión no se hace responsable de cualquier uso que se haga de la información contenida en ella.



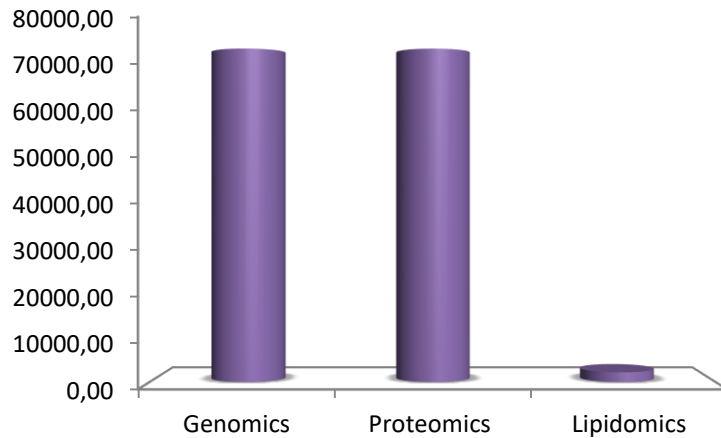
# Lipidómica basada en Espectrometría de Masas

Elisabete Maciel,  
Eliana Alves,  
Pedro Domingues,  
Rosário Domingues

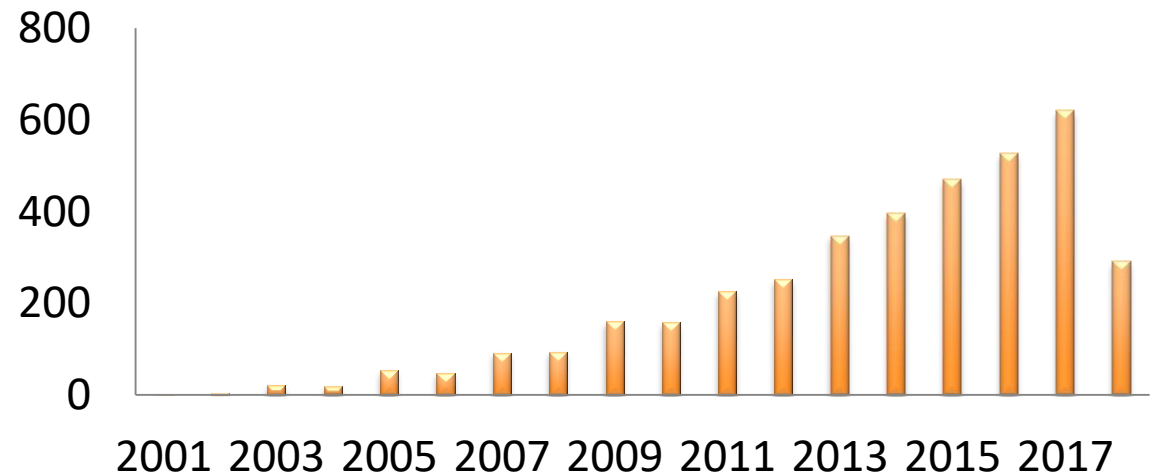


# Lipidómica

Nº de artículos publicados por área “ómica”  
(Scopus)

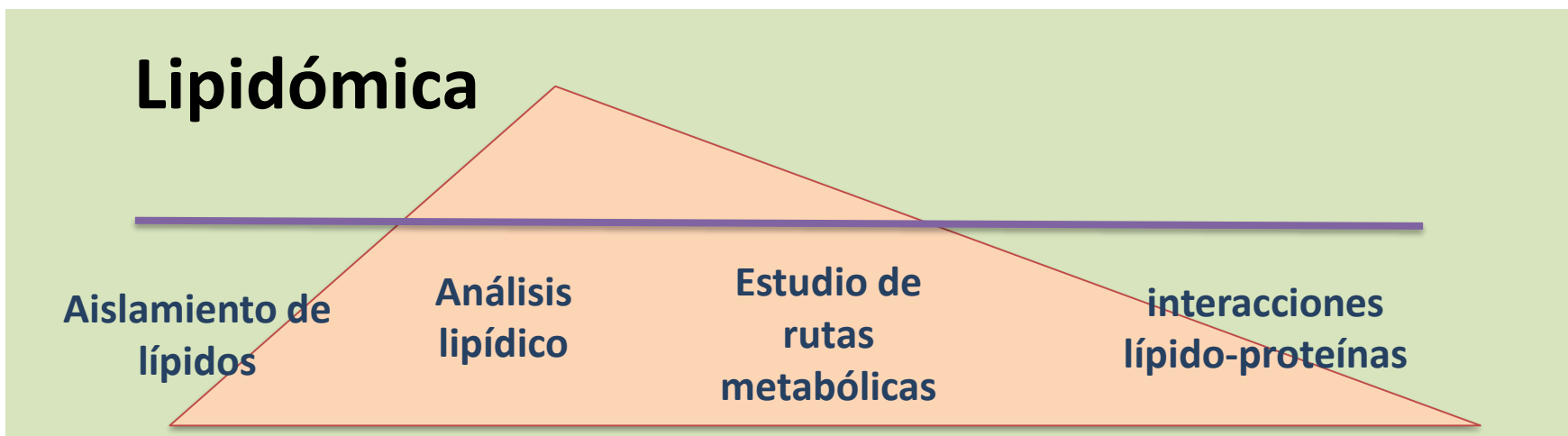


nº de publicaciones



# Lipidómica

- La **caracterización completa** de las especies moleculares lipídicas y de sus **funciones biológicas** con respecto a la expresión y función de las proteínas que intervienen en el metabolismo lipídico, en particular la regulación génica (AOCS Lipids Library)
- Análisis de perfil lipídico y su relación con la fisiología y la patofisiología celular



# Lipidómica

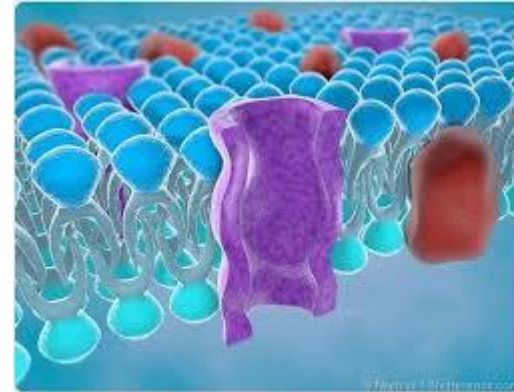
---

- Determinación del perfil del lipidoma celular
- Dominios y dinámicas de los lípidos de membranas
- Funciones regulatorias (ej. de señalización) de los lípidos
- Integración de las ómicas + interacción del complemento celular + mecanismos para la formación de células/organismos



# ¿Por qué son tan importantes los lípidos?

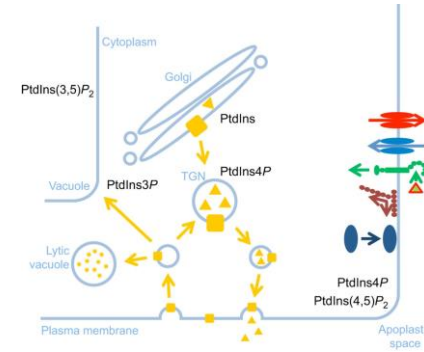
## Membranas



## Regulación Celular

mensajeros de  
señalización, hormonas, ...

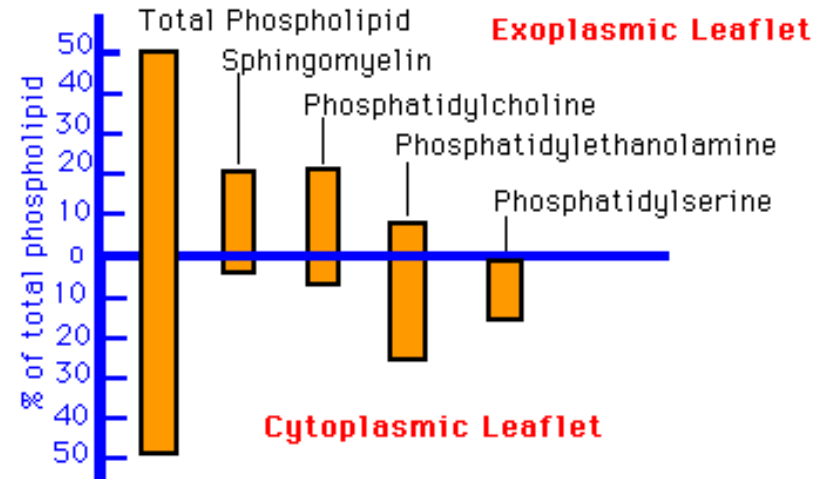
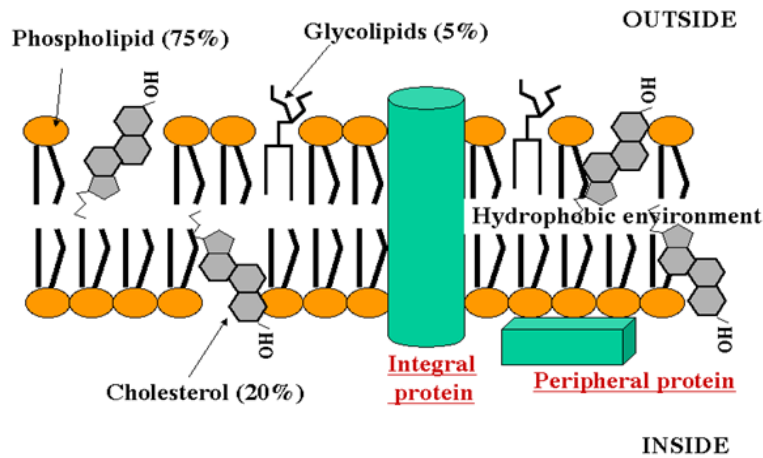
## Metabolismo Energético/Reservas



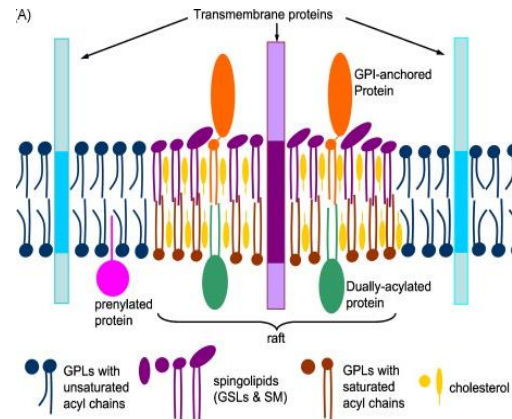
La disfunción de la señalización y el metabolismo lipídicos desempeña un papel fundamental en la **salud y las enfermedades**

# Lípidos - Membrana Celular

## Asimetría de la membrana



## Dominios de la membrana Balsas lipídicas



# Elaboración de perfiles lipídicos en células, tejidos y biofluidos

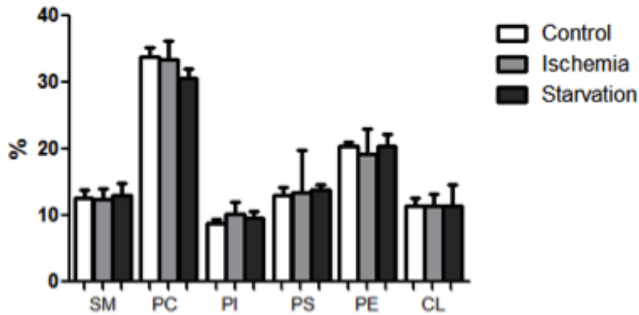
Cada tipo de célula, tejido y fluido corporal tiene un perfil lipídico característico con una composición lipídica definida

Identificación de todos los lípidos celulares -

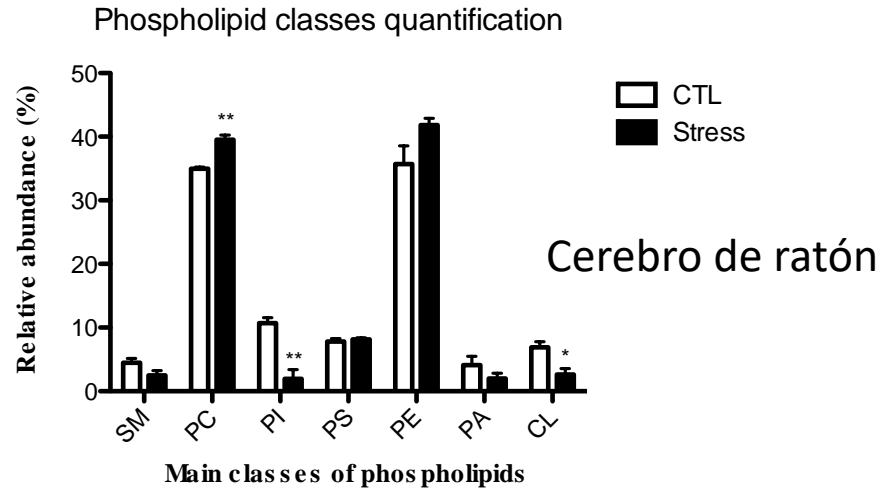
## lipidoma

Glycerolipids	% of total acyl lipid content		
	Chloroplast thylakoid	inner	
		mitochondrial membrane	plasma membrane
MGDG	51%	0	0
DGDG	26	0	0
SQDG	7	0	0
PC	3	27	32
PS	0	25	0
PG	9	0	0
PE	0	29	46
PI	1	0	19
CL	0	20	0

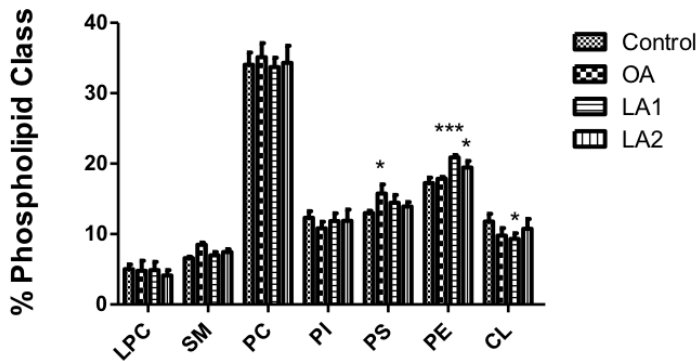
# Perfil de clases fosfolipídicas



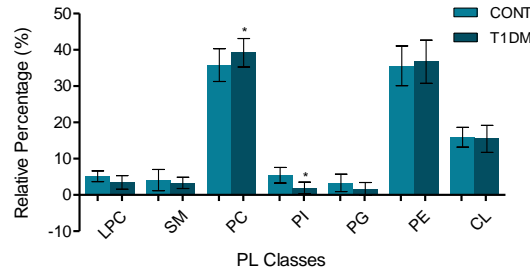
Línea celular de cardiomiocitos



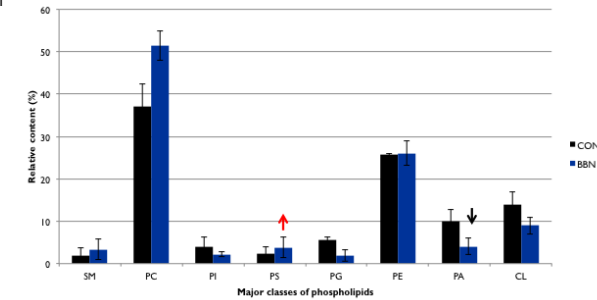
Cerebro de ratón



Líneas celulares de cáncer de mama

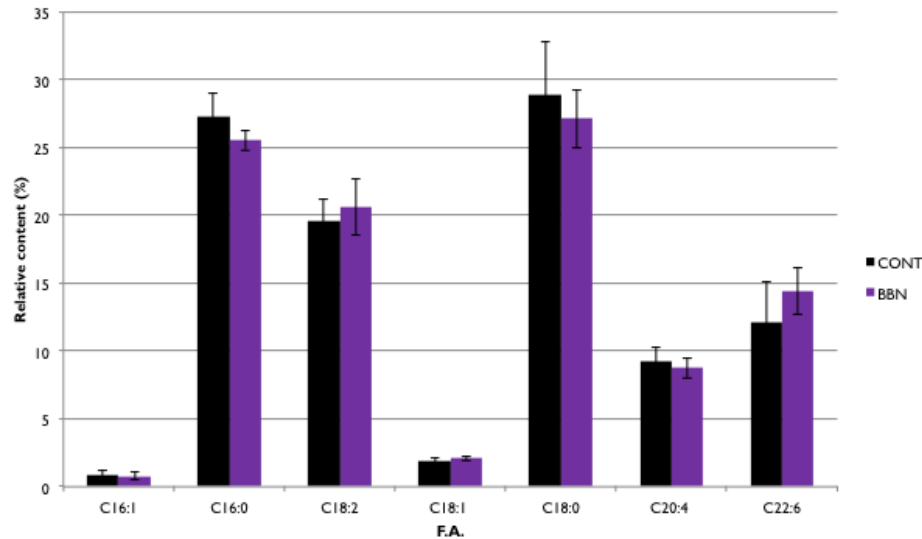


Mitocondria del corazón

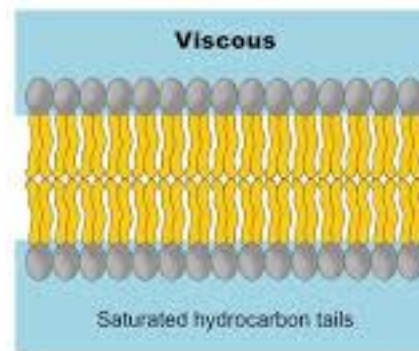
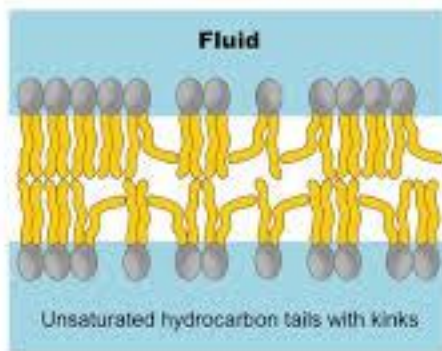


Mitocondria del músculo

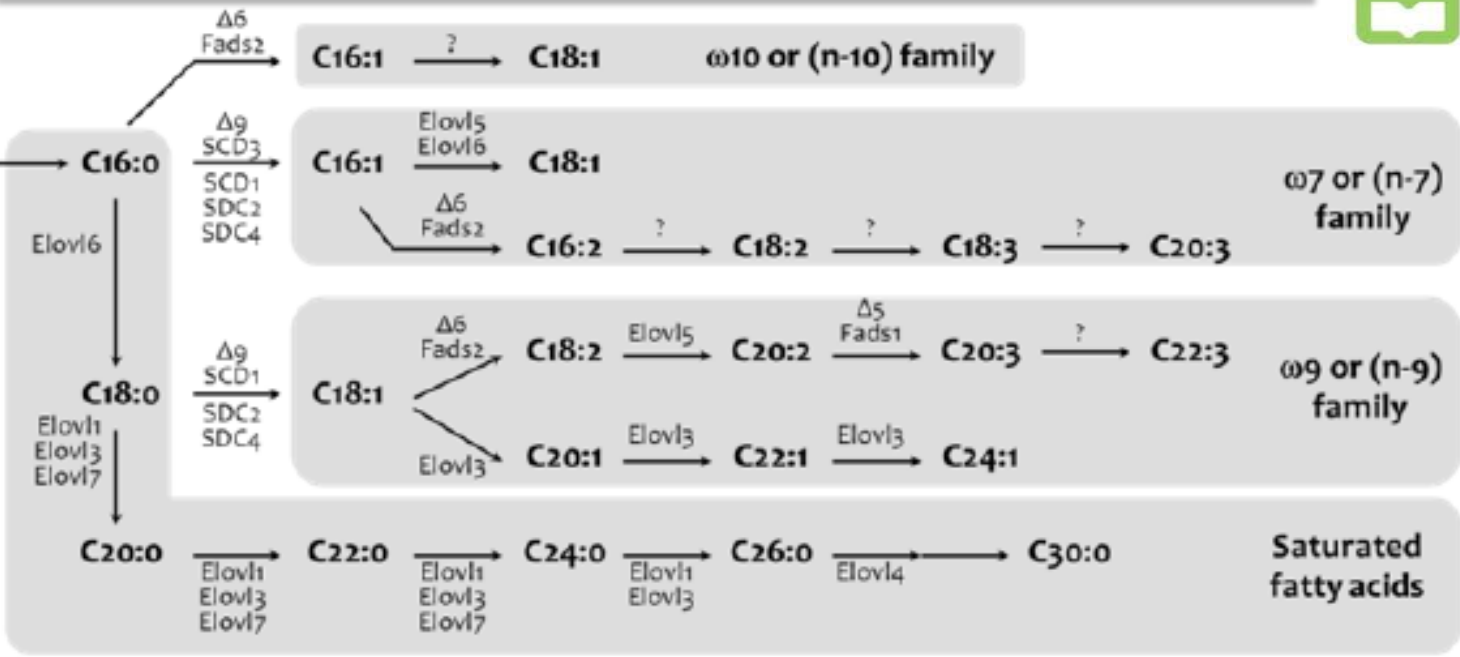
# Perfil de ácidos grasos



Efectos de la composición de ácidos grasos fosfolipídicos en propiedades de la membrana



Fatty acid synthase



Diet



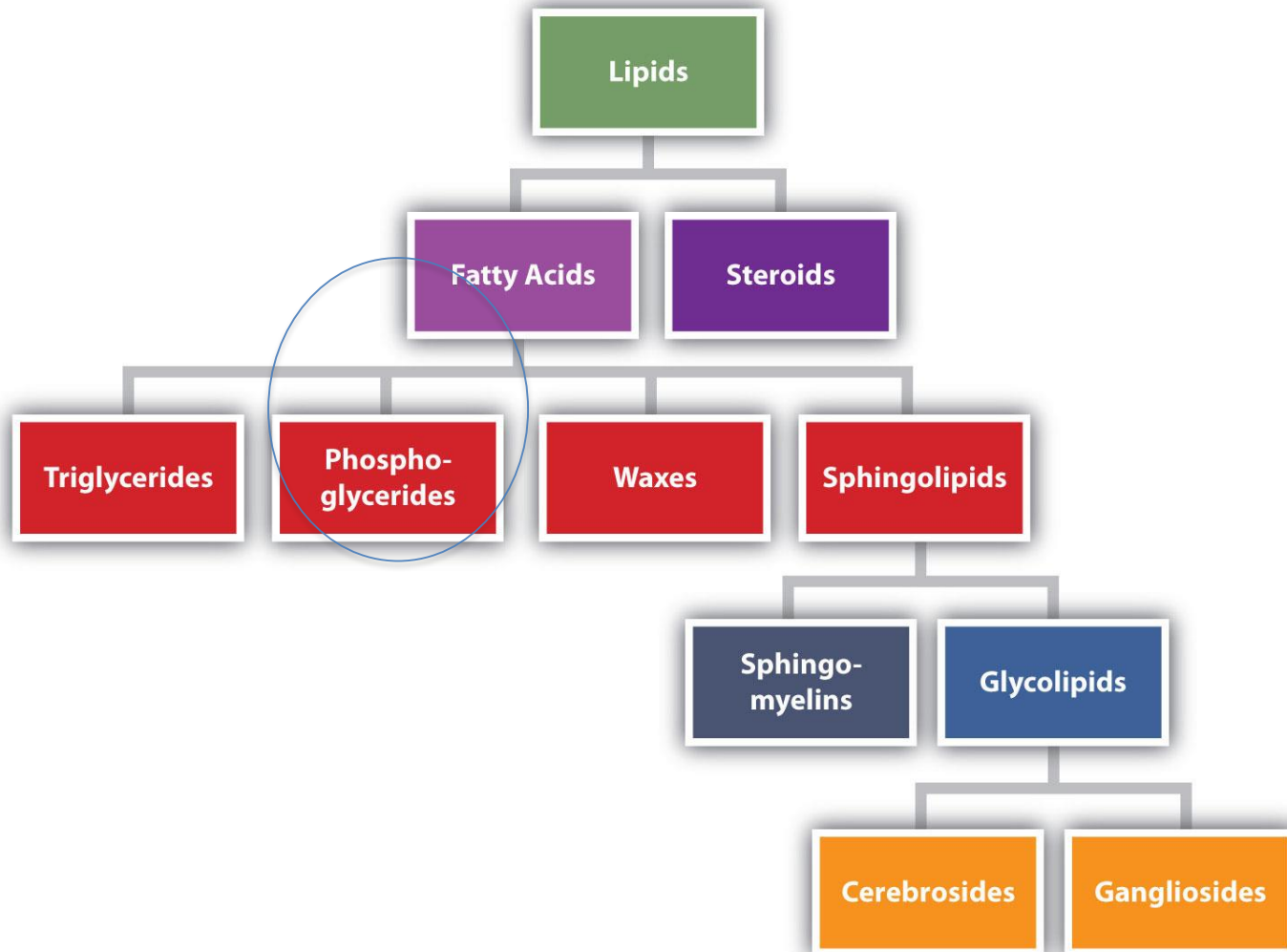
# ¿Cuáles son los grandes retos de la lipidómica?

---

AAcLifeSci 



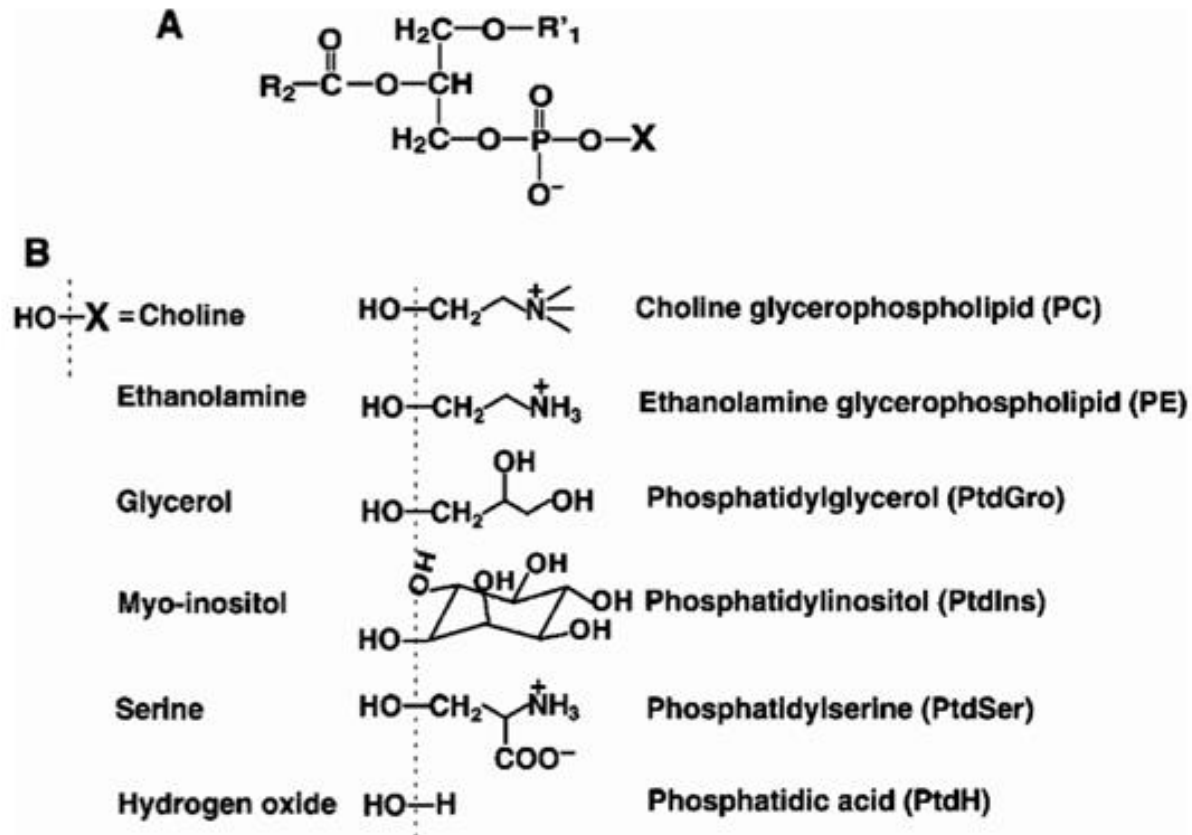
# Complejidad estructural de los lípidos



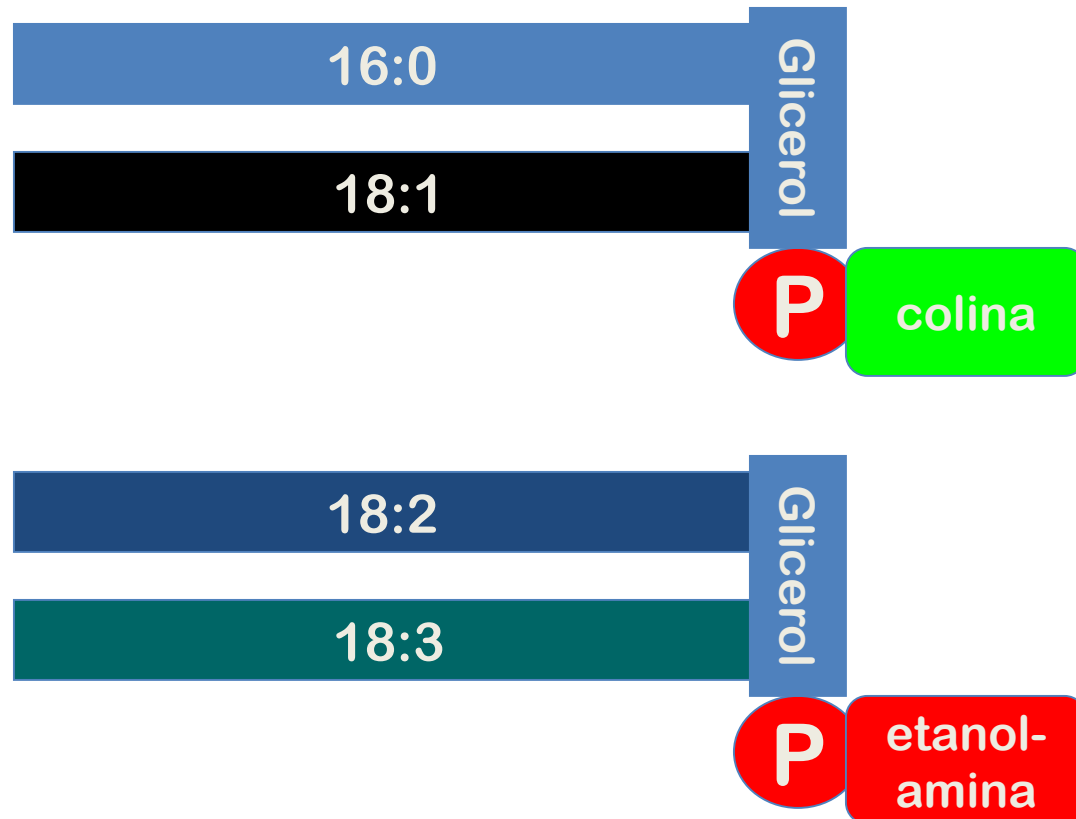


# Especies Moleculares de Fosfolípidos/Glicerolípidos

También denominados  
glicerofosfolípidos

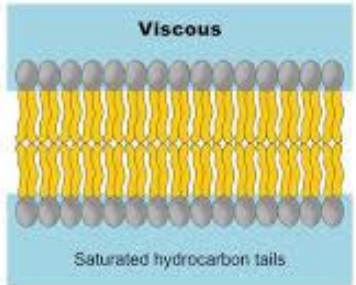
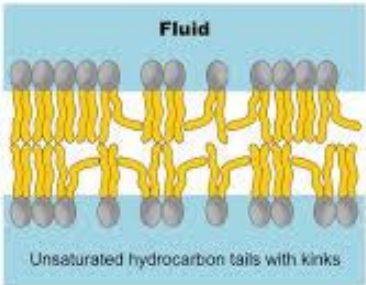
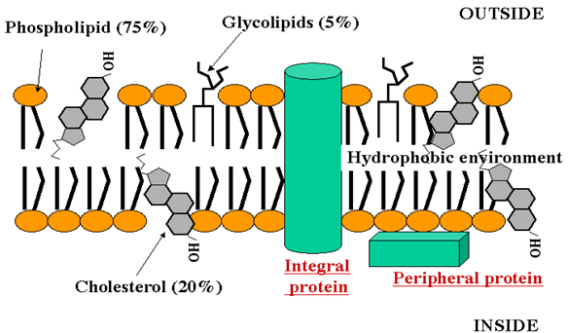
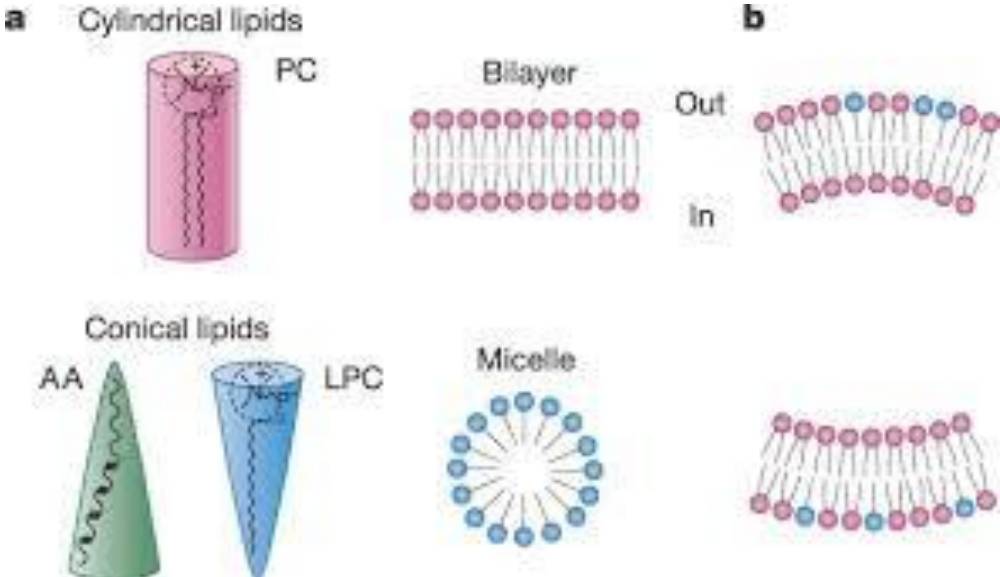


# Especies Moleculares de Fosfolípidos/Glicerolípidos

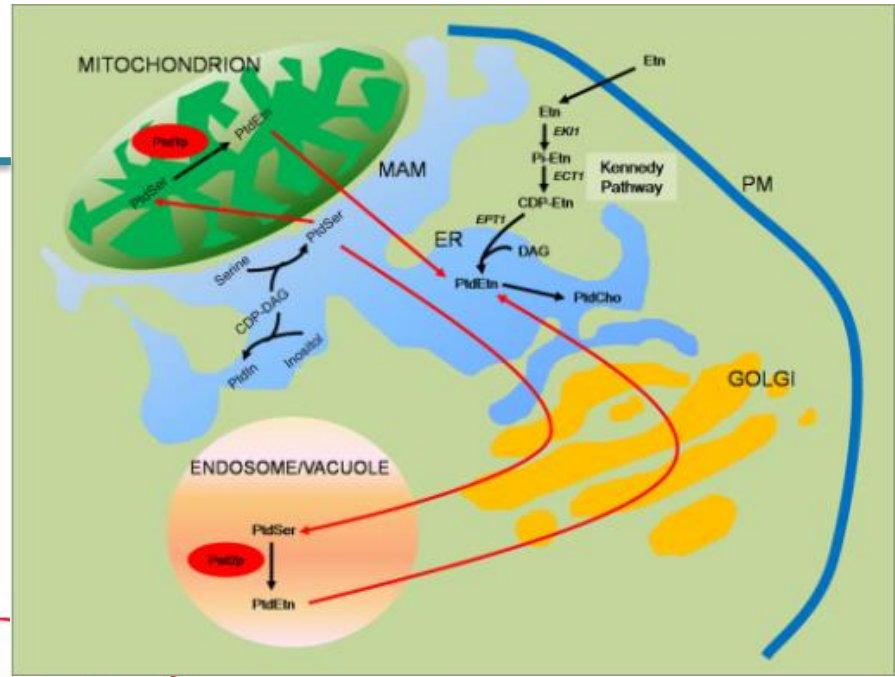
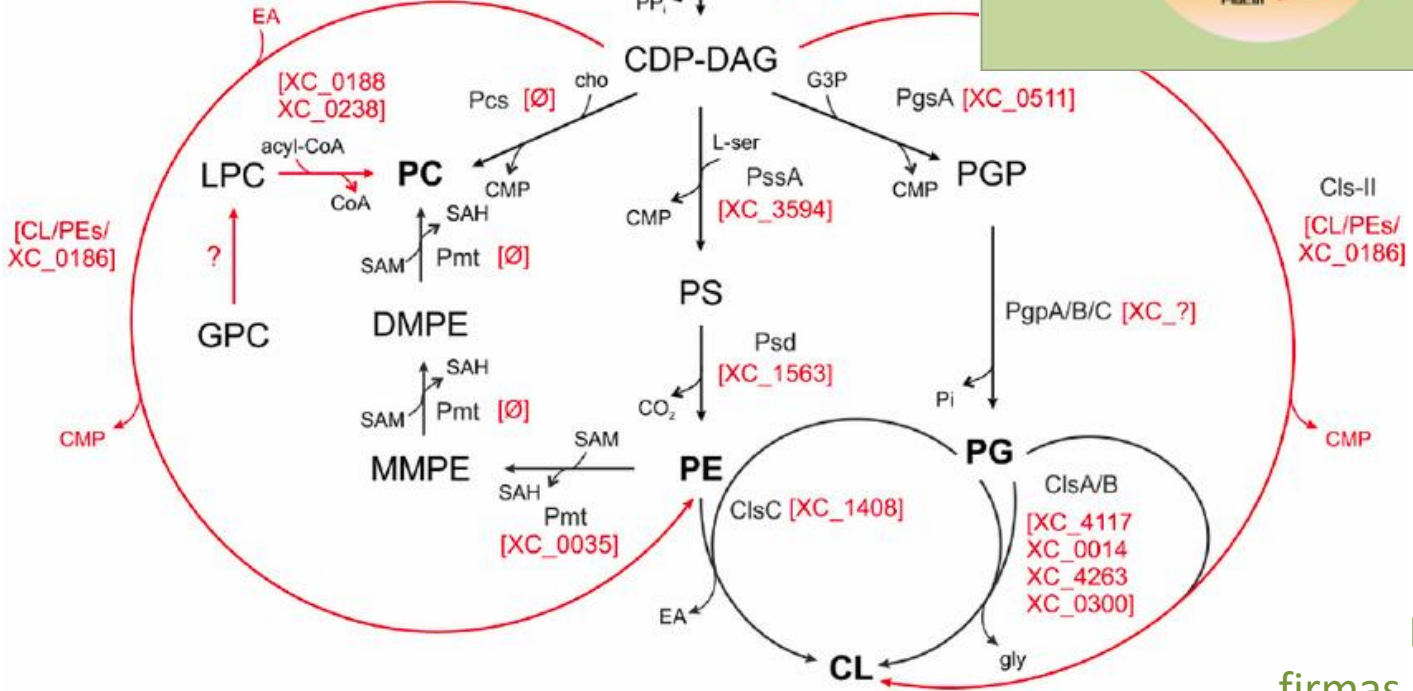
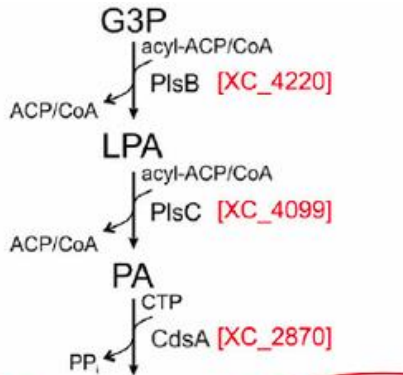


*PC, PE, PS, NPE, PG, CL, PI, PIP, PIP<sub>2</sub>, PIP<sub>3</sub>, LisoPL.....*

# Composición de los fosfolípidos y ácidos grasos y propiedades de la membrana



# Biosíntesis de los Fosfolípidos



Definir y regular firmas lipidómicas específicas

# Desviaciones en el lipidoma

---

## Patología

- Alteración en las rutas metabólicas
- Modificación oxidativa de algunos lípidos

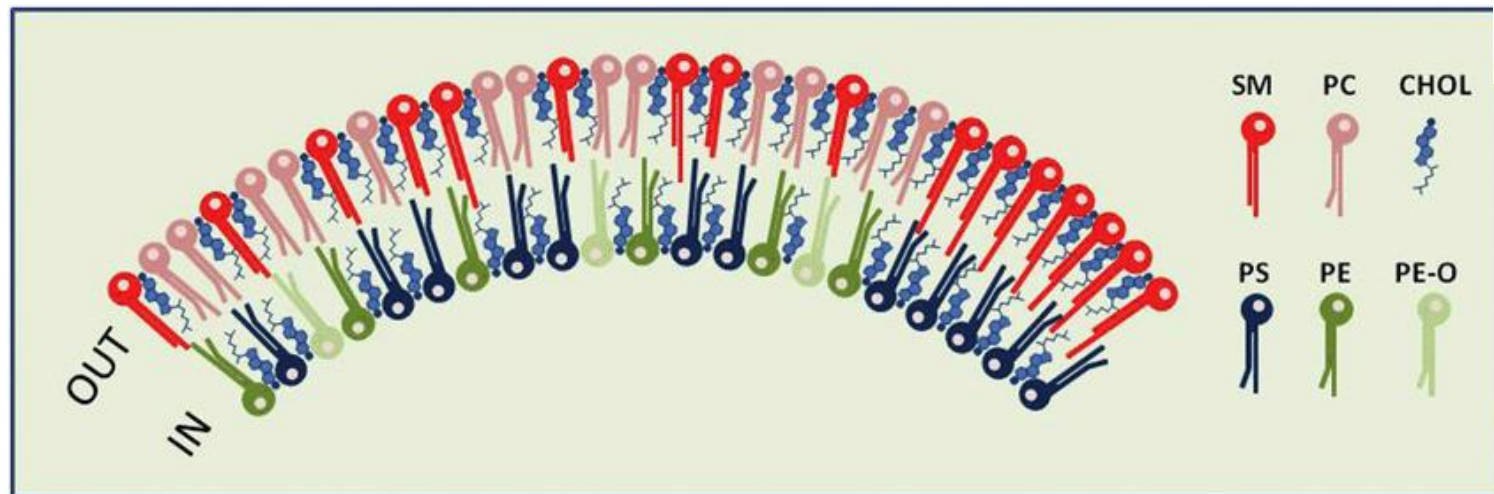
Otros:

- **Dieta** –fuente de distintos lípidos

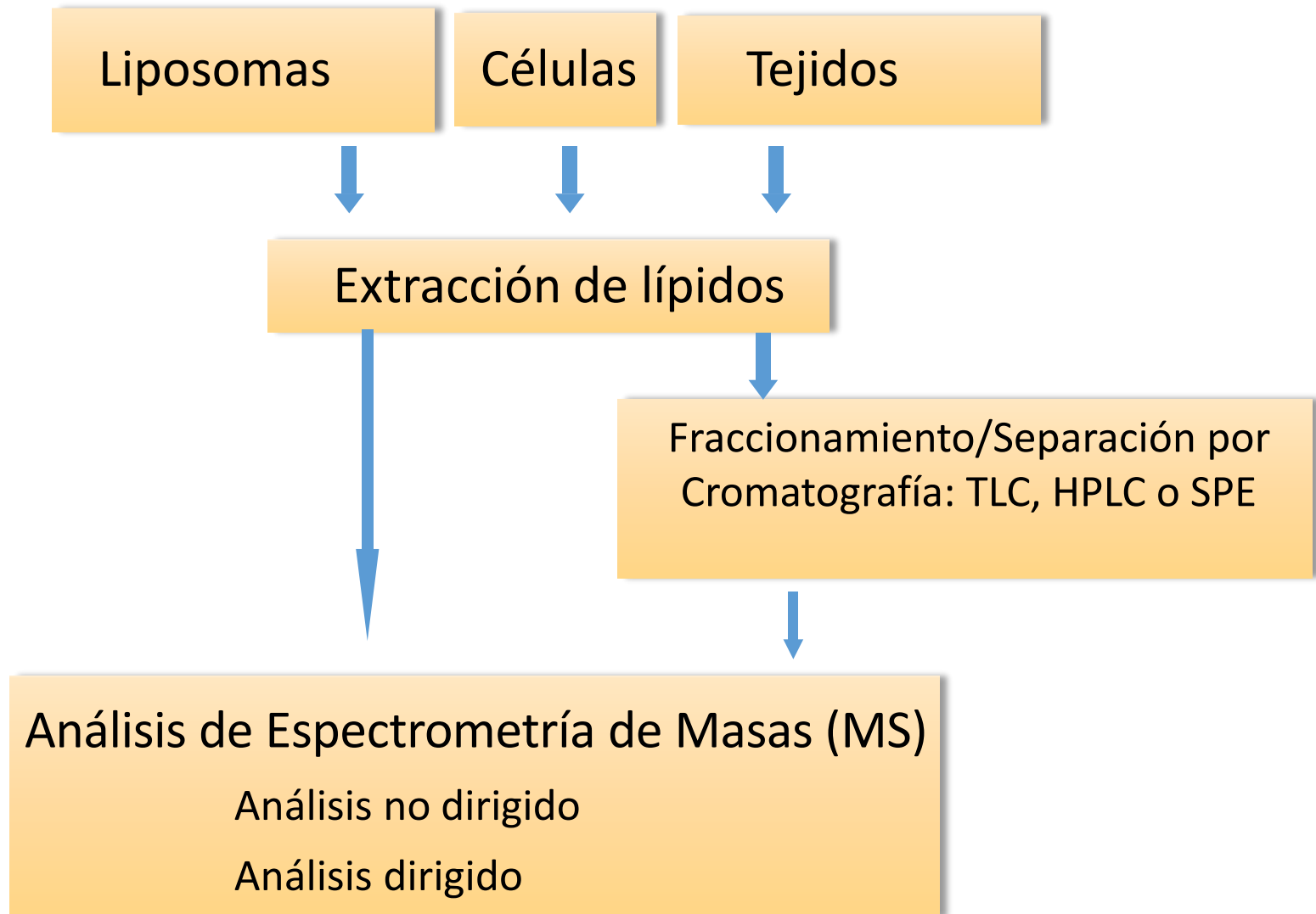
Importancia:

- Nuevos biomarcadores
- Nuevas técnicas terapéuticas
- Nuevas aplicaciones biotecnológicas

# Estrategias analíticas lipidómicas para abordar la complejidad del lipidoma

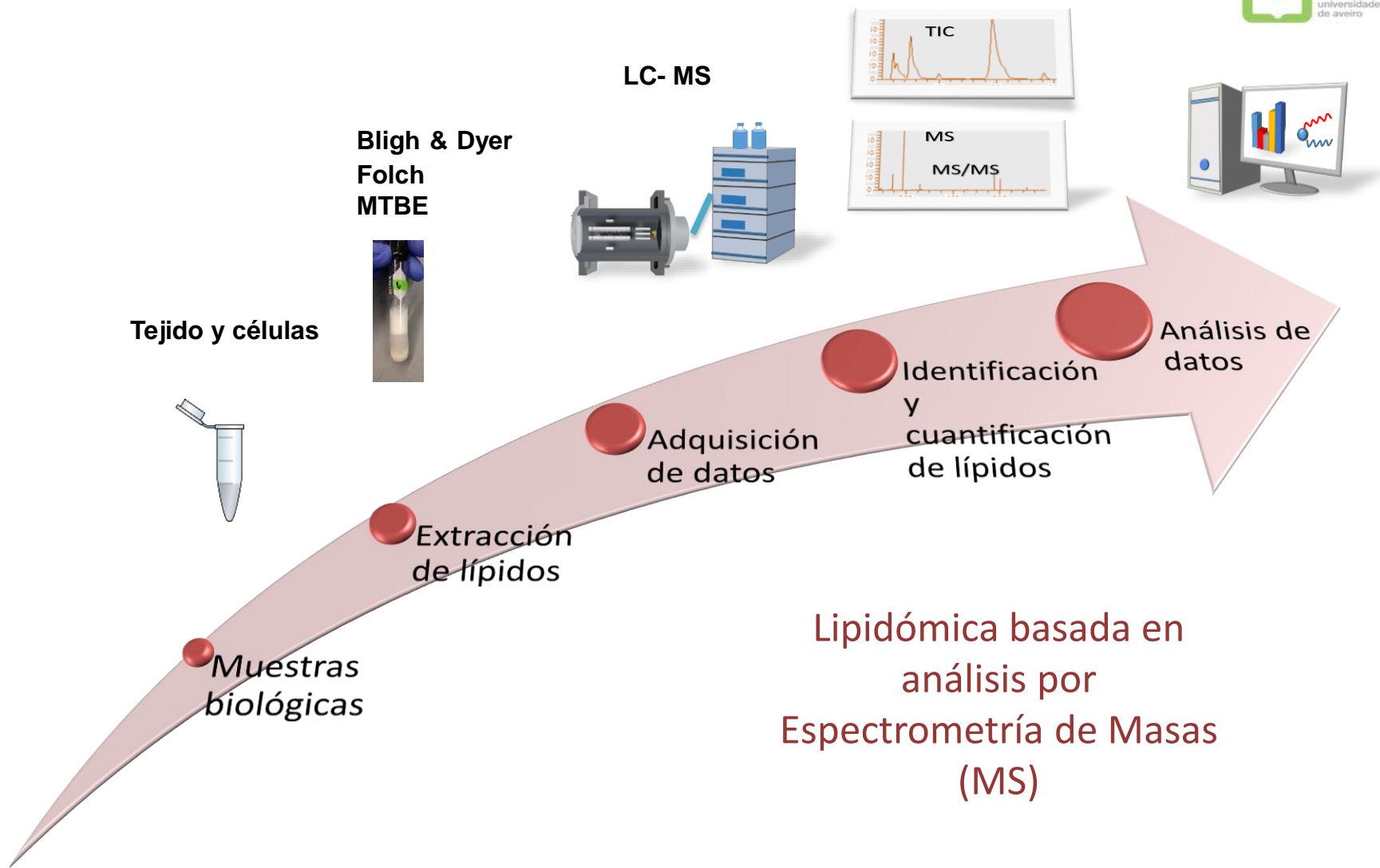


# Enfoque lipidómico





# Flujo de trabajo de la lipidómica



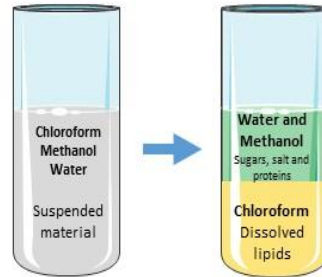




**Muestras biológicas**

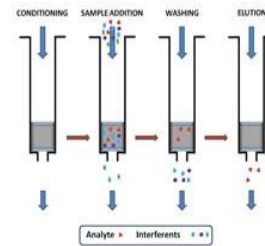
**Lipid extraction and fractionation**  
**Lipid class separation**  
**MS analysis**

**Extracción en fase líquida  
Metanol/cloroformo**



y/o

**Fraccionamiento de lípidos**



**Cromatografía en capa fina TLC**



**Cromatografía de líquidos**



HPLC *Fase inversa*  
 nLC *Fase normal*  
 UPLC *Interacción hidrofílica*

**Adquisición de datos en Espectrometría de Masas**

**Ionización**

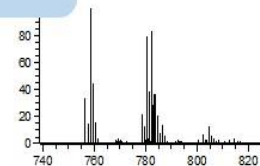
ESI  
MALDI  
APCI/APPI  
SIMS

**Análisis de iones**

*Resolución de masa baja*  
*Resolución de masa alta*

**Detección**

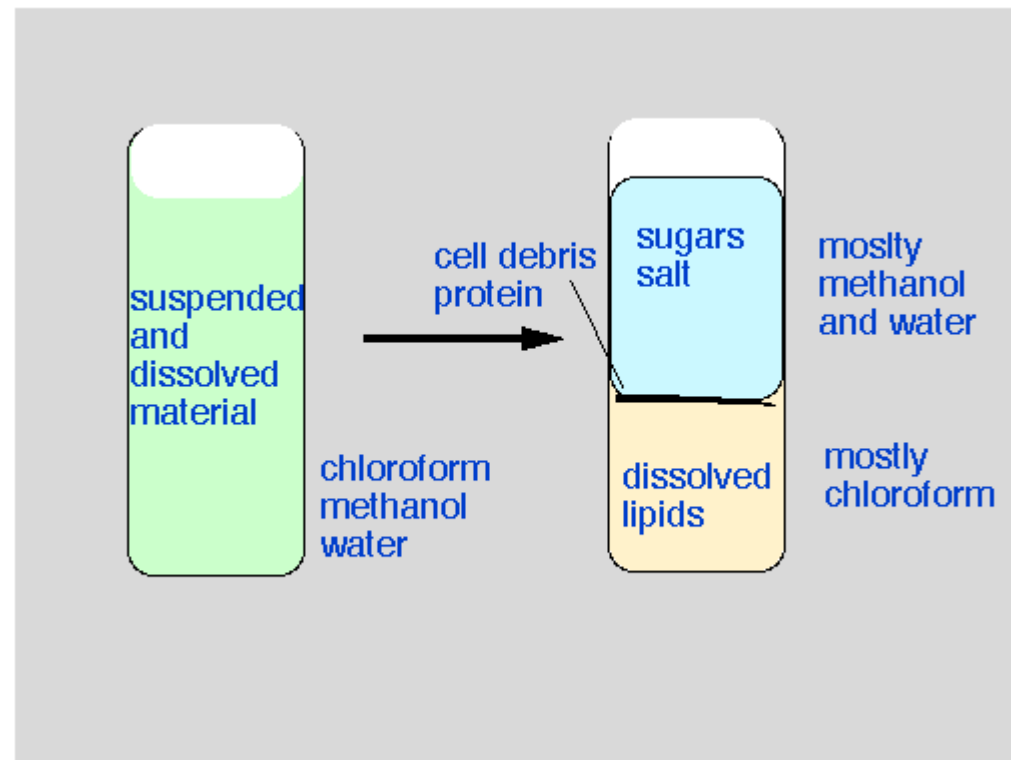
TIC/RIC  
MS  
MS/MS



# Extracción de lípidos

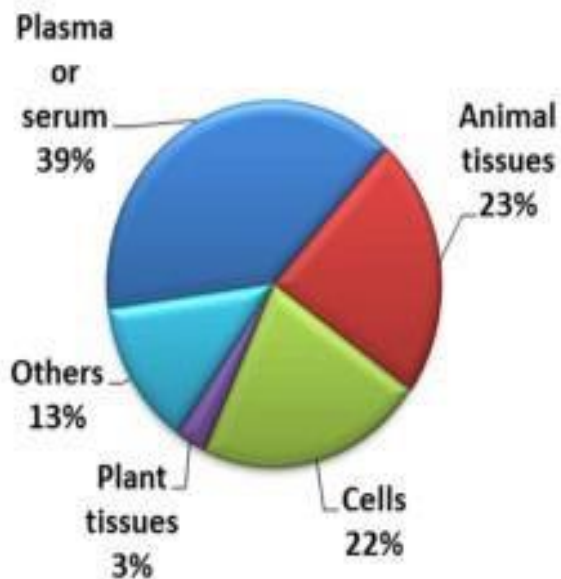
Extracción química usando disolventes orgánicos:

- ❑ Método Folch ( $\text{CHCl}_3:\text{CH}_3\text{OH}$  2:1)
- ❑ Bligh y Dyer ( $\text{CHCl}_3:\text{CH}_3\text{OH}$  1:2)
- ❑ otros

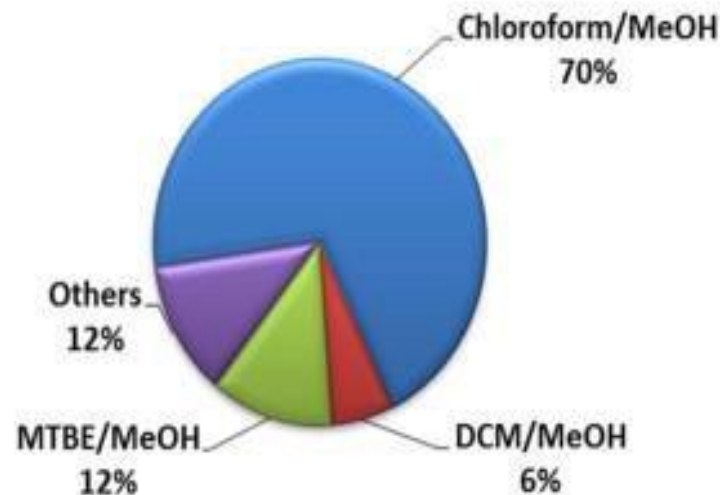


# Extracción de lípidos

## Analyzed Matrices

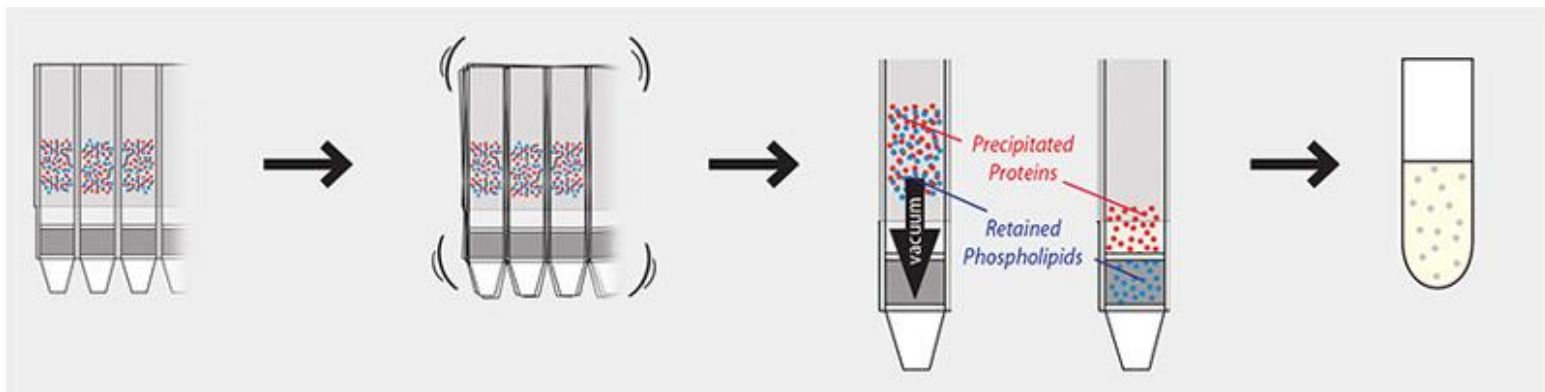
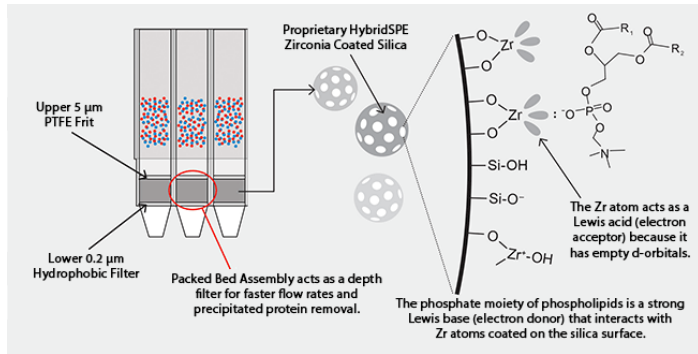


## Extraction protocols



# Extracción de lípidos

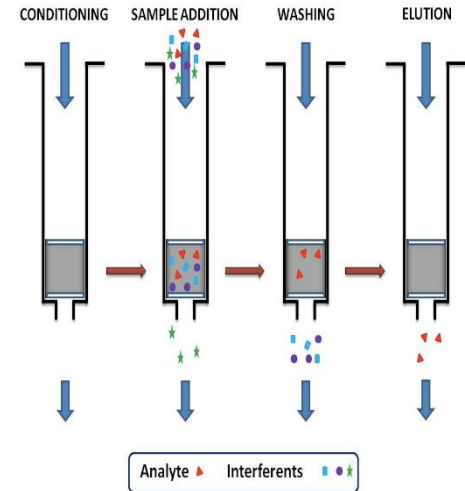
## Extracción selectiva de fosfolípidos del plasma utilizando Hybrid SPE



# Fraccionamiento de extractos lipídicos

## Extracción de fase sólida

Para separar lípidos neutros de polares  
Lípidos neutros (TG) de polares (PL)

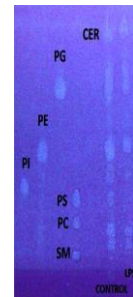


## Métodos cromatográficos

TLC (Cromatografía en capa fina)

HPLC (cromatografía líquida de alta eficacia)

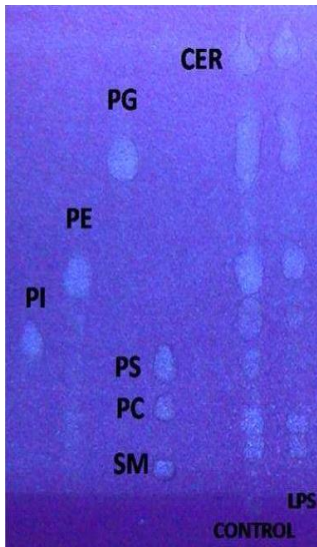
Para separar clases de lípidos/especies moleculares



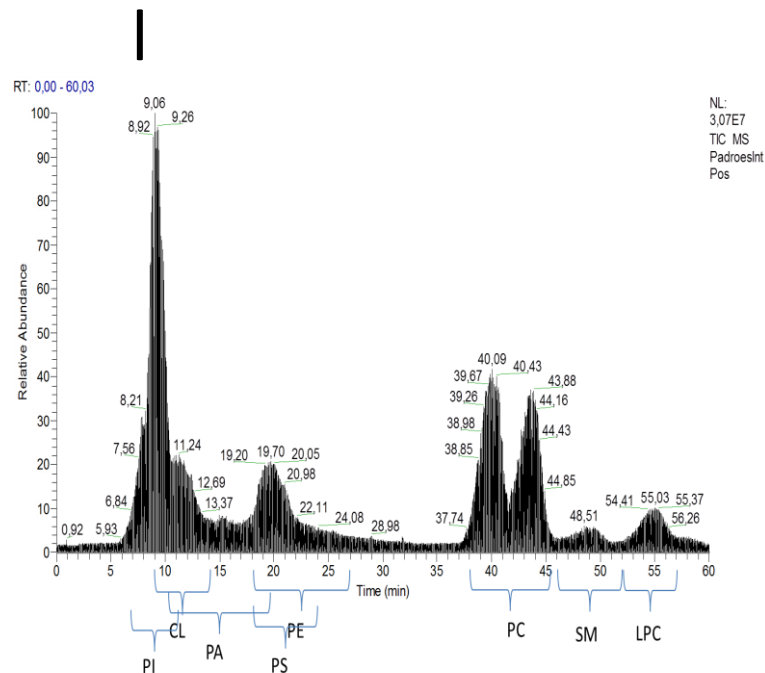
# Separación de clases fosfolipídicas

Las clases fosfolípidas se pueden separar, en función de su polaridad, mediante:

TLC

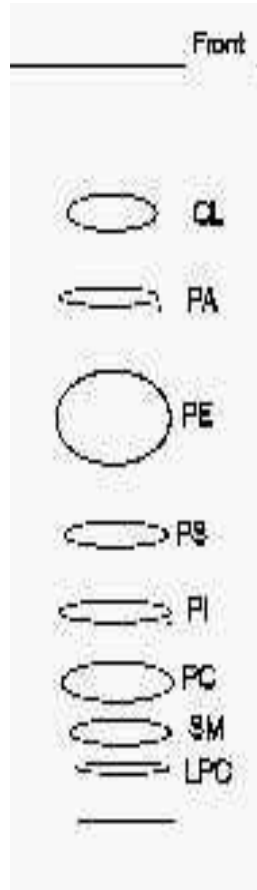


HPLC

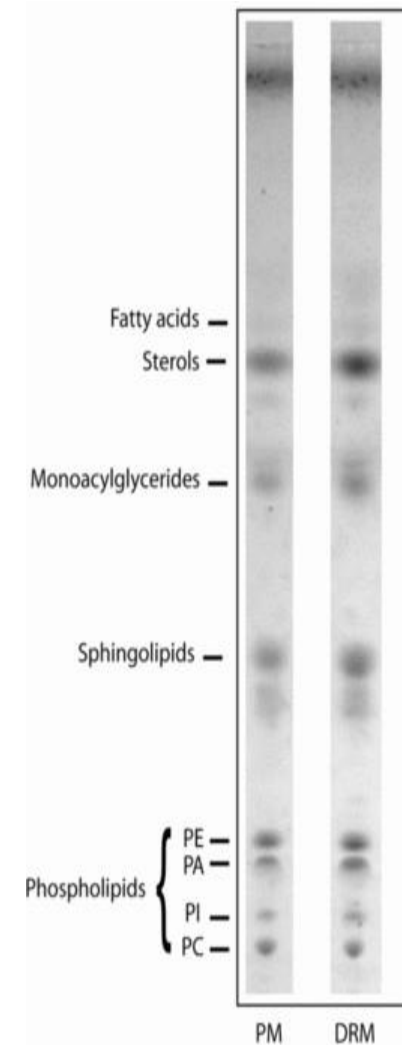


# TLC – Cromatografía en capa fina

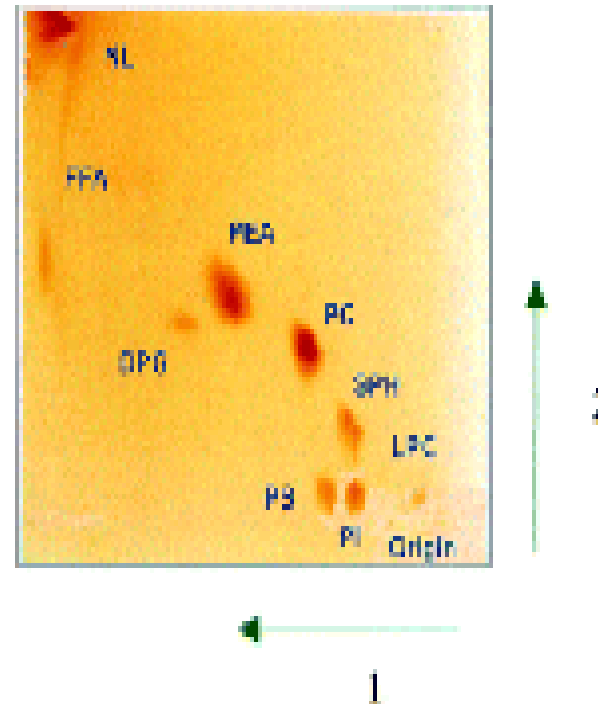
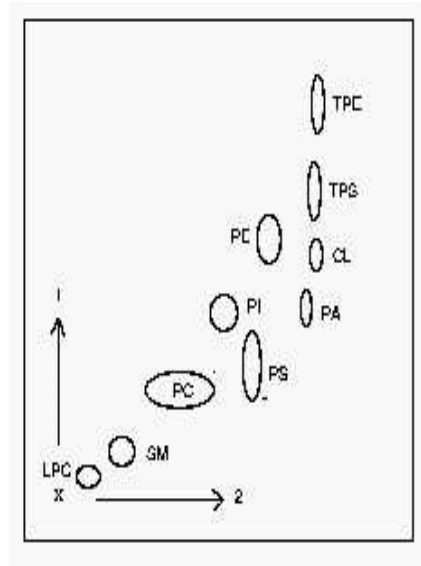
## Diferentes sistemas de elución - diferentes perfiles TLC



- CL-Cardiolipina
- PA-Ácido fosfatídico
- PE-Fosfatidiletanolamina
- PS-Fosfatidilserina
- PI-Fosfatidilinositol
- PC-Fosfatidilcolina
- SM-Esfingomiélinea
- LPC-Lisofosfatidilcolina



# TLC 2-D



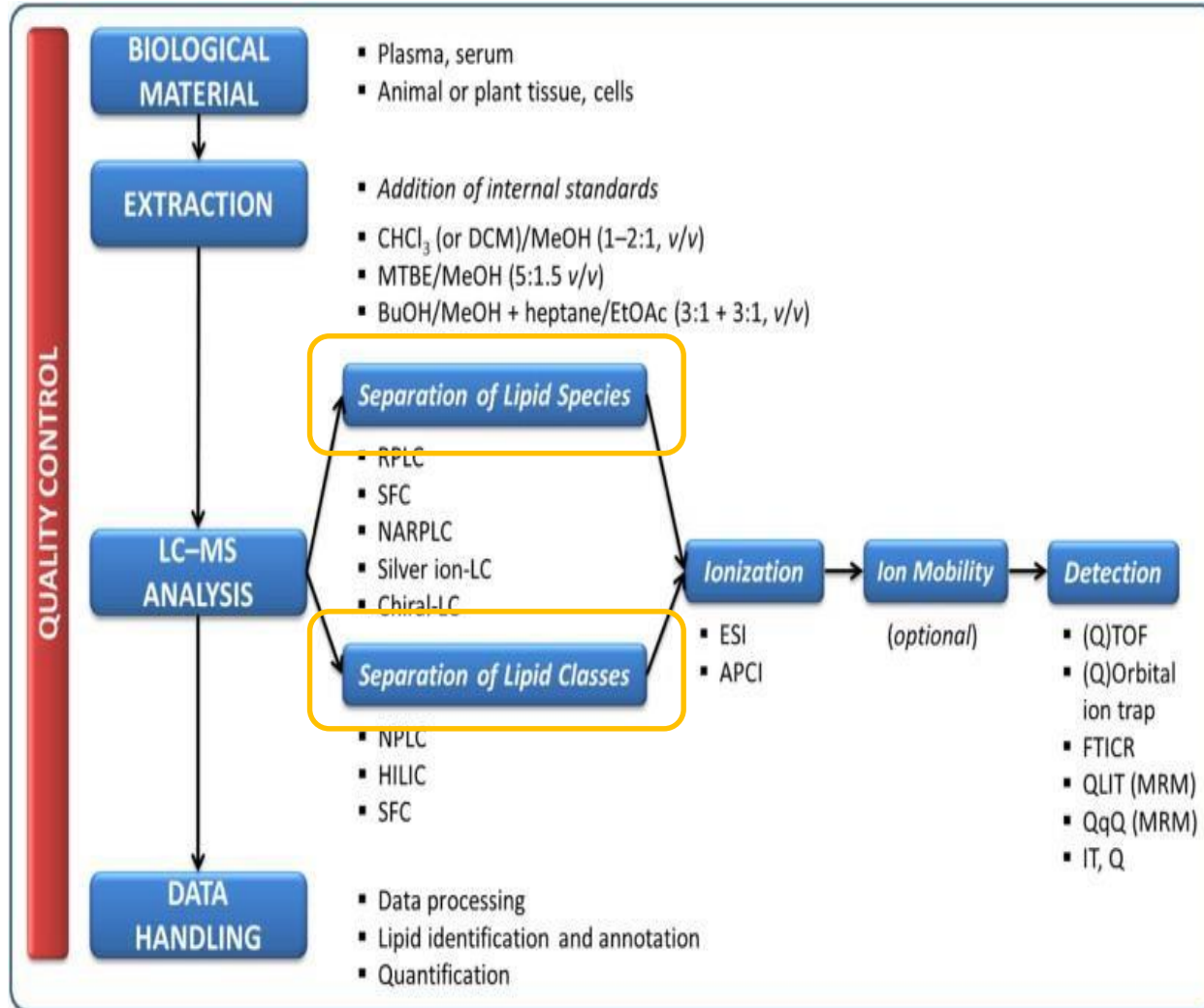
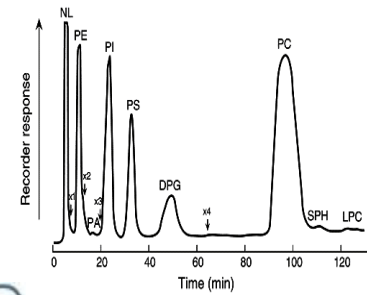
Dos sistemas disolventes distintos

Prof Valerian Kagan lab

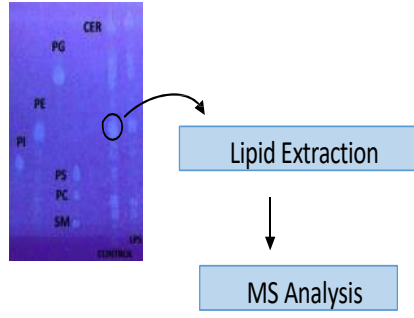




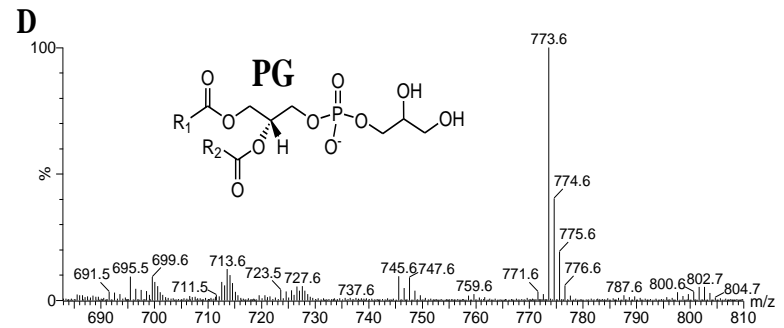
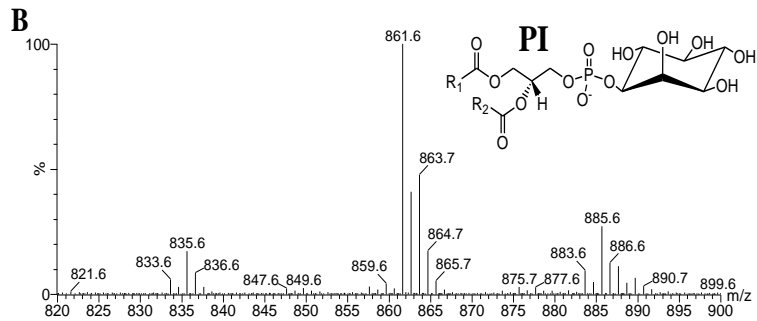
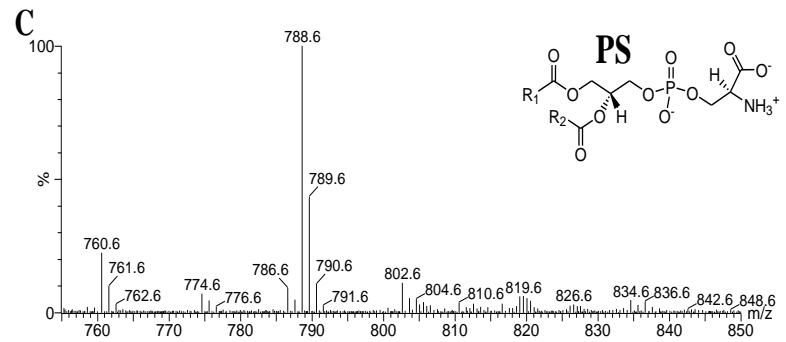
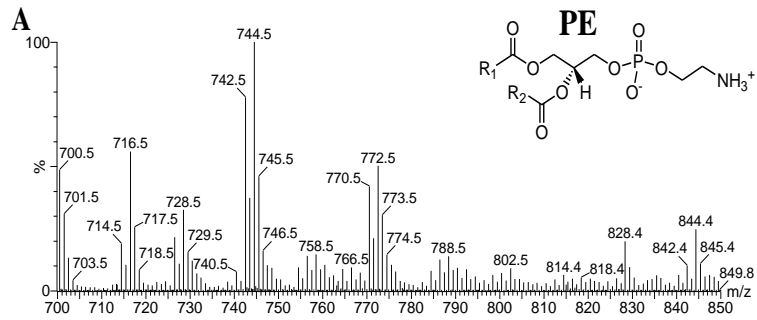
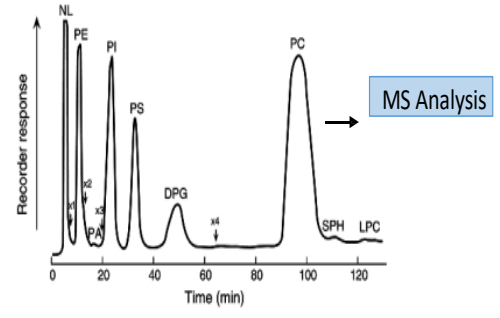
# HPLC-MS



### off-line TLC-MS



### on-line HPLC-MS

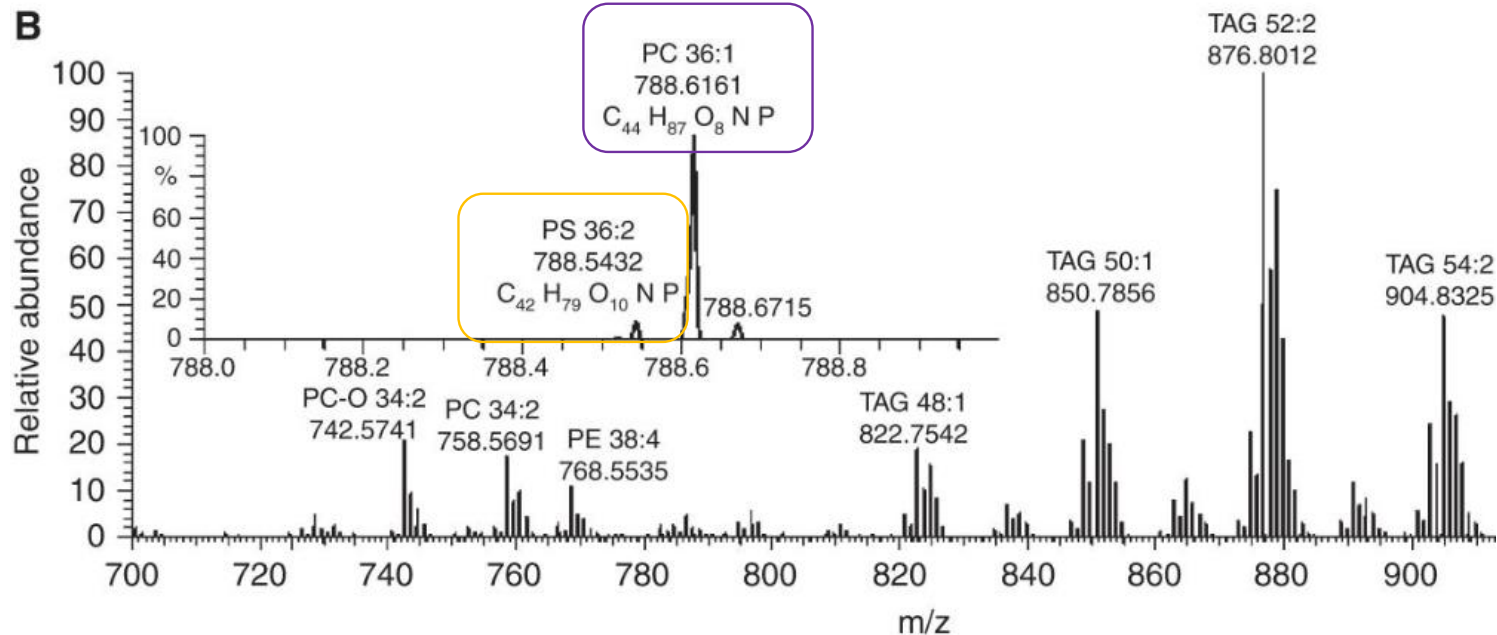


# ANÁLISIS DE DATOS DE MS:

## Iones formados durante la ionización de lípidos

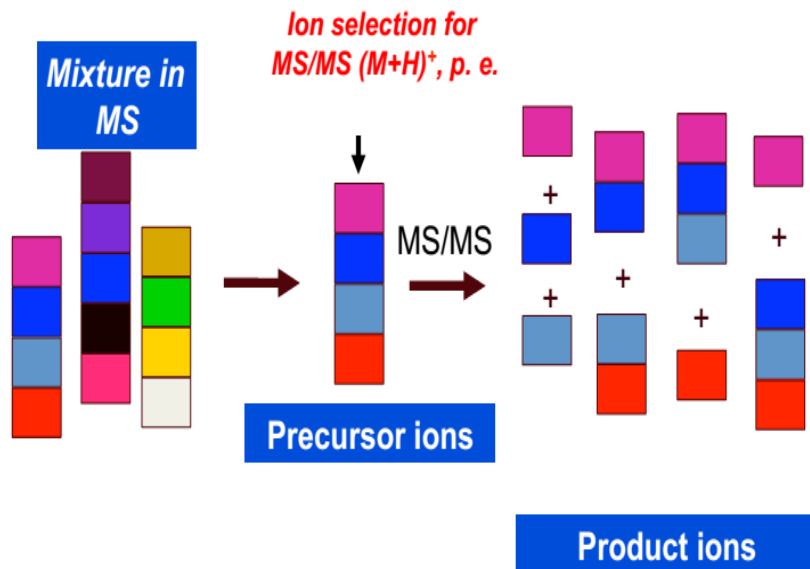
Lipid class	Positive mode	Negative mode
LPC, PC	$[M + H]^+$ , $[M + Na]^+$	$[M - H]^-$ , $[M + HCOO]^-$ , $[M + CH_3COO]^-$
<b>LPE, PE</b>	<b><math>[M + H]^+</math>, <math>[M + Na]^+</math></b>	<b><math>[M - H]^-</math></b>
PG	$[M + H]^+$ , $[M + NH_4]^+$ , $[M + Na]^+$	$[M - H]^-$
PI	$[M + H]^+$ , $[M + NH_4]^+$ , $[M + Na]^+$	$[M - H]^-$
<b>PS</b>	<b><math>[M + H]^+</math></b>	<b><math>[M - H]^-</math></b>
PA		$[M - H]^-$
CE	$[M + NH_4]^+$ , $[M + Na]^+$	
SM	$[M + H]^+$	$[M + HCOO]^-$ , $[M + CH_3COO]^-$
Cholesterol	$[M - H_2O + H]^+$	
MG, DG, TG	$[M + NH_4]^+$ , $[M + Na]^+$	
MGDG, DGDG, SQDG	$[M + NH_4]^+$ , $[M + Na]^+$	$[M - H]^-$
Fatty acids		$[M - H]^-$
CL	$[M + H]^+$ , $[M + NH_4]^+$ , $[M + Na]^+$	$[M - H]^-$ , $[M - 2H]^{2-}$
Cer, GluCer, LacCer	$[M + H]^+$ , $[M + NH_4]^+$ , $[M + Na]^+$	$[M - H]^-$ , $[M + HCOO]^-$ , $[M + CH_3COO]^-$

- Alta exactitud de masa:
  - cálculo de peso molecular
  - composición elemental y determinación de la fórmula molecular
  - estructura molecular
  - Podrían distinguirse iones moleculares de especies isobáricas (mismo valor  $m/z$  pero distintas fórmula y estructura moleculares)

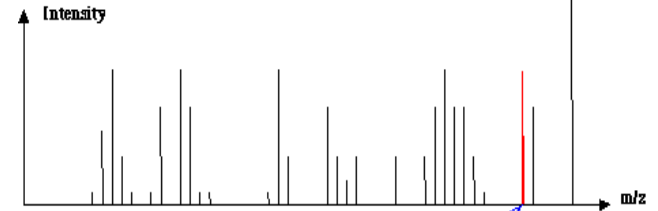


## Fragmentación:

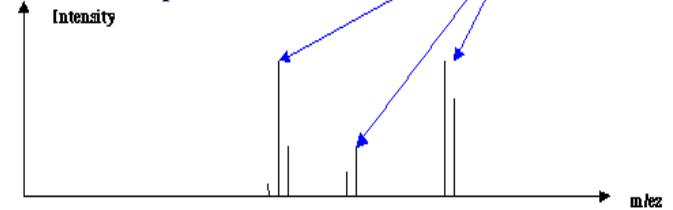
- Selección del ion de interés en MS
- Formación de iones fragmento en MS/MS
- Información estructural



*MS Spectrum*



*MS/MS Spectrum*



La interpretación del espectro de MS/MS es como resolver un puzle



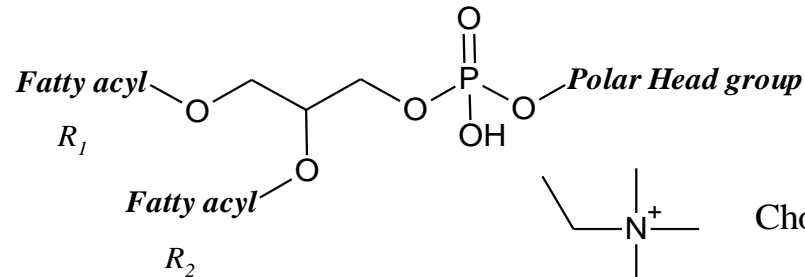
Nos permite obtener información estructural sobre el compuesto inicial

# Espectrometría de masas en tándem (MS/MS)

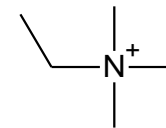
AALifeSci 



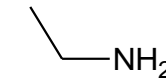
# Glicerofosfolípidos o fosfolípidos (PL)



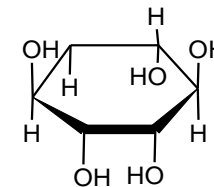
$R_1$  and  $R_2$  are used to designate undefined alkyl groups at the *sn*-1 and *sn*-2 positions, respectively



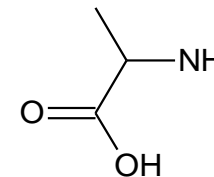
Choline



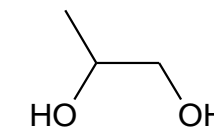
Ethanolamine



Inositol



Serine



Glycerol



Hidrógeno

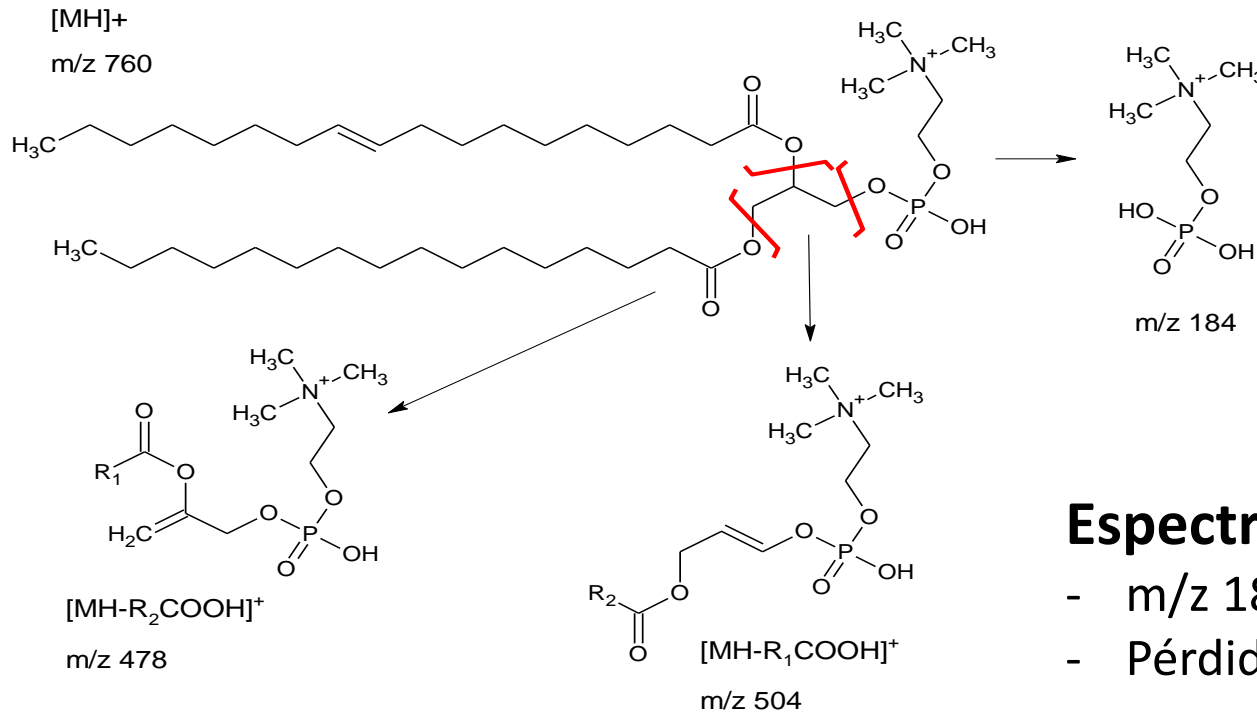
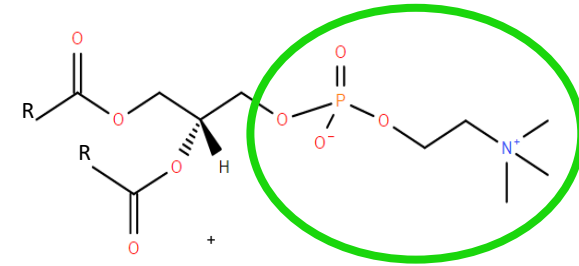
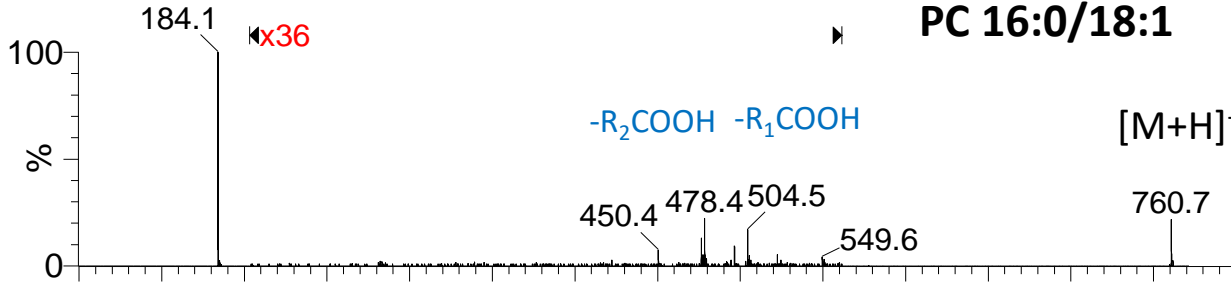
Información necesaria para confirmar:

- Grupo de cabeza polar
- Ácidos grasos

La fragmentación depende de:

- El tipo de ion precursor
- La energía de colisión
- otros

# Fosfatidilcolina– MS/MS [M+H]<sup>+</sup>



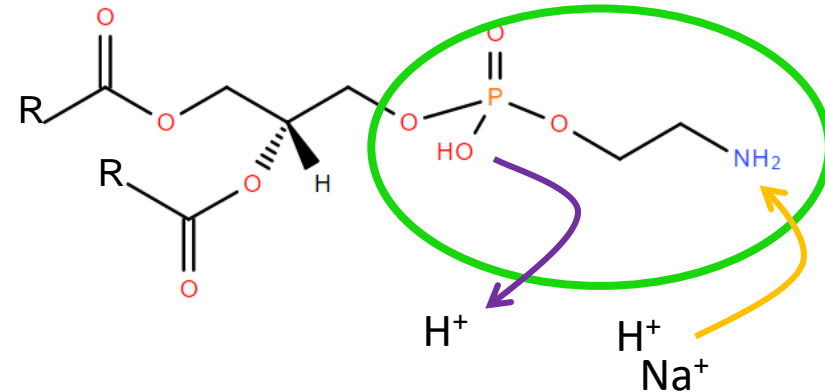
## Espectro MS/MS

- m/z 184
- Pérdida de ácidos grasos



# Fosfatidiletanolamina- PE

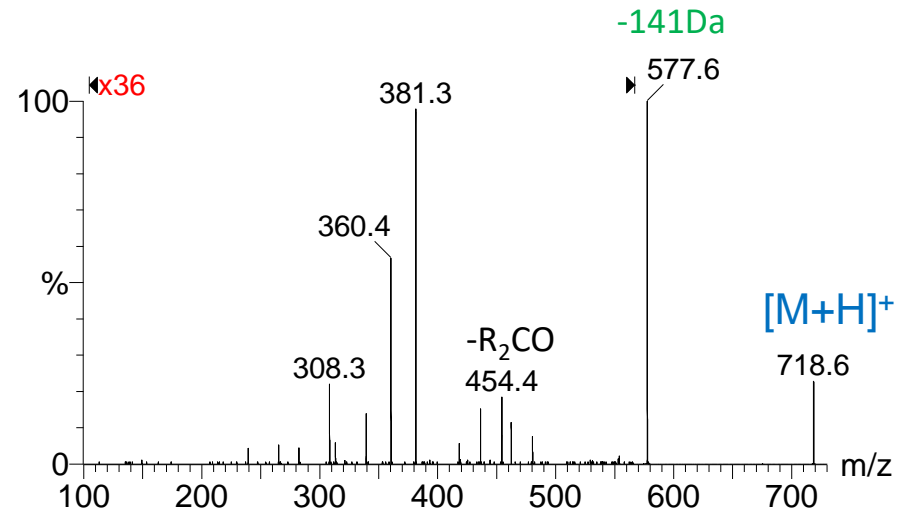
- Modo Positivo  $[M+H]^+$
  - Modo Negativo  $[M-H]^-$
- ESI-MS/MS  $[M+H]^+$



Pérdida característica de **141 Da**

Pérdida de  $RCOOH$  y  $RC=O$   
( $R_1=CO^+$  y  $R_2=CO^+$ ).

PE 16:0/18:1

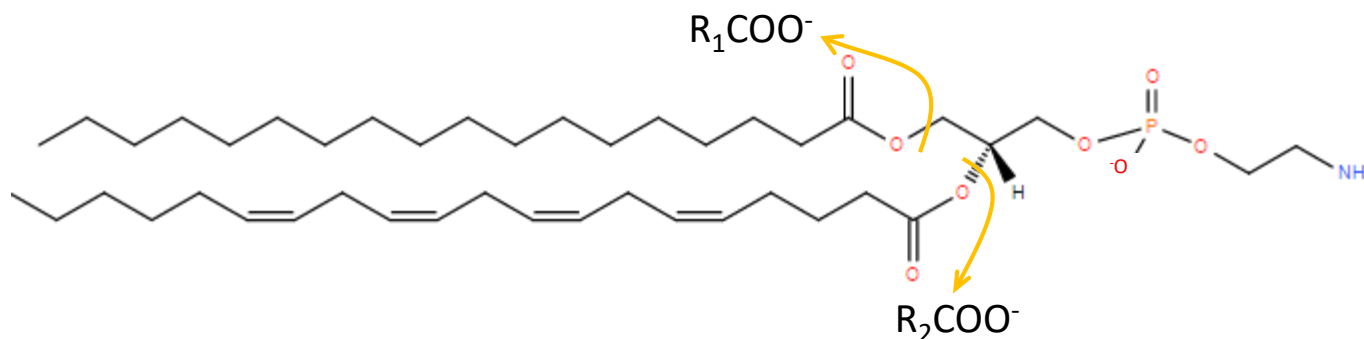
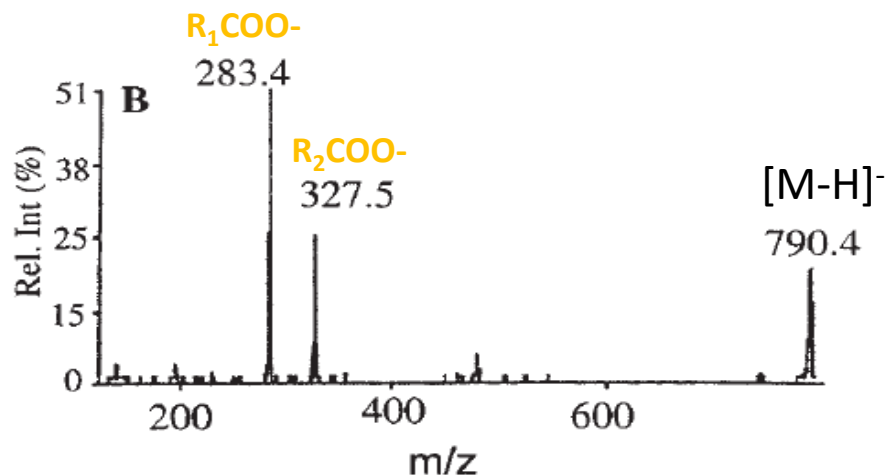


# Fosfatidiletanolamina – MS/MS

## ESI-MS/MS $[M-H]^-$

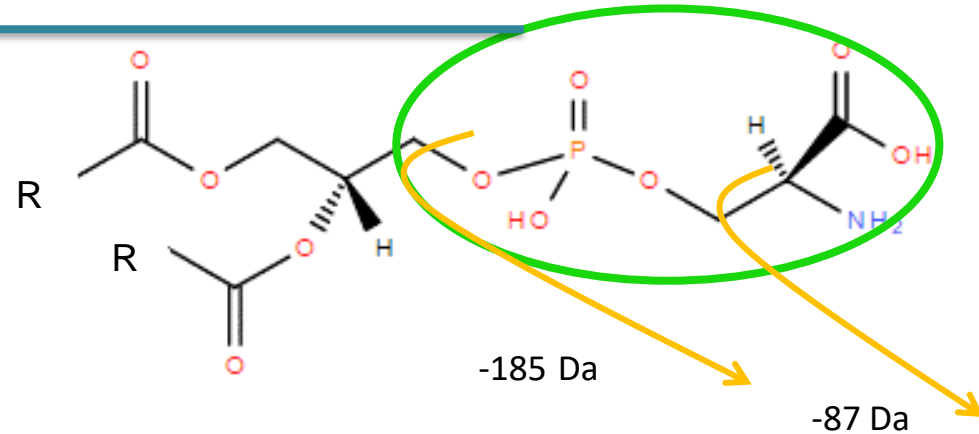
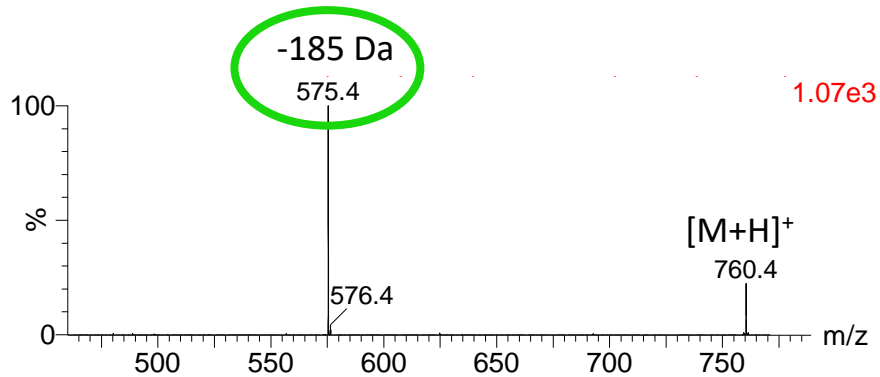
Espectro MS/MS en modo negativo

- Ion  $R_1COO^-$
- Ion  $R_2COO^-$



# Fosfatidilserinas – PS

## ESI-MS/MS $[M+H]^+$ POPS

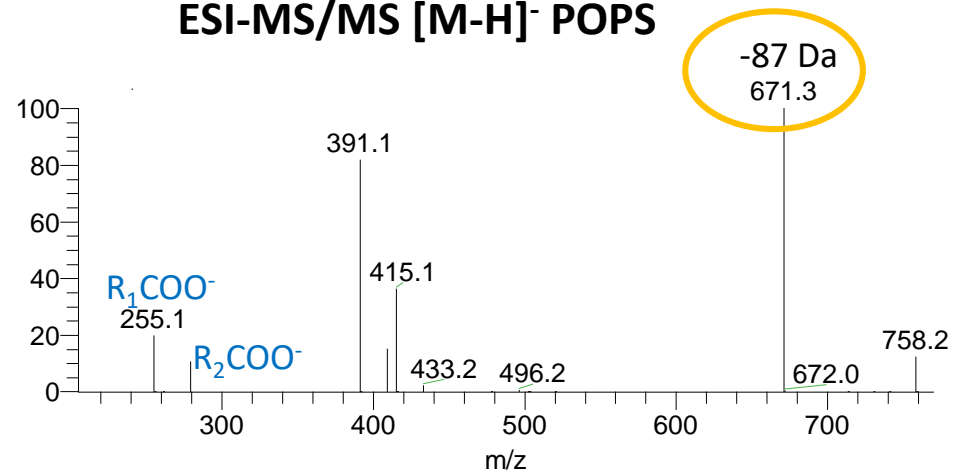


MS/MS  $[M+H]^+$  **Pérdida característica de 185 Da**

## ESI-MS/MS de $[M-H]^-$

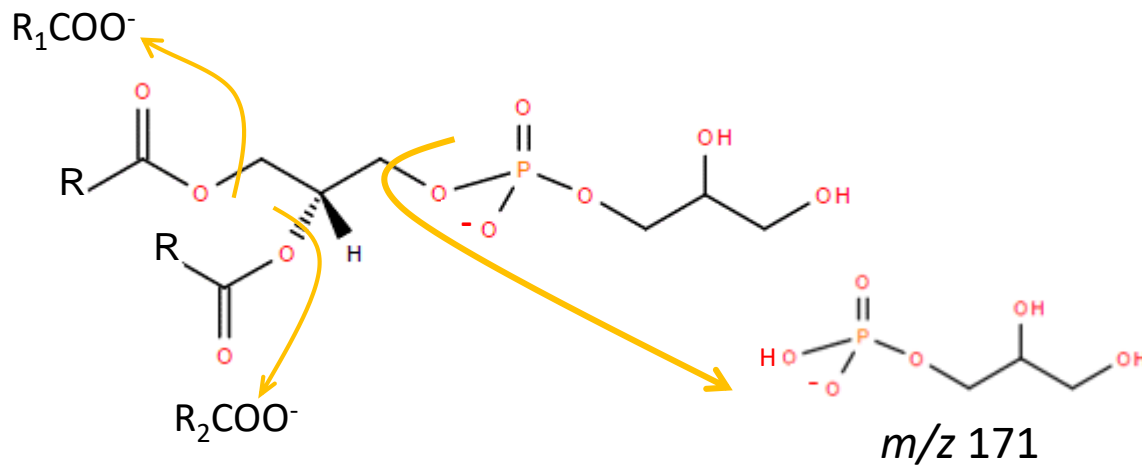
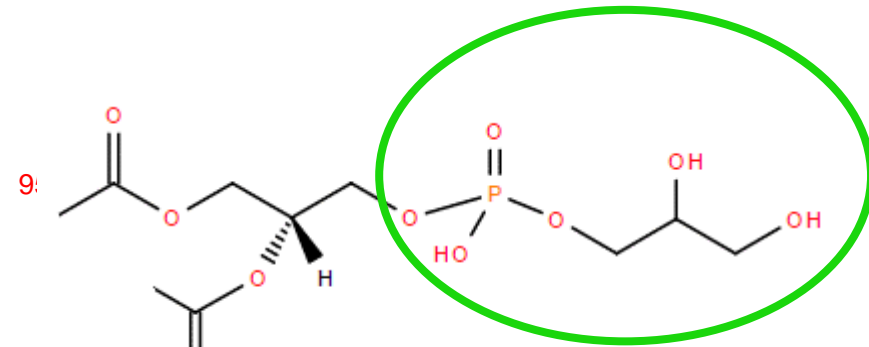
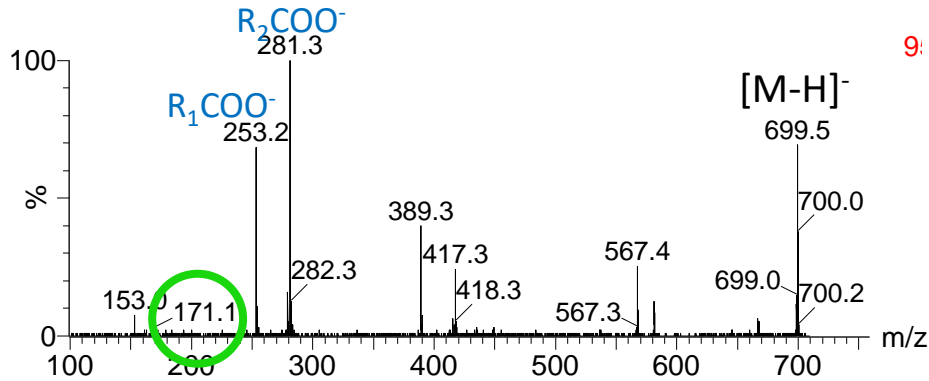
- Pérdida de 87 Da
- Ion  $R_1COO^-$
- Ion  $R_2COO^-$

## ESI-MS/MS $[M-H]^-$ POPS



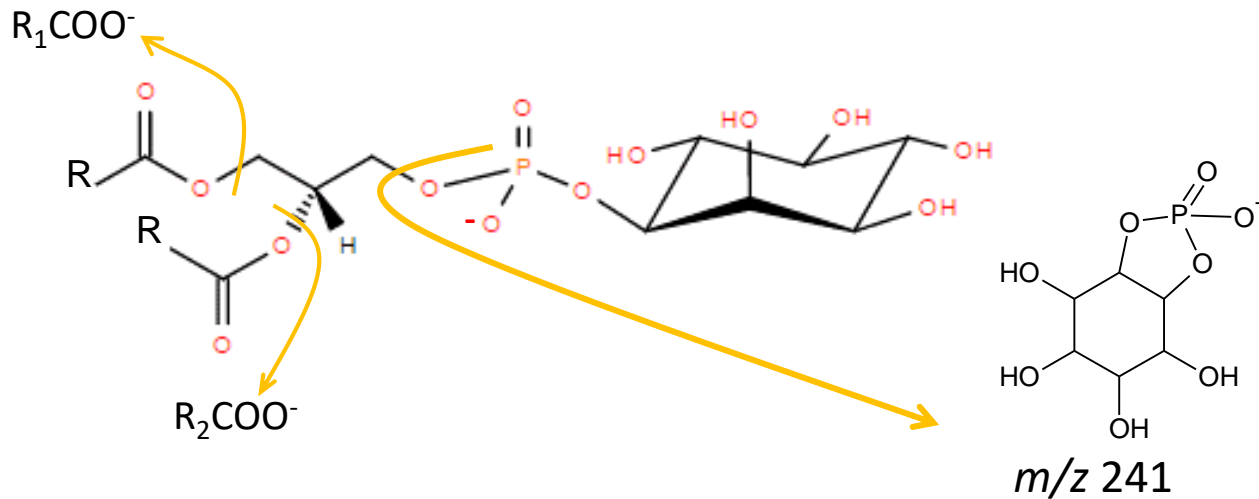
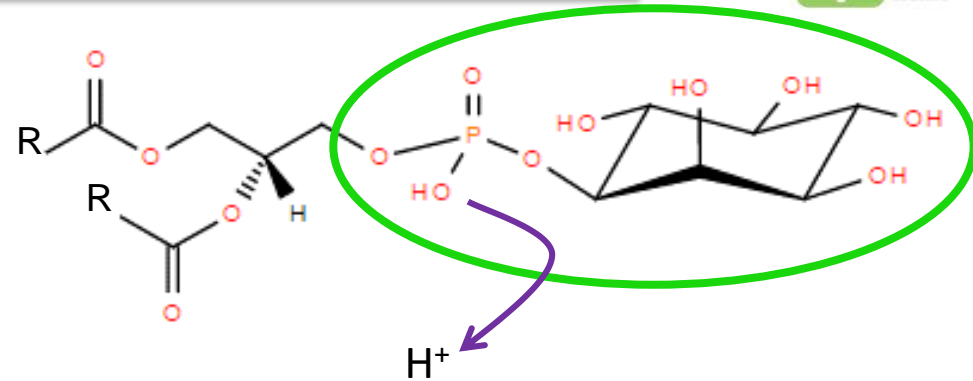
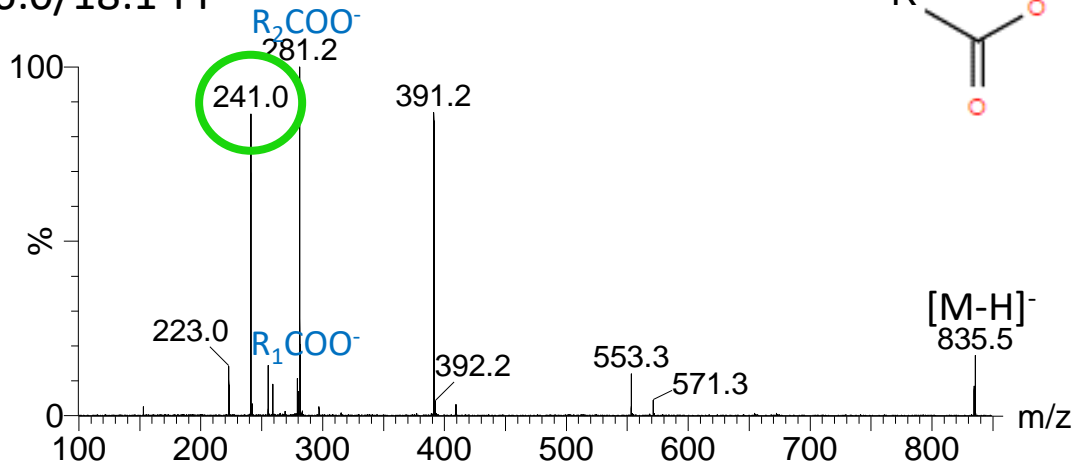
# Fosfatidilglicerol - PG

Espectro de ESI-MS/MS em modo negativo



# Fosfatidilinositol -PI

ESI-MS/MS de  $[M-H]^-$   
 16:0/18:1-PI



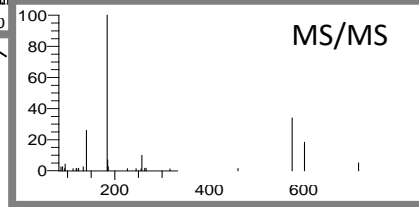
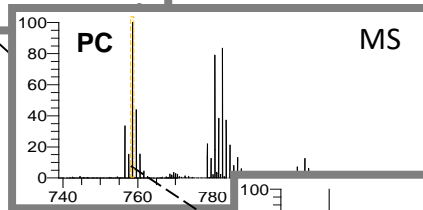
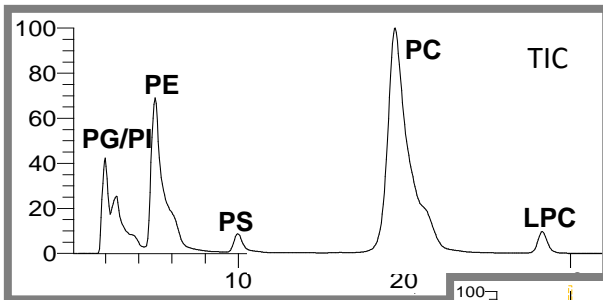
# Clases de fosfolípidos y MS/MS

	Positive ion	Negative ion
Headgroup		
Phosphatidylcholine	Precursor ion $m/z$ 184	—
Phosphatidylserine	Neutral loss 185 Da	Neutral loss 87 Da
Phosphatidylethanolamine	Neutral loss 141 Da	—
Phosphatidylinositol	—	Precursor ion $m/z$ 241
Sphingomyelin	Precursor ion $m/z$ 184	

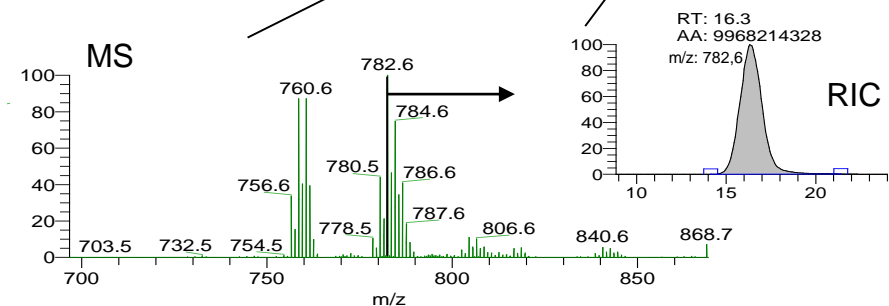
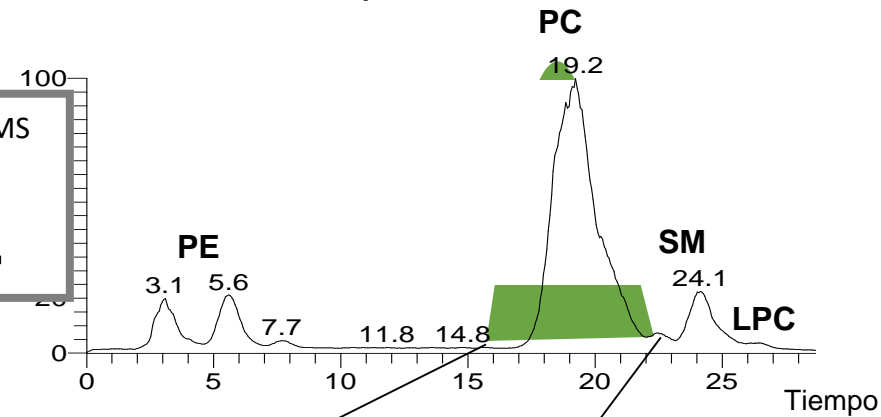
## Información sobre la composición de acilos grasos

- Modo positivo
  - Pérdida de  $\text{RCOOH}$  y  $\text{R=C=O}$
- Modo negativo
  - Formación de aniones carboxilados  $\text{RCOO}^-$

# Análisis de Lipidómica no Dirigida



TIC  
Modo ionización positivo



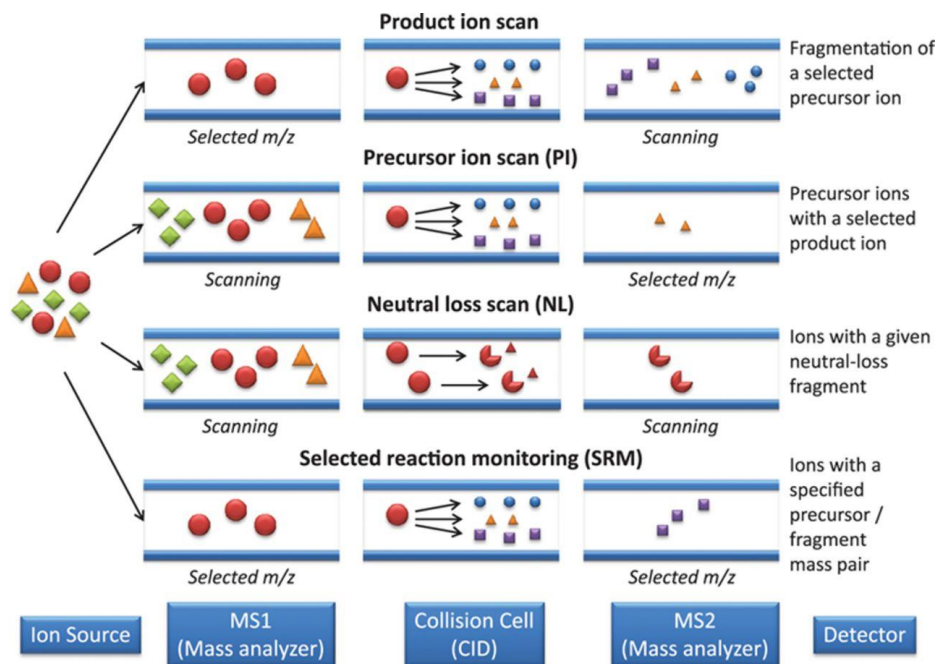
Herramientas bioinformáticas:

- Mzmine
- Compound Discover
- Lipidizer
- LipidBlast
- Otros

# Análisis de Lipidómica Dirigida

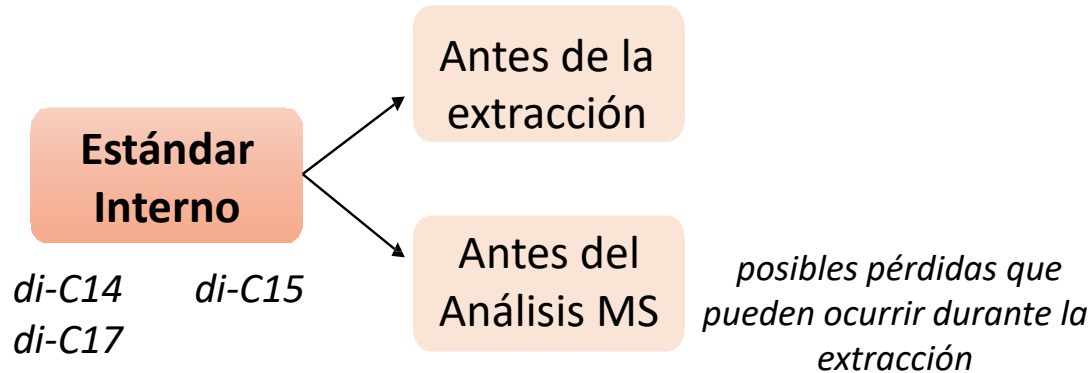
## Lipidómica por Infusão Directa

	Positive ion	Negative ion
Headgroup		
Phosphatidylcholine	Precursor ion $m/z$ 184	—
Phosphatidylserine	Neutral loss 185 Da	Neutral loss 87 Da
Phosphatidylethanolamine	Neutral loss 141 Da	—
Phosphatidylinositol	—	Precursor ion $m/z$ 241
Sphingomyelin	Precursor ion $m/z$ 184	



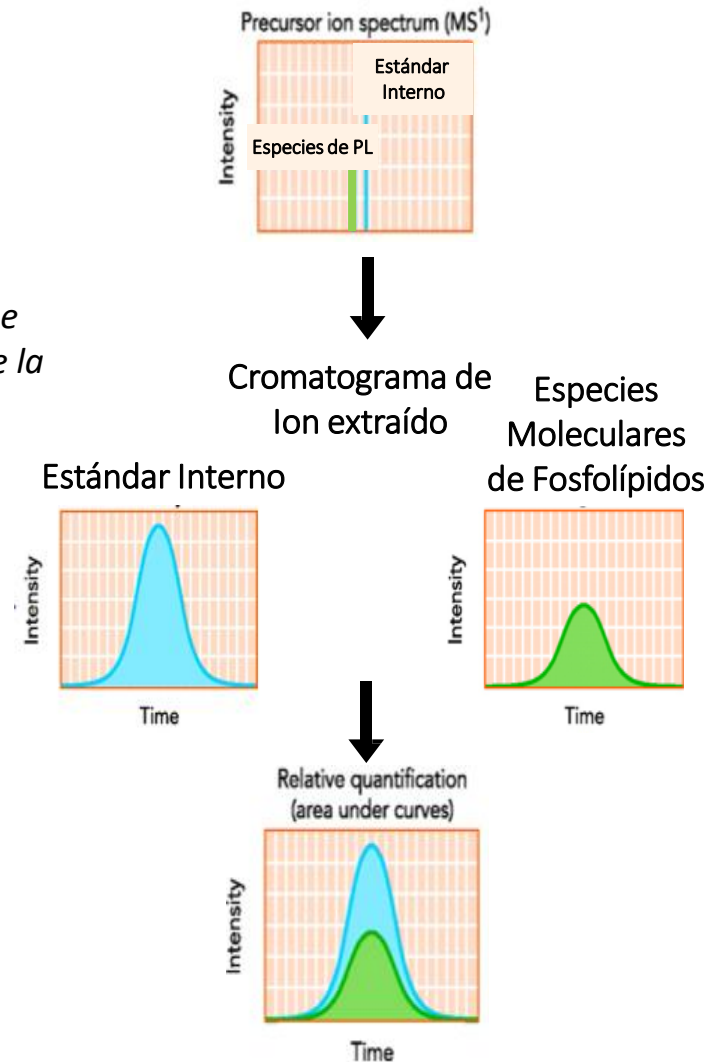


# Cuantificación mediante LC MS



- Especies de lípidos distintas presentan distintas respuestas cuantitativas
- Un estándar interno por cada clase de fosfolípido

Normalización de cada especie molecular al estándar interno de la clase correspondiente





Este proyecto ha sido financiado con ayuda de la Comisión Europea.

La presente publicación recoge únicamente las opiniones de los autores, por lo que la Comisión no se hace responsable de cualquier uso que se haga de la información contenida en ella.



# Proteómica basada en la espectrometría de masas

Pedro Domingues  
Rosário Domingues  
Rita Ferreira  
Tânia Melo  
Eliana Alves  
Elizabete Maciel



Erasmus+

AACLifeSci



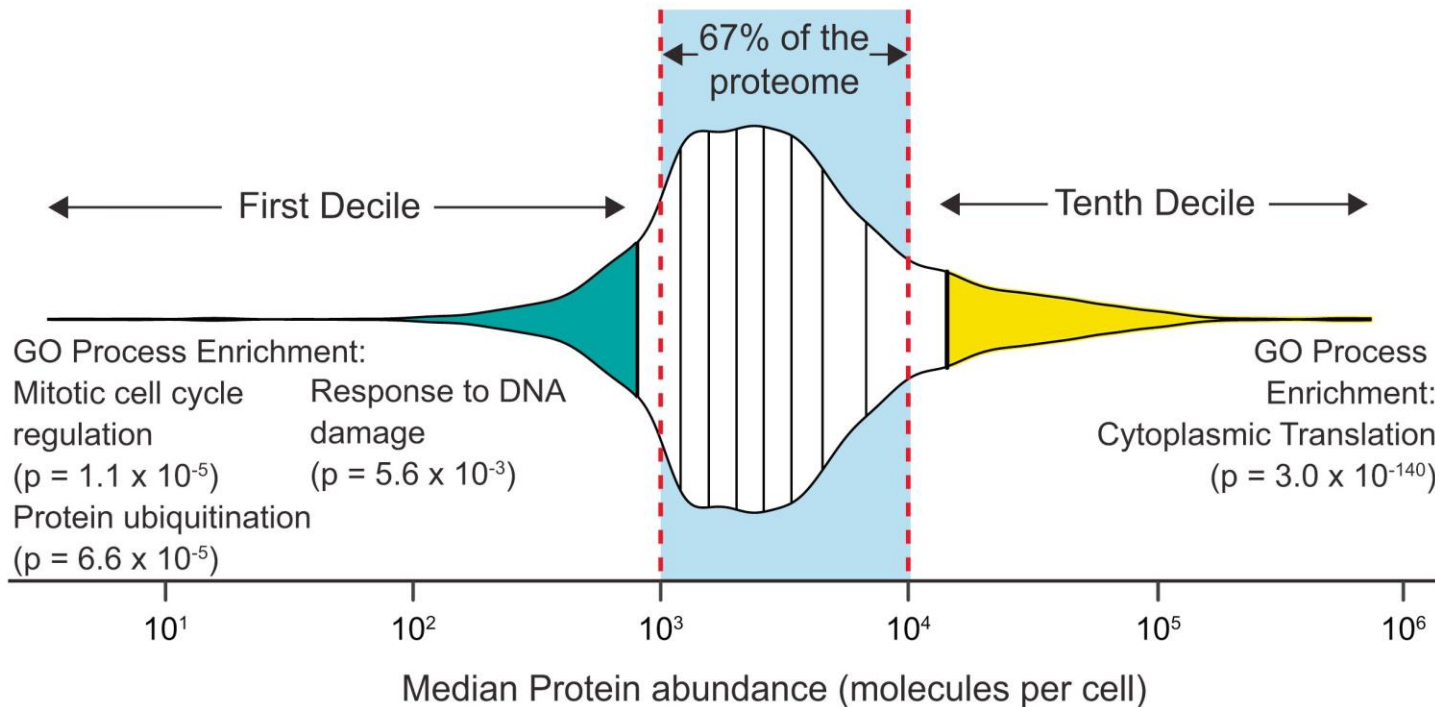
# Proteoma

*S. Cerevisiae*

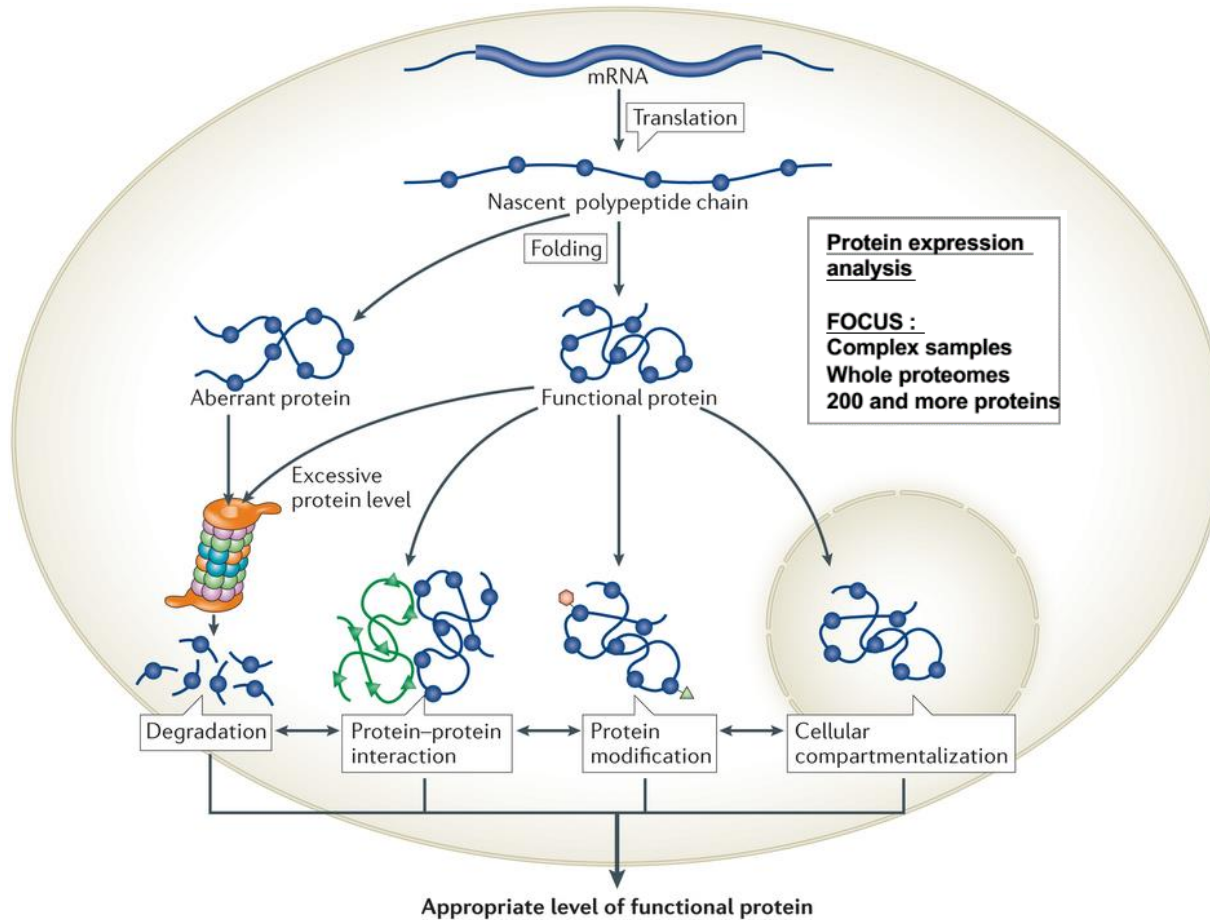
Número de proteínas (proteoma): 5.858

Proteínas/células totales: 42 millones

**c**



# Proteómica



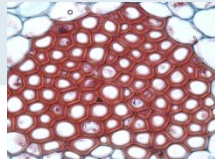
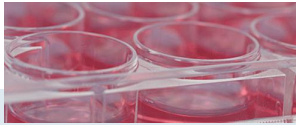
**Protein expression analysis**  
**FOCUS :**  
Complex samples  
Whole proteomes  
200 and more proteins

**Interaction / Functional Proteomics**  
**FOCUS :**  
Subcellular fraction  
Organelle  
Protein Complex  
  
1-200 proteins

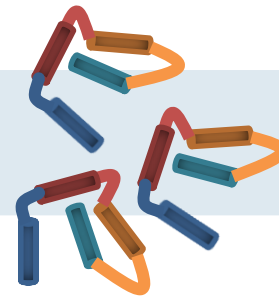
**PTM analysis**  
**FOCUS :**  
Single protein  
  
1-20 proteins

Nature Reviews | Molecular Cell Biology

# Proteómica



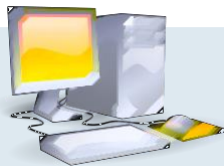
Muestra  
Tejido o cultivo celular



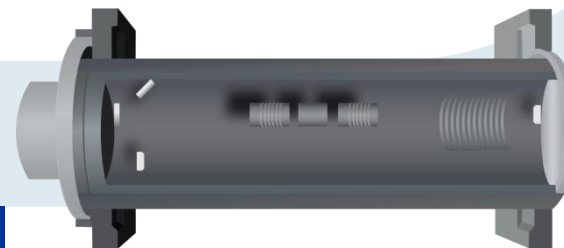
Extracción de proteínas y  
purificación



Digestión de la proteína y  
purificación de péptidos

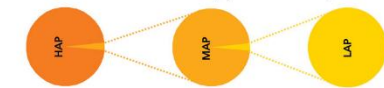
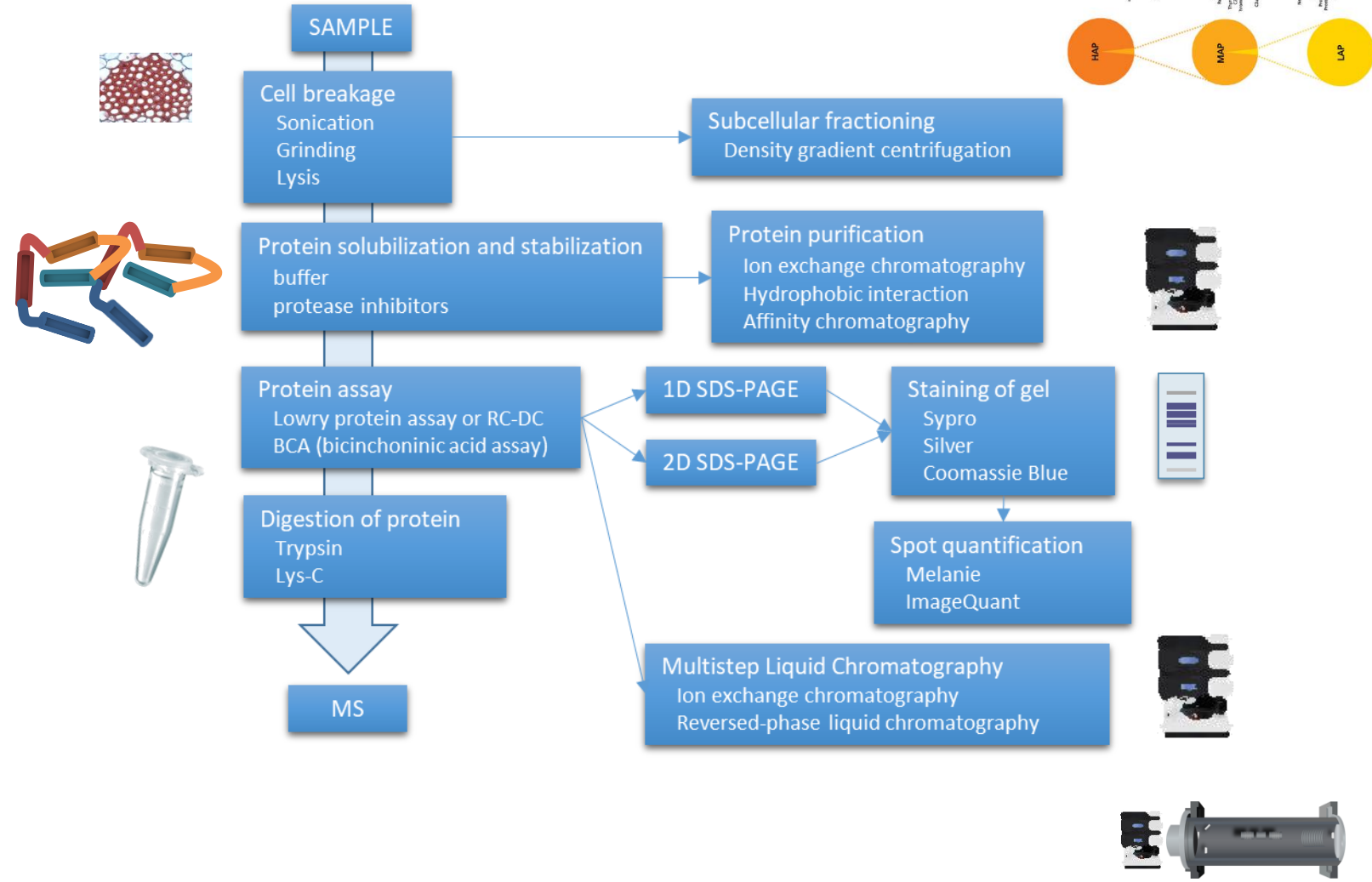
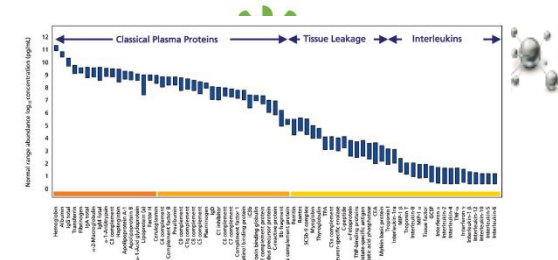


Análisis de datos y  
automatización



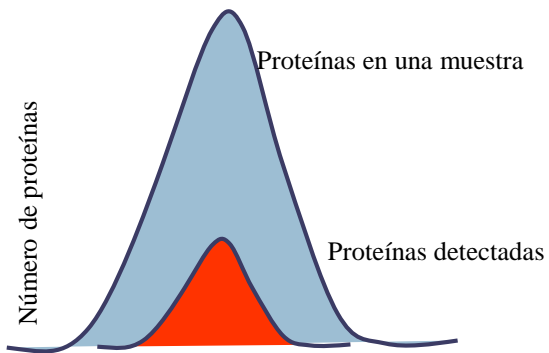
Análisis por LC-MS

# Proteómica basada en MS



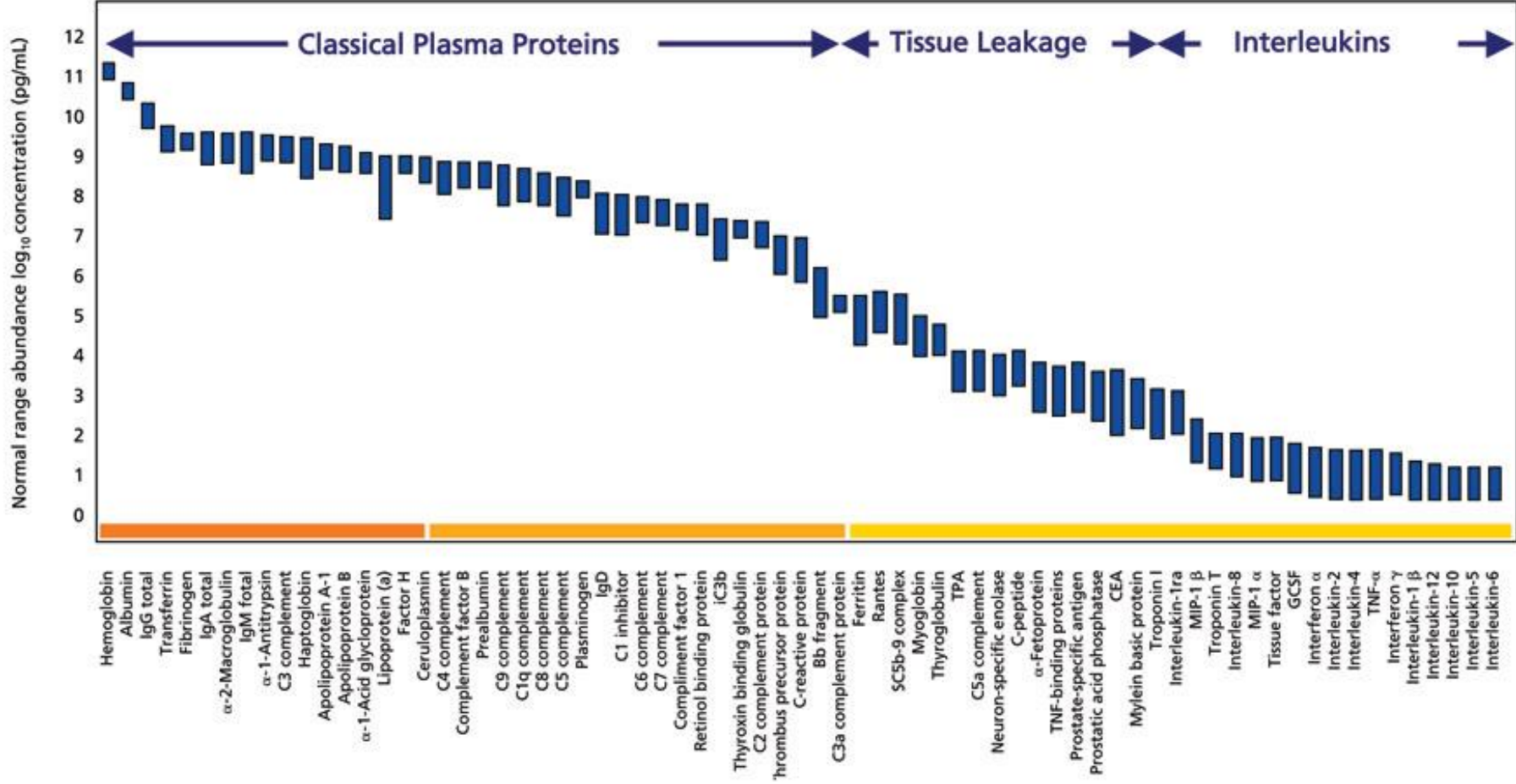
# Sensibilidad

- **Sensibilidad**
  - Respuesta del sistema en una curva de calibración
- **Límite de detección**
  - Relación señal/ruido (S/N)
  - Depende de la matriz
- En las ómicas:

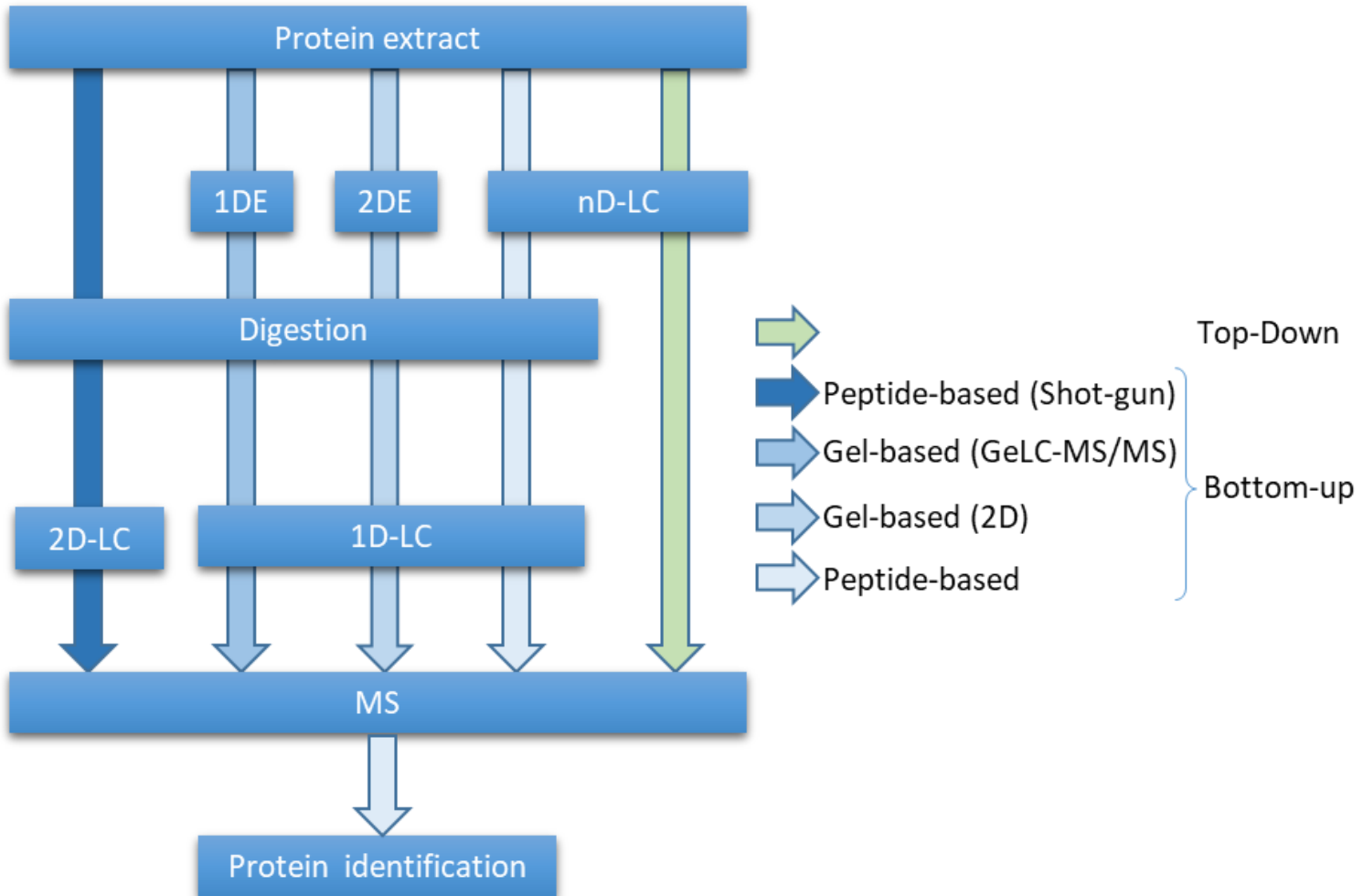


$$\text{Sensibilidad} = \frac{\text{n}^{\circ} \text{ resultados verdaderos no rechazados}}{\text{n}^{\circ} \text{ total de verdaderos}}$$

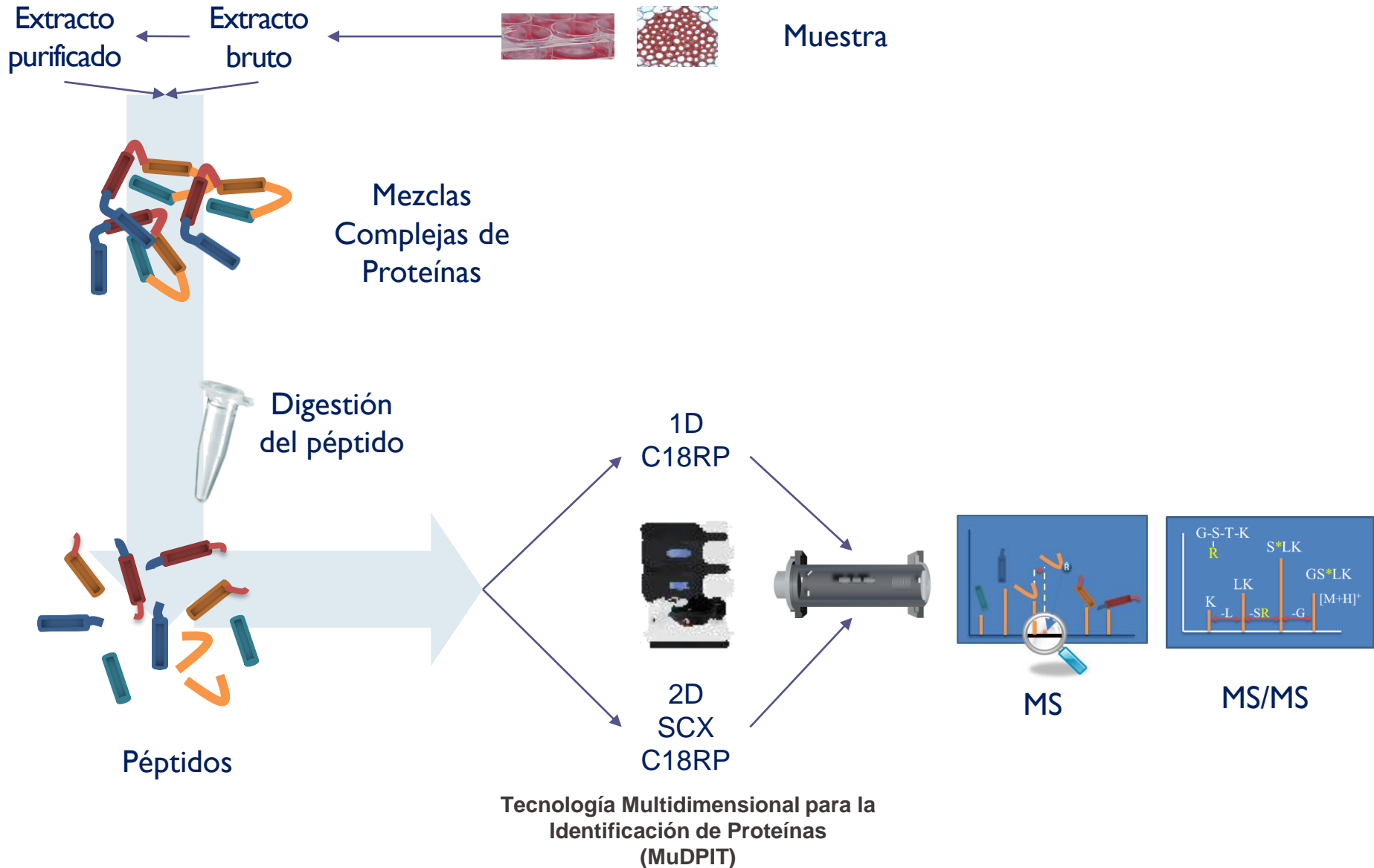




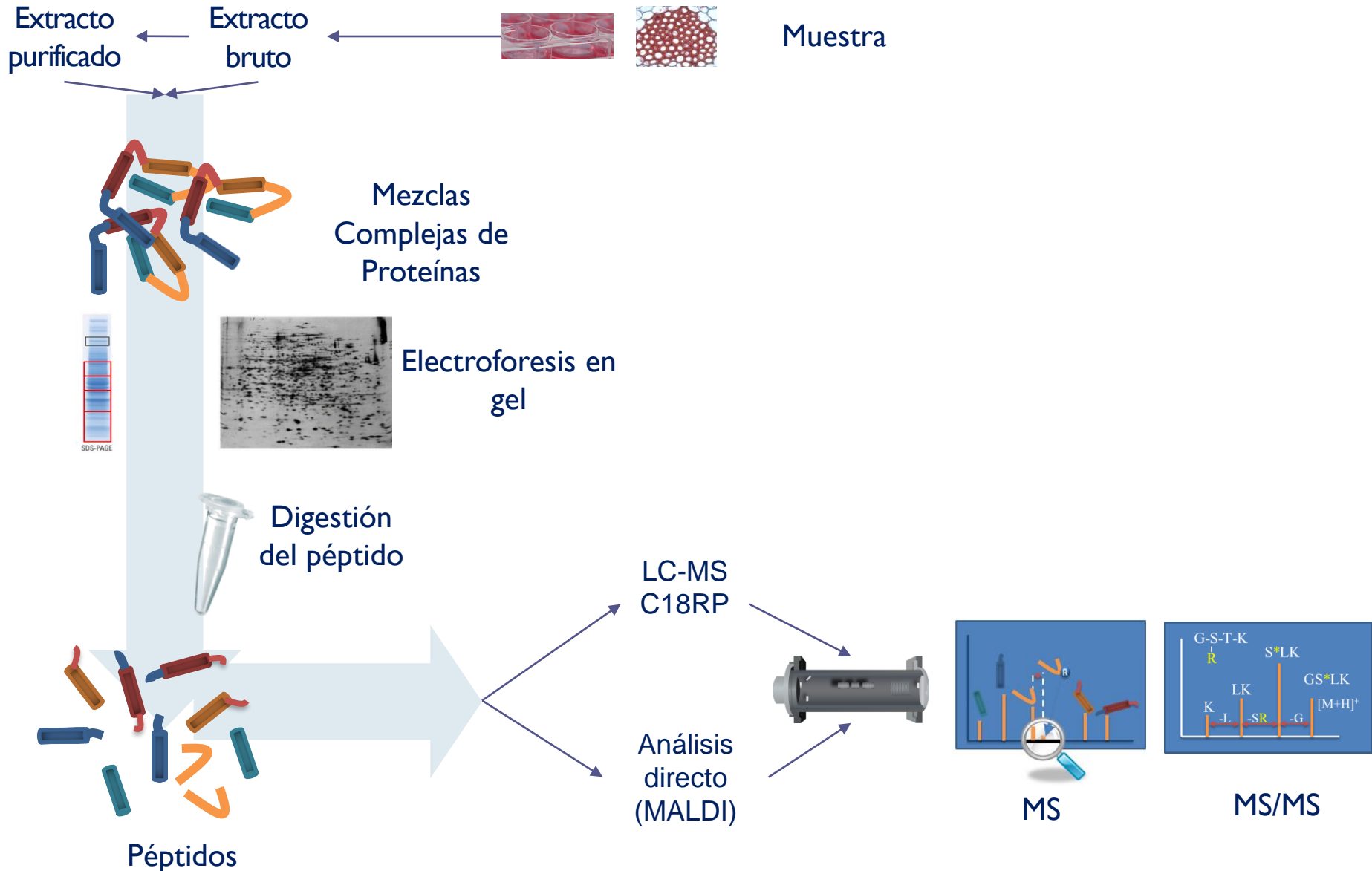
# Enfoque basado en péptidos vs. basado en proteínas

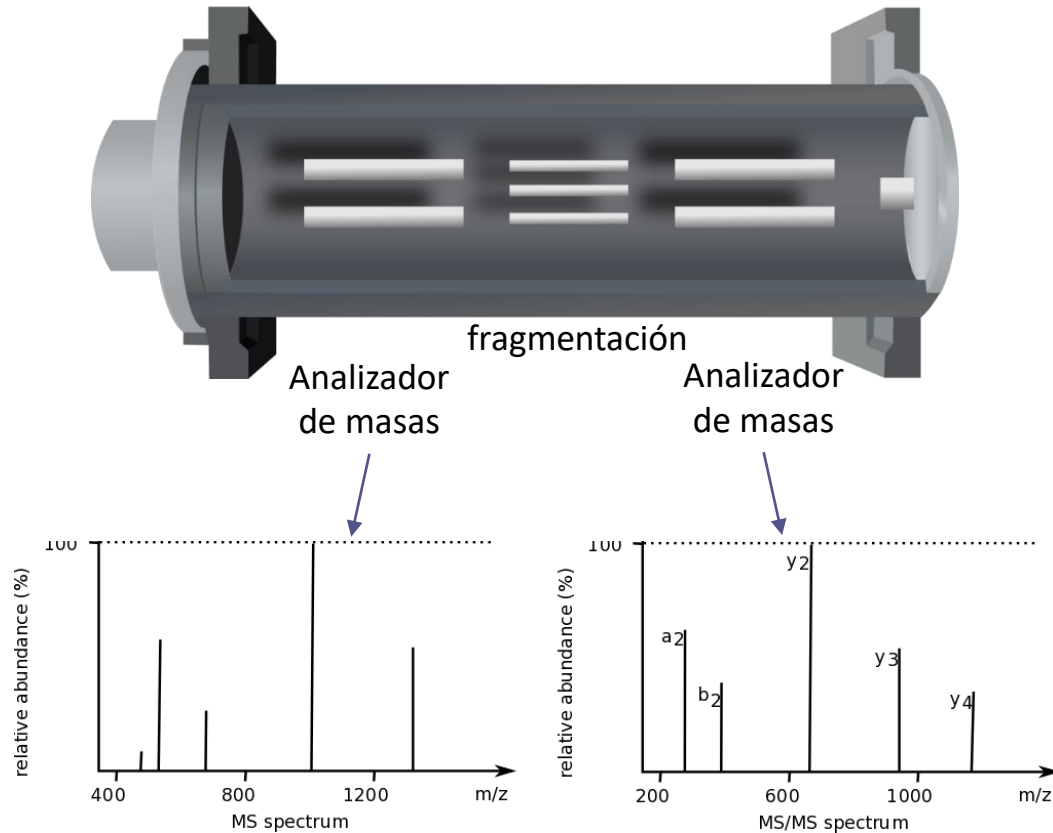


# Enfoque basado en péptidos (por infusión directa)



# Enfoque basado en la proteína (Electroforesis)

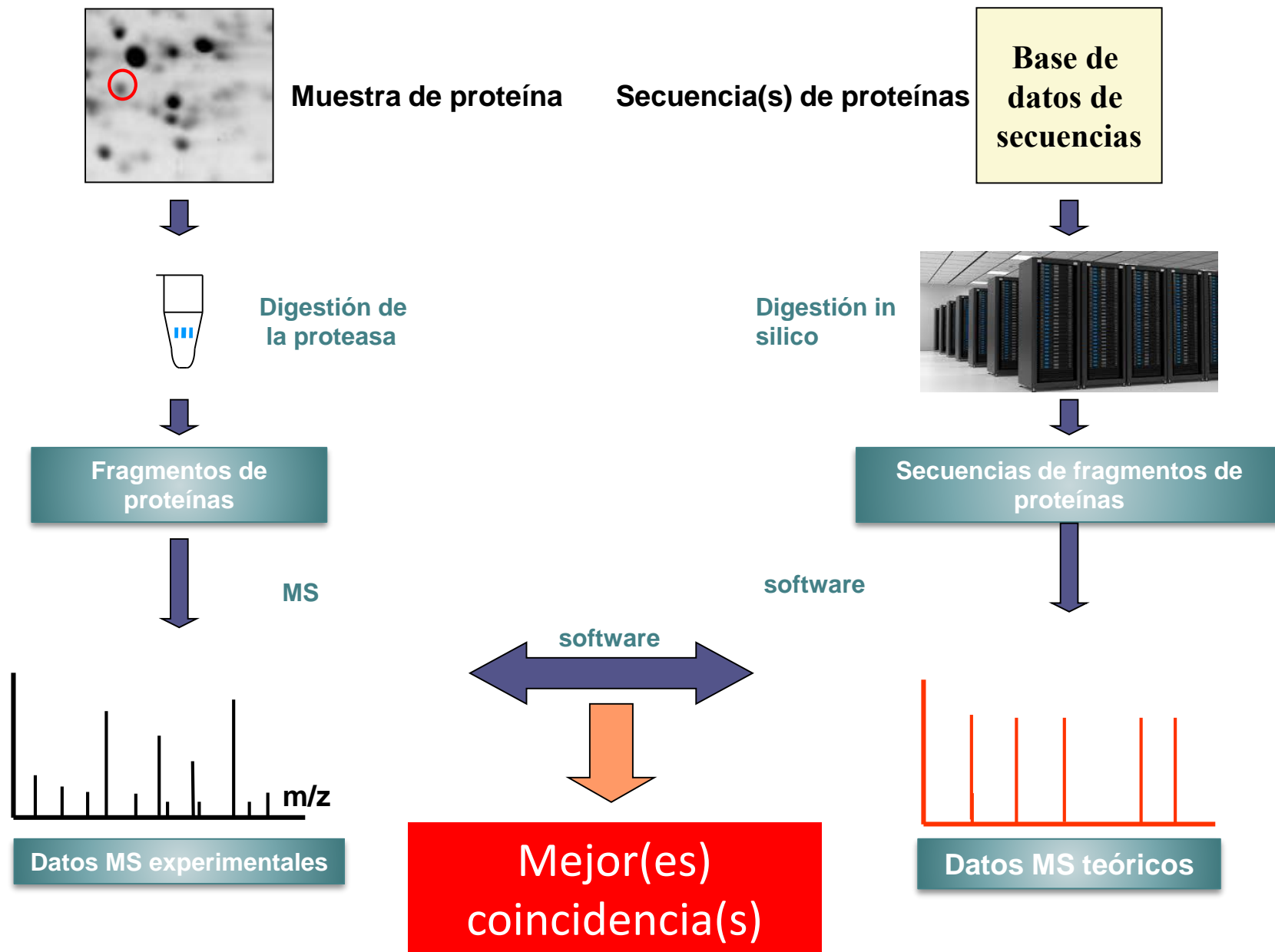


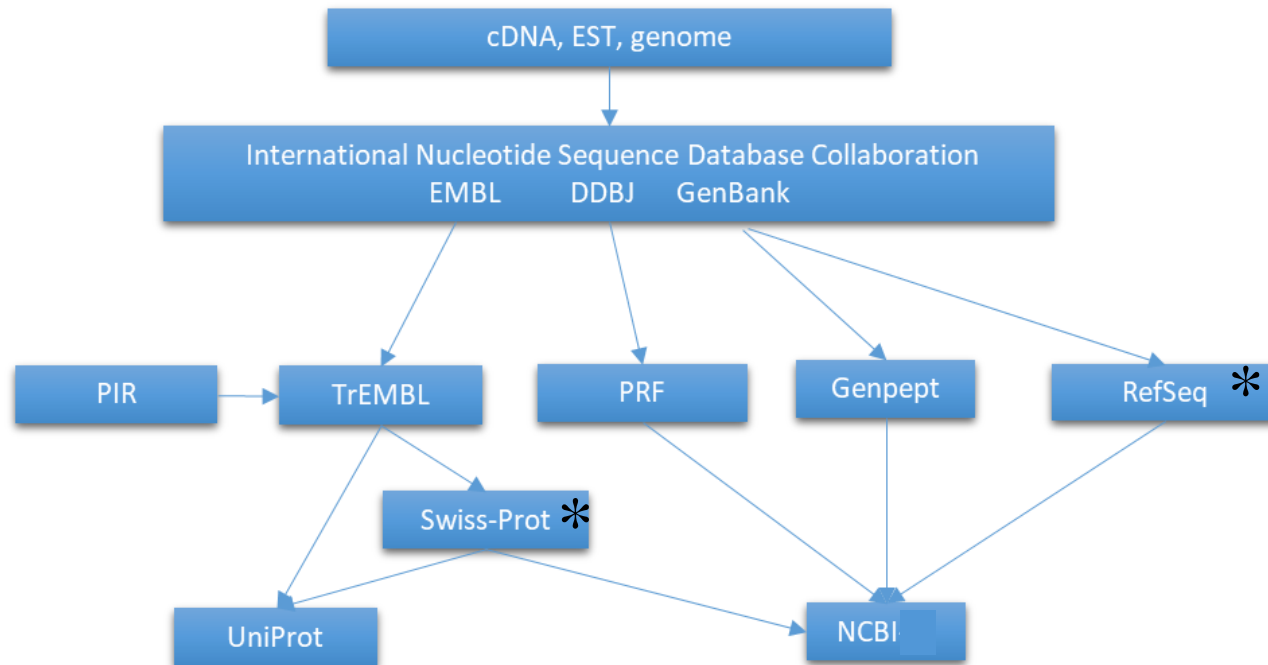


PMF - Huella Peptídica

PFF - Huella de  
Fragmentación  
Peptídica

# Identificación de proteínas a partir de MS: concepto PMF



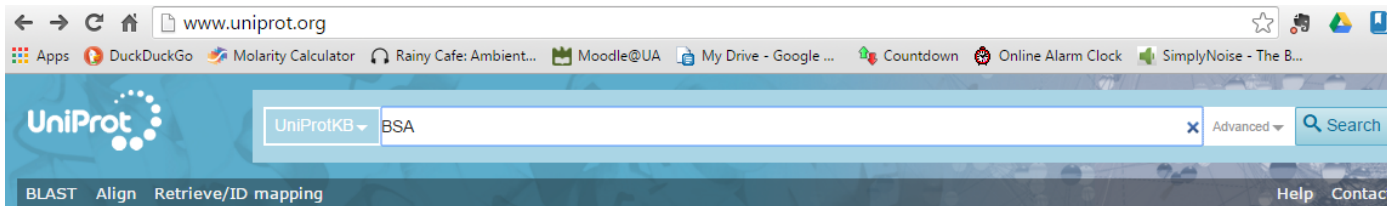


**UniProt:** Swiss-Prot + TrEMBL + (PIR)

**NCBI-nr:** Swiss-Prot + GenPept + (PIR) + RefSeq + PDB + PRF

\* base de datos de proteínas *curada* y no redundante

# Bases de datos de secuencias de proteínas - UniProt



The screenshot shows the UniProt website interface. At the top, there is a search bar with 'UniProtKB' selected and 'BSA' entered. To the right of the search bar are buttons for 'Advanced' and 'Search'. Below the search bar, there are navigation links for 'BLAST', 'Align', and 'Retrieve/ID mapping'. On the far right, there are links for 'Help' and 'Contact'.

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

**UniProtKB**  
UniProt Knowledgebase

**Swiss-Prot**  
(549,646)  
Manually annotated and reviewed.

**TrEMBL**  
(52,783,601)  
Automatically annotated and not reviewed.


**UniRef**  
Sequence clusters



**UniParc**  
Sequence archive



**Proteomes**



**Supporting data**

Literature citations 	Taxonomy 	Subcellular locations 
Cross-ref. databases 	Diseases XXX	Keywords 

**News**

[BLOG](#) [Twitter](#) [Facebook](#) [RSS](#)

**Forthcoming changes**  
Planned changes for UniProt

---

**UniProt release 2015\_10**  
The smell of the sea in UniProtKB | Cross-references to WBParaSite | Removal of the cross-references to CYGD | UniParc cross-reference t...

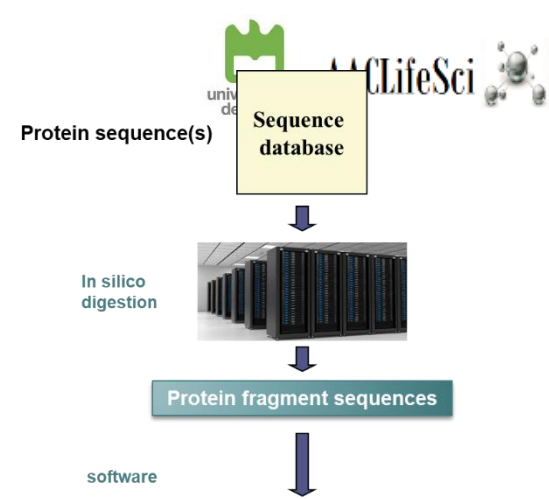
---

**UniProt release 2015\_09**  
Life (and death) in 2D | 27 new species in ...

[News archive](#)



# Digestión proteica (in silico)



Display None Sequence status: Complete. Sequence processing: The displayed sequence is further processed into a mature form.

P02769-1 [UniParc] FASTA Add to basket

Length: 607  
Mass (Da): 69,293  
Last modified: February 1, 1996 - v4  
Checksum: 39167DFE768585D4

PeptideMass GO

```

10      20      30      40      50
MKWVTFISLL LFFSSAYSRG VFRDRTHKSE IAHRFKDLGE EHFKGLVLIA
60      70      80      90     100
FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
110     120     130     140     150
VASLRETYGD MADCCCKQEP ERNECFLSHK DDSFDLPKLL POPNLTCEDE
160     170     180     190     200
KADEKIFWGG YLVEIARRHP YFYAPELLEY ANKYNVGFQE CQCAEDKGGC
210     220     230     240     250
LLPKIETIRE KVLASSARRR LRCASIQKFG ERALKANSVA RLSQKFPFAE
260     270     280     290     300
FVEVTKLVTD LTKVHKECCG GDLLCEADDR ADLAKYIDNI QDTISSKLE
310     320     330     340     350
CCDKPILLEK HCTIAEVEKDA IPENLPPLTA DFAEDKDVCK NVQEAQDAFL
360     370     380     390     400
GSFLYEYSRR HPEYAVSLL RLAKVEYATL EECCKADDPH ACYSTVFDKLL
410     420     430     440     450
KHLVDEPNIL IQNCDQFEK LGEVGFQNAL IVRYTRKVPQ VSTPTLVEVS
460     470     480     490     500
RSLGKVGTRC CTKPESERHP CTEDVLSLIL NRLCVLHEK PVSEKTKCC
510     520     530     540     550
TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT
560     570     580     590     600
ALVELLKHKP KATEEQKTV MENFVAFVCK CCAADDKCAE FAVEGPKLVV
STQTALA
    
```

STB ExpASY Bioinformatics Resource Portal

PeptideMass

**PeptideMass**

The entered protein is: P02769

The selected enzyme is: Trypsin

Maximum number of missed cleavages (MC): 2

All cysteines in reduced form.

Methionines have not been oxidized.

Displaying peptides with a mass bigger than 500 Dalton.

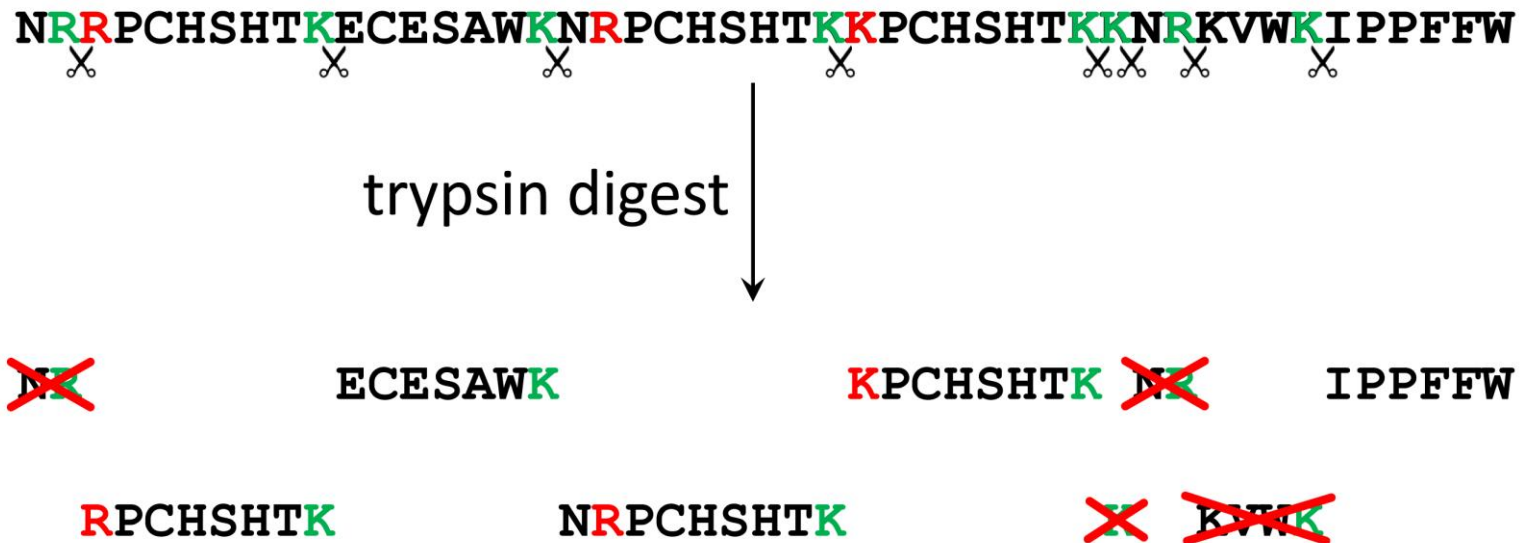
Using monoisotopic masses of the occurring amino acid residues and giving peptide masses as [M+H]<sup>+</sup>.

You have selected ALBU\_BOVIN (P02769) from UniProtKB/Swiss-Prot.

Serum albumin precursor (BSA) (Allergen Bos d 6)  
Signal and propep in positions 1-24 have been removed.

- Chain Serum albumin at positions 25 - 607 [Theoretical pI: 5.60 / Mw (average mass): 66432.96 / Mw (monoisotopic mass): 66389.86]

mass	position	#MC	modifications	peptide sequence
4910.3837	45-88	2	PHOS: 82	4990.3500 GLVLIAFSQYLQQCPFDEHV KLVNELTEFAKTCVADESHA GCEK
4535.2954	37-75	2		DLGEEHFKGLVLIAFSQYLQ QCPFDEHVKLVNELTEFAK
4246.0511	508-544	2	PHOS: 512	4326.0174 RPCFSALTPDETYVPKAFDE KLFTFHADICTLPDTEK
4185.9547	169-204	2		HPYFYAPELLEYANKYNGVF QECCQAEKGAACLLPK
4110.9707	300-336	2		ECCDKPILLEKSHCIAEVEKD AIPENLPPLTADFAEDK
3837.8131	66-100	2	PHOS: 82, 89	3997.7458 LVNELTEFAKTCVADESHAG CEKSLHTLFGDELCK
3758.8958	402-433	2		HLVDEPNILIQNCDQFEKL GEGVFNALIVR
3665.8460	35-65	2		FKDLGEEHFKGLVLIAFSQY LQQCPFDEHVK
3659.6722	168-197	2		RHPYFYAPELLEYANKYNGV FQCCQAEK
3579.8555	45-75	1		GLVLIAFSQYLQQCPFDEHV KLVNELTEFAK
3523.6874	460-489	2		CCTKPESERMPCTEDVLSLI LNRLCVLHEK
3503.5711	169-197	1		HPYFYAPELLEYANKYNGVF QECCQAEK
3444.4810	267-297	2	PHOS: 296	3524.4473 ECCHGDLLECADDRADLAKY ICDNQDTISSK
3419.6068	499-528	2	PHOS: 512	3499.5731 CCTESLVNRRPCFSALTPDE TYVPKAFDEK
3397.6289	310-340	2		SHCIAEVEKDAIPENLPPLT ADFAEDKDVCK
3390.6826	37-65	1		DLGEEHFKGLVLIAFSQYLQ QCPFDEHVK



La tripsina corta las proteínas principalmente en el extremo carboxílico (o «extremo C-terminal») de los aminoácidos:

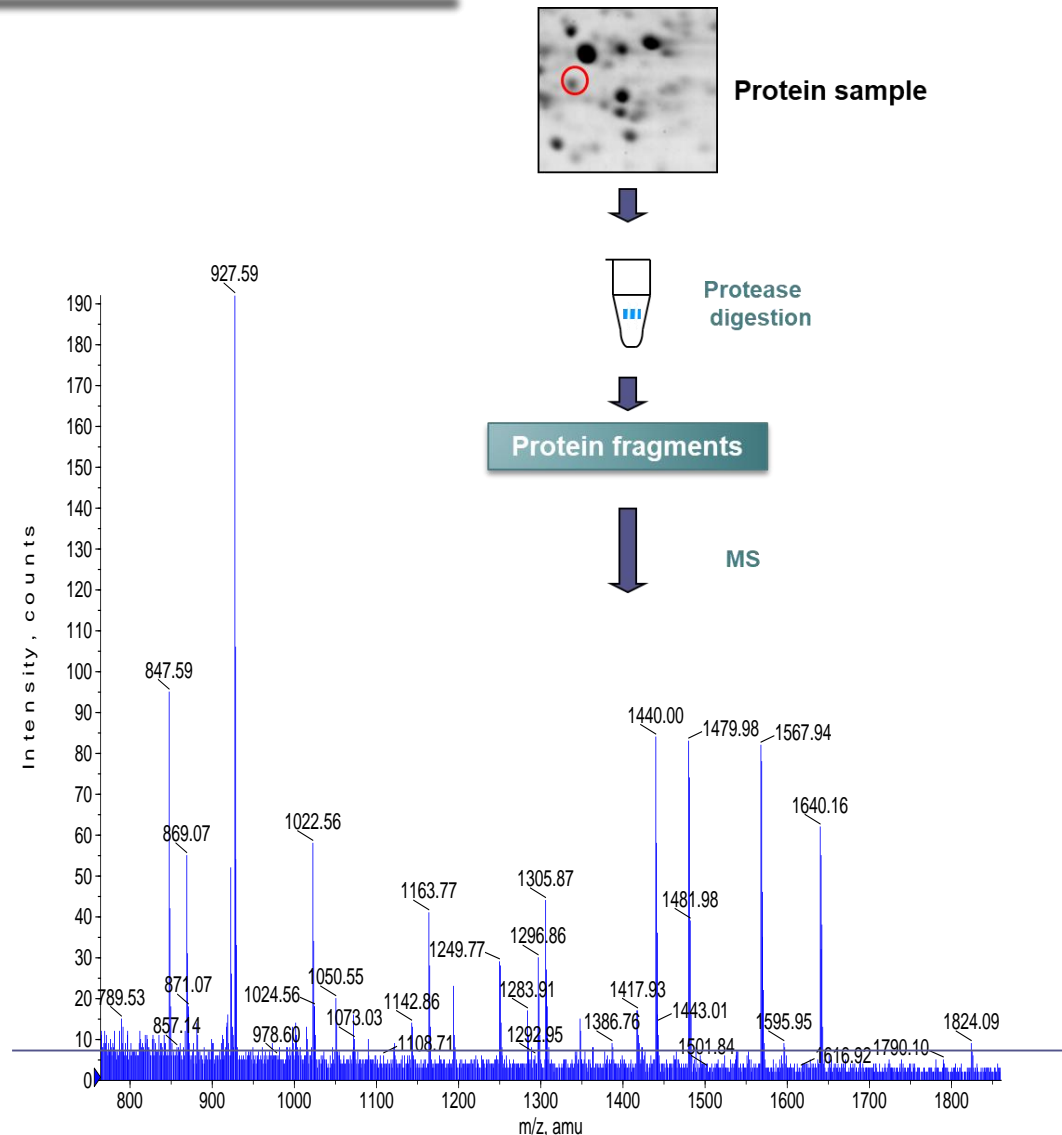
- lisina (K) y
- arginina (R)
- excepto cuando una de ellas está unida a una prolina (P) C-terminal.

# PMF (Huella Peptídica)

## Lista de picos extraída (m/z)

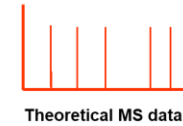
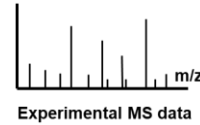
COM=10 pmol digest of Sample BSA  
MASS=Monoisotopic  
USERNAME=Pedro Domingues  
USEREMAIL=p.domingues@ua.pt  
TITLE= Cmpd 7, +MS, 16.8 min  
847,50413  
868,97220  
922,46673  
923,48150  
927,49393  
1022,45510  
1050,45330  
1163,63123  
1164,65310  
1193,60273  
1249,62173  
1250,71030  
1296,75560  
1297,74990  
1305,71663  
1416,79290  
1439,81233  
1479,79593  
1482,75830  
1567,74323  
1639,93833  
1823,86000

umbral



- Mascot
- X!Tandem,
- MS-GF+,
- MS Amanda,
- MyriMatch,
- Comet,
- Tide,
- Andromeda
- OMSSA
- ProteinProspector
- Sequest
- PEAKS
- PRIDE
- etc.

## Motores de búsqueda de proteínas



[Home](#) [Mascot database search](#) [Products](#) [Technical support](#) [Training](#) [News](#) [Blog](#) [Newsletter](#) [Contact](#)

[Access Mascot Server](#) | [Database search help](#)

Mascot database search > Access Mascot Server

### Access Mascot Server

You are welcome to submit searches to this free Mascot Server. Searches of MS/MS data are limited to 1200 spectra and some functions, such as no enzyme searches, are unavailable. Automated searching of batches of files is not permitted. If you want to automate search submission, perform large searches, search additional sequence databases, or customise the modifications, quantitation methods, etc., you'll need to [license your own](#), in-house copy of Mascot Server.

#### Peptide Mass Fingerprint

The experimental data are a list of peptide mass values from the digestion of a protein by a specific enzyme such as trypsin.

[Perform search](#) | [Example of results report](#) | [Tutorial](#)

#### Sequence Query

One or more peptide mass values associated with information such as partial or ambiguous sequence strings, amino acid composition information, MS/MS fragment ion masses, etc. A super-set of a sequence tag query.

[Perform search](#) | [Example of results report](#) | [More information](#)

#### MS/MS Ions Search

Identification based on raw MS/MS data from one or more peptides.

[Perform search](#) | [Example of results report](#) | [Tutorial](#)

#### More info

- > [Mascot overview](#)
- > [Search parameter reference](#)
- > [Data file format](#)
- > [Results report overview](#)



# PMF (Huella Peptídica)



Lista de picos extraída (m/z)

- 847,50413
- 868,97220
- 922,46673
- 923,48150
- 927,49393
- 1022,45510
- 1050,45330
- 1163,63123
- 1164,65310
- 1193,60273
- 1249,62173
- 1250,71030
- 1296,75560
- 1297,74990
- 1305,71663
- 1416,79290
- 1439,81233
- 1479,79593
- 1482,75830
- 1567,74323
- 1639,93833
- 1823,86000

## MASCOT Peptide Mass Fingerprint

**Your name** pedro **Email** p.domingues@ua.pt

**Search title**

**Database(s)** SwissProt  
NCBItr  
contaminants  
cRAP

**Enzyme** Trypsin

**Allow up to** 1 missed cleavages

**Taxonomy** Mammalia (mammals)

**Fixed modifications** -- none selected --

Display all modifications

**Variable modifications** Oxidation (M)

**Protein mass**  kDa

**Peptide tol. ±** 5 ppm

**Mass values**  MH<sup>+</sup>  M<sub>r</sub>  M-H<sup>-</sup>

**Monoisotopic**  Average

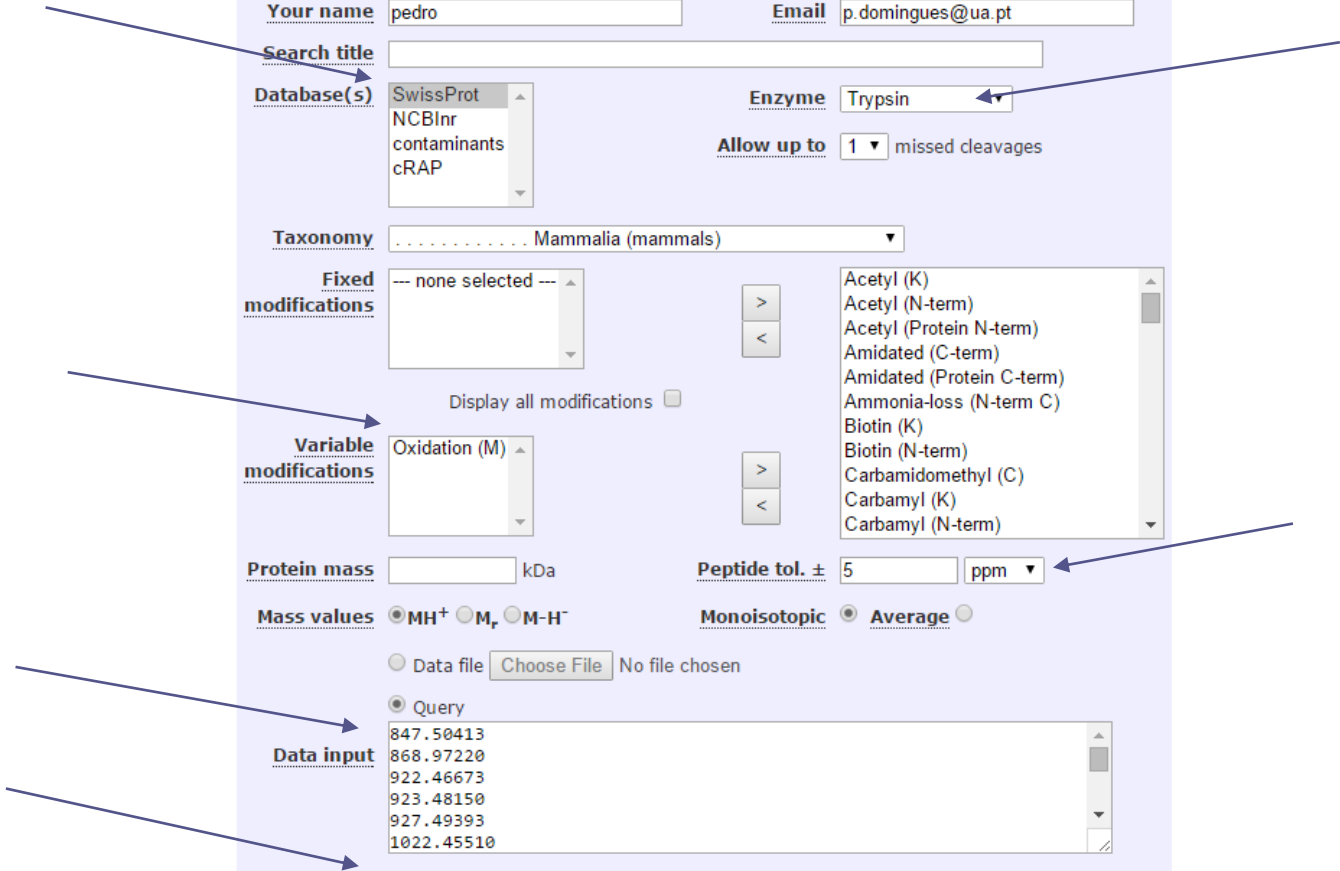
Data file  No file chosen

Query

**Data input** 847.50413  
868.97220  
922.46673  
923.48150  
927.49393  
1022.45510

**Decoy**

**Report top** AUTO hits



# PMF (resultados de Mascot)

## Señuelo

1. Durante la búsqueda, cada vez que se evalúa una secuencia proteica de la base de datos objetivo, se genera automáticamente y se evalúa una secuencia señuelo de la misma longitud.

1. Secuencia inversa
2. Secuencia aleatoria

2. Las coincidencias y puntuaciones de las secuencias señuelo se registran por separado en el archivo de resultados.

3. Cuando se completa la búsqueda, los números de coincidencias y la tasa de descubrimientos falsos se registran en el encabezado de resultados.

## Contaminantes

Secuencias de contaminantes comunes, como queratinas, BSA y tripsina.

Expect es el número de veces que cabe esperar que se obtenga una puntuación igual o mayor, por simple azar (cuanto menor, mejor)

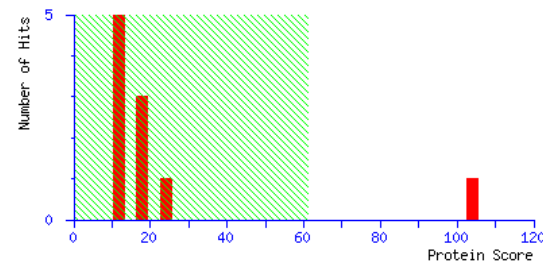
## MATRIX SCIENCE Mascot Search Results

User : pedro  
Email : p.domingues@ua.pt  
Search title :  
Database : SwissProt 2015\_10 (549646 sequences; 195983064 residues)  
Taxonomy : Mammalia (mammals) (66401 sequences)  
Timestamp : 20 Oct 2015 at 11:43:21 GMT  
Top Score : 104 for **ALBU\_BOVIN**, Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

	SwissProt	Decoy
Protein hits above identity threshold	1	0
Highest scoring protein hit	104	20

## Mascot Score Histogram

Protein score is  $-10 \cdot \log(P)$ , where P is the probability that the observed match is a random event. Protein scores greater than 61 are significant ( $p < 0.05$ ).



## Concise Protein Summary Report

Format As: Concise Protein Summary [Help](#)

Significance threshold  $p < 0.05$  Max. number of hits: AUTO

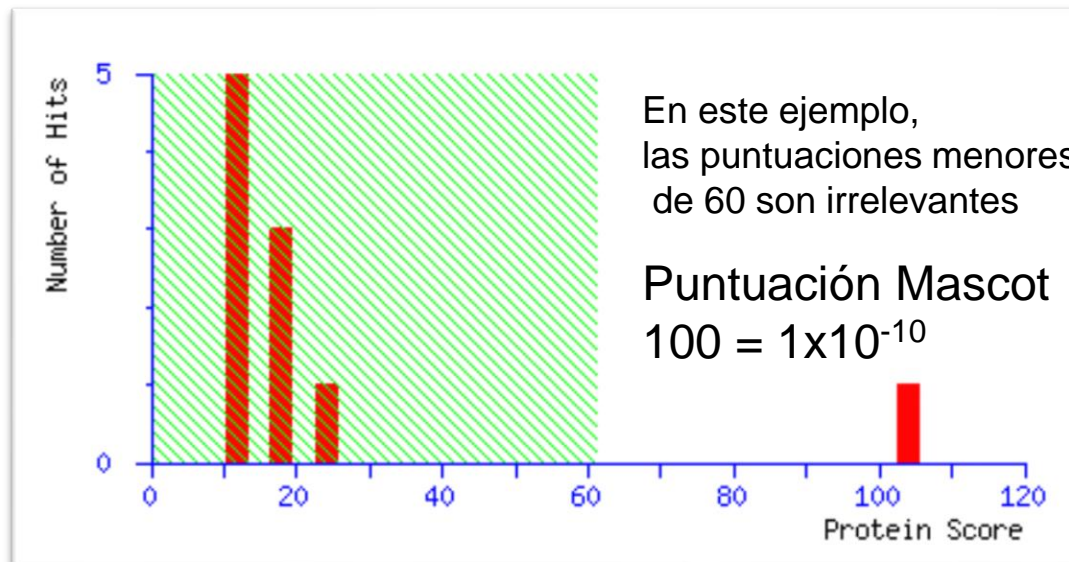
Preferred taxonomy: All entries

Re-Search All Search Unmatched

1.	<a href="#">ALBU_BOVIN</a>	Mass: 69248	Score: 104	Expect: 2.6e-06	Matches: 11
	Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4				
	<a href="#">ALBU_FELCA</a>	Mass: 68615	Score: 24	Expect: 2.6e+02	Matches: 4
	Serum albumin OS=Felis catus GN=ALB PE=1 SV=1				
	<a href="#">ALBU_CAPHI</a>	Mass: 10048	Score: 24	Expect: 2.8e+02	Matches: 2
	Serum albumin (Fragments) OS=Capra hircus GN=ALB PE=1 SV=2				
	<a href="#">ALBU_SHEEP</a>	Mass: 69143	Score: 22	Expect: 4e+02	Matches: 4
	Serum albumin OS=Capra hircus GN=ALB PE=1 SV=1				

# PMF (Puntuación Mascot)

- La Puntuación Mascot se representa mediante  $S = -10 \cdot \log(P)$ , donde
  - **P es la probabilidad de que la coincidencia observada sea un evento aleatorio**
  - $P = E \cdot N^{-1}$ 
    - E=valor esperado
    - N=número de proteínas en la base de datos
- La importancia de ese resultado depende del tamaño de la base de datos que esté siendo examinada. Mascot sombrea en verde los resultados irrelevantes utilizando un umbral de  $E=0,05$





# PMF (resultados de Mascot)

## MATRIX SCIENCE MASCOT Search Results

### Protein View: ALBU\_BOVIN

Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Database: SwissProt  
Score: 104  
Expect: 2.6e-06  
Nominal mass (M<sub>r</sub>): 69248  
Calculated pI: 5.82  
Taxonomy: [Bos taurus](#)

Sequence similarity is available as [an NCBI BLAST search of ALBU\\_BOVIN against nr.](#)

### Search parameters

Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.  
Variable modifications: [Oxidation \(M\)](#)  
Mass values searched: 22  
Mass values matched: 11

### Protein sequence coverage: 18%

Matched peptides shown in **bold red**.

```
1 MKWVTFISLL LFFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA
51 FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
101 VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLDCDF
151 KADEKKFWGK YLYEIARRHP YFYAPELLYY ANKYNQVQFE CCQAEDKGAC
201 LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE
251 FVEVTKLVTD LTKVHKCCCH GDLLCACDDR ADLAKYICDN QDTISSKLKE
301 CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVKC NYQEAKDAFL
351 GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL
401 KHLVDEPQNL IKQNCQDFEK LGEYGFQNAL IVRYIRKVPQ VSTPTLVEVS
451 RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLOVLHEKI PVSEKVIKCC
501 TESLVNRRPC FSALTFDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT
551 ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV
601 STQTALA
```

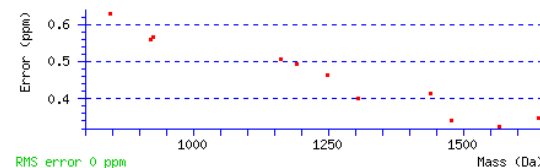
Unformatted sequence string: [607 residues](#) (for pasting into other applications).

Sort peptides by  Residue Number  Increasing Mass  Decreasing Mass

Show predicted peptides also

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	M	Peptide
25 - 34	1193.6027	1192.5955	1192.5949	0.49	1	R.DTHKSEIAHR.F
35 - 44	1249.6217	1248.6145	1248.6139	0.46	1	R.FKDLGEEHFK.G
66 - 75	1163.6312	1162.6240	1162.6234	0.50	0	K.LVNELTEFAK.T
161 - 167	927.4939	926.4867	926.4861	0.57	0	K.YLYEIAR.R
205 - 211	922.4667	921.4595	921.4589	0.56	1	K.IETMREK.V + Oxidation (M)
242 - 248	847.5041	846.4969	846.4963	0.63	1	R.LSQKFPK.A
347 - 359	1567.7432	1566.7360	1566.7354	0.33	0	K.DAFLGSLYEYSR.R
360 - 371	1439.8123	1438.8051	1438.8045	0.41	1	R.RHPEYAVSVLLR.L
402 - 412	1305.7166	1304.7094	1304.7088	0.40	0	K.HLVDEPQNLIK.Q
421 - 433	1479.7959	1478.7887	1478.7881	0.34	0	K.LGEYGFQNALIVR.Y
437 - 451	1639.9383	1638.9311	1638.9305	0.35	1	R.KVPQVSTPTLVEVSR.S

No match to: 868.9722, 923.4815, 1022.4551, 1050.4533, 1164.6531, 1250.7103, 1296.7556, 1297



AC P02769; A5PJX3; O02787; P04277; Q3SZR2;  
DT 21-JUL-1986. Integrated into UniProtKB/Swiss-Prot.

# PMF (Huella Peptídica)

## Lista de picos extraída (m/z)

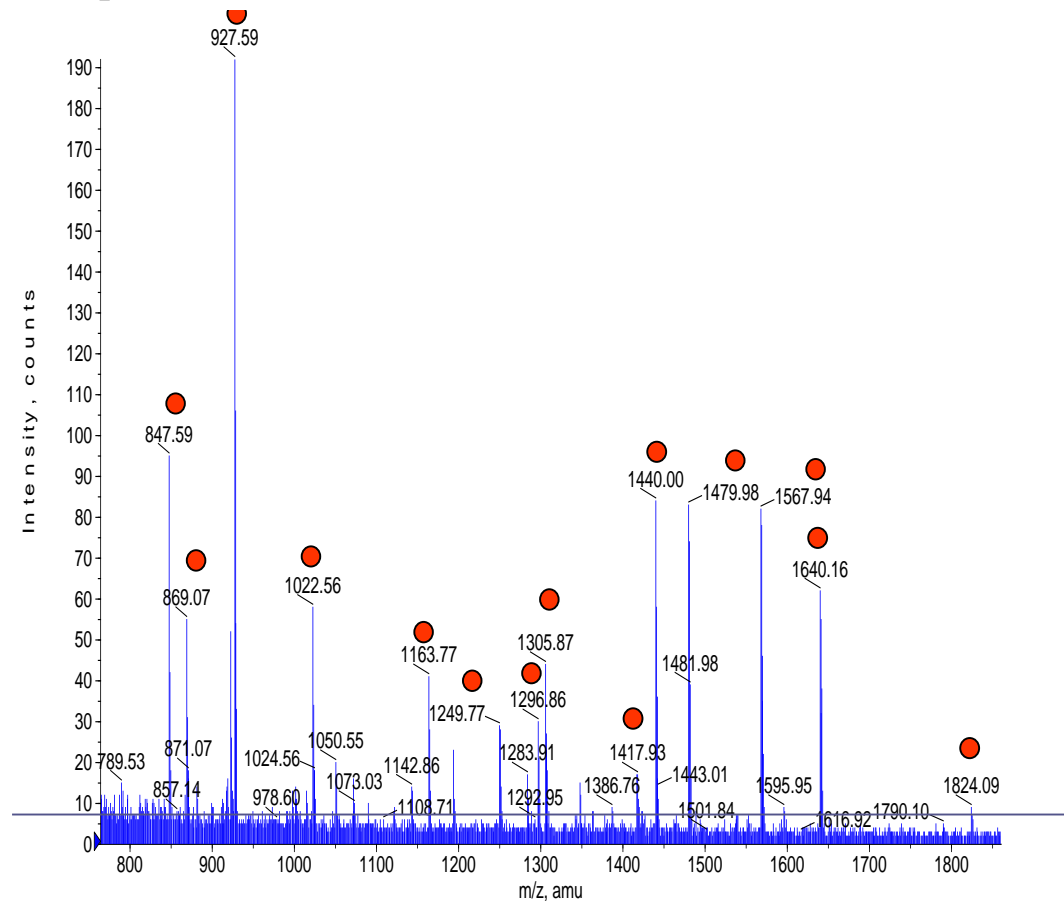
COM=10 pmol digest of Sample BSA  
MASS=Monoisotopic  
USERNAME=Pedro Domingues  
USEREMAIL=p.domingues@ua.pt  
TITLE= Cmpd 7, +MS, 16.8 min  
847,50413  
868,97220  
922,46673  
923,48150  
927,49393  
1022,45510  
1050,45330  
1163,63123  
1164,65310  
1193,60273  
1249,62173  
1250,71030  
1296,75560  
1297,74990  
1305,71663  
1416,79290  
1439,81233  
1479,79593  
1482,75830  
1567,74323  
1639,93833  
1823,86000

Valores de masa buscados: 22

Valores de masa compatibles: 11

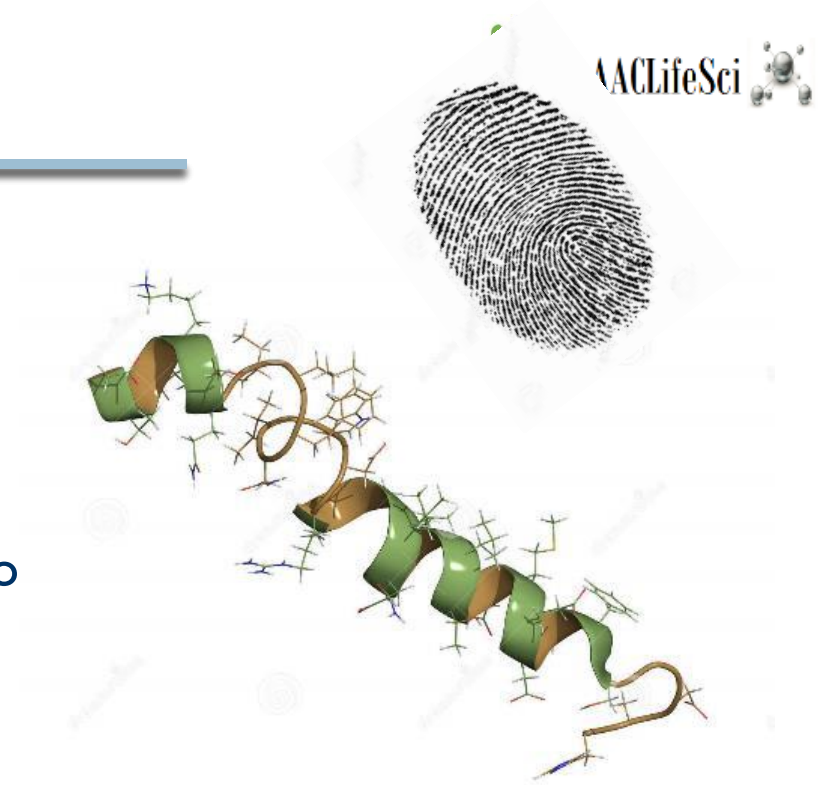
1305.0 counts.

umbral

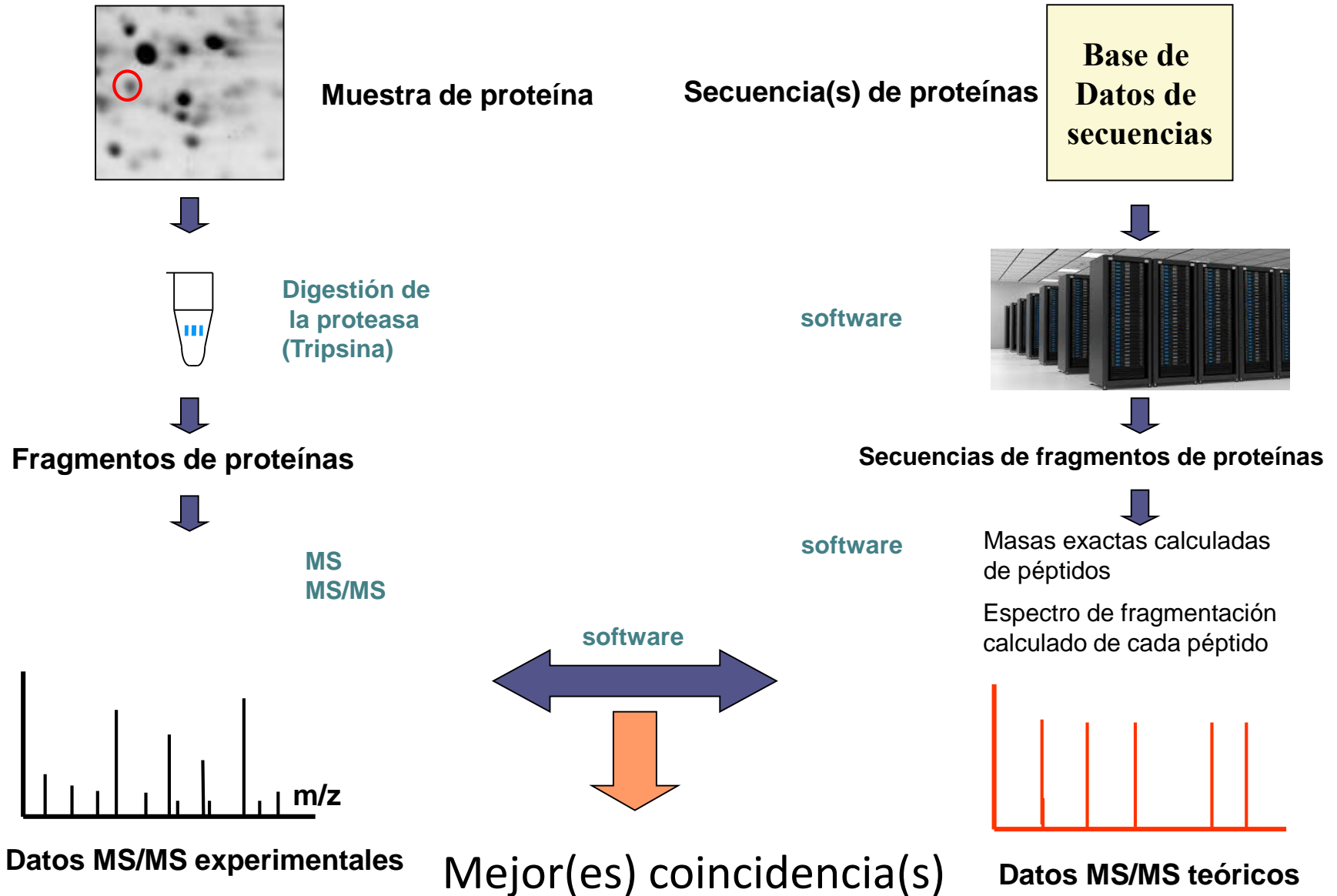


# PMF (Huella Peptídica)

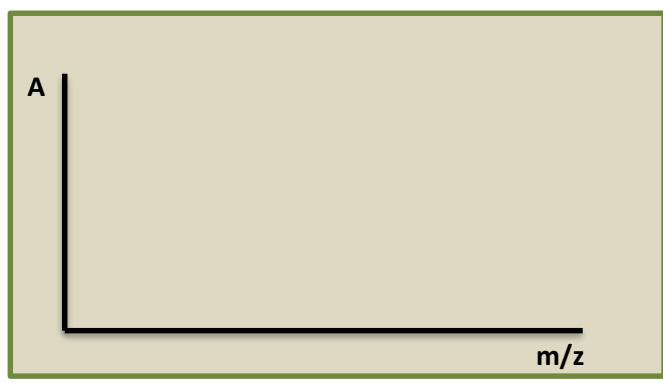
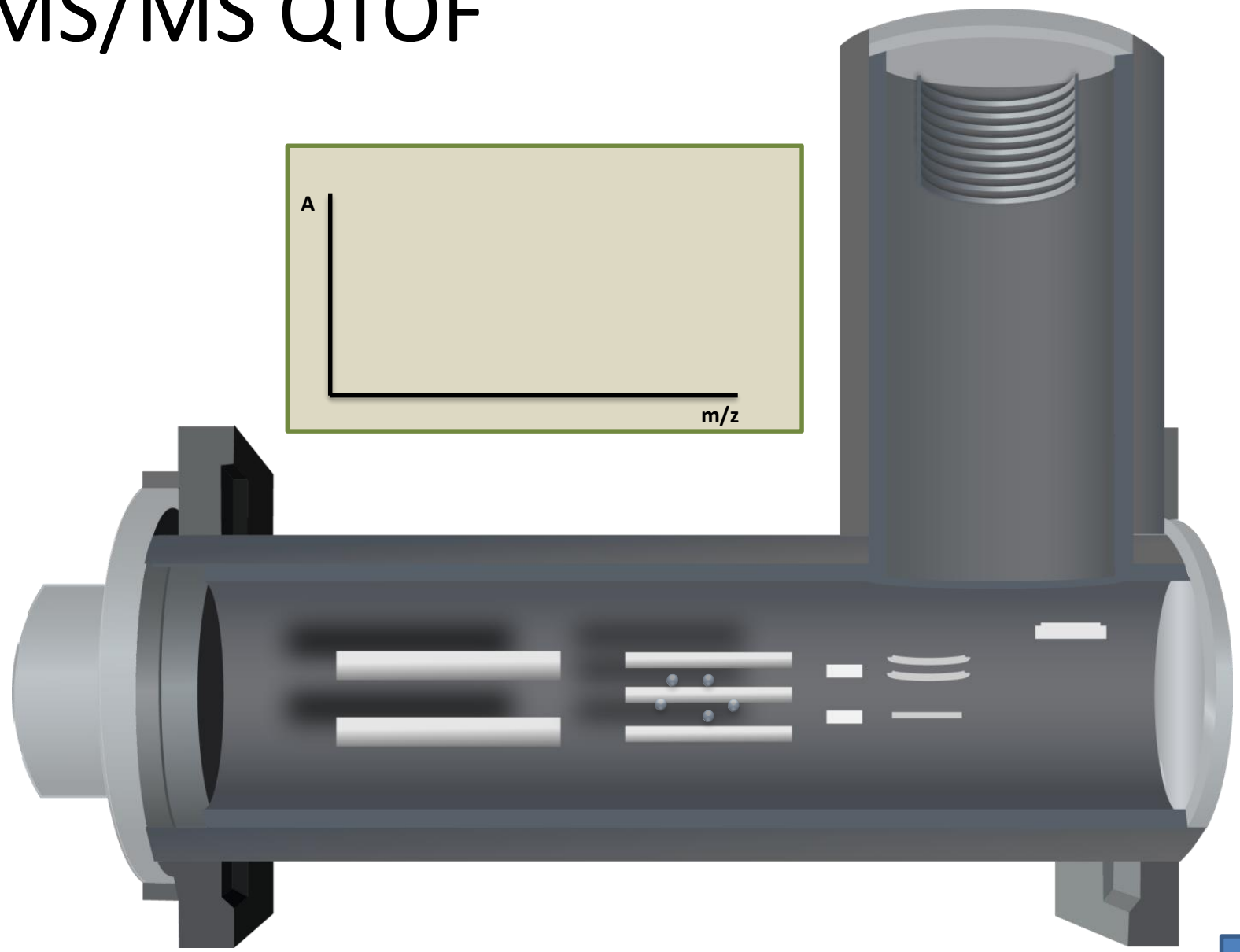
- Análisis rápido y sencillo
- Alta sensibilidad
- No adecuado para mezclas
  - especialmente un componente secundario
- Se necesita una alta exactitud de masa



# Identificación de proteínas a partir de MS: concepto PFF

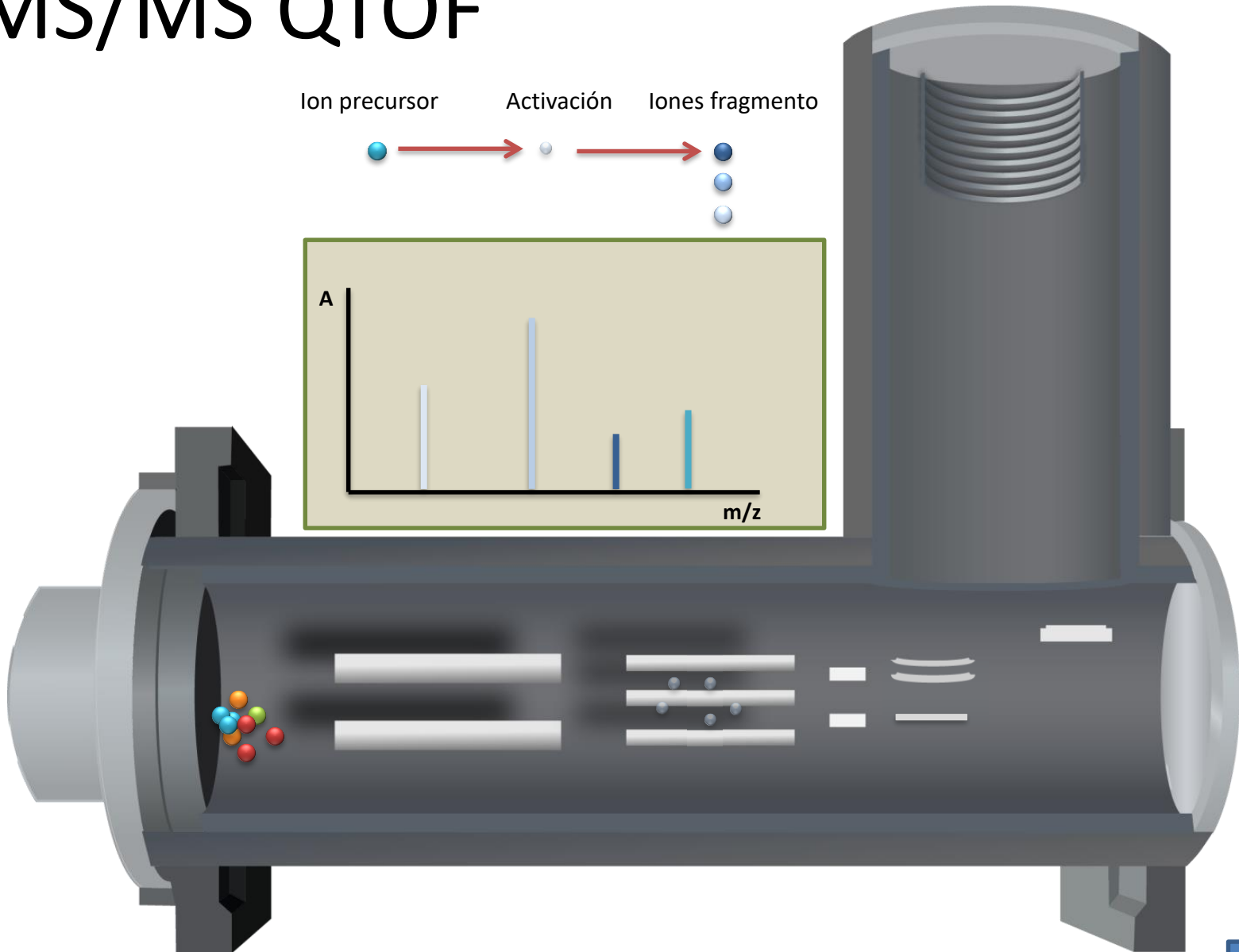
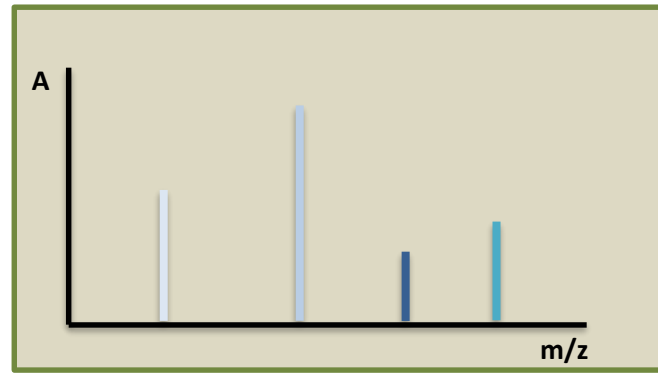


# MS/MS QTOF

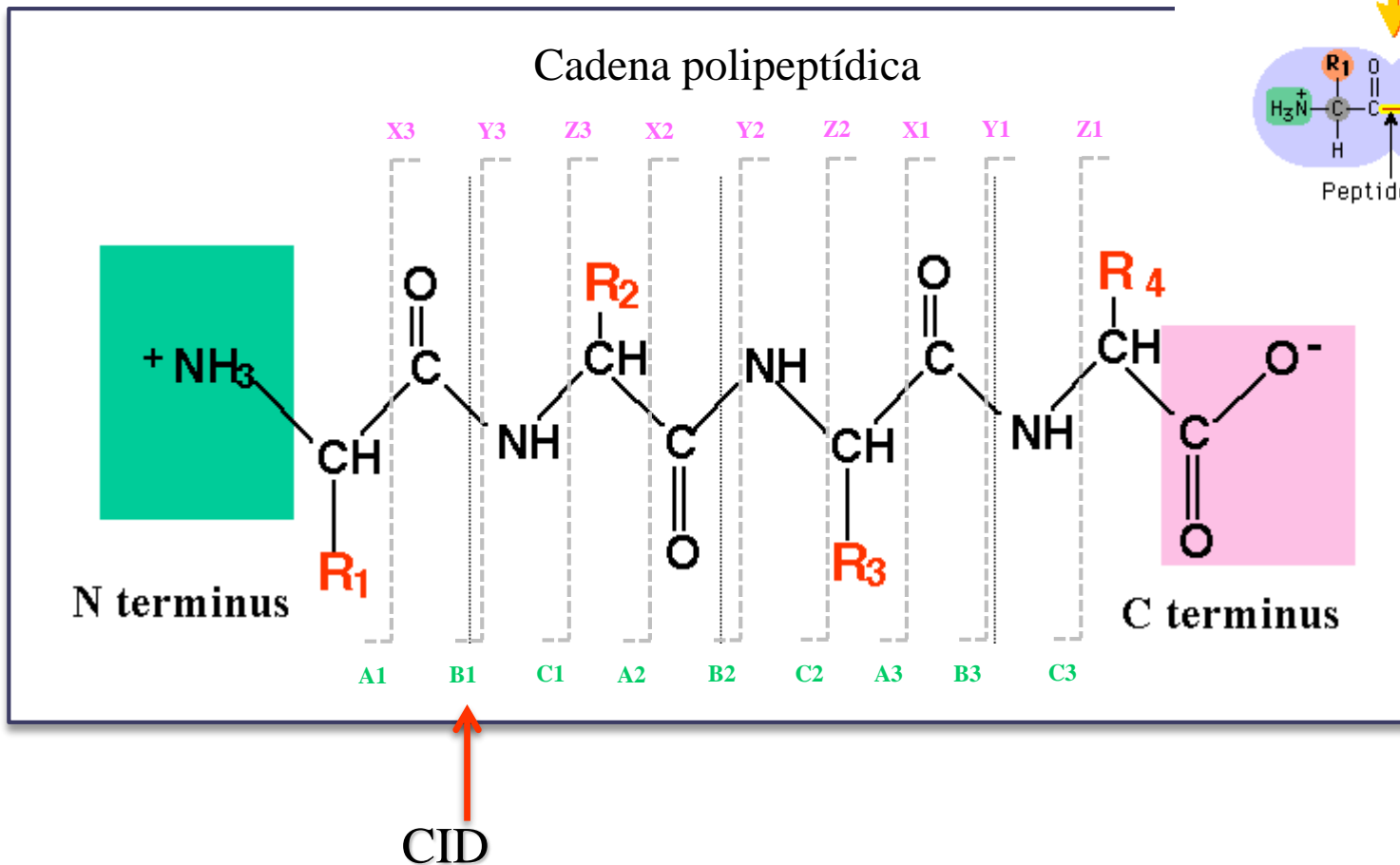


# MS/MS QTOF

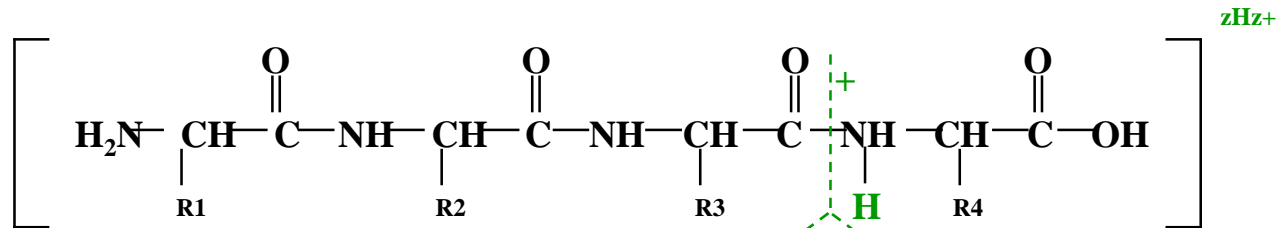
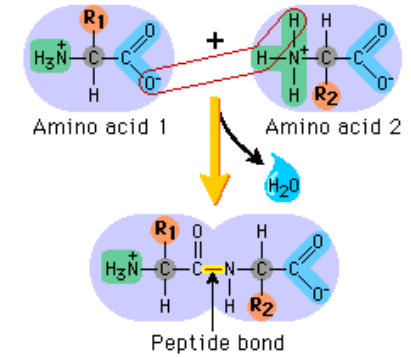
Ion precursor      Activación      Iones fragmento



# Fragmentación de péptidos



# Fragmentación dirigida por la carga

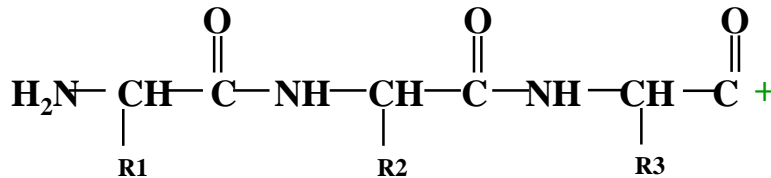


**b** formación de ion

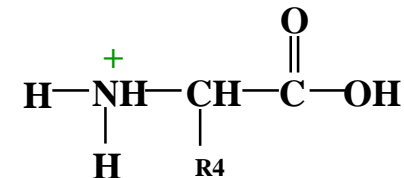
y/o

**y** formación de ion

**b3**



+  
Neutro bombeado mediante sistema de vacío

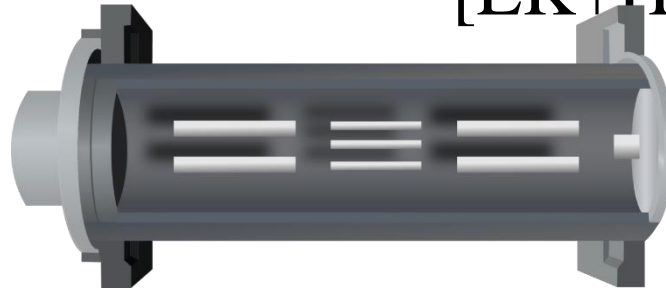
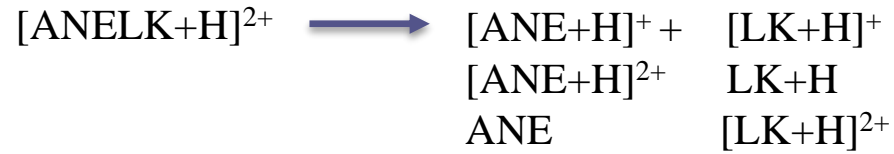


+  
Neutro bombeado mediante sistema de vacío

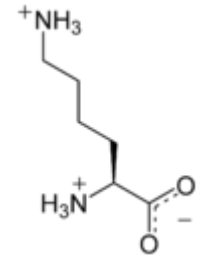
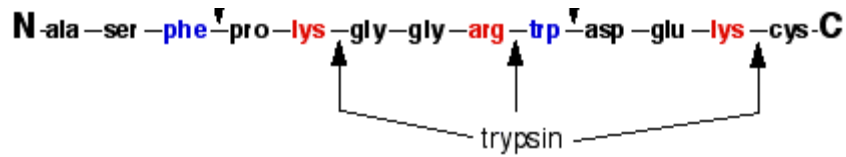
**y1**



# Pares de iones complementarios b/y (múltiples iones cargados)

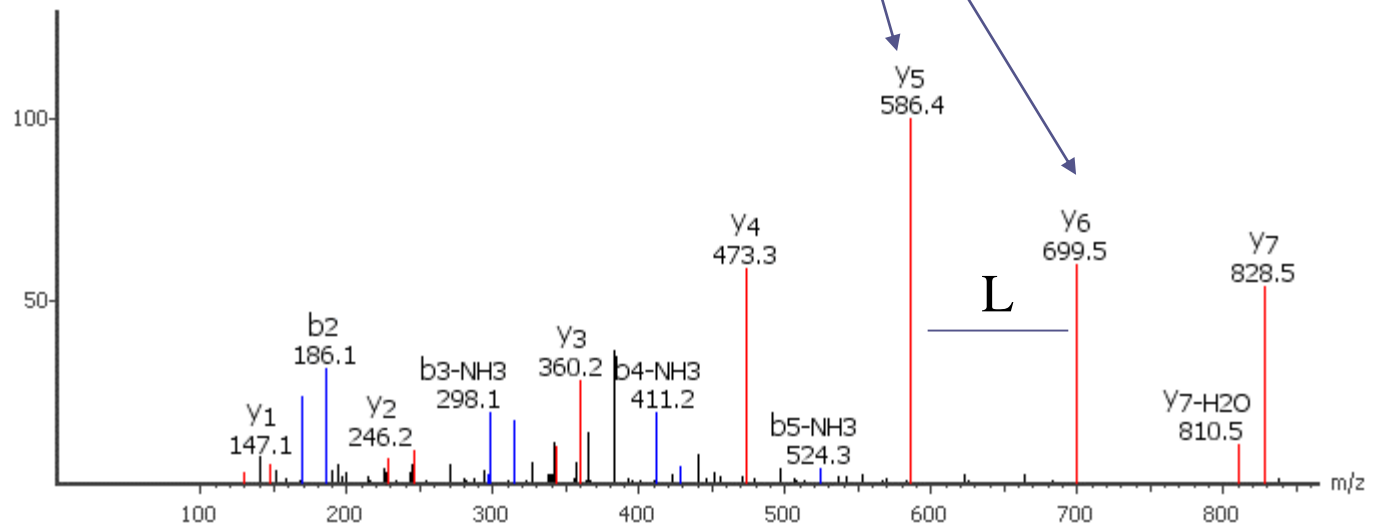


# Pares de iones complementarios b/y



Lis (K)

b <sub>1</sub>	A NELLNVK	Y <sub>8</sub>
b <sub>2</sub>	AN ELLNVK	Y <sub>7</sub>
b <sub>3</sub>	ANE LLLNVK	Y <sub>6</sub>
b <sub>4</sub>	ANEL LLNVK	Y <sub>5</sub>
b <sub>5</sub>	ANELL LNVK	Y <sub>4</sub>
b <sub>6</sub>	ANELL NVK	Y <sub>3</sub>
b <sub>7</sub>	ANELLN VK	Y <sub>2</sub>
b <sub>8</sub>	ANELLNV K	Y <sub>1</sub>



# Modelos de fragmentación peptídica in silico

## Fragment Ion Calculator Results

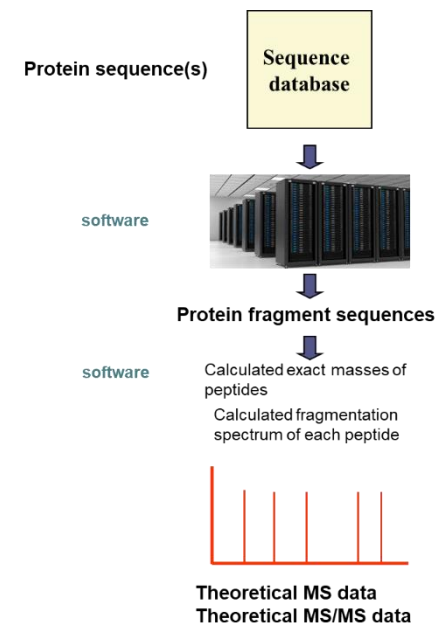
Sequence: AEFVEVTK, pI: 4.53158

### Fragment Ion Table, monoisotopic masses

Seq	#	B	Y	# (+1)
A	1	72.04444	922.48807	8
E	2	201.08703	851.45095	7
F	3	348.15544	722.40836	6
V	4	447.22386	575.33995	5
E	5	576.26645	476.27153	4
V	6	675.33486	347.22894	3
T	7	776.38254	248.16053	2
K	8	904.47750	147.11285	1

### Mass/Charge Table

	Mass	
	Mono	Avg
(M)	921.48079	922.04593
(M+H) <sup>+</sup>	922.48807	923.05320
(M+2H) <sup>2+</sup>	461.74769	462.03026
(M+3H) <sup>3+</sup>	308.16757	308.35595
(M+4H) <sup>4+</sup>	231.37751	231.51879



<http://db.systemsbioology.net:8080/proteomicsToolkit/FragIonServlet.html>

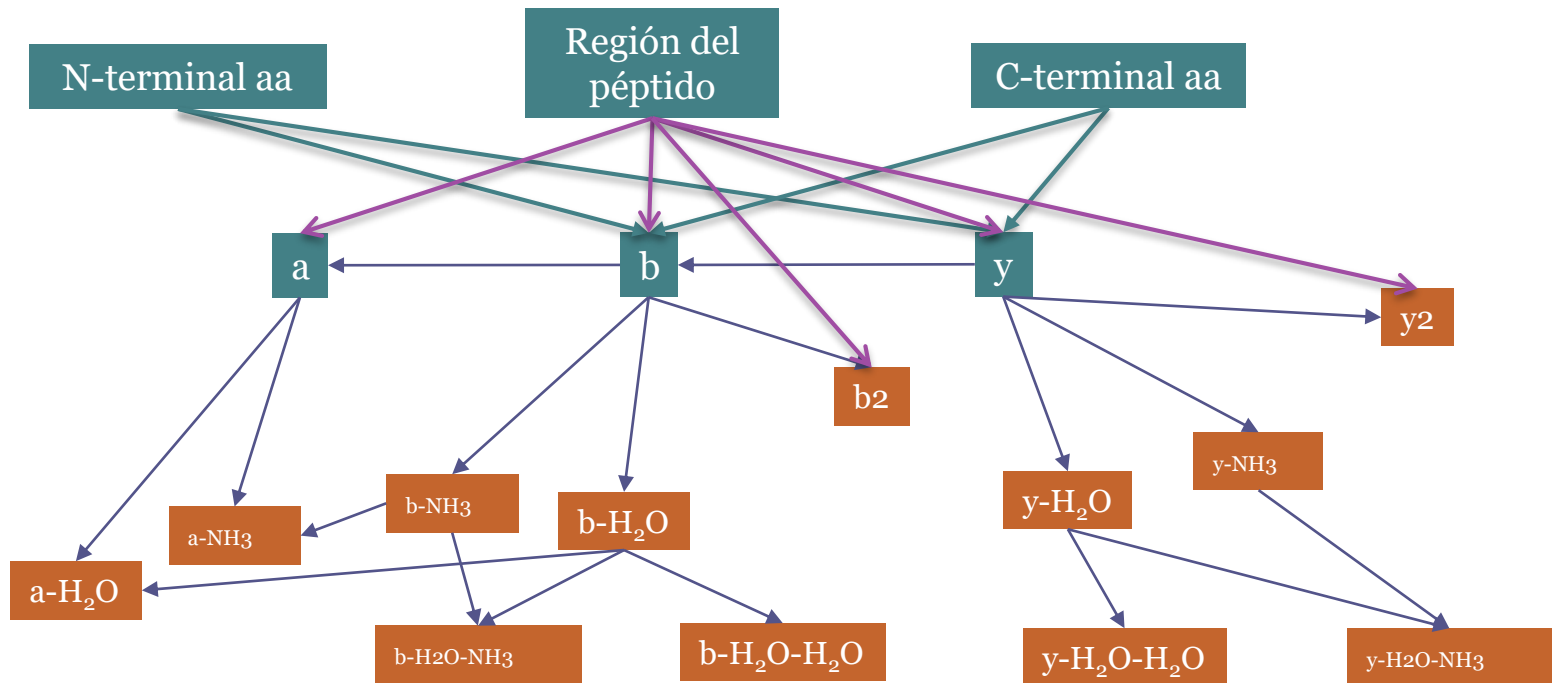
**AEFVEVTK**

# Modelación precisa de la fragmentación peptídica

Método de ionización

Método de fragmentación

Energía de fragmentación



- (1) correlaciones entre iones fragmento;
- (2) dependencias debidas a la posición relativa del extremo de corte del péptido;
- (3) influencia de aminoácidos complementarios en el extremo de corte.

- Mascot
- X!Tandem,
- MS-GF+,
- MS Amanda,
- MyriMatch,
- Comet,
- Tide,
- Andromeda
- OMSSA
- ProteinProspector
- Sequest
- PEAKS
- PRIDE
- ....



[Home](#) [Mascot database search](#) [Products](#) [Technical support](#) [Training](#) [News](#) [Blog](#) [Newsletter](#) [Contact](#)

[Access Mascot Server](#) | [Database search help](#)

[Mascot database search](#) > [Access Mascot Server](#)

## Access Mascot Server

You are welcome to submit searches to this free Mascot Server. Searches of MS/MS data are limited to 1200 spectra and some functions, such as no enzyme searches, are unavailable. Automated searching of batches of files is not permitted. If you want to automate search submission, perform large searches, search additional sequence databases, or customise the modifications, quantitation methods, etc., you'll need to license your own, in-house copy of Mascot Server.

### Peptide Mass Fingerprint

The experimental data are a list of peptide mass values from the digestion of a protein by a specific enzyme such as trypsin.

[Perform search](#) | [Example of results report](#) | [Tutorial](#)

### Sequence Query

One or more peptide mass values associated with information such as partial or ambiguous sequence strings, amino acid composition information, MS/MS fragment ion masses, etc. A super-set of a sequence tag query.

[Perform search](#) | [Example of results report](#) | [More information](#)

### MS/MS Ions Search

Identification based on raw MS/MS data from one or more peptides.

[Perform search](#) | [Example of results report](#) | [Tutorial](#)

### More info

- > [Mascot overview](#)
- > [Search parameter reference](#)
- > [Data file format](#)
- > [Results report overview](#)



# PFF (Huella de Fragmentación Peptídica)



Protease  
digestion  
(Trypsin)



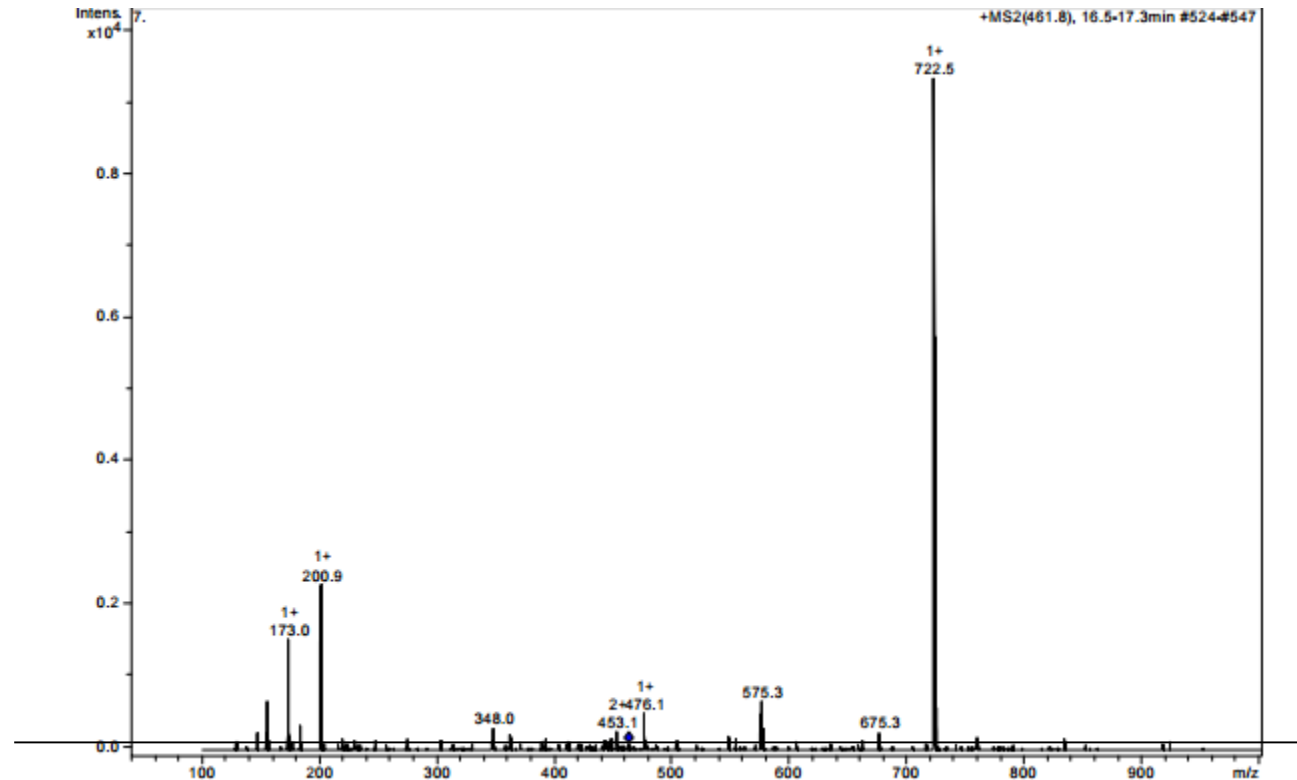
Protein fragments

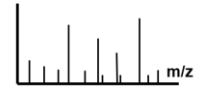


## MS/MS 461.7

COM=10 pmol digest of Sample BSA  
MASS=Monoisotopic  
USERNAME=Pedro Domingues  
USEREMAIL=p.domingues@ua.pt  
BEGIN IONS  
TITLE= Cmpd 7, +MSn(461.8), 16.8 min  
PEPMASS=461.7491 18565  
CHARGE=2+  
147.05 229  
154.98 787 1+  
172.95 1734 1+  
183.87 410 1+  
200.91 2479 1+  
218.92 144  
248.16 150  
347.22 250  
347.96 200  
476.17 670 1+  
548.13 180  
575.33 672  
576.37 393 1+  
675.33 231  
722.45 13125 1+  
723.14 357 1+  
759.39 159  
833.47 141  
851.45 149  
904.13 432 1+  
END IONS

Umbral





Experimental MS/MS data



Best Match(es)



Theoretical MS/MS data



Search this site 

Home Mascot database search Products Technical support Training News Blog Newsletter Contact

Access Mascot Server | Database search help

Mascot database search > Access Mascot Server > MS/MS Ions Search

## MASCOT MS/MS Ions Search

Your name	<input type="text" value="pedro"/>	Email	<input type="text" value="p.domingues@ua.pt"/>
Search title	<input type="text"/>		
Database(s)	<input type="text" value="SwissProt (AA)"/>	>	<
			<ul style="list-style-type: none"><li>Amino acid (AA)</li><li>contaminants</li><li>cRAP</li><li>NCBIprot</li><li>Nucleic acid (NA)</li><li>Environmental_EST</li><li>Fungi_EST</li><li>Human_EST</li><li>Invertebrates_EST</li><li>Mammals_EST</li></ul>
Taxonomy	<input type="text" value="..... Mammalia (mammals)"/>		
Enzyme	<input type="text" value="Trypsin"/>	Allow up to	<input type="text" value="1"/> missed cleavages
Quantitation	<input type="text" value="None"/>		
Fixed modifications	<input type="text" value="--- none selected ---"/>	>	<
			<ul style="list-style-type: none"><li>Acetyl (K)</li><li>Acetyl (N-term)</li><li>Acetyl (Protein N-term)</li><li>Amidated (C-term)</li><li>Amidated (Protein C-term)</li><li>Ammonia-loss (N-term C)</li><li>Biotin (K)</li><li>Biotin (N-term)</li><li>Carbamidomethyl (C)</li><li>Carbamyl (K)</li><li>Carbamyl (N-term)</li></ul>
	Display all modifications <input type="checkbox"/>		
Variable modifications	<input type="text" value="--- none selected ---"/>	>	<
Peptide tol. ±	<input type="text" value="5"/> ppm	# <sup>13</sup> C	<input type="text" value="0"/>
MS/MS tol. ±	<input type="text" value="5"/> ppm		
Peptide charge	<input type="text" value="2+"/>	Monoisotopic	<input checked="" type="radio"/> Average <input type="radio"/>
Data file	<input type="button" value="Choose File"/> Mascot MSMS2.txt	Precursor	<input type="text"/> m/z
Data format	<input type="text" value="Mascot generic"/>	Error tolerant	<input type="checkbox"/>
Instrument	<input type="text" value="Default"/>	Report top	<input type="text" value="AUTO"/> hits
Decoy	<input checked="" type="checkbox"/>		
	<input type="button" value="Start Search ..."/>		<input type="button" value="Reset Form"/>



# PFF (resultados de Mascot)

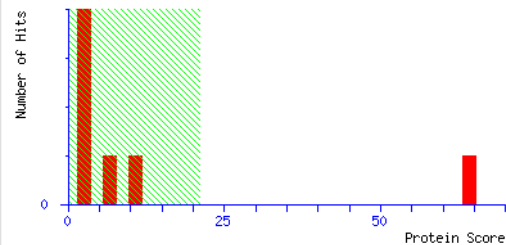
## MASCOT SCIENCE Mascot Search Results

User : Pedro Domingues  
Email : p.domingues@ua.pt  
Search title : 10 pmol digest of Sample BSA  
MS data file : new 1.txt  
Database : SwissProt 2015\_10 (549646 sequences; 195983064 residues)  
Taxonomy : Mammalia (mammals) (66401 sequences)  
Timestamp : 21 Oct 2015 at 13:31:53 GMT  
Protein hits : [ALBU\\_BOVIN](#) Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

	SwissProt	<a href="#">Decoy</a>	False discovery rate
Peptide matches above identity threshold	1	0	0.00 %
Peptide matches above homology or identity threshold	1	0	0.00 %

### Mascot Score Histogram

Ions score is  $-10 \cdot \log(P)$ , where P is the probability that the observed match is a random event. Individual ions scores  $> 21$  indicate identity or extensive homology ( $p < 0.05$ ). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



### Peptide Summary Report

Format As	Peptide Summary	<a href="#">Help</a>		
Significance threshold p<	0.05	Max. number of hits	AUTO	
Standard scoring	<input checked="" type="radio"/> MudPIT scoring <input type="radio"/> Ions score or expect cut-off	0	Show sub-sets	0
Show pop-ups	<input checked="" type="radio"/> Suppress pop-ups <input type="radio"/> Sort unassigned	Decreasing Score	Require bold red	<input type="checkbox"/>
Preferred taxonomy	All entries			

**Error tolerant**

1. [ALBU\\_BOVIN](#) Mass: 69248 Score: 64 Matches: 1(1) Sequences: 1(1)  
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide	
<input checked="" type="checkbox"/>	<a href="#">1</a>	461.7491	921.4836	921.4807	3.15	0	64	3e-06	1	U	K.AEFVEVTK.L

### Search Parameters

Type of search : MS/MS Ion Search  
Enzyme : Trypsin  
Mass values : Monoisotopic  
Protein Mass : Unrestricted  
Peptide Mass Tolerance :  $\pm 5$  ppm  
Fragment Mass Tolerance :  $\pm 0.5$  Da  
Max Missed Cleavages : 1  
Instrument type : Default  
Number of queries : 1

Mascot: <http://www.matrixscience.com/>

# PFF (resultados de Mascot)

## MATRIX SCIENCE MASCOT Search Results

### Protein View: ALBU\_BOVIN

Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Database: SwissProt  
Score: 61  
Nominal mass ( $M_r$ ): 69248  
Calculated pI: 5.82  
Taxonomy: [Bos taurus](#)

Sequence similarity is available as [an NCBI BLAST search of ALBU\\_BOVIN against nr.](#)

#### Search parameters

MS data file: new 1.txt

Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.

#### Protein sequence coverage: 1%

Matched peptides shown in **bold red**.

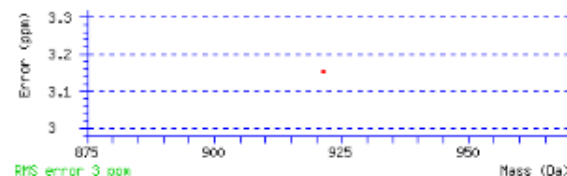
```
1 MKWVIFISLL LFFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA
51 FSQYLQCCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
101 VASLRETYGD MADCCCKQEP ERNECFLSHK DDSPDLPLK FDPNTLCDEF
151 KADEKKFWGK YLYEIAARRHP YFYAPELLYY ANKYNGVFQE CCQAEKDGAC
201 LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFFKAE
251 FVEVTKLVTD LTKVHKECOH GDLEECADDR ADLAKYICDN QDTISSKLKE
301 CCKDKLLEKS HCIAEVEKDA IPENLPLPLA DFAEDKDVCK NYQEAKDAFL
351 GSFLYEYSRR HPEYAVSVLL RLAKYEYATL EECCKADDPH ACYSTVFDKL
401 KHLVDEPQNL IKQNCQDFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS
451 RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT FVSEKVTKCC
501 TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT
551 ALVELLKHKP KATEEQKLTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV
601 SIQTALA
```

Unformatted sequence string: [607 residues](#) (for pasting into other applications).

Sort peptides by  Residue Number  Increasing Mass  Decreasing Mass

Show predicted peptides also

Query	Start - End	Observed	Mr (expt)	Mr (calc)	ppm	M Score	Expect	Rank	U	Peptide
<a href="#">1</a>	249 - 256	461.7491	921.4836	921.4807	3.15	0	61	6.1e-06	1	<b>U K.AEFVEVTK.L</b>



AC P02769; A5FJXS; O02787; P04277; Q382R2;  
DI 21-JUL-1986, integrated into UniProtKB/Swiss-Prot.  
DT 01-FEB-1996, sequence version 4.  
DT 14-OCT-2015, entry version 148.  
DE RecName: Full=Serum albumin;  
DE AltName: Full=BSA;  
DE AltName: Allergen=Bos d 6;  
DE Flag: Precursor;

# PFF (resultados de Mascot)

## MATRIX SCIENCE Mascot Search Results

### Peptide View

MS/MS Fragmentation of **AEFVEVTK**

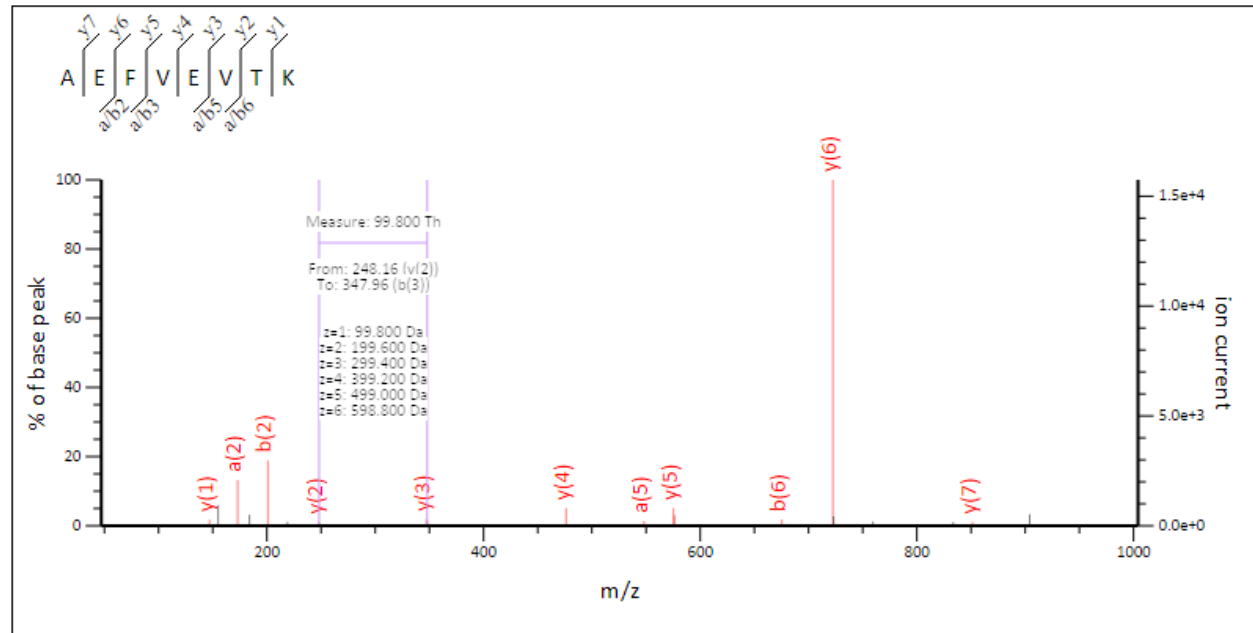
Found in **ALBU\_BOVIN** in **SwissProt**, Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Match to Query 1: 921.483648 from(461.749100,2+) intensity(18565.0000) index(0)

Title: Cmpd 7, +MSn(461.8), 16.8 min

Data file new 1.txt

COM=10 pmol digest of Sample BSA  
MASS=Monoisotopic  
USERNAME=Pedro Domingues  
USEREMAIL=p.domingues@ua.pt  
BEGIN IONS  
TITLE= Cmpd 7, +MSn(461.8), 16.8 min  
PEPMASS=461.7491 18565  
CHARGE=2+  
147.05 229  
154.98 787 1+  
172.95 1734 1+  
183.87 410 1+  
200.91 2479 1+  
218.92 144  
248.16 150  
347.22 250  
347.96 200  
476.17 670 1+  
548.13 180  
575.33 672  
576.37 393 1+  
675.33 231  
722.45 13125 1+  
723.14 357 1+  
759.39 159  
833.47 141  
851.45 149  
904.13 432 1+  
END IONS



47.05 to 1004.13



# 2-PFF (resultados de Mascot)

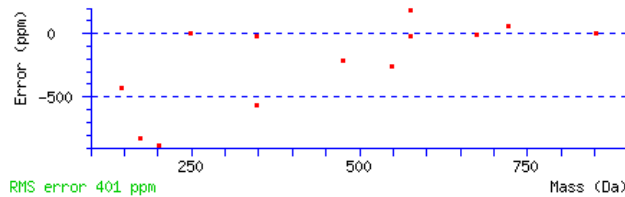
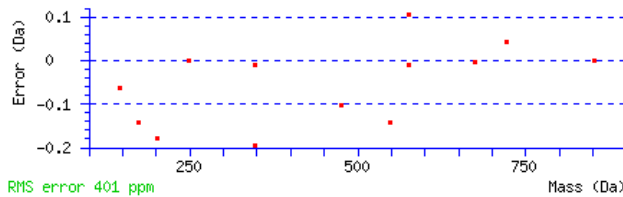
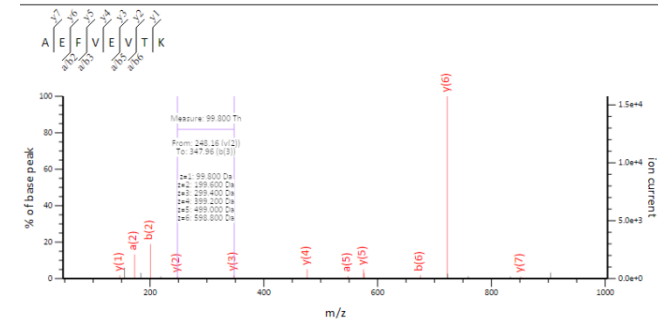
Label all possible matches  Label matches used for scoring

Monoisotopic mass of neutral peptide Mr(calc): 921.4807

Ions Score: 64 Expect: 2.5e-05

Matches : 13/56 fragment ions using 19 most intense peaks [\(help\)](#)

#	a	a <sup>++</sup>	b	b <sup>++</sup>	Seq.	y	y <sup>++</sup>	y <sup>*</sup>	y <sup>+++</sup>	#
1	44.0495	22.5284	72.0444	36.5258	A					8
2	<b>173.0921</b>	87.0497	<b>201.0870</b>	101.0471	E	<b>851.4509</b>	426.2291	834.4244	417.7158	7
3	320.1605	160.5839	<b>348.1554</b>	174.5813	F	<b>722.4083</b>	361.7078	705.3818	353.1945	6
4	419.2289	210.1181	447.2238	224.1155	V	<b>575.3399</b>	288.1736	558.3134	279.6603	5
5	<b>548.2715</b>	274.6394	<b>576.2664</b>	288.6368	E	<b>476.2715</b>	238.6394	459.2449	230.1261	4
6	647.3399	324.1736	<b>675.3348</b>	338.1710	V	<b>347.2289</b>	174.1181	330.2023	165.6048	3
7	748.3876	374.6974	776.3825	388.6949	T	<b>248.1605</b>	124.5839	231.1339	116.0706	2
8					K	<b>147.1128</b>	74.0600	130.0863	65.5468	1



NCBI BLAST search of [AEFVEVTK](#)

(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)

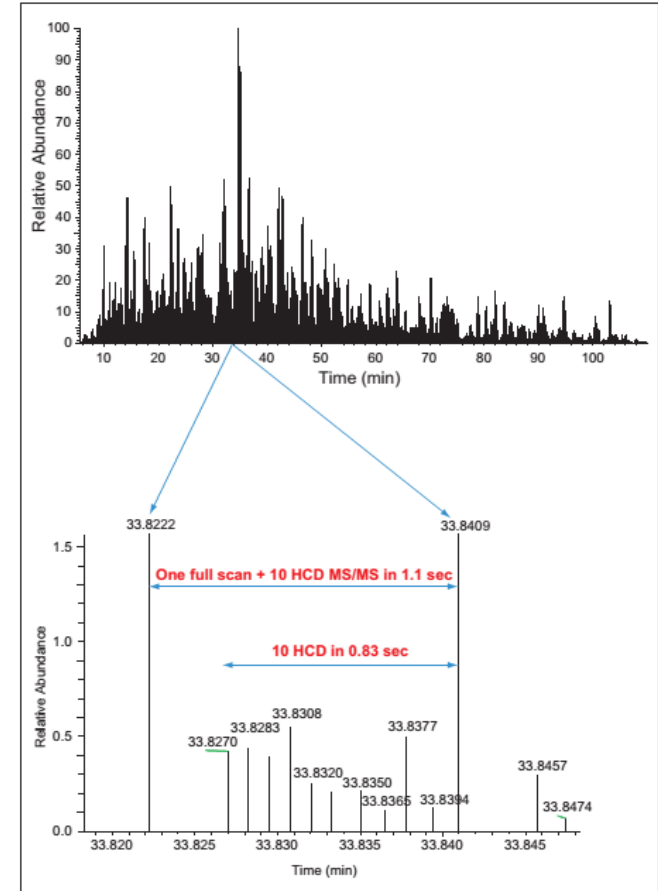
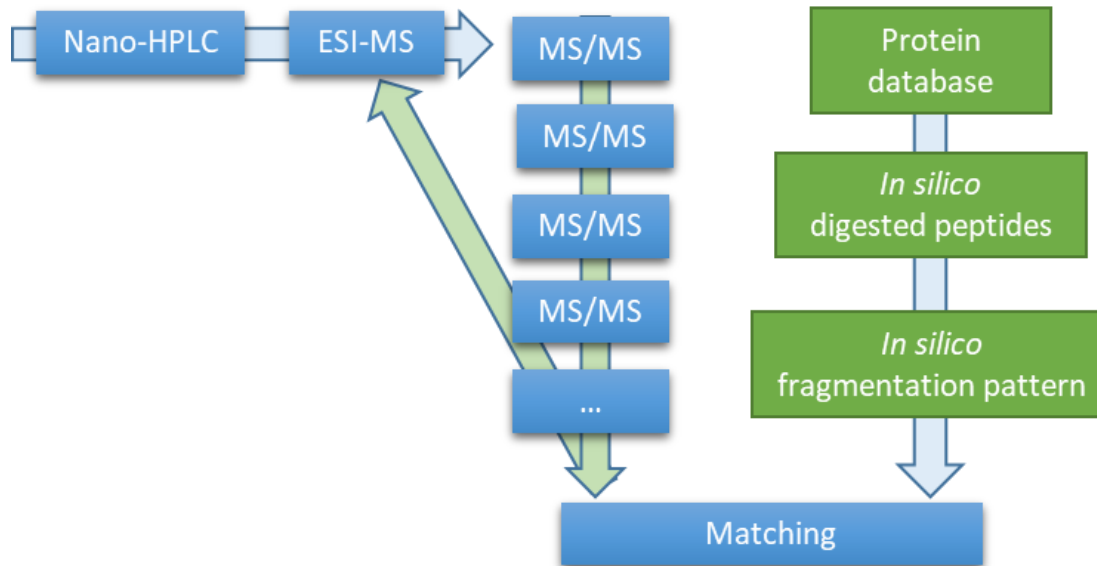
Other BLAST [web gateways](#)

All matches to this query

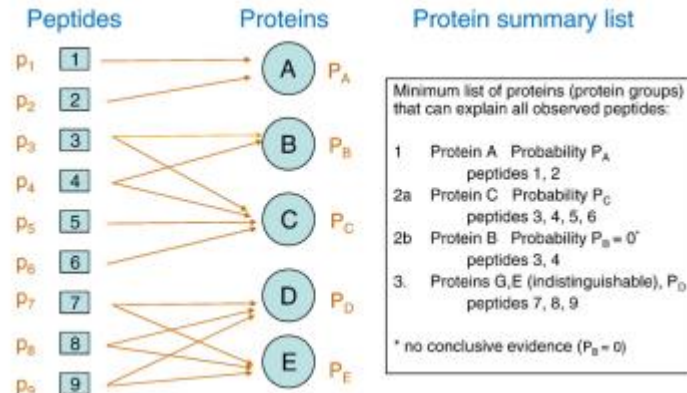
Score	Mr(calc)	Delta	Sequence
64.2	921.4807	0.0029	<a href="#">AEFVEVTK</a>
14.1	921.4807	0.0029	<a href="#">YIADLEAK</a>
12.8	921.4807	0.0029	<a href="#">EADLFISK</a>
12.8	921.4807	0.0029	<a href="#">VTDFLAEK</a>
11.8	921.4841	-0.0005	<a href="#">VTMETLTK</a>
11.5	921.4807	0.0029	<a href="#">AEALYDIK</a>
11.5	921.4807	0.0029	<a href="#">AEAYLLDK</a>
11.5	921.4807	0.0029	<a href="#">AEYIADIK</a>

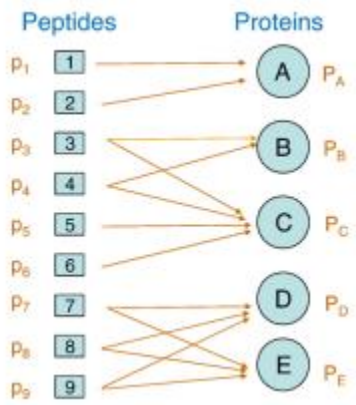
- Conjuntos de datos ortogonales y niveles de confianza
  - Db : 100.000 secuencias
  - 500 espectros
- Probabilidad de que un espectro (cualquiera) coincida «accidentalmente» con un secuencia (coincidencia errónea) :
  - $1/100.000 \times 500 = 5,10^{-3}$  (0,005)
- Probabilidad de que dos espectros coincidan «accidentalmente» con la misma secuencia (coincidencia errónea):
  - $5,10^{-3} \times 5,10^{-3} = 2,5 \times 10^{-5}$
- **Confianza de identificación mucho mayor al coincidir al menos 2 péptidos con la misma secuencia proteica**

# Espectrometría de Masas en Tándem Dependiente de Datos



- Fácilmente automatizada para alto rendimiento
- Puede obtener coincidencias a partir de datos marginales
- Puede resultar lenta
- Extenso conjunto de datos
- MS/MS es identificación peptídica
- Proteínas por inferencia.



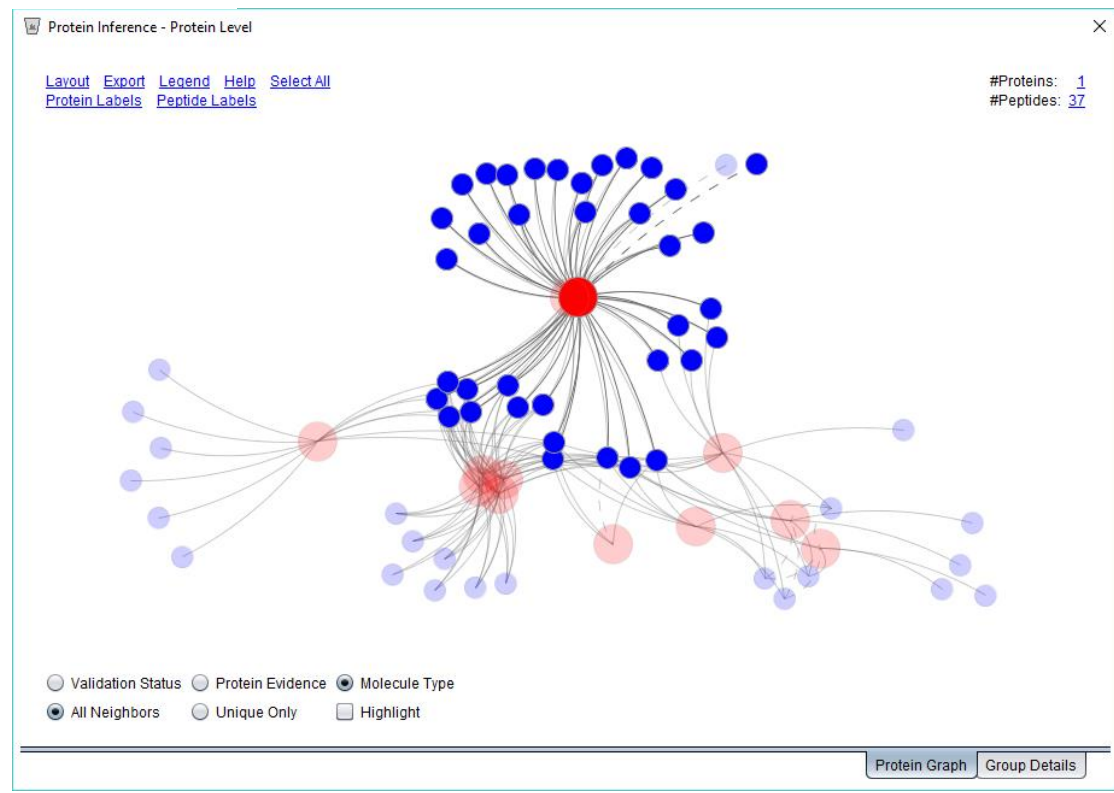


Protein summary list

Minimum list of proteins (protein groups) that can explain all observed peptides:

1. Protein A Probability  $P_A$   
peptides 1, 2
- 2a. Protein C Probability  $P_C$   
peptides 3, 4, 5, 6
- 2b. Protein B Probability  $P_B = 0^*$   
peptides 3, 4
3. Proteins G,E (Indistinguishable),  $P_D$   
peptides 7, 8, 9

\* no conclusive evidence ( $P_B = 0$ )

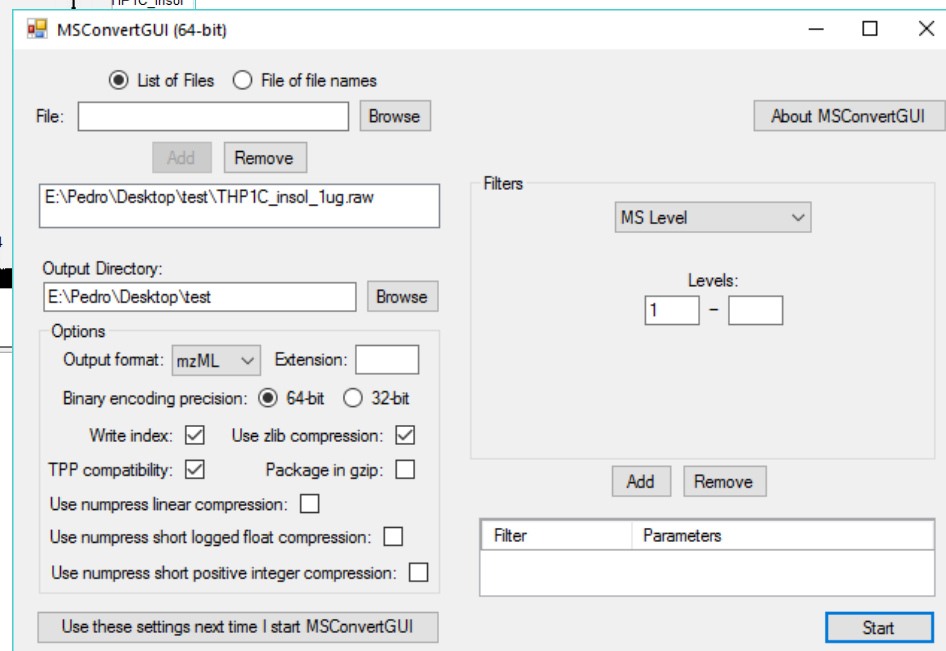
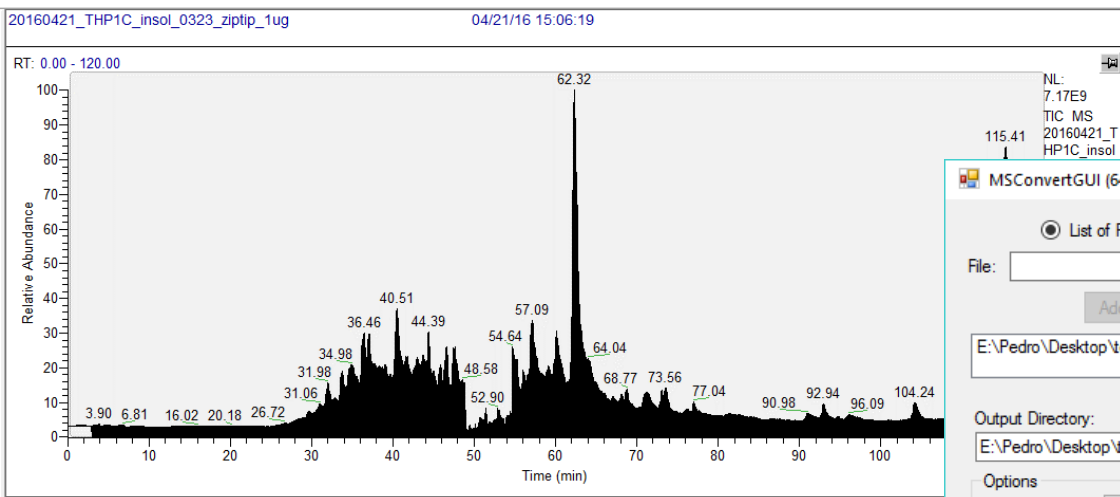




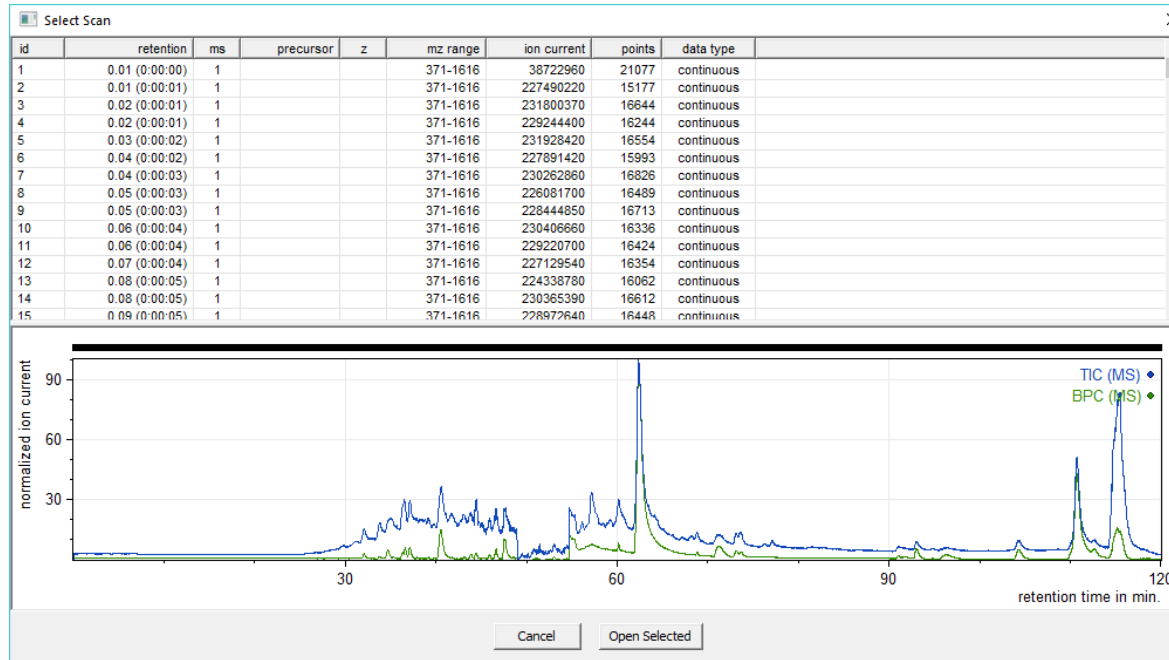
# Bioinformática Práctica

# Transformación de datos a mzML con MSConverter en ProteoWizard

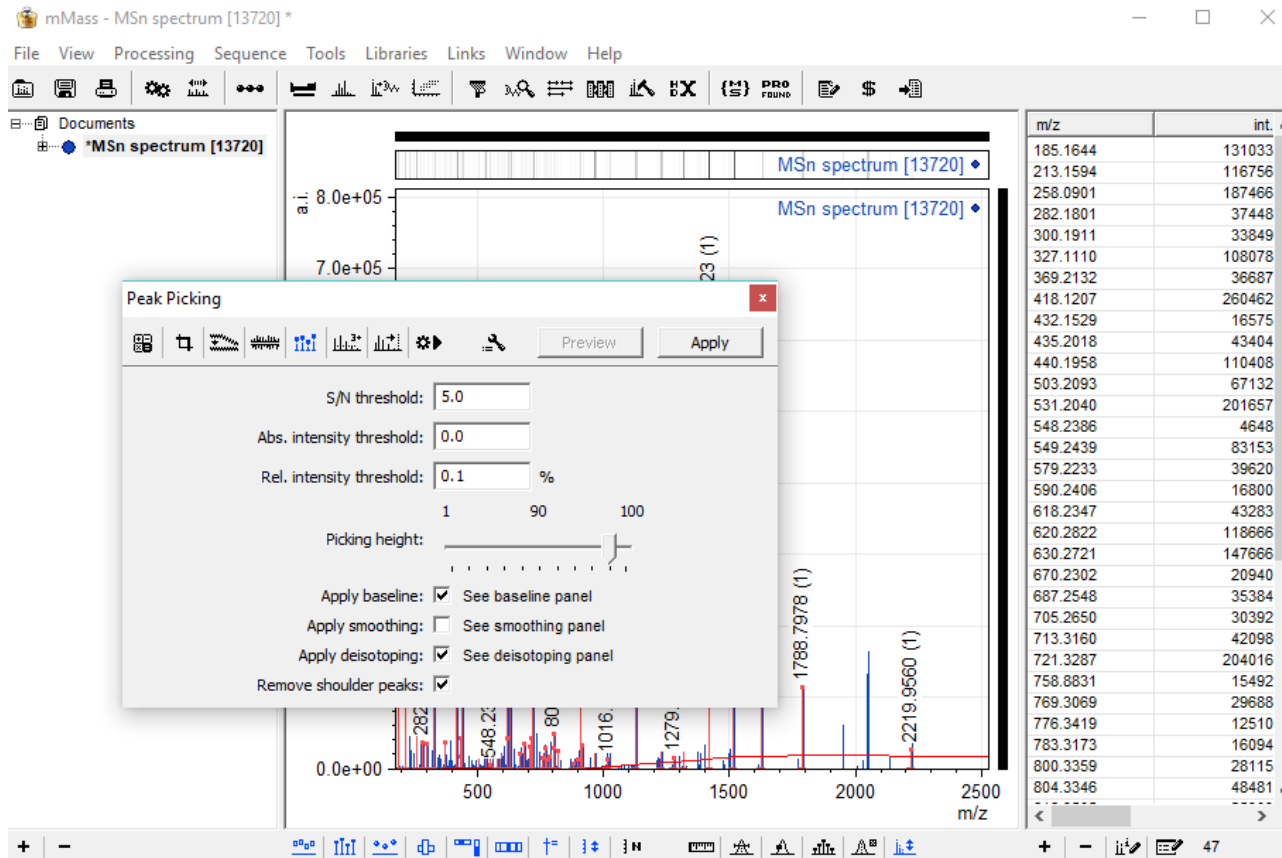
- 1) Para ver los datos en la aplicación mMass, aunque se pueda leer el formato de archivo mgf que se utilice en SearchGUI, es preferible convertir los datos BRUTOS a formato mzML.



# Datos observados en la aplicación mmass



- 1) Observa la frecuencia de espectros MS y MS/MS en la lista de adquisición
- 2) Observa el espectro MS con un RT de 47,50
- 3) Busca MSMS con m/z 1216,56(+2) con un RT de 47,53
- 4) Busca MS/MS en Mascot



# Busca MS/MS en Mascot

Mascot - MS/MS Ion Search

Server: Matrix Science Search

Title: MSn spectrum [13720]

Name: pedro E-mail: p.domingues@ua.pt

Taxonomy: Homo sapiens (human)

Database: SwissProt Enzyme: Trypsin Misc.: 2

Fixed modifications: (0)

- Acetyl (K)
- Acetyl (N-term)
- Acetyl (Protein N-term)
- Amidated (C-term)
- Amidated (Protein C-term)
- Ammonia-loss (N-term C)
- Biotin (K)

Variable modifications: (3)

- ICPL:13C(6)2H(4) (K)
- ICPL:13C(6)2H(4) (N-term)
- ICPL:13C(6)2H(4) (Protein N-term)
- ICPL:2H(4) (K)
- ICPL:2H(4) (Protein N-term)
- ITRAQ4plex (K)
- ITRAQ4plex (N-term)

Show hidden modifications

Precursor m/z: 1216.55735674 Mass type: Monoisotopic

Precursor tolerance: 10.0 ppm Charge: 2+

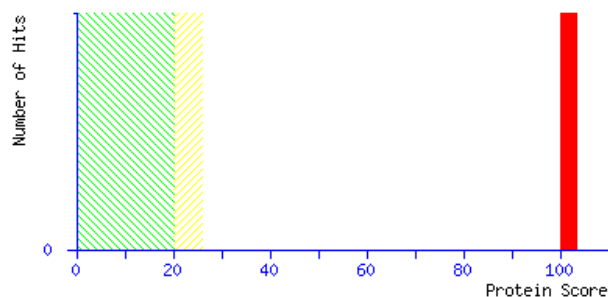
MS/MS tolerance: 0.05 Da Instrument: Default

Quantitation: None

Decoy  Error tolerant Report: AUTO

## Mascot Score Histogram

Ions score is  $-10 \cdot \log(P)$ , where P is the probability that the observed match is a random event. Individual ions scores > 20 indicate peptides with significant homology. Individual ions scores > 26 indicate identity or extensive homology ( $p < 0.05$ ). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



## Peptide Summary Report

Format As	Peptide Summary	<a href="#">Help</a>
Significance threshold p<	0.05	Max. number of hits
Standard scoring	<input checked="" type="radio"/> MudPIT scoring	Display non-significant matches
Show pop-ups	<input checked="" type="radio"/> Suppress pop-ups	Sort unassigned
Preferred taxonomy	All entries	Require bold red

Select All   Select None   Search Selected    Error tolerant

1. [HS90B\\_HUMAN](#)   Mass: 83212   Score: 102   Matches: 1(1)   Sequences: 1(1)

Heat shock protein HSP 90-beta OS=Homo sapiens GN=HSP90AB1 PE=1 SV=4

Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
<input checked="" type="checkbox"/> 1	1216.5574	2431.1002	2431.0970	1.28	0	102	1.4e-09	1	U	R.LVSSPCCIVTSTYGWTANMER.I + 2 Carbamidomethyl (C)

## Search Parameters

Type of search : MS/MS Ion Search  
 Enzyme : Trypsin  
 Variable modifications : [Acetyl \(K\)](#), [Carbamidomethyl \(C\)](#), [Oxidation \(M\)](#)  
 Mass values : Monoisotopic  
 Protein Mass : Unrestricted  
 Peptide Mass Tolerance : ± 10 ppm  
 Fragment Mass Tolerance : ± 0.05 Da  
 Max Missed Cleavages : 2  
 Instrument type : Default  
 Number of queries : 1



## Sequence Modifications

Mo. Protein Digest

Un. Posi. Mass:  Mo  Av Max charge:  Digest Match Annotate

All C Enzyme:  Misc.:  Mass range:  -   Ignore mods Coverage: 0/99 %

slice	mis.	m/z	z	sequence	error
[73-82]	0	1194.6477	1	k IDIPNPOER t	

Sequence Editor

Sequence type:   Cyclic Pattern

## Peptide Fragmentation

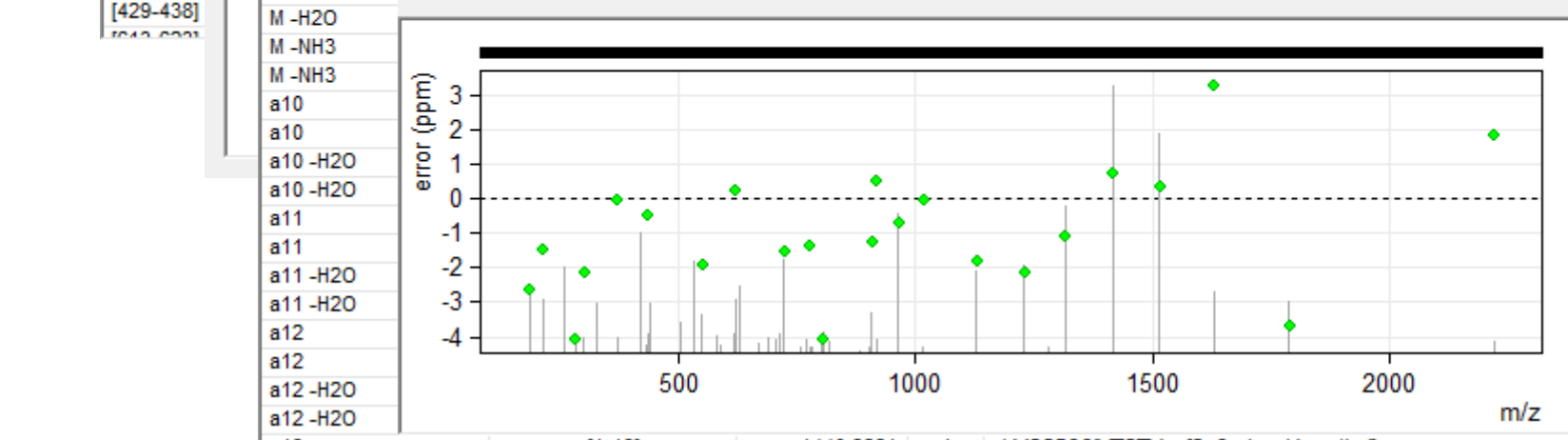
Mo. Un. Mass:  Mo  Av Max charge:  Fragment Match Annotate

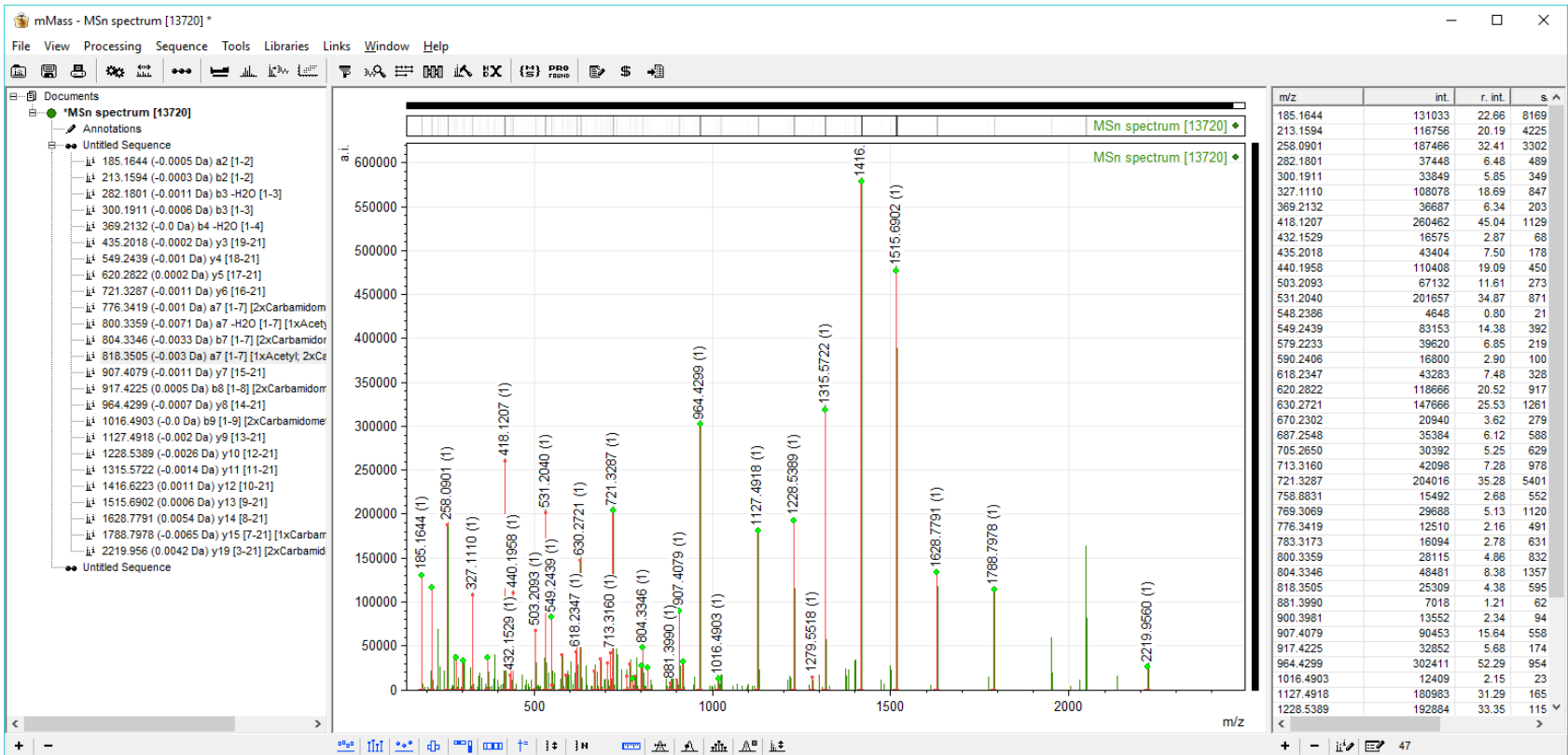
LV  M  a  b  c  int-a  N-ladder  -H2O  -CO  Defined losses  +H2O  Allow scrambling

## Match Fragments

ion  M  im Tolerance:   Da  ppm  Ignore charge Match Calibrate

Remove:  Annotated  Matched  Unselected  Isotopes  Unknown







- Herramientas
  - *Mass calculator*
  - *Mass to formula*

# Generación de base de datos FASTA a partir de UniProt (SwissProt)

- 1) Descarga ahora el archivo FASTA del proteoma de HOMO sapiens.
  - a) Esto se debe hacer en la página de taxonomía del sitio web de Uniprot y buscando Homo Sapiens

The screenshot shows the UniProt Taxonomy results page for 'human'. The search bar contains 'Taxonomy human' and 'Search'. Below the search bar, there are navigation links: 'BLAST', 'Align', 'Retrieve/ID mapping', 'Peptide search', 'Help', and 'Contact'. The main heading is 'Taxonomy results' with a link to 'About Taxonomy'. A 'Filter by' section includes a 'Download' button and a pagination indicator '1 to 25 of 8,220'. A 'Taxons with:' section lists 'entries in UniProtKB (7,305)', 'reviewed entries in UniProtKB (531)', and 'a proteome (1,051)'. The main content area shows a 'Repeat search in UniProtKB (2,393,722)' button and a 'Taxon' list. The selected taxon is 'Homo sapiens (Human)', with a detailed taxonomic path: 'Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo'. Below the taxon list, there are links for 'Proteomes (1) UniProtKB (156,851)- Reviewed (Swiss-Prot) (20,168) Unreviewed (TrEMBL) (136,683)'.

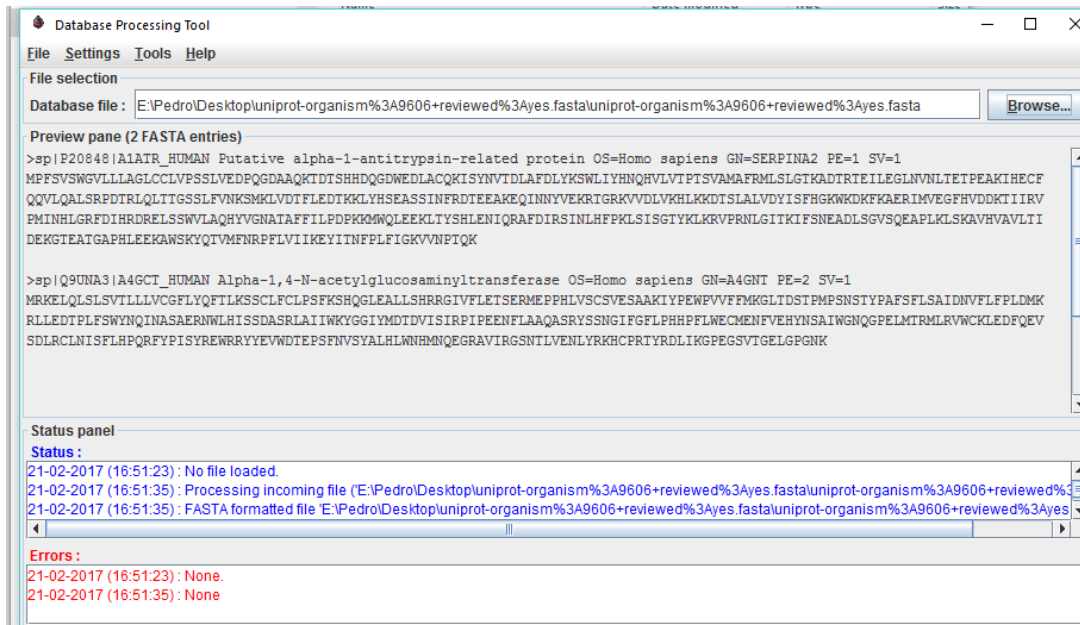
The screenshot shows the UniProtKB results page for the search query 'reviewed:yes AND organism:"Homo sapiens (Human) [9606]"'. The search bar contains 'UniProtKB reviewed:yes AND organism:"Homo sapiens (Human) [9606]"' and 'Search'. Below the search bar, there are navigation links: 'BLAST', 'Align', 'Retrieve/ID mapping', 'Peptide search', 'Help', and 'Contact'. The main heading is 'UniProtKB results' with a link to 'About UniProtKB' and a 'Basket' button. A 'Filter by' section includes a 'Download' button, an 'Add to basket' button, and a 'Columns' button. A pagination indicator shows '1 to 25 of 20,168'. A 'Download' dropdown menu is open, showing options: 'Download selected (0)', 'Download all (20168)', 'Format: FASTA (canonical)', 'Compressed', 'Uncompressed', and 'Preview first 10'. The main content area shows a table of results with columns: 'Entry', 'Entry', 'Gene names', 'Organism', and 'Length'. The table contains the following data:

Entry	Entry	Gene names	Organism	Length
P20848	A1AT	SERPINA2 ARGs, ATR, PIL, SERPINA2P	Homo sapiens (Human)	420
Q9UNA3	A4GD	A4GNT	Homo sapiens (Human)	340
Q96GX2	A7L3B_HUMAN	Putative ataxin-7-like protein 3B	Homo sapiens (Human)	97
Q8N5Z0	AADAT_HUMAN	Kynurenine/alpha-aminoadipate amino...	Homo sapiens	425

- 2) A continuación debes descargar el archivo FASTA (de Swiss-Prot) revisado

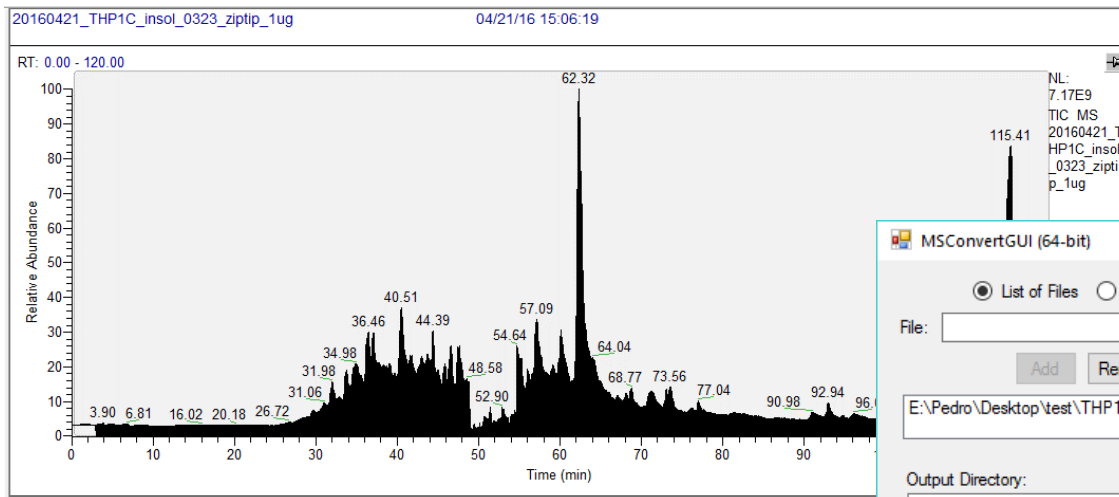
# Herramienta de procesamiento *Dbtoolkit dataBase*

Puedes ver y editar la información del archivo FASTA con la *dataBase Processing Tool* (herramienta de procesamiento de bases de datos) de Dbtoolkit



# Transformación de datos a MGF con MSConverter en ProteoWizard

- Utiliza MSConverter en ProteoWizard para convertir los datos de HPC-MS adquiridos en la Orbitrap (archivo de datos BRUTOS) a un formato que se pueda leer con SearchGUI (archivo de datos MGF).



MSConvertGUI (64-bit)

List of Files  File of file names

File:  Browse

Add Remove

Output Directory:  Browse

Options

Output format:  Extension:

Binary encoding precision:  64-bit  32-bit

Write index:  Use zlib compression:

TPP compatibility:  Package in gzip:

Use numpress linear compression:

Use numpress short logged float compression:

Use numpress short positive integer compression:

Filters

MS Level  -

Add Remove

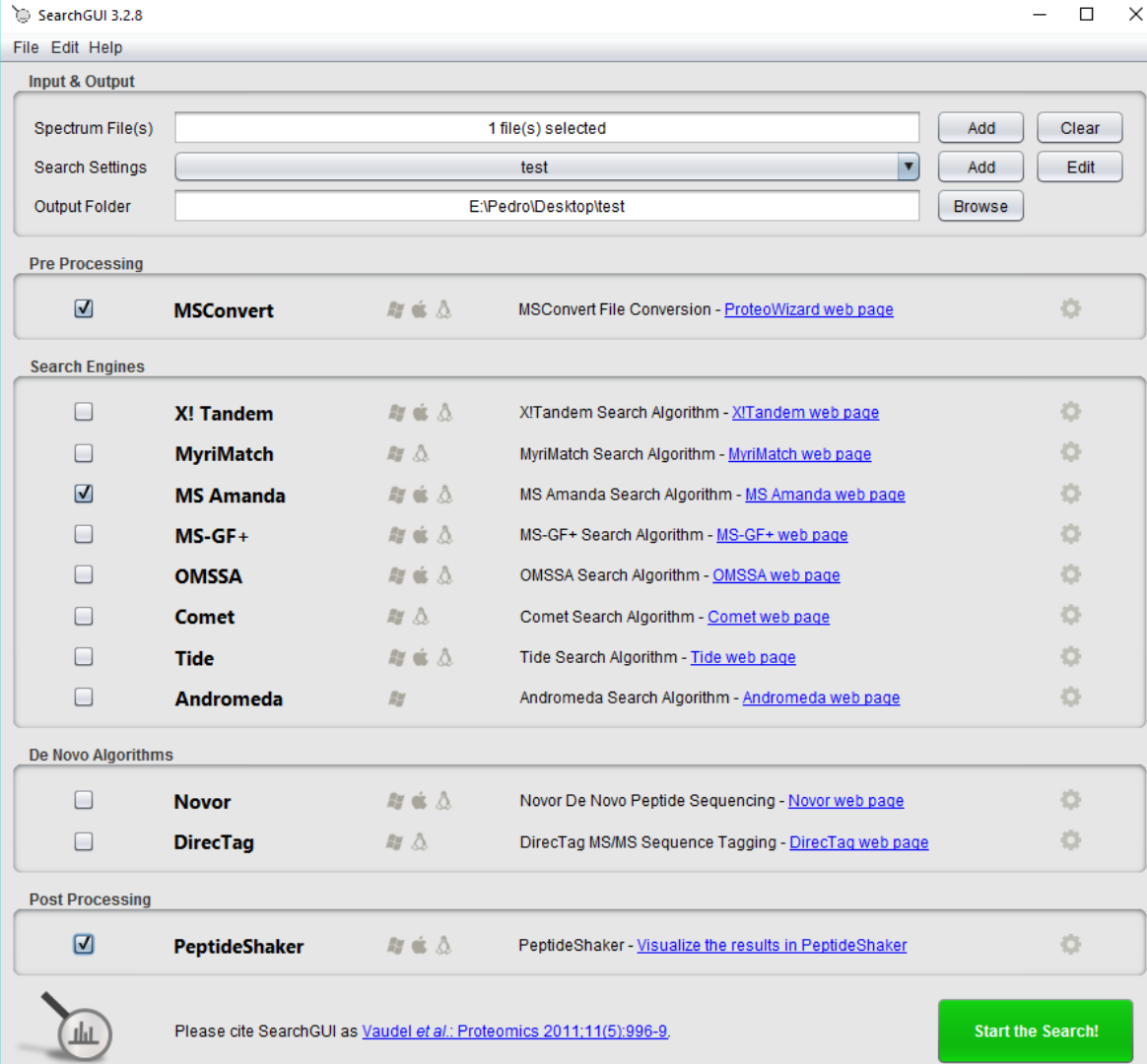
Filter Parameters

Start

Use these settings next time I start MSConvertGUI

# Motores de búsqueda: SearchGUI

- 1) Abre SearchGUI.
- 2) En los ajustes de la búsqueda, edita y cubre el formulario del modo que se muestra.
- 3) Se escogieron modificaciones de variables (¿por qué?).
- 4) Puedes configurar también el mezclador de péptidos para abrir el archivo de resultados, tal y como se muestra abajo.
- 5) SearchGUI preguntará si quieres crear un archivo fasta concatenated\_target\_decoy. Di sí (¿por qué es importante hacerlo?).



SearchGUI 3.2.8

File Edit Help

**Input & Output**

Spectrum File(s) 1 file(s) selected Add Clear

Search Settings test Add Edit

Output Folder E:\Pedro\Desktop\test Browse

**Pre Processing**

**MSConvert** MSConvert File Conversion - [ProteoWizard web page](#)

**Search Engines**

**X! Tandem** XITandem Search Algorithm - [XITandem web page](#)

**MyriMatch** MyriMatch Search Algorithm - [MyriMatch web page](#)

**MS Amanda** MS Amanda Search Algorithm - [MS Amanda web page](#)

**MS-GF+** MS-GF+ Search Algorithm - [MS-GF+ web page](#)

**OMSSA** OMSSA Search Algorithm - [OMSSA web page](#)

**Comet** Comet Search Algorithm - [Comet web page](#)

**Tide** Tide Search Algorithm - [Tide web page](#)

**Andromeda** Andromeda Search Algorithm - [Andromeda web page](#)

**De Novo Algorithms**

**Novor** Novor De Novo Peptide Sequencing - [Novor web page](#)

**DirectTag** DirectTag MS/MS Sequence Tagging - [DirectTag web page](#)

**Post Processing**

**PeptideShaker** PeptideShaker - [Visualize the results in PeptideShaker](#)

Please cite SearchGUI as [Vaudel et al.: Proteomics 2011;11\(5\):996-9.](#)

Start the Search!

## Selecciona las opciones de búsqueda y la base de datos FASTA

Se escogieron modificaciones de variables (¿por qué?)

Search Settings - test

Database (FASTA)

**Modifications**

Fixed Modifications (0)

Name	Mass

Variable Modifications (3)

Name	Mass
Oxidation of M	15.99
Carbamidomethylation of C	57.02
Acetylation of K	42.01

Most Used Modifications

Name	Mass
Acetylation of protein N-term	42.01
Deamidation of N	0.98
Deamidation of Q	0.98
Phosphorylation of S	79.97
Phosphorylation of T	79.97
Phosphorylation of Y	79.97
Pyroglutamine from E	-18.01
Pyroglutamine from Q	-17.03
Pyroglutamine from carbamidomethylated C	-17.03
TMT 10-plex of K	229.16
TMT 10-plex of peptide N-term	229.16
TMT 6-plex of K	229.16
TMT 6-plex of peptide N-term	229.16
iTRAQ 4-plex of K	144.10

**Protease & Fragmentation**

Digestion

Enzyme

Specificity

Max Missed Cleavages

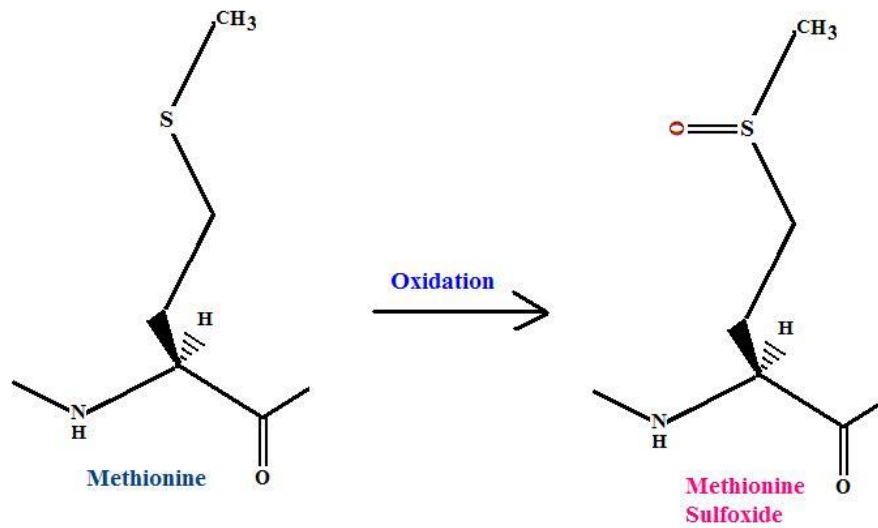
Fragment Ion Types

Precursor m/z Tolerance

Fragment m/z Tolerance

Precursor Charge  -

Isotopes  -



lysine acetylation iodoac x K-acetylation and its enz x Overalkylation of a Prote x

← → ↻ Not secure | pubs.acs.org/doi/pdf/10.1021/ac0103423?src=recsys

**analytical chemistry**

Search Citation Subject **Advanced Search**

Enter search text / DOI Anywhere Search

Anal. Chem.  All Publications/Website

[Browse the Journal](#)
[Articles ASAP](#)
[Current Issue](#)
[Multimedia](#)
[Submission & Review](#)
[Open Access](#)
[About the Journal](#)

Article 
[← Previous Article](#)
[Next Article →](#)
[Table of Contents](#)

## Overalkylation of a Protein Digest with Iodoacetamide

Emily S. Boja and Henry M. Fales\*  
National Heart, Lung and Blood Institute, National Institutes of Health, 10 Center Drive, MSC 1676, Bethesda, Maryland 20892

*Anal. Chem.*, 2001, 73 (15), pp 3576–3582  
DOI: 10.1021/ac0103423  
Publication Date (Web): June 30, 2001  
Copyright Not subject to U.S. Copyright. Published 2001 American Chemical Society

**i** Your current credentials do not allow retrieval of the full text.

**Abstract**

Cystine linkages in proteins are often opened with reducing agents, sometimes to improve their digestion, often to eliminate disulfide linkages from complicating analysis of the digest. After reduction, the sulfhydryls are usually reacted with iodoacetamide (IAM), iodoacetic acid (IAA), or another electrophile to prevent reformation of disulfide linkages in a random manner. When the amount of protein may be reliably estimated, side reactions from excess IAM or IAA can be avoided. When this is not so, removal of excess iodoalkane can be accomplished by HPLC, by dialysis, or simply by allowing a reducing thiol to consume any excess. In mass spectrometric analysis of proteins isolated by 1D or 2D gels, removal of the excess iodoalkane is often accomplished simply by washing the gel prior to proteolytic digestion. During a recent study of the glutathionylation site mapping of actin, IAM was used to block any residual sulfhydryl groups remaining on the protein so that they would not displace glutathione from its initial site. In addition, to avoid losses due to actin polymerization during dialysis, the IAM was allowed to remain during the digestion. This further ensured that any sulfhydryl groups liberated during the digestion would be similarly blocked by the IAM. Under these conditions, we observed the peptides to undergo N- as well as S-carbamidomethylation. In examining a series of other peptides alkylated with IAM in this way, we have found N-alkylation to be the rule rather than the exception and even O-alkylation

**Article Options**

- PDF (140 KB)
- PDF w/ Links (0 KB)
- Full Text HTML
- Abstract
- Figures
- References
- Citing Articles

[★ Add to Favorites](#)  
[Download Citation](#)  
[Email a Colleague](#)  
[Order Reprints](#)  
[Rights & Permissions](#)  
[Citation Alerts](#)

**avast**

Add to

**SCIFID**

- Retrieve Detailed Re
- Retrieve Substances
- Retrieve All Referenc
- Retrieve All Referenc

Anyone can see what you do online  
Keep your browsing safe and private.

**HIDE ME NOW**



## Selecciona las opciones de peptideshaker

PeptideShaker Settings

**PeptideShaker**

Location

**Project Details**

Project Name

Sample Name  Replicate

**Output**

Output File

**Advanced Settings (see help for details)**

Project

Mascot Files

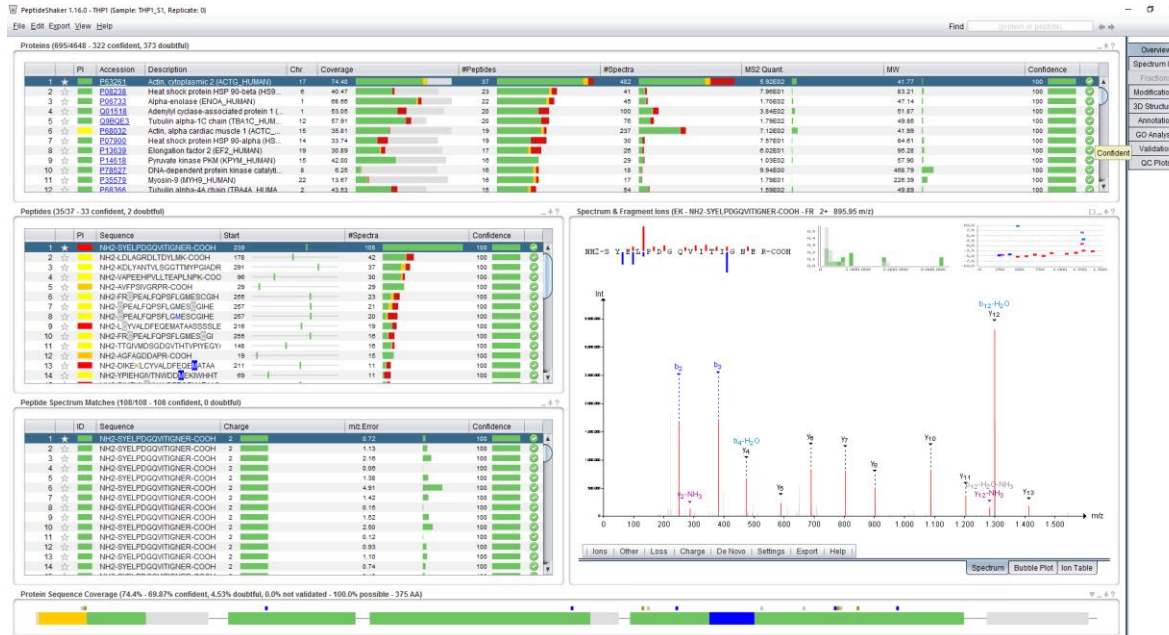
# Generación y evaluación de resultados de resultados: PeptideShaker para la visualización, y validación, de péptidos y proteínas. Análisis PTM

## 1) Después de realizar la búsqueda con searchGUI

- (+/- 3 minutos aproximadamente con un Intel I7-6700K de 16MB de RAM), se abrirá la ventana de resultados en la plataforma peptideshaker.

## 2) Aquí podrás ver que se han identificado en torno a 700 proteínas, aunque 373 han sido clasificadas como dudosas (¿por qué?).

## 3) Podrás ver además información sobre los péptidos identificados para cada proteína y los espectros de masas con el patrón de fragmentación anotado.



## Una asignación espectro-péptido (PSM)

Quality Control (beta)

**General Settings**

Mark as Doubtful

- Hits obtained on small databases (<1000 protein sequences)
- Datasets with a low number of target hits
- Hits near the confidence threshold (margin= 1 x resolution)

**Protein Filters**

	Name	Description
1	>=2 confident peptides	Number of confident peptides filter
2	>=2 confident spectra	Number of confident spectra filter

**Peptide Filters**

	Name	Description
1	One confident PSM	Number of confident PSMs filter

**PSM Filters**

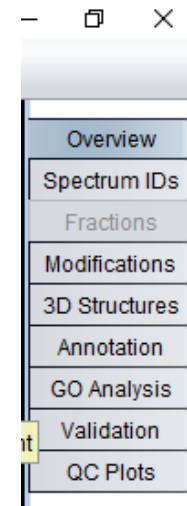
	Name	Description
1	Fragment Ion Sequence Coverage	Sequence coverage filter by fragment ions
2	Mass deviation	Precursor m/z deviation probability

Right-click in the tables to edit the filters.

OK Cancel

# Análisis de datos: información proteica, análisis de rutas metabólicas y ontología de genes

Explora las opciones avanzadas de análisis de datos del peptideshaker abriendo la pestaña de modificaciones, y la pestaña «GO Analysis».



# Análisis de datos: información proteica, análisis de rutas metabólicas y ontología de genes

En la pestaña «Annotation» podrás anotar información para cada proteína. No obstante, si quieres realizar anotaciones de varias proteínas, será necesario que exportes tus resultados (informe de proteínas por defecto) y que hagas clic en el enlace web que hay al lado del recurso, y que sigas las instrucciones que aparecen en la página web del recurso.

PeptideShaker 1.16.0 - THP1 (Sample: THP1\_S1, Replicate: 0)

File Edit Export View Help Find (protein or peptide)

### Basic Protein Annotation

Accession: P63261  
Description: Actin, cytoplasmic 2 (ACTG\_HUMAN)  
Gene Name: ACTG1 Chromosome: 17  
Taxonomy: Homo sapiens  
Database: UniProt

### Protein Annotation - Help

**Single Protein**  
To access the annotations for the currently selected protein, simply click the button corresponding to the wanted resource.

**Multiple Proteins**  
To get the list of all validated proteins in your project click [here](#).  
Advanced export options: *Export > Identification Features*.

To query using multiple proteins, click the [web](#) link next to the resource and follow the instructions provided at the resource web page.

### UniProt - protein knowledgebase

[Search UniProt](#) High-quality protein sequence and functional information. [web](#)

### neXtProt - human protein knowledgebase

[Search neXtProt](#) High-quality human protein information and annotation. [web](#)

### STRING - protein interaction

[Search STRING](#) Known and Predicted Protein-Protein Interactions. [web](#)

### QuickGO - gene ontology terms and annotations

[Search QuickGO](#) Web-based browser for Gene Ontology terms. [web](#)

### DASTy - protein sequence features

[Search DASTy](#) Web client for visualizing protein sequence feature information. [web](#)

### Reactome - pathway database

[Search Reactome](#) Manually curated and peer-reviewed pathway database. [web](#)

### DAVID - functional annotation

[Search DAVID](#) Database for Annotation, Visualization and Integrated Discovery. [web](#)

### IntAct - protein interaction

[Search IntAct](#) Analysis tools for protein interaction data. [web](#)

### InterPro - predictive protein signatures

[Search InterPro](#) Integrated database of predictive protein signatures. [web](#)

### PDB - protein data bank

[Search PDB](#) Biological macromolecular resource. [web](#)

[PICR - Protein Identifier Cross-Reference Service](#)

Overview  
Spectrum IDs  
Fractions  
Modifications  
3D Structures  
Annotation  
GO Analysis  
Validation  
QC Plots

## Muestra los resultados de este análisis:

- Número de proteínas identificadas
- Las 3 proteínas más abundantes
- Selecciona un grupo basado en el análisis GO
  - Muestra el análisis String (redes funcionales de asociación de proteínas) de este grupo
  - Anota una proteína importante (nodo central)

# Licencia de Creative Commons

Este material electrónico están bajo la licencia Creative Commons Attribution-NonCommercial 3.0 CC BY-NC 3.0.

