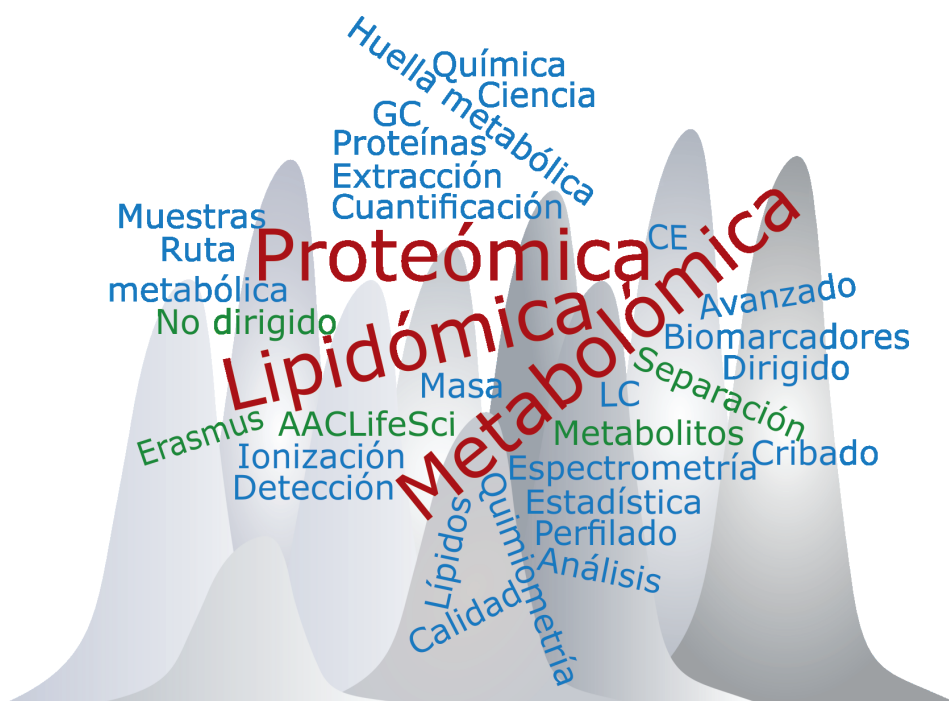


# AAALifeSci

## Química Analítica Avanzada en Ciencias de la Vida

*Advanced Analytical Chemistry for Life Sciences*

Editado por:  
Pedro Domingues, Antonia García, Elżbieta Skrzydlewska





# Química Analítica Avanzada en Ciencias de la Vida

Editado por  
Pedro Domingues, Antonia García, Elzbieta Skrzydlewska

**AACLifeSci**  
**Manual complementario para el Curso AACLifeSci**

Este proyecto ha sido financiado con ayuda de la Comisión Europea.  
Esta publicación refleja las opiniones de los autores y la Comisión no puede ser considerada responsable de ningún uso que se pueda hacer de la información que figura en la misma.

ISBN: 978-84-17591-06-9.  
Editorial: Liberlibro.com A.C.

## Índice

<b>Prólogo: Plan de estudios de AACLifeSci</b>	<b>6</b>
I. Justificación	6
II. Objetivos y resultados del curso	8
III. Formato y procedimientos	9
Organización del programa de doctorado de la Universidad de Medicina de Białystok y cómo se implementará AACLifeSci	9
Organización del programa de doctorado de la CEU y cómo se implementará AACLifeSci	11
Organización del programa de doctorado en Bioquímica de la UA y cómo se implementará AACLifeSci	13
IV. Lecturas de recursos adicionales	14
<b>Módulo 1 Técnica de separación/Cromatografía y Espectrometría de Masas para las Ciencias de la Vida</b>	<b>16</b>
I. Justificación	16
II. Objetivos y resultados del curso	16
II. Contenido del curso	17
1. Espectrometría de masas	18
1.1. Introducción a la espectrometría de masas	18
1.2. Técnicas de ionización en la espectrometría de masas biológica	19
1.3. Analizadores de masas en espectrometría de masas	22
1.4. Espectrometría de masas en tándem	26
1.5. Interpretación de los espectros de masas	28
2. Técnicas de separación acopladas a la espectrometría de masas	30
2.1. Cromatografía líquida-espectrometría de masas (LC-MS)	30
2.1.1. Introducción a la LC-MS	31
2.1.2. Principios teóricos de la cromatografía líquida	31
2.1.2.1. El proceso cromatográfico	34
2.1.2.2. Cromatografía de fase normal (NP)	35
2.1.2.3. Cromatografía de fase inversa (RP)	36
2.1.2.4. Cromatografía de interacción hidrofílica (HILIC)	36
2.1.2.5. Cromatografía de intercambio de iones (IEC)	37
2.1.3. Instrumentación en LC-MS	37
2.2. Cromatografía de gases-espectrometría de masas (GC-MS)	40
2.2.1. Introducción a la GC-MS	40
2.2.2. Principios teóricos de la GC-MS	40
2.2.2.1. El proceso cromatográfico	40
2.2.2.2. Fases de columna	41

2.2.2.3. GC×GC/MS	43
2.2.2.4. Pretratamiento y derivación de muestras	45
2.2.3. Acoplamiento GC-MS	48
2.3. Electroforesis capilar-espectrometría de masas (CE-MS)	49
2.3.1. Introducción a la CE-MS	49
2.3.2. Principios teóricos de la CE-MS	50
2.3.2.1. Electroforesis de zona capilar	54
2.3.2.2. Cromatografía capilar electrocinética micelar	54
2.3.2.3. Electro cromatografía capilar	54
2.3.2.4. Electroforesis capilar en gel	55
2.3.3. Instrumentation	55
3. Sesiones prácticas	58
3.1. Adquisición de un espectro de masas ESI-MS	58
3.2. Adquisición de un espectro de masas ESI-MS/MS	60
3.3. Interpretación del espectro de masas	62
3.4. Observación de la ejecución de una GC-MS y configuración de la adquisición por monitorización selectiva de iones	62
3.5. Observación de la ejecución de una HPLC-MS y configuración de la adquisición por MS/MS dependiente de datos	68
<b>Módulo 2 Metabolómica</b>	<b>74</b>
I. Justificación	74
II. Objetivos y resultados del curso	74
III. Contenido del curso	75
1. Introducción a la metabolómica	76
2. Enfoques analíticos en la metabolómica	78
2.1. Flujo de trabajo del estudio metabolómico	79
2.1.1. Diseño del experimento	79
2.1.2. Recogida de muestras	80
2.1.3. Extracción de metabolitos	80
2.1.4. Separación y detección	81
2.2. Procedimiento de Control de Calidad y Garantía de Calidad en la metabolómica	82
3. Procesamiento de datos e identificación de metabolitos	84
3.1. Flujo del procesamiento de datos	84
3.1.1. Procesamiento de datos brutos	85
3.1.2. Preprocesamiento de datos	85
3.1.3. Pretratamiento de datos	91
3.2. Identificación de metabolitos	92
3.2.1. Identificación en GC-EI-MS	92



3.2.2. Identificación de metabolitos en experimento de metabolómica de LC o CE (ESI)-MS: Trabajando con bases de datos	94
3.2.3. Bases de datos centradas en el metabolito sobre compuestos para búsquedas simples de PM	95
3.2.4. Bases de datos espectrales centradas en el metabolito	96
3.2.5. Mediadores	97
3.3. Análisis estadístico	98
3.3.1. PCA	99
3.3.2. PLS-DA	100
3.3.3. OPLS-DA	100
3.3.4. Gráfica S-plot, intervalo <i>Jack-knife</i> y Variable Importancia en la Proyección (VIP)	101
3.3.5. Validación de los modelos estadísticos multivariantes	101
4. Análisis de datos	102
4.1. De la identificación de datos a las rutas metabólicas: Evaluación de datos inteligente pero asistida para la evaluación de la relevancia biológica	102
4.1.1. Bases de datos de rutas metabólicas	103
4.1.2. Análisis de rutas metabólicas	104
4.1.3. Enriquecimiento de metabolitos	104
4.1.4. Análisis (topológico) de rutas metabólicas	106
4.2. Validación de biomarcadores: Asignación de relevancia a los metabolitos	106
5. Sesiones prácticas sobre metabolómica	107
5.1. Metabolómica dirigida y no dirigida	107
5.2. Preprocesamiento con Workflow4Metabolomics (W4M)	108
5.2.1. Conversión de archivos MS	108
5.2.2. Inspección de datos brutos	109
5.2.3. Subir datos a Galaxy	110
5.2.4. Preprocesamiento de datos de LC-MS a través de XCMS	114
5.3. Estadística	128
5.4. Anotación	129
<b>Módulo 3 Lipidómica</b>	<b>130</b>
I. Justificación	130
II. Objetivos y resultados del curso	130
III. Contenido del curso	131
1. Introducción a la lipidómica	132
1.1. Enfoques analíticos en la lipidómica	134
1.1.1. Preparación de muestras: extracción y fraccionamiento de lípidos	134
1.1.2. Identificación de lípidos a partir de MS: conceptos	136

1.1.2.1. Enfoques de la lipidómica no dirigida basada en LC-MS	138
1.1.2.2. Enfoques de la lipidómica dirigida de infusión directa	139
1.2. Cuantificación	140
1.3. Procesamiento de datos e identificación/cuantificación de lípidos	141
1.4. Sesiones prácticas	143
1.4.1. Identificación del perfil de fosfolípidos de monofitos THP1 mediante un enfoque lipidómico basado en HILIC-LC-MS	143
1.5. Lecturas de recursos adicionales	145
<b>Módulo 4 Proteómica</b>	<b>146</b>
I. Justificación	146
II. Objetivos y resultados del curso	146
III. Contenido del curso	147
1. Introducción a la proteómica	148
2. Enfoques analíticos en la proteómica	149
2.1. Preparación de muestras	149
2.2. Identificación de proteínas a partir de MS: conceptos	150
2.3. Enfoques de la espectrometría de masas para la identificación de proteínas	152
2.4. Proteómica cuantitativa	153
2.5. Identificación de PTM	156
2.6. Detección y cuantificación de la localización de proteínas subcelulares	158
2.7. Detección y cuantificación de interacciones proteicas	159
3. Procesamiento de datos e identificación de proteínas	160
3.1. Bases de datos de secuencias de proteínas	160
3.2. Motores de búsqueda de bases de datos de proteínas	161
3.3. Análisis de datos	161
4. Sesiones prácticas	162
4.1. Identificación de proeínas empleando los enfoques PMF y PFF con MASCOT	162
4.2. Identificación de proteínas a partir de un experimento de LC-MS	166
4.2.1. Transformación de datos a mzML con MSConverter en ProteoWizard	167
4.2.2. Generación de la base de datos FASTA a partir de Uniprot (SwissProt)	168
4.2.3. Motores de búsqueda: SearchGUI	169
4.2.4. Generación y evaluación de resultados: PeptideShaker para la visualización de péptidos y proteínas, y validación. Análisis PTM	170
4.2.5. Análisis de datos: información proteica, análisis de rutas metabólicas y ontología de genes	171

## Prólogo: Plan de estudios de AACLifeSci

Pedro Domingues<sup>1</sup>, Antonia García<sup>2</sup>, Elzbieta Skrzydlewska<sup>3</sup>

<sup>1</sup> *Unidad de Espectrometría de Masas, Departamento de Química, Universidad de Aveiro, 3810-193 Aveiro, Portugal, CESAM (Centro de Estudos do Ambiente e do Mar)*

<sup>2</sup> *Centro de Metabolómica y Bioanálisis (CEMBIO), Facultad de Farmacia, Universidad San Pablo CEU; Boadilla del Monte, 28668 Madrid, España*

<sup>3</sup> *Departamento de Química Analítica, Universidad Médica de Białystok, 15-089 Białystok, Polonia*

### I. Justificación

Considerando el crecimiento exponencial que ha experimentado la investigación genómica durante la presente década, cabría esperar que se desarrollase con rapidez la investigación de las ómicas basadas en la espectrometría de masas. Sin embargo, a pesar de haberse producido un desarrollo rápido de la investigación proteómica, la implementación de otras ómicas, como la lipidómica y la metabolómica, no ha tenido todavía un alcance amplio. Este desarrollo diferenciado se debe a distintos factores, como por ejemplo, el desarrollo insuficiente de las herramientas bioinformáticas, o la falta de profesionales cualificados en este campo. Es en este contexto donde se desarrolla el curso AACLifeSci, con el objetivo de facilitar una formación avanzada a estudiantes de doctorado en las áreas de ciencias biomoleculares y de la salud, en técnicas cromatográficas acopladas a la espectrometría de masas, así como en la lipidómica, la metabolómica y la proteómica.

A día de hoy, se considera que la principal causa del incremento de la morbilidad y la mortalidad en Europa es el rápido desarrollo de las enfermedades propias de la civilización. En este contexto, uno de los remedios posibles apunta a un diseño cualitativo y cuantitativo de la investigación que permita comprender e identificar biomarcadores y descubrir nuevos métodos terapéuticos. Las ómicas, un campo emergente en el estudio de la biología que engloba otras disciplinas, como la genómica, la proteómica, la lipidómica o la metabolómica, son una herramienta esencial que se emplea actualmente para explicar la base bioquímica del funcionamiento de un organismo mediante la caracterización y cuantificación masiva de moléculas biológicas. Toda vez que los compuestos simples no son biomarcadores efectivos, las ómicas ofrecen nuevas oportunidades de alcanzar una comprensión más pormenorizada de las enfermedades. Esta comprensión mejora la eficacia, la monitorización, y conduce a una terapia personalizada.

Existe un desequilibrio en la enseñanza y el aprendizaje de las ómicas en las universidades europeas. Las materias de las ómicas se imparten normalmente a los alumnos a través de unidades curriculares que se pueden designar colectivamente como Química Analítica Avanzada (AAC, por sus siglas en inglés), y que constituyen la base fundamental de los cursos de doctorado en el campo de la bioquímica. Con todo, el número de centros y equipos docentes especializados con

acceso a los equipos necesarios dedicados a ofrecer un enfoque integral es limitado. Además, existe una carencia de docentes con titulaciones en AAC, de materiales didácticos o de un sistema que permita el intercambio de experiencias a todos los niveles. Por tanto, solo unos pocos alumnos de doctorado pueden obtener una formación adecuada en AAC, análisis instrumental o procesado de datos, o adquirir las competencias necesarias para integrar la AAC con las necesidades de los servicios sanitarios, los laboratorios de diagnóstico de los hospitales, las empresas o el mercado laboral.

Química Analítica Avanzada en Ciencias de La Vida (AACLifeSci, por sus siglas en inglés) fue un proyecto educativo y científico coordinado por la doctora Elżbieta Skrzydlewska, decana de la Facultad de Farmacia con el Departamento de Medicina de Laboratorio de la Universidad Médica de Białystok (MUB), financiado por el programa Erasmus+ e implementado mediante la colaboración de tres universidades:

- Universidad Médica de Białystok, Polonia, Facultad de Farmacia con la Sección de Medicina de Laboratorio, Departamento de Química Analítica (Doctora Elżbieta Skrzydlewska).
- Universidad de Aveiro, Portugal, Departamento de Química, Unidad de Espectrometría de Masas (Doctor Pedro Domingues).
- Universidad San Pablo CEU, España, Centro de Metabolómica y Bioanálisis (CEMBIO), Facultad de Farmacia (Doctora Antonia García).

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

La finalidad del proyecto fue adaptar los programas de estudio de las tres universidades colaboradoras a las necesidades de las industrias vinculadas a la biología molecular y los servicios sanitarios en el uso de técnicas de Química Analítica Avanzada relacionadas con las ómicas mediante el intercambio de experiencias a nivel europeo. El proyecto incluyó la preparación del equipo docente de las tres universidades colaboradoras para la formación de estudiantes de doctorado en el uso de técnicas de Química Analítica Avanzada en la metabolómica, la lipidómica y la proteómica, mediante la introducción de modificaciones en los programas de estudio y mediante la creación de recursos educativos, entre los que se incluye un libro complementario para el curso y materiales electrónicos para los alumnos. Las actividades del proyecto comprendían actividades de aprendizaje, formación y docencia basadas en la colaboración e incluían un intercambio de experiencias y buenas prácticas, creando una red de cooperación estable y activa, y el desarrollo personal y social de los participantes del proyecto.

La totalidad del curso se organiza en cuatro módulos, lo que permite un aprendizaje estructurado de los contenidos y facilita la organización de seminarios y clases prácticas. El curso cubrirá, dentro del primer módulo, los aspectos principales de la espectrometría de masas molecular, incluyendo las innovaciones más recientes en el diseño de equipos, con sus ventajas y desventajas, así como la comprensión del análisis de los espectros de masas para la elucidación estructural. También se tratarán los fundamentos de las técnicas de separación de alta eficacia acopladas a la espectrometría de masas, en particular los mecanismos de separación y la instrumentación. Este módulo proporcionará a los alumnos la base analítica para que puedan continuar desarrollando

su conocimiento de las disciplinas ómicas espectrometría de masas (MS). Los módulos siguientes cubrirán los fundamentos de los estudios metabolómicos, proteómicos y lipidómicos. Estos módulos presentarán los enfoques analíticos, métodos de procesado y análisis de datos y procesos para cada una de las metodologías de las ómicas. Se describirán también algunos de los recursos bioinformáticos que se emplean actualmente y los alumnos tendrán la oportunidad de probarlos en clases prácticas.

## II. Objetivos y resultados del curso

### *Objetivos*

El objetivo del curso es proporcionar a los estudiantes competencias teóricas y prácticas en las áreas de la química, la instrumentación bioquímica y la bioinformática aplicables a la cromatografía y las técnicas de separación, así como a la espectrometría de masas, para las ciencias de la vida, la metabolómica, la lipidómica y la proteómica. Los estudiantes adquieren experiencia en las distintas áreas de las ómicas y también cómo se aplican en la vida real a la investigación científica. Se ofrece información teórica introductoria sobre los objetivos y métodos de la cromatografía y la espectrometría de masas, así como sobre cada una de las disciplinas ómicas, y todo ello será objeto de un análisis pormenorizado. Se proporciona una explicación sobre el modo en que se aplican estos métodos en la ciencia, sus ventajas y limitaciones prácticas, así como los problemas y las cuestiones científicas que abordan. Los conocimientos prácticos que se adquirirán se centrarán principalmente en el uso de bases de datos de libre acceso y herramientas de software gratuito para el procesado y análisis de datos cuando se empleen estos enfoques.

Al finalizar este curso, los alumnos deberían ser capaces de reconocer los problemas científicos de la era postgenómica y saber cómo obtener y cómo realizar un análisis simple de los datos de espectrometría de masas y ómicas. La comprensión de los principios, métodos y usos de las técnicas ómicas más recientes contribuirá de manera significativa al futuro del alumno tanto el campo académico como industrial.

### *Resultados específicos del aprendizaje:*

Tras terminar el curso, el alumno será capaz de:

1. realizar una revisión crítica de los analizadores de masas y métodos de ionización disponibles, conociendo sus ventajas y desventajas, en particular los análisis críticos de espectrometría de masas en tándem.
2. reconocer modelos de fragmentación de moléculas orgánicas, principalmente de metabolitos pequeños, péptidos y lípidos.
3. discutir ampliamente los diferentes mecanismos de separación acoplados a MS en metabolómica, proteómica y lipidómica.

4. conocer y entender los diferentes métodos de tratamiento de muestras y sus limitaciones para cada una de las aplicaciones ómicas.
5. identificar todas las etapas de un estudio ómico y sus diferentes enfoques en la metabolómica, la lipidómica y la proteómica.
6. discutir el uso de software público en el reprocesado de datos y análisis de rutas metabólicas.
7. adquirir competencias básicas en el uso de bases de datos, al igual que otros recursos de acceso gratuito.
8. explicar a personas no especialistas la forma en que se puede esperar que las disciplinas «ómicas» proporcionen información de valor en las diferentes áreas de las Ciencias de la Vida.

### III. Formato y procedimientos

El presente curso se organiza en cuatro módulos independientes estrechamente relacionados:

Modulo 1 - Técnicas de separación/cromatográficas y Espectrometría de Masas para las Ciencias de la Vida;

Modulo 2 - Metabolómica;

Modulo 3 - Lipidómica;

Modulo 4 - Proteómica.

Cada uno de estos módulos examina los aspectos teóricos y técnicos principales que resultan esenciales para una comprensión avanzada de los enfoques ómicos para la resolución de problemas de las ciencias en la vida real. Se considera que el conocimiento y las competencias que se adquieren durante este curso se corresponden con un nivel de conocimiento equivalente al doctorado.

El curso se organiza en talleres. Cada uno de los módulos comprenderá un taller de, por lo menos, una hora de duración y dos horas de clases prácticas tuteladas en las que los alumnos pueden adquirir un conocimiento práctico y competencias en aplicaciones que sean esenciales para cada módulo. Cada módulo es independiente. No obstante, el conocimiento de las materias del Módulo 1- Cromatografía/Técnicas de separación y Espectrometría de Masas para las Ciencias de la Vida es esencial para la comprensión de los módulos siguientes.

### Organización del programa de doctorado de la Universidad Médica de Białystok y cómo se implementará AACLifeSci

La Facultad de Farmacia con la Sección de Medicina de Laboratorio de la Universidad de Białystok ofrece dos cursos distintos de doctorado: uno en el área de ciencias médicas, dentro del campo de la biología médica, y otro en el área de las ciencias farmacéuticas.

El programa de doctorado en **Ciencias Médicas** se organiza en cuatro años. En el segundo año se incluye el curso de Técnicas analíticas modernas en ciencias biomédicas, que se ha ampliado de 15 a 30 horas en base a los contenidos de AACLifeSci. Este curso se corresponde con 3 ECTS.

El presente curso pretende proporcionar a los alumnos estrategias y métodos modernos y especializados de análisis bioquímico utilizados principalmente en los laboratorios de investigación.

Tras terminar el curso, el alumno deberá:

- haber adquirido un conocimiento de las técnicas de preparación de muestras empleadas en la metabolómica, la lipidómica y la proteómica;
- conocer y ser capaz de describir las técnicas de separación actuales, incluidas las técnicas cromatográficas y de electromigración;
- conocer la metodología de investigación empleada en el análisis clásico y ómico;
- demostrar un conocimiento de conceptos y usos prácticos de la evaluación bioestadística de los resultados de investigación;
- ser capaz de proponer una técnica analítica para solución de un problema científico concreto desde el campo de las ciencias biomédicas;
- ser capaz de realizar un análisis metabolómico/lipidómico/proteómico;
- ser capaz de interpretar los resultados obtenidos a través de métodos estadísticos;
- ser capaz de aplicar y extraer conclusiones de la investigación para la resolución de problemas.

El programa de doctorado en **Ciencias Farmacéuticas** se organiza también en cuatro años. En el segundo año se incluye el curso *Técnicas analíticas avanzadas* en investigación ómica, que se imparte en 15 horas en base a los contenidos de AACLifeSci. Este curso se corresponde con 1 ECTS. No obstante, este curso viene precedido del titulado *Elementos de análisis farmacéutico moderno*, donde se incluye el conocimiento de las técnicas cromatográficas/de separación y el análisis de espectrometría de masas y objetivo. Por tanto, en el segundo curso se incluye información que se refiere únicamente al análisis ómico. Se dedican un total de 30 horas para AACLifeSci que se corresponden con 4 ECTS.

MUB / Ciencias Médicas /	Horas totales	Espectrometría de masas y técnicas de separación	Metabolómica	Lipidómica	Proteómica
		Técnicas analíticas modernas en ciencias biomédicas	Técnicas analíticas avanzadas en investigación ómica		
Seminarios y talleres	6h	2h	1h	1h	2h
Clases prácticas	24h	10h	4h	5h	5h
Aprendizaje autónomo del alumno	57h	15h	10h	14h	18h
Esfuerzo total del alumno	87h	27h	15h	20h	25h

El presente curso pretende proporcionar a los alumnos estrategias y métodos de análisis ómico utilizados en los laboratorios de investigación.

Tras terminar el curso, el alumno deberá:

- conocer los métodos de preparación de muestras biológicas de análisis analíticos y ómicos;
- conocer la metodología de investigación empleada en el análisis clásico y ómico;

- demostrar un conocimiento de conceptos y usos prácticos de la evaluación bioestadística de los resultados de investigación;
- ser capaz de realizar un análisis metabolómico/lipidómico/proteómico;
- ser capaz de interpretar los resultados obtenidos a través de métodos estadísticos;
- ser capaz de aplicar y extraer conclusiones de la investigación para la resolución de problemas.

MUB /Ciencias Farmacéuticas/	Horas totales	Espectrometría de masas y técnicas de separación	Metabolómica	Lipidómica	Proteómica
		Elementos del análisis farmacéutico moderno	Técnicas analíticas avanzadas en investigación ómica		
Seminarios y talleres	8h	5h	1h	1h	1h
Clases prácticas	22h	10h	4h	4h	4h
Aprendizaje autónomo del alumno	30h	14h	3h	5h	6h
Esfuerzo total del alumno	58h	29h	8h	10h	11h

## Organización del programa de doctorado de la Universidad San Pablo-CEU y cómo se implementará

La Facultad de Farmacia de la Universidad San Pablo-CEU ofrece un máster denominado DESCUBRIMIENTO DE FÁRMACOS, para alumnos de posgrado de Química, Farmacia y otras Ciencias de la Salud. El máster interuniversitario nace de la colaboración entre tres universidades: Universidad Complutense, Universidad CEU San Pablo y Universidad de Alcalá, todas ellas ubicadas en la Comunidad Autónoma de Madrid. Estas universidades han estado colaborando en un programa de doctorado denominado QUÍMICA MÉDICA, que contó con mención de calidad y con Mención a la Excelencia del Ministerio. Durante el curso 2016-17 se impartió la cuarta edición de este máster.

El programa máster cubre un año académico de 60 ECTS. En el segundo semestre se incluye la materia optativa: TÉCNICAS BIOANALÍTICAS AVANZADAS, que comprende 3 ECTS con 30 horas de asistencia presencial dentro del campus.

El objetivo del curso es proporcionar al alumno conocimiento y competencias sobre las técnicas modernas de bioanálisis, basadas principalmente en técnicas de separación de alta resolución acopladas a la espectrometría de masas. Se ofrece además conocimiento sobre los métodos de purificación y preconcentración. Comprende aprendizaje teórico y práctico trabajando con muestras reales en un laboratorio moderno.



Tras terminar el curso, el alumno deberá:

- ser capaz de escoger y aplicar los métodos analíticos apropiados, considerar los fundamentos, la instrumentación, el alcance y las aplicaciones de los métodos utilizados actualmente en la caracterización y análisis de muestras farmacéuticas y biológicas.
- ser capaz de escoger y aplicar el método analítico apropiado para el análisis de ingredientes activos y sus metabolitos en muestras biológicas, así como el análisis de ingredientes activos en concentraciones muy bajas en el producto farmacéutico final.
- identificar todas las etapas del estudio y sus diferentes enfoques en la metabolómica

CEU /DESCUBRIMIENTO DEFÁRMACOS/	Horas totales	Espectrometría de masas y acoplamientos	Cromatografía y CE en bioanálisis	Metabolómica	Purificación y concentración de muestras
<b>Técnicas bioanalíticas avanzadas</b>					
Seminarios y talleres	24h	11h	8h	3h	2h
Clases prácticas	6h	2h	2h		2h
Aprendizaje autónomo del alumno	60h				
Esfuerzo total del alumno	90h				
ECTS	3				

La materia TÉCNICAS BIOANALÍTICAS AVANZADAS se impartirá en función de los contenidos de AACLifeSci en el segundo semestre del curso académico y cubrirá los mismos 3 ECTS.

Las clases se organizarán del siguiente modo:

CEU	Horas totales	(1) MS y (2) Técnicas de separación	Metabolómica	Lipidómica	Proteómica
Seminarios y talleres	22h	6h LC: 2h GC: 1h CE: 1h	4h	4h	4h
Clases prácticas	8h	2h	2h	2h	2h
Aprendizaje autónomo del alumno	60h	30h	10h	10h	10h
Esfuerzo total del alumno	90h	42h	16h	16h	16h
ECTS	3	1,4	0,53	0,53	0,53

## Organización del programa de doctorado en Bioquímica de la UA y cómo se implementará AACLifeSci

El programa de doctorado en Bioquímica de la Universidad de Aveiro se organiza en cuatro años. El primer año incluye siete unidades curriculares, cuatro de ellas opcionales. Dos de estas unidades curriculares opcionales (cursos avanzados de Bioquímica I y cursos avanzados de Bioquímica II) se basarán en los contenidos de AACLifeSci. Estas unidades curriculares se componen de tres módulos opcionales, cada uno de ellos equivalente a 3 ECTS. Para completar cada una de estas unidades curriculares, los alumnos deben completar dos de estos módulos. Se realizará una evaluación de cada módulo y se obtendrá la calificación final a partir de una media ponderada de la evaluación obtenida en dos módulos. Los cursos y los respectivos módulos son:

### Cursos avanzados en Bioquímica I:

El presente curso pretende proporcionar a los alumnos competencias avanzadas en estrategias y métodos de análisis bioquímico modernos y especializados utilizados en laboratorios de investigación y clínicos y en las industrias. Tras terminar el curso, el alumno deberá:

- Comprender los fundamentos y las aplicaciones de las técnicas de cromatografía acoplada a espectrometría de masas.
- Ser capaz de evaluar de manera crítica los métodos utilizados en los estudios lipidómicos y metabolómicos.
- Ser capaz de planificar experimentos lipidómicos y metabolómicos.
- Ser capaz de analizar datos lipidómicos y metabolómicos.

UA	Horas totales	Cromatografía y Espectrometría de Masas para las Ciencias de la Vida	Lipidómica	Metabolómica
Seminarios y talleres	20h	10h	10h	10h
Clases prácticas		5h	5h	5h
Aprendizaje autónomo del alumno	150h	75h	75h	75h
Esfuerzo total del alumno	180h	90h	90h	90h
ECTS	6	3	3	3

### Cursos avanzados en Bioquímica II:

El presente curso pretende proporcionar a los alumnos competencias avanzadas en estrategias y métodos modernos y especializados de análisis bioquímico en el campo de las ómicas, utilizados para la proteómica, la glicómica y la bioinformática.

Tras terminar el curso, el alumno deberá ser capaz de:

- Evaluar de manera crítica qué métodos deberían emplearse para resolver problemas concretos de la proteómica y la glicómica.
- Planificar experimentos proteómicos y glicómicos.
- Analizar los datos adquiridos en estudios proteómicos y glicómicos.
- Saber cómo seleccionar y cómo usar herramientas y algoritmos adecuados en estas aplicaciones bioquímicas.

UA	Horas totales	Proteómica	Glicómica	Bioinformática
Lectures and workshops	20h	10h	10h	10h
Practical Lessons		5h	5h	5h
Student centred learning	150h	75h	75h	75h
Total student effort	180h	90h	90h	90h
ECTS	6	3	3	3

En líneas generales, AACLifeSci estará totalmente integrada en cuatro módulos del programa de doctorado en Bioquímica de la Universidad de Aveiro, cada uno de ellos equivalente a 3 ECTS, constituyendo un total de 12 ECTS.

Módulo 1- Técnicas cromatográficas/de separación y Espectrometría de Masas para las Ciencias de la Vida (3 ECTS)

Módulo 2- Metabolómica (3 ECTS)

Módulo 3- Lipidómica (3 ECTS)

Módulo 4- Proteómica (3 ECTS)

#### IV. Lectura de recursos adicionales

Módulo 1

1. H.J. Hübschmann. Handbook of GC-MS: Fundamentals and Applications (3rd Edition). Somerset, NJ, USA: Wiley, 2015. ProQuest ebrary. Web. 8, 2015. . Somerset, NJ, EE. UU.: Wiley, 2015. ProQuest ebrary. Web. 8, 2015.
2. L.R. Snyder, J.J. Kirkland, J.W. Dolan. Introduction to modern liquid chromatography. Wiley, 2011.
3. E. De Hoffmann, V. Stroobant. Mass spectrometry: principles and applications. Wiley, 2007.
4. D.R. Baker. Capillary electrophoresis. Wiley, 1995.
5. [https://www.agilent.com/cs/library/usermanuals/public/G7100-90001\\_CEMSAAnalysis\\_ebook.pdf](https://www.agilent.com/cs/library/usermanuals/public/G7100-90001_CEMSAAnalysis_ebook.pdf)

#### Módulo 2

1. A. García, J. Godzien, Á. López-González, C. Barbas. Capillary electrophoresis mass spectrometry as a tool for untargeted metabolomics. *Bioanalysis*, 2017, 9(1):99-130.

#### Módulo 3

1. L. Feng, G.D. Prestwich (ed.) *Functional Lipidomics*. Taylor & Francis, Boca Raton, 2006.
2. D. Armstrong. *Lipidomics: Volumes 1e 2: Methods and Protocols (Methods in Molecular Biology)*. Humana Press, 2009, 1.
3. U. Loizides-Mangold. On the future of mass-spectrometry-based lipidomics. *FEBS J.*, 2013, 280(12):2817-2.
4. Harkewicz, E.A. Dennis. Applications of mass spectrometry to lipids and membranes. *Ann. Rev. Biochem.*, 2011, 80:301-25.

#### Módulo 4

1. J. Lovric. *Introducing Proteomics: From concepts to sample separation, mass spectrometry and data analysis*, Wiley. ISBN: 978-0-470-03523-8.
2. R. Aebersold, M. Mann. Mass-spectrometric exploration of proteome structure and function, *Nat. Rev. Mol. Cell. Biol.*, 2015, 16(5):269-80.
3. A.M. Silva, R. Vitorino, M.R. Domingues, C.M. Spickett, P. Domingues. Post-translational Modifications and Mass Spectrometry Detection, *Free Radic. Biol. Med.*, 2013, 65:925-41.

# Módulo 1- Técnicas de separación/Cromatografía y Espectrometría de Masas en las Ciencias de la Vida

Wojciech Luczaj<sup>1</sup>, Antonia García<sup>2</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*

<sup>2</sup>*Centro de Metabolómica y Bioanálisis (CEMBIO), Facultad de Farmacia, Universidad San Pablo CEU; Boadilla del Monte, 28668 Madrid, España*

## I. Justificación

El primer módulo del curso AACLifeSci se refiere a los aspectos principales de la espectrometría de masas molecular. Incluye una perspectiva general de las innovaciones más recientes en el diseño de equipos y sus ventajas, junto con la comprensión del análisis de los espectros de masas para la elucidación estructural. También se tratarán los fundamentos de las técnicas de separación de alta eficacia acopladas a la espectrometría de masas, en particular los mecanismos de separación y la instrumentación. El objetivo de este módulo es proporcionar a los alumnos las herramientas analíticas esenciales para aumentar su conocimiento de las ómicas vinculadas a la espectrometría de masas.

## II. Objetivos y resultados del curso

### **Objetivos**

El objetivo del módulo es proporcionar a los estudiantes competencias teóricas y prácticas en las áreas de la química, la instrumentación bioquímica y la bioinformática aplicables a la cromatografía y espectrometría de masas y que se aplican a las ciencias de la vida. Se facilitará una descripción teórica introductoria sobre el modo en que se aplican estos métodos en la ciencia, particularmente en relación con la investigación ómica, sus ventajas y limitaciones, así como los problemas y las cuestiones científicas que abordan. Los conocimientos prácticos que se adquirirán se centrarán principalmente en el uso de bases de datos de libre acceso y herramientas de software gratuito dentro de estos enfoques.

### **Resultados específicos del aprendizaje:**

Tras terminar el curso, el alumno deberá ser capaz de:

1. describir los principios de ionización, separación y detección de moléculas mediante espectrometría de masas;
2. describir los principios de la espectrometría de masas en tándem, la instrumentación y su aplicación;

3. identificar moléculas pequeñas empleando datos de espectrometría de masas;
4. describir los principios de separación, optimización y detección de compuestos a través de métodos de cromatografía líquida, cromatografía de gases y electroforesis capilar;
5. predecir de qué manera los cambios en las condiciones experimentales pueden influir en la separación con estos métodos;
6. describir el principio y el diseño de los componentes principales de instrumentación analítica de separación y espectrometría de masas;
7. escoger la técnica más apropiada, tanto para los objetivos de separación como de detección;
8. seleccionar la instrumentación apropiada para cada aplicación «ómica», justificándolo en función del rendimiento y de los requisitos que se exijan.

### III. Contenido del curso

#### Módulo 1 – Técnicas de separación/Cromatografía y Espectrometría de Masas para las Ciencias de la Vida

1. Espectrometría de masas
  - 1.1. Introducción a la espectrometría de masas
  - 1.2. Técnicas de ionización en espectrometría de masas biológicas
  - 1.3. Analizadores de masas en espectrometría de masas
  - 1.4. Espectrometría de masas en tándem
  - 1.5. Interpretación de los espectros de masas
2. Técnicas cromatográficas acopladas a la espectrometría de masas
  - 2.1. Cromatografía líquida-espectrometría de masas (LC-MS)
    - 2.1.1. Introducción a la LC-MS
    - 2.1.2. Principios teóricos de la cromatografía líquida
      - 2.1.2.1. El proceso cromatográfico
      - 2.1.2.2. Cromatografía de fase normal (NP)
      - 2.1.2.3. Cromatografía de fase inversa (RP)
      - 2.1.2.4. Cromatografía de interacción hidrofílica (HILIC)
      - 2.1.2.5. Cromatografía de intercambio de iones (IEC).
    - 2.1.3. Instrumentación en LC-MS
  - 2.2. Cromatografía de gases-espectrometría de masas (GC-MS)
    - 2.2.1. Introducción a la GC-MS
    - 2.2.2. Principios teóricos de la GC-MS
      - 2.2.2.1. El proceso cromatográfico
      - 2.2.2.2. Fases de columna
      - 2.2.2.3. GC×GC/MS
      - 2.2.2.4. Pretratamiento y derivación de muestras
    - 2.2.3. Acoplamiento GC-MS

- 2.3. Electroforesis capilar-espectrometría de masas (CE-MS)
  - 2.3.1. Introducción a la CE-MS
  - 2.3.2. Principios teóricos de la CE-MS
    - 2.3.2.1. Electroforesis capilar zonal
    - 2.3.2.2. Cromatografía electrocinética micelar
    - 2.3.2.3. Electro cromatografía capilar
    - 2.3.2.4. Electroforesis capilar en gel
  - 2.3.3. Instrumentación
- 3. Sesiones prácticas
  - 3.1. Adquisición de un espectro de masas ESI-MS
  - 3.2. Adquisición de un espectro de masas ESI-MS/MS
  - 3.3. Interpretación de los espectros de masas
  - 3.4. Observación de un análisis por GC-MS y configuración de la adquisición por monitorización selectiva de iones
  - 3.5. Observación de un análisis por HPLC-MS y configuración de la adquisición por MS/MS dependiente de datos

## 1. Espectrometría de masas

### 1.1. Introducción a la espectrometría de masas

Desde los años sesenta, la espectrometría de masas MS se ha convertido en una herramienta analítica estándar para el análisis de compuestos orgánicos, a pesar de que, en un principio, las aplicaciones a las áreas de la biología fueron escasas. En los quince últimos años los espectrómetros de masas se han vuelto mucho más fáciles de usar, con un avance enorme en la instrumentación, y hoy representa una tecnología analítica fundamental en el estudio de las proteínas, los lípidos y los metabolitos. La función actual de la MS en la bioquímica y en las ciencias de la vida deriva del desarrollo de las nuevas técnicas de ionización, la mejora de los analizadores, los métodos de fragmentación, así como del desarrollo notable de las técnicas y los métodos de separación. En este curso se revisarán las características más importantes de la MS moderna y su acoplamiento con las distintas técnicas de separación.

La MS es una técnica analítica en la que interviene el estudio de iones en fase gaseosa a partir del analito, el cual se separa de acuerdo con su relación masa/carga ( $m/z$ ) y su abundancia. Los componentes principales de un espectrómetro de masas son la fuente de ionización, el analizador de masas y el detector. El analizador y el detector se encuentran siempre en condiciones de alto vacío que se obtienen empleando varias bombas (neumáticas y turbobombas) para no influir en el patrón de fragmentación de las moléculas y no incrementar el ruido de fondo.

En pocas palabras, los flujos de trabajo de la MS consisten en la ionización de la muestra de analitos en una fuente de iones, la separación de moléculas ionizadas y sus fragmentos ionizados de acuerdo con su relación masa/carga ( $m/z$ ) con un analizador, la detección de los iones, y el análisis e interpretación del espectro de masas obtenido. Los espectrómetros de masas en tándem poseen la capacidad adicional de seleccionar iones e inducir su fragmentación para obtener información estructural detallada de las especies seleccionadas (MS/MS).

## 1.2. Técnicas de ionización en espectrometría de masas biológica

Existen varias fuentes de ionización que permiten el análisis de compuestos orgánicos. La fuente tradicional, utilizada para el análisis de pequeñas moléculas volátiles, es el Impacto Electrónico o Electro-ionización (EI). Esta fuente requiere que el analito sea volátil y termoestable. Actualmente se emplea particularmente en la combinación con la cromatografía de gases (GC). En esta fuente, las moléculas gaseosas colisionan con electrones energéticos, lo que produce su ionización y una mayor fragmentación. El patrón de fragmentación característico es muy reproducible, y se utiliza con frecuencia con fines de identificación a través de bases de datos espectrales. En la ionización EI, como consecuencia de la retirada de un solo electrón, se genera un ion molecular, catión radical ( $M^+$ ), con una relación  $m/z$  que se corresponde con la masa nominal (masa entera) del compuesto (Figura 1.1).

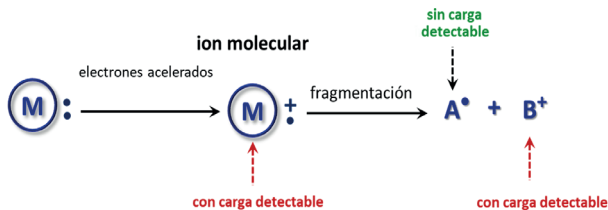


Figura 1.1. Ionización con electrones (EI) y fragmentación subsiguiente.

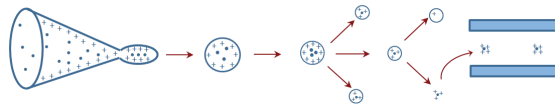
El principal inconveniente de la fuente de EI es que es necesario que el analito esté en estado gaseoso antes de la ionización y, por tanto, se excluyen los compuestos polares, no volátiles y térmicamente lábiles a menos que se puedan derivatizar para modificar sus propiedades. Otro problema es la ausencia del ion molecular en el espectro de algunos compuestos.

La espectrometría de masas se ha vuelto muy accesible para los compuestos no volátiles y térmicamente inestables tras el desarrollo de nuevos sistemas de ionización a finales de la década de los ochenta. En esa época se desarrollaron la ionización por electrospray (ESI) y la desorción/ionización láser asistida por matriz (MALDI). Ambas técnicas de ionización son técnicas «suaves»



con aplicaciones universales en el análisis de moléculas biológicas no volátiles, térmicamente inestables e, incluso, de masa elevada. A día de hoy, ambos, ESI y MALDI, son las técnicas de ionización con un uso más extendido para los análisis LC-MS de moléculas biológicas como péptidos, proteínas o lípidos. Son técnicas de ionización suaves, puesto que principalmente producen iones con muy poca fragmentación. Las dos son fuentes de desorción, pudiendo funcionar con fases líquidas o sólidas, y, normalmente, no requieren de derivatización química como los análisis por GC-MS.

El primero en utilizar la técnica de ionización por electrospray fue John Fenn en 1984. En el electrospray, un líquido que contiene el analito fluye a través de un capilar metálico al que se le aplica un alto voltaje para producir iones, tabajando a presión atmosférica. Se forma un spray de gotitas cargadas y, a continuación, se someten a desolvatación, normalmente a través de una corriente de  $N_2$ . Las gotas se vuelven más pequeñas produciéndose un fenómeno denominado «fisión de Coulomb» cuando la tensión superficial del líquido se ve superada por la repulsión entre cargas de la superficie de las gotitas. Cuando las gotitas son lo suficientemente pequeñas se generan iones de analito en estado gaseoso, normalmente a través de un proceso de evaporación iónica (Figura 1.2).



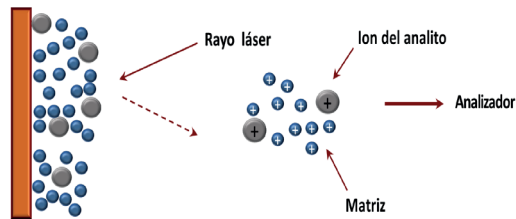
**Figura 1.2.** El mecanismo de ionización por electrospray (ESI).

En la ESI, la ionización se produce directamente a partir de una disolución, por lo que moléculas térmicamente lábiles se pueden ionizar sin degradación. La respuesta del analito depende de la concentración, por lo que es posible trabajar en LC con flujos bajos (entre 5 y 100  $\mu\text{L min}^{-1}$ ), o incluso en un flujo de sub-microlitro en fuentes de nano electrospray, y utilizando una columna muy estrecha. Los disolventes típicos son el agua, el metanol y el acetonitrilo.

Por lo general se utilizan disolventes con tensión y viscosidad superficiales bajas y de alta conductividad. Es posible añadir modificadores para promover la ionización del analito en la disolución. Para un modo de ionización positivo se suele utilizar un 0,1% de ácido fórmico o acético, y en modo negativo formiato o acetato de amonio, toda vez que el agua pura o los disolventes orgánicos no polares resultan menos efectivos.

El MALDI (Matrix Assisted Laser Desorption Ionization –ionización por desorción láser asistida por matriz–) fue desarrollada por Karas y Hillenkamp, y por Tanaka, en 1988. Se trata de un método de ionización suave que se emplea principalmente con biomoléculas como proteínas, péptidos, oligonucleótidos, polisacáridos y polímeros sintéticos. En este método, la ionización se obtiene mezclando la disolución de la muestra con un exceso molar grande de un material matriz. Seguidamente se evapora el disolvente y se irradian cristales de la matriz muestra con un rayo láser de

pulso corto. La matriz puede absorber una alta cantidad de energía de la radiación láser y transferirla a las moléculas de la muestra que son desorbidos quedando como iones intactos en fase gaseosa. El mecanismo de ionización conlleva la desorción e ionización del analito mediante una reacción de transferencia de protones con moléculas de la matriz (Figura 1.3).



**Figura 1.3.** El mecanismo de la técnica de desorción/ionización láser asistida por matriz (MALDI).

Las matrices que habitualmente se utilizan para el MALDI son compuestos orgánicos pequeños como el ácido 2,5-dihidroxibenzoico (DHB), el nicotínico, el sinapínico o el 3,5-dimetoxi-4-hidroxicinámico. La selección de la matriz depende de las propiedades del analito y el láser, que normalmente se trata de un láser de nitrógeno (337 nm). Se debe hacer hincapié en la importancia del tipo de matriz. Por ejemplo, el DHB es apto para mezclas de lípidos analizadas en modo positivo y negativo, en tanto que la p-nitroanilina permite una detección más sensible de fosfatidiletanolaminas en modo de ionización negativo.

Los métodos de ionización MALDI y ESI ofrecen ventajas comunes, entre las que destaca principalmente la alta sensibilidad, por lo que normalmente tan solo se necesitan unos pocos picomoles del analito por análisis. Por otro lado, debido a la sencillez de su funcionamiento, estas técnicas se utilizan ampliamente para obtener un examen rápido de una muestra. Cuando se utiliza el MALDI existe un importante inconveniente, y es que, normalmente, los compuestos de la matriz están ionizados, lo que genera una señal de fondo alta, dificultando el análisis de la región baja de  $m/z$  (por debajo de 800 Da).

La ionización química a presión atmosférica (APCI) es otra técnica de ionización suave que se puede utilizar para la ionización de analitos no volátiles neutros. El APCI es un método de ionización vinculado a la ESI, donde el flujo de líquido se dispersa en pequeñas gotitas, tal y como se explicó antes. En el caso del APCI, interviene en este proceso un calefactor y un potencial alto mediante un electrodo de descarga corona que ioniza el gas nebulizador formando un plasma. A continuación, las moléculas de la muestra se ionizan dentro del plasma mediante procesos de transferencia de protones a través del eluyente. En general, el APCI se adapta mejor a los compuestos lipofílicos y se puede combinar satisfactoriamente con la cromatografía líquida de alto rendimiento (HPLC) utilizando columnas analíticas cuando el ESI no ofrezca una ionización adecuada. En la Tabla 1.1 figuran las características principales y se comparan distintas técnicas de ionización.

**Tabla 1.1.** Caracterización de las distintas técnicas de ionización.

Técnica de ionización	Analitos habituales	Rango de masas	Introducción de la muestra	Ventajas	Inconvenientes
EI	Compuestos relativamente pequeños, no polares, volátiles y termoestables	<1 kDa	GC o líquido/sólido	Analitos no polares no hay supresión de iones, fácil acoplamiento con GC, librerías de espectros	Método duro, necesita muestras volátiles, necesita estabilidad térmica, compuestos de bajo peso molecular
ESI	Péptidos, proteínas, azúcares, nucleótidos de compuestos polares	<200 kDa	LC o disolución	Compuestos termolábiles, compuestos de alto PM, sensibilidad, fácil interacción con LC, método de ionización suave iones multicargados	Analitos ionizables, sensible a sales, supresión de iones
MALDI	Péptidos, proteínas, azúcares, nucleótidos de compuestos polares	<500 kDa	Muestra mezclada con una matriz sólida	Compuestos termolábiles, compuestos de alto PM, sensibilidad, menos sensible a sales, método de ionización suave	Amplia variedad de matrices, dificultades en el análisis cuantitativo, supresión de iones
APCI	Compuestos neutros	<1 kDa	LC o disolución	Permite velocidades de flujo mayores, fácil interacción con LC, compuestos termoestables método de ionización suave	Necesita solubilidad en disolventes polares, sensible a sales, supresión de iones

### 1.3. Analizadores de masas en espectrometría de masas

Un espectro de masas consiste en un haz de muchos iones representado en valores de  $m/z$ , producido tras la ionización del analito. Para obtener un espectro de masas es necesario separar los iones de las diferentes relaciones  $m/z$  y, a continuación, medir las intensidades relativas de cada ion. En la MS, los analizadores de masas son los componentes que separan los iones en función de su relación masa/carga ( $m/z$ ). Existen diversos analizadores de masas que varían en resolución, exactitud de masa, rango dinámico y capacidad de realizar experimentos de MS en tándem. Estos conceptos se pueden definir como:

**Resolución:** El término resolución se emplea en numerosas áreas de interés analítico y, con mayor frecuencia, se refiere a la capacidad de diferenciar entre señales estrechamente relacionadas.

Estas señales, en MS, se consideran relaciones  $m/z$  de iones y las ecuaciones para una resolución  $R$  y un poder de resolución  $RP$  se describen a continuación:

$$R = \Delta m/m$$

$$RP = m_1/(m_2 - m_1),$$

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

Otra expresión es  $RP = m/Dm$ , donde  $m$  es la relación  $m/z$  medida y  $Dm$  (en Da) normalmente se mide utilizando la anchura del pico a un porcentaje específico de la altura del pico, normalmente el ancho total a la mitad del máximo de altura (FWHM, por sus siglas en inglés).

**Exactitud:** Es la proximidad de la masa experimental (masa precisa) al valor verdadero (masa exacta). En espectrometría de masas se determina en ppm a través de la siguiente ecuación:

$$\text{Error} = (\text{masa exacta medida} - \text{masa monoisotópica teorica}) \times 10^6 / \text{masa monoisotópica exacta}$$

Un instrumento que puede medir masas diferenciando solo masas enteras, muestra una exactitud de masas nominales o exactitud de masa unidad. Una mayor resolución está correlacionada con una mayor exactitud de masa, y esto incrementa las capacidades de identificación de un instrumento de MS

**Rango dinámico lineal:** definido como el rango sobre el que la señal de iones es directamente proporcional a la concentración de analitos. El rango lineal es crucial para obtener mediciones precisas, especialmente en el análisis de cuantificación.

Existen muchos dispositivos distintos de separación de masas que se emplean en MS, conocidos como analizadores, y cada uno de ellos presenta sus ventajas e inconvenientes. Los analizadores que con mayor frecuencia se aplican al análisis biológico son los cuadrupolos (Q), trampas de iones (IT), analizadores de tiempo de vuelo (TOF) y, más recientemente, los orbitraps.

El *cuadrupolo* tal y como sugiere su nombre, se compone de cuatro barras paralelas, como se observa en la Figura 1.4. Se aplica potencial de corriente continua (DC) y radiofrecuencia (RF) a los pares de barras opuestos del cuadrupolo. Si se escogen valores de DC y RF específicos, los iones con una relación  $m/z$  particular tendrán trayectorias estables y pasarán a través del analizador al detector, en tanto que no se transmitirá ninguno de los demás iones. Estas condiciones son la base para el modo de monitorización de iones seleccionado (SIM, por sus siglas en inglés), que es el modo de más habitual de trabajo con analizador de cuadrupolo para el análisis cuantitativo, ya que se consigue una enorme mejora en la relación señal/ruido. Por otra parte, se puede considerar el cuadrupolo un instrumento de escaneo cuando se utiliza en modo escaneo completo. En este modo, a través del escaneo de la DC y la RF en distintas relaciones, se transmiten todos los iones, una relación  $m/z$  de cada vez, al detector. Como resultado del modo de escaneo completo, se pueden obtener los espectros completos de las masas de resolución unidad.

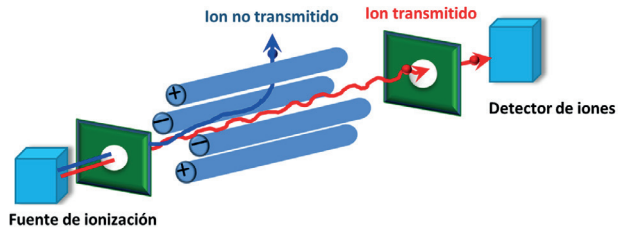


Figura 1.4. Esquema de un analizador de masas de cuadrupolo (Q).

El analizador de masas de trampa de iones (IT, por sus siglas en inglés), tal y como sugiere su nombre, atrapa los iones mediante campos cuadrupolares. Hay dos tipos de estos analizadores: Trampas de iones de 2D y 3D, que atrapan iones en dos o tres dimensiones, respectivamente. La trampa de iones de 3D, llamada también trampa de iones de cuadrupolo, se compone de cuatro electrodos: dos electrodos de anillo y dos electrodos elipsoides con el extremo encapsulado (tal y como se observa en la Figura 1.5). Todos los iones introducidos en la trampa de iones quedan atrapados a través de campos eléctricos cuadrupolares tridimensionales oscilantes. Sin embargo, cuando se aplica un voltaje específico al electrodo de anillo, los iones con un valor  $m/z$  específico se transmitirán al detector. Al alterar el valor de tensión, se transmiten diferentes iones al detector. En la trampa de iones de 2D, conocida también como trampa de iones lineal, los iones quedan atrapados a través de un campo de radiofrecuencia bidimensional, que proporciona un cuadrupolo de cuatro barras. Dentro de la trampa de iones de 2D, se aplican potenciales de bloqueo a los electrodos en la entrada y al final del cuadrupolo, para confinar los iones. Ambos analizadores de trampa de iones son instrumentos de baja resolución. La trampa de iones de 2D, sin embargo, posee mayores capacidades de almacenamiento de iones y de velocidad de escaneado. En las dos tecnologías, el gas helio está presente para controlar la trayectoria de los iones dentro de la trampa y llevar a cabo experimentos de  $MS^n$ .

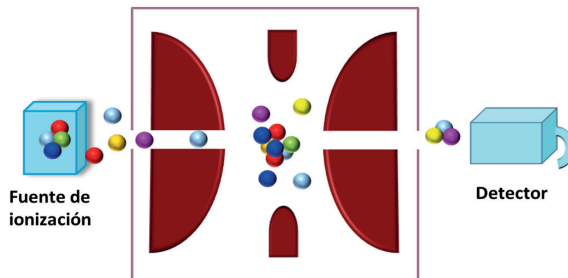
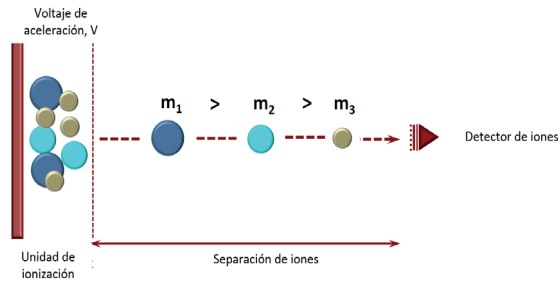


Figura 1.5. Esquema del analizador de masas de trampa de iones (IT).

El analizador de *tiempo de vuelo* (TOF, por sus siglas en inglés) (Figura 1.6) se basa en el hecho de que cuando se aceleran con un alto voltaje todos los iones producidos en la fuente de un espectrómetro de masas, dichos iones poseerán la misma energía cinética por unidad de carga ( $E_k$ ). En consecuencia, la velocidad de cada ion estará vinculada a la relación  $m/z$  ( $E_k=1/2(m/z)$

$v^2$ ). Así, el tiempo necesario para que los iones se desplacen a través del tubo de vuelo (tubo en condiciones de muy alto vacío) del espectrómetro de masas estará también vinculado a la relación  $m/z$  del ion. Transcurrido ese tiempo, en el que todos los iones lleguen al detector, se obtiene un espectro de masas completo.



**Figura 1.6.** El mecanismo de separación de iones en el analizador de masas de tiempo de vuelo (TOF)..

Una de las ventajas más importantes del instrumento TOF, si se compara con los instrumentos de cuadrupolo y de trampa de iones, es que se pueden determinar con mayor resolución y exactitud de las relaciones  $m/z$  de los iones. Se pueden obtener mayores resoluciones si se incrementa la longitud del tubo de vuelo, de manera que los iones necesiten más tiempo para desplazarse desde la fuente al detector. Otra solución es utilizar espejos de iones en los instrumentos TOF, los denominados reflectrones, que reflejan el haz de iones. El reflectrón incrementa también la resolución al disminuir la dispersión de la energía cinética. Por otra parte, los instrumentos TOF son instrumentos de escaneo rápido, limitados por el tiempo que tarde el ion más pesado en llegar al detector. La otra ventaja del analizador TOF es que el TOF lineal no tiene ninguna limitación en el rango de masas, por lo que resulta el más útil para el análisis de proteínas intactas.

Más recientemente se ha desarrollado un nuevo analizador: el Orbitrap, que consiste en un electrodo con forma de barril. Los iones quedan atrapados radialmente en torno a un electrodo central, y la transformada rápida de Fourier calcula los valores de  $m/z$  a partir de las frecuencias de oscilación de los iones atrapados. Este detector se caracteriza por su alta resolución, hasta 500.000 FWHM, alta exactitud ( $<1\text{ppm}$  con calibración interna) y sensibilidad. Resumiendo, en la Tabla 1.2 se expone una comparativa de los distintos analizadores de MS y sus principales características.

**Tabla 1.2.** Caracterización de los distintos analizadores de MS.

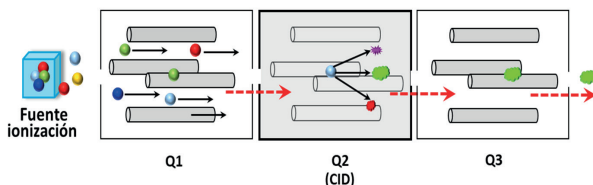
Analizador	Q	IT	TOF	Orbitrap
Velocidad de adquisición (Hz)	2-10	2-10	10-100	1-18
Exactitud de masa (ppm)	baja	baja	1-10 <sup>a</sup> ppm	1-5 ppm
Rango de masas( <i>m/z</i> )	<3000	<6000	<100.000 <sup>a</sup> ilimitado	<6000
Resolución <sup>b</sup>	unidad	unidad	<50.000 <sup>a</sup>	<500.000 <sup>b</sup>
Ventajas	Bajo coste, fácil interacción con varias técnicas de ionización mayor rango dinámico	Bajo coste, fácil interacción con varias técnicas de ionización, MS <sup>n</sup>	Escaneo rápido, alto rango de masas, alta exactitud de masas	Alta exactitud de masas, cambio rápido de polaridad
Inconvenientes	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	Alto coste, rango dinámico menor que Q	Velocidad de escaneo menor que QTOF, alto coste, rango dinámico menor que Q

a)En configuración híbrida/ con reflectrones; b) resolución de trabajo (FWHM)

## 1.4. Espectrometría de masas en tándem

La espectrometría de masas en tándem (MS/MS) es una técnica para descomponer iones seleccionados (iones precursores) en fragmentos (iones del producto), aportando información valiosa sobre la estructura química del ion precursor. La MS/MS se puede llevar a cabo con un espectrómetro de masas en tándem, que consiste en dos analizadores de masas separados mediante una celda de colisión que contiene un gas de colisión, donde normalmente se produce una disociación por colisión inducida (CID, por sus siglas en inglés). Los experimentos de MS en tándem se pueden emplear para enfoques cuantitativos orientados, en los que normalmente no se requiere una alta resolución ni exactitud de masa, o con fines de identificación en los que normalmente se requiere una alta resolución y exactitud de masa.

El triple cuadrupolo (*QqQ*) probablemente sea el instrumento de MS/MS más empleado en enfoques orientados al análisis cuantitativo. Este dispositivo, tal y como sugiere su nombre, comprende tres cuadrupolos en serie (Figura 1.7). El segundo cuadrupolo (*Q2*) no se utiliza como unidad de separación de masas, sino como célula de colisión (*q*), donde normalmente se produce una disociación por colisión inducida (CID). Así, en el módulo *Q2*, los iones transmitidos a través del primer cuadrupolo (*Q1*) se fragmentan, y los iones producto generados se transmiten a continuación al tercer cuadrupolo (*Q3*).



**Figura 1.7.** La espectrometría de masas en tándem basada en el funcionamiento de un triple cuadrupolo.

Además de los triples cuadrupolos, otros instrumentos híbridos de MS/MS frecuentemente utilizados, que utilizan diferentes analizadores, son los instrumentos de cuadrupolo tiempo de vuelo (QqTOF) y QqOrbitrap. Otros instrumentos que se utilizan para la espectrometría de masas en tándem son los instrumentos de trampas de iones, en los que se producen las tres etapas de análisis dentro del mismo analizador (la trampa de iones). Además, se pueden utilizar otros métodos de fragmentación como el de Disociación por Captura de Electrones (ETD) o el de Disociación por colisión de alta energía (HCD), cada uno de ellos con sus propias aplicaciones específicas.

Se pueden llevar a cabo cuatro experimentos principales diferentes de MS/MS con espectrómetros de masas en tándem, aunque tres de ellos solo se pueden realizar utilizando instrumentos de triple cuadrupolo.

*Escaneo de iones producto:* el primer módulo del espectrómetro de masas en tándem (Q1) se utiliza para aislar el ion de interés (ion precursor). Seguidamente, el ion seleccionado se fragmenta en una celda de colisión (q) mediante colisión con un gas ultrapuro. El tercer espectrómetro de masas (Q3) produce un espectro de masas mediante el escaneo de los iones producto generados a partir del ion precursor dentro de la célula de colisión. Este experimento se puede llevar a cabo en todos los tipos de espectrómetros de masas en tándem.

*Escaneo de ion precursor:* el tercer módulo del espectrómetro de masas (Q3) solo permite la transmisión de un ion (fragmento) producto seleccionado, con un valor  $m/z$  conocido, mientras que el primer módulo (Q1) escanea sobre el rango de masas de interés, lo que permite que se transmitan y fragmenten secuencialmente todos los iones precursores dentro de la célula de colisión (q). La señal procedente del detector se registra si los iones transmitidos por Q1 originan el producto monitorizado en Q3. Este experimento solo se puede llevar a cabo en instrumentos de triple cuadrupolo.

*Modo de escaneo de pérdida neutra:* En este modo, los espectrómetros de masas Q1 y Q3 escanean de manera vinculada de una forma en que la diferencia de  $m/z$  en Q1 y de  $m/z$  en Q3 se corresponde con la masa de un fragmento neutro que se haya perdido en la fragmentación dentro de la célula de colisión (q). Los iones producto generados como consecuencia de la pérdida de cualquier otra masa no se transmiten por Q3. Este experimento solo se puede llevar a cabo en instrumentos de triple cuadrupolo.

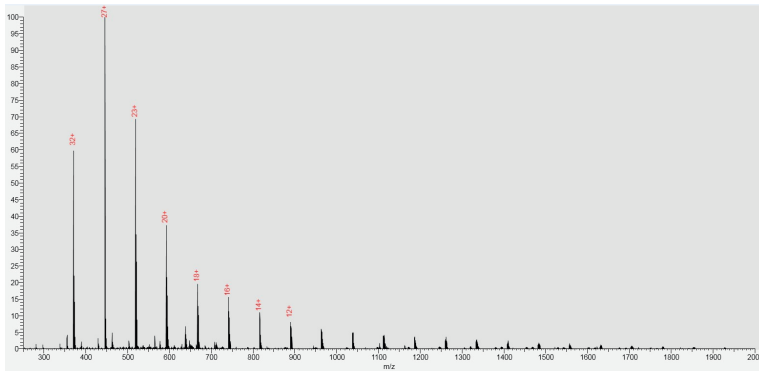


*Monitorización de reacción seleccionada (SRM) y monitorización de reacción múltiple (MRM):* Los dos son experimentos habituales de MS/MS empleados en enfoques analíticos cuantitativos. El modo SRM monitoriza la fragmentación de un ion precursor seleccionado hasta un ion producto seleccionado. Esto se lleva a cabo configurando cada uno de los módulos del espectrómetro de masas en tándem para transmitir un solo ion, es decir, el ion precursor, a través de Q1 y el ion producto a través de Q3. El modo MRM es la monitorización de la reacción seleccionada aplicada a múltiples iones producto a partir de uno o más iones precursores. En este enfoque se monitorizan varios pares «producto-precursor». La combinación de la separación por cromatografía de líquidos con la identificación de SRM/MRM se puede utilizar para incrementar la especificidad de la medición, eliminando transiciones «producto-precursor» que interfieran. Un inconveniente importante de este enfoque es que no resulta adecuado para un análisis global, puesto que no se podrían proporcionar los pares específicos de iones de transición. Con respecto a la detección de MS, la MRM es preferible a la SIM, toda vez que el modo MRM ofrece una mayor sensibilidad, mayor selectividad y reproducibilidad, y, así, a día de hoy, es el tipo más popular de configuración experimental con equipos de triple cuadrupolo.

## 1.5. Interpretación de los espectros de masas

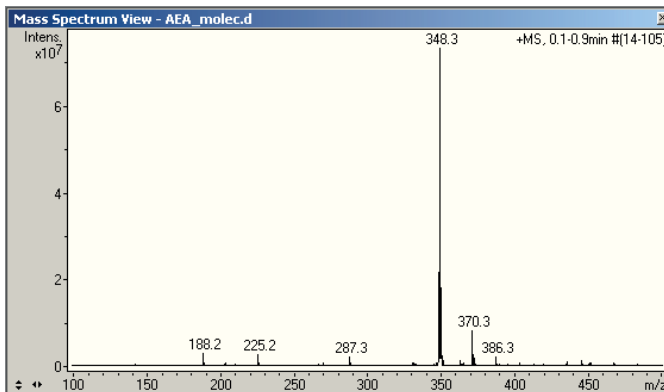
En esta sección se describen los principios fundamentales de la interpretación de los espectros de masas considerando las técnicas de ionización descritas en la sección 1.2. Los espectros de masas producidos por los métodos de ionización ESI, MALDI y APCI pueden originar tanto iones positivos como negativos.

El modo de ionización ESI puede producir múltiples iones cargados. El número de carga depende de la estructura del compuesto, en concreto, del número de posibilidades de ganar o perder protones, en modos de ionización positiva y negativa, respectivamente. Esta propiedad resulta especialmente útil para la interpretación de los espectros de masas de compuestos de alto peso molecular, como las proteínas intactas (Figura 1.8), mientras que los compuestos de bajo peso molecular cuentan normalmente con menos sitios potenciales para la protonación o desprotonación y, por tanto, sus espectros muestran menos iones cargados múltiples.



**Figura 1.8.** Espectro de masas de tiorredoxina de ESI (12 kDa) como ejemplo de un espectro de masas de ESI típico con múltiples iones cargados [Quadrupole-Orbitrap, Thermo Fisher, Q Exactive Plus].

En el caso de compuestos de bajo peso molecular, con tan solo un grupo polar, sus espectros de ESI muestran predominantemente iones moleculares, normalmente  $[M+H]^+$  o  $[M-H]^-$ , por tanto, es fácil de determinar el peso molecular de las especies individuales. Debido a que algunos analitos ionizan mejor en modo positivo que en negativo, mientras que en otros es al contrario, algunas veces vale la pena analizar las muestras utilizando los dos modos de ionización. Por otra parte, debe considerarse la presencia de aductos del analito con especies como  $Na^+$ ,  $K^+$ ,  $NH_4^+$  en modo positivo, y aniones procedentes de la fase móvil como el formiato o acetato en modo de ionización negativa.



**Figura 1.9.** Espectro de masas de anandamida de ESI ionizado en modo positivo; la señal a 348,3  $m/z$  se corresponde con el ion molecular cargado positivamente  $[M+H]^+$  de anandamida [IonTrap, Bruker Daltonics, Esquire 6000].

Por último, se analizarán algunos aspectos útiles en la interpretación de los espectros. En los espectros de masas se pueden observar varios tipos de iones distintos a los iones moleculares. A estos iones se les llama «iones fragmento» y su abundancia depende de varias variables, como el método de ionización, los métodos de fragmentación, la presencia en la molécula de grupos

lábiles, etc. Los métodos de ionización suave producen normalmente muy poca, o incluso ninguna, fragmentación. En estos casos, los iones precursores normalmente se fragmentan en una célula de colisión cuando se trabaja en modo de espectrometría de masas en tándem. Por otra parte, la ionización con electrones (EI) produce una fragmentación extensa y reproducible. Normalmente, el patrón de fragmentación de moléculas de la misma familia (por ejemplo, péptidos, lípidos, etc.) es similar y su conocimiento resulta esencial para identificar estas clases. Esto se abordará más adelante. Por ejemplo, cuando se utiliza la EI como fuente de ionización, los compuestos que contienen cadenas de hidrocarburos dan origen a una serie de iones con una diferencia de masas entre ellas de 14 Da, lo que equivale a (-CH<sub>2</sub>-). La pérdida de agua se observa por una pérdida de 18 Da que es normal en compuestos que contienen grupos hidroxilos. Existen muchos otros patrones de fragmentación habituales como estos que el científico de espectrometría de masas emplea para identificar compuestos.

## 2. Técnicas de separación acopladas a la espectrometría de masas

### 2.1. Cromatografía líquida-espectrometría de masas (LC-MS)

Tradicionalmente, en la cromatografía de líquidos, los compuestos se identificaban por su tiempo de retención ( $t_R$ ) en un cromatograma. Sin embargo, en el caso de muestras con compuestos caracterizados por un  $t_R$  similar, o de varias muestras complejas, la separación cromatográfica no permite su identificación. Por ello normalmente se necesita más información, empleando una técnica adicional de elucidación estructural. Por este motivo, la aplicación de técnicas acopladas, como la LC-MS, es deseable toda vez que la MS ofrece uno de los métodos de detección más sensibles de los que dispone en la actualidad, al mismo tiempo que permite la identificación simultánea de los compuestos. Una de las mayores ventajas de la LC-MS es el alto número de especies detectables y el menor efecto matriz al atenuarse la supresión de iones mediante la fase de separación. La MS proporciona información sobre el peso molecular y la estructura de los compuestos, a partir de picogramos o, incluso, femtogramos de material. Además de esto, garantiza la más alta selectividad al permitir la monitorización de iones o de iones fragmento característicos procedentes del compuesto a analizar. Por tanto, la LC-MS es una técnica analítica ideal tanto para el análisis cualitativo como cuantitativo.

## 2.1.1. Introducción a la LC-MS

La separación cromatográfica de compuestos de una mezcla, en combinación con la capacidad de identificación del espectrómetro de masas, ofrece ciertas ventajas. La LC-MS permite diferenciar numerosos compuestos con  $t_R$  similar pero con diferente  $m/z$  o patrón de fragmentación. La especificidad única y la alta selectividad de la LC-MS posibilitan la cuantificación, algo que no es posible con tan solo la cromatografía. Las técnicas de ionización suave (sección 1.2), que son las técnicas de ionización más extensamente utilizadas en LC-MS, permiten estudiar una amplia variedad de compuestos, desde fármacos y metabolitos de bajo peso molecular (por debajo de 1.000 Da) a biopolímeros de alto peso molecular (más de 100.000 Da). A través de la espectrometría de masas se puede obtener el peso molecular del analito a determinar, junto con la información estructural que podría generarse, ayudando a la identificación. Asimismo, gracias a su alta selectividad y sensibilidad, la espectrometría de masas posibilita una cuantificación muy exacta y precisa.

Sin embargo, a pesar de las muchas ventajas que ofrecen los análisis LC-MS, existen algunos problemas que se deben tener en cuenta. Por ejemplo, diferentes especies poseen diferente capacidad de ionizarse, de lo que se derivan disimilitudes en sensibilidad, aunque se encuentren en concentraciones similares. Además, los análisis por LC-MS pueden resultar complicados por el fenómeno conocido como «supresión de iones» (descrito pormenorizadamente en la sección 1.2), que conlleva la reducción de la señal de un analito particular debido a sustancias que coeluyen, lo que afecta al proceso de ionización.

Un sistema cromatográfico típico contiene cuatro componentes principales: inyector de muestras, fase móvil, fase estacionaria y detector. Si se consideran las diferentes técnicas cromatográficas, la importancia de estos cuatro componentes puede ser distinta. Por lo general, en la cromatografía líquida, la función del inyector es la introducción de los analitos en el sistema mediante la fase móvil que fluye. A día de hoy, la válvula de seis puertos con el inyector de bucle es la más utilizada. Las fases móvil y estacionaria, como partes más importantes de un sistema cromatográfico, son responsables del proceso de separación. En la cromatografía líquida de alta eficacia/rendimiento (HPLC), la fase móvil es un disolvente o sistema de disolventes que se suministra con una velocidad de flujo constante bajo alta presión (normalmente un máximo de 400 bares), mientras que la fase estacionaria está empaquetada en una columna cromatográfica de una manera que permite aguantar esta alta presión

## 2.1.2. Principios teóricos de la cromatografía líquida

La cromatografía es una técnica de separación instrumental que tiene por objeto la separación de los componentes de una mezcla, lo que permite su identificación y cuantificación. La identificación normalmente se basa en las características de retención cromatográfica. La definición que da de la cromatografía la Unión Internacional de Química Pura y Aplicada

(IUPAC, por sus siglas en inglés) es la siguiente: «La cromatografía es un método físico de separación, en el que los componentes a separar se distribuyen en dos fases, una estacionaria (fase estacionaria) y otra que se mueve (fase móvil) en una dirección definida». La fase móvil se describe como «un fluido que penetra a través o a lo largo del lecho estacionario, en una dirección determinada». Puede ser un líquido, un gas o un fluido supercrítico, mientras que la fase estacionaria puede ser un sólido, un gel o un líquido. «Si es un líquido, puede estar adherido sobre un sólido y este sólido puede o no contribuir al proceso de separación».

En una separación cromatográfica, las moléculas del analito se distribuyen entre la fase móvil y la fase estacionaria en la medida en que depende principalmente de su estructura química. La velocidad de los componentes de la muestra en la columna es una función de la distribución de estos componentes en dos fases que permanecen en equilibrio. Los componentes con una mayor afinidad con la fase estacionaria se mueven más lentamente que los componentes que poseen una mayor afinidad para la fase móvil. En consecuencia, la separación es el resultado de distintas velocidades de migración provocadas por distintos valores del coeficiente de partición ( $K_s$ ). La ecuación de Nernst puede expresar el coeficiente de partición:

$$K_s = C_s/C_M$$

donde  $C_s$  y  $C_M$  representan la concentración de las sustancias en las fases estacionaria y móvil, respectivamente. Una afinidad mayor para el material de la fase estacionaria resultará en un valor mayor de  $K_s$ , cuanto mayor sea el valor del tiempo de retención ( $t_R$ ).

El «tiempo de retención» ( $t_R$ ) es el tiempo que le lleva a un analito eluir de una columna cromatográfica con una fase móvil y llegar al detector. Debido a que la longitud de la columna y la velocidad de flujo de una fase móvil afectan a  $t_R$ , se ha introducido otro parámetro, el «factor de retención» ( $k$ ). Este factor relaciona el tiempo de retención de un analito con el tiempo que le lleva a un compuesto no retenido, que no interactúa con la fase estacionaria y eluye de la columna en el tiempo específico ( $t_0$ ). En ocasiones, el  $t_0$  se denomina el tiempo muerto de la columna. A la diferencia entre  $t_R$  y  $t_0$  se le suele llamar el tiempo de retención corregido o ajustado. Por tanto, el factor de retención se define mediante la siguiente ecuación:

$$k = (t_R - t_0)/t_0$$

Los compuestos de una mezcla se separan cuando interactúan de manera diferenciada con la fase móvil y estacionaria. Así, el tiempo que necesitan las sustancias para atravesar la distancia desde el punto de inyección de la muestra hasta el detector varía. Hay dos extremos de dicha interacción. Si todos los analitos poseen una afinidad total para la fase móvil y no interactúan con la fase estacionaria, no se separarán y llegarán muy rápido y al mismo tiempo al detector (tiempo de la fase móvil o tiempo muerto). Si todos los analitos interactúan fuertemente con la fase estacionaria y no interactúan con la fase móvil, quedarán retenidos en la columna y

no llegarán al detector. En consecuencia, el conocimiento de las propiedades químicas de los compuestos analizados es crucial si se va a considerar un proceso de separación cromatográfica. Si se conocen las propiedades, al cambiar las condiciones cromatográficas, por ejemplo, las propiedades de las fases estacionaria y móvil, es posible conseguir la separación deseada. El éxito de la separación de compuestos presentes en una mezcla se relaciona con una serie de parámetros mostrados en la Tabla 1.3.

**Tabla 1.3.** Los parámetros principales de la separación cromatográfica.

Parámetro	Símbolo	Definición
<b>Tiempo de retención</b>	$t_R$	Tiempo que le lleva a un analito eluir de una columna cromatográfica y llegar al detector
<b>Tiempo muerto de la columna</b>	$t_0$	Tiempo que le lleva a un compuesto no retenido que no interactúe con la fase estacionaria.
<b>Tiempo de retención corregido</b>	$(t_R - t_0)$	La diferencia entre el tiempo de retención y el tiempo muerto de la columna
<b>Factor de retención</b>	$k$	La relación entre el tiempo de retención reducido y el tiempo muerto de la columna
<b>Selectividad (factor de separación)</b>	$\alpha$	La relación entre los factores de retención de dos compuestos separados
<b>Resolución</b>	$R$	La capacidad de separar dos picos cromatográficos
<b>Número de platos</b>	$N$	El número de pequeñas capas, a lo largo de la columna cromatográfica, en las que se produce un equilibrio de distribución del analito entre la fase estacionaria y la móvil

Conviene señalar que para obtener una resolución suficiente en un tiempo razonable del análisis, los valores  $k$  deben situarse en un rango de entre 1 y 10. La eficiencia de la separación de dos compuestos (A y B) se define mediante el «factor de separación» ( $\alpha$ ), que representa la relación entre sus factores de retención  $k_B$  y  $k_A$  (el coeficiente de distribución para la primera elución de los dos componentes):

$$\alpha = k_B/k_A$$

La separación de dos compuestos cuando  $\alpha > 1$ , donde cuanto mayor es el valor de  $\alpha$ , mayor será la separación. Esta posibilidad se describe a través de otro parámetro llamado «resolución» ( $R$ ), que se obtiene mediante la ecuación:

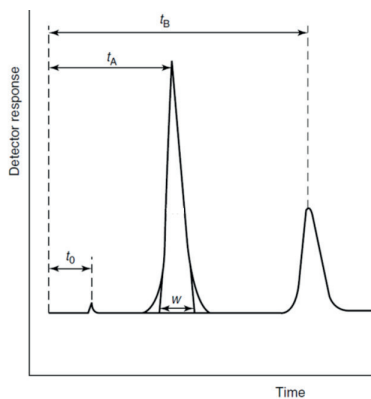
$$R = (t_B - t_A) / 0.5(w_A + w_B)$$

donde  $w_A$  y  $w_B$  representan las anchuras en la base de los picos correspondientes a dos compuestos A y B, dos picos inmediatamente adyacentes.

El cromatograma es la gráfica de la respuesta del detector frente al tiempo de elución, y la descripción gráfica de los parámetros antes mencionados se representa en la Figura 1.10.

Con respecto a la columna cromatográfica, su rendimiento es un parámetro importante que afecta a la separación cromatográfica. El rendimiento de la columna, conocido también como eficacia, está relacionado con el número de platos teóricos ( $N$ ). El número de platos teóricos es una medida de las bondades de la columna. Este parámetro se refiere al número de pequeñas capas, a lo largo de la columna cromatográfica, en las que se produce un equilibrio de distribución de la muestra entre la fase estacionaria y la móvil. El número de platos depende de la longitud de la columna ( $L$ ), al contrario que el segundo parámetro, la «altura del plato» ( $H$ ), definido como la eficacia de la columna en la longitud del plato teórico. La regla general de la cromatografía es que cuanto menor sea la altura del plato mayor será la resolución. Del mismo modo, también es un criterio deseable un número alto de platos. La relación entre  $N$  y  $H$  se describe a continuación:

$$H = L/N$$



**Figura 1.10.** Cromatograma que muestra un tiempo de retención muerto  $t_0$ , tiempos de retención  $t_A$  y  $t_B$ , de compuestos A y B, y  $w$ , anchura del pico.

### 2.1.2.1. El proceso cromatográfico en HPLC

Diversos mecanismos de retención tienen lugar en la HPLC. Sin embargo, las interacciones particulares se basan en las polaridades relativas de compuestos separados. Las propiedades de la fase móvil resultan esenciales para la separación cromatográfica. Si la fase móvil proporciona una separación insuficiente de analitos polares se retendrán durante más tiempo en la columna los analitos no polares, y viceversa. No obstante, no siempre es posible conseguir una separación adecuada mediante una fase móvil que contenga un solo disolvente. Generalmente, cuando se da esta situación, se utiliza una mezcla de disolventes. Una separación obtenida con una fase móvil de

composición constante se define como «elución isocrática», mientras que, cuando se altera durante el análisis la composición de la fase móvil, se define como «elución en gradiente». En la elución en gradiente, se va aumentando a lo largo del análisis la proporción de un disolvente que se sabe que es más eficaz para eluir los compuestos más retenidos.

Si bien existen muchas condiciones cromatográficas diferentes, no siempre es posible obtener una separación suficiente de todos los componentes de una mezcla. En consecuencia, dependiendo del tipo de analito y de la polaridad de las fases estacionaria y móvil, se utilizarán modos diferentes de cromatografía líquida (Tabla 1.4).

**Tabla 1.4.** Comparativa de los distintos modos de cromatografía líquida.

Modo LC	Fase Móvil	Fase estacionaria	Tipo de compuestos separados
NP	disolventes 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

### 2.1.2.2. Cromatografía de fase normal (NP)

La separación cromatográfica en fase normal se deriva de interacciones polares de compuestos separados con la fase estacionaria de una columna. El término «normal» se refiere al sistema en el que la fase estacionaria es polar, mientras que la fase móvil es no polar. En este modo, se retienen más los componentes polares y, al incrementar la polaridad de la fase móvil, puede disminuir la retención. Cuanto más polar sea la fase móvil, más rápido eluirán los analitos de la columna. La sílice no modificada, con sus grupos silanol libres como grupos funcionales, es la fase estacionaria más habitual utilizada en cromatografía líquida NP. La fase móvil en NP se basa normalmente en hidrocarburos, diclorometano, acetato de etilo u otro disolvente inmiscible con agua. Desde el punto de vista de la partícula, el principal motivo para utilizar cromatografía NP es incrementar la retención de componentes polares y eluir compuestos hidrofóbicos. Hay que mencionar que la cromatografía NP también resulta útil en la separación de isómeros geométricos y posicionales.



### 2.1.2.3. Cromatografía de fase inversa (RP)

La cromatografía de fase inversa (RP) es la HPLC más habitual. En este modo, la separación se basa en las interacciones de analitos con una fase estacionaria no polar o de baja polaridad y una fase móvil líquida polar. En la RP, la fase móvil es más polar que la fase estacionaria. Así, los compuestos más polares son eluidos primero y antes que los menos polares. Por otra parte, en la RP, la separación cromatográfica se basa en diferencias en la hidrofobicidad de los componentes de la mezcla.

Existen muchas fases estacionarias no polares utilizadas en la HPLC de fase inversa. La fase estacionaria de las columnas de RP que más se utilizan normalmente contiene hidrocarburos de cadena larga enlazados covalentemente a la superficie de la sílice. La fase estacionaria de RP más popular y universal es la C18 u ODS, que contiene grupos octadecilsilanos. El sistema disolvente habitual en RP comprende dos componentes principales, agua o una disolución tampón, y disolventes orgánicos, entre los cuales, los más utilizados son el metanol o el acetonitrilo. La cromatografía RP es una técnica muy completa, apta para la separación de compuestos no polares y de baja polaridad incluso durante un único análisis.

Tanto la NP-LC como la RP-LC se han utilizado con distintas finalidades en análisis de muestras biológicas. Por ejemplo, una columna de NP se podría aplicar para separar clases individuales de lípidos basadas en los grupos de cabeza polar, mientras que la separación de especies de lípidos en la columna de RP se basa en sus distintas hidrofobicidades (cadenas de ácidos grasos). Por otra parte, la cromatografía RP en gradiente también resulta útil para la determinación de perfiles de metabolitos en los estudios metabolómicos.

### 2.1.2.4. Cromatografía de interacción hidrofílica (HILIC)

La cromatografía de interacción hidrofílica (HILIC) es otro modo de elución para los compuestos polares como la cromatografía NP. Utiliza fases estacionarias hidrofílicas, con una fase móvil consistente en un sistema disolvente típico para RP, el más utilizado el acetonitrilo, con una pequeña cantidad de agua. Con frecuencia se añaden a la fase móvil sales como el acetato o el formiato de amonio para incrementar la polaridad y la fuerza iónica. Normalmente las columnas de HILIC contienen superficies polares de sílice o se pueden derivatizar a fases ligadas tipo amino o amida.

El mecanismo de separación del modo HILIC es complejo y se basa en un sistema de extracción líquido/líquido con una formación de capa de agua sobre la superficie de la fase estacionaria polar y la fase móvil orgánica. Asimismo, otros mecanismos además de la interacción hidrofílica con la capa rica en agua y la fase móvil son las interacciones electrostáticas y las interacciones polares con la fase estacionaria. Por tanto, en HILIC, los compuestos polares interactúan con la fase estacionaria, mientras que los menos polares se distribuyen más en la fase móvil. Al intervenir fases móviles altamente orgánicas en la separación de la HILIC, la técnica se puede adaptar fácilmente

a la MS. Además, el uso de disolventes orgánicos aumenta la sensibilidad de la MS debido a una disminución de la supresión iónica.

### 2.1.2.5. Cromatografía de intercambio iónico (IEC)

La cromatografía de intercambio iónico (IEC) conlleva la separación de compuestos iónicos e ionizables. Emplea materiales de empaque que contienen grupos funcionales iónicos, normalmente con cargas opuestas a los analitos. En la IEC se utiliza una fase estacionaria aniónica para la separación de cationes o moléculas cargadas positivamente, mientras que para las moléculas o aniones con carga negativa se recurre a la fase estacionaria catiónica. En este modo cromatográfico, la fase móvil es altamente polar, basada con más frecuencia en agua, con algún tampón o sales. El principal factor responsable de la elución de compuestos de la columna de IEC es el aumento de la fortaleza iónica de la fase móvil.

La IEC resulta útil para las separaciones biomoleculares tanto grandes como pequeñas, tales como aminoácidos, ácidos carboxílicos o aminos. Conviene señalar que, debido a un fenómeno de supresión iónica, resulta relativamente difícil acoplar directamente la IEC, que utiliza una fase móvil con alta fuerza iónica, a un espectrómetro de masas.

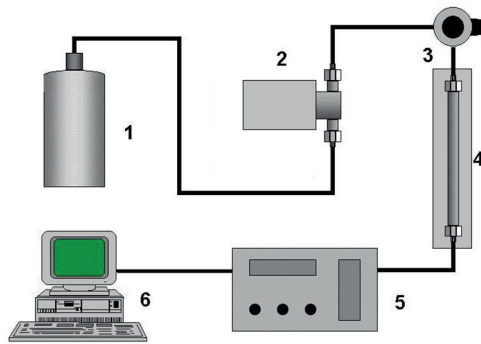
### 2.1.3. Instrumentación para LC-MS

En la Figura 1.11 se muestra un diagrama de bloques de un sistema de LC donde se representan sus componentes principales.

#### a) Bomba

La principal función de una bomba es proporcionar un flujo estable que varía en función de la interfaz que se utilice en la LC-MS y de los parámetros de la columna cromatográfica. Por ejemplo, en el caso de la fuente de ESI, cuando se utilice un diámetro de columna de 2,1 mm, se optará normalmente por una velocidad de flujo de entre 0,2 y 0,4 mL/min, mientras que se preferirá la velocidad de flujo de 1 mL/min para la APCI con una columna convencional de 4,6 mm.

Por otra parte, cuando se combine la LC con la MS, la bomba no deberá generar ningún pulso, sino proporcionar una velocidad de flujo constante de la fase móvil. La pulsación del flujo dificulta la interpretación de los resultados obtenidos. En consecuencia, para prevenir la formación de burbujas en la fase móvil, se incluye un desgasificador en cada sistema LC-MS.



**Figura 1.11.** Diagrama de bloques de un sistema de LC (1 – depósito con la fase móvil, 2 - bomba, 3 - inyector, 4 – columna cromatográfica, 5 - detector, 6 – ordenador).

### b) Inyector

En la LC se utiliza casi exclusivamente un solo tipo de inyector, conocido como el «inyector de bucle» (o inyector de válvula de seis puertos). En este inyector, se introduce la muestra con una microjeringa en una fase móvil que llenaba un bucle de un volumen nominal. 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. En el momento de la inyección, la fase móvil fluye a través del bucle, descargando sus contenidos en la columna. Los problemas importantes que presentan los inyectores son que deben permitir inyecciones con una alta reproducibilidad y exactitud, evitando la presencia de burbujas de aire y pulsos.

### c) Columnas

El formato de una columna de HPLC se refiere a la longitud de la columna, el diámetro de la columna y el tamaño de partícula de la fase estacionaria. Las columnas habituales de HPLC, capaces de aguantar presiones de 400 bares, tienen un diámetro de 4,6 mm, su longitud varía entre 100 y 250 mm, y el tamaño de partícula oscila normalmente entre los 3 y los 5  $\mu\text{m}$ . Sin embargo, el desarrollo de bombas de cromatografía líquida de ultra alto rendimiento (UPLC), capaces de bombear la fase móvil a una presión de 1200 bares, permite utilizar un tamaño de partícula reducido en columnas cromatográficas que mejoran la resolución cromatográfica. Por otra parte, la disminución de la velocidad de flujo de la fase móvil, a consecuencia de la disminución del tamaño de partícula y diámetro de la columna, confiere una mayor sensibilidad y menores límites de detección ofrecidos por los sistemas de nano-LC, que necesitan cantidades traza de muestra para el análisis. Están disponibles distintos tipos de columnas de HPLC, en relación a los distintos formatos, tal y como figura en la Tabla 1.5.

**Tabla 1.5.** Formatos de columnas de HPLC y su nomenclatura.

Descripción	Dimensión (i.d.)	Velocidad de flujo óptima
Columna calibre nano	0,075 mm	0,3 $\mu\text{L}/\text{min}$
Columna capilar	0,5 mm 1 mm	10 $\mu\text{L}/\text{min}$ 50 $\mu\text{L}/\text{min}$
Columna de calibre micro	2,1 mm	250 $\mu\text{L}/\text{min}$
Columna de calibre estrecho (pequeño)	3 mm	500 $\mu\text{L}/\text{min}$
Columna de calibre normal	4,6 mm	1250 $\mu\text{L}/\text{min}$

**Para saber más:**

G. Rozing. Trends in HPLC column formats - microbore, nanobore and smaller. LC GC EUROPE, 2003 16.6A:14-19.

S.R. Wilson, T. Vehus, H.S. Berg, E. Lundanes. Nano-LC in proteomics: recent advances and approaches. Bioanalysis, 2015, 7(14):1799-1815.

N. Gray, M.R. Lewis, R.S. Plumb, I.D. Wilson, J.K. Nicholson. High-throughput microbore UPLC-MS metabolic phenotyping of urine for large-scale epidemiology studies. J. Proteome Res., 2015, 14(6):2714-2721.

**d) Detectores**

Los detectores utilizados en LC se caracterizan por distintos parámetros, incluida la selectividad, el límite de detección y la sensibilidad. La *selectividad* define la capacidad de un detector de determinar un analito de interés sin interferencias derivadas de la matriz, los disolventes u otras sustancias presentes en el sistema. El término *sensibilidad* se emplea con frecuencia en lugar del término *límite de detección*, a pesar de que este último se refiere a la concentración más baja de un analito que se puede diferenciar de manera fiable del ruido de fondo. Sin embargo, el término *sensibilidad* se refiere a la respuesta del detector que esté relacionada con la concentración de un analito que llegue al detector. Estos parámetros son fundamentales a la hora de considerar el rendimiento general de un método analítico. Con todo, hay que tener en cuenta que la sensibilidad más alta no proporciona el límite más bajo de detección/determinación porque las interferencias presentes en la muestra pueden transmitir también señales fuertes.

Se pueden utilizar numerosos detectores conjuntamente con la LC, incluidos los detectores UV, de fluorescencia, electroquímicos, de conductividad, de índice de refracción y de MS. Cada uno de ellos posee sus propias ventajas e inconvenientes, aunque los detectores de LC más utilizados son el UV y el de MS, utilizándose con frecuencia un detector UV como detector de apoyo. Cuando se utiliza un detector UV, si se conoce la longitud de onda de máxima absorción del analito ( $\lambda_{\text{max}}$ ), se puede monitorizar, y el detector puede ser selectivo para ese analito. Lamentablemente, el detector UV no posee una alta

selectividad porque muchas moléculas orgánicas absorben radiación UV a la misma longitud de onda. En ese sentido, no cabe duda de que la MS es una técnica mucho más fiable para obtener información analítica, especialmente en mezclas complejas. La detección de MS tan solo precisa de cantidades en picogramos de muestra para proporcionar información precisa sobre el peso molecular y la estructura del compuesto. Los espectrómetros de masas son detectores ideales tanto para los enfoques analíticos cualitativos como cuantitativos, por lo que, a día de hoy, se considera el detector de preferencia para su acoplamiento a la LC.

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

### 2.2.1. Introducción a la GC-MS

La cromatografía de gases acoplada a la espectrometría de masas (GC-MS) es una combinación de dos técnicas instrumentales avanzadas para el análisis de compuestos orgánicos. La GC-MS es la técnica más efectiva para el análisis de compuestos orgánicos volátiles de matrices complejas en una amplia variedad de concentraciones. Resulta especialmente ventajosa en la determinación de compuestos de bajo peso molecular, polaridad media o baja, termoestables con puntos de ebullición por debajo de 350-400 °C, con diferentes concentraciones que oscilan entre ppb y ppm. El desarrollo de nuevas tecnologías en las columnas empleadas ha sido importante en el análisis de compuestos de alta temperatura de ebullición (hasta 500 °C). Las posibilidades analíticas de la GC-MS dependen, principalmente, del rendimiento del espectrómetro de masas. Con la disminución en los límites de detección y el aumento en el número de compuestos estudiados, el análisis de los datos puede resultar complicado. Por ello se hacen necesarios nuevos desarrollos. La técnica GC-MS se caracteriza por su alta selectividad y sensibilidad. Es una técnica esencial para la monitorización de una amplia variedad de aplicaciones, desde el análisis medioambiental a la bioquímica. Los enfoques metabólicos basados en GC-MS son ampliamente utilizados para comprender los procesos que provocan el desarrollo de enfermedades, como cánceres o dolencias cardiovasculares.

### 2.2.2. Principios teóricos de la GC-MS

#### 2.2.2.1. El proceso cromatográfico

En el sistema cromatográfico de gases (GC), la muestra a analizar podría ser un gas, un líquido o moléculas adsorbidas en una superficie después de la microextracción en fase sólida (SPME). El principio de separación en la GC depende de la transferencia de una sustancia (como vapor) empleando un gas portador (fase móvil) a través de una columna capilar, que normalmente estará recubierta internamente con una película polimérica fina. Durante la transferencia a la GC, los compuestos de la muestra se volatilizan mediante una rápida exposición a una zona previamente

calentada a una temperatura relativamente alta (200-300 °C) y se mezclan con una corriente de gas portador puro (Ar, He, N<sub>2</sub> o H<sub>2</sub>).

En la Figura 1.12 se muestra un diagrama de un sistema GC, donde se representan sus componentes principales, a saber, el gas portador, el puerto de inyección, el horno, la columna y el detector.

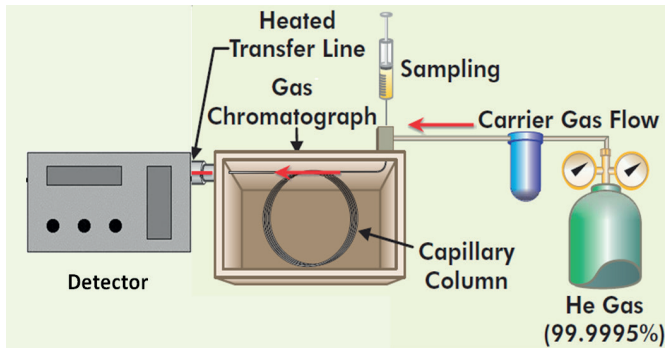


Figura 1.12. Diagrama del sistema GC (modificado a partir de *de.leco-europe.com*).

El helio es el gas portador más habitual en GC-MS y es compatible con el detector. Existen tres modos principales de inyección: split, especial para columnas capilares; splitless, la mejor para niveles traza de analitos de alta temperatura de ebullición en disolventes muy volátiles y por último, inyección en columna, para compuestos térmicamente inestables.

Los componentes de la muestra separada en la columna van llegando sucesivamente al detector, generando las señales eléctricas, que son proporcionales a la cantidad del compuesto separado. Se pueden describir dos modos de elución según la temperatura del horno: isotérmica (temperatura constante durante el análisis) y con programación de temperatura. En esta última, la temperatura aumenta durante la separación para incrementar la presión de vapor del analito y reducir los tiempos de elución de la mayoría de los compuestos retenidos.

Los detectores tradicionales utilizados en GC responden a diferencias en las propiedades fisicoquímicas del gas portador con y sin las sustancias eluidas de la columna, por ejemplo, los detectores de ionización de llama (FID), captura electrónica (ECD) o conductividad térmica (TCD). Los cambios registrados pueden ser proporcionales bien a la concentración, o a la masa del analito en un gas portador. Características importantes del detector son su elevada sensibilidad, detectabilidad (límite de detección), alta estabilidad de la señal, señal de línea base reproducible y un amplio intervalo lineal. Los datos resultantes se procesan para generar el cromatograma.

### 2.2.2.2. Fases de columna

Escoger el tipo de columna tiene una influencia decisiva en la calidad de la separación de los componentes de la mezcla, que es consecuencia del análisis cromatográfico. Existen dos tipos de columnas: empaquetadas (analíticas, microempaquetadas) y tubulares abiertas, denominadas con

frecuencia capilares (capilar, microcapilar). Las columnas empaquetadas se llenan con partículas y las columnas capilares son tubos capilares tubulares abiertos. En las columnas empaquetadas, las partículas sólidas (adsorbentes de carbono, sílice, alúmina, tamices moleculares, polímeros sintéticos porosos, entre otros) sirven como apoyo o se integran en un líquido (siliconas, escualeno, polietilenglicol, entre otros), actuando como fase estacionaria para separar componentes dentro de la GC.

Los adsorbentes son fases estacionarias menos habituales por la menor reproducibilidad de los resultados, el mayor tiempo de retención y la aparición de «colas» con eficiencias de separación mucho más bajas. Actualmente se utilizan con frecuencia columnas capilares recubiertas con una fase estacionaria líquida. Las fases estacionarias líquidas deberían ser químicamente inertes, capaces de disolver componentes separados, altamente selectivas para los componentes de la mezcla, deberían tener una baja volatilidad y poseer elevada estabilidad térmica bajo las condiciones de operación de la columna. Resultan especialmente aptas para la separación de componentes gaseosos con una alta eficacia de separación.

La característica más importante de la mayoría de las fases estacionarias líquidas que influye en las propiedades de separación es la polaridad. La eficacia de la separación en la columna capilar depende también de los siguientes parámetros y propiedades: dimensiones de la columna (longitud-L y diámetro interno-ID) y grosor de la fase. La longitud comercial estándar de las columnas capilares es de 15, 30 y 60 m. El diámetro interno de las columnas capilares tiene dimensiones estándar: 0,15; 0,25; 0,32 y 0,53 mm. Las columnas para su uso en GC-MS se adaptan especialmente a este sistema de detección y se caracterizan por su bajo sangrado.

Se suele escoger la polaridad de la fase estacionaria líquida después de considerar la polaridad de los analitos de la mezcla. Por regla general, para la separación de compuestos no polares se debe utilizar la fase no polar, y para la separación de los compuestos polares, una fase polar. En las columnas no polares (por ejemplo, escualeno, 100% metilpolisiloxano, (5%-fenil)-metilpolisiloxano), la elución de analitos se realiza de acuerdo con su volatilidad, es decir, por orden creciente de sus puntos de ebullición. En las columnas polares, por ejemplo, polietilenglicol (Carbowax), ((70% cianopropil) fenil-metilpolisiloxano), el factor determinante para la separación de analitos son su punto de ebullición y su polaridad (el momento dipolar- $\mu$ ) y la fortaleza de las interacciones con la fase estacionaria (dipolo, enlace de hidrógeno). Las columnas polares, como la de cianopropilfenil-polisiloxano, se caracterizan también por una alta selectividad, vinculada al alto contenido de grupos fenólico y cianico (Tabla 1.6). La temperatura de trabajo máxima permitida se corresponde con fases estacionarias no polares. Se podrán utilizar otras fases, dependiendo de la finalidad: alcoholes, sacáridos, compuestos nitrosos, aminas alifáticas, sales de ácidos inorgánicos (por ejemplo, nitrato de plata), entre otros. Existen también otras fases conocidas, cuyo mecanismo de separación conlleva varias interacciones de los compuestos con las partículas individuales de la fase estacionaria o una estructura ordenada, consiguiendo así una selectividad de separación considerable. Estas fases incluyen fases estacionarias quirales, fases líquidas iónicas, etc.

Tabla 1.6. Clasificación de fases estacionarias de GC

FASE LIGADA	USO GENERAL DE LA FASE	TEMP (°C)
Metilpolisiloxano	baja selectividad, separación según puntos de ebullición, estabilidad termal excelente, recomendada para análisis medioambientales, análisis de trazas, métodos EPA, pesticidas, PCB, análisis de alimentos y fármacos, fase más utilizada en GC	50-325
Metilpolisiloxano 5% fenil	selectividad ligeramente mayor que el metilpolisiloxano, estabilidad termal excelente, inertidad perfecta para compuestos básicos, fase completa para análisis medioambientales, análisis de trazas, métodos EPA, pesticidas, PCB, análisis de alimentos y fármacos, fase estándar con una amplia variedad de aplicaciones	50-325
Metilpolisiloxano 5% fenil	mayor selectividad que para el metilpolisiloxano, buena estabilidad termal, ideal para muestras biomédicas como esteroides o azúcares, fase completa para análisis medioambientales, análisis de ultratrazas, métodos EPA, pesticidas, PCB, análisis de alimentos y fármacos,	40-325
Metilpolisiloxano 6% cianopropil 94% fenil	selectividad nominal para compuestos polarizables y polares, buena estabilidad termal, fase de uso general pero especialmente idónea para análisis medioambientales sofisticados (por ejemplo, métodos EPA para PAH, PCB y pesticidas)	20-240
Metilpolisiloxano 7% cianopropil, 7% fenil	alta selectividad, buena estabilidad termal, ideal para muestras de azúcares derivados y medioambientales, recomendada para análisis de alcoholes, oxigenados y PCB	40-280
Metilpolisiloxano 25% cianopropil, 25% fenil	alta selectividad, fase polar, buena estabilidad termal, ideal para moléculas polares (por ejemplo, ácidos grasos, esteroides, acetatos de alditol, derivados de azúcares)	40-240
Poliétilenglicol	selectividad única para moléculas con tipo de enlace de hidrógeno, buena estabilidad termal, ideal para muestras oxigenadas, pero susceptible a degradación por oxígeno, recomendada para análisis de disolventes y alcoholes, ideal para soluciones acuosas, no recomendada para el análisis de mezclas que contengan reactivos de sililación	20-260
Metilpolisiloxano >70% cianopropil	alta selectividad, fase polar, buena estabilidad termal, ideal para análisis medioambientales difíciles y para ésteres metílicos de los ácidos grasos (FAME)	20-260
Polisiloxano >90% cianopropil	alta selectividad, alta fase polar, buena estabilidad termal, ideal para análisis de FAME, dioxinas, furanos, y para la rápida separación de compuestos aromáticos, perfumes, petroquímicos y otros compuestos que sean difíciles de resolver a través de columnas convencionales.	40-280

### 2.2.2.3. GC×GC/MS

La cromatografía de gases desempeña un papel importante entre las técnicas cromatográficas debido a su capacidad de alta resolución, flexibilidad y por su amplia variedad de aplicaciones. A pesar de ello, falla en el análisis de muestras muy complejas por causa de una capacidad de picos limitada (el número de picos cromatográficos, que pueden estar completamente separados usando tan solo una columna de la cromatografía). A medida que los componentes van migrando a través de la columna, las bandas cromatográficas se ensanchan, lo que genera un número limitado de bandas que pueden estar completamente separadas a la salida de la columna. Esta limitación no se puede superar modificando los parámetros cromatográficos. La única solución efectiva es utilizar una separación bidimensional (GC×GC), donde se separan de nuevo los componentes parcialmente separados de la primera columna de la cromatografía, mediante un mecanismo diferente, en una segunda columna de GC.

Desde su introducción en 1991, se ha producido un rápido desarrollo en la técnica GC×GC. Hoy en día, la GC×GC se emplea de manera habitual en muchos campos diferentes, abarcando una amplia variedad de aplicaciones. La cromatografía de gases bidimensional completa acoplada a espectrometría de masas (GC×GC/MS) es una tecnología emergente que ofrece una separación bidimensional y un perfil espectral de masas completo, basado en las coordenadas del tiempo de retención de los compuestos, dentro del espacio de separación bidimensional. Los mecanismos de separación en ambas dimensiones tienen que ser completamente independientes entre ellos, es



decir, ortogonales. La GC×GC posee todas las ventajas de la técnica de la cromatografía de gases unidimensional («normal»): está completamente automatizada y es un análisis sensible de compuestos detectados. Asimismo, las separaciones en 2D ofrecen mejor información y los compuestos con una estructura similar eluyen juntos en el cromatograma 2D.

Al principio, la muestra introducida se separa en una columna de GC capilar de alta resolución (1D), como en la GC clásica. Después de salir de la primera columna, el eluyente no se dirige directamente al detector, sino al modulador, donde se concentra en fracciones pequeñas, a intervalos regulares y cortos, en un período menor que el de un ancho del pico principal, según el índice de modulación empleado. A continuación, el material recogido se inyecta en una segunda columna para la separación rápida de la segunda dimensión (2D) (Figura 1.13). Después de inyectar la muestra, se reinicia la recogida en el modulador de la nueva fracción. Al mismo tiempo, los componentes anteriores de las fracciones se separan mediante la segunda columna. El mecanismo de esta separación es independiente de la separación de la primera columna. Es entonces posible separar los componentes que coeluyen al final de la primera columna. Los compuestos al salir de la segunda columna se dirigen al detector, lo que genera un registro de una serie de cromatogramas muy cortos (por ejemplo, 3s). El proceso de recogida de fracciones y dosificación de la muestra en la columna en 2D se repite hasta el final del análisis. Esto mantiene la separación de picos alcanzada en la primera columna, proporcionando simultáneamente una separación adicional de analitos en la segunda columna.

Las columnas de la cromatografía en ambas dimensiones deberían caracterizarse por distintos mecanismos de retención, es decir, deben ser columnas de distinta selectividad. Para una combinación no polar/polar (NP/P), los compuestos que coeluyen pueden presentar puntos de ebullición similares en el momento de la elución, pero con distinta «polaridad». La separación en 2D se basa en las diferencias en los coeficientes de actividad entre los solutos y la fase polar, permitiendo que se separen compuestos de distinta polaridad. Además de la mayor capacidad de picos, la ventaja que presenta este enfoque es que la dimensión de separación adicional reduce la sobrecarga de picos eluidos, permitiendo así la resolución de especies menos abundantes.

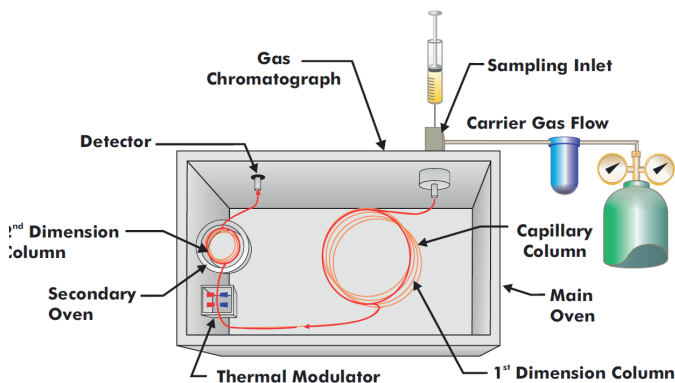
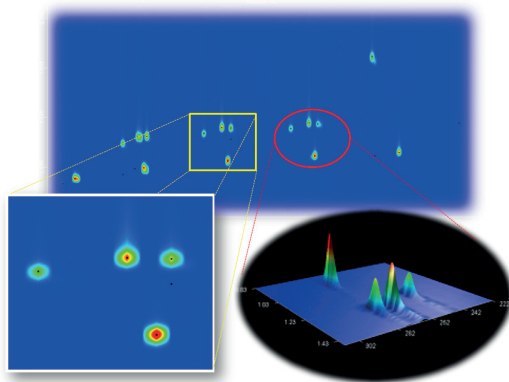


Figura 1.13. Diagrama del sistema GC×GC (*de.leco-europe.com*).

Los cromatogramas obtenidos a través de GCxGC contienen mucha más información que en 1D, lo que permite una identificación más fácil y precisa de sustancias desconocidas. En consecuencia, es necesario un detector muy eficiente. El detector tiene que ser capaz de recoger datos a una frecuencia de por lo menos 50 Hz. Para reproducir exactamente la forma de un solo pico, hay que registrar al menos diez veces el pico, y los picos eluidos de la segunda columna tienen un ancho de base mínimo de 150-400 ms. La selección de detectores cromatográficos que podrían satisfacer este criterio resulta insuficiente. Los detectores más utilizados en cromatografía GCxGC son el FID (máx. 300 Hz), el  $\mu$ -ECD (50-100 Hz) y el TOF (máx. 500 Hz).

En conclusión, un plano de retención tiene más capacidad de picos que un eje de retención y, en consecuencia, puede acomodar y resolver las mezclas altamente complejas asociadas, por ejemplo, a la determinación de perfiles de muestras biológicas complejas, perfiles de metabolitos, etc. En general, las posibilidades de uso de GCxGC se pueden dividir en tres áreas: análisis de huella metabólica de matrices muy complejas, análisis dirigido o clásico e identificación de compuestos desconocidos. A continuación se muestra en la Figura 1.14 un ejemplo de un cromatograma 2D de GC.



**Figura 1.14.** Cromatograma de la mezcla de compuestos volátiles (alcanos, alcoholes, aldehídos, cetonas y aminas) analizados en GC 2D [Pegasus 4D; Leco].

#### 2.2.2.4. Pretratamiento y derivatización de muestras

Se trata de la etapa más importante y difícil en el análisis GC, y que resulta crucial para la precisión del resultado, especialmente por lo que respecta a las muestras de matrices complejas. El pretratamiento de muestras para el análisis GC suele incluir las siguientes etapas: limpieza, cambio de la matriz, extracción, derivatización y preconcentración. Los problemas obvios de la GC y la GC-MS residen en el contenido de las muestras en componentes no volátiles y la limpieza de las muestras, lo que está relacionado con el enriquecimiento de ingredientes traza. No existe un único método universal para la preparación de muestras que resulte apropiado para todos los materiales analizados. La preparación de la muestra depende de numerosos factores,

principalmente si se encuentra en estado gaseoso, líquido o sólido.

Antes del análisis GC se pueden llevar a cabo distintos métodos de extracción de compuestos volátiles y semivolátiles de material biológico: extracción clásica líquido-líquido (LLE), extracción continua líquido-líquido (LLCE), extracción de fluidos supercríticos (SFE), extracción líquida presurizada (PLE), extracción asistida por microondas (MAE) y extracción asistida por ultrasonido (USE) (Figuras 1.15 y 1.16), y la mayoría de ellas se han automatizado completamente. De entre ellas, la extracción de fase sólida (SPE) permite una reducción significativa del tiempo de preparación de la muestra, además del volumen de disolventes utilizados y de residuos peligrosos. En este método, los analitos son extraídos en un sistema líquido-sólido, según la distribución del analito entre la muestra líquida y el sorbente sólido.

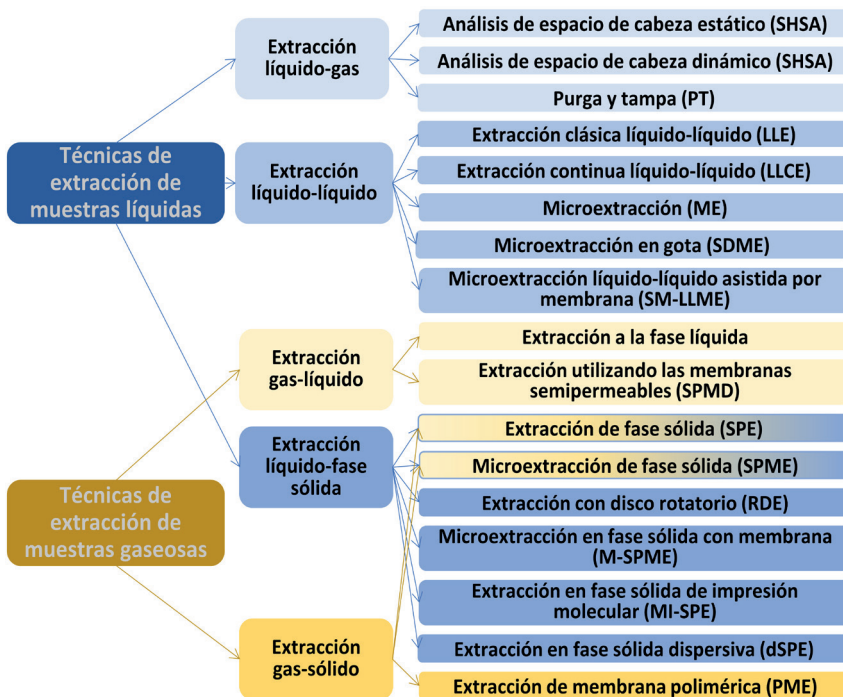


Figura 1.15. Métodos de extracción de muestras líquidas y gaseosas.

El principio de separación por SPE depende principalmente de la naturaleza del sorbente. Las interacciones entre el analito y los adsorbentes sólidos polares (por ejemplo, gel de sílice, óxido de aluminio) se determinan mediante enlaces de hidrógeno así como fuerzas dipolo-dipolo, dipolo-dipolo inducido y de dispersión (fuerzas van der Waals). Este tipo de extracción se basa en los mismos principios que la cromatografía de adsorción. Cuando el adsorbente es sílice con un grupo

polar enlazado químicamente, por ejemplo, un grupo amino (sistema de fase normal), o no polar, por ejemplo, un grupo octadecil (de fase inversa), el principio de separación es el mismo que en la cromatografía de partición. Las alternativas a la SPE son las técnicas de microextracción, a saber, la SPME (microextracción de fase sólida) y LPME (microextracción de fase líquida). La extracción SPME comprende la sorción de cantidades microscópicas de compuestos orgánicos sobre una capa cilíndrica fina de la fase estacionaria que recubre la fibra de vidrio o cuarzo. La LPME, en cambio, es un procedimiento de pretratamiento de muestras que minimiza el uso de disolventes, puesto que tan solo requiere unos pocos microlitros de disolvente. Una ventaja esencial de los métodos de microextracción es el bajo límite de detección (5-50 ppt) y el poco tiempo de preparación de la muestra (aprox. 10 minutos). Para el análisis de compuestos volátiles también se utilizan las técnicas de «espacio de cabeza» (HS) y de «purga y trampa» (PT).

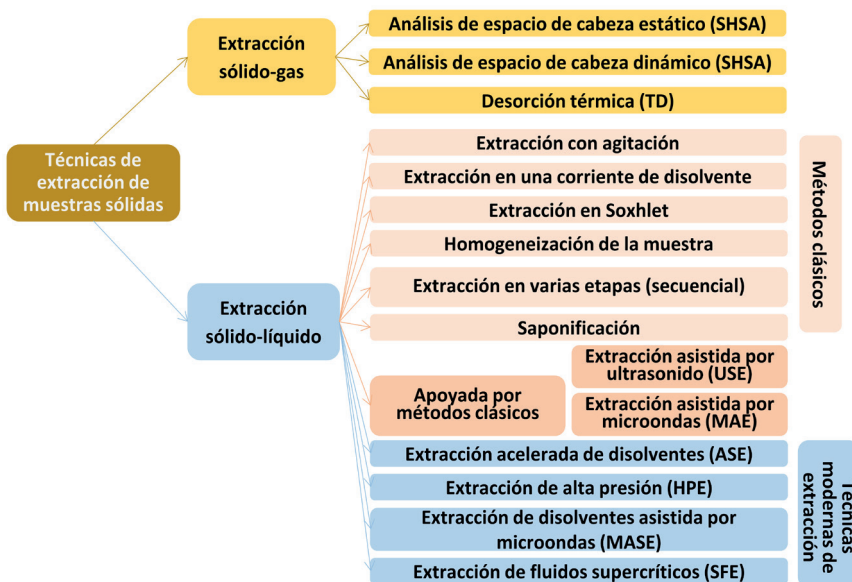


Figura 1.16. Métodos de extracción de muestras sólidas.

El análisis directo de mezclas de compuestos en GC es complicado a causa de las interacciones entre los compuestos o entre los compuestos y la fase estacionaria de la columna de GC. Estas interacciones pueden provocar una mala resolución del pico y la aparición de picos con cola («tailing»), lo que hace que resulte poco práctico o complicado realizar una identificación de pico apropiada. La conversión a productos derivados reducirá las interacciones que interfieren con el análisis. También se recurre a la derivatización para cambiar la volatilidad, incrementar la estabilidad térmica, evitar la degradación durante el proceso cromatográfico, incrementar

la sensibilidad o especificidad del ensayo introduciendo grupos funcionales apropiados o bloqueando los grupos funcionales de los analitos. Para la derivatización de analitos se utilizan compuestos orgánicos, normalmente ejecutando reacciones de alquilación, acilación o sililación en compuestos que contienen grupos funcionales con hidrógenos activos, por ejemplo, -COOH, -OH, -NH y -SH (Tabla 1.7).

**Tabla 1.7.** Derivatización previa.

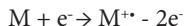
Procedimiento	Grupo funcional – Tipo de compuesto	Derivado	Reactivo
Sililación	-OH - alcoholes, fenoles	Trimetilsilil éteres	Bistrimetilsililacetamida (BSA)
	-CO -cetonas, oximas, esteroides		Bistrimetilsililfluoroacetamida (BSTFA)
	-COOH - aminoácidos , ácidos grasos, cannabinoles, esteroides		Hexametildisilazano (HMDS)
	-(CH <sub>2</sub> OH) <sub>n</sub> -azúcares		N- metil-N-t-butildimetilsilil-trifluoroacetamida (MTBSTFA)
	-NH <sub>2</sub> , -NH -aminas, nucleósidos de urea		N-metiltrimetilsililtrifluoroacetamida (MSTFA)
	-SH -mercaptanos		Trimetilclorosilano (TMCS)
	-CONH <sub>2</sub> , -CONH <sub>2</sub> -imidias, proteínas	Trimetilsilil amidias	Trimetilsilildietilamina (TMA -DEA)
			Trimetilsilimidazol (TMSI)
			Reactivos halo-metilsilios
Alquilación	-OH - alcoholes, fenoles	Ésteres metílicos (DMF)	Bromuro de Bencilo
	-CO -aldehidos	Trifluoroacetatos (TFAA)	Trifluoruro de boro (BF <sub>3</sub> ) en metanol o butanol
	-COOH - aminoácidos, ácidos grasos	Ésteres metílicos (BF <sub>3</sub> -metanol)	Diazometano (N <sub>2</sub> CH <sub>2</sub> )
	-NH <sub>2</sub> , -NH <sub>2</sub> -aminas, aminoazúcares	Éteres de pentafluorobencilo (PFBBr)	Dimetilformamida (DMF)
	-CONH <sub>2</sub> -amidias	Metilamidias (TMAH)	Bromuro de pentafluorobencilo (PFBBr)
	-SH -mercaptanos	Ésteres metílicos (DMF)	Pentafluorobencil-hidroxilamina clorhidrato (PFBHA)
			Anhídrido de tetrabutilamonio (TFAA)
Acilación	-OH - alcoholes, fenoles	Pentafluoropropionados (PFPA)	Hidróxido de trimetilanolino (TMAH)
	-(CH <sub>2</sub> OH) <sub>n</sub> -azúcares	Trifluoroacetamidias (MBTFA)	Anhídrido heptafluorobutirico (HFBA)
	-NH <sub>2</sub> , -NH <sub>2</sub> -aminas	Trifluoroacetamidias (TFAA)	N-Metil-bis(trifluoroacetamida) (MBTFA)
	-CONH <sub>2</sub> -amidias	Trimetilsilil éteres (MBTFA)	Cloruro de pentafluorobenzilo (PFBCl)
	-SH -mercaptanos		Pentafluoropropanol (PFPOH)
			Pentafluoropropanilimidazol (PFPI)
			Anhídrido pentafluoro propiónico (PFPA)
			Anhídrido trifluoroacético (TFAA)
		Trifluoroacetilimidazol (TFAI)	

Los reactivos de sililación son los más utilizados en GC. Estos reactivos incluyen, entre otros, trimetilclorosilano (TMCS), bistrimetilsililtrifluoro-acetamida (BSTFA), N-metil-trimetilsililtrifluoroacetamida (MSTFA), N-metil-N-t-butildimetilsililtrifluoro-acetamida (MTBSTFA). Los compuestos que contienen varios grupos funcionales distintos pueden precisar de múltiples etapas de derivatización.

### 2.2.3. Acoplamiento GC-MS

En GC-MS, la esencia del análisis depende del proceso de ionización de la molécula, con la consiguiente fragmentación. El método de ionización más empleado en GC-MS es el Impacto Electrónico o Electro-ionización (EI), de la que hemos hablado en la sección 1.2. Para obtener

espectros MS reproducibles se emplea una energía de 70 eV. En la fuente de EI, las moléculas del analito pierden electrones de valencia para generar cationes radicales y, seguidamente, sufren fragmentación de acuerdo con determinadas reglas de fragmentación:



Además de la fuente convencional de iones de la EI para la GC-MS, se puede recurrir también a la ionización química (CI, por sus siglas en inglés). El método de ionización es relativamente blando, y normalmente se puede observar el ion molecular de los analitos. Sin embargo, es menos universal, y se ionizan menos compuestos cuando se emplea este método de ionización, por lo que su uso no está tan extendido. Los iones formados de la cámara de ionización se extraen a través de una serie de electrodos que concentran los iones y los aceleran hasta un analizador de masas.

Toda vez que los analizadores de masas separan iones cargados en función de su relación  $m/z$ , registrando la masa y abundancia de los distintos iones, en los equipos de GC-MS se pueden utilizar muchos tipos de analizadores de masas, en particular, los espectrómetros de masas en tándem, como ya se dijo antes. La GC-MS es una herramienta esencial para la identificación y cuantificación de metabolitos por los patrones de fragmentación molecular reproducibles que genera. Si bien la aplicación de la GC se limita a compuestos volátiles, una parte significativa de pequeños metabolitos moleculares se sitúa dentro del ámbito de actuación de la separación GC. Con todo, dado que la mayoría de los metabolitos tienen grupos funcionales polares, la forma de los picos y las recuperaciones que se obtienen se consideran con frecuencia inaceptables debido a la adsorción de la columna. En consecuencia, normalmente es necesario proteger estos grupos funcionales de metabolitos con derivatización química.

**Para más información:**

H.J. Hübschmann. Handbook of GC-MS: Fundamentals and Applications (3rd Edition). Somerset, NJ, USA: Wiley, 2015. ProQuest ebrary. Web. 8, 2015.

X. Guo. Advances in Gas Chromatography, Chapter 4: "Gas Chromatography in Metabolomics Study" de Y. Qiu y D. Reed. InTech, 2014.

## 2.3. Electroforesis capilar-espectrometría de masas (CE-MS)

### 2.3.1. Introducción a la CE-MS

La electroforesis capilar acoplada a la espectrometría de masas (CE-MS) es el resultado del acoplamiento de una técnica de separación de alta eficacia y resolución con las potentes capacidades de la espectrometría de masas, y se considera una técnica ortogonal para la GC o la LC. La técnica se introdujo a finales de la década de los ochenta y representa justamente el acoplamiento más novedoso que puede ofrecer información importante sobre la composición de pequeños compuestos como los metabolitos polares y cargados. Es por este motivo, que se ha venido utilizando en distintas áreas de investigación de la metabolómica, como la metabolómica biomédica, clínica y vegetal. A través de

la CE-MS se pueden analizar muestras de sangre (suero y plasma), orina y otros biofluidos, cultivos celulares, tejidos y plantas con un tratamiento de la muestra mínimo. A través de la CE-MS se analizan con frecuencia muchos tipos distintos de analitos, incluidos aminoácidos, ácidos orgánicos de cadena corta (SCOA, por sus siglas en inglés), derivados de ácidos grasos (acilcarnitinas), nucleótidos, fosfatos de azúcares, nucleótidos y compuestos vinculados al metabolismo de las poliaminas. Además, también pueden ser objeto de análisis pequeños péptidos y proteínas (después de hidrólisis o proteínas nativas).

### **Ventajas de la CE-MS**

La CE-MS es muy apta para muestras acuosas. Por otra parte, resulta muy apropiada para muestras con una alta concentración de sales, como medios de cultivos o extractos celulares.

- Se requiere un volumen de muestra bajo. Esto es muy importante en muchos experimentos cuando solo se dispone de una mínima cantidad de muestra (animales pequeños, lágrimas, líquido cefalorraquídeo y otros fluidos similares, etc.).
- Adecuada para analitos hidrofílicos.
- Separación muy eficiente y un mecanismo ortogonal respecto a las técnicas cromatográficas.
- Una vez detectado el analito, se puede lavar el capilar muy fácilmente y acondicionarlo para el análisis siguiente.
- Es una técnica analítica verde. El consumo de reactivos y disolventes orgánicos es muy bajo y menor que el que se necesita en la LC.

### **Inconvenientes de la CE-MS**

- Menor reproducibilidad y robustez de la CE en lo que respecta a los tiempos de migración y la respuesta del detector.
- Baja sensibilidad debido a la cantidad de muestra introducida en el sistema, en torno a nL.
- Se considera el acoplamiento menos robusto.
- No se dispone de librerías de espectros. Identificación de compuestos desconocidos a partir de librerías y bases de datos de masas precisos disponibles a nivel interno.

## **2.3.2. Principios teóricos de la CE-MS**

La separación por electroforesis capilar se basa en una separación inducida por electrones dentro de los capilares de sílice de poros estrechos, normalmente con una ID de 50  $\mu\text{m}$  y 100 cm de longitud. El movimiento de los analitos a lo largo del capilar lleno de un medio líquido conductor y bajo la influencia del campo eléctrico, se basa en la relación carga/masa, y esto proporciona a esta técnica sus propiedades únicas, con sus fortalezas e inconvenientes. Normalmente se utiliza sílice fundida sin recubrir además de recubierta (recubrimientos neutros o positivos) para distintas aplicaciones, aunque es más común la primera.

Los tiempos de migración, y no los tiempos de retención, y los electroferogramas (representación de la respuesta del detector en función de los tiempos de migración) son algunos términos utilizados habitualmente.

## PRINCIPIOS DE SEPARACIÓN

Hay dos fuerzas que regulan el análisis CE: El flujo electroosmótico (EOF) y la movilidad electroforética (EM). El flujo electroosmótico es la consecuencia de la carga de superficie sobre la pared interna del capilar. Con los capilares sin recubrir, los grupos silanoles (Si-OH) en la superficie se ionizan parcialmente a grupos silanatos cargados negativamente (Si-O<sup>-</sup>) a pH > 2. Estas cargas negativas atraen a las cargas positivas del tampón del capilar formando dos capas internas tal y como se muestra en la Figura 1.17. La «capa fija» se compone de cargas positivas más próximas a iones negativos, mientras que la «capa móvil» está conformada por cargas más dispersas. Cuando se aplica el campo eléctrico, la capa móvil de cationes es atraída al electrodo negativo, los cationes se solvatan y arrastran con ellos a la totalidad de la disolución del tampón. El resultado es un flujo neto dedisolución de tampón en la dirección del electrodo negativo, lo que provoca el flujo electroosmótico. Este tipo de análisis se denomina «electroforesis capilar normal». Existe también otro modo llamado «electroforesis capilar invertida», en el que se cambia la polaridad de los electrodos. En este caso, el electrodo de entrada es el cátodo, y se debe suprimir o invertir el flujo, por ejemplo, añadiendo aminas cuaternarias, o utilizando capilares recubiertos con carga positiva.

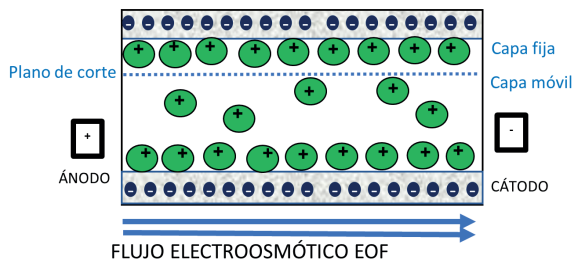


Figura 1.17. Representación del flujo electroosmótico de un capilar.

Entre las dos capas existe un plano de corte que se caracteriza por una diferencia potencial, denominada «potencial zeta» ( $\zeta$ ), debido a un desequilibrio eléctrico de las cargas. El potencial zeta es proporcional al flujo electroosmótico y al grosor de la doble capa, e inversamente proporcional a la constante dieléctrica del tampón, según se escribe en la ecuación siguiente:

$$\zeta = 4\pi\delta e/\epsilon$$

donde  $\delta$  es el grosor de la capa doble difusa,  $e$  es la carga por área de superficie unitaria y  $\epsilon$  es la constante dieléctrica del tampón.

Existen diferentes parámetros que influyen en la velocidad del EOF, como el pH del tampón, la tensión y la temperatura. Un aumento del pH del tampón produce un aumento del EOF a causa de la disociación del Si-OH a Si-O<sup>-</sup> en la pared capilar interna que incrementa el potencial zeta ( $\zeta$ ). El campo eléctrico (tensión por unidad de longitud) tiene un efecto significativo sobre la velocidad del flujo electroosmótico, cuanto mayor sea el campo eléctrico, mayor será la velocidad del flujo



electroosmótico e influye directamente también sobre el flujo electroforético, lo que reduce además el tiempo de análisis. No obstante, una tensión alta puede producir el calentamiento por efecto Joule, que aumentaría la temperatura dentro del capilar. Es importante controlar este parámetro porque una temperatura alta podría destruir y desnaturalizar las muestras, y, además, se podría formar el gradiente de viscosidad del tampón de manera perpendicular al capilar, lo que incrementaría la difusión y la dispersión. La ecuación de Smoluchowski define la velocidad del flujo electroosmótico:

$$v_{\text{eof}} = \epsilon \zeta E / 4 \pi \eta$$

donde  $v_{\text{eof}}$  es la velocidad del flujo electroosmótico;  $\epsilon$  es la constante dieléctrica del tampón;  $E$  es el campo eléctrico aplicado en voltios/cm;  $\eta$  es la viscosidad del tampón,  $\zeta$  es el potencial zeta medido en el plano de corte.

El tampón tiene movilidad electroosmótica ( $\mu_{\text{eof}}$ ) que depende de las características del tampón, al igual que la constante y la viscosidad dieléctricas, y es independiente del campo eléctrico aplicado tal y como se describe en la ecuación siguiente:

$$\mu_{\text{eof}} = \epsilon \zeta / 4 \pi \eta$$

donde  $\mu_{\text{eof}}$  es la movilidad electroosmótica del tampón,  $\epsilon$  es la constante dieléctrica,  $\eta$  es la viscosidad del tampón;  $\zeta$  es el potencial zeta medido en el plano de corte.

El perfil del flujo electroosmótico es diferente del de un fluido que se mueva bajo presión (como en la HPLC). En la figura 1.18, el primer diagrama muestra el perfil plano del flujo en la CE causado por el movimiento de todas las cargas hacia los electrodos. El segundo diagrama muestra el perfil en forma parabólica del flujo en la HPLC causado por la fricción de la pared de la columna que ralentiza el flujo más próximo a los lados de la columna. La ventaja del perfil del flujo plano es que todas las moléculas de solutos tienen la misma velocidad dentro del capilar, obteniéndose así picos estrechos con una alta eficiencia. En el perfil del flujo con forma parabólica de la HPLC, las moléculas de solutos del centro del tubo se mueven más rápido que aquellas más próximas a la pared de la columna, obteniéndose picos relativamente anchos.

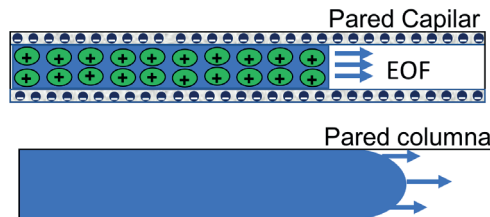


Figura 1.18. Diagramas de flujo en CE y HPLC.

La movilidad electroforética es una fuerza adicional presente solo en moléculas cargadas y que puede arrastrar a los iones hacia el electrodo con la carga opuesta. La velocidad con la que se mueve el soluto en función de la carga y bajo un campo eléctrico aplicado se obtiene con la ecuación:

$$v_{ep} = \mu_{ep} E$$

donde  $v_{ep}$  es la velocidad electroforética del soluto,  $\mu_{ep}$  es la movilidad electroforética del soluto y  $E$  es un campo eléctrico aplicado.

La movilidad electroforética ( $\mu_{ep}$ ) es directamente proporcional a la carga del soluto e inversamente proporcional al tamaño molecular y también a la viscosidad del medio electroforético, tal y como se describe en la ecuación siguiente:

$$\mu_{ep} = Q/6\pi r\eta$$

donde  $\mu_{ep}$  es la movilidad electroforética del soluto,  $Q$  es la carga del soluto;  $r$  es el radio iónico del soluto, y  $\eta$  es la viscosidad del medio.

La última ecuación demuestra que la razón por la que las moléculas pequeñas con una alta carga se mueven a través del capilar más rápido que las moléculas grandes con una carga más baja es la diferencia de movilidad electroforética ( $\mu_{ep}$ ) que presentan entre ellas.

La movilidad total del soluto ( $v_{tot}$ ) dentro del capilar se define mediante la ecuación:

$$v_{tot} = v_{ep} + v_{eof}$$

En la electroforesis capilar con una polaridad normal, todas las moléculas tienen la misma velocidad de flujo electroosmótico ( $v_{eof}$ ), tal y como se ve en la figura 1.19, pero eluyen en momentos distintos. Los cationes eluyen primero porque tienen velocidad electroforética ( $v_{ep}$ ) dirigida al mismo electrodo del flujo electroosmótico; las moléculas neutras tienen la misma velocidad del flujo electroosmótico porque no están cargadas; y por último eluyen los aniones porque la dirección de la velocidad electroforética ( $v_{ep}$ ) apunta al electrodo opuesto (ánodo).

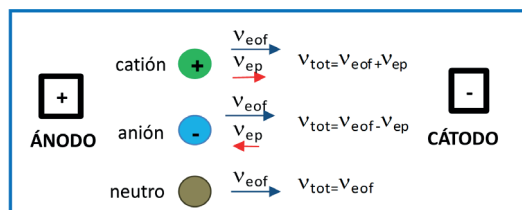


Figura 1.19. Movilidad total del soluto dentro del capilar.

Existen numerosos factores que influyen en las separaciones electroforéticas: el pH del tampón de separación, que influye en la carga y el radio de las moléculas solvatadas; la concentración del tampón, que el EOF disminuye con la raíz cuadrada de su valor; la viscosidad del medio de separación; la temperatura; el campo eléctrico aplicado, etc.

En la CE se pueden distinguir distintos modos de separación: electroforesis capilar zonal (CZE), cromatografía capilar electrocinética micelar (MEKC), electrocromatografía capilar (CEC), electroforesis capilar en gel (CGE), entre otros. La CZE es apta para compuestos polares y cargados, y se trata del modo más habitual de CE, principalmente cuando se acopla a la MS. La CE es también apta para compuestos neutros que utilizan micelas en el electrolito de fondo (MECK). Existen otros aditivos que pueden modificar la selectividad de la separación; por ejemplo, las ciclodextrinas para separaciones quirales.

### 2.3.2.1. Electroforesis de zona capilar

Es el modo más fácil y más extendido, pero solo se pueden separar compuestos con carga. La sección anterior se ocupó casi totalmente de esta forma. Los analitos se mueven de un extremo del capilar al otro de acuerdo con la suma vectorial de la movilidad electroforética y electroosmótica.

### 2.3.2.2. Cromatografía capilar electrocinética micelar

Introducida por Terabe en 1984 para la separación de compuestos neutros en lugar de la CZE. La MECK incluye detergentes, generalmente con carga, en el tampón de separación y en una concentración suficientemente alta para formar micelas. Estos detergentes pueden ser compuestos catiónicos, aniónicos, no iónicos o zwitteriónicos, aunque cuando se acoplen a la CE-MS tienen que ser volátiles. Estos agregados macromoleculares con un núcleo interno hidrofóbico y su superficie polar pueden interactuar con el analito mediante otro mecanismo como fuerzas de reparto y de Coulomb. Los compuestos más hidrofóbicos interactúan más claramente con la micela y su velocidad a través del capilar es menor. Cuando no interactúa con la micela, la molécula del analito migrará con el EOF. En todos los casos se puede recurrir a variaciones en la concentración del tampón, en el pH, la temperatura o al uso de aditivos como la urea, iones metálicos o selectores quirales, para afectar la selectividad. Al igual que en la CZE, se han utilizado, con resultados satisfactorios, modificadores como el metanol, el acetonitrilo y el 2-propanol.

### 2.3.2.3. Electrocromatografía capilar

Es una forma de cromatografía líquida miniaturizada. Los capilares se llenan con partículas de fase inversa con base de sílice de tipo HPLC, de 1-5  $\mu\text{m}$  de tamaño, haciendo las veces de fase estacionaria. Por otra parte, se utilizan fases móviles convencionales para las separaciones de tipo RP, como disoluciones tampón en mezclas orgánico/agua. Actúa un mecanismo de reparto de los solutos entre la fase móvil y la estacionaria, y con metabolitos

cargados, la velocidad electroforética adicional modificará la separación. El sistema requiere una bomba adicional para empujar la fase móvil a través del capilar lleno. El EOF generado en el capilar lleno de partículas tiene un perfil de velocidad de flujo idéntico al de un capilar tubular abierto (CZE).

#### 2.3.2.4. Electroforesis capilar en gel

Es utilizada ampliamente en la investigación biológica molecular para moléculas diferenciadas únicamente por tamaño y no por su relación masa/carga, como las proteínas y los ácidos nucleicos (ADN). El capilar se llena con un gel o disolución viscosa. En la CGE los polímeros pueden estar reticulados covalentemente, con enlaces de hidrógeno o simplemente disueltos en soluciones tampón. Con frecuencia se suprime el EOF, por lo que la migración de los analitos solamente se produce mediante electroforesis. En un medio viscoso o de red polimérica, el movimiento de los analitos se ve entorpecido, ya que las moléculas más grandes se ralentizan más que las moléculas más pequeñas. Para obtener una relación carga/masa más uniforme con las proteínas se suele realizar un pretratamiento con el detergente SDS. Normalmente las proteínas se unen a un número constante de moléculas de SDS por unidad de longitud, de manera similar a la electroforesis del gel de SDS-poliacrilamida.

#### 2.3.3. Instrumentación

El sistema de electroforesis capilar consiste en un capilar, un vial de entrada, un vial de salida, un vial de muestra, un sistema de refrigeración capaz de eliminar el calor producido dentro del capilar, dos electrodos (un ánodo y un cátodo), una fuente de alto voltaje y el detector, tal y como se muestra en la Figura 1.20. El capilar, el vial de entrada y el vial de salida se llenan con el mismo tampón, que suele ser una disolución acuosa. Normalmente el ánodo y el cátodo son electrodos de paladio, conectados al suministro eléctrico, y el detector se suele situar cerca del final del capilar (próximo al vial de salida).

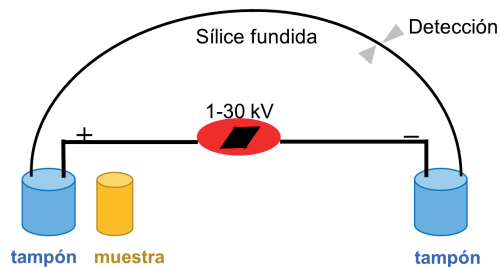


Figura 1.20. El esquema de la electroforesis capilar.

Trabajando en modo CZE en CE-MS se utiliza un capilar de sílice fundida. Este capilar puede ser un capilar desnudo o puede estar recubierto con un polímero (neutros o positivos) la pared interna para reducir la electroósmosis a niveles muy bajos, o incluso invertirla. El capilar se coloca dentro de los dos viales donde está también el electrodo, y seguidamente se carga la muestra. La carga se produce sustituyendo uno de los viales con el vial de muestra e inyectando la muestra mediante gravedad, aplicando una presión externa (hidrodinámica) o aplicando un campo eléctrico (inyección electrocinética). A continuación, se coloca el capilar dentro del vial de tampón y se aplica el campo eléctrico para la separación de los componentes. Existen diferentes sistemas de refrigeración disponibles capaces de reducir el calor producido por causa del efecto Joule dentro del capilar mientras se aplicaba el alto voltaje.

Los sistemas de detección más utilizados en la electroforesis capilar son el detector Ultravioleta/Visible (UV-VIS), el detector de matriz de fotodiodos (PDA), detector de fluorescencia inducida por láser (LIF) y la espectrometría de masas (MS) (tabla 1.8). Con mucha diferencia, las mayores capacidades se demuestran mediante la CE-MS, por cuanto se puede obtener información sobre el tiempo de migración bajo las condiciones de análisis y el espectro de masas de los compuestos que llegan al detector. La fuente de ionización más habitual cuando se emplean detectores de espectrometría de masas es la ionización por electrospray (ESI).

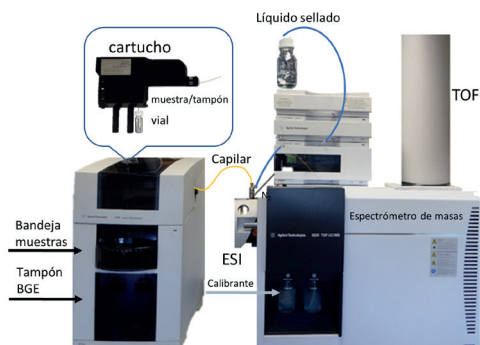
**Tabla 1.8.** Diferencias entre detectores en la CE.

Detector	Ventajas	Características	Límite del rango de detección (M)
Absorbancia UV/vis	<ul style="list-style-type: none"> <li>• La posibilidad de detección directa e indirecta</li> <li>• Detector muy habitual</li> </ul>	- universal	-10 <sup>-3</sup> -10 <sup>-6</sup> para la detección de compuestos aromáticos
LIF (fluorescencia inducida por láser)	<ul style="list-style-type: none"> <li>• Información cualitativa y cuantitativa</li> <li>• Altamente sensible y altamente selectivo</li> </ul>	- selectivo	-10 <sup>-6</sup> - 10 <sup>-9</sup>
MS	<ul style="list-style-type: none"> <li>• Información cualitativa y cuantitativa</li> <li>• Altamente sensible y altamente selectivo</li> </ul>	- universal - selectivo	~10 <sup>-5</sup> (depende del tipo de MS y de metabolitos)

Cuando se combina con otras técnicas de detección, como la UV o la fluorescencia, la entrada y la salida del capilar se sumergen en los viales con el electrolito de fondo (BGE, por sus siglas en inglés). Sin embargo, cuando se utiliza la MS como detector, el vial de salida se sustituye con la entrada a la MS. En este caso es necesario utilizar una interfaz que se ajuste a las dos tensiones y mantener un spray estable con un volumen muy bajo de fluido (Figura 1.21). En la CE-ESI-MS, el volumen introducido en el capilar se mide en nanolitros y, para mantener un fluido y spray constantes, se suele utilizar un fluido o líquido adicional (SL, del inglés sheath flow o sheath liquid). El flujo adicional, añadido en un rango de entre nL/min y mL/min, permite la formación de un spray e iones estables, y normalmente se compone de una

mezcla 1:1 de agua-metanol con un 0,1% de ácido acético o ácido fórmico. Conseguir que el electrospray funcione de una forma estable depende de un equilibrio de múltiples parámetros, como la posición de los capilares, la velocidad y composición del flujo del líquido adicional, la velocidad del flujo adicional de gas y las condiciones de la ESI. A pesar de los numerosos desarrollos de interfaces sin flujo adicional (sheathless) con una sensibilidad más alta demostrada, continúa estando comprometida la robustez.

Con respecto al analizador de masas, todos los instrumentos comercialmente disponibles se pueden conectar a la CE, siendo los instrumentos analizadores de tiempo de vuelo (TOF) los que tienen mayor número de aplicaciones publicadas. Esto es así, principalmente, por la alta velocidad de escaneo necesaria para obtener un número suficiente de puntos a lo largo del pico, pero también por la exactitud de las masas y la alta resolución de estos instrumentos.



**Figura 1.21.** Representación gráfica de la CE acoplada a la MS, donde se describen las partes más importantes del equipo.

#### Para saber más:

D.R. Baker. Capillary electrophoresis. Wiley, 1995.

S. Terabe, K. Otsuka, T. Ando. Electrokinetic chromatography with micellar solution and open-tubular capillary. *Anal. Chem.*, 1985, 57 (4):834-841.

T. Soga, Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita, T. Nishioka. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J Proteome Res.*, 2003, 2(5):488-94.

A. García, S. Naz, C. Barbas. Metabolite fingerprinting by capillary electrophoresis-mass spectrometry. *Methods Mol Biol*, 2014, 1198:107-23.

A. García, J. Godzien, Á. López-González, C. Barbas. Capillary electrophoresis mass spectrometry as a tool for untargeted metabolomics based on GC-(EI)-Q-MS. *Anal. Chim. Acta*, 2015, 900:21-35.

## 3. Sesiones prácticas

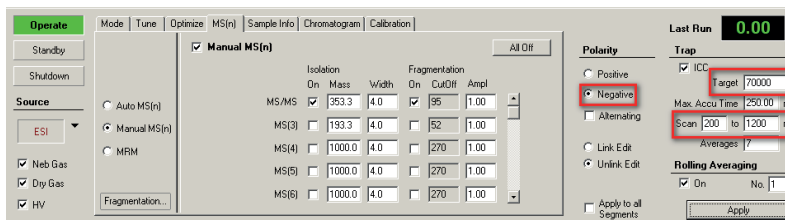
### 3.1. Adquisición de un espectro de masas ESI-MS

Tras la presentación de los principios operativos de una espectrometría de masas con trampa de iones con una fuente de iones ESI, se analizará una disolución de la muestra utilizando una bomba de infusión directa. Se analizará una disolución conteniendo rutina, un antioxidante de bajo peso molecular con una masa molecular conocida (o unadisolución de otro estándar comercialmente disponible) para obtener un espectro MS. La adquisición se llevará a cabo tanto en el modo positivo como negativo para escoger las mejores condiciones, en función de las propiedades de la rutina.

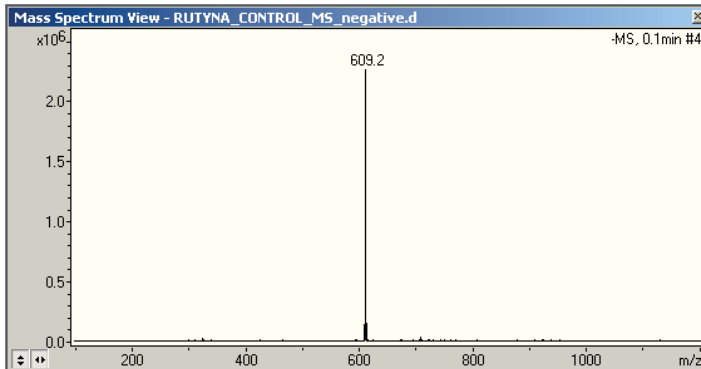
Asimismo, tras la presentación del principio operativo y la construcción básica del espectrómetro de masas Orbitrap con fuente de iones ESI, se analizará una muestra de tiorredoxina mediante infusión directa. Esta etapa permite introducir a los alumnos espectros de masas de compuestos de alto peso molecular, con la posibilidad de observar iones con múltiples cargas. Los espectros ESI-MS obtenidos se registrarán y utilizarán con fines de interpretación en el punto 3.3.

#### Experimento de ESI-MS:

1. Prepare una disolución en dimetilsulfóxido (DMSO) de rutina (masa molecular 610,52) a una concentración de 1 mg/mL
2. Disuelva con ACN/H<sub>2</sub>O (60/40 v/v) para obtener unadisolución de rutina de 10ug/mL (en 1% de DMSO).
3. ¿Por qué se requiere la disolución de tan solo el 1% de DMSO?
4. Introduzca la disolución de muestra en la fuente de ESI con bomba de jeringa de infusión directa preparada a una velocidad de flujo de 150µl/h.



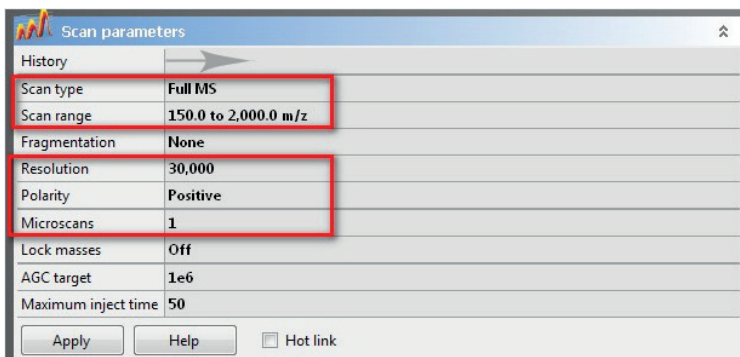
5. Registre el espectro de masas durante 30 s utilizando los parámetros que se facilitan a continuación en la figura.
6. Encuentre el ion molecular de la rutina y explique el valor  $m/z$  correspondiente



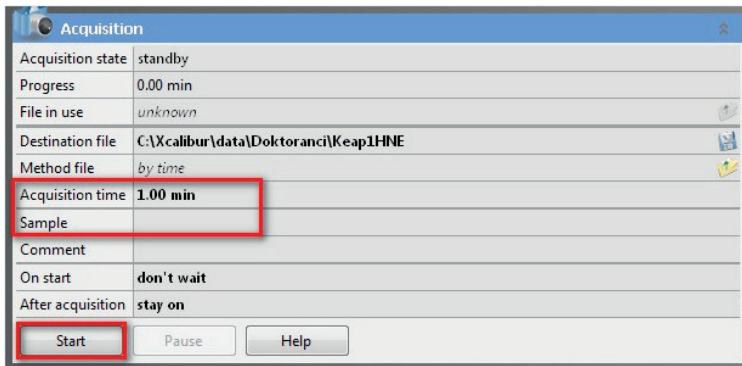
Espectros ESI-MS de la rutina en modo negativo

### Experimento de MS ESI-OrbiTrap:

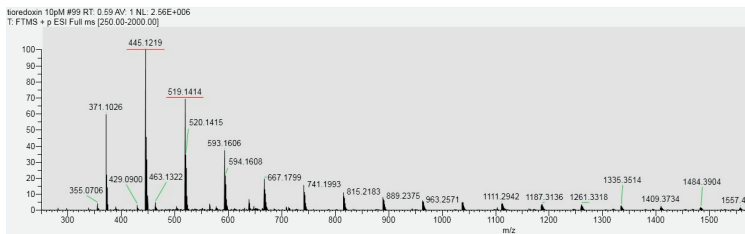
1. Prepare una disolución de tiorredoxina (masa molecular de 12kDa) a una concentración de 10 $\mu$ M.
2. Disuelva con ACN/H<sub>2</sub>O (50/50 v/v) + 0,1% de ácido fórmico (FA) para obtener una disolución de tiorredoxina de 10pM.
3. ¿Por qué se requiere una disolución de 0,1% de FA?
4. Introduzca la disolución en la fuente de ESI con bomba de jeringa de infusión directa preparada a una velocidad de flujo de 2 $\mu$ L/min.
5. Registre el espectro de masas durante 60 s utilizando los parámetros que se facilitan a continuación en la figura.







6. Abra los resultados obtenidos. Conociendo la masa de la molécula del biopolímero tioredoxina (12kDa), calcule la carga de cada ion obtenido.



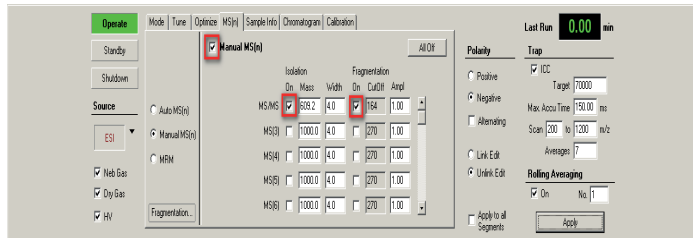
p. ej.	ion 445.12	ion 519.14
	$m/z = (M+z)/z$	$m/z = (M+z)/z$
	$445.12 = (12000+z)/z$	$519.14 = (12000+z)/z$
	$z=12000/444.12$	$z=12000/518.12$
	$z=27$	$z=23$

### 3.2. Adquisición de un espectro de masas ESI-MS/MS

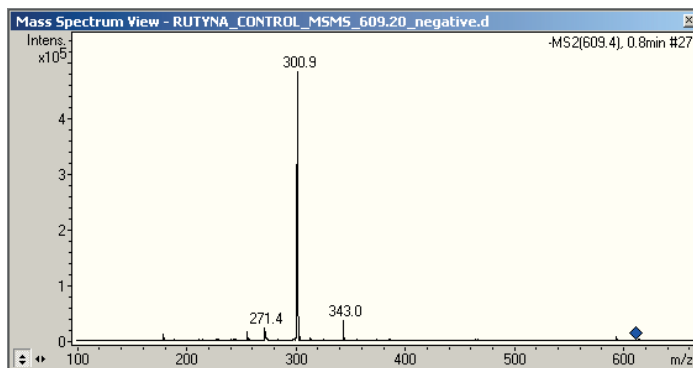
En esta parte de la sesión práctica, se analizará la misma disolución de rutina empleada en el punto 3.1 mediante infusión directa ESI-MS/MS a través del sistema de trampa de iones ESI. En esta etapa, los alumnos aprenderán qué relaciones  $m/z$  de iones (con respecto a iones moleculares e iones fragmento) se deben seleccionar para aplicarse al enfoque cuantitativo típico utilizando el modo MRM. Seguidamente, utilizando el software especializado, programarán la selección de un ion precursor, la fragmentación de este ion y la medida de la relación masa/carga de los iones producto formados. Los espectros ESI-MS/MS obtenidos se registrarán y utilizarán con nuevos fines de interpretación previstos en el punto 3.3.

**Experimento de ESI-MS/MS:**

1. Introduzca la disolución de rutina del punto 2 en la fuente de ESI con bomba de jeringa preparada a una velocidad de flujo de 150µL/h.
2. Marque la opción de aislamiento y fragmentación en la pestaña MS(n), escribir la  $m/z$  correspondiente al ion molecular de la rutina.



3. Registre el espectro de masas durante 30 s con los parámetros que se facilitan a continuación en la figura.



Espectros ESI-MS/MS de rutina en modo negativo

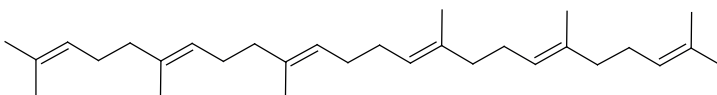
4. Observe detenidamente el espectro ESI-MS/MS obtenido.
5. ¿Qué ion se debería seleccionar para el análisis cuantitativo de la rutina en el modo MRM?
6. ¿Por qué?

### 3.3. Interpretación del espectro de masas

Los alumnos observarán los espectros de masas (obtenidos en los puntos 3.1 y 3.2) e identificarán los picos correspondientes a la  $m/z$  del ion molecular y sus iones fragmento característicos después de un examen pormenorizado de la estructura de los compuestos analizados. Asimismo, se estudiarán de manera crítica los espectros en relación a la presencia de posibles aductos con sodio o potasio e iones con múltiples cargas.

### 3.4. Observación de la ejecución de una GC-MS y configuración de la adquisición por monitorización selectiva de iones

El objetivo de este ejercicio es introducir a los alumnos una perspectiva práctica del método analítico de la cromatografía de gases acoplada a espectrometría de masas. Por tanto, la determinación cualitativa y cuantitativa de un compuesto típico no volátil (por ejemplo, escualeno como ingrediente de muestras de aceites comestibles, como el aceite de amaranto, de espino amarillo o de onagra) se llevará a cabo mediante el método GC-SIM MS.



Escualeno –  $C_{30}H_{50}$ , odo-trans-Escualeno, (6E,10E,14E,18E)-2,6,10,15,19,23-Hexamiltetracosa-2,6,10,14,18,22-hexaeno

#### **Experimento de GC-MS:**

1. Prepare una serie de estándares de escualeno en concentración de: 0,5 mg /mL, 1,0 mg/mL, 2 mg /mL, 5 mg /mL y 10 mg /mL en hexano.
2. Transfiera 10  $\mu$ L de muestra del aceite al vial de 2mL de la GC.
3. Añada 40  $\mu$ L del reactivo de sililación (BSTFA:TMCS; 99:1) en el vial.
4. Tape el vial y calentar a 80°C durante 1 hora.
5. Deje enfriar a temperatura ambiente.
6. Inyecte 1  $\mu$ L de la muestra en columna capilar de HP-5ms (0,25-mm; 0,25- $\mu$ m, 30-m) del sistema GC/MS con cuadrupolo MS/MS de 7890A GC – 7000 (Agilent Technologies, Palo Alto, CA).
7. Antes de analizar las muestras, configure los valores de software de los parámetros clave del análisis GC-MS que se facilitan a continuación:

**# Selection**

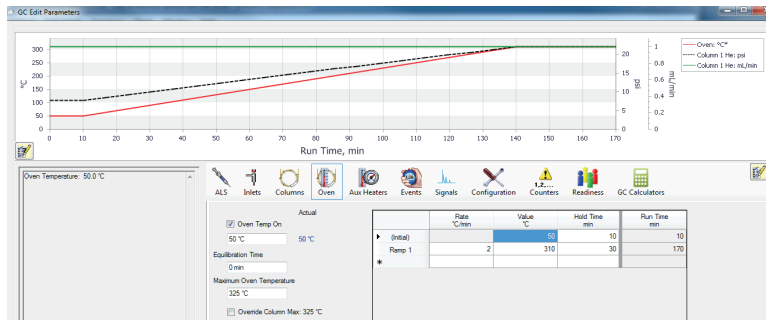
- Agilent 190915-433, 1371, 48429
- HP-5MS 5% Phenyl Methyl Silox
- 60 °C-325 °C (325 °C) 30 m x 250 µm x 0.25 µm
- In: Front SS Inlet He
- Out: MSD
- GGQ Collision Cell EPC

**Control Mode**  On

Parameter	Setpoint	Actual
Flow	1 mL/min	1 mL/min
Pressure	7.6522 psi	7.7 psi
Average Velocity	36.445 cm/sec	(Initial): 0 min He @ 50 °C Oven Out: MSD 30 m x 250 µm x 0.25 µm
Holdup Time	1.3719 min	

Constant Flow

Programación del horno: 50°C (10 min), a razón de 2°C/min a 310°C (30min)



Entrada (Split/Splitless): temperatura 270°C, split – 20:1

**Split-Splitless Inlet**

Parameter	Setpoint	Actual
Heater	270 °C	270 °C
Pressure	7.6522 psi	7.7 psi
Total Flow	24 mL/min	24 mL/min
Septum Purge Flow	3 mL/min	3 mL/min

Septum Purge Flow Mode: Standard

Mode: Split

Split Ratio: 20 : 1

Split Flow: 20 mL/min

Gas Saver:  On

20 mL/min After: 3 min

Temperatura de la línea de transferencia: 260°C

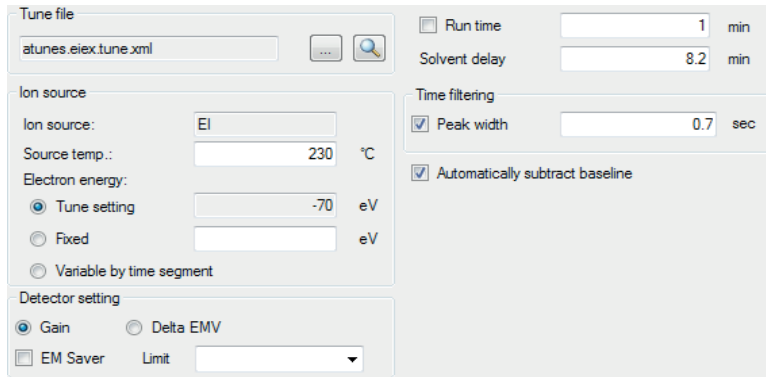
**Thermal Aux 2 (MSD Transfer Line)**

Actual

On

260 °C 259.8 °C

Fuente de iones (EI):230°C, Fuente de energía: - 70eV, Tiempo de espera para la elución del disolvente (Solvent delay): 8,2 min



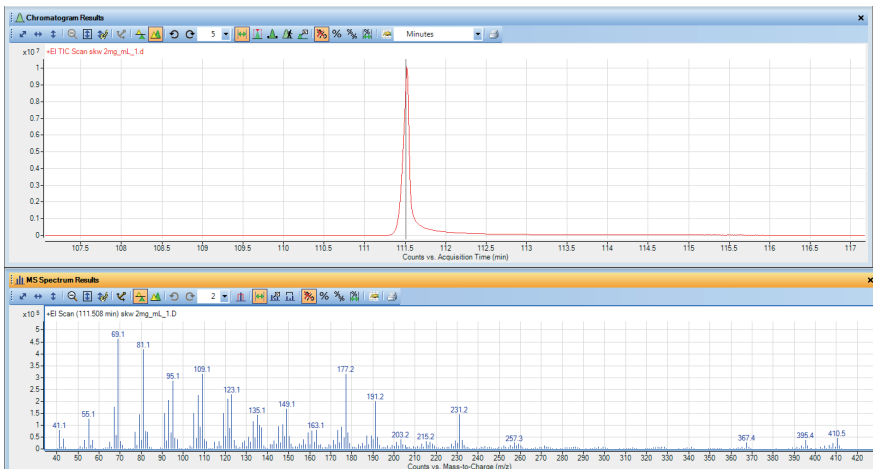
- En la ventana «Parámetros de MS-SIM/Scan», elija el «modo Scan» para obtener un cromatograma de iones totales (TIC, por sus siglas en inglés) en el rango de  $m/z$  (40-750).

Acquisition Instrument Chromatogram

Scan segments

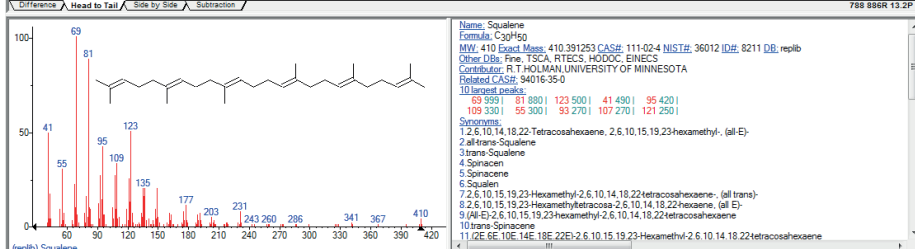
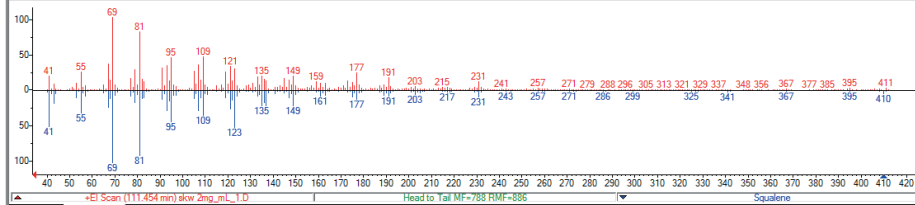
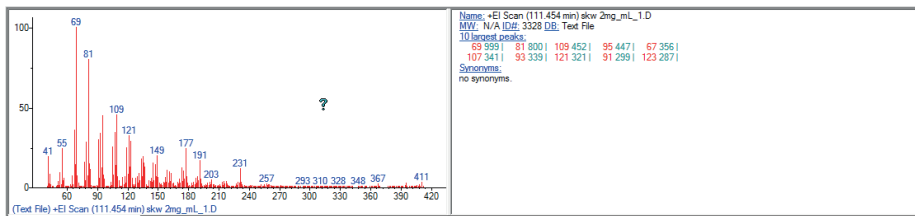
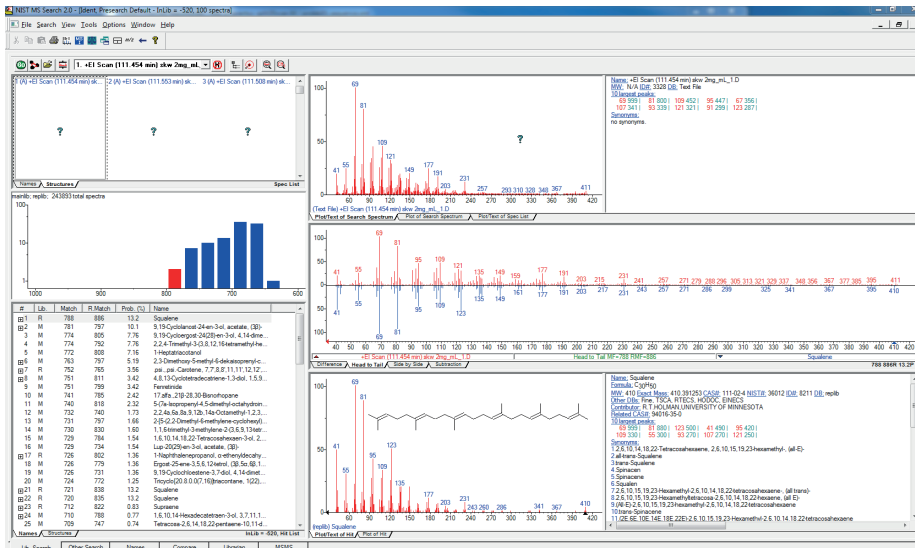
Segment name	Start mass	End mass	Scan time (ms)	Data Samples	Expected Scan Time (ms)
	40	750	476	4	466

- Realice un análisis cualitativo utilizando el escaneo anterior, registre el cromatograma del estándar de escualeno a una concentración de, por ejemplo, 2 mg/mL (en hexano)



Cromatograma y espectro de masas del estándar de escualeno.

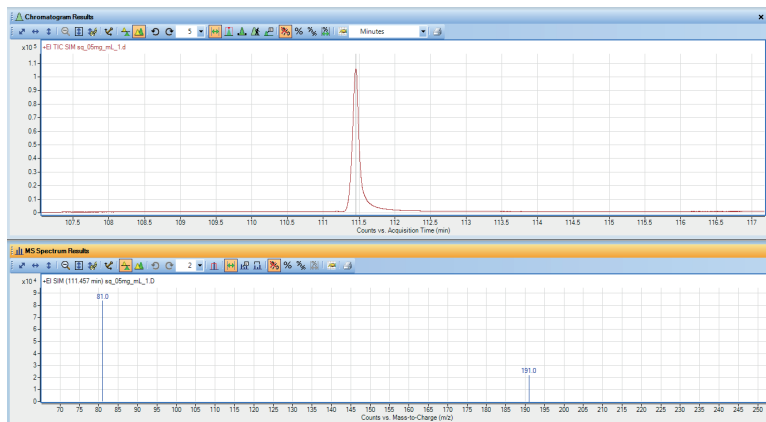
10. Búsqueda en la biblioteca (NIST MS) para el espectro obtenido; utilice el comando «Search Using NIST Program» (Búsqueda con el programa NIST).



- Para un análisis cuantitativo, cree un método de monitorización de iones seleccionados en MS (MS-SIM) mediante la selección en el software de iones a  $m/z$  191 y 81, que se corresponde con fragmentos característicos de la estructura del escualeno.

Acquisition					
Instrument		Chromatogram			
Scan segments					
	Compound name	ISTD?	Mass	MS1 resolution	Dwell
▶	squalene	<input type="checkbox"/>	191	Wide	15
	squalene	<input type="checkbox"/>	81	Wide	15

- Con el método SIM creado, registre el cromatograma del estándar de escualeno a una concentración de, por ejemplo: 0,5 mg/mL (en hexano).



Cromatograma de monitorización selectiva de iones y espectrometría en modo «SIM» del estándar de escualeno.

Se utiliza como «ion cualificador» un ion a  $m/z$  191, y como «ion cuantificador» un ion a  $m/z$  81 (el cualificador solo se emplea para determinar una relación de cualificador observada).

- En el siguiente paso, registre los cromatogramas de los estándares de escualeno a una concentración de: 0,5 mg/mL, 1,0 mg/mL, 2 mg/mL, 5 mg/mL y 10 mg/mL en hexano.

Quantifier						
Name	TS	Scan	Type	MZ	Uncertainty	
▶ squalene	1	SIM	Target	81.0	Relative	
Qualifier						
MZ	Rel. Resp.	Uncertainty	Area Sum			
191.0	6.7	20.0	<input type="checkbox"/>			

14. Cree los niveles de calibración en el software.

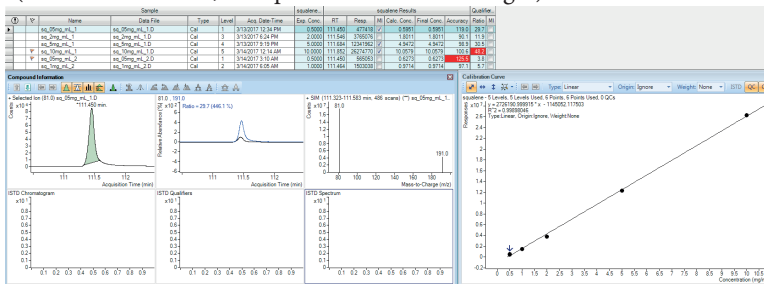
Quantifier				
Name	TS	Scan	Type	Units
squalene	1	SIM	Target	mg/ml

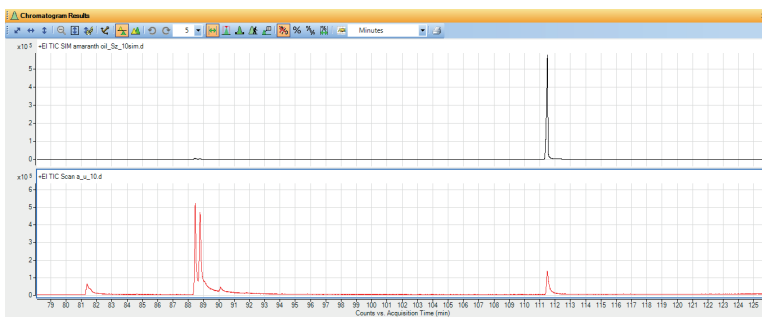
Calibration			
Level	Conc.	Response	Enable
1	0.5000	477418	<input checked="" type="checkbox"/>
2	1.0000	1503038	<input checked="" type="checkbox"/>
3	2.0000	3765076	<input checked="" type="checkbox"/>
4	5.0000	12341962	<input checked="" type="checkbox"/>
5	10.0000	26274770	<input checked="" type="checkbox"/>

15. Después de aplicar este método al lote -batch- (en la parte del software de Análisis Cuantitativo), genere la Curva de Calibración con el comando «Analyze Batch» (analizar lote) del menú.

16. Mediante el método SIM mencionado, registre los cromatogramas de las muestras sililadas de aceites (aceite de amaranto, de espino amarillo o de onagra).



Comparación de los cromatogramas de aceite de amaranto registrados en modo SIM y Scan.



17. Seleccione todos los cromatogramas registrados de muestras de aceites de la ruta metabólica y presione OK para añadirlos al archivo por lotes previamente creado.

18. A continuación, haga clic en el icono «Analyze Batch» (analizar lote) para obtener una tabla con los resultados cuantitativos.



### 3.5. Observación de la ejecución de una LC-MS y configuración de la adquisición por MS/MS dependiente de datos

Se analizará el extracto de lípidos de plasma en modo MS/MS dependiente de datos utilizando el sistema LC-MS- QTOF siguiendo la optimización previa de parámetros críticos para este tipo de análisis, como por ejemplo, el nivel basal, el número de precursores seleccionados o el número de escaneos. Durante la práctica, se proporcionará también el principio operativo y la construcción básica del sistema HPLC y el QTOF, como espectrómetro de masas híbrido. Durante esta parte de la sesión práctica, también se tratará la importancia de la lista de masas de compuestos preferidos y descartados en la adquisición dependiente de datos.

1. Realice un análisis de extracto de lípidos de plasma en modo dependiente de datos utilizando el método «Phospholipid\_HILIC\_negative.m» del menú contextual principal del programa Mass Hunter Acquisition Agilent. Los parámetros de la parte LC del método se muestran a continuación:

La fase móvil A consiste en un 25% de agua, 50% de acetonitrilo y 25% (v/v) de metanol, con 10 mM de acetato de amonio. La fase móvil B consiste en un 60% de acetonitrilo y un 40% de metanol, con 10 mM de acetato de amonio.

**+ Timetable (12/100 events)**

Time [min]	A [%]	B [%]	Flow [ml/min]	Max. Pressure Limit [bar]
0.00	0.00	100.00	0.030	600.00
8.00	0.00	100.00	0.030	600.00
15.00	40.00	60.00	0.030	600.00
40.00	0.00	100.00	0.030	600.00
45.00	0.00	100.00	0.030	600.00

- Simultáneamente, considere y analice de forma crítica los valores de los principales parámetros de ionización ESI.

Ion Source: Dual AJS ESI  
 Ion Polarity: Negative  
 Data Storage: Both  
 LC Stream: MS

Stop Time:  
 No Limit/As Pump  
 Stop Time: 30 min

Time Segment and Experiment #

Time (min)	Expt
0	1

Cycle Time: 1.3 s

Binary Pump | Column Comp. | DAD | **Q-TOF**

General | **Source** | Acquisition | Ref Mass | Chromatogram

Dual AJS ESI (Seg)

Gas Temp: 250 °C | 0 °C  
 Drying Gas: 13 l/min | 0.0 l/min  
 Nebulizer: 60 psig | 0 psig  
 Sheath Gas Temp: 125 °C | 0 °C  
 Sheath Gas Flow: 2 l/min | 0.0 l/min

Dual AJS ESI (Expt)

VCap: 3000 V | Capillary: 0.000 uA  
 Nozzle Voltage (Expt): 0 V  
 Chamber: 0.00 uA

MS TOF (Expt)

Fragmentor: 175 V  
 Skimmer: 65 V  
 OCT 1 RF Vpp: 750 V

Se debe seleccionar el tipo de análisis dependiente de datos en la pestaña de adquisición del programa Mass Hunter Agilent.

Binary Pump | Column Comp. | DAD | **Q-TOF**

General | Source | **Acquisition** | Ref Mass | Chromatogram

Mode:  
 MS (Seg)  
 **Auto MS/MS (Seg)**  
 Targeted MS/MS (Seg)

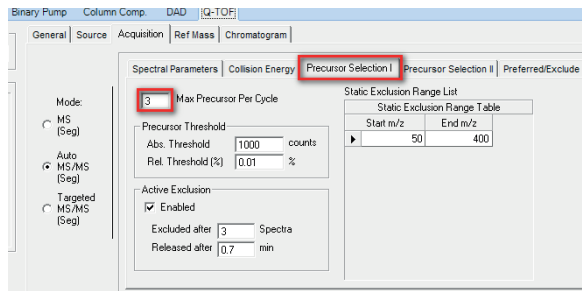
Spectral Parameters | Collision Energy | Precursor Selection I | Precursor Selection II | Preferred/Exclude

On	Prec. m/z	Delta m/z (ppm)	Z	Prec. Type	Ret. Time	Delta Ret. Time (min)	Iso. Width	Collision Energy
<input checked="" type="checkbox"/>	679.76	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	649.71	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	612.19	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	611.17	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	610.14	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	476.42	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	475.53	100	1	Exclude	0		Narrow (~1.3 m/z)	
<input checked="" type="checkbox"/>	474.32	100	1	Exclude	0		Narrow (~1.3 m/z)	

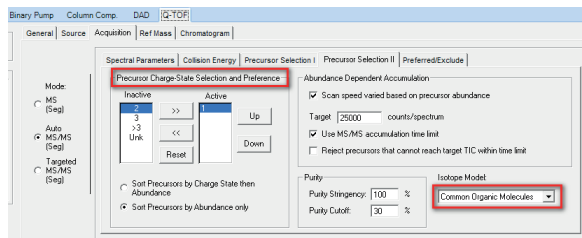
Default Values:  
 Delta m/z: 100 ppm  
 Delta Ret. Time: min  
 Use Preferred ion list only

- Analice de forma crítica parámetros importantes para el análisis dependiente de datos, como por ejemplo:

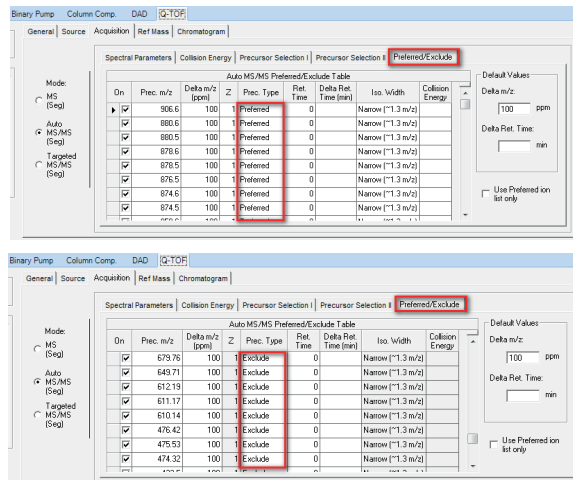
- El número de precursores seleccionados



- El patrón isotópico

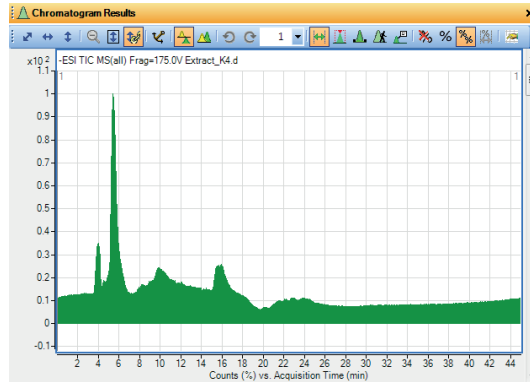


- La creación de listas de valores  $m/z$  preferidos y descartados

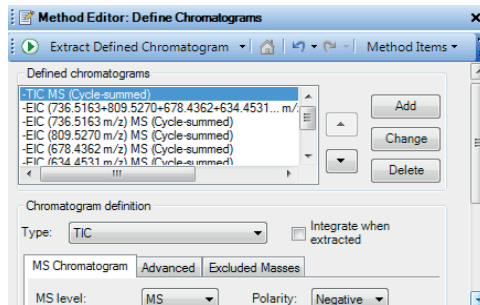


¿Cuál es la importancia y el significado de estas listas en el caso de un análisis de datos dependientes?

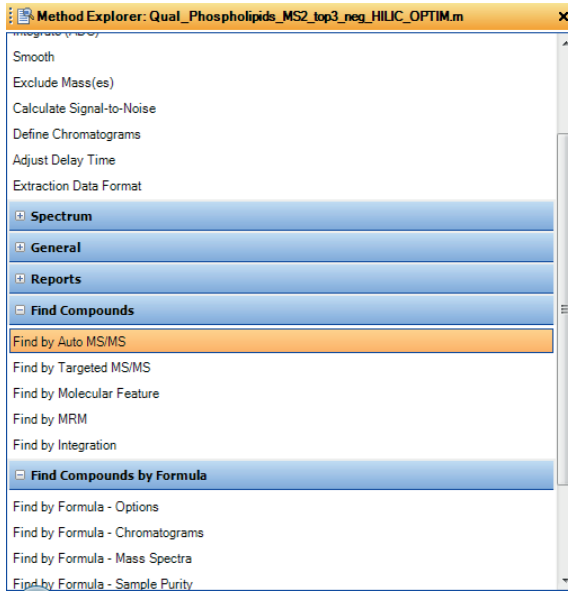
4. Genere un Cromatograma de Iones Totales (TIC) para obtener un perfil de fosfolípidos de la muestra de plasma analizada.



5. Genere Cromatogramas de Ion Extraído (EIC, por sus siglas en inglés) mediante el comando «Extract defined chromatogram» (extraer cromatograma definido) del programa Mass Hunter para cada estándar interno de cada clase de fosfolípidos que se haya añadido a la muestra para estimar los rangos de tiempos de retención que permitan la asignación de especies identificadas a una clase particular.



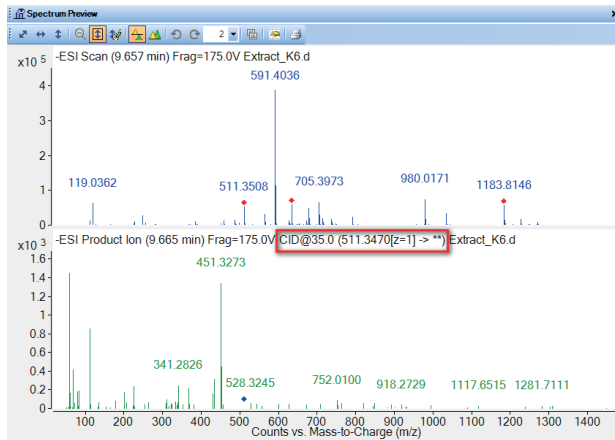
- Genere una lista de compuestos con la función «Find compound by Auto MS/MS» (encontrar compuesto mediante MS/MS automática) del menú contextual «Find Compound» (encontrar compuesto)



The image shows a 'Compound List' window with a table of detected compounds. The table has columns for 'Show/Hide', 'Cpd', 'File', 'RT', and 'm/z'. The 'Show/Hide' column contains checkboxes, all of which are checked. The 'Cpd' column contains numbers from 1 to 22. The 'File' column contains 'Extract\_K6.d' for all entries. The 'RT' column contains retention times, and the 'm/z' column contains mass-to-charge ratios. The 12th row is highlighted in blue.

Show/Hide	Cpd	File	RT	m/z
<input checked="" type="checkbox"/>	1	Extract_K6.d	0.062	494.3586
<input checked="" type="checkbox"/>	2	Extract_K6.d	0.811	819.5143
<input checked="" type="checkbox"/>	3	Extract_K6.d	0.822	718.555
<input checked="" type="checkbox"/>	4	Extract_K6.d	0.946	656.8809
<input checked="" type="checkbox"/>	5	Extract_K6.d	1.034	764.5413
<input checked="" type="checkbox"/>	6	Extract_K6.d	1.119	690.5284
<input checked="" type="checkbox"/>	7	Extract_K6.d	1.18	906.5547
<input checked="" type="checkbox"/>	8	Extract_K6.d	1.229	824.4565
<input checked="" type="checkbox"/>	9	Extract_K6.d	1.362	482.9254
<input checked="" type="checkbox"/>	10	Extract_K6.d	1.465	848.5168
<input checked="" type="checkbox"/>	11	Extract_K6.d	1.506	621.4743
<input checked="" type="checkbox"/>	12	Extract_K6.d	1.721	690.5607
<input checked="" type="checkbox"/>	13	Extract_K6.d	1.75	819.5143
<input checked="" type="checkbox"/>	14	Extract_K6.d	1.899	706.4321
<input checked="" type="checkbox"/>	15	Extract_K6.d	2.266	848.4113
<input checked="" type="checkbox"/>	16	Extract_K6.d	2.41	876.4235
<input checked="" type="checkbox"/>	17	Extract_K6.d	2.416	522.4201
<input checked="" type="checkbox"/>	18	Extract_K6.d	2.465	718.5351
<input checked="" type="checkbox"/>	19	Extract_K6.d	2.829	819.5143
<input checked="" type="checkbox"/>	20	Extract_K6.d	2.866	880.5361
<input checked="" type="checkbox"/>	21	Extract_K6.d	3.125	718.555
<input checked="" type="checkbox"/>	22	Extract_K6.d	3.15	611.3993

- Encuentre compuestos a una  $m/z$  de, por ejemplo, 591,4036
- Asigne el tiempo de retención para cada uno de estos compuestos.
- Asigne los valores  $m/z$  de los iones fragmento característicos de cada uno de estos compuestos del punto 2 y utilícelos.



**Para más información:**

- E. de Hoffmann, V. Stroobant. Mass spectrometry: principles and applications. Wiley, 2007.  
H.J. Hübschmann. Handbook of GC-MS: Fundamentals and Applications (3rd Edition). Wiley, 2015.  
J. Kool, W.M. Niessen. Analyzing biomolecular interactions by mass spectrometry. Wiley, 2015.

## Módulo 2

# Metabolómica

Coral Barbas, Danuta Dudzik, M<sup>a</sup> Fernanda Rey-Stolle, Francisco J. Rupérez, Antonia García

*Centro de Metabolómica y Bioanálisis (CEMBIO), Facultad de Farmacia, Universidad San Pablo CEU; Madrid, España*

### I. Justificación

El objetivo del módulo es que los alumnos de posgrado adquieran competencias y conocimientos avanzados en la disciplina «ómica» más reciente. La metabolómica es el análisis sistemático de un conjunto completo de metabolitos intra y extracelulares de una muestra biológica. El fin último de la metabolómica es ayudar a comprender las alteraciones de los metabolitos y el mecanismo molecular de un estado de enfermedad subyacente, así como el descubrimiento de nuevos biomarcadores clínicos de diagnóstico y pronóstico. Este módulo se divide en cuatro secciones que cubren los aspectos principales de la metabolómica: 1) Objetivos y definiciones, 2) Enfoques, diseño experimental, requisitos especiales, herramientas analíticas, control de calidad y flujos de trabajo en la metabolómica, 3) Pretratamiento de datos, procesamiento de datos, análisis estadístico e identificación de biomarcadores, y, en la última parte, 4) Interpretación bioquímica, análisis de rutas metabolómicas y validación de biomarcadores junto con varias aplicaciones.

El curso se impartirá mediante una combinación de seminarios, talleres, tutorías y sesiones informales, entre las que se incluyen clases prácticas y de demostración, con datos reales, donde se utilizan recursos bioinformáticos de acceso libre que se aplican en la actualidad. Todos estos recursos docentes permitirán la consolidación de los conocimientos y competencias específicas.

### II. Objetivos y resultados del curso

#### **Objetivos**

El objetivo de este curso es proporcionar a los alumnos una comprensión de los conceptos y enfoques fundamentales para los estudios metabolómicos. Los alumnos adquirirán conocimientos teóricos y prácticos de la metabolómica no dirigida, entre los que se incluyen flujos de datos, etapas y requisitos, análisis de datos metabolómicos, preprocesado, estadística, identificación y anotación. Destacando también la necesidad de validar los biomarcadores y de la investigación futura de las hipótesis biológicas establecidas.

#### **Resultados del aprendizaje:**

Tras terminar el curso, el alumno será capaz de:

1. Definir y aplicar terminología común de la metabolómica.

2. Analizar pormenorizadamente los diferentes mecanismos de separación acoplados a la MS para la metabolómica.
3. Identificar todas las etapas del estudio metabolómico y sus diferentes enfoques.
4. Elegir entre los distintos enfoques para resolver un problema concreto.
5. Conocer y entender los diferentes métodos de tratamiento de muestras y sus limitaciones en la metabolómica.
6. Conocer y entender las distintas etapas del procesado de datos.
7. Usar software público en el procesamiento de datos y en el análisis de rutas metabólicas.
8. Conocer los diferentes enfoques para el análisis estadístico aplicado a la metabolómica.
9. Adquirir competencias básicas en el uso de bases de datos de metabolitos, como recursos de acceso gratuito.
10. Explicar a personas no especialistas en el campo la forma en la que esta disciplina «ómica» puede proporcionar información de valor en las diferentes áreas de las Ciencias de la Vida.
11. Comunicar y justificar conclusiones de forma clara e inequívoca ante especialistas y no especialistas.
12. Continuar el proceso de aprendizaje, mayormente, de forma autónoma.

### III. Contenido del curso

#### Módulo 2 – Metabolómica

1. Introducción a la metabolómica
2. Enfoques analíticos en la metabolómica
  - 2.1. Flujo de trabajo del estudio de la metabolómica
    - 2.1.1. Diseño del experimento
    - 2.1.2. Recogida de muestras
    - 2.1.3. Extracción de metabolitos
    - 2.1.4. Separación y detección
  - 2.2. Procedimiento de control de calidad y garantía de calidad en la metabolómica
3. Procesamiento de datos e identificación de metabolitos
  - 3.1. Flujo del procesamiento de datos
    - 3.1.1. Procesamiento de datos brutos
    - 3.1.2. Preprocesamiento de datos
    - 3.1.3. Pretratamiento de datos
  - 3.2. Identificación de metabolitos
    - 3.2.1. Identificación en GC-EI-MS
    - 3.2.2. Identificación de metabolitos en experimentos de metabólica de LC o CE-(ESI)MS : trabajando con bases de datos
    - 3.2.3. Bases de datos centradas en el metabolito sobre compuestos para búsquedas simples de PM



- 3.2.4. Bases de datos espectrales centradas en el metabolito
- 3.2.5. Mediadores
- 3.3. Análisis estadístico
  - 3.3.1. PCA
  - 3.3.2. PLS-DA
  - 3.3.3. OPLS-DA
  - 3.3.4. S-Gráfica S-Plot, intervalo Jack-knife y Variable importancia en la proyección (VIP)
  - 3.3.5. Validación de los modelos estadísticos multivariantes
- 4. Análisis de datos
  - 4.1. De la identificación de datos a las rutas metabólicas: Evaluación de datos inteligente pero asistida para la evaluación de la relevancia biológica
    - 4.1.1. Bases de datos de rutas metabólicas
    - 4.1.2. Análisis de rutas metabólicas
    - 4.1.3. Enriquecimiento de metabolitos
    - 4.1.4. Análisis (topológico) de rutas metabólicas
  - 4.2. Validación de biomarcadores: Asignación de relevancia a los metabolitos
- 5. Sesiones prácticas sobre metabolómica
  - 5.1. Metabolómica dirigida y no dirigida
  - 5.2. Preprocesamiento de datos con Workflow4Metabolomics (W4M)
    - 5.2.1. Conversión de archivos MS
    - 5.2.2. Inspección de datos brutos
    - 5.2.3. Subir datos a Galaxy
    - 5.2.4. Preprocesamiento de datos de LC-MS a través de XCMS
  - 5.3. Estadísticas
  - 5.4. Anotación

## 1. Introducción a la metabolómica

La metabolómica está relacionada con el análisis sistemático del conjunto completo de metabolitos intra y extracelulares en un sistema biológico dado (incluidos sistemas microbianos, vegetales y mamíferos). Los metabolitos son sustancias intermedias con un bajo peso molecular (normalmente < 1.500 Da) que constituyen los elementos básicos para otros muchos componentes biológicos (por ejemplo, proteínas, ADN, ARN). Participan en reacciones metabólicas generales y son esenciales para la regulación, el crecimiento y funcionamiento normal de la célula. Los metabolitos, resultantes de los procesos transcripcionales, traduccionales y post-traduccionales, funcionan como los exponentes más claros de las alteraciones que ocurren en el cuerpo en respuesta a un proceso patológico. Por ello, la metabolómica ofrece una ventaja única para proporcionar una imagen directa del estado actual y el fenotipo de un organismo. Comprender el metaboloma, definido como el conjunto completo de metabolitos de una célula, tejido

u organismo, es indispensable para obtener una visión más global del funcionamiento de las células, los tejidos y organismos. En la mayoría de los casos, la metabolómica consiste en un estudio diferencial de huellas metabólicas (fingerprinting) generadas a partir de subgrupos de observaciones de «control» y «prueba» para encontrar diferencias en sus perfiles en respuesta a estímulos externos (patologías, los efectos del estrés bioquímico o ambiental, procesamiento alimentario, etc.). En la Tabla 2.1 se incluye una relación de términos específicos propios del campo de la metabolómica.

**Tabla 2.1.** Definiciones relacionadas con la metabolómica.

<b>Término</b>	<b>Definición</b>
<b>Metabolómica</b>	El estudio del conjunto completo de metabolitos que varían en función de la fisiología, el estado de desarrollo o patológico de la célula, tejido, órgano u organismo.
<b>Metabolitos</b>	Pequeñas moléculas, sustancias intermedias con un bajo peso molecular (< 1.500 Da), base fundamental de otros componentes biológicos, esenciales para la regulación, el crecimiento y funcionamiento normal de la célula.
<b>Metaboloma</b>	El producto final resultante del genoma, el conjunto completo de todos los metabolitos de bajo peso molecular presentes en la célula u organismo. El metaboloma se divide en exometaboloma (metabolitos extracelulares) y endometaboloma (metabolitos intracelulares).
<b>Metabotipo</b>	El fenotipo metabólico
<b>Huella metabólica</b>	Enfoque objetivo, global, de alto rendimiento para la clasificación de muestras basado en patrones o «huellas» de metabolitos que cambian en respuesta a las perturbaciones de enfermedades, ambientales o genéticas con el objetivo final de identificar metabolitos discriminatorios.
<b>Determinación del perfil metabólico</b>	Identificación y cuantificación o semicuantificación de un grupo específico de metabolitos que estén relacionados con una ruta metabólica concreta.
<b>Huella metabólica</b>	Análisis de los (exo)metabolitos segregados/excretados por un organismo en condiciones controladas.
<b>Metabolómica no dirigida</b>	De alcance global, libre de hipótesis, con el objetivo de medir simultáneamente tantos metabolitos como sea posible de muestras biológicas.
<b>Metabolómica dirigida</b>	Implica el uso de un método analítico particular para la determinación y cuantificación de un conjunto pequeño de metabolitos conocidos.
<b>Ruta metabólica</b>	Un conjunto de reacciones químicas dentro de la célula.

La metabolómica ya ha demostrado ser un método eficaz, consolidado y valioso de investigación de la salud y la enfermedad de los seres humanos, su envejecimiento, estilo de vida y el descubrimiento y desarrollo de fármacos. Por ello, el fin último de la metabolómica es ayudar a comprender las alteraciones de los metabolitos y el mecanismo molecular que subyacen en el estado de enfermedad, así como el descubrimiento de nuevos biomarcadores clínicos de diagnóstico y pronóstico. En los últimos años de manera conjunta con el perfeccionamiento de la tecnología analítica, se ha incrementado el número de estudios que abordan el análisis del metaboloma.

## 2.Enfoques analíticos en la metabolómica

Dependiendo de los objetivos del estudio metabolómico, se pueden aplicar dos estrategias analíticas: el enfoque dirigido y el enfoque no dirigido (Figura 2.1).

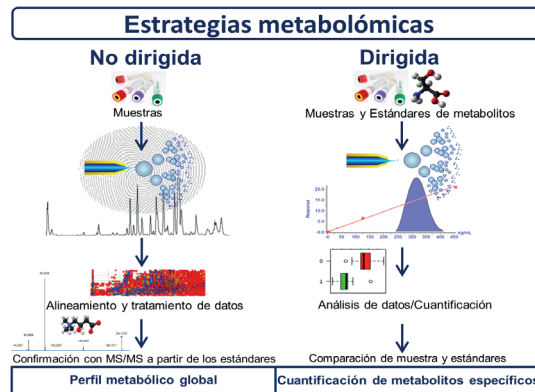


Figura 2.1. El enfoque analítico no dirigido y dirigido para estudios metabolómicos.

### Enfoque dirigido

La metabolómica dirigida se refiere a estudios cuyo objetivo es medir moléculas específicas conocidas (normalmente en número <20), centrándose en una o más rutas metabólicas relacionadas que hayan sido definidas en estudios previos como biológicamente relevantes. De este modo, se aplican métodos con altos niveles de especificidad, precisión y exactitud para definir cantidades absolutas de cada metabolito; por ejemplo, la cuantificación de aminoácidos utilizando métodos de monitorización de la reacción seleccionada con MS de triple cuadrupolo (QqQ). La cuantificación absoluta del compuesto objetivo se realiza mediante la incorporación de estándares internos apropiados y basando la cuantificación en las curvas de calibración construidas.

### Enfoque no dirigido

La metabolómica no dirigida, que se refiere al estudio de las huellas metabolómicas, tiene un alcance

global y su objetivo principal es medir de forma simultánea el mayor número posible de metabolitos de cada muestra sin distorsión, en un solo experimento. En la estrategia no dirigida, los metabolitos que van a ser identificados no se conocen antes del estudio. La metabolómica no dirigida se aplica a estudios libres de hipótesis, que constituye la fase de descubrimiento, con el objetivo de definir nuevos cambios en el metaboloma no observados con anterioridad. Esto puede relacionarse con el funcionamiento biológico de un organismo y con los mecanismos que están detrás de los cambios metabólicos alterados relacionados con el desarrollo de un estado de enfermedad. Este enfoque conduce a la generación de una nueva hipótesis que deberá ser validada en el estudio posterior, incluyendo un enfoque dirigido.

## 2.1. Flujo de trabajo del estudio metabolómico

En la Figura 2.2 se muestra un flujo de trabajo típico de un análisis metabolómico no dirigido y que está directamente vinculado a un diseño experimental apropiado. Los puntos cruciales del flujo de trabajo comprenden (1) el diseño experimental; (2) la recogida de muestras; (3) el protocolo de extracción; (4) la adquisición de datos; (5) el procesamiento y análisis de datos; y (6) la identificación de metabolitos que permita la interpretación biológica. Cada una de las etapas del flujo de trabajo metabolómico reviste una gran importancia. Por ello, hay que aplicar procedimientos de garantía de calidad para reducir variaciones preanalíticas y experimentales no deseadas y para garantizar la fiabilidad de los estudios metabolómicos.



Figura 2.2. Flujo de trabajo típico para estudios metabolómicos no dirigidos.

### 2.1.1. Diseño del experimento

Un diseño experimental ha de describir pormenorizadamente cómo se realizará el estudio. Deberá cubrir todos los aspectos relevantes del flujo de trabajo, desde el muestreo hasta el análisis final de datos. Un diseño experimental implica que todas las muestras que vayan a ser estudiadas son escogidas de una manera bien meditada. Si las muestras ya han sido recogidas (por ejemplo, cohortes de muestras de estudios epidemiológicos) es importante saber (1) cómo se han recogido y guardado; (2) cómo se pueden encajar las muestras de control con las de tratamiento; y (3) si hay

un fenotipo claro entre los grupos de control y experimentales. El diseño experimental debe incluir una descripción clara del número de muestras, réplicas y grupos estudiados para garantizar que se obtienen datos con suficiente poder estadístico. Los estudios basados en un modelo experimental, por ejemplo, en sistemas animales o de cultivos celulares, dentro de un entorno de laboratorio bien controlado, normalmente requieren un tamaño de muestra pequeño (normalmente 6-12) para identificar cambios estadísticamente sólidos. Sin embargo, los estudios basados en la población humana en general son más complejos e incluyen muchas más variables que podrían sesgar el estudio diseñado. Se deben considerar factores relacionados con el estilo de vida (por ejemplo, la dieta y el ejercicio), la demografía (por ejemplo, el origen étnico) y la fisiología (por ejemplo, el género, la edad, el IMC), además de datos clínicos (por ejemplo, el historial médico, medicaciones), los cuales permitirán identificar y evitar variables de confusión. Es necesario un diseño experimental eficaz y minucioso para poder garantizar que incluso las variaciones pequeñas del metaboloma biológicamente relevantes puedan ser medidas con precisión y sean significativamente mayores que las alteraciones introducidas por el sesgo experimental.

### 2.1.2. Recogida de muestras

Seguir las características requeridas que se establecen en el diseño del experimento. Recoger todas las muestras con los mismos contenedores apropiados, cerrarlas bien, identificarlas y guardarlas a  $-80^{\circ}$  siempre que sea posible. Descongelar en el mismo día del análisis y mantenerlas en hielo.

### 2.1.3. Extracción de metabolitos

La extracción de metabolitos es otra fase que debe estar bien controlada para obtener resultados de estudio fiables. Las muestras normalmente se analizan sin grandes preparaciones y el método de extracción depende de la matriz de la muestra y del sistema analítico aplicado.

**LC-MS.** Para un análisis basado en LC-MS, las muestras que contengan proteínas requieren una fase de desproteización, que se refiere al proceso de eliminación del complemento proteínico. Por ello, normalmente se añade un disolvente orgánico frío (por ejemplo, metanol o acetonitrilo) o una mezcla de disolventes (por ejemplo, metanol/etanol (1:1, v/v)) para precipitar las proteínas, que se pueden separar entonces mediante una nueva centrifugación. El sobrenadante que resulta de la centrifugación contiene metabolitos extraídos y es apto para el análisis LC-MS.

**GC-MS.** Para la GC-MS, además de una nueva desproteización, deben llevarse a cabo procedimientos más exhaustivos, ya que la GC no es compatible con compuestos no volátiles y térmicamente lábiles. Entre estos procesos se incluyen la protección de grupos funcionales cetona mediante metoximación y derivatización para incrementar la volatilidad de los metabolitos. El proceso de derivatización disminuye el punto de ebullición de muchos metabolitos endógenos, lo que los convierte en lo suficientemente volátiles para permitir separaciones cromatográficas.

Los compuestos que contienen grupos funcionales con hidrógenos activos, como -SH (tioles), -OH (alcoholes, polioles, fenoles, enoles), -NH (aminas, amidas) y -COOH (ácidos carboxílicos) tienen una importancia fundamental en el análisis GC-MS, debido a la tendencia de estos grupos funcionales a formar enlaces de hidrógeno intermoleculares. Estos enlaces de hidrógeno intermoleculares afectan a la volatilidad inherente de los compuestos, a su tendencia a interactuar con los materiales de empaque de la columna y su estabilidad térmica. Así, tras la desproteinización, las muestras deben secarse antes de añadir cualquier agente derivatizante. La metoximación utiliza clorhidrato de O-metoxiamina en piridina para estabilizar fracciones de carbonilo suprimiendo la tautomería ceto-enólica y la formación de estructuras acetales (o cetales) múltiples. Asimismo, ayuda a reducir el número de derivados de azúcares reductores y solo genera dos formas del derivado -N=C<, sin/anti. El proceso de sililación subsiguiente comprende la sustitución de los átomos de hidrógeno activos con un grupo alquilsililo, por ejemplo, el Trimetilsililo (-SiMe<sub>3</sub>). La modificación del grupo funcional de una molécula por derivatización química posibilita el análisis de muchos compuestos polares, incrementando su volatilidad, estabilidad térmica y la simetría del pico cromatográfico.

**CE-MS.** La preparación típica de una muestra para el análisis CE-MS consiste normalmente en dos etapas; a saber, la dilución de la muestra y la desproteinización. Para una mejor sensibilidad, la fortaleza iónica de la disolución de muestra debe ser menor que el electrolito de fondo que rellena el capilar.

#### 2.1.4. Separación y detección

La elección de la plataforma analítica puede influenciar significativamente la cobertura del metaboloma obtenido en estudios no dirigidos. La espectrometría de masas (MS) y la resonancia magnética nuclear (NMR) son los métodos de detección más empleados en el análisis del metaboloma. La NMR resulta muy útil para la caracterización estructural de compuestos desconocidos. No obstante, se caracteriza también por una baja sensibilidad y, a mayores, unos costes de instalación mucho más elevados si se compara con las técnicas basadas en MS. Por otro lado, las ventajas más importantes de la MS son su alta sensibilidad y su alto rendimiento. El metaboloma se compone de compuestos químicos extremadamente diversos que hacen prácticamente imposible captar simultáneamente el metaboloma completo presente en la muestra en una sola plataforma analítica. Por tanto, la combinación de técnicas de separación (cromatografía líquida, cromatografía de gases y electroforesis capilar) con la MS amplía enormemente la capacidad del análisis químico de muestras biológicas altamente complejas y permite obtener una cobertura metabólica completa. La GC-MS ofrece la capacidad de analizar metabolitos volátiles de baja polaridad como algunas grasas y ésteres, así como metabolitos de alta polaridad como los aminoácidos y ácidos orgánicos convertidos en derivados volátiles, mientras que los metabolitos de alto peso molecular apolares y semipolares (por ejemplo esfingolípidos o glicerosfolípidos) se suelen estudiar aplicando la LC-MS. Asimismo, la CE-MS ofrece un enfoque complementario a la LC-MS para el análisis de

aniones, cationes y partículas neutras. Estos métodos analíticos han sido desarrollados y validados para ofrecer datos reproducibles y fiables. En la Tabla 2.2 se muestran las características principales de las técnicas de separación acopladas a MS para estudios metabolómicos

**Tabla 2.2.** Técnicas de separación acoplada a la espectrometría de masas para la metabolómica.

Técnica de análisis	Aplicación	Ventajas	Inconvenientes
GC-MS	Separación, identificación y cuantificación de metabolitos menos polares volátiles o térmicamente estables.	Alta resolución cromatográfica, disponibilidad de un amplio espectro de librerías para la identificación de metabolitos.	Imposibilidad de analizar metabolitos termolábiles y de alto peso molecular; es necesaria derivatización para los metabolitos no volátiles.
LC-MS	Separación, identificación y cuantificación de grupos muy amplios de metabolitos, dependiendo del tipo de columna y fase móvil.	Alta sensibilidad, gran capacidad de muestra, no se requiere derivatización, capacidad para analizar compuestos termolábiles.	Disponibilidad limitada de librerías comerciales, restricción de eluyentes de LC, efecto matriz, potencial de identificación limitado a menos que se utilice una técnica MS-MS.
CE-MS	Separación, identificación y cuantificación de metabolitos polares e ionizados, utilizando volúmenes de muestra reducidos.	Alta resolución y análisis rápido, útil con muestras biológicas complejas, aun cuando se trate de volúmenes pequeños.	Disponibilidad limitada de librerías comerciales. Incompatibilidad con tampones, límites en la detección. Potencial de identificación limitado, a menos que se utilice una técnica MS-MS.

**Para saber más:**

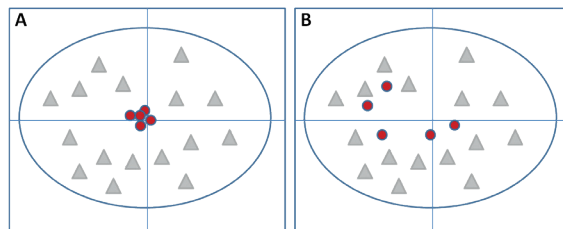
S.G. Villas-Bôas, S. Mas, M. Åkesson, J. Smedsgaard, and J. Nielsen, Mass spectrometry in metabolome analysis. *Mass spectrometry reviews* 24:613-646, 2005

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

En todas las estrategias analíticas (dirigidas y no dirigidas) es necesario contar con un procedimiento de garantía de calidad (QA, por sus siglas en inglés), especialmente cuando se trata de estudios a gran escala. En estudios dirigidos donde se adopten plataformas de MS, se aplican habitualmente muestras de control de calidad (QC, por sus siglas en inglés) para determinar desde un punto de vista cuantitativo la exactitud y la precisión del método analítico. En el contexto de la metabolómica, se aplican muestras de QC con el fin de (1) controlar el rendimiento del sistema analítico; (2) equilibrar el sistema para conseguir condiciones de reproducibilidad absoluta; (3)

corregir niveles de desviación pequeños en la señal medida durante el análisis (a nivel interno de los lotes) y entre lotes analíticos; (4) integrar los datos procedentes de distintos lotes analíticos; (5) calcular la precisión de la medición metabólica de inyecciones repetidas de QC.

Las muestras de QC se analizan al inicio y al final de la ejecución de un análisis, y en intervalos regulares durante todo el análisis para monitorizar la estabilidad y reproducibilidad del proceso analítico. Las muestras de QC deben ser representativas de la composición cualitativa y cuantitativa de las muestras que se estén analizando en el estudio. Se supone que la muestra de QC contendrá una concentración media de todos los componentes que estén presentes en las muestras de la investigación. La reproducibilidad de los datos del experimento se verá afectada por fuentes de variación aleatorias y sistemáticas. Se han desarrollado varios métodos de post adquisición de datos basados en los resultados del QC dirigidos a garantizar la calidad de los experimentos de metabolómica. La evaluación de la calidad de los datos incluye una perspectiva general de la precisión analítica que se basa en (1) el examen de los datos brutos (raw data) adquiridos; (2) la representación gráfica de la suma de las intensidades de las características metabólicas de todas las muestras experimentales y de QC; (3) la comprobación de desviaciones de señal o la pérdida o variación de sensibilidad en el procedimiento de QC a través de la predicción en un modelo multivariante no supervisado (Análisis de Componentes Principales, PCA-X); (4) el cálculo del coeficiente de variación CV (desviación estándar relativa porcentual  $-\% \text{RSD}-$ ) de cada característica metabólica detectada en las muestras de QC. Con el cálculo del CV de cada característica metabólica (posible metabolito) de todas las muestras de QC del experimento se obtiene una medida cuantitativa de la variabilidad. La aceptación de los límites del CV, por debajo del 20% para datos de LC-MS y CE-MS, y por debajo del 30% para la GC-MS, permitirá filtrar las matrices de datos complejas y eliminar metabolitos que muestren una variación inaceptable. En la Figura 2.3 se muestra el posible resultado de la predicción de muestras de QC en un PCA-X multivariante.



**Figura 2.3.** Representación esquemática de la predicción de muestras de QC en un modelo multivariante de PCA-X. Panel 3.A. Se puede observar una gran aglomeración de la muestra de QC, indicando el resultado analítico preciso. Panel 3.B. La dispersión de las muestras de QC indica la pérdida de la estabilidad y reproducibilidad del sistema dentro del ensayo analítico.



### 3. Procesamiento de datos e identificación de metabolitos

La investigación metabolómica genera conjuntos de datos complejos que afectan a cientos o incluso miles de metabolitos. El tratamiento correcto de los datos es una etapa fundamental de enorme impacto en el alcance y la calidad de los resultados metabolómicos. Con todo, el procesamiento de datos metabolómicos, el tratamiento de datos y la identificación de metabolitos son probablemente también otras de las etapas más problemáticas de este enfoque, debido a distintas características intrínsecas: los datos generados a partir de señales instrumentales son ruidosos y contienen una gran cantidad de variables altamente correlacionadas en comparación con el número de individuos. No obstante, existen varias herramientas avanzadas para el preprocesamiento y análisis de datos metabolómicos, suministrados comercialmente (por ejemplo, Agilent Technologies) o de forma gratuita en plataformas en línea (por ejemplo, XCMS, MZmine, MetAlign, MetaboAnalyst, MeltDB o la recientemente desarrollada W4M).

#### 3.1. Flujo del procesamiento de datos

En los estudios metabolómicos es importante separar la variación biológica de interés de las fuentes de variabilidad no deseadas introducidas en el experimento. Así, un análisis completo de los datos metabolómicos requiere una estrategia exhaustiva de análisis de datos que a menudo es única y que requiere software de análisis de datos especializado que permita el análisis quimiométrico, bioinformático y estadístico. El flujo de trabajo general del análisis de datos metabolómicos se basa en cuatro etapas principales relacionadas con (1) el procesamiento de datos brutos; (2) el preprocesamiento de datos; (3) el pretratamiento de datos y (4) el tratamiento de datos (Figura 2.4).

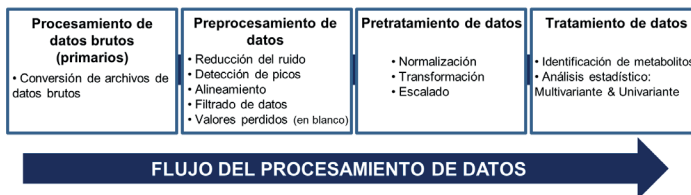


Figura 2.4. Etapas principales del flujo de trabajo del análisis de datos metabolómicos.

Los estudios metabolómicos no dirigidos se caracterizan por la medición simultánea de un gran número de metabolitos de cada muestra. Es por ello que se genera una gran cantidad de datos. En los estudios metabolómicos no dirigidos, las cuestiones principales que se deben considerar se refieren a dos factores esenciales que introducen sesgo en los datos: la variación biológica y los errores técnicos.

### 3.1.1. Procesamiento de datos brutos

Una vez generados los datos, se debe organizar el resultado para extraer datos de los archivos de datos brutos y para convertir las mediciones en datos científicos disponibles. Los datos brutos con formatos comerciales pueden ser convertidos a formatos libres mediante el uso de software suministrado por distribuidores instrumentales o a través de plataformas de acceso gratuito y código abierto para el preprocesado y análisis de datos metabólicos (por ejemplo, XCMS).

### 3.1.2. Preprocesamiento de datos

El procesamiento de datos se refiere a todas las acciones que afectan directamente a los datos brutos antes de continuar con su análisis. El problema principal es aumentar la calidad de la señal y reducir los posibles sesgos presentes en los datos brutos. Esta etapa incluye: (1) la reducción del ruido; (2) la detección y deconvolución de picos; (3) el alineamiento; (4) las técnicas de imputación de valores perdidos y (5) la filtración de datos.

#### ***(1) Reducción del ruido***

Se recurre a una corrección de partida, la filtración de ruido, para eliminar artefactos de baja frecuencia y diferencias aparentes que generan el protocolo experimental y la variación instrumental. Esta etapa facilita la detección de picos y reduce potencialmente la detección de falsos positivos. Existe algún software que ha integrado en una sola función los algoritmos de filtración y detección de picos.

#### ***(2) Detección y deconvolución de picos***

El propósito de la detección de picos y la recogida de picos (deconvolución) es identificar o cuantificar las señales de los espectros que proceden de los compuestos detectados en la muestra.

Cada pico del cromatograma o el electroferograma aparece representando la intensidad (abundancia) en el eje Y y el tiempo de retención (RT) o el tiempo de migración (MT) en el eje X. Esas intensidades se miden en cada escaneo, en un conjunto de varios miles, dentro del rango de masas establecido durante la ventana de tiempo de detección. Por otra parte, el gráfico obtenido por MS, el espectro de masas, ofrece información dual además del tiempo de separación. En muestras complejas, es frecuente que la coelución de diferentes compuestos dé un solo pico y, además, que el espectro de masas a lo largo del pico sea la suma de todos los compuestos coeluyentes. La deconvolución es el proceso de separar computacionalmente componentes coeluyentes y crear un espectro puro para cada componente a través de la extracción de iones de un cromatograma de iones totales (TIC) complejo, aun cuando la señal del compuesto objetivo se encuentre a niveles ínfimos.

Los picos se detectan en todo el espectro y sus áreas se integran para proporcionar la cuantificación del metabolito subyacente. Como resultado de este proceso, las matrices de datos

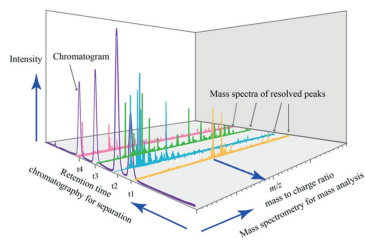
complejas tridimensionales de MS se convierten en listas de características metabólicas anotadas que se obtienen en sus RT o MT correspondientes adoptando la forma de una relación masa/carga ( $m/z$ ) y una intensidad relativa de los compuestos medidos.

El método de detección de picos de señal debe identificar las señales verdaderas y evitar los falsos positivos. La metodología difiere dependiendo de que las fuentes de ionización sean duras (GC-EI-MS o LC-ESI-MS) o blandas (CE-ESI-MS)

### GC-EI-MS

Cuando se trabaja con una fuente de EI, el espectro de masas que se obtiene es fundamentalmente una huella de la molécula que se puede usar para identificar cualquier compuesto. En último término, la GC/MS proporciona un gráfico en 3D con un cromatograma y un espectro para cada componente separado. La primera separación de picos se deriva de la resolución cromatográfica, en la que se separan físicamente los analitos de la muestra siguiendo el orden del punto de ebullición además de su interacción selectiva con la fase estacionaria de la columna analítica, eluyendo a tiempos de retención distintos, aunque existen analitos que se eluyen en tiempos de retención similares. La segunda separación viene provocada por la resolución espectral, en la que los analitos que se eluyen en tiempos de retención similares (porque no tienen una resolución cromatográfica completa) se separan con el analizador de masas en función de sus distintas (y únicas) relaciones masa/carga a través.

El cromatograma de la corriente de iones totales (TIC) representa la suma de intensidades en todo el rango de masas que se detecten en cada punto del análisis. En el cromatograma de ion extraído (EIC), necesario para la ejecución de un análisis cromatográfico se recuperan («extraen») del conjunto total de datos uno o más valores  $m/z$  que representen uno o más analitos de interés. En la Figura 2.5 se muestran los datos tridimensionales obtenidos por GC/MSMS.



**Figura 2.5.** Datos tridimensionales de cromatograma GC/MS: Tiempo de retención, espectro de MS e intensidad.

La utilidad de los estudios metabólicos depende en gran medida del número de metabolitos identificados y de los vínculos a su interpretación biológica. La fragmentación de los metabolitos durante la EI es muy característica de la estructura química, de modo que estos espectros de masas se pueden emplear para la identificación de compuestos a través de librerías de espectros de masas. Este procedimiento puede dar resultados satisfactorios si el espectro se corresponde con el componente puro sin contaminación de masas que pertenezcan a otras fuentes (coelución, fondo, sangrado de columna). Además en muestras complejas, se puede producir la elución simultánea

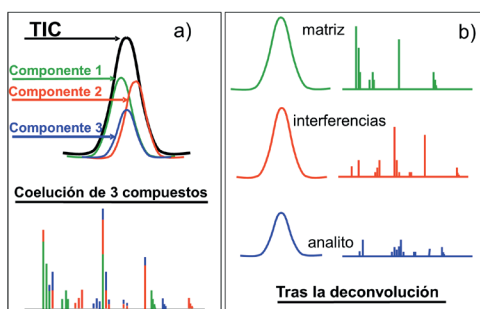
de múltiples analitos, ocultando especies individuales, y el cromatograma TIC suele proporcionar una información limitada.

Cuando se extraen espectros de masa para identificar componentes desconocidos, puede resultar difícil apreciar si existe coelución con otros componentes. Una coelución produciría un espectro de masas combinado y cuando este espectro de masas combinado se busque en una librería, lo más probable es que encuentre coincidencias de escasa calidad, ninguna coincidencia potencial o una identificación incorrecta. Existen diversas formas de buscar coeluciones, y esto, por ejemplo, se puede hacer de forma manual haciendo clic por el pico para ver si el espectro de masas cambia a medida que el cursor se desplaza por el pico o extrayendo los iones fundamentales para producir cromatogramas de ion extraído que, cuando se superponen, tienen que mostrar las mismas formas y tiempos de retención, indicando que todos ellos pertenecen al pico del compuesto de interés.

### Deconvolución de espectros en GC-EI-MS

Existen formas de mejorar la calidad del espectro de masas antes de la identificación por búsquedas en librerías de espectros. La más común es la sustracción del fondo (background) en ambos extremos del pico. Sin embargo, en la mayor parte de las funciones de software, la sustracción del fondo y la extracción manual de iones pueden constituir un proceso largo y engorroso, especialmente cuando el cromatograma presenta muchos picos y hay muchas muestras de análisis. Para identificar y extraer información cuantitativa de los metabolitos correspondientes, se debe construir el espectro para cada uno de los metabolitos. En el procesamiento de datos de GC-EI-MS, la etapa de construcción del espectro se denomina deconvolución. Por lo tanto, como ya se explicó antes, la deconvolución es el proceso de separar computacionalmente componentes coeluyentes y crear un espectro puro para cada componente. Concretamente, para cada EIC observado que resulte de dos o más componentes, la deconvolución calcula la contribución de cada componente al EIC.

El software libre utilizado en esta técnica para la extracción de iones, incluso a niveles ínfimos, es AMDIS (Automated Mass Spectrometry Deconvolution and Identification System), desarrollado por el Instituto Nacional de Normas y Tecnología de EE. UU. (NIST, por sus siglas en inglés). En la siguiente figura se muestra el resultado del proceso de deconvolución.

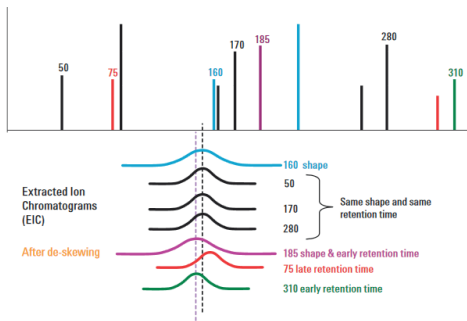


**Figura 2.6.** a) Antes y b) Después del proceso de deconvolución (extraído de <https://www.agilent.com/cs/library/Support/Documents/f05017.pdf>).

El proceso global de deconvolución en AMDIS se compone de cuatro etapas secuenciales: 1) Análisis del ruido, 2) Percepción de componentes, 3) Forma del modelo y 4) Deconvolución espectral.

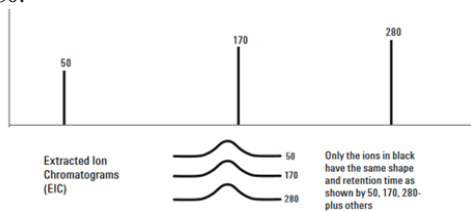
En la primera fase, se extraen las características del ruido de un archivo de datos de GC-MS calculando el factor de ruido que se utilizará para la representación de la magnitud de la señal en unidades de ruido. En la segunda fase, búsqueda de componentes, se buscan los componentes cromatográficos individuales. La justificación para la percepción de componentes es que un componente existe cuando se maximiza simultáneamente una magnitud suficiente de iones. En la tercera etapa, se determinan los picos modelo que se utilizarán en la siguiente etapa para la deconvolución. La forma del modelo para cada componente percibido se recoge como la suma de los cromatogramas de iones individuales que se maximizan conjuntamente y cuyos valores de agudeza (sharpness) se sitúan dentro del 75% del valor máximo para este componente. En la última etapa, la deconvolución propiamente dicha, se extraen para cada componente los espectros «purificados» procedentes de los cromatogramas de iones individuales a través de las formas de modelos y del método de mínimos cuadrados.

AMDIS considera las formas de los picos de todos los iones extraídos y los tiempos de retención (RT) de sus vértices. En el ejemplo de la Figura 2.7, tan solo se superponen para mayor claridad algunos de los cromatogramas de ion extraído (EIC). El EIC del ion 160 tiene el mismo RT que los iones 50, 170 y 280, pero la forma de su pico es diferente. El ion 185 tiene una forma de pico distinta y un RT menor. Los iones 75 y 310 tienen formas de pico similares pero distintos RT.



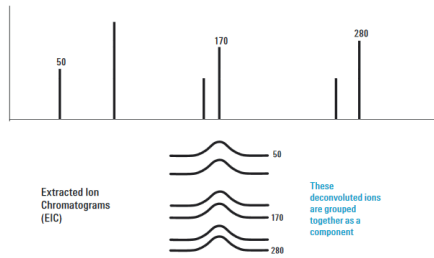
**Figura 2.7.** Cromatogramas de ion extraído de determinados iones (imagen proporcionada por Agilent).

En la Figura 2.8 figuran los EIC después de eliminar las distintas formas de pico o RT que resultan ser los iones 50, 170 y 280.



**Figura 2.8.** Iones con la misma forma y tiempo de retención (imagen proporcionada por Agilent).

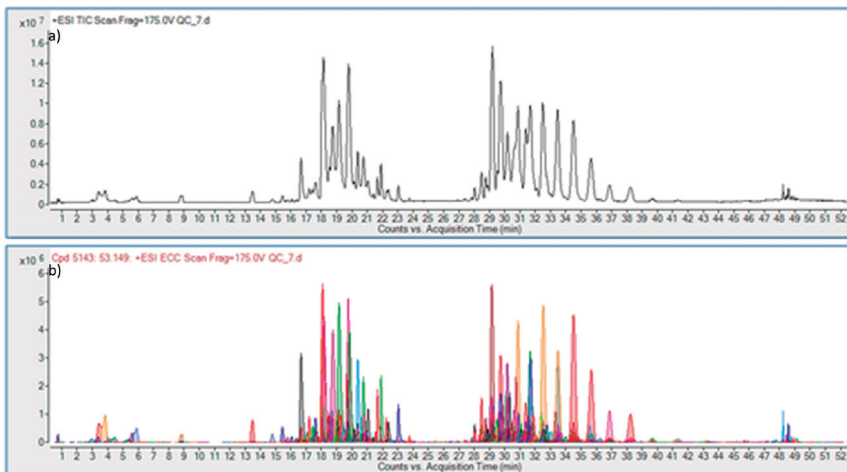
La Figura 2.9 muestra todos los iones con formas de pico y RT similares dentro de los criterios fijados previamente por el analista. AMDIS los agrupa y los designa como un componente.



**Figura 2.9.** Espectro de masas de un componente tras la deconvolución de AMDIS (imagen proporcionada por Agilent).

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

Los métodos basados en picos son los algoritmos más comunes para la detección de características en estudios de MS. A la hora de determinar si los distintos iones son de la misma característica metabólica, estos algoritmos, como por ejemplo el Molecular Feature Extractor (Agilent), consideran la exactitud de las mediciones de masas para iones vinculados a un grupo por envolvente de estado de carga, distribución isotópica, y las posibles relaciones químicas. Puede considerar iones vinculados como aductos: aductos de protones, sodio, potasio y amonio en la ionización positiva o pérdida de un protón, aductos con formiato, cloruro, etc., en el modo de ionización negativa. En la Figura 2.10 se muestra el cromatograma antes y después de la deconvolución.



**Figura 2.10.** Comparativa de dos cromatogramas del mismo análisis de una mezcla compleja a) Cromatograma de Iones Totales b) Cromatogramas de cada uno de los compuestos obtenidos después de la deconvolución.

### **(3) Alineamiento**

El alineamiento de espectros es una de las etapas principales del preprocesamiento en los estudios metabolómicos donde se realizan múltiples análisis de muestras. Cuando se analizan múltiples espectros, la posición de los picos correspondiente a la misma característica metabólica se puede ver afectada por desplazamientos no lineales. En los estudios basados en MS, los desplazamientos de los picos se observan principalmente a través del eje del RT. Por tanto, se debe aplicar el alineamiento espectral (coincidencia de pico por  $m/z$  y RT entre todas las muestras) para corregir esa variabilidad en las muestras analizadas, puesto que puede afectar considerablemente a la calidad del estudio. Los algoritmos de alineamiento de espectros se pueden dividir en dos grupos principales: (1) aquellos en los que se alinean los datos antes de la detección de picos; y (2) métodos de alineamiento basados en el pico, donde se alinean en todas las muestras los picos espectrales detectados.

### **(4) Imputación de valores en blanco**

Lamentablemente, no todos los métodos de preprocesamiento de datos generan un conjunto completo de datos y tratar con valores en blanco (vacíos) en una matriz de datos supone un problema considerable en los estudios metabolómicos. Existen varias fuentes de celdas con valores en blanco, como por ejemplo: (1) límites de la detección computacional; (2) Baja intensidad de la señal; (3) error de medición; (4) errores de deconvolución/alineamiento; (5) errores en la identificación de las señales del fondo; o (6) simplemente, la ausencia del metabolito en la muestra. En general, se podrían identificar tres tipos de valores perdidos: (1) datos perdidos completamente al azar (MCAR, por sus siglas en inglés), cuando los datos perdidos no están relacionados con ninguna variable o muestra observada; (2) datos perdidos al azar (MAR), cuando los datos perdidos están relacionados con una o más variables observadas pero no con la muestra; (3) datos perdidos sin ser al azar (MNAR), cuando los datos perdidos están relacionados con la propia muestra. Los valores perdidos suponen un serio problema para continuar con el procesamiento y el análisis estadístico de los datos. Los valores perdidos se originan por distintos factores, en particular concentraciones por debajo de los límites de detección en los sistemas de MS, pero también el error de identificación que cometa el software. Por ello, se han desarrollado distintos algoritmos para la imputación de valores perdidos (MVI). Estos métodos incluyen: (1) sustitución por cero; (2) sustitución por media/mediana, valor mínimo, mitad del valor mínimo; (3) el k-medias vecino más cercano (kNN); (4) el PCA Probabilístico (PPCA); (5) el bosque aleatorio (RF); (6) el método PCA bayesiano (BPCA).

### **(5) Filtrado de datos**

La matriz de datos metabolómicos normalmente se construye con miles de características de metabolitos de múltiples muestras. Antes del análisis estadístico, esta muestra de datos tiene que

someterse a una nueva filtración para reducir el número de variables. Existen diferentes maneras de filtrar los datos, pero la más generalizada es la de sustituir las características metabólicas con una mala reproducibilidad. En los estudios de metabolómica se emplean dos estrategias principales: (1) la filtración por presencia, basada en el porcentaje de las muestras presentes en el grupo de muestras; (2) la filtración basada en los procedimientos de garantía de calidad y el umbral del valor de la desviación estándar relativa calculado para cada característica metabólica detectada en las muestras de control de calidad. Asimismo, se deben depurar los datos de sesgos experimentales sustrayendo el valor obtenido para las muestras en blanco.

### 3.1.3. Pretratamiento de datos

Existen varios factores (por ejemplo, variaciones experimentales y biológicas no deseadas y errores técnicos) que pueden causar dificultades en la identificación de diferencias significativas en los perfiles metabólicos. Para eliminar determinados tipos de variaciones no deseadas, se adoptan la corrección de las desviaciones de la señal y los métodos de escalado.

#### a) Normalización

La normalización se aplica para corregir diferencias de intensidad de picos no deseadas y para estabilizar la variación dentro del conjunto de datos. La variación experimental no deseada puede surgir de un error humano (por ejemplo, en la extracción y preparación de muestras), dentro de la variación del instrumento (por ejemplo, cambios de temperatura dentro del instrumento, la degradación de la muestra o la pérdida de rendimiento del instrumento durante un análisis largo de muestras), por la variación entre distintos instrumentos, distintos lotes, distintos laboratorios y distintas plataformas analíticas. También es común la variación biológica desconocida (por ejemplo, en el número de células o en la concentración del biofluido) y puede confundirse con los factores de interés. La normalización de esas interferencias se puede llevar a cabo aplicando un estándar interno como referencia para el cálculo de los errores analíticos observados. Los métodos de normalización que no se basan en estándares internos con frecuencia aplican como factor de normalización la suma, la media o la mediana de las respuestas de todos los metabolitos de una muestra. Además existen otros métodos como, por ejemplo, la normalización del cociente de probabilidad (PQN), la normalización del cambio múltiple promedio (FC), la normalización cuantil o la normalización del alisado de diagramas de dispersión localmente ponderados (LOESS). La elección de la técnica de normalización depende de un conjunto de presupuestos sobre las mediciones metabólicas y la variabilidad biológica del conjunto de datos.

#### b) Transformación

Las abundancias de metabolitos en una matriz de datos normalmente tienen una distribución sesgada a la derecha, por lo que con frecuencia se aplica una transformación logarítmica antes del análisis estadístico.



### c) Escalado

El escalado se realiza para ajustar las diferencias de cambio múltiplo entre metabolitos que puedan venir causadas por la existencia de grandes diferencias en la variación de las respuestas medidas. Los métodos de escalado dividen cada variable por un factor, el factor de escalado, que para cada variable es distinto. Distintos métodos de escalado se han ido aplicando en la metabolómica, incluyendo entre otros el autoescalado, el escalado Pareto, el escalado de rango y el escalado VAST.

#### Para más información:

- O. Fiehn. Metabolomics--the link between genotypes and phenotypes. *Plant Mol. Biol.*, 2002, 48(1-2):155-71.
- W.B. Dunn, D.I. Broadhurst, H.J. Atherton, et al. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.*, 2011, 40(1):387-426.
- W.B. Dunn, I.D. Wilson, A.W. Nicholls, and D. Broadhurst. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis*, 2012, 4:2249-2264.
- A. Alonso, S. Marsal and , A. Julià. Analytical methods in untargeted metabolomics: state of the art in 2015. *Frontiers Bioeng. Biotech.*, 2015, 3:23.
- R. Di Guida, J. Engel, J.W. Allwood, et al. Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. *Metabolomics*, 2016, 12(5):1-14.

## 3.2. Identificación de metabolitos

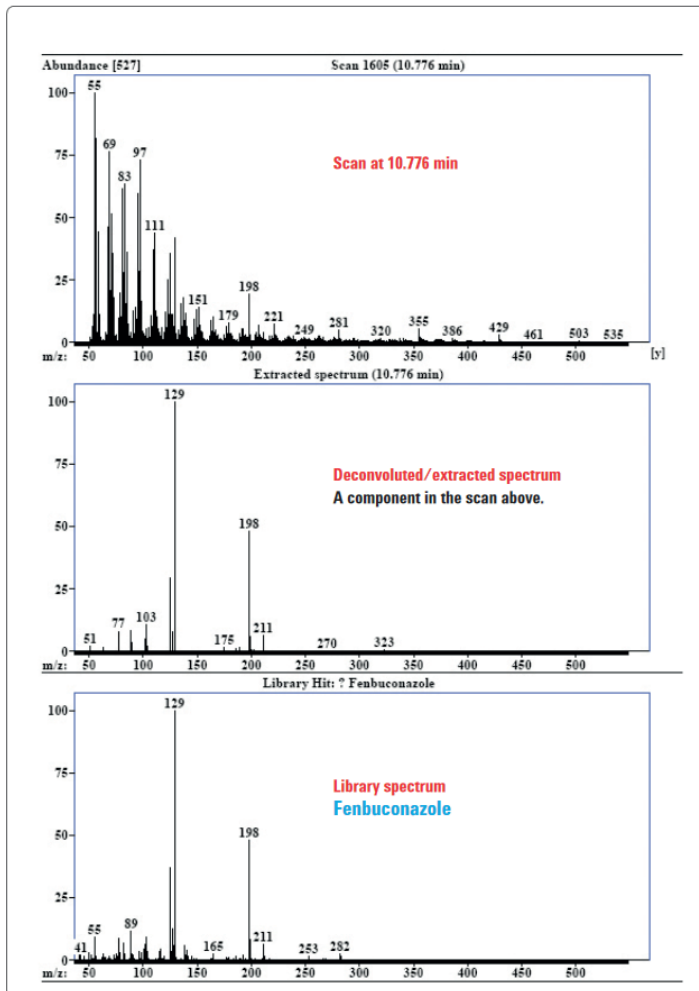
Con frecuencia, una de las etapas más problemáticas es la identificación de metabolitos. Según Donald Rumsfeld (Secretario de Defensa de EE. UU.): « (...) como sabemos, hay conocidos conocidos; hay cosas que sabemos que sabemos. También sabemos que hay conocidos no conocidos, es decir, sabemos que hay cosas que no sabemos. Sin embargo, también hay no conocidos no conocidos; los que no sabemos que no sabemos»

### 3.2.1. Identificación en GC-EI-MS

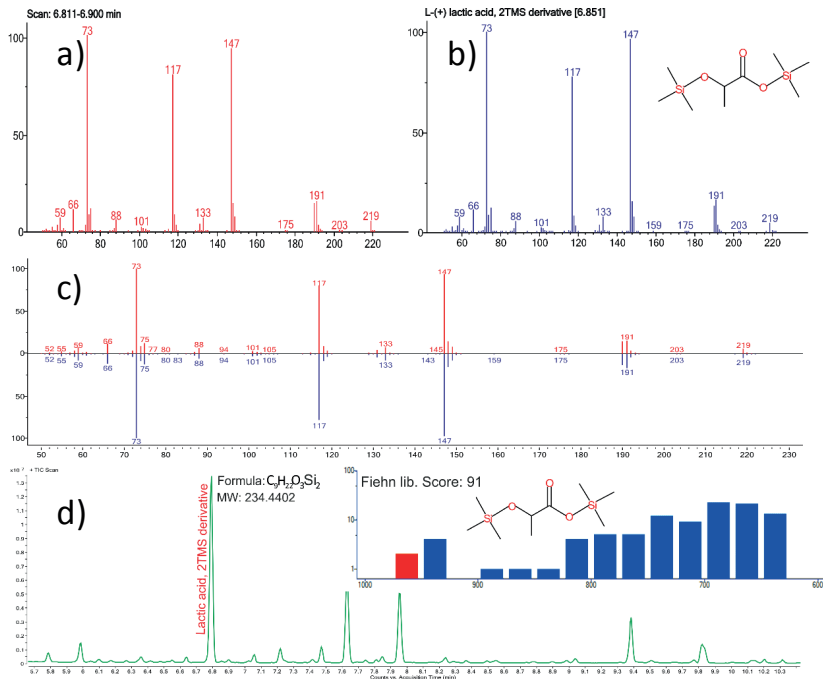
En GC-EI-MS, la supuesta identificación de metabolitos se podría realizar empleando dos criterios: buscando el espectro puro obtenido a través de la deconvolución con las librerías espectrales comerciales y haciendo coincidir el RT con una biblioteca de espectros con RT bloqueado para todas las entradas (tiempo de retención constante para el mismo analito obtenido a través del mismo método de análisis). En función del factor de coincidencia de la búsqueda, se pueden identificar o marcar los compuestos objetivo en un TIC complejo. Se realizará una búsqueda para cada componente con una biblioteca de tiempos de retención (llamada biblioteca objetivo). Existen librerías de espectros objetivo de metabolómica, librerías RTL (acrónimo inglés para «librería de tiempos de retención bloqueados») que se han desarrollado expresamente para facilitar la identificación de metabolitos. Debido a su alta reproducibilidad, resolución de los picos cromatográficos y la existencia de librerías de espectros de masas, la GC-EI-MS está considerada el patrón de referencia para la búsqueda metabolómica.

La potencia de deconvolución se aprecia al comparar los dos espectros de arriba de la Figura 2.11. El escaneo bruto o escaneo no deconvolucionado original aparece arriba. El escaneo limpio, que es el componente deconvolucionado, aparece en el medio. El escaneo de abajo es el compuesto identificado en una biblioteca espectral. Sin la deconvolución, el analista compararía visualmente el escaneo bruto con fondo sustraído y los escaneos de librerías para la confirmación.

Con este tipo de comparación sería difícil, cuando no imposible, decir que está presente el Fenbuconazol, el compuesto objetivo de este ejemplo.



**Figura 2.11.** Comparación de espectros bruto, deconvolucionado y de librería (imagen proporcionada por Agilent Technologies).



**Figura 2.12.** Identificación de un metabolito en un cromatograma de GC-EI-MS haciendo coincidir el RT y el espectro con una librería objetivo. a) Espectro obtenido, b) Espectro de la librería objetivo, c) Diagrama de comparación de espectros, d) Cromatograma TIC con la puntuación de búsqueda en una escala de 100.

El número de compuestos identificados será invariablemente mayor cuando los experimentos se realicen con instrumentación de masas precisa utilizando también bases de datos de masas precisas hechas a medida. Además, una instrumentación de masas precisa facilitará la identificación de no conocidos a través de la generación de fórmulas de picos moleculares y sus fragmentos.

### 3.2.2. Identificación de metabolitos en experimento de metabolómica de LC o CE-[ESI]-MS: Trabajando con bases de datos

Comúnmente se acepta que en un experimento metabolómico normal existen 3 niveles de identificación de metabolitos, del menos al más identificado: aproximadamente el 50% de los compuestos serían compuestos sin identificar («desconocidos»), pero dentro del otro 50% pueden existir compuestos supuestamente identificados en una clase de compuesto, gracias a un proceso de coincidencias de consultas donde se utilizan distintas bases de datos. Esta identificación se puede refinar con más datos instrumentales como el tiempo de retención o la fragmentación MS/MS. Los compuestos se pueden identificar positivamente todavía más cuando se analiza un estándar conocido bajo las mismas condiciones y los mismos resultados de la coincidencia del desconocido y del estándar.

Las bases de datos para metabolómica han experimentado una importante evolución en los últimos años, en extensión (número de metabolitos), profundidad (campos de información por metabolito) y funcionalidad (tipos de consultas). Por ejemplo, la Human Metabolome DataBase –base de datos del metaboloma humano– (HMDB) se lanzó por primera vez en 2006 con 2.180 metabolitos y actualmente cuenta con más de 100 campos de información completamente consultables para cada compuesto en una lista que comprende más de 42.000 metabolitos.

No obstante, los trabajos con bases de datos en el campo de la metabolómica están lejos de estar tan estandarizados como en las demás «ómicas», y los investigadores de este campo tienen que trabajar con distintas bases de datos y combinar la información que proporcionan, dado que una lista de bases de datos relacionados con la metabolómica se puede ampliar hasta más de 50 recursos en línea distintos. Este es el motivo por el que en los últimos años han aparecido distintos mediadores para permitir que los investigadores gestionen información procedente de diferentes fuentes.

Las bases de datos que se pueden utilizar en la metabolómica se pueden clasificar, según su tipo de información, como centradas en la ruta metabólica o centradas en el metabolito. Y, a su vez, este último tipo se puede continuar subdividiendo en bases de datos espectrales o bases de datos de compuestos. No se puede considerar esta clasificación una organigrama con límites perfectamente definidos, ya que se están modificando continuamente y parte de la información se puede encontrar en distintas bases de datos.

El siguiente apartado trata de resumir algunas de las características más importantes de algunas de las bases de datos más relevantes actualmente en el campo de la metabolómica. Es importante subrayar que, aunque se utilizan en investigaciones metabolómicas, el propósito principal de algunas bases de datos no es el campo de la metabolómica; este es el caso de PubChem o KEGG, por poner dos ejemplos.

### 3.2.3. Bases de datos centradas en el metabolito sobre compuestos para búsquedas simples de PM

**ChEBI** (entidades químicas de interés biológico, por sus siglas en inglés) es una iniciativa del Instituto Europeo de Bioinformática, perteneciente al Laboratorio Europeo de Biología Molecular (EMBL). Incluye compuestos de distintas fuentes para proporcionar descripciones más estandarizadas de entidades moleculares (cerca de 45.000 han recibido una calificación de «3 estrellas» según el curador) que hacen posible que otras bases de datos realicen la anotación de sus entradas de una manera uniforme. ChEBI se centra en la anotación manual de alta calidad, en evitar la redundancia y el suministro de ontología química. Se pueden hacer búsquedas por nombre, fórmula o estructura.

**PubChem** es la base de datos de los Institutos Nacionales de la Salud de EE.UU. que contiene muchos millones (cerca de 100) de compuestos (de menos de 1.000 átomos) y sustancias (más de 200 millones). Los datos provienen de 80 proveedores/depositantes distintos. Las sustancias son «impuras/réplicas», mientras que los compuestos son entidades únicas (el identificador de compuestos PubChem arroja un único compuesto). Las entradas incluyen sinonimias, propiedades

químicas, la estructura química con identificadores SMILES e InChI, bioactividad y enlaces a compuestos relacionados estructuralmente y a otras bases de datos del Centro Nacional para la Información Biotecnológica de EE. UU. (NCBI, por sus siglas en inglés), como PubMed. Se pueden hacer búsquedas por nombre, fórmula, rango de PM, estructura, recuento de enlaces de hidrógeno dador/aceptor o XlogP.

**ChemSpider** es una base de datos de estructuras químicas que proporciona un acceso rápido a más de 58 millones de estructuras, propiedades e información asociada. Al integrar y enlazar compuestos de aproximadamente 500 fuentes de datos, ChemSpider posibilita que los investigadores obtengan la perspectiva más completa de los datos químicos de acceso libre a partir de una única búsqueda en línea. Perteneció a la Royal Society of Chemistry. Se puede buscar por nombre, sinonimia, InChI, estructura, número de registro, SMILES y propiedades calculadas (pero no por fórmula o masa). Los datos incluyen nombres, sinonimias, artículos de Wikipedia, descripciones, fuentes de datos, proveedores, patentes, artículos, propiedades, Encabezados de Temas Médicos (MeSH), enlaces farmacológicos y espectros (UV, IR, NMR, MS) procedentes de otros sitios web.

### 3.2.4. Bases de datos espectrales centradas en el metabolito

Estas bases de datos no solo son compatibles con búsquedas de PM o rangos de PM, sino que también admiten búsquedas de iones precursores (positivos, negativos, neutros), búsquedas de listados de picos (a partir de datos de MS o MS/MS) así como coincidencias espectrales de MS/MS. Estas bases de datos, en lugar de ser simples herramientas de búsqueda de PM, están ideadas para la identificación metabolómica y de compuestos basada en la MS.

**BioCyc**, desarrollada por SRI International (Menlo Park, California), es un conjunto de bases de datos curadas para distintos organismos. Las bases de datos se organizan según la cuantía de actualizaciones manuales que hayan recibido. Las bases de datos de Nivel 1, como EcoCyc (para *E. Coli*) y HumanCyc, han sido objeto de una amplia curación, mientras que la mayor parte de las bases de datos BioCyc (de Nivel 2 y 3) han sido obtenidas computacionalmente. Estas bases de datos se aplican especialmente a la identificación de metabolitos de organismos específicos y a la reconstrucción metabólica utilizando la búsqueda de rutas metabólicas.

**HMDB** es una base de datos dedicada al metabolismo humano que ha sido desarrollada con el apoyo de Canadian Institutes of Health Research, Alberta Innovates - Health Solutions y el Metabolomics Innovation Centre. Por cada entrada de datos, se proporciona información sobre las características químicas, biológicas y clínicas, además de bibliográficas que incluyen asociaciones de enfermedades de las que se informa, enzimas relacionadas y transportadores, además de enlaces a bases de datos externas como KEGG.

**Komic Market** (Kazusa Omics Data Market) es una base de datos de anotaciones de metabolitos de picos de MS detectados en estudios metabolómicos. Proviene del proyecto Development of Fundamental Technologies for Controlling the Material Production Process of Plants (Desarrollo de tecnologías fundamentales para el control del proceso de producción material de las plantas), con el

apoyo de la New Energy and Industrial Technology Development Organisation de Japón.

**LipidBank** es la base de datos oficial de la Japanese Conference on the Biochemistry of Lipids – Conferencia japonesa sobre la bioquímica de lípidos– (JCBL). Esta base de datos está dedicada a los lípidos neutros. Cubre varias clases distintas y toda la información molecular es objeto de curación manual y aprobada por expertos en investigación de lípidos. Cada entrada incluye el nombre de un lípido, su estructura molecular, información espectral y referencias bibliográficas.

**LipidMaps** se creó a través de una beca de investigación colaborativa a gran escala («Glue Grant») del Instituto Nacional de Ciencias Médicas Generales de los EE.UU., perteneciente a los NIH. Su objetivo es ofrecer identificación y cuantificación de los lípidos de los mamíferos, en particular la cuantificación de los cambios que se producen en respuesta a perturbaciones. En este recurso se incluye también la base de datos LipidMaps Proteome Database (LMPD)

**MassBank** es un registro público de datos espectrales de masas que tiene por base compartir identificaciones y elucidaciones estructurales de compuestos químicos que se detecten mediante espectrometría de masas. Se puede acceder a MassBank a través de dos dominios: uno japonés (<http://massbank.jp>) y otro europeo (<http://massbank.eu>) (NORMAN MassBank). La herramienta está implantada en los dos dominios, pero algunas funciones solo están disponibles en el japonés.

**METLIN** es una marca registrada del Scripps Research Institute, que desarrolla y aplica tecnologías basadas en la espectrometría de masas para entender el metabolismo. Incluye informática de procesamiento de datos en la nube (XCMS) y espectrometría de masas de imagenología de nanoestructuras (NIMS). Con casi 1.000.000 de entradas de compuestos reales (no de predicción), constituye una de las mayores bases de datos disponibles. Las entradas en METLIN incluyen metabolitos, lípidos, esteroides, metabolitos de plantas y bacteriales, péptidos pequeños y metabolitos y agentes tóxicos de fármacos exógenos. También se incluye IsoMETLIN, un módulo para la metabolómica basada en isótopos.

**MycompoundID** es un recurso web desarrollado por la Universidad de Alberta para la identificación de compuestos en función de sus propiedades químicas, incluida la masa exacta. Es posible realizar distintas búsquedas, en particular búsquedas MS, MS2 y PEP de péptidos sin marcaje y con marcaje dimetilico y datos MS con marcaje isotópico químico. Las búsquedas se realizan en una biblioteca metabolómica de base empírica (EML, por sus siglas en inglés) formada por 8.021 metabolitos endógenos humanos conocidos y sus productos metabólicos previstos, entre los que se incluyen 375.809 compuestos procedentes de una reacción metabólica y 10.583.901 de dos reacciones. Se generan compuestos previstos in silico a partir de entradas de HMDB.

### 3.2.5. Mediadores

**CEU MassMediator**, un proyecto desarrollado conjuntamente por el CEMBIO y el Laboratorio de Bioingeniería de la Facultad Politécnica de la Universidad CEU San Pablo de Madrid, es una herramienta que realiza una búsqueda automatizada a través de fuentes de datos externas (HMDB, KEGG, LipidMaps, METLIN y MINE) y que ofrece las posibles identificaciones para una masa dada (unificando en una sola correspondencia los resultados similares obtenidos en más de una

base de datos).

**CSI:FINDERID** es una base de datos específica para la identificación MSn. Es compatible con búsquedas avanzadas en picos en MS. Es un proyecto desarrollado conjuntamente entre la Universidad Friedrich Schiller, en Alemania, y el Helsinki Institute for Information Technology de la Universidad Aalto, en Finlandia, que combina la computación mediante árboles de fragmentación y el aprendizaje automático para mejorar el porcentaje total de moléculas identificadas y la precisión de la identificación.

**MAGMa** es una herramienta de anotación desarrollada dentro del proyecto eMetabolomics, creado por el Netherlands eScience Center de la Universidad de Wageningen, en colaboración con el Netherlands Metabolomics Centre. Se pueden descargar datos MSn como un árbol jerárquico de picos de fragmentos, a partir de la  $m/z$  o fórmulas químicas, y las moléculas candidatas se recuperan automáticamente de PubChem, KEGG o HMDB. Se calcula una puntuación de coincidencia basada en la calidad de la explicación de los picos de fragmentos.

**MassTRIX** es una herramienta en línea para la anotación de datos de espectrometría de masas de alta exactitud. Los resultados se presentan en mapas de rutas metabólicas KEGG específicas para cada organismo y se puede añadir cualquier otra información genómica o transcriptómica. La herramienta fue desarrollada en el Helmholtz Zentrum München, en una colaboración entre Philippe Schmitt-Kopplin y Karsten Suhre.

### 3.3. Análisis estadístico

Una vez obtenida la matriz con datos limpios, en los que se contenga la lista de posibles compuestos y sus abundancias en cada muestra, el siguiente paso es la elaboración del análisis estadístico utilizando herramientas quimiométricas, que ofrece descripciones basadas en modelos de la variación biológica del sistema objeto de estudio. En el análisis de datos de la metabolómica no dirigida se utilizan dos análisis estadísticos distintos: análisis de datos univariante y análisis de datos multivariante (MVDA).

El análisis de datos univariante parte de la hipótesis de que solo hay un factor con influencia en la respuesta de las variables. En una enfermedad compleja intervienen y se ven afectadas al mismo tiempo distintas rutas biológicas reguladas por múltiples variables. El análisis estadístico tradicional suele transformar todo los problemas en problemas univariantes, incluso los que son multivariantes por naturaleza. Por esta razón, el MVDA es una herramienta estadística adecuada para la interpretación de datos de un estudio metabolómico. Sintetiza tablas de datos con muchas variables y pocas observaciones y funciona reduciendo el número de variables y clasificando los datos. El MVDA ofrece modelos estadísticos que diferencian específicamente representantes de metabolitos de interés (picos anotados) y que se pueden identificar posteriormente de manera definitiva, química o estructuralmente. El primer paso del MVDA es la creación de la matriz X, donde las muestras se organizan en filas y todas las variables (los compuestos) en columnas. Seguidamente, siempre es necesario realizar un pretratamiento (escalado y transformación) de los

datos para mejorar la información respectiva. Una vez hecho esto, los datos se analizan mediante el análisis de componentes principales (PCA) no supervisado, el análisis discriminante supervisado mediante regresión de mínimos cuadrados parciales (PLS-DA), y el análisis discriminante por mínimos cuadrados parciales ortogonales (OPLS-DA). Ambos métodos reducen la matriz de datos original de tal manera que se puedan revelar patrones subyacentes.

### 3.3.1. PCA

Es un método de reducción de dimensiones ampliamente utilizado en la exploración y visualización de datos. Fue propuesto por primera vez en 1901 por Pearson. El PCA busca revelar patrones subyacentes comprimiendo los datos, tratando de retener la mayor cantidad posible de la información original. El PCA representa el punto de partida natural de cualquier análisis de datos multivariante. Resulta útil cuando lo que se quiere es ofrecer una perspectiva gráfica global de todas las muestras que estén en la matriz de datos, revelando valores atípicos, grupos, aglomerados, similitudes/disimilitudes y variables dominantes. La posición de cada muestra en el diagrama de dispersión tipo se utiliza para relacionarlas entre sí: las muestras que se encuentran próximas entre ellas poseen un perfil multivariante similar; en cambio, las muestras que están lejos entre ellas poseen propiedades dispares.

#### ***PCA y Detección de Valores Atípicos***

Los valores atípicos se definen como puntos de datos cuyos valores difieren del de la mayoría de los demás puntos de datos. Tienen influencia sobre la validez de los resultados metabólicos, ya que alteran la variación de los datos y su distribución, reduciendo así el poder estadístico del análisis de datos. Los valores atípicos pueden ser biológicos o analíticos, dependiendo de la fuente. Los biológicos son difíciles de reconocer. Son el resultado de variaciones biológicas aleatorias o inducidas entre las muestras que ocurren con frecuencia durante el estudio de enfermedades complejas con un grupo heterogéneo de pacientes. Además se debe incluir en el análisis estadístico la información procedente de valores atípicos biológicos aleatorios. En el caso contrario, los valores atípicos analíticos causan una distorsión elevada de la información biológica y se deben excluir del análisis estadístico. Podrían tener diferentes orígenes: muestreo, almacenamiento, tratamiento de muestras, análisis, búsqueda de características o proceso de deconvolución/identificación. La decisión de mantener o eliminar los valores atípicos tiene que estar siempre justificada.

Una vez obtenido el modelo de PCA, se pueden destacar los valores atípicos mediante la representación gráfica de la intensidad de los compuestos totales detectados en oposición al orden de análisis y observando el estadístico T2 de Hotelling y la opción «Distancia al Modelo» (disponible en SIMCA P+). Son necesarias distintas gráficas y análisis de datos brutos y filtrados para individualizar y eliminar los valores atípicos. En una gráfica de puntuaciones S-plot (scores plot) de PCA, los valores especialmente atípicos aparecen como puntos muy alejados de la frontera elíptica, representando el 95% de los intervalos de confianza de la variación del modelo. El gráfico del estadístico T2 de Hotelling muestra la distancia desde el origen en el plano del modelo (espacio



de puntuación) para cada observación seleccionada (representada como una columna). Si se localiza un valor muy por encima del límite crítico, eso quiere decir que no sigue la norma en el rango de componentes seleccionado dentro del espacio de puntuación. Por este motivo, la probabilidad de que las observaciones sean de la misma clase que las otras muestras es menor del 5%.

### 3.3.2. PLS-DA

Es necesario diferenciar entre los grupos y maximizar sus diferencias. Modela la relación entre dos o más clases de datos a través de una serie de ajustes por mínimos cuadrados locales. El PLS (análisis por mínimos cuadrados) evalúa una relación entre una matriz descriptiva  $X$  y una matriz de respuesta  $Y$ . La diferencia fundamental entre el PLS y el PCA (análisis de componentes principales) es que el PLS es una técnica «supervisada», que utiliza información adicional para producir un modelo estadístico, mientras que el PCA es una técnica «no supervisada» y no requiere más introducciones de datos.

Puede recurrirse al PLS-DA y al PCA como métodos de reducción de dimensiones en modelos predictivos, antes del análisis discriminante lineal (PLS-DA, PCA-DA). El enfoque estadístico bilineal del análisis discriminante por mínimos cuadrados parciales (PLS-DA) es uno de los métodos supervisados más populares empleados en la metabolómica. En el PLS-DA, la matriz  $X$  contiene las variables de datos, mientras que la matriz  $Y$  contiene la variable de clase por la que se seleccionan valores para que sean los descriptores de la clase.

### 3.3.3. OPLS-DA

Evolución del PLS, se trata de un método de regresión lineal que ha sido empleado con éxito en simulaciones de predicción de distintas aplicaciones biológicas y bioquímicas. Una de las ventajas del método OPLS es su capacidad para modelar datos con variables con ruido y multicolineales, como por ejemplo datos metabólicos espectrales. En términos simples, el OPLS utiliza la información de la matriz  $Y$  para descomponer la matriz  $X$  en bloques de variación estructurada correlacionados y ortogonal respecto a  $Y$ . Al igual que el PLS-DA, se puede utilizar el OPLS para la discriminación (OPLS-DA). La ventaja principal del OPLS-DA es la posibilidad de separar la variación predictiva de la no predictiva (ortogonal). No obstante, el OPLS-DA, al contrario que el PLS-DA, no se puede emplear para analizar más de dos clases al mismo tiempo.

Cuando se construye un modelo (PCA, PLS-DA, etc.), existen dos parámetros de calidad del modelo ( $R^2$  y  $Q^2$ ) con importancia estadística y que se pueden ser considerados el primer paso de la validación.

$-R^2$  es el porcentaje de variación del conjunto de entrenamiento (training set) explicado por el modelo. Mide en qué medida el modelo se ajusta a los datos. Un buen modelo tiene un gran  $R^2$ , lo que se traduce en una buena reproducibilidad.

$-Q^2$ , que predice la variación, proporciona información sobre la capacidad del modelo de predecir nuevos datos. Un gran  $Q^2$  indica una buena predictibilidad.

Para asegurar que el modelo tenga una capacidad para diagnósticos, los índices  $R^2$  y  $Q^2$  tienen que ser altos y no podrán variar en más de 0,2 - 0,3. La desproporción entre estos dos valores es la primera advertencia de que el modelo se podría haber formado con un sobreajuste. La hipótesis de partida es que, para datos biológicos, los valores aceptables son:  $R^2 > 0,7$  y  $Q^2 > 0,4$  [35].

### 3.3.4. Gráfica S-plot, intervalo Jack-knife y Variable Importancia en la Proyección (VIP)

El OPLS-DA es el punto de partida para realizar la gráfica S-plot, los análisis Jack-Knife y la proyección VIP para la identificación de biomarcadores indicativos. Los intervalos de confianza de S-plot y Jack-Knife son herramientas útiles para seleccionar grupos diferenciados de metabolitos. De forma resumida, la S-plot es una representación visual, en forma de diagrama de dispersión, de la covarianza y correlación entre los metabolitos y la designación de clase que se muestran en una gráfica de puntuaciones. En la S-plot se pueden desplegar las aportaciones de las distintas fuentes a la distribución de una muestra de datos en una variable dada.

La técnica Jack-knife, que adopta la forma de un diagrama de columnas, se emplea para construir el intervalo de confianza para los modelos de parámetros estimados. El intervalo de confianza refleja la estabilidad y la incertidumbre de la variable y debería tener un tamaño reducido. A partir del análisis Jack-Knife se puede crear una lista indicativa de los metabolitos más relevantes de los grupos diferenciados.

La puntuación VIP muestra la contribución de cada variable predictiva al modelo. En SIMCA P+, los diagramas VIP se ordenan en función de la importancia de las variables; así, el valor VIP se utiliza con frecuencia para la selección de variables. Para construir un modelo preciso, normalmente se seleccionan los metabolitos con una puntuación VIP mayor que 1.

### 3.3.5. Validación de los modelos estadísticos multivariantes

Para evitar el riesgo de sobreajuste, toda vez que los resultados obtenidos tras el MVDA son sensibles a las correlaciones por azar, es necesario validar el modelo estadístico. Las herramientas de validación, como (a) el test de permutación y (b) la validación cruzada, se utilizan en la mayoría de los casos para proporcionar una evaluación objetiva del rendimiento y la estabilidad de un modelo.

(a) Test de Permutación: Para evaluar la importancia de una clasificación se recurre al test de permutación. La asignación de una clase se puede permutar varias veces y, por cada permutación, se puede construir un modelo entre los datos y la asignación de clase permutada. La discriminación entre las clases del modelo, basada en la asignación de la clase permutada, se compara con la discriminación del modelo basado en la clasificación original. El test de permutación resulta más adecuado para conjuntos de datos extensos que comprendan más de 50 muestras.

(b) Validación cruzada: Está aceptado como el método más adecuado para los conjuntos de datos medianos y pequeños. En la validación cruzada, el primer paso es dividir las muestras en varios grupos. Seguidamente, se retienen esos grupos de manera secuencial al mismo tiempo que se utilizan las muestras restantes para construir modelos de validación cruzada. Se predicen las muestras retenidas para cada modelo y se mide la capacidad predictiva. Se repite este proceso hasta que se retengan y predigan todas las muestras. Para cruzar los datos en rondas sucesivas utilizando el enfoque «excluir 1/3», se divide el conjunto de datos aleatoriamente en tres partes y se excluye 1/3 de las muestras para construir un modelo con los 2/3 restantes de las muestras. A continuación, se predicen las muestras excluidas mediante este nuevo modelo utilizando un análisis de la puntuación de la predicción T, y el procedimiento se repite hasta que hayan sido objeto de predicción todas las muestras por lo menos una vez. Si todos los datos se predicen correctamente, el modelo se puede considerar válido.

**Para más información:**

E. Saccenti, H.C. Hoefsloot, A.K. Smilde, J.A. Westerhuis, M.M. Hendriks. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics*, 2014, 10:361-374.

K. Varmuza, and P. Filzmoser. Introduction to multivariate statistical analysis in chemometrics. CRC press, 2009.

J. Godzien, M. Ciborowski, S. Angulo, and C. Barbas. From numbers to a biological sense: How the strategy chosen for metabolomics data treatment may affect final results. A practical example based on urine fingerprints obtained by LC MS. *Electrophoresis*, 2013, 34:2812-2826.

J. Trygg, E. Holmes, T. Lundstedt. Chemometrics in metabonomics. *J. Proteome Res.*, 2007, 6:469-479.

## 4. Análisis de datos

### 4.1. De la identificación de datos a las rutas metabolómicas: Evaluación de datos inteligente pero asistida para el análisis de la relevancia biológica.

La lista o listas de identificadores de compuestos (todavía no es una lista de compuestos) debe ser objeto de curación, donde se debe intentar siempre eliminar o consolidar las asignaciones. En el caso de la LC-MS o la CE-MS, entre los aspectos que se deben tener en cuenta están los aductos y las múltiples especies cargadas, los fragmentos (pérdidas neutras, productos de descomposición, recomposiciones), el perfil isotópico y los picos del ruido.

Asimismo, se debe realizar una revisión cuidadosa de las identificaciones para eliminar asignaciones imposibles. Se pueden establecer dos tipos de posibilidades para rechazar identificaciones: el conocimiento de la relación entre las propiedades del compuesto presuntamente identificado (polaridad, carga, aductos, etc.) y las condiciones experimentales (disolvente de extracción, fase móvil, tipo de cromatografía, condiciones de ionización, etc.); y el conocimiento del diseño del experimento (organismo, tratamiento, muestra, etc.).

Una vez curada la lista, estas identificaciones (y abundancias relativas, en algunos casos) se pueden pasar por un nuevo conjunto de herramientas bioinformáticas que permita evaluar

la importancia bioquímica de esos cambios. Estas herramientas son complementarias a la información bioquímica recogida en los libros de texto, combinada con nuevos enfoques para ese conocimiento, como el Roche Biochemical Pathways o el IUBMB-Sigma-Nicholson Metabolic Pathways Chart, mapas metabólicos completos a los que se puede acceder a través de Internet.

Asimismo, las herramientas bioinformáticas relacionadas con las rutas bioquímicas se pueden dividir en dos grupos: las bases de datos de rutas metabólicas, con características similares a las de las bases de datos de compuestos, pero con registros de rutas metabólicas completas en lugar de compuestos; y las herramientas de análisis de rutas metabólicas, capaces de aplicar algoritmos matemáticos para evaluar la importancia de los cambios observados, según el enriquecimiento de las señales y la cantidad y relevancia de los metabolitos que cambia.

#### 4.1.1. Bases de datos de rutas metabólicas

Una ruta bioquímica puede definirse perfectamente como una serie de reacciones que convierten un conjunto de sustratos en un conjunto de productos. Pero ésta es una definición subjetiva y no estándar, ya que la selección de metabolitos que se incluyen en una ruta metabólica particular es arbitraria, y las rutas metabólicas tienen distintos tamaños (número de metabolitos y reacciones que intervienen) y a menudo se solapan, ya que un metabolito puede formar parte de distintas rutas metabólicas. Por otra parte, las rutas metabólicas que se establecen en una base de datos pueden ser diferentes de las establecidas en otra.

A pesar de los problemas intrínsecos que existen de manera general en todas las bases de datos de rutas metabólicas, lo cierto es que constituyen una excelente fuente de datos biológicos que sirve para relacionar metabolitos con genes, proteínas y enfermedades, señalando así eventos y procesos, y ofreciendo también distintas herramientas que permiten la visualización y el mapeo de genes/metabolitos, habitualmente con múltiples especies.

Algunas de las bases de datos de rutas metabólicas más importantes son:

**KEGG** (Kyoto Encyclopedia of Genes and Genomes) es una base de datos que sirve para comprender funciones y utilidades de alto nivel del sistema biológico en estudio, como pueden ser la célula, el organismo y el ecosistema, ofreciendo información desde el nivel genómico y molecular. Es una representación informática del sistema biológico que se organiza en bloques que son componentes moleculares fundamentales como genes y proteínas (información genómica) y sustancias químicas (información química) que se integran junto con el conocimiento de los diagramas de conexión molecular de redes de interacción, reacción y relación (información de sistemas). Contiene además información sobre enfermedades y fármacos (información médica) como posibles perturbaciones del sistema biológico. Cuenta con 508 mapas de rutas metabólicas, 17.931 compuestos y 11.015 glicanos, además de 10.476 reacciones bioquímicas.

**Recon 2** cuenta con 1.789 genes implicados en la codificación de enzimas, 7.440 reacciones y 2.626 metabolitos únicos distribuidos en ocho compartimentos celulares. Se puede explorar el contenido de la reconstrucción mediante la búsqueda/exploración de metabolitos y reacciones.

La base de datos Small Molecule Pathway Database (SMPDB) se compone de aproximadamente 900 rutas metabólicas de pequeñas moléculas desarrolladas manualmente, de las cuales 384 se corresponden con rutas metabólicas de fármacos, 232 enfermedades y 220 rutas metabólicas generales. Representa compartimentos celulares, orgánulos, localización de proteínas y estructuras cuaternarias para convertir listas de genes, proteínas o sustancias químicas en rutas metabólicas o diagnósticos de enfermedades.

**PathWhiz** es un servidor web diseñado para permitir la creación de diagramas de rutas metabólicas a color y biológicamente precisos legibles por máquina e interactivos, con un visor al estilo de Google Maps.

### 4.1.2. Análisis de rutas metabólicas

Con los datos metabólicos, se pueden considerar dos tipos de análisis de rutas metabólicas: el enriquecimiento y el análisis topológico. El objetivo del enriquecimiento de metabolitos es comprobar si en los datos existen grupos biológicamente significativos de metabolitos que estén enriquecidos de forma considerable. Esta significancia está relacionada con las rutas metabólicas implicadas, la localización de esos metabolitos o la enfermedad que pueda existir. El análisis topológico (análisis de rutas metabólicas *sensu stricto*) se desarrolla para ampliar y mejorar el análisis del enriquecimiento del conjunto de metabolitos para rutas metabólicas mediante la consideración de estructuras de rutas metabólicas y sirviendo de apoyo para la visualización de rutas metabólicas.

### 4.1.3. Enriquecimiento de metabolitos

**iMet-Q** (intelligent Metabolomic Quantitation) es una herramienta automática con interfaces intuitivas para cuantificar metabolitos en datos obtenidos con un análisis completo de cromatografía líquida-espectrometría de masas (LC-MS). iMet-Q cuenta con un procedimiento de cuantificación completo para la eliminación del ruido, la detección y el alineamiento de picos. Asimismo, proporciona los estados de carga y las ratios de isótopos de los picos de metabolitos detectados con el fin de facilitar la identificación del metabolito.

**MBRole** (Metabolite Biological Role) realiza análisis de sobrerrepresentación (enriquecimiento) de anotaciones categóricas para un conjunto de compuestos de interés. Estas anotaciones categóricas se corresponden con la información biológica y química disponible en varias bases de datos y software. Aunque se centra en las anotaciones cualitativas, MBRole es lo suficientemente general y versátil para realizar análisis de enriquecimiento funcional en cualquier muestra metabolómica (incluidas anotaciones biológicas y químicas adicionales para metabolitos humanos), por lo que complementa el software existente para el campo en auge de la metabolómica.

**BiNChE** es una herramienta de análisis de enriquecimiento para moléculas pequeñas basada en la ontología de ChEBI. BiNChE muestra un gráfico interactivo que se puede exportar como imagen de alta resolución o en formatos de red. La herramienta ofrece análisis simples, ponderados y

de fragmentos basados bien en la ontología de roles de ChEBI o en la ontología estructural de ChEBI. BiNChE sirve de ayuda en la exploración de grandes conjuntos de pequeñas moléculas que se generan dentro de la metabolómica o de otros contextos de investigación de la biología de sistemas. Esta herramienta de código abierto proporciona un acceso web fácil y altamente interactivo a análisis de enriquecimiento con la herramienta ontológica ChEBI, y se ofrece también como biblioteca independiente.

**IMPALA** es una herramienta web para el análisis conjunto de rutas metabólicas de datos transcriptómicos o proteómicos y metabolómicos. IMPALA ejecuta análisis de sobrerrepresentación o enriquecimiento con listas de metabolitos y genes especificadas por el usuario utilizando más de 3.000 rutas metabólicas previamente anotadas procedentes de 11 bases de datos. De este modo, se pueden identificar rutas metabólicas que se pueden desregular a nivel transcripcional, a nivel metabólico, o a ambos niveles. Se combinan los indicios que existan de desregulación de la ruta metabólica, lo que permite la identificación de nuevas rutas metabólicas con una actividad modificada que no aparecerían destacadas de aplicarse el análisis a tan solo uno los niveles funcionales.

**MPEA** (Metabolite Pathway Enrichment Analysis) es una herramienta rápida para el análisis funcional y la interpretación biológica de datos de caracterización metabólica. MPEA sigue el concepto del análisis de enriquecimiento de conjunto de genes (GSEA, por sus siglas en inglés) y comprueba si los metabolitos implicados en alguna ruta metabólica predefinida se encuentran hacia el principio (o el final) de una lista de compuestos consultados que han sido clasificados. En concreto, MPEA ha sido diseñada para tratar con relaciones muchos a muchos (many-to-many) que se puedan dar entre los compuestos consultados y las anotaciones de metabolitos.

**MSEA** (Metabolite Set Enrichment Analysis) es una herramienta web que ayuda a identificar e interpretar patrones en los cambios de concentraciones de metabolitos en un contexto biológicamente significativo para estudios metabolómicos de humanos y mamíferos. Para el desarrollo de MSEA ha sido clave la creación de una biblioteca de aproximadamente 1.000 conjuntos de metabolitos predefinidos que comprenden diversas rutas metabólicas, estados de enfermedad, biofluidos y localizaciones de tejidos. MSEA también es compatible con conjuntos de metabolitos definidos por el usuario o personalizados para análisis más especializados. MSEA ofrece tres análisis de enriquecimiento distintos para estudios metabolómicos, entre los que se incluye el análisis de sobrerrepresentación (ORA), la elaboración de perfiles de muestras únicas (SSP) y el análisis de enriquecimiento cuantitativo (QEA). ORA solo precisa de los nombres de los compuestos, mientras que SSP y QEA necesitan los nombres de los compuestos y sus concentraciones. MSEA genera gráficas o tablas fáciles de entender integradas con enlaces a imágenes de rutas metabólicas relevantes y descriptores de enfermedades. Para estudios metabolómicos no realizados con mamíferos o estudios más especializados, MSEA permite a los usuarios proporcionar sus propios conjuntos de metabolitos para el análisis de enriquecimiento. El servidor MSEA también es compatible con la conversión entre nombres de metabolitos comunes, sinonimia e identificadores de las bases de datos más importantes.

#### 4.1.4. Análisis (topológico) de rutas metabólicas

La estructura de las rutas biológicas representa nuestro conocimiento de las relaciones complejas entre moléculas (activación, inhibición, reacción, etc.). Sin embargo, ni el análisis de sobrerrepresentación ni el análisis de enriquecimiento de rutas tienen en cuenta la estructura de la ruta a la hora de determinar qué rutas tienen más posibilidades de intervenir en las condiciones objeto de estudio. Resulta obvio que los cambios que se den en las posiciones clave de una red tendrán un mayor impacto en la ruta que otros cambios que se produzcan en posiciones marginales o relativamente aisladas. Existen dos mediciones consolidadas de la centralidad del nodo para estimar la importancia del nodo: la intermediación (betweenness) y la centralidad de grado (degree centrality). La primera se centra en la estructura de la ruta metabólica desde la parte relativa al nodo hasta la parte general, mientras que la segunda se centra en las conexiones locales inmediatas.

Existen muchas herramientas informáticas comerciales de análisis de rutas metabólicas, como por ejemplo Pathway Studio, MetaCore, Ingenuity Pathway Analysis, etc. En comparación con estas, el análisis de rutas de **MetaboAnalyst** es una herramienta web gratuita, diseñada para el análisis de datos metabolómicos. Utiliza las rutas metabólicas de alta calidad KEGG como fuente de conocimiento base. Integra dentro del análisis de rutas metabólicas diversos métodos consolidados (por ejemplo, el análisis univariante o el análisis de sobrerrepresentación), así como conceptos/algoritmos novedosos (GlobalTest, GlobalAncova, análisis de topología de rutas). Asimismo, MetPa cuenta con un sistema de visualización interactiva al estilo de Google Maps para ayudar a los usuarios a comprender los resultados de sus análisis.

#### 4.2. Validación de biomarcadores. Asignación de relevancia a los metabolitos

Una vez que se proponga un biomarcador (o un grupo de ellos) como conjunto de metabolitos relevantes que podrían clasificar las muestras según las condiciones objeto de estudio, será necesario configurar un experimento dirigido y cuantitativo, aplicado a un grupo de muestras más grande (pacientes, en un estudio clínico, por ejemplo), preferiblemente independientes de aquellas seleccionadas para la ejecución de la fase de descubrimiento. Los resultados de este experimento se analizan para evaluar la validez de dicho biomarcador y la metodología más utilizada para ello se basa en la evaluación de las curvas de la característica operativa del receptor (ROC, por sus siglas en inglés).

El análisis de la curva ROC es el método más objetivo y estadísticamente válido para la evaluación del rendimiento del biomarcador. Permite construir un modelo predictivo a partir de una o más variables, que se puede utilizar para clasificar nuevos sujetos en grupos específicos (por ejemplo, sanos/enfermos). Una curva ROC es una representación gráfica que muestra el rendimiento de un sistema clasificador binario a medida que varía su umbral de discriminación. La curva se crea representando gráficamente la tasa de verdaderos positivos (TPR) contraponiéndola a la tasa de falsos positivos (FPR) para varios valores de umbral. En el aprendizaje automatizado, la tasa de verdaderos positivos se conoce también como sensibilidad, memoria (recall) o probabilidad de

detección. La tasa de falsos positivos se conoce también como efecto colateral (fall-out) o probabilidad de falsa alarma y se puede calcular como  $(1 - \text{especificidad})$ .

El área bajo la curva (AUC, por sus siglas en inglés) oscila entre 0,5 y 1,0 y el objetivo del investigador será maximizar el AUC por debajo de la curva ROC, al mismo tiempo que minimiza el número de metabolitos utilizados en el panel de biomarcadores.

La herramienta principal para construir curvas ROC era Roccet, que a día de hoy se integra dentro del cuadro de herramientas de **MetaboAnalyst**

## 5. Sesiones prácticas sobre metabolómica

Las sesiones prácticas comenzarán con una visita introductoria al laboratorio y la identificación de las herramientas analíticas principales: homogeneizadores de tejidos, fuentes de ultrasonidos, ultracentrifugadoras, SpeedVac y corriente de nitrógeno para concentraciones, insertos y viales, tapones, septa, cápsulas etc., junto con el instrumental analítico asociado a la espectrometría de masas que se utiliza principalmente en la metabolómica. GC-MS-Q; LC-MS-QTOF; LC-MS-QqQ; CE-MS-TOF.

### 5.1. Metabolómica dirigida y no dirigida

Se leerá y analizará de manera interactiva con los alumnos en el laboratorio un protocolo para metabolómica basada en LC-MS con muestras de plasma. En una sesión tutorial se mostrarán todos los pasos dados en un entorno profesional, utilizando tan solo software de código abierto y gratuito. El estudio metabolómico constará de dos grupos: CASO y CONTROL, con seis muestras cada uno más cuatro muestras de control de calidad (QC, por sus siglas en inglés) que se analizarán mediante LC-MS con ionización positiva y negativa.

The screenshot shows the Workflow4Metabolomics 3.0 website. The header includes the 'Wm' logo and the text 'Workflow4metabolomics'. A navigation menu on the left lists various sections. The main content area is titled 'Workflow4Metabolomics 3.0' and contains a welcome message: 'Welcome to the collaborative portal dedicated to metabolomics data processing, analysis and innovation for the metabolomics community.' Below this, there is a note about the software being a pre-release and a request for feedback. The bottom section features a four-step onboarding process: STEP 1: Request an account, STEP 2: Connect to Galaxy, STEP 3: Upload your data, and STEP 4: Make reproducible science. Each step is accompanied by a small icon or logo.



## 5.2. Preprocesamiento con Workflow4Metabolomics (W4M)

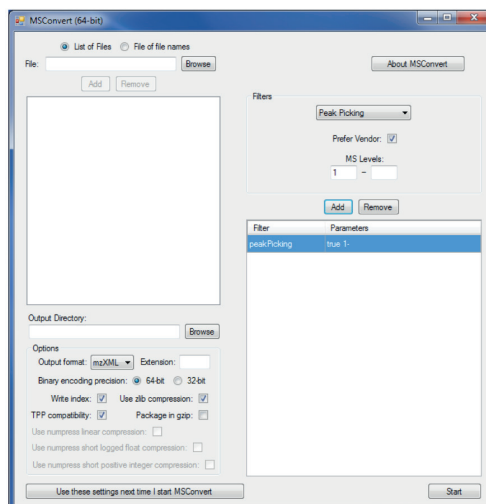
Este tutorial mostrará cómo preparar un conjunto de datos y utilizar la infraestructura en línea **Workflow4Metabolomics**, desarrollada en Galaxy. W4M es un proyecto colaborativo entre plataformas dedicadas a la metabolómica (la infraestructura francesa MetaboHUB) y a la bioinformática (IFB: Institut Français de Bioinformatique) y proporciona una plataforma de código abierto para la metabolómica informática.

Antes de comenzar, cada alumno deberá crear una cuenta en **Workflow4Metabolomics**. W4M: <http://workflow4metabolomics.org/>. Haz clic en «STEP 1: Request an account» (solicitar cuenta) y rellena el formulario. Se facilitarán las credenciales en el plazo de un día hábil.

### 5.2.1. Conversión de archivos MS

Dentro del entorno de W4M, los archivos MS tienen que convertirse a formato abierto como .mzML, .mzXML, .mzData y .cdf. Si el software comercial (normalmente suministrado con el equipo) no incluye una herramienta de exportación, se puede utilizar otro software de código abierto como ProteoWizard (<http://proteowizard.sourceforge.net/>) para la conversión de datos brutos.

Se recomienda utilizar solo letras (en minúscula o mayúscula), dígitos y caracteres “\_” (sin espacios) en todos los nombres proporcionados (nombres de muestras, de archivos, de carpetas, variables, etc.), así como convertir todos los archivos brutos a **mzXML** en modo centroide (archivos más pequeños), de manera que los archivos sean compatibles con el método xcms centWave.



Parámetros recomendados para la conversión de archivos MS para W4M

## 5.2.2. Inspección de datos brutos

Los investigadores usan normalmente el software proporcionado por el proveedor de la instrumentación de MS, pero también existen alternativas de código abierto, como **InsilicosViewer**, un visor gratuito para visualizar datos MS en mzXML que posee la mayor parte de las funciones que ofrecen sus homólogos comerciales. Esta herramienta incluye un software lector de mzXML/mzData desarrollado por el Institute of System Biology (<http://www.systemsbiology.org>)

Abre tus datos brutos en formato .mzXML.

1. Selecciona archivo > Abre archivo.
2. Ve a la carpeta que contenga los archivos de datos en formato mzXML.
3. Selecciona tus datos.
4. Haz clic en Abrir.

El archivo de datos aparecerá en la ventana de Datos Brutos (Raw Data). (ver Figura)

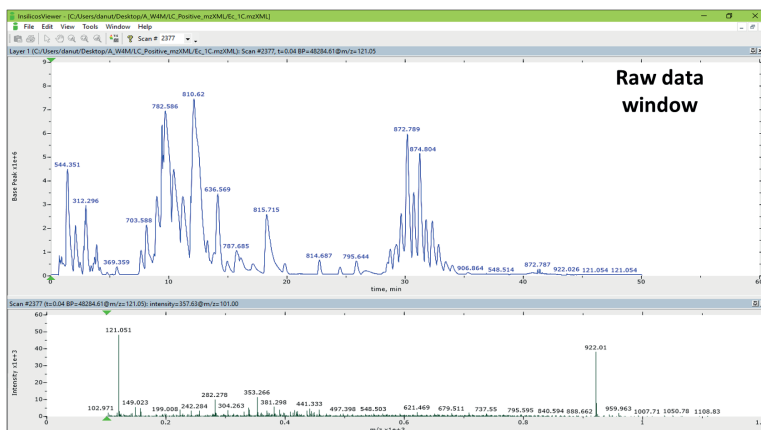
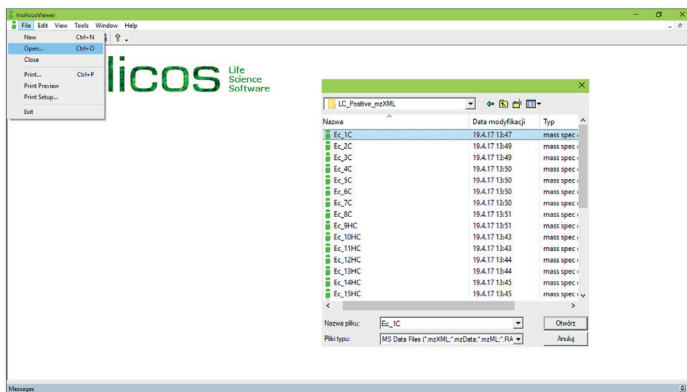
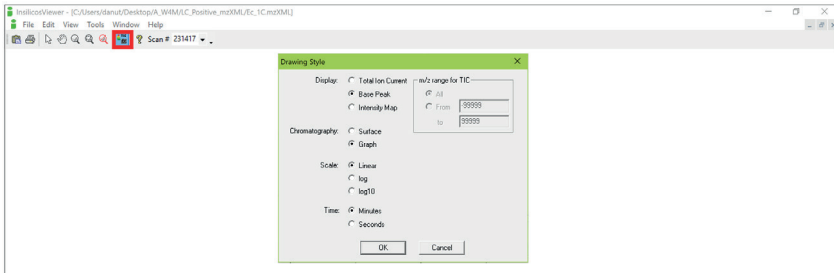
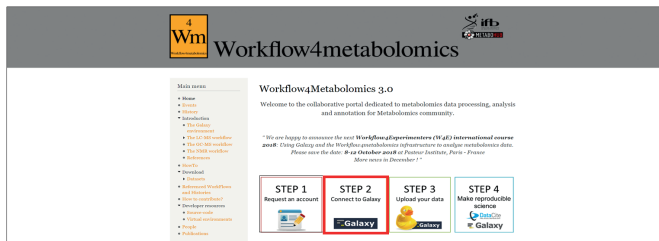


Figura. Ventana de datos brutos.

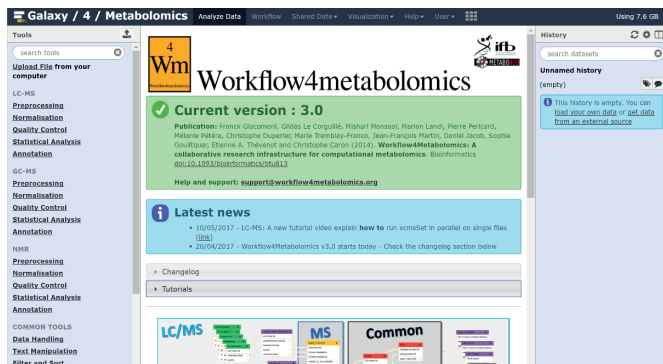
Para mostrar más estilos de gráficas, haz clic en el icono «TIC» o en «Tools > Drawing style»



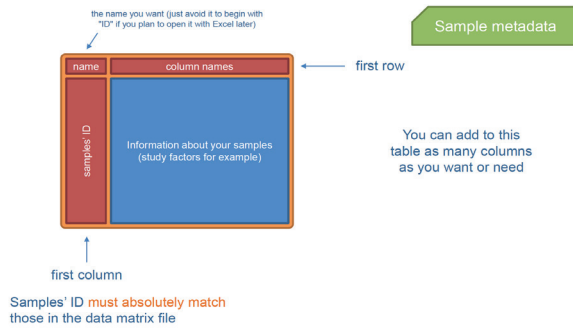
### 5.2.3. Subir datos a Galaxy



Aparecerá la ventana principal.



Los (archivos de) datos que se vayan a subir deben ir acompañados de la información del archivo de metadatos de las muestras (información sobre tus muestras) y se debe proporcionar esta información al sistema de acuerdo con unos criterios estrictos:



En el caso particular de las muestras que se vayan a utilizar en esta sesión práctica, la información sobre las muestras siguió la estructura que se describe en la tabla:

Nombre Muestra	clase	polaridad	Tipo Muestra	lote	Orden inyección	Dieta
QC	uno	positiva	pool	B1	1	NA
C1	uno	positiva	muestra	B1	7	C
HC3	uno	positiva	muestra	B1	10	HC
BL	uno	positiva	en blanco	B1	12	NA
...	...	...	...	...	...	...

Los archivos deben estar separados por tabulaciones: archivos TSV o TXT con una tabulación como separación.

Hay dos métodos para subir datos: subir un conjunto de datos individuales o crear antes un archivo .zip, con una estructura predefinida de subdirectorios. Nosotros pondremos en práctica tan solo la primera opción:

- Archivo único (recomendado): Puedes seleccionar un archivo único como entrada. De este modo, podrás ejecutar varios xcmsSet de forma paralela y utilizar el «xcms.xcmsSet Merger» antes del «xcms.group».
- Archivo zip: Puedes seleccionar un archivo zip que contenga tus archivos de entrada: myinputs.zip (donde figuren todas tus condiciones en forma de subdirectorios).

Archivo zip: Pasos para crear el archivo zip

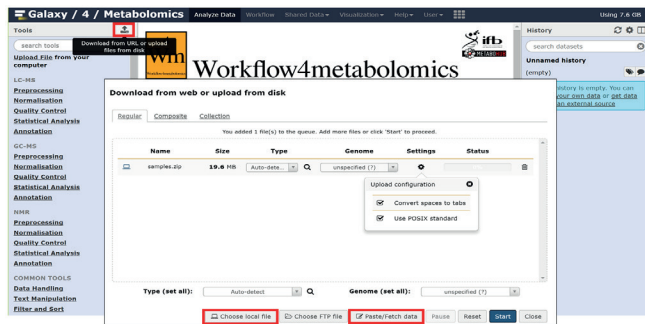
## Paso 1: Crear tu directorio y jerarquizar los subdirectorios

**MUY IMPORTANTE:** Si comprimes tus archivos en Windows, debes usar el software 7-Zip (<http://www.7-zip.org/>), de lo contrario, no podrás descomprimir correctamente tu archivo zip en la plataforma W4M (error de compresión).

## Paso 2: Crea tu archivo zip

### Importación de datos <2Gb

Para importar los datos, utiliza la opción «Paste/Fetch data» (pegar/recuperar datos) o «Choose local file» (escoger archivo local). Desde la opción «Upload configuration» (subir configuración) haz clic en «Convert spaces to tabs» (convertir espacios a etiquetas) a las etiquetas «Start» (iniciar)

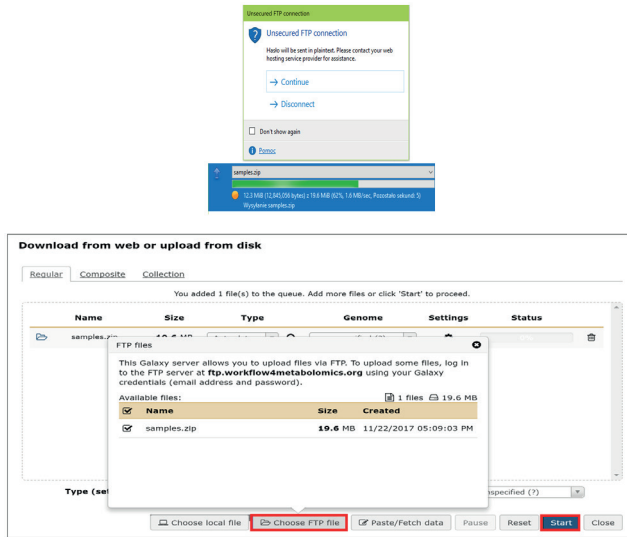


y «Close» (cerrar).



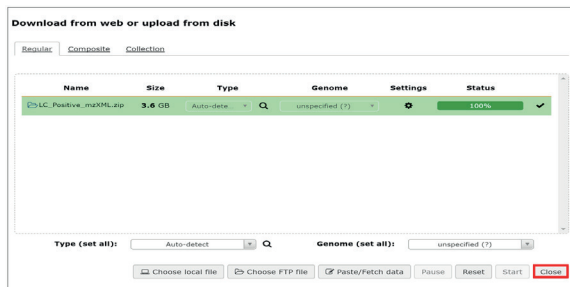
### Importación de datos <2Gb

**Paso 1.** Choose a FTP Client > Go to Cyberduck > Open Connection > Server (Elige un cliente FTP > Ir a Cyberduck > Abrir conexión > Servidor): <ftp.workflow4metabolomics.org>, e introduce tu nombre de usuario y contraseña; a continuación haz clic en «Open Connection» (abrir conexión) y copia y pega tus archivos

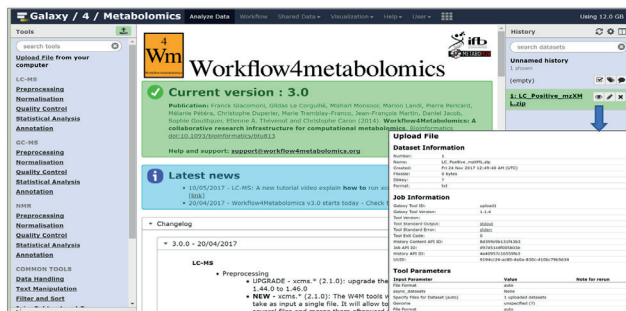


Cuando finalice la transferencia, vuelve a seleccionar «Choose FTP file > select your data > Start» (Elige un archivo FTP > selecciona tus datos > Inicio).

Tus datos se subirán a la plataforma WM.

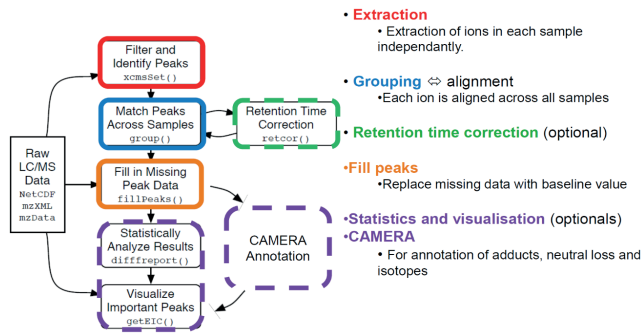


Ahora puedes ver los detalles de tus datos subidos



## 5.2.4. Preprocesamiento de datos de LC-MS a través de XCMS

XCMS es un paquete de software R especializado en la extracción de picos y alineamiento de tiempos de retención de archivos de adquisiciones de cromatografías de gases o líquidas acopladas a espectrometría de masas (GC-MS y LC-MS).



### Extracción de picos - xcmsSet

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

Esta herramienta se utiliza para el preprocesamiento de datos de archivos múltiples de LC/MS (en formatos NetCDF, mzXML y mzData). Extrae iones de cada muestra de manera independiente utilizando un modelo estadístico; los picos se filtran y se integran

### Archivos de entrada:

Parameter : num + label	Format
OR : Zip file	zip
OR : Single file	mzXML, mzML, mzData, netCDF

**xcms.xcmsSet Filtration and Peak Identification using xcmsSet function** 🔗 Versions ⌵ Options

from `xcms` R package to preprocess LC/MS data for relative quantification and statistical analysis  
(Galaxy Version 2.1.0)

File(s) from your history containing your chromatograms  

  
Single file mode for the format: mxmzml, mzml, mzdata and netcdf. Zip file mode for the format: no\_unzip.zip, zip. See the help section below.

Scan range option

Extraction method for peaks detection

[method] See the help section below

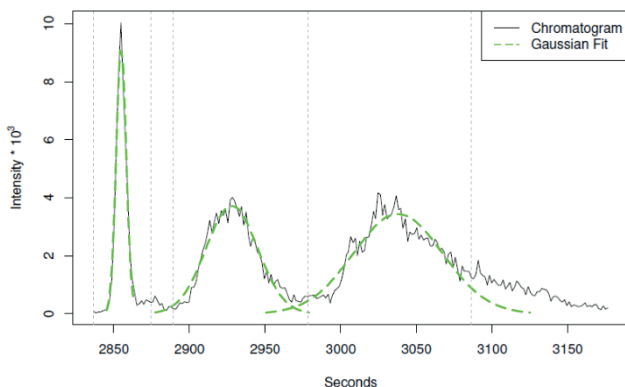
Max tolerated ppm m/z deviation in consecutive scans in ppm  
  
 [ppm]

Min,Max peak width in seconds  
  
 [peakwidth]

Advanced options

**Parámetros:****Método de extracción para la detección de picos:**

- **Matched Filter:** Matched Filter: especializado en datos MS de centroide o de un perfil de baja resolución.
- **Centwave:** especializado en datos de centroides de alta resolución. El algoritmo tiene como objetivo detectar «trazas de masa» o «regiones de interés (ROI)», que se definen como regiones con una desviación de  $m/z$  menor que la definida en escaneos consecutivos. Esta desviación tiene que ser inferior al valor del parámetro «ppm».



- **MSW:** Transformación continua de la ondícula (onda pequeña). Basado en la ondícula, se utiliza en datos de infusión directa y puede utilizarse para localizar picos cromatográficos en diferentes escalas.

Puedes comprobar el tipo de tus datos en tu archivo de datos brutos. .

```
<MS1CentroidDataAbsThreshold>200</MS1CentroidDataAbsThreshold>
<MS1CentroidDataRELThreshold>0.010</MS1CentroidDataRELThreshold>
<MS2CentroidDataAbsThreshold>5</MS2CentroidDataAbsThreshold>
<MS2CentroidDataRELThreshold>0.010</MS2CentroidDataRELThreshold>
<TimeSegment>
  <Index>1</Index>
  <StartTime>0.0</StartTime>
  <DiverterValveState>MS</DiverterValveState>
  <StorageMode>Centroid</StorageMode>
  <IonMode>Dual ES1</IonMode>
  <TimeSourceParameter>
    <Td>DGasHeater</Td>
    <Value>350</Value>
  </TimeSourceParameter>
  <TimeSourceParameter>
    <Td>DGasFlow</Td>
    <Value>10.0</Value>
  </TimeSourceParameter>
  <TimeSourceParameter>
    <Td>NebulizerPressure</Td>
    <Value>40</Value>
  </TimeSourceParameter>
</TimeSegment>
```

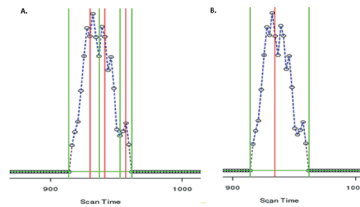
**Desviación ppm  $m/z$  máxima tolerada en escaneos consecutivos en ppm (en relación con la  $m/z$ )**  
Fluctuación del valor  $m/z$  (ppm) de escaneo a escaneo. Este parámetro tiene que ser fijado según la exactitud del espectrómetro de masas



**Ancho de pico mín. y máx. en segundos: (en relación con el RT)**

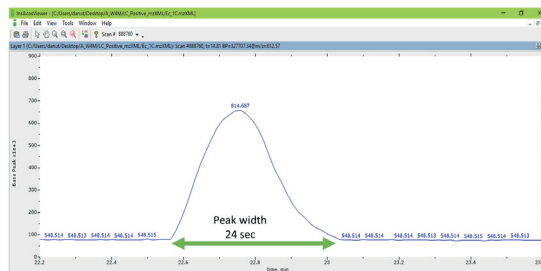
El propósito principal del parámetro que mide el ancho de pico es estimar de una manera aproximada el rango de anchura del pico; este parámetro no constituye un umbral. Las ondículas utilizadas para la detección de picos se calculan a partir de este parámetro.

Importante: No elegir un pico mínimo demasiado bajo, ya que no incrementará la sensibilidad, sino que dividirá los picos. .



**Figura.** Ejemplo: pico de 45s de ancho A. Al utilizar el ancho de pico (20,60), el pico se divide en tres picos, cada uno de ellos detectado como un pico separado de 10s de ancho: B. Al utilizar un ancho de pico de entre (20-120), el pico quedará intacto.

Puedes comprobar este parámetro en tus datos brutos. Calcula el ancho de pico para un pico estrecho y un pico ancho.



**Opciones avanzadas:**

**Advanced options**

show

**Signal/Noise threshold**

10

[snthresh] Signal to noise ratio cutoff

**Minimum difference in m/z for peaks with overlapping retention times**

-0.0015

[mzdiff] Can be negative to allow overlap

**peak limits method**

peak limits based on smoothed 2nd derivative (less precise)

[Integrate]

**Prefilter step for the first phase**

3,1000

[prefilter] Separate by coma k,l. Mass traces are only retained if they contain at least 'k' peaks with intensity >= 'l'

**Noise filter**

0

[noise] optional argument which is useful for data that was centroided without any intensity threshold, centroids with intensity smaller than 'noise' are omitted from ROI detection

## Diferencia mínima en $m/z$ para picos con tiempos de retención solapados (en relación con el $m/z$ y el RT):

Diferencia mínima de  $m/z$  para picos con un RT solapado (pico coeluyente). Tiene que ser negativa para permitir el solapamiento.

## Prefiltrado (en relación con la intensidad):

Un pico tiene que estar presente en  $n$  escaneos con una intensidad mayor que  $k$ .

Una vez fijados todos los parámetros, presiona «Execute» y espera... Este paso lleva tiempo.

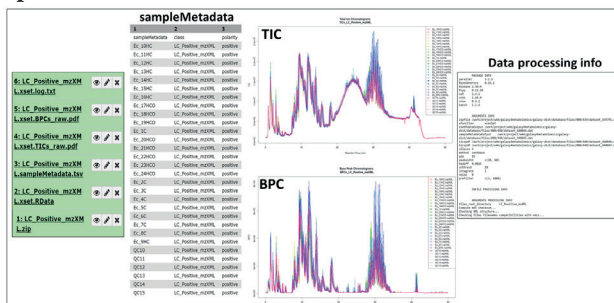
## Archivos de salida (Historial):

**xset.TICs\_raw.pdf** gráfica «Cromatogramas de Iones Totales» en formato PDF.

**xset.BPCs\_raw.pdf** gráfica «Cromatogramas de Pico Base» en formato PDF con cada muestra de clase confrontada.

**sampleMetadata.tsv** Archivo tabular que contiene información de cada muestra; está asociado con la clase y la polaridad (positiva, negativa y mixta). Este archivo es necesario para los pasos ANOVA y PCA del flujo de trabajo.

**xset.RData:** formato rdata.xcms.raw. Archivo rdata necesario para el segundo paso del flujo de trabajo «xcms.group».



El archivo de salida es un archivo **xset.RData**. Puedes continuar tu análisis en la herramienta **xcms.group**

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

Después de emparejar los picos en grupos, XCMS puede usar estos grupos para identificar y corregir desviaciones correlacionadas en el tiempo de retención de un análisis a otro

Seguidamente, se pueden utilizar los picos alineados para llevar a cabo un segundo pase de agrupamiento de picos que será más preciso que el primero.

Se puede repetir todo el proceso de forma reiterada. No todos los grupos de picos serán útiles para la identificación de variaciones en el tiempo de retención. Algunos grupos pueden ser picos perdidos que pertenecen a una gran parte de muestras, ofreciendo así una imagen incompleta de la variación en ese punto temporal. De cualquier modo, pueden existir otros grupos que contengan múltiples picos procedentes de la misma muestra, lo que es una señal de una agrupación incorrecta.

	pool1B1			pool1B2			pool1B3		
	mz	rt	int	mz	rt	int	mz	rt	int
Independent peak lists	196.0905	66.6	7810936	196.0910	66.7	11733921	196.0902	66.6	7933326
	158.1180	67.4	71736	342.0310	69.0	74594	158.1178	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
Group ions by m/z	196.0905	66.6	7810936	196.0910	66.7	11733921	196.0902	66.6	7933326
	158.1180	67.4	71736	342.0310	69.0	74594	158.1178	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
Group ions by RT	196.0905	66.6	7810936	196.0910	66.7	11733921	196.0902	66.6	7933326
	158.1180	67.4	71736				158.1178	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>Resulting matrix</b>									
	mz	rt	pool1B1	pool1B2	pool1B3				
	196.0905	66.6	7810936	11733921	7933326				
	158.1176	67.4	71736		82969				
	342.0308	21.3			2581				
	342.0309	66.3	202268	74594					
	267.0581	65.5	282039	260877					
	283.0319	65.2		424631	357448				

Tras la identificación de picos con xcmsSet, esta herramienta agrupa los picos que representan el mismo analito en todas las muestras a través de bins (regiones) de *m/z* superpuestos y del cálculo de las distribuciones de los picos redondeados en el tiempo cromatográfico. Permite rechazar características que se detectan solo parcialmente dentro de las réplicas de una clase de muestras.

**Archivos de entrada:**

<b>Parameter : num + label</b>	<b>Format</b>
Or : RData file	rdata.xcms.raw
Or : RData file	rdata.xcms.retcor

**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.xcims.raw, rdata.xcims.group, rdata.xcims.retcor or rdata dataset available. ▼  
 output file from another function xcims (xcimsSet, retcor etc.)

**Method to use for grouping**  
 density ▼  
 [method] See the help section below

**Bandwidth**  
  
 [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**  
  
 [minfrac] In at least one of the sample groups for it to be a valid group

**Width of overlapping m/z slices**  
  
 [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**  
  
 [max]

**Get a Peak List**

Parámetros

### Método para la agrupación

**mzClust.** Ejecuta un alineamiento de alta resolución en muestras espectrales únicas almacenadas en el archivo RData generadas por la herramienta xcimsSet.

**Densidad.** Agrupa picos de todas las muestras a través de bins  $m/z$  de solapamiento y del cálculo de las distribuciones de los picos redondeados en el tiempo cromatográfico.

**Nearest.** Agrupa picos en las muestras creando una lista de picos dominantes y asignando los picos correspondientes de todas las muestras. Se inspira en el algoritmo de alineamiento de mzMine.

### Ancho de banda [bw] (en relación con el RT)

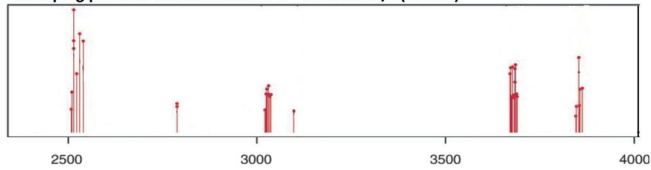
La desviación estándar del metapico gaussiano que agrupa picos.

### Ancho de los cortes de $m/z$ superpuestos [mzwid] (en relación con la $m/z$ )

Tamaño de los cortes (bins) de  $m/z$ . Rango de la  $m/z$  a incluir en un grupo. Depende de la exactitud del espectrómetro de masas.

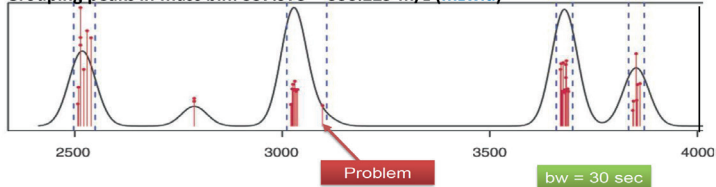
Se ejecuta un binning del dominio masa. **mzwid** define el tamaño del bin. A continuación, para cada bin de  $m/z$  se consideran todos los iones de todas las muestras para todos los tiempos de retención. Se utiliza el método de estimación de densidad de Kernel para detectar una región de tiempos de retención con una alta densidad de iones.

Grouping peaks in mass bin: 337.975 – 338.225 m/z (mzwid)

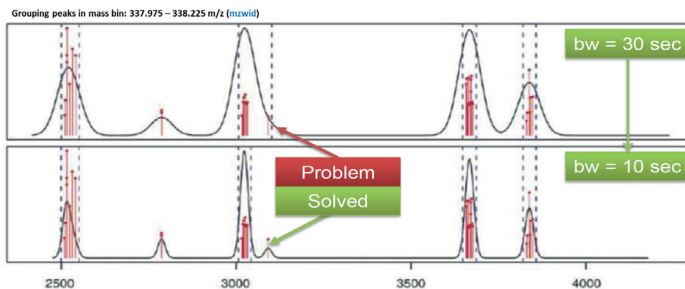


- Un modelo gaussiano agrupa los picos con RT similares.
- La inclusión de los iones en un grupo viene definida por la desviación estándar del modelo gaussiano (**ancho de banda**) que se corresponde con el parámetro **bw** (podría interpretarse como ventana de RT).
- Las líneas de rayas verticales indican que la característica es válida y se conservará en la matriz de datos.
- Para que sea válida, el número de picos de un grupo debe ser mayor que el porcentaje del número total de muestras. Este umbral está definido por el parámetro «minfrac» (fracción mínima).

Grouping peaks in mass bin: 337.975 – 338.225 m/z (mzwid)



- Disminuir el bw permite separar estos dos grupos.



- La  $m/z$  y el RT resultantes de la característica correspondiente a la mediana de la  $m/z$  y el RT de todos los iones, agrupada como una sola característica.

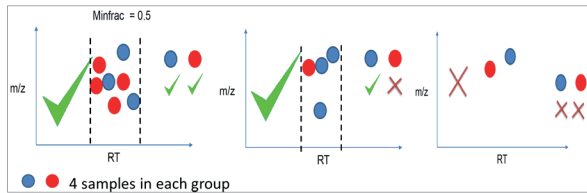
**Fración mínima de muestra necesaria [minfrac] (en relación con las muestras)**

Fración mínima del número total de muestras para que se pueda considerar válido un grupo (muestra mínima detectada en al menos una clase para que se pueda considerar un grupo).

Minfrac=0,5 se corresponde con el 50%. (n=10 minfrac=0,5 encontrado en al menos 10 de cada 50)

**Máx (en relación con el número de iones)**

Un número máximo de grupos detectados en cortes individuales de m/z.



Archivos de salida (Historial)

**xset.group.RData:** formato rdata.xcms.group

**xset.group.Rplots.pdf**

Fichero Rdata que será necesario para el tercer y cuarto paso del flujo de trabajo (xcms.retcor y xcms.fillpeaks).

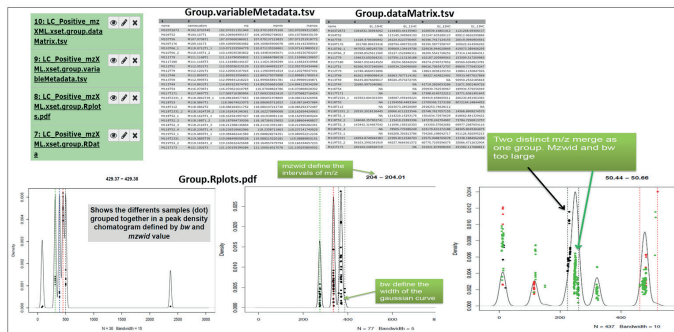
Galaxy 4 / Metabolomics

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

History

- 10: LC\_Positive\_mzXML.xset.group.dataMatrix.tsv
- 9: LC\_Positive\_mzXML.xset.group.variableMetadata.tsv
- 8: LC\_Positive\_mzXML.xset.group.Rplots.pdf
- 7: LC\_Positive\_mzXML.xset.group.RData
- 6: LC\_Positive\_mzXML.xset.RData
- 5: LC\_Positive\_mzXML.xset.RData
- 4: LC\_Positive\_mzXML.xset.RData
- 3: LC\_Positive\_mzXML.xset.RData
- 2: LC\_Positive\_mzXML.xset.RData
- 1: LC\_Positive\_mzXML.xset.RData



xcms.retcor (Alineamiento)

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

Después de emparejar los picos en grupos, XCMS puede usar estos grupos para identificar y corregir desviaciones correlacionadas en el tiempo de retención de un análisis a otro. Seguidamente, se pueden utilizar los picos alineados para llevar a cabo un segundo pase de agrupamiento de picos que será más preciso que el primero. Se puede repetir todo el proceso de forma reiterada. No todos los grupos de picos serán útiles para la identificación de variaciones en el tiempo de retención. Algunos grupos pueden ser picos perdidos que pertenecen a una gran parte de muestras, ofreciendo así una imagen incompleta de la variación en ese punto temporal. De cualquier modo, pueden existir otros grupos que contengan múltiples picos procedentes de la misma muestra, lo que es una señal de una agrupación incorrecta.

Archivos de entrada

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

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

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

**Method to use for retention time correction**  
 peakgroups  
 [method] See the help section below

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

**Number of extra peaks to allow in retention time correction correction groups**  
 1  
 [extra]

**Number of missing samples to allow in retention time correction groups**  
 1 **Number of admitted missing well behaved peak in a group.**  
 [missing]

**Advanced options**  
 hide

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

Execute

**Parámetros:**

**Método**

**peakgroups**

xcms ignora estos grupos al considerar solamente los grupos de picos con buen comportamiento (well-behaved) que no aparezcan (missing) en como máximo una muestra y que tengan como máximo un pico extra. (Estos valores se pueden cambiar con los comandos **missing** y **extra**.)

Para cada uno de esos grupos con buen comportamiento, el algoritmo calcula el RT medio y, para cada muestra, una desviación de esa media. Dentro de una muestra, la desviación observada varía en el tiempo de un modo no lineal. Esos cambios se aproximan mediante una técnica de regresión polinomial local aplicada a la función **loess**. Por defecto, el ajuste de la curva se realiza

a través de mínimos cuadrados en todos los puntos de datos. No obstante, se puede activar la detección y eliminación de valores atípicos ajustando el comando **family** a **symmetric**.

### obiwarp

Calcula para cada muestra las desviaciones del RT utilizando el código de obiwarp en «<http://obiwarp.sourceforge.net/>». Esta función puede alinear múltiples muestras con una *estrategia center-star*. Ordered Bijective Interpolated Warping (OBI-Warp) alinea matrices a lo largo de un eje único utilizando el algoritmo Dynamic Time Warping (DTW) y una función de warp interpolada y biyectiva. OBI-Warp saca provecho de la capacidad de alineamiento no lineal y global de DTW y añade el resultado discreto y no-biyectivo de DTW para ofrecer interpolantes naturales que puedan ser utilizados con múltiples conjuntos de datos.

**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

Archivos de salida (Historial)

**xset.group.retcor.TICs\_corrected.pdf** TIC gráfico de TIC en formato PDF, corregido después de un paso retcor.

**xset.group.retcor.BPCs\_corrected.pdf** gráfico de BPC en formato PDF, corregido después de un paso retcor.

**xset.group.retcor.RData: rdata.xcms.retcor** archivo de Rdata que será necesario para el paso xcms.group del flujo de trabajo.

El archivo de salida es un archivo **xset.retcor.RData**. Puedes continuar tu análisis utilizándolo en la herramienta **xcms.fillPeaks**.

**Tras un paso retcor, es obligatorio realizar un paso group. De no ser así, el resto del flujo de trabajo no funcionará con el archivo de RData. (El agrupamiento de picos inicial pierde su validez y queda descartado)**



**Galaxy / 4 / Metabolomics** Analyze Data Workflow Shared Data Visualization Help User 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.Merge** 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.fillGaps** 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:

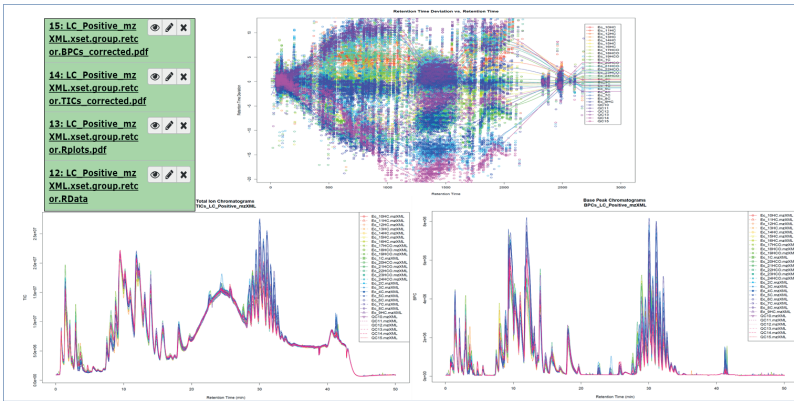
- 12: LC\_Positive\_mzXML.xset.group.retcor.RData
- 13: LC\_Positive\_mzXML.xset.group.retcor.Rplots.pdf
- 14: LC\_Positive\_mzXML.xset.group.retcor.TICs\_corrected.pdf
- 15: LC\_Positive\_mzXML.xset.group.retcor.BPCs\_corrected.pdf
- 16: xset.log.txt

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 3.71 GB

Unnamed history 14 shown, 1 hidden

- 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.dataMatrix.tsv
- 9: LC\_Positive\_mzXML.xset.group.variableMetadata.tsv
- 8: LC\_Positive\_mzXML.xset.group.Rplot.s.pdf
- 7: LC\_Positive\_mzXML.xset.group.RData



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

Tools search tools

**xcms.xcmsSet** function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis

**xcms.xcmsSet.Merge** 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.fillGaps** 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

CAMERA.combineAnnotations

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 3.73 GB

Unnamed history 18 shown, 2 hidden

- 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

**A group step**

## 2.xcms.fillpeaks

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

**Integrar áreas de picos perdidos.** Para cada muestra, identifica grupos de picos donde esa muestra no esté representada. Para cada uno de estos grupos de picos, integra la señal en la región de ese grupo de picos y crea un nuevo pico.

Dependiendo del tipo de datos brutos, existen dos métodos diferentes disponibles. Para cubrir datos de GCMS/LCMS, el método «chrom» integra datos brutos en el dominio cromatográfico, mientras que «MSW» se utiliza para listas de picos sin información de RT como los que provienen de espectros de infusión directa.

Archivos de entrada

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**

No rdata.xcms.group or rdata dataset available. ▼

output file from another xcms function (group)

**Filling method**

chrom ▼

[method] See the help section below

**Get a Peak List**

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

**Parámetros:**

**Método de cumplimentación**

- **chrom**

Este método produce valores de intensidad para las muestras perdidas integrando datos brutos en una región de un grupo de picos. En un grupo dado, se definen los puntos de RT de inicio y finales para la integración a través de los puntos de inicio y finales de la mediana de los otros picos detectados. Los valores  $m/z$  de inicio y finales se determinan de forma similar. Las intensidades pueden ser todavía igual a cero, lo que supone una intensidad bastante inusual para un pico. Este es el caso si, por ejemplo, se definiera el umbral de intensidad en los datos brutos y el área de integración no contuviera realmente intensidades brutas, o si una muestra fuera mal calibrada, esos puntos de datos brutos estarán entonces (apenas) fuera del área de integración.

Es importante que, cuando se disponga de la corrección de los datos de RT, la información de

alineamiento se utilice para integrar de forma más precisa la región apropiada de los datos brutos. Si el RT corregido se encuentra más allá del final de los datos brutos, el valor será un NaN (not-a-number /no-un-número).

- **MSW**

«MSW» se utiliza para listas de picos sin información de RT como las que provienen de espectros de infusión directa.

### Obtener una lista de picos

Cuando sea «verdadero» (true), el módulo generará dos archivos adicionales correspondientes a la lista de picos: el archivo de metadatos variable (correspondiente a la información de iones extraídos, como puede ser la masa o el RT), y la matriz de datos (correspondiente a las intensidades relacionadas).

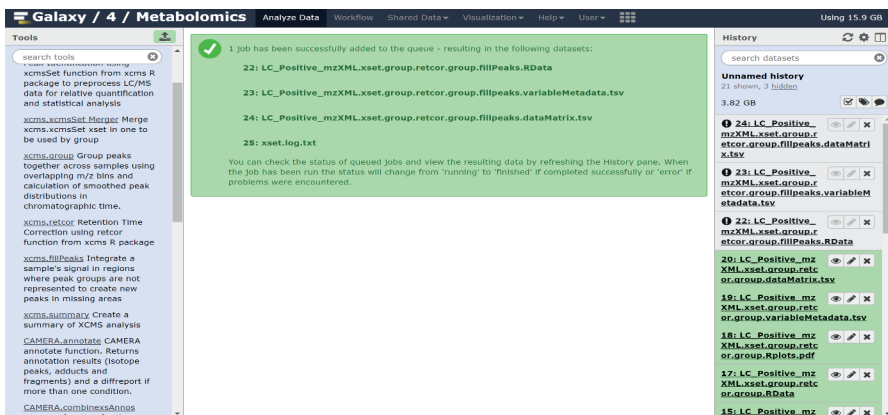
### Cifras decimales para valores [de masa o de tiempo de retención] en identificadores

Los identificadores de iones se construyen como MxxxTyyy, donde «xxx» es la masa media del ion y «yyy» es el RT medio del ion.

Se utilizan dos parámetros para ajustar el número de decimales deseado en identificadores para masa y para RT, respectivamente. Estos parámetros no afectan a decimales en columnas que no pertenezcan al identificador.

**Valores de intensidad registrados.** Este parámetro determina qué valores deberían ser registrados como intensidades en la tabla matrizDatos; se corresponde con el intervalo de xcms «intval».

- **into:** área integrada del pico (bruto) original
- **maxo:** intensidad máxima del pico (bruto) original
- **intb:** área integrada del pico corregida en la línea base (solo está disponible cuando se realice la detección del pico a través de «findPeaks.centWave»)



### Archivos de salida

**xset.fillPeaks.RData:** formato rdata.xcms.fillpeaks

**Archivo Rdata** que se usará en el paso **CAMERA.annotate** o **xcms.summary** del flujo de trabajo.

**xset.variableMetadata.tsv:** formato tabular. Tabla que contiene información sobre iones; se puede utilizar como una entrada de los módulos **Quality\_Metrics** o **\*\*Generic\_filter\***.

**xset.dataMatrix.tsv:** formato tabular. Tabla que contiene información sobre las intensidades de los iones; se puede utilizar como una entrada de los módulos **Quality\_Metrics** o **\*\*Generic\_filter\***.

El archivo de salida es un archivo **xset.fillPeaks.RData**. Puedes continuar tu análisis utilizándolo en la herramienta **CAMERA.annotate** o **xcms.summary**.

variableMetadata.tsv					dataMatrix.tsv				
name	formula	z	m/z	intensity	name	z	m/z	intensity	
24:LC_Positive_mz	ML1130719	10	103.0719	1000000000	ML1130719	10	103.0719	1000000000	
XML_xset_group.ret	ML1130719	10	103.0719	1000000000	ML1130719	10	103.0719	1000000000	
or.group.fillpeaks.dataMatrix.tsv	ML1130719	10	103.0719	1000000000	ML1130719	10	103.0719	1000000000	
Y	ML1130719	10	103.0719	1000000000	ML1130719	10	103.0719	1000000000	
23:LC_Positive_mz	ML1130719	10	103.0719	1000000000	ML1130719	10	103.0719	1000000000	
XML_xset_group.ret	ML1130719	10	103.0719	1000000000	ML1130719	10	103.0719	1000000000	
or.group.fillpeaks.variableMetadata.tsv	ML1130719	10	103.0719	1000000000	ML1130719	10	103.0719	1000000000	
22:LC_Positive_mz	ML1130719	10	103.0719	1000000000	ML1130719	10	103.0719	1000000000	
XML_xset_group.ret	ML1130719	10	103.0719	1000000000	ML1130719	10	103.0719	1000000000	
or.group.fillpeaks.RData	ML1130719	10	103.0719	1000000000	ML1130719	10	103.0719	1000000000	

24:LC\_Positive\_mz  
XML\_xset\_group.ret  
or.group.fillpeaks.dataMatrix.tsv  
Y

6,929 lines  
format: tabular, database: 2

PACKAGE INFO  
version 3.2.5  
BioCGenerics 0.16.1  
BiocS2 2.10.0  
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

Exported data matrix

TIN	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
MS07267	1030212	1030213	1030214	1030215	1030216	1030217	1030218	1030219	1030220	1030221	1030222	1030223	1030224	1030225	1030226	1030227	1030228	1030229	1030230

### 3. xcms.summary

xcms.summary Create a summary of XCMS analysis

Esta herramienta ofrece un resumen en HTML que sintetiza tu análisis a través de las herramientas [W4M] XCMS y CAMERA.

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

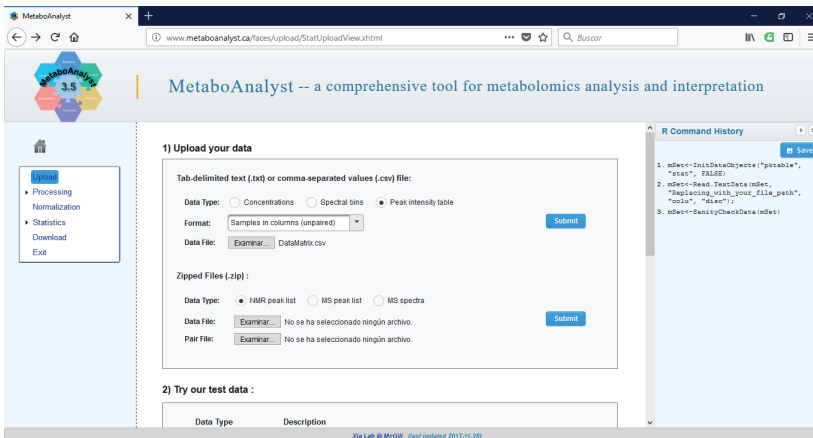
xset RData file

No rdata.xcms.raw, rdata.xcms.group, rdata.xcms.retcor, rdata.xcms.fillpeaks, rdata...

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

### 5.3. Estadísticas

Análisis estadístico univariante y multivariante: Se ejecutarán el PCA y el PLS-DA con Metaboanalyst (<https://metaboanalyst.ca>). El objetivo de MetaboAnalyst es ofrecer una herramienta gratuita, intuitiva y de fácil acceso para analizar datos procedentes de datos metabolómicos de alto rendimiento. Ha sido diseñada para solucionar dos tipos habituales de problemas: 1) identificar características que sean significativamente diferentes entre dos condiciones (descubrimiento de biomarcadores); 2) usar datos metabolómicos para predecir las condiciones objeto de estudio (clasificación). Por otra, MetaboAnalyst proporciona también herramientas para la identificación de compuestos y el mapeo de rutas metabólicas para la anotación de características significativas. Ve a la página principal y haz clic en >>click here to start<< (haz clic aquí para empezar).



Selecciona el tipo adecuado de Dato y Formato y envía («Submit») los datos. Como no existen valores perdidos en los datos facilitados, se puede saltar el siguiente paso.

En cada paso existen múltiples opciones para el análisis estadístico. Se recomienda a los alumnos que prueben alternativas a aquellas propuestas durante este curso.

Al tratarse de plasma, no se requiere normalización externa.

Trabajaremos después de la transformación logarítmica y el escalado Pareto.

Los alumnos intentarán identificar compuestos significativos tras un análisis ANOVA y post hoc unidireccional, y después también de un PLS-DA.

Se evaluarán opciones de aglomeración (clustering).

Los resultados se combinarán en una hoja de cálculo de MS Excel, de donde se obtendrá una lista de *m/z* significativas.

## 5.4. Anotación

Identificación a través CEU Mass Mediator, <http://ceumass.eps.uspceu.es/>

Análisis de rutas: integración del análisis de enriquecimiento y el análisis de topología de rutas.  
Visualización de los resultados obtenidos para el organismo modelo, con Metaboanalyst.

**Para saber más:**

E.G. Armitage, F. J. Rupérez, C. Barbas. Metabolomics of diet-related diseases using mass spectrometry. *Trends Anal. Chem.*, 2013, 52:61-73.

A. Mastrangelo, A. Ferrarini, F. Rey-Stolle, A. García, C. Barbas. From sample treatment to biomarker discovery: A tutorial for untargeted metabolomics based on GC-(EI)-Q-MS. *Anal. Chim. Acta*, 2015, 900:21-35.

A. García, S. Naz, C. Barbas. Metabolite fingerprinting by capillary electrophoresis-mass spectrometry. *Methods Mol Biol*, 2014, 1198:107-223.

S. Naz, A. García, M. Rusak, C.Barbas. Method development and validation for rat serum fingerprinting with CE-MS: application to ventilator-induced-lung-injury study. *Anal Bioanal Chem.*, 2013, 405(14):4849-4858.

W4M Workflow4Metabolomics. <http://workflow4metabolomics.org/>

## Módulo 3

### Lipidómica

Elisabete Maciel<sup>1</sup>, Eliana Alves<sup>2</sup>, Pedro Domingues<sup>2</sup>, Rosário Domingues<sup>1,2</sup>

*Mass Spectrometry Center, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal*

<sup>1</sup> CESAM (Centro de Estudos do Ambiente e do Mar)

<sup>2</sup> QOPNA (Química Orgânica, Produtos Naturais e Agroalimentares)

#### I. Justificación

La Lipidómica se puede definir como el análisis a gran escala de la composición lipídica de muestras biológicas. Desde la última década ha ido creciendo el interés por la lipidómica, debido al desarrollo de distintas técnicas analíticas, especialmente la espectrometría de masas. Por otra parte, han contribuido a este desarrollo, el reconocimiento del papel fundamental que juegan los lípidos en el mantenimiento de la homeostasia celular y el desarrollo de un gran número de patologías. El análisis lipidómico puede ofrecer perspectivas novedosas de los procesos bioquímicos, identificando nuevos biomarcadores lipídicos de la enfermedad y reuniendo pruebas de rutas metabólicas modificadas que puedan resultar útiles para estrategias terapéuticas innovadoras.

La primera parte de este módulo se centrará en las principales estrategias analíticas para identificar el perfil lipídico de muestras biológicas. Incluirá el estudio de las principales técnicas empleadas hasta la fecha, incluidos protocolos de extracción de lípidos, métodos de separación/fraccionamiento de lípidos, utilización de la espectrometría de masas en el análisis de lípidos y plataformas bioinformáticas para análisis de datos. La segunda parte de este módulo se ocupa de un estudio del caso y ha sido diseñado para que los alumnos se familiaricen con el análisis lipidómico. El curso se basa en clases expositivas e interactivas con los alumnos, sesiones demostrativas/prácticas y análisis de casos prácticos que permitirán consolidar los conceptos y alcanzar unos resultados de aprendizaje específicos.

#### II. Objetivos y resultados del curso

##### **Objetivos**

El objetivo de este curso es proporcionar a los alumnos una visión avanzada e integrada de las estrategias analíticas que se pueden emplear en el estudio del lipídoma. Los alumnos adquirirán conocimientos teóricos y prácticos sobre la importancia de la lipidómica en el contexto de las ciencias bioanalíticas. Abarcará conceptos básicos y avanzados de las características estructurales y la función de los lípidos. Se destacarán especialmente los conceptos prácticos relacionados con las metodologías lipidómicas, donde se incluirán distintas etapas de extracción, fraccionamiento y análisis estructural de lípidos, centrados en enfoques basados en la espectrometría de masas.

También se pretende contextualizar las potencialidades del análisis de las desviaciones en el lipidoma, asociadas a distintas patologías, como las patologías cardiovasculares, las enfermedades neurodegenerativas, el cáncer, la diabetes y la inflamación, para el desarrollo de nuevos biomarcadores o estrategias terapéuticas.

### **Resultados del aprendizaje:**

Tras terminar el curso, el alumno deberá ser capaz de:

1. Definir y aplicar terminología común en lipidómica;
2. Reconocer los principios de la mayoría de las técnicas lipidómicas;
3. Comprender los distintos flujos de trabajo de la lipidómica basada en la espectrometría de masas;
4. Analizar datos lipidómicos;
5. Comunicar y justificar conclusiones de forma clara e inequívoca ante especialistas y no especialistas;
6. Continuar el proceso de aprendizaje, mayormente, de forma autónoma.

## **III. Contenido del curso**

### **Módulo 3 - Lipidómica**

#### **1. Introducción a la Lipidómica**

##### **1.1. Enfoques analíticos en la Lipidómica**

1.1.1. Preparación de muestras: extracción y fraccionamiento de lípidos.

1.1.2. Identificación de lípidos a partir de MS: conceptos

1.1.2.1. Enfoques de la lipidómica no dirigida basada en LC-MS

1.1.2.2. Enfoques de la lipidómica dirigida de infusión directa

##### **1.2. Cuantificación**

1.3. Procesamiento de datos e identificación/cuantificación de lípidos

##### **1.4. Sesiones prácticas**

1.4.1. Identificación del perfil de fosfolípidos de monocitos THP1 mediante un enfoque lipidómico basado en HILIC-LC-MS.

1.5. Lectura de recursos adicionales.



## 1. Introducción a la Lipidómica

Se ha definido la Lipidómica como «la caracterización completa de las especies moleculares lipídicas y de sus funciones biológicas con respecto a la expresión y función de las proteínas implicadas en el metabolismo lipídico, e incluyendo igualmente la regulación génica» (en <http://lipidlibrary.aocs.org/Analysis/content.cfm?ItemNumber=39284>, abril de 2018).

Los lípidos son una clase de biomoléculas que incluye una enorme diversidad de especies moleculares distribuidas en distintas clases de lípidos, presentes en todos los seres vivos, y con distintas funciones. Esta variedad es todavía mayor cuando se consideran las modificaciones estructurales que se derivan de la peroxidación y glicación lipídicas. Los lípidos desempeñan diversas funciones, por ejemplo, como centros estructurales de las membranas celulares, donde resultan esenciales como moduladores de la actividad de las proteínas y como reconocimiento celular y moléculas de señalización. Los lípidos participan en la regulación de procesos celulares importantes, como la proliferación, la apoptosis, el metabolismo y la migración celulares. La caracterización completa del lipidoma representa un gran reto, no solo por la diversidad estructural, sino también debido a un gran número de especies moleculares lipídicas que existen a muy bajas concentraciones.

La biosíntesis y el metabolismo de los lípidos están regulados por un conjunto de enzimas que permiten mantener la diversidad estructural que existe en las clases lipídicas. La biosíntesis de los lípidos se produce principalmente en el retículo endoplasmático y en la mitocondria, y, tras ello, los lípidos se pueden translocar a otras membranas y orgánulos. El desequilibrio en el perfil lipídico de las células afecta a la red de señalización lipídica y puede estar ligado al inicio y desarrollo de varias enfermedades en seres humanos. Por tanto, la lipidómica, como herramienta emergente de gran potencial para la decodificación de la desviación en el lipidoma, ayuda a entender los procesos fisiopatológicos de la enfermedad, a definir nuevos biomarcadores de diagnóstico o a definir nuevas estrategias terapéuticas.

Para la identificación del perfil lipídico de una muestra biológica es necesario utilizar un conjunto de procedimientos de tratamiento de muestras y un conjunto de técnicas analíticas que es necesario planificar de acuerdo con los objetivos de la investigación. En general, el flujo de trabajo de la lipidómica se inicia con la extracción de todos los lípidos, seguida de fraccionamiento, análisis adecuado mediante espectrometría de masas y procesamiento de los datos (Figura 3.1).

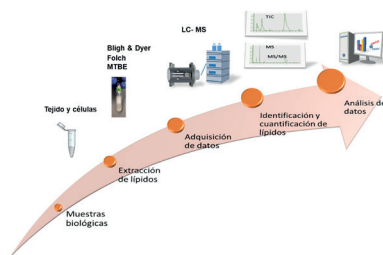


Figura 3.1. Flujo de trabajo básico en la lipidómica.

La primera etapa del flujo de trabajo en la lipidómica es la extracción de los lípidos totales en las muestras biológicas. Se pueden extraer diferentes tipos de muestras, concretamente, muestras celulares, tejidos y biofluidos. Dado que el extracto lipídico se compone de una gran variedad de lípidos, es conveniente, en algunos casos concretos, desarrollar una etapa de purificación, por ejemplo, separando el extracto total de lípidos en las distintas clases de lípidos. El fraccionamiento se puede llevar a cabo empleando métodos de separación como la cromatografía en capa fina (TLC) y la cromatografía líquida de alto rendimiento (HPLC), o mediante una extracción en fase sólida (SPE). La identificación de especies moleculares lipídicas en cada clase puede realizarse, por tanto, a través de la espectrometría de masas. La identificación de ácidos grasos esterificados en lípidos se puede realizar mediante una cromatografía de gases acoplada a la espectrometría de masas después de procedimientos de hidrólisis y derivatización. Debido al desarrollo de instrumentos y métodos de espectrometría de masas, la identificación y caracterización de especies moleculares lipídicas han experimentado mejoras enormes en sensibilidad y velocidad de análisis. Desde los años noventa, el desarrollo del método de la ionización por electrospray (ESI) ha permitido el acoplamiento de los espectrómetros de masas a la cromatografía líquida de alta eficacia (HPLC-MS), un avance esencial para el estudio de muestras complejas. La comercialización de espectrómetros de masas de alta resolución, como el Q-TOF y el Q-Orbitrap, posibilitó la adquisición de datos con alta sensibilidad. Los análisis basados en la LC-MS permiten identificar y cuantificar rápidamente miles de especies lipídicas, lo que resulta esencial para el análisis lipídico de muestras biológicas.

Dependiendo del objetivo principal del estudio, se pueden considerar dos estrategias generales distintas: análisis no dirigido y dirigido. La lipidómica no dirigida intenta abarcar un área amplia, identificando y cuantificando tantas especies lipídicas como sea posible. Asimismo, busca determinar una firma lipídica característica de un tejido, una célula, un orgánulo o unos biofluidos. Puede ser empleada además para pronosticar biomarcadores de enfermedades, comparando los datos de lipidómica de condiciones saludables con condiciones de enfermedad. Normalmente recurre a plataformas LC-MS para analizar los extractos lipídicos totales en un único análisis y generar una cantidad importante de datos que requieren de un análisis bioinformático y posterior validación. La información recogida de enfoques no dirigidos puede continuar utilizándose para diseñar estrategias de lipidómica dirigida. Los análisis de lipidómica dirigida suelen tener por objetivo la detección y cuantificación de un panel de lípidos específicos. Se puede emplear para identificar cambios concretos en una clase lipídica, y, para ello, emplea enfoques de espectrometría de masas dirigida, como el escaneo de iones precursores (PIS) o el escaneo de pérdida neutra (NLS). La cuantificación de especies moleculares lipídicas normalmente se realiza mediante monitorización múltiple de reacciones (MRM) en espectrómetros de masas QqQ.

## 1.1. Enfoques analíticos en la lipidómica

### 1.1.1. Preparación de muestras: extracción y fraccionamiento de lípidos

El flujo de trabajo en lipidómica se inicia con la obtención de un extracto lipídico total de la matriz biológica, como biofluidos, células, tejidos u otras matrices, empleando disolventes orgánicos (Figura 3.2). Se pueden utilizar varios métodos comunes para la extracción de lípidos, empleando disolventes orgánicos con distintas polaridades, que se pueden adaptar en función del tipo de muestras, como tejidos, extractos celulares o fluidos biológicos. Los métodos más populares de extracción de lípidos son el método de Folch (cloroformo/metanol (2:1, v/v)) y el método de Bligh y Dyer (cloroformo: metanol (1:2, v/v)). El primero es el que se utiliza con mayor frecuencia para extraer biofluidos (plasma, suero y orina), mientras que el segundo se emplea mayormente para extraer lípidos de tejidos y sedimentos celulares. En estos métodos, los lípidos se recuperan de la fase inferior, en la fase de cloroformo. Por otra parte, se han desarrollado otros métodos para evitar el uso de disolventes halogenados, como el cloroformo. Por ejemplo, el método MTBE utiliza éter metil-terbutílico (MTBE)/metanol/agua (10:3:2.5, v/v). Este método tiene además la ventaja de que la fase orgánica es la fase superior, facilitando la recuperación del extracto lipídico. Además, este método permite recuperaciones más rápidas y limpias y se puede adaptar para la elaboración de perfiles por infusión directa («shotgun»), utilizando MTBE/metanol (1:1). Para analizar un extracto lipídico rico en lípidos neutros como los triglicéridos (TG), se deben escoger disolventes menos polares, como el hexano.

La extracción en fase sólida (SPE) es otro tratamiento que se puede emplear también para separar fosfolípidos (PL) del plasma. En este procedimiento, se trata el plasma con acetonitrilo (ACN) acidificado para precipitar las proteínas y, a continuación, se separa la fase superior utilizando un cartucho SPE de fase iónica. Seguidamente se eluyen los fosfolípidos con acetonitrilo/hidróxido amónico (5%). Se emplean cartuchos de aminopropil para separar fracciones de colesterol, triglicéridos y fosfolípidos, en tanto que se recomienda el uso de cartuchos de gel de sílice SPE para separar lípidos neutros (ácidos grasos libres, colesterol y triglicéridos) y las fracciones de fosfolípidos.

Las clases de lípidos también se pueden obtener vía TLC. En varios estudios lipidómicos se han utilizado TLC uni o bidimensionales. El sistema disolvente se puede elegir en función de las clases de lípidos que se vayan a separar. Normalmente, las diferentes clases de fosfolípidos se fraccionan mediante una TLC unidimensional que emplea un sistema de disolventes compuesto de cloroformo/etanol/agua/trietilamina (35:30:7:35 v/v). Las manchas se detectan después de rociar con spray el plato de TLC con una disolución de primulina en acetona y de visualizarlos bajo una lámpara UV. Se identifica cada una de las clases de fosfolípidos comparando los factores de retención (Rf) con los estándares de los fosfolípidos aplicados en la misma placa. Se puede cuantificar el contenido de fósforo de cada mancha, lo que permite determinar el contenido relativo de cada clase de fosfolípido en el extracto lipídico total. Además, se pueden raspar las manchas de la placa y se pueden recuperar lípidos de cada mancha utilizando disolventes orgánicos apropiados

y continuar con su análisis. Por otra parte, también se puede utilizar la TLC bidimensional (2-D TLC) para separar lípidos. La combinación de sistemas disolventes para la 2-D TLC se elige en función de las clases de lípidos que se vayan a aislar. A pesar de que la 2-D TLC mejora en gran medida los resultados de la separación, presenta algunos inconvenientes cuando se compara con una TLC unidimensional, puesto que solo se puede aplicar una muestra en la placa y, por tanto, no es posible la aplicación simultánea de muestras y estándares. Además, la 2-D TLC consume más tiempo.

Los extractos o fracciones de lípidos totales obtenidos de la SPE o la TLC son posteriormente analizados a través de espectrometría de masas dirigida o no dirigida. Los tres enfoques más populares que utilizan la MS para el análisis de extractos/fracciones de lípidos son: la infusión directa, utilizando estrategias de tipo «shotgun» (por infusión directa), o las plataformas de LC-MS. A pesar de las distintas posibilidades, actualmente el enfoque más utilizado es el de los análisis de extractos lipídicos totales mediante LC-ESI-MS. En ambas plataformas es necesario poseer conocimientos en análisis e interpretación de datos de espectrometría de masas para identificar cada clase de lípido.

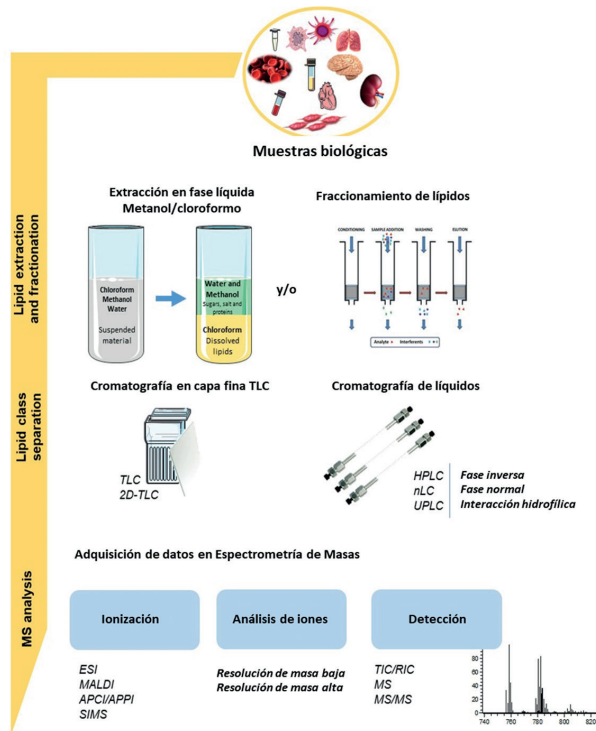


Figura 3.2. Preparación de muestras en la lipidómica: de la extracción y el fraccionamiento de lípidos a su análisis.

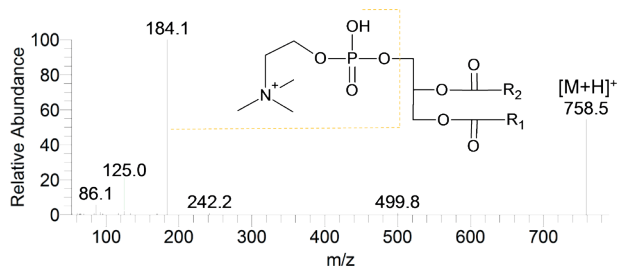
### 1.1.2. Identificación de lípidos a partir de MS: conceptos

El análisis de lípidos mediante espectrometría de masas necesita la información obtenida con la interpretación de los espectros de la MS y la MS/MS. Los fosfolípidos son el principal componente de los extractos lipídicos totales procedentes de las células y los tejidos. Teniendo en cuenta esto, los espectros de masas se adquieren en modos de ionización positiva y negativa, dependiendo de las características estructurales de las especies moleculares de fosfolípidos (PL). En el caso de los fosfolípidos: las fosfatidilcolinas (PC), fosfatidiletanolaminas (PE) y esfingomielinas (SM) normalmente se identifican en espectros MS adquiridos en modo de ionización positiva. En este modo, estos fosfolípidos muestran los correspondientes iones  $[M+H]^+$ , pero, en algunos casos, se pueden observar en baja abundancia iones  $[M+Na]^+$ . Los fosfatidilinositoles (PI), las fosfatidilserinas (PS), los fosfatidilgliceroles (PG) y las cardiolipinas (CL) se identifican preferentemente en espectros MS en modo de ionización negativa y se asignan como iones  $[M-H]^-$ . Las fosfatidiletanolaminas (PE) se pueden ver también en los datos de la MS en el modo negativo como  $[M-H]^-$ . En el caso de las CL, se observan también iones de doble carga ( $[M-2H]^{2-}$ ). En el caso de los enfoques basados en la LC-MS, en los datos de la MS en modo negativo, se puede ver el uso de eluyentes con acetato de amonio o formiato de amonio, fosfatidilcolinas (PC) o esfingomielinas (SM) como aductos con aniones de acetato o aniones de formiato,  $[M+CH_3COO]^-$  o  $[M+COO]^-$ , respectivamente (Tabla 1). También se puede ver el PI como iones  $[M+NH_4]^+$ . Gracias a este conocimiento se puede calcular el peso molecular de cada especie molecular de fosfolípido. En la actualidad, la adquisición de datos de MS de alta resolución en los espectrómetros de masas Orbitrap, que permiten adquirir datos de MS con una resolución mayor, de 70.000 a 500.000, proporciona información sobre los valores exactos de las masas de los iones con una exactitud de masa por debajo de 2 ppm.

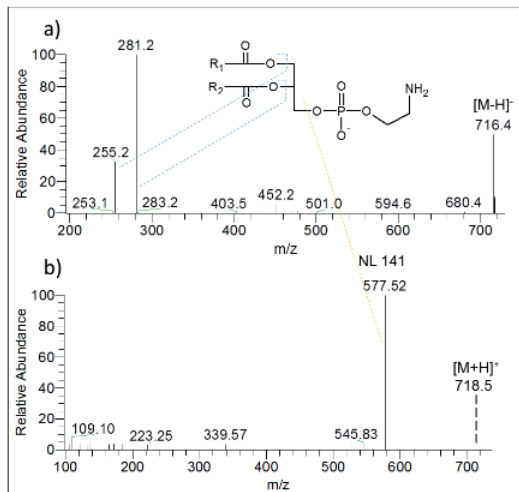
**Tabla 3.1.** Tipo de iones y aductos formados durante la ionización por electrospray y condiciones de escaneo para un análisis independiente de iones moleculares de lípidos polares. Los iones más abundantes formados en la ESI-MS aparecen en negrita.

Clase de lípido	Modo ion positivo	Modo ion negativo
Fosfatidilcolina (PC)	<b><math>[M+H]^+</math></b> , $[M+Na]^+$ ,	$[M+Ac-H]^-$
Fosfatidiletanolamina (PE)	$[M+H]^+$ , $[M+Na]^+$	$[M-H]^-$
Fosfatidilglicerol (PG)	$[M+NH_4]^+$ , $[M+Na]^+$	<b><math>[M-H]^-</math></b>
Fosfatidilinositol (PI)	$[M+NH_4]^+$	<b><math>[M-H]^-</math></b>
Fosfatidilserina (PS)	$[M+H]^+$	<b><math>[M-H]^-</math></b>
Cardiolipina (CL)		<b><math>[M-H]^-</math></b> , $[M-2H]^{2-}$

Los espectros de masas en tándem (MS/MS) de cada ion son necesarios para poder confirmar la identidad de cada molécula. En los espectros MS/MS es posible observar iones fragmento típicos de cada clase de fosfolípidos, lo que permite identificar el grupo de cabeza polar de los PL. Determinadas pérdidas neutras de los ácidos grasos como cetona o ácido (en modo positivo) o la formación de iones carboxilados,  $\text{RCOO}^-$ , en modo negativo, permiten la identificación de la composición de los acilos grasos de cada especie molecular de PL (Figuras 3.3 y 3.4).



**Figura 3.3.** Espectro MS/MS de  $[\text{M}+\text{H}]^+$  de PC, que muestra el producto típico de la clase de PC de  $m/z$  184.



**Figura 3.4.** Los espectros MS/MS de las fosfatidiletanolaminas (PE) en modo de ionización negativo (a), que muestran los aniones carboxilados típicos  $\text{RCOO}^-$  que permiten identificar las cadenas de acilos grasos en las moléculas de PE, y en modo ionizado positivo (b), que muestran la pérdida neutra típica de 141 Da, habitual en la clase PE, que, en este caso, conduce a la formación de un ion fragmento abundante a  $m/z$  577,5.

En la Tabla 3.2 se muestran los iones y las pérdidas neutras más comunes de cada clase de PL

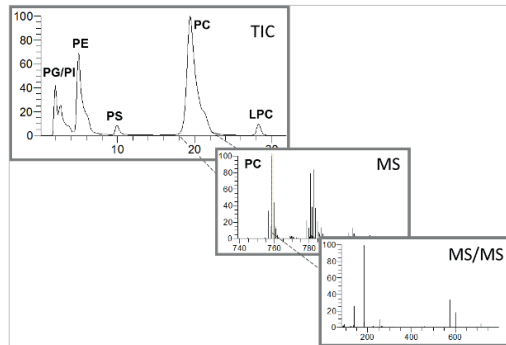
**Tabla 3.2.** Fragmentación típica de lípidos polares observada en los espectros MS/MS. PIS- escaneo de iones precursores, NLS- escaneo de pérdida neutra.

Clase de fosfolípido	Iones detectados en MS	Ion positivo	Ion negativo
Fosfatidilcolina (PC)	$[M+H]^+$	PIS $m/z$ 184	-
Fosfatidiletanolamina (PE)	$[M+H]^+$ $[M-H]^-$	NLS 141 Da	-
Fosfatidilserina (PS)	$[M+H]^+$ $[M-H]^-$	NLS 185 Da	NLS 87 Da
Fosfatidilglicerol (PG)	$[M-H]^-$	-	NLS 74 Da
Fosfatidilinositol (PI)	$[M-H]^-$	-	PIS $m/z$ 241
Cardiolipina (CL)	$[M-H]^-$ $[M-2H]^{2-}$		
Esfingomielina (SM)	$[M+H]^+$	PIS $m/z$ 184	-

### 1.1.2.1. Enfoques de la lipidómica no dirigida basada en LC-MS

En la lipidómica no dirigida, el análisis se realiza por LC-MS. Este enfoque tiene la ventaja de poder separar y caracterizar extractos de lípidos complejos. La resolución de lípidos en sus clases representativas y especies moleculares basadas en la espectrometría de masas combinada con cromatografía, ofrece la posibilidad de separar y concentrar distintas clases en función de sus propiedades fisicoquímicas. El uso de plataformas de LC-MS ha mejorado la resolución y sensibilidad del análisis lipidómico.

A través de las plataformas de LC-MS, la identificación de especies lipídicas se basa en el tiempo de retención (RT) y la interpretación de los datos de MS y MS/MS (Figura 3.5). La LC se realiza en fase inversa (RP), en fase normal (NP) o mediante cromatografía de interacción hidrofílica (HILIC). El RT depende de las condiciones experimentales, como el tipo de columna de LC, la velocidad de flujo, el tipo de eluyente, etc. La capacidad y selectividad de la columna son variables y dependen mayormente del fabricante de columnas, mientras que la eficiencia y resolución se pueden controlar, en cierta medida, a través de las condiciones de elución que se empleen.



**Figura 3.5.** Datos del análisis por LC-MS de un análisis de extractos lipídicos a través de la plataforma de HILIC-MS. Ejemplo de un Cromatograma de Iones Totales (TIC) de LC y un espectro MS de una clase PC, que muestra los iones  $[M+H]^+$  de todas las especies moleculares de esta clase de PL. El espectro MS/MS de un ion precursor seleccionado del espectro MS, muestra iones producto que permiten confirmar las características estructurales de la especie molecular.

En la Lipidómica, las columnas más utilizadas son las basadas en NP o HILIC, lo que permite la separación de las clases de lípidos de una sola vez. Se han utilizado para explorar marcadores potenciales de enfermedad o identificar desviaciones en rutas metabólicas asociadas a determinadas patologías. En estas columnas, las especies de lípidos eluyen en función de su hidrofiliidad, la cual depende principalmente de las propiedades de la cabeza polar. Las columnas de fase inversa C18 separan lípidos en función de sus propiedades hidrofóbicas, lo que depende principalmente del número de carbonos y el grado de saturación de los sustituyentes del acilo graso. Por tanto, los lípidos que contienen cadenas de acilos grasos más largas y saturadas eluyen más tarde que aquellos que contienen cadenas de acilos más cortas y poliinsaturadas. Los análisis basados en LC con C18 se emplean normalmente después de un fraccionamiento y aislamiento previos de la clase de interés, o adoptando enfoques de espectrometría de masas dirigida.

Los enfoques no dirigidos permiten la identificación y cuantificación de un gran número de especies lipídicas. Por tanto, el análisis de estos datos requiere la identificación y cuantificación rápidas de cientos de especies de lípidos, lo que entraña no poca dificultad, aunque se están desarrollando herramientas bioinformáticas para el procesamiento, la organización, el análisis y la visualización de datos.

### 1.1.2.2. Enfoques de la lipidómica dirigida de infusión directa

Como ya se dijo antes, los enfoques dirigidos utilizan patrones de fragmentación típicos o iones característicos de clases de PL para definir un ion determinado o precursor, o escaneos de pérdidas neutras. Por ejemplo, el escaneo de iones precursores del ion a  $m/z$  184 se utiliza para detectar especies de PC, y los escaneos de pérdidas neutras, por ejemplo, NL de 141 Da para detectar PE,



se utilizan normalmente en lipidómica por infusión directa. La selectividad de estos enfoques analíticos resulta muy útil y recurre a rutas metabólicas de fragmentación específicas del grupo de cabeza único de cada clase de fosfolípido (Tabla 1). Estos enfoques se pueden ejecutar bien por infusión directa en espectrómetros de masas ESI-QqQ o Q-trap, o bien utilizando plataformas de LC-MS. El enfoque de monitorización de reacciones múltiples (MRM), que normalmente se realiza con espectrómetros de triple cuadrupolo, es un análisis por MS dirigido que filtra pares de un ion precursor/ion fragmento específicos. Este modo MS/MS se utiliza generalmente para cuantificar compuestos, normalmente a través de una curva de calibración.

Una importante ventaja que presenta la lipidómica por infusión directa es que se pueden observar en un solo espectro de masas, sin separación, todas las especies moleculares de una clase lipídica. Este enfoque es muy rápido, permitiendo el análisis de cientos de muestras en un corto espacio de tiempo. El principal inconveniente de este enfoque es la presencia de fenómenos de supresión iónica, lo que puede reducir considerablemente la sensibilidad. Asimismo, no es posible, en algunos casos, definir NLS o PIS específicos o, en otros, no son tan específicos como para que no puedan causar interferencias (por ejemplo, PIS de 184 son habituales para la PC y la SM), Tabla 2.

**Para más información:**

- U. Loizides-Mangold. On the future of mass-spectrometry-based lipidomics. *FEBS J.*, 2013, 280(12):2817-29.
- X. Han. Lipidomics for studying metabolism. *Nature Rev. Endocrin.*, 2016, 12(11):668-679.
- L. Yang et al. Recent advances in lipidomics for disease research. *J. Separation Sci.*, 2016, 39:38-50.
- M. Holčapek, G. Liebisch, K. Ekroos. Lipidomic Analysis. *Anal. Chem.*, 2018, 2090:4249-4257.

## 1.2. Cuantificación

En lipidómica se pueden alcanzar distintos niveles de cuantificación. La cuantificación relativa de los PL como clases se puede llevar a cabo después de la separación de cada clase de PL mediante cromatografía en capa fina (TLC), de la identificación de cada clase de lípido cotejándola con los estándares de los PL, y de una cuantificación adicional del fósforo en cada mancha.

A nivel molecular, se puede abordar la cuantificación relativa o absoluta mediante enfoques de LC-MS o infusión directa («shotgun lipidomics»). La cuantificación relativa de especies moleculares se puede llevar a cabo para identificar el perfil de una clase específica así como las variaciones de los perfiles de la misma clase en distintas condiciones biológicas. Para ello, y después de la identificación de cada especie molecular de PL, es necesario calcular el área de picos de cada ion. Para determinar la abundancia relativa de cada especie de lípido dentro de la clase seleccionada, es necesario normalizar cada pico mediante la suma total de las áreas de todos los picos. Se pueden añadir a la muestra (células, tejidos o biofluidos) estándares internos antes de la extracción de lípidos o se pueden añadir a los extractos lipídicos totales, justo antes de realizar el análisis por LC-MS. En este caso, se pueden normalizar los picos utilizando el área de picos del estándar lipídico. En conjunto, todo ello permite estimar la abundancia relativa de cada especie

lipídica dentro de una clase de lípido específica.

Para conseguir la cuantificación absoluta de cada especie molecular mediante el análisis LC-MS, es necesario construir una curva de calibración para cada especie que queramos cuantificar. El enfoque de monitorización de reacciones múltiples (MRM), que normalmente se realiza con espectrómetros de triple cuadrupolo, es un análisis MS dirigido que filtra pares de un ion precursor/ion fragmento específicos. Este modo MS/MS se utiliza generalmente para la cuantificación de compuestos, normalmente a través de una curva de calibración.

La desventaja de los enfoques cualitativos y cuantitativos cuando se aplican los enfoques de NP o HILIC es que no pueden discriminar especies de PL de las mismas clases con el mismo peso molecular. Por ejemplo, PC(36:4), pueden ser PC(16:0/20:4), PC(18:1/ 18:1) o PC(18:0/ 18:2). En este caso, tan solo cabe deducir que se está realizando la cuantificación de PC(34:2). En este caso, se puede emplear un enfoque alternativo que emplee prepurificación y columnas C18.

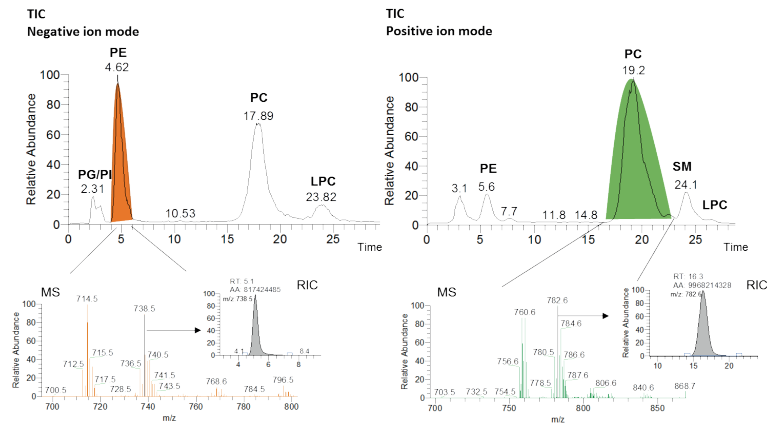
**Para más información:**

X. Han, K. Yang, R.W. Gross. Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses. *Mass Spectrom. Rev.*, 2012, 31:134-78.

### 1.3. Procesamiento de datos e identificación/cuantificación de lípidos.

La lipidómica basada en LC-MS de alto rendimiento genera una cantidad considerable de datos que requiere de herramientas bioinformáticas para poder llevar a cabo el análisis e integración de toda la información. Además, la lipidómica suele requerir bases de datos de extenso contenido para que sirvan de apoyo a la identificación de lípidos. Habida cuenta de que no existen herramientas bioinformáticas universales para el análisis automatizado, la interpretación y cuantificación de datos de LC-MS continúan siendo un problema.

El análisis de datos lipidómicos debería comenzar con la identificación de especies lipídicas que se puedan reunir a partir de la identificación basada en el RT y los valores  $m/z$  del ion molecular en los datos brutos de MS y en el análisis por MS/MS (Figura 3.6).



**Figura 3.6.** Cromatograma LC durante el análisis lipidómico del extracto total de queratinocitos. En esta Figura se destaca el pico del cromatograma del que eluyeron las clases lipídicas de PE y PC, así como los espectros MS correspondientes. Al seleccionar cada ion, es posible representar gráficamente el cromatograma de ion reconstruido (RIC, por sus siglas en inglés). Se puede utilizar el área de cada pico para la cuantificación.

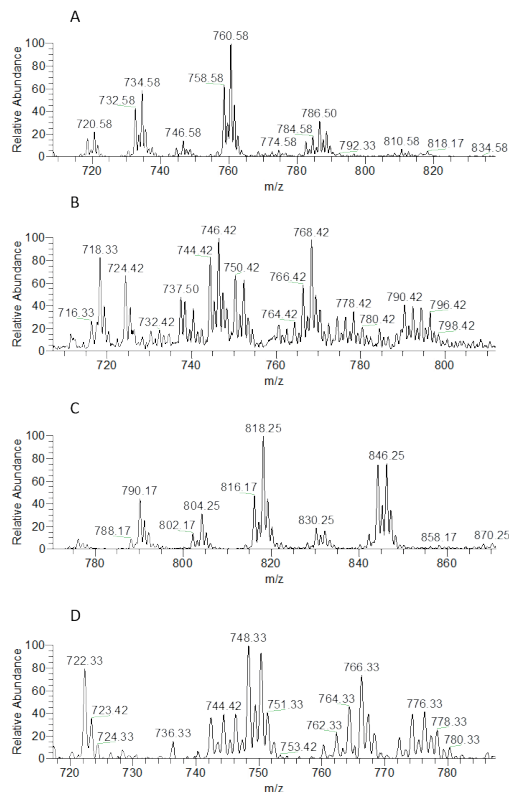
La identificación de lípidos normalmente se realiza a través de un análisis manual de datos, con la ayuda de consultas de bases de datos de masas exactas y fragmentación. Actualmente no existe ninguna clasificación de lípidos o base de datos de compuestos que tenga un carácter universal, aunque hay unas cuantas bases de datos lipídicas que pueden ayudar en esta tarea, como por ejemplo LipidBank (<http://lipidbank.jp/>), LIPIDATA (<http://lipidata.in/>) o LIPID MAPS (<http://www.lipidmaps.org/>). Por otra parte, casi todos los proveedores de espectrómetros de masas han desarrollado y comercializado su software para el almacenamiento y tratamiento de los datos lipidómicos. Los dos paquetes de software disponibles para el análisis de datos MS de lípidos, tanto comerciales como gratuitos, han sido desarrollados para determinados tipos de aplicación y modos de adquisición de datos. El software comercializado del que se dispone ahora es LipidView™ (de Sciex), que ha sido desarrollado para el escaneo de múltiples iones precursores y pérdidas neutras, LipidSearch™ (de Thermo Scientific) para datos lipidómicos basados en LC/MS, junto con los datos de masas exactas de alta resolución generados por los espectrómetros de masas basados en Orbitrap™. Existen también algunas herramientas y librerías de software libre y código abierto para el análisis de datos lipidómicos de MS. Las más utilizadas son LipidXplorer ([https://wiki.mpi-cbg.de/wiki/lipidx/index.php/Main\\_Page.](https://wiki.mpi-cbg.de/wiki/lipidx/index.php/Main_Page.)), Alex software ([http://mslipidomics.info/contents/?page\\_id=133](http://mslipidomics.info/contents/?page_id=133)), Lipid Blast (<http://fiehnlab.ucdavis.edu/projects/LipidBlast>) y MS-DIAL ([http://prime.psc.riken.jp/Metabolomics\\_Software/MS-DIAL/](http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/)). Permiten el análisis cualitativo y cuantitativo de datos de lípidos, adquiridos a través de distintos enfoques y distintos espectrómetros de masas. Tanto LipidXplorer como ALEX han sido diseñadas para lipidómica por infusión directa, empleando espectrómetros de masas de alta resolución. LipidBlast ha sido desarrollada para el análisis de lípidos a través de experimentos de MS/MS y búsquedas de librerías espectrales de masas, empleando instrumentos de baja o alta resolución. MS-DIAL se creó para tratar con experimentos de MS/MS dependientes e independientes de datos.

Tras la cuantificación, para continuar el procesamiento de datos, se normalizan los datos y se someten a análisis estadístico, utilizando normalmente los mismos enfoques que se describen en el capítulo sobre la metabolómica de este libro.

## 1.4. Sesiones prácticas

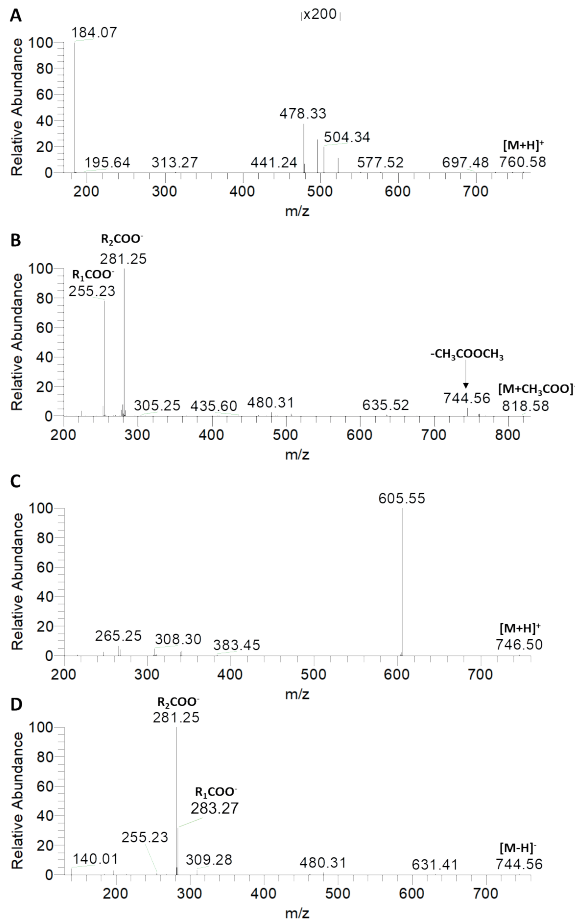
### 1.4.1. Identificación del perfil de fosfolípidos de monocitos THP1 mediante un enfoque lipidómico basado en HILIC-LC-MS

El extracto lipídico obtenido de líneas de células monocíticas humanas THP1 se analizó a través de LC-ESI-MS y LC-ESI-MS/MS en modo ion positivo y negativo. En la Figura 3.7 se muestran los espectros LC-ESI-MS obtenidos en modo de ionización positivo, en tiempos de retención (RT) de 22 min (A) y de 7 min (B), y los obtenidos en modo negativo en RT de 22 min (C) y en RT de 7 min (D).



**Figura 3.7.** Los espectros obtenidos después del análisis por LC-ESI-MS del extracto lipídico total en modo de ionización positivo con un tiempo de retención de 22 min (A) y 7 min (B) y los obtenidos en modo negativo con un tiempo de retención de 22 min (C) y 7 min (D).

Para obtener información sobre las clases de PL es necesario proceder con el estudio de la fragmentación de los iones moleculares seleccionados. En la Figura 3.8 se muestran los espectros ESI-MS/MS adquiridos para los iones observados a  $m/z$  760,58 (Fig. 3.8A), a  $m/z$  818,58 (Fig. 3.8B), a  $m/z$  746,5 (Fig. 3.8C) y a  $m/z$  744,56 (Fig. 3.8D)



**Figura 3.8.** Los espectros ESI-MS/MS de los iones observados a  $m/z$  760,58 (A),  $m/z$  818,58 (B),  $m/z$  746,50 (C) y  $m/z$  744,56 (D). Los espectros A y C fueron adquiridos en modo de ionización positivo, mientras que los espectros B y D fueron adquiridos en modo de ionización negativo.

## Preguntas

1. Basándonos en las rutas metabólicas de fragmentación típica de los fosfolípidos, ¿qué clases de fosfolípidos tienen el tiempo de retención de 7 min y 22 min en el cromatograma obtenido por LC-MS?
2. Considerando los espectros ESI-MS/MS obtenidos en 7 minutos tanto en modo negativo como positivo, proponga una estructura para esta especie de fosfolípido.
3. ¿Qué otras clases de fosfolípidos se podrán esperar del análisis LC-ESI-MS del extracto lipídico de THP1s?
4. Describa otro enfoque que permita evaluar los cambios lipídicos en las líneas de células THP1

## 1.5. Lectura de recursos adicionales

J.D. Martins, E.A. Maciel, A. Silva, I. Ferreira, P. Domingues, B.M. Neves, M.T. Cruz, M.R.M. Domingues. Phospholipidomic profile variation on THP-1 cells exposed to skin or respiratory sensitizers and respiratory irritant. *Journal of Cell Physiology*, 2016, 231(12):2639-51.

M. Pulfer, R.C. Murphy. Electrospray mass spectrometry of phospholipids. *Mass Spectrometry Reviews*, 2003, 22(5):332-64.

Y.H. Rustam, E. Gavin. Analytical Challenges and Recent Advances in Mass Spectrometry Based Lipidomics. *Anal. Chem.*, 2018, 90:374-397.

K. Yang, X. Han. Lipidomics: Techniques, Applications, and Outcomes Related to Biomedical Sciences. *Trends in Biomedical Sciences*, 2018, 41:954-969.

H.C. Lee, T Yokomizo. Applications of mass spectrometry-based targeted and nontargeted lipidomics. *Biochem Biophys Res Commun*, 2018, doi: 10.1016/j.bbrc.2018.03.081.

S. Sethi, E. Brietzke. Recent advances in lipidomics: Analytical and clinical perspectives. *Prostaglandins Other Lipid Mediat.*, 2017, 128-129:8-16.7.

Gracias a la Fundação para a Ciência e a Tecnologia (FCT, Portugal), la Unión Europea, el Quadro de Referência Estratégica Nacional (QREN), el Programa Operacional Potencial Humano (POPH), el Fundo Europeu de Desenvolvimento Regional (FEDER) y el Programa Operacional Factores de Competitividade (COMPETE) por financiar la unidad de investigación QOPNA (Química Orgânica de Produtos Naturais e Agroalimentares), y a RNEM (REDE/ 1504/REM/2005) por la Red Nacional de Espectrometría de Masas de Portugal.

## Module 4

### Proteomics

Tània Melo<sup>1</sup>, Rita Ferreira<sup>1</sup>, Rosário Domingues<sup>1,2</sup>, Pedro Domingues<sup>1</sup>,

*Centro de Espectrometria de Massa, Departamento de Química, Universidade de Aveiro, 3810-193 Aveiro, Portugal*

<sup>1</sup> QOPNA (*Química Orgânica de Produtos Naturais e Agroalimentares*)

<sup>2</sup> CESAM (*Centro de Estudos do Ambiente e do Mar*)

#### I. Justificación

Proteómica es el análisis a gran escala de las proteínas. Esta información es, a día de hoy, fundamental para la comprensión de los mecanismos moleculares que están detrás del papel que juegan las proteínas en las condiciones patofisiológicas, con el objetivo de mejorar el tratamiento, identificar los biomarcadores proteicos y señalar hechos, así como interacciones proteicas. La primera parte de este módulo se centra en las metodologías actuales que se emplean en el análisis e identificación de proteínas. Comprende el estudio de las principales tecnologías que intervienen, a saber, los métodos de separación de proteínas, los enfoques basados en la espectrometría de masas para el análisis proteico y el análisis de bases de datos de proteínas. La segunda parte de este módulo se ocupa de un estudio del caso y ha sido diseñado para que los alumnos se familiaricen con el software de análisis proteico de libre acceso. El curso se basa en clases expositivas e interactivas con los alumnos, sesiones demostrativas/prácticas y análisis de casos prácticos que permitirán consolidar resultados específicos.

#### II. Objetivos y resultados del curso

##### **Objetivos**

El objetivo de este curso es proporcionar a los alumnos una comprensión de los conceptos y enfoques fundamentales para el análisis del proteoma. Los alumnos adquirirán un conocimiento básico de la teoría y la práctica de la proteómica y el uso de estos métodos dentro de la investigación biomédica, incluidos los métodos de extracción, purificación, identificación y cuantificación de proteínas y el uso de herramientas bioinformáticas proteómicas.

##### **Resultados del aprendizaje:**

Tras terminar el curso, el alumno será capaz de:

1. Definir y aplicar terminología común de la proteómica
2. Reconocer los principios de la mayoría de las técnicas proteómicas.

3. Comprender los distintos flujos de trabajo de la proteómica basada en la espectrometría de masas.
4. Elegir entre los distintos enfoques proteómicos para resolver un problema concreto.
5. Utilizar herramientas bioinformáticas proteómicas.
6. Interpretar los datos que se deriven de la identificación de proteínas.
7. Comunicar y justificar conclusiones de forma clara e inequívoca ante especialistas y no especialistas.
8. Continuar el proceso de aprendizaje, mayormente, de forma autónoma.

### III. Contenido del curso

#### Module 4- Proteómica

1. Introducción a la Proteómica
2. Enfoques analíticos en la proteómica
  - 2.1. Preparación de muestras
  - 2.2. Identificación de proteínas basada en MS: conceptos
  - 2.3. Enfoques de la espectrometría de masas para la identificación de proteínas
  - 2.4. Proteómica cuantitativa
  - 2.5. Identificación de PTM
  - 2.6. Detección y cuantificación de la localización de proteínas subcelulares
  - 2.7. Detección y cuantificación de interacciones proteicas
3. Procesamiento de datos e identificación de proteínas.
  - 3.1. Bases de datos de secuencias de proteínas
  - 3.2. Motores de búsqueda de bases de datos de proteínas
  - 3.3. Análisis de datos
4. Sesiones prácticas
  - 4.1. Identificación de proteínas empleando los enfoques PMF y PFF con MASCOT
  - 4.2. Identificación de proteínas a partir de un experimento por LC-MS
    - 4.2.1. Transformación de datos a mzML con MSConverter en ProteoWizard )
    - 4.2.2. Generación de base de datos FASTA a partir de UniProt (SwissProt)
    - 4.2.3. Motores de búsqueda: SearchGUI
    - 4.2.4. Generación y evaluación de resultados: PeptideShaker para la visualización de péptidos y proteínas, y validación. Análisis PTM
    - 4.2.5. Análisis de datos: información proteica, análisis de rutas metabólicas y ontología de genes



## 1. Introducción a la Proteómica

En pocas palabras, la proteómica se define como el estudio a gran escala de las proteínas. La proteómica se nutre de una variedad de nuevas tecnologías recientes que ofrecen una determinación rápida y precisa del proteoma. Esta tarea es muy compleja por las limitaciones técnicas y biológicas. A nivel biológico, la existencia de un amplio rango de concentraciones proteicas en las muestras biológicas constituye una de las dificultades más importantes. El alto rango de concentración se manifiesta en varias dificultades técnicas, en concreto, la necesidad del fraccionamiento y enriquecimiento de las proteínas menos abundantes y la importancia de desarrollar métodos rápidos y sensibles para el procesamiento de grandes cantidades de muestras. Por otra parte, a nivel biológico, el proteoma, definido como las proteínas expresadas por un genoma, una célula, un tejido u organismo en un momento determinado, es altamente dinámico y puede cambiar con la edad, las condiciones de estrés, los nutrientes, etc. Asimismo, las modificaciones postraduccionales (PTM) contribuyen en gran medida a incrementar la complejidad del problema.

Existen diferentes enfoques proteómicos que proporcionan datos que permiten abordar cuestiones biológicas distintas. El análisis de la expresión proteómica busca determinar todas las proteínas que se expresen en una condición dada. La proteómica cuantitativa tiene por objeto medir la abundancia de proteínas expresadas, por ejemplo, comparando grupos experimentales y muestras de control. Otros enfoques incluyen la identificación de PTM y el estudio de interacciones proteína-proteína:

- Detección y cuantificación de niveles de proteínas
- Detección y cuantificación de modificaciones de proteínas
- Detección y cuantificación de la localización de proteínas subcelulares
- Detección y cuantificación de interacciones proteicas

Para cada una de estas cuestiones biológicas podemos utilizar distintos flujos de trabajo de proteómica basada en espectrometría de masas. No obstante, el flujo de trabajo básico de la proteómica incluye normalmente la extracción de proteínas del organismo o tejido objeto de estudio, la digestión enzimática de la proteína a través de proteasas y la purificación de los péptidos, el análisis de los péptidos mediante espectrometría de masas en tándem y el análisis de los datos con herramientas bioinformáticas (Figura 4.1).

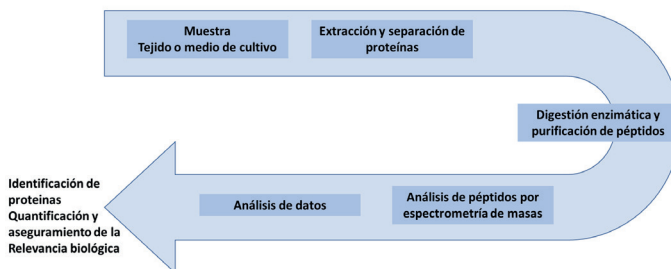


Figura 4.1. Flujo de trabajo básico en proteómica.

Los enfoques de la investigación proteómica pueden ser variados y se pueden clasificar, dependiendo de su objetivo, en proteómica de exploración (o descubrimiento) y proteómica dirigida. En la proteómica de exploración, se emplea un número limitado de muestras a analizar para optimizar la identificación de proteínas, si bien se pretende identificar tantas proteínas como sea posible. Las estrategias de proteómica dirigida limitan el número de proteínas que serán objeto de seguimiento para alcanzar la máxima sensibilidad y el máximo rendimiento. Con todo, y por regla general, cuanto mayor sea el número de proteínas a analizar, y cuanto mayor sea el rango dinámico de una muestra, más difícil será identificarlas y medir su presencia cuantitativamente.

## 2. Enfoques analíticos en la proteómica

### 2.1. Preparación de muestras

La preparación del análisis de muestras proteicas es esencial en los estudios proteómicos. La calidad de la muestra de proteínas y la reproducibilidad de los métodos empleados para la extracción y la separación de proteínas afectan considerablemente a la calidad de los datos proteómicos. La extracción y purificación de proteínas es una tarea muy compleja, al ser las proteínas un grupo muy heterogéneo de compuestos, con diferencias de tamaño, carga, solubilidad y concentración. El rango de concentraciones proteicas en una célula es de varios órdenes de magnitud, y más de ocho órdenes de magnitud en proteínas séricas. En estos momentos, no existe ningún método estándar de preparación de muestras para los estudios proteómicos, lo que hace que resulte muy difícil la comparación entre estudios. Sin embargo, aunque los protocolos pueden ser muy diferentes, todos ellos comparten una filosofía común, tal y como se muestra en la Figura 4.2. En la figura aparece el procedimiento general de preparación de muestras y se destacan algunas de las distintas opciones disponibles.

Al principio, el enfoque clásico de la proteómica solía ser la electroforesis bidimensional en gel (2-DE) y el análisis por espectrometría de masas (MS) MALDI-TOF. Más recientemente, los análisis incluyen 1D SDS-PAGE seguido de LC-ESI-MS/MS GeLC-MS/MS. Con todo, a causa de los muchos problemas asociados con las técnicas basadas en gel, como puede ser la precipitación de proteínas, y las dificultades asociadas con la resolución y la reproducibilidad, en los últimos años se han desarrollado distintos enfoques sin gel. Normalmente incluyen el uso de cromatografía líquida de alta resolución en varias etapas que permiten el fraccionamiento previo de las proteínas y presentan alta sensibilidad en su identificación.

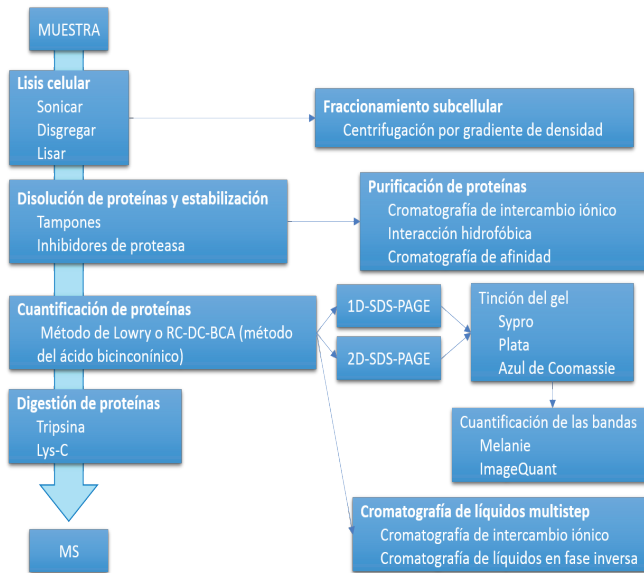


Figura 4.2. Procedimiento general de preparación de muestras, donde se destacan las distintas opciones disponibles.

**Para más información:**

J. Lovric. *Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis*. ISBN: 978-0-470-03524-5, Chapter 2.1, 2011.

## 2.2. Identificación de proteínas a partir de MS: conceptos

Existen varias estrategias para la identificación de proteínas a través de la espectrometría de masas. Estos métodos empleados para la identificación de proteínas se designan en función de los enfoques «ascendente» (bottom-up) y «descendente» (top-down) (Figura 4.3). El enfoque «ascendente» es el enfoque proteómico tradicional, y en este enfoque, las proteínas se digieren enzimáticamente, y los péptidos resultantes se analizan a través de la espectrometría de masas y la espectrometría de masas en tándem, y la proteína se identifica en función de la identidad de los péptidos.

Para la digestión enzimática de las proteínas, se pueden utilizar distintas enzimas o una combinación de enzimas, aunque la que se escoge con mayor frecuencia es la tripsina. La digestión de las proteínas separadas mediante un enfoque basado en gel normalmente se realiza directamente en el gel, después de cortar coágulos que contienen proteínas a identificar (digestión en gel) y, seguidamente, se extraen los péptidos. Cuando se utiliza un enfoque de separación basado en líquido, normalmente se digieren las proteínas mediante un protocolo de digestión endisolución, seguido de una purificación de los péptidos antes del análisis mediante MS.

El uso de la separación cromatográfica de los péptidos normalmente requiere incrementar el número de proteínas que se identifican en la muestra. Esto se debe a la existencia del «efecto de supresión iónica» cuando se emplean técnicas de ionización suave (MALDI y ESI), y consiste en la

supresión de la señal de los péptidos con menor afinidad protónica y que sean más hidrofílicos. Esta separación se suele realizar utilizando nanoLC-MS con columna de fase inversa C18, y normalmente los péptidos son eluidos con un gradiente binario de disolventes consistente en agua y acetonitrilo y un reactivo de par iónico (por ejemplo, ácido fórmico). En muestras complejas, se emplean gradientes más largos, o se puede combinar cromatografía de intercambio iónico con cromatografía en fase inversa, tanto en línea como fuera de línea, utilizando distintas columnas (2DLC) o en la misma columna (Tecnología Multidimensional de Identificación de Proteínas, MudPIT). Cuando se utiliza un enfoque proteómico que no ejecuta previamente ninguna separación de las proteínas antes de la digestión (enfoque de infusión directa), normalmente se hace necesario, para obtener mejores resultados, recurrir al uso de la separación 2D-LC de péptidos.

Después de la separación de los péptidos, estos se ionizan, empleando ESI o MALDI como fuentes de ionización, y se analizan mediante espectrometría de masas (MS) y espectrometría de masas en tándem (MS/MS). El efecto de discriminación en el análisis de los distintos péptidos en MALDI y ESI genera distintos porcentajes de cobertura de las proteínas cada uno de ellos. Normalmente, el ESI ofrece un porcentaje de cobertura ligeramente mejor, aunque MALDI es menos sensible a la contaminación y al efecto de supresión iónica.

En el enfoque descendente, se ioniza una proteína purificada, utilizando normalmente una fuente ESI y es fragmentada a través de los métodos de disociación por captura de electrones (ECD) o disociación por transferencia de electrones (ETD). A continuación, se analizan los iones fragmento utilizando analizadores de alta resolución como FT-ICR u orbitraps. Esta metodología suele requerir un mayor grado de purificación de las proteínas, normalmente a través de cromatografía líquida multidimensional. Además requiere una mayor cantidad de proteínas y el uso de instrumentación costosa.

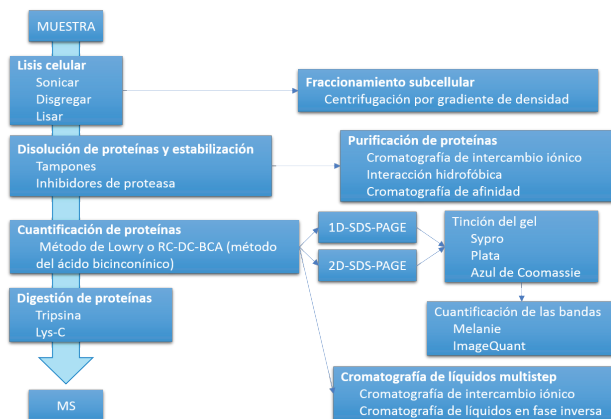


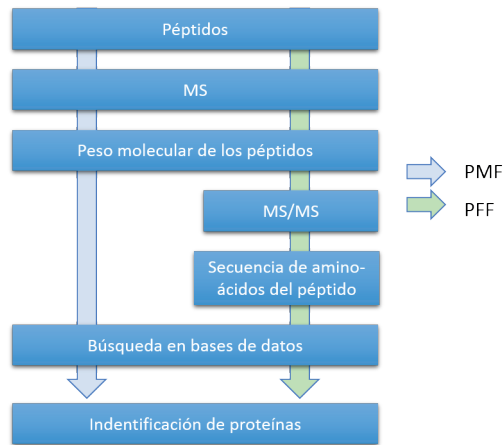
Figura 4.3. Enfoques de identificación de proteínas a partir de MS.

**To know more:**

Z. Szabo, T. Janaky. Challenges and developments in protein identification using mass spectrometry. Trends Anal. Chem., 2015, 69:76-87.

### 2.3. La espectrometría de masas en la identificación de proteínas

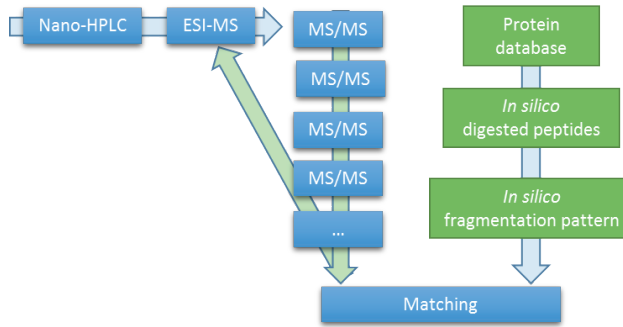
Se utilizan dos técnicas principales, la huella peptídica (peptide mass fingerprinting PMF son sus siglas en inglés) o la huella de fragmentación peptídica (peptide fragmentation fingerprinting PFF son sus siglas en inglés), para la identificación de proteínas mediante el enfoque «ascendente» (Figura 4.4). La técnica PMF identifica las proteínas comparando el peso molecular de los péptidos que se originan de la digestión enzimática con el peso molecular teórico de los péptidos generados in silico de bases de datos de proteínas o ADN. Los datos experimentales son una lista de valores de masas de péptidos procedentes de la digestión de una proteína a través de una enzima específica, como la tripsina, adquirida en un instrumento de alta resolución y exactitud. Se puede utilizar esta técnica en muestras simples de proteínas al permitir un análisis rápido, si se cumple que se observan varios péptidos tripticos (idealmente más de cinco). Se le asigna una puntuación según las proteínas identificadas mediante este método con base en distintos algoritmos. No obstante, normalmente estos se basan en la cobertura de secuencias de proteínas, en la observación de péptidos largos, que no suelen estar presentes en múltiples proteínas, y en la exactitud de masa de la observación.



**Figura 4.4.** Principales técnicas empleadas para la identificación de proteínas que utilizan el enfoque «ascendente»: la huella peptídica (PMF) o la huella de fragmentación peptídica (PFF).

La técnica PFF identifica las proteínas comparando el patrón de fragmentación de la MS/MS de los péptidos tripticos con el espectro de fragmentación teórico generado in silico de una base de datos de proteínas o ADN. Esta técnica está considerada «de referencia» para la identificación de proteínas. Identifica proteínas por inferencia y resulta muy útil para la identificación de proteínas en muestras muy complejas, puesto que se automatiza con facilidad para un alto rendimiento y puede obtener coincidencias de datos limitados, lo que al mismo tiempo, permite identificar muchas modificaciones variables. A la hora de emplear este enfoque, normalmente se diseña un experimento de LC-MS/MS para obtener la más alta sensibilidad y un gran número de datos

MS/MS. Para ello, se programa un experimento de MS/MS dependiente de datos, en el que el espectrómetro de masas adquiere de manera automática datos MS/MS de un número  $n$  ( $n=10-20$ ) de los péptidos más intensos observados en el barrido de MS previo (Figura 4.5).



**Figura 4.5.** Experimento de MS/MS dependiente de datos en enfoque proteómico de huella de fragmentación peptídica (PFF).

**Para más información :**

J. Lovric. *Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis*. ISBN: 978-0-470-03524-5, Chapters 4.2 and 4.3, 2011.

## 2.4. Proteómica cuantitativa

Aunque el objetivo de algunos experimentos proteómicos es identificar las proteínas que se encuentran presentes en una muestra, hoy en día la mayoría de los experimentos se refieren al estudio de los cambios que se producen a nivel de expresión, cuando los sujetos están en condiciones diferentes, por ejemplo, muestras de individuos sanos frente a muestras de individuos con enfermedad. Para estas aplicaciones, la cuantificación precisa de las proteínas es esencial y constituye un aspecto central de la proteómica.

Actualmente existen varios ensayos de espectrometría para la detección y medición de la cantidad de una proteína individual en una disolución. Sin embargo, estos métodos se basan en la purificación previa de la proteína de interés. Hay también varios métodos consolidados para la cuantificación de proteínas individuales, ya sea en una disolución o utilizando un ensayo de fase sólida, que se basa en el uso de anticuerpos marcados. Sin embargo, estos se basan en la calidad del anticuerpo y la fortaleza de la unión antígeno-anticuerpo y son métodos de bajo rendimiento. Con todo, estos anticuerpos marcados se pueden agrupar en microarrays analíticos de proteínas, lo que posibilita un análisis cuantitativo de alto rendimiento. Por desgracia, los microarrays que proporcionan mejores resultados tan solo contienen un pequeño número de anticuerpos bien caracterizados.

El análisis cuantitativo de alto rendimiento y a gran escala en la proteómica se realiza actualmente en dos enfoques distintos: el basado en la separación de proteínas de alta resolución mediante electroforesis bidimensional (2D) en gel o el basado en la abundancia relativa de los iones de péptidos en el espectro de masas (Figura 4.6). A pesar de que permite la cuantificación simultánea de miles de proteínas, el uso de electroforesis bidimensional en gel representa un obstáculo para el análisis a gran escala de proteínas debido a las dificultades de resolución, reproducibilidad y coincidencia de manchas. Estas dificultades se pueden de algún modo obviar utilizando electroforesis 2D diferencial en gel con tinción fluorescente (2-D DIGE). En la 2D DIGE, las proteínas de distintas muestras se marcan con marcajes fluorescentes (Cy2, Cy3 y Cy5) y, seguidamente, se combinan antes de la separación y cuantificación mediante electroforesis bidimensional en gel. Este enfoque cuenta con la ventaja de minimizar la variabilidad de los patrones de manchas y el número de geles de un experimento.

El análisis de péptidos mediante espectrometría de masas no es, básicamente, un análisis cuantitativo. Esto se debe a múltiples factores, entre los que encontramos las diferencias en las eficiencias de ionización de los péptidos, el efecto de supresión cuando se utilizan técnicas de ionización suave y el rango dinámico limitado de algunos analizadores. Por tanto, se han implementado distintos enfoques para cuantificar proteínas con espectrometría de masas. Estos se pueden clasificar en metodologías de cuantificación relativa y absoluta (Figura 4.6). Los enfoques de cuantificación relativa comparan la abundancia relativa en una muestra de péptidos individuales con los de muestras diferentes, mientras que los enfoques de cuantificación absoluta determinan la cantidad o la concentración de masas exactas de una proteína.

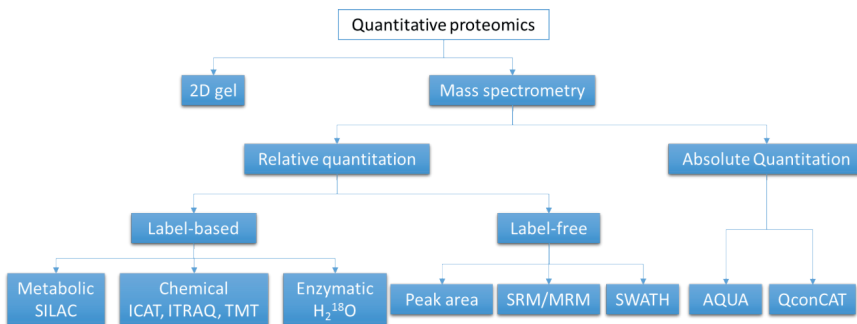


Figura 4.6. Métodos para análisis cuantitativos en la proteómica.

Los procedimientos de cuantificación relativa se basan en experimentos de LC-MS y LC-MS/MS y se pueden llevar a cabo utilizando enfoques con marcajes y sin marcajes. Los enfoques sin marcajes dependen del recuento espectral o de la medición de las intensidades de los picos de los péptidos. El enfoque del recuento espectral se basa en el número de péptidos identificados de una proteína dada para determinar su abundancia relativa, empleando distintos algoritmos. Estos algoritmos, por ejemplo, emPAI, calculan la abundancia prevista de las proteínas considerando la secuencia y longitud de la proteína. La cuantificación relativa mediante la medida de intensidades de picos de

péptidos se lleva a cabo comparando la intensidad relativa de los iones de péptidos de MS de una proteína en distintas muestras. Por lo general, las áreas de los picos de los distintos péptidos de un experimento de LC-MS, normalmente más de tres, se integran, y se comparan los valores de cada péptido.

Los enfoques con marcajes se basan en la introducción de diferentes marcajes de masa que alteran la masa de la proteína o el péptido. Después de esto, las muestras se combinan en una única muestra, y se realiza la cuantificación relativa comparando la abundancia relativa de cualquiera de los iones de los péptidos (de experimentos de MS) o los iones producto de los péptidos (de experimentos de MS/MS) de los iones con su marcaje diferencial. Estos marcajes se pueden introducir metabólicamente, química o enzimáticamente, a nivel de las proteínas o de los péptidos, durante la preparación de la muestra. El marcaje metabólico se refiere a un marcaje isotópico estable con aminoácidos en un cultivo celular (SILAC), mediante el cultivo de células u organismos en medios con un ( $2\text{H}$ ,  $15\text{N}$ ,  $13\text{C}$  y  $18\text{O}$ ) marcado. Estas muestras procedentes de diferentes condiciones se pueden juntar y analizar de manera simultánea, por lo que se evita los problemas de cuantificación más habituales.

En el marcaje químico, se introduce el marcaje isotópico en las proteínas o péptidos tripticos a través de una reacción química. Por ejemplo, cuando se utiliza marcaje de afinidad codificado por isótopos (ICAT), se recurre a un marcaje que se une a residuos de cisteína con reactivo «pesado» ( $\text{C}13$ ) o «ligero» ( $\text{C}12$ ). Se modifican las dos muestras antes del agrupamiento y seguidamente se combinan para la digestión, purificación de péptidos mediante cromatografía con avidina y análisis por espectrometría de masas. Los marcajes de masa isobáricos, como el marcaje de masas en tándem (tandem mass tag, TMT son sus siglas en inglés) y los marcajes isobáricos para la cuantificación relativa y absoluta (isobaric tags for relative and absolute quantification, ITRAQ son sus siglas en inglés) son marcajes de compensación de masas que se emplean para sintetizar los péptidos tripticos marcados que posean la misma masa y propiedades cromatográficas. A continuación se identifican los distintos marcajes de masa a través de una espectrometría de masas en tándem, lo que permite determinar cuantitativamente la abundancia relativa de proteínas de hasta ocho muestras.

El marcaje enzimático se basa en el uso de proteasas para incorporar un marcaje de masa  $^{18}\text{O}$  en el extremo carboxilo de los péptidos. Esta modificación en el grupo carboxilo del C-terminal de fragmentos proteolíticos conlleva la sustitución de dos átomos  $^{16}\text{O}$  por dos átomos  $^{18}\text{O}$  en el momento en que se produce la digestión en presencia de  $\text{H}_2^{18}\text{O}$ . Estos péptidos se pueden identificar entonces fácilmente en los espectros MS de muestras marcadas por la presencia de un desplazamiento de masa de 4Da del grupo carboxilo de los péptidos proteolíticos del grupo carboxilo del C-terminal  $\text{H}_2^{16}\text{O}$ .

La proteómica de cuantificación absoluta con espectrometría de masas normalmente se basa en el uso de péptidos estables como estándares internos, similares a los obtenidos con tripsina con marcaje isotópico. La cuantificación absoluta se consigue añadiendo a la muestra una cantidad conocida de estos estándares, antes del experimento de LC-MS. Estos péptidos, conocidos como péptidos de cuantificación absoluta (absolute quantification, AQUA son sus siglas en inglés), tienen el mismo tiempo de retención que los péptidos que proceden de la digestión triptica, y la cuantificación absoluta se consigue a través de una curva estándar que proporciona la cuantificación



absoluta del péptido objetivo. Existen enfoques alternativos que utilizan proteínas artificiales hechas de péptidos concatenados (enfoque concatémico (QconCAT)) o el uso de estándares proteicos para la cuantificación absoluta (protein standards for absolute quantification, PSAQ son sus siglas en inglés). En el primer enfoque, cada péptido es un estándar interno (sustituto) que representa una proteína específica, mientras que en el segundo enfoque se recurre a una proteína que está marcada isotópicamente y que se corresponde con el análogo recombinante expresado de la proteína a cuantificar. Sin embargo, estos métodos resultan costosos y limitados por su bajo rendimiento y su uso se restringe a un número pequeño de experimentos dirigidos.

**Para más información:**

J. Lovric. *Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis*. ISBN: 978-0-470-03524-5, Chapters 4.4 and 5.3, 2011.

## 2.5. Identificación de PTM

Las modificaciones postraduccionales (PTM) son modificaciones químicas de cadenas de proteínas después de la traducción, responsables del cambio de tamaño, composición, función y localización de las proteínas. Más de 900 PTM distintas han sido identificadas e incluidas en la base de datos UNIMOD ([www.unimod.org/](http://www.unimod.org/)). Las modificaciones postraduccionales más habituales (Tabla 4.1) incluyen la fosforilación, acetilación, oxidación, alquilación, metilación y la formación de puentes disulfuros. A día de hoy, la espectrometría de masas es el mejor enfoque analítico para la secuenciación de péptidos e identificación de PTM, puesto que posibilita el análisis de un gran número de muestras con una buena relación coste-eficacia.

**Tabla 4.1.** Relación de las modificaciones postraduccionales más comunes.

Tipo de PTM	Sustrato
N-glicosilación	Asparagina y lisina
O-glicosilación	Lisina, prolina, serina, treonina y tirosina
C-glicosilación	Triptófano
Fosforilación	Serina, treonina, tirosina, aspartato, histidina o cisteína
Acetilación	Extremo N-terminal de algunos residuos y ramificación de lisina o cisteína
Amidación	Generalmente en el extremo C-terminal de un péptido activo maduro después de la ruptura oxidativa de la última glicina
Hidroxilación	Generalmente la ramificación de asparagina, aspartato, prolina o lisina
Metilación	Generalmente en la fenilalanina N-terminal, la ramificación de lisina, arginina, histidina, asparagina o glutamato, y cisteína C-terminal

Las PTM normalmente se identifican a través de espectrometría de masas, mediante la observación de un desplazamiento de masa en los espectros de masas de los péptidos trípticos (Tabla 4.2). El desplazamiento de masa se deriva de una modificación química que se produce en la ramificación de residuos de aminoácido. Con todo, para la identificación inequívoca de la localización de las PTM, se hacen necesarios experimentos de espectrometría de masas en tándem (MS/MS). En estos datos MS/MS, el desplazamiento de masa detectado en el ion precursor (péptido obtenido por la digestión tríptica de la proteína modificada) se identifica en los iones fragmento que tienen el residuo de aminoácido modificado. Sin embargo, los péptidos modificados pueden tener otras rutas modificadas de fragmentación que dificulten la identificación del lugar de la modificación. Estas rutas de fragmentación son la pérdida de la modificación bien como molécula neutra o como residuo cargado.

La mayoría de las PTM son modificaciones de abundancia baja y son lábiles en análisis de MS/MS. Además, muchas de estas modificaciones son hidrofílicas, lo que podría hacer problemáticas la manipulación y purificación de muestras antes de la MS. Por otra parte, la presencia de modificaciones de proteínas puede afectar a la eficiencia de las rupturas de proteasas, generando productos de péptidos no previstos o de gran tamaño. Determinadas modificaciones proteicas reducirán la eficacia de la ionización y detección en la MS, mientras que la modificación de proteínas en múltiples lugares hace que la interpretación de los conjuntos de datos MS/MS resulte complicada y difícil. Por estas razones, suele ser útil considerar y explorar varios enfoques para el mapeo de PTM dentro de la proteómica. La purificación bioquímica de compartimentos celulares, orgánulos o complejos proteicos, o de proteínas individuales, es un enfoque muy útil para el análisis de modificaciones de proteínas, puesto que reduce la complejidad de la muestra proteica. Es también muy importante usar métodos de enriquecimiento de proteínas y péptidos modificados para reducir la complejidad de la muestra y, en consecuencia, incrementar el número de PTM identificados. De este modo, se han desarrollado diversos métodos para el enriquecimiento selectivo de proteínas modificadas para la identificación y cuantificación de PTM, entre los que se incluyen, por ejemplo, la inmunoprecipitación de proteínas o el enriquecimiento de PTM. El enriquecimiento de PTM se puede llevar a cabo a nivel de las proteínas, por ejemplo, utilizando anticuerpos, cromatografía de afinidad por iones metálicos (IMAC-fosfoproteínas) o lectinas inmovilizadas (glicoproteínas). El enriquecimiento de PTM se puede llevar a cabo a nivel de los péptidos, por ejemplo, utilizando anticuerpos ( $^{14}\text{C}$ ,  $^{3}\text{H}$ , pTyr), IMAC y  $\text{TiO}_2$  (fosfopéptidos) o lectinas inmovilizadas (glicopéptidos).

**Tabla 4.2.** Desplazamiento de masa observado en el análisis por MS y por MS/MS de péptidos trípticos modificados.

Tipo de PTM	$\Delta$ Masa Monoisotópica (Da)
Hidroxilación	15.9949
Fosforilación	
pTyr	79.9663
pSer,	79.9663
pThr	79.9663
Acetilación	42.0105
Metilación	14.0156
Sulfatación (sTyr)	79.9568
Deamidación	0.9840
Acilación	
Farnesil	204.1878
Miristoil	210.1983
Palmitoyl	238.2296
Glicosilación	
N-	>800
O-	203, >800
Formación de enlace disulfuro	-2.0157
Nitración de tirosina	44.9850

**Para más información:**

A.M.N. Silva, R. Vitorino, M.R.M. Domingues, C.M. Spickett, P. Domingues. Post-translational Modifications and Mass Spectrometry Detection. *Free Radic. Biol. Med.*, 2013, 65:925-941.

J. Lovric. *Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis*. ISBN: 978-0-470-03524-5, Chapter 5.4, 2011.

## 2.6. Detección y cuantificación de la localización de proteínas subcelulares

El campo de la proteómica de orgánulos ha ido emergiendo con el objetivo de revelar las funciones de las proteínas de cada orgánulo. Sin embargo, la elucidación de la distribución subcelular de las proteínas bajo condiciones diferentes plantea una dificultad importante en la biología celular, debido a la naturaleza compartimental y dinámica de la localización de las proteínas. Tradicionalmente, las localizaciones subcelulares de las proteínas se deducían a partir de métodos más tradicionales, como la microscopía de inmunofluorescencia, el marcaje con la proteína fluorescente verde o el fraccionamiento bioquímico. No obstante, se han ido desarrollando flujos de trabajo de proteómica cuantitativa para identificar con fiabilidad la proteína de orgánulos completos, y para atribuirles a las proteínas una localización subcelular. El enfoque típico se asienta en estrategias de fraccionamiento y enriquecimiento subcelular, normalmente a través de la centrifugación de gradientes de densidad o kits comerciales. Se pueden enriquecer también los

orgánulos celulares mediante inmunoprecipitación con anticuerpos específicos dirigidos contra epítopos presentes en la superficie del orgánulo. Las mezclas de proteínas obtenidas de las distintas fracciones de orgánulos se identifican y cuantifican entonces mediante los enfoques proteómicos tratados anteriormente. A pesar de que los enfoques basados en orgánulos pueden ofrecer una información valiosa sobre compartimentos subcelulares específicos en condiciones de aislamiento, también es importante estudiar la localización de las proteínas en el conjunto total de la célula con el fin de obtener una visión sistémica global de la organización del proteoma.

**Para más información:**

R. Drissi, M.-L. Dubois, F.-M. Boisvert. Proteomics methods for subcellular proteome analysis. *FEBS J.*, 2013, 280(22):5626-5634.

## 2.7. Detección y cuantificación de interacciones proteicas

La función de las proteínas dentro de la célula depende de la interacción entre varias clases diferentes de proteínas y de la formación de complejos multiproteicos. Estos complejos proteicos son altamente dinámicos y su composición va cambiando con el tiempo y con el estado de la célula. Se emplean tres enfoques principales para el estudio de la interacción de las proteínas, que son el «pull-down» por afinidad, el marcaje de proximidad y la elaboración de perfiles de correlación de proteínas. El pull-down por afinidad emplea anticuerpos específicos para aislar la proteína de interés y las proteínas con las que interacciona, que seguidamente se identifican y cuantifican mediante espectrometría de masas. El marcaje de proximidad es una técnica basada en la transferencia covalente de marcajes con biotina de una proteína a las proteínas próximas con las que pueda potencialmente interaccionar. A continuación, se identifica el marcaje con biotina en las proteínas utilizando la espectrometría de masas en tándem. La elaboración de perfiles de correlación de proteínas se basa en la separación a través de distintos métodos, como la centrifugación de gradientes de densidad o la cromatografía nativa, y parte de la hipótesis de que las proteínas interaccionando coeluirán.

Las interacciones proteína-proteína en la señalización celular están mediadas con frecuencia por secuencias de aminoácidos específicos que interactúan. A pesar de la importancia capital de estas interacciones en la señalización celular, la identificación de proteínas asociadas unidas a péptidos continúa siendo una tarea muy complicada, donde se recurre, principalmente, al uso de herramientas bioinformáticas, como STRING (<http://string-db.org/>).

**Para saber más:**

Syafrizayanti, C. Betzen, J.D. Hoheisel, D. Kastelic. Methods for analyzing and quantifying protein-protein interaction. *Expert. Rev. Proteomics*, 2014(2), 11(1):107-20.

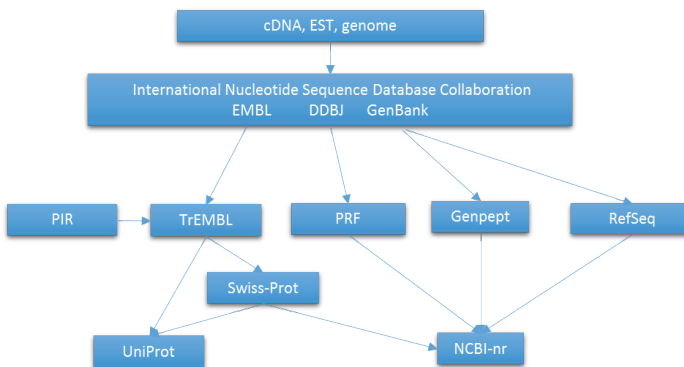
### 3. Procesamiento de datos e identificación de proteínas

#### 3.1. Bases de datos de secuencias de proteínas

Como hemos explicado anteriormente, las bases de datos de secuencias de proteínas son esenciales en la investigación proteómica con base en la espectrometría de masas. Actualmente, las principales fuentes de datos sobre secuencias de proteínas son traducciones de secuencias de nucleótidos depositadas en las bases de datos GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), EMBL (<http://www.ensembl.org/index.html>) o DDBJ (<http://www.ddbj.nig.ac.jp>). Estas bases de datos se integran en una iniciativa denominada International Nucleotide Sequence Database Collaboration (<http://www.insdc.org>) que aglutina secuencias de nucleótidos y anotaciones de esas entidades.

La base de datos de proteínas NCBI (Entrez Proteins, <http://www.ncbi.nlm.nih.gov/entrez>) posee un conjunto completo de proteínas deducidas. Utiliza información de la Protein Research Foundation (PRF), Genpet (secuencias no revisadas), RefSeq (secuencias de proteínas anotadas manualmente y revisadas), NCBI (National Center for Biotechnology Information) y Swiss-Prot (secuencias de proteínas anotadas manualmente y revisadas) (Figura 4.7). No obstante, dado que esta base de datos asigna a cada secuencia de proteínas un único número de identificación de gen (gi), resulta muy grande y redundante. NCBI también posee una base de datos de proteínas no redundante (nr), donde se fusionaron las secuencias y las secuencias de fragmentos para formar una sola entrada.

UniProt Knowledgebase (UniProtKB) contiene información sobre proteínas amplia y curada, incluyendo funciones, clasificación y referencias cruzadas, y se divide en dos bases de datos distintas. UniProtKB/Swiss-Prot, que se anota manualmente y se revisa, y UniProtKB/TrEMBL, que es la sección no revisada de UniProt.



**Figura 4.7.** Bases de datos de secuencias de proteínas más importantes y sus relaciones.

**Para saber más:**

C. Brooksbank, M.T. Bergman, R. Apweiler, E. Birney, J. Thornton. The European Bioinformatics Institute's data resources 2014. *Nucleic Acids Res.*, 2014, 42(D1):D18-D25.

### 3.2. Motores de búsqueda de bases de datos de proteínas

Como ya se dijo anteriormente, para identificar proteínas es necesario explorar bases de datos de proteínas para buscar coincidencias con las masas de péptidos (enfoque PMF) o las secuencias de péptidos (enfoque PFF). Para esto se han diseñado varias herramientas bioinformáticas, que se designan colectivamente como motores de búsqueda de bases de datos de proteínas. Estos motores de búsqueda pueden ser comerciales (Mascot (Matrix Science), Proteome Discoverer (Thermo), ProteinPilot (Sciex), ProteinLynx (Waters), ProteinScape (Bruker) y otros) o libres (Mascot (búsqueda en línea limitada) ProteinProspector, Andromeda, OMSSA, SEQUEST, X!Tandem, Amanda, y muchos otros). La correcta identificación de las proteínas a través de MS depende de numerosos factores, como por ejemplo, el algoritmo que empleen los motores de búsqueda y los factores experimentales que tengan una influencia en el contenido de la información de los datos de la MS (tratado ya antes). Así, estos motores de búsqueda no tienen por qué necesariamente identificar las mismas proteínas en una muestra y, de hecho, se suelen observar diferencias. Aun con todo, existen algunos motores de búsqueda, por ejemplo SearchGUI, que permiten combinar resultados de múltiples motores de búsqueda. En proteómica se adquiere una cantidad ingente de datos y, por ello, estos buscadores pueden pasar varios días por experimento, especialmente cuando se emplean experimentos de LC-MS/MS de larga duración, o cuando se ejecuta una búsqueda múltiple de PTM.

**Para saber más:**

Z.F. Yuan, S. Lin, R.C. Molden, B.A. García. Evaluation of proteomic search engines for the analysis of histone modifications. *J. Proteome Res.*, 2014, 13:4470-4478.

### 3.3. Análisis de datos

El principal problema de la proteómica basada en la espectrometría de masas está relacionado con el análisis e interpretación de los datos adquiridos. Este análisis incluye la integración de metainformación taxonómica y funcional procedente de muestras que contienen varios cientos de proteínas, de una forma que permita interpretar los experimentos. Para ello, una vez finalizada la identificación y cuantificación de proteínas, es necesario proceder con el análisis funcional de las proteínas diferenciales relevantes para el estudio. Existen varias herramientas bioinformáticas gratuitas que pueden ayudar en esta tarea, incluidos software gratuitos para todo el flujo de trabajo (pipelines). Estos software proteómicos (OpenMS Proteomics Pipeline, Trans-proteomic pipeline o PeptideShaker) ejecutan diferentes tareas analíticas, entre ellas, la de ayudar en el análisis de datos. El análisis de datos incluye la identificación de PTM, la comparativa e interpretación bioquímica de datos cuantitativos, la identificación de procesos biológicos e interacciones proteicas.

El primer paso para la interpretación funcional de la lista de proteínas resultante es conectar el identificador de proteínas con sus términos de Ontología de Genes asociados. El sistema de clasificación ontológica de genes (<http://www.geneontology.org/>) define conceptos o clases que se emplean para describir la función del gen y clasificar las funciones del gen desde tres perspectivas

diferentes: la función molecular del producto génico, el componente celular donde los productos génicos están activos y el proceso biológico donde los productos génicos están activos. El sistema de clasificación PANTHER (Protein Analysis Through Evolutionary Relationships –Análisis de proteínas a través de relaciones evolutivas–) (<http://www.pantherdb.org/about.jsp>), que forma parte del Proyecto Genoma de referencia en la ontología de genes, es una herramienta valiosa para la clasificación de proteínas y sus genes que facilita el análisis de alto rendimiento de rutas metabólicas. STRING (<http://string-db.org/>) y Cytoscape (<http://www.cytoscape.org/>) son dos herramientas bioinformáticas esenciales que se utilizan para el análisis de interacciones proteína-proteína y el diseño de redes de proteínas. Seguidamente, estos datos se pueden clasificar en categorías para ayudar al científico a interpretar los resultados de un experimento. Reactome (<http://www.reactome.org/>) ofrece información de la ruta metabólica. Por último, Uniprot (<http://www.uniprot.org/>) es también una fuente de información funcional de proteínas de alta calidad.

**Para más información:**

S.W. Haga, H.F. Wu. Overview of software options for processing, analysis and interpretation of mass spectrometric proteomic data. *J. Mass Spectrom.*, 2014, 49:959-969.

## 4. Sesiones prácticas

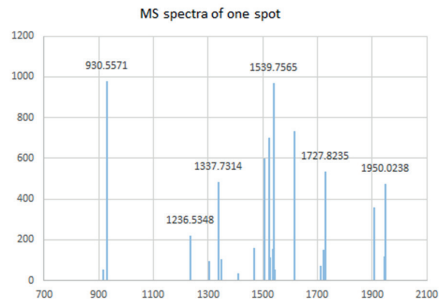
### 4.1. Identificación de proteínas empleando los enfoques PMF y PFF con MASCOT

Descripción del experimento

1. Se cargaron 120 µg de proteína de línea celular humana de riñón de embrión humano (HEK 293) en tiras de IPG, y se realizó un análisis 2-DE;
2. Tras la detección y escisión del gel con una punta de la pipeta, las manchas de proteínas fueron sometidas a digestión con tripsina después de la reducción y alquilación de los residuos de cisteína empleando para ello ditioneitol (DTT) e iodoacetamida (IAA);
3. Los péptidos trípticos fueron liofilizados y resuspendidos en 10 µL de una mezcla acuosa con 50% de acetonitrilo/0,1% de ácido fórmico. Las muestras se mezclaron (1:1) con una disolución matriz saturada de ácido  $\alpha$ -ciano-4-hidroxicinámico, y se rociaron alícuotas (0,5 µL) sobre la placa para el análisis de la muestra por MALDI;
4. Se obtuvieron espectros de masas de péptidos en un espectrómetro de masas MALDI-TOF/TOF (4700 Proteomics Analyzer, Applied Biosystems, Foster City, CA, EE. UU.) en modo reflectrón ion positivo;
5. Para cada mancha de muestra se creó un método de adquisición dependiente de datos para seleccionar los dos picos más intensos para la adquisición subsiguiente de datos MALDI-TOF/TOF MS/MS, excluyendo los de la matriz, los debidos a la autólisis de la tripsina o los picos de acrilamida.

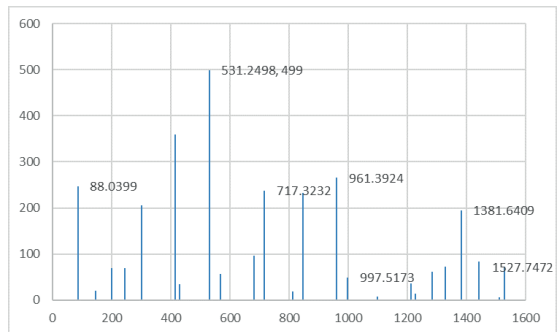
Espectros MS de una mancha:

<i>m/z</i>
916.4055
930.5571
1236.5348
1303.6696
1337.7314
1349.7409
1408.7048
1467.6749
1507.6632
1507.6686
1522.6397
1523.7924
<b>1527.7532</b>
1538.7780
1539.7565
1544.7386
1615.9085
1712.7815
1723.8684
1727.8235
1907.0442
1944.9352
1950.0238



MS/MS del ion a *m/z* 1527.7532

MS/MS	AR
88.0399	246
201.1301	70
302.1732	204
416.2143	358
531.2498	498
717.3232	237
846.3714	232
961.3924	266
1098.4522	7
1211.5352	35
1282.5703	61
1381.6409	195
1509.7398	5
1527.7472	72





### a) Identificación de una proteína a través del enfoque PMF

1. Vaya al sitio web de MASCOT: ([http://www.matrixscience.com/search\\_form\\_select.html](http://www.matrixscience.com/search_form_select.html))
2. Seleccione ejecutar una búsqueda en la opción «Peptide Mass Fingerprint» –huella peptídica– ([http://www.matrixscience.com/cgi/search\\_form.pl?FORMVER=2&SEARCH=PMF](http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF))
3. Cumplimente el formulario. Para la introducción de datos puede copiar y pegar la m/z de los péptidos o escoger el archivo de datos Mascot MS.txt
4. Debe reflexionar y discutir sobre lo siguiente:
  - a. ¿Por qué se seleccionaron las bases de datos Swiss Prot y de contaminantes?
  - b. ¿Por qué se permitieron dos cortes enzimáticos omitidos (missed cleavages)?
  - c. ¿Por qué se permitieron dos modificaciones de variables?
  - d. ¿Por qué la tolerancia es de 25 ppm?
  - e. ¿Qué es el «decoy»?

5. Después de realizar la búsqueda, se obtienen los siguientes resultados de la búsqueda de Mascot:

- Examine detenidamente los resultados. ¿Qué significan? Seleccione la proteína con la puntuación más alta y observe el Protein View: datos HS90A\_HUMAN.

### b) Identificación de una proteína a través del enfoque PFF

- Seleccione ejecutar una búsqueda en la opción «Peptide Mass Fingerprint» –huella peptídica– ([http://www.matrixscience.com/cgi/search\\_form.pl?FORMVER=2&SEARCH=MIS](http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS))
- Cumplimente el formulario. Para la introducción de datos puede escoger el archivo de datos Mascot MSMS.txt. Presione «iniciar búsqueda...»

**MASCOT MS/MS Ions Search**

Your name: Pedro Email: p.dominguez@ua.pt

Search title:

Database(s): contaminants (AA) SwissProt (AA)

Amino acid (AA): cRAP, NCBIprot, Nucleic acid (NA), Environmental\_EST, Fungi\_EST, Human\_EST, Invertebrates\_EST, Mammals\_EST, Mus\_EST

Taxonomy: Homo sapiens (human)

Enzyme: Trypsin Allow up to: 2 missed cleavages

Quantitation: None

Fixed modifications: none selected

Variable modifications: none selected

Peptide tol. ±: 25 ppm # 3.7C 0 MS/MS tol. ±: 50 mmu

Peptide charge: 1+ Monoisotopic: Average

Data file: Choose File Mascot MSMS.txt Precursor: m/z

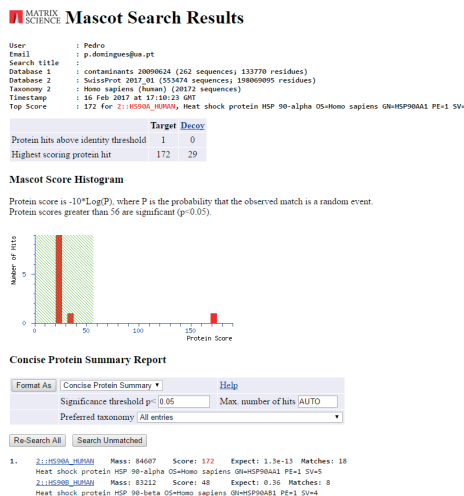
Data format: Mascot generic Error tolerant:

Instrument: Default Report top: AUTO hits

Decoy:  Report top: AUTO hits

Start Search ... Reset Form

- Examine detenidamente los resultados (Mascot Search Results). ¿Qué significan? Seleccione la proteína con la puntuación más alta y observe el Protein View: datos HS90A\_HUMAN



4. Seleccione el número 1 de la búsqueda y observe el Protein View: Observe el patrón de fragmentación de los péptidos. ¿Por qué los iones b son más intensos que los iones y?

## 4.2. Identificación de proteínas a partir de un experimento de LC-MS

Para este tutorial necesitará descargar los siguientes archivos:

- a) Pida a su instructor los datos primarios (brutos) del experimento (1,2 GB) y el archivo transformado (MGF, 100 MB).
- b) ProteoWizard - Se trata de una interfaz de código abierto para la conversión de software en formato MS. Solo necesitará esto si descarga el archivo primario. Además, SearchGUI (ver abajo) incluye ya esta aplicación. <http://proteowizard.sourceforge.net/>
- c) mMass (opcional) - Se trata de una interfaz de espectrometría de masas de código abierto que se puede utilizar para la visualización de espectros. Cuenta también con varias herramientas distintas que incluyen herramientas de proteómica <http://www.mmass.org/>
- d) Dbtoolkit - Se trata de un editor de datos FASTA de código abierto para la manipulación de bases de datos de secuencias fasta <https://github.com/compomics/dbtoolkit>
- e) SearchGUI - Se trata de una interfaz de código abierto para la configuración y funcionamiento de motores de búsqueda de identificación proteómica. <http://compomics.github.io/projects/searchgui.html>
- f) Peptide shaker - Se trata de una plataforma independiente de motores de búsqueda para la interpretación de resultados de identificación proteómica. <http://compomics.github.io/projects/peptide-shaker.html>

Además, es muy importante que lea con atención todo el tutorial del mezclador de péptidos en <https://compomics.com/bioinformatics-for-proteomics/>

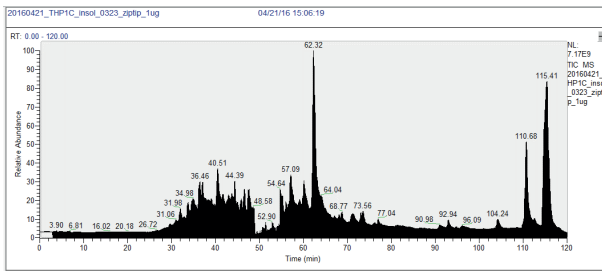
### Descripción del experimento

1. Para una identificación rápida de las proteínas intracelulares más abundantes de la Línea Celular THP1 (línea celular monocítica humana), las células se rasparon de la placa empleando 500 µL de PBS/pocillo y se sometieron a ultrasonidos en un baño a 53 kHz y 37°C durante 30 minutos. Se precipitaron las proteínas añadiendo a las suspensiones del lisado celular dos volúmenes de acetona. Se incubaron las suspensiones de proteínas a -20°C y a continuación se centrifugaron. Los pellets de proteínas secos fueron resuspendidos en 30 µL de tampón de almacenamiento de muestra. Las cantidades de las proteínas extraídas se midieron a través del método de Lowry.
2. La digestión triptica en disolución se realizó con tripsina. Los residuos de cisteína de la proteína se redujeron con DTT y se alquilaron con iodoacetamida. Los péptidos tripticos fueron liofilizados y resuspendidos en una disolución de 5% de acetonitrilo/0,1% de ácido fórmico.
3. Se analizaron 250 ng de extracto de proteína de la muestra con un QExactive Orbitrap que

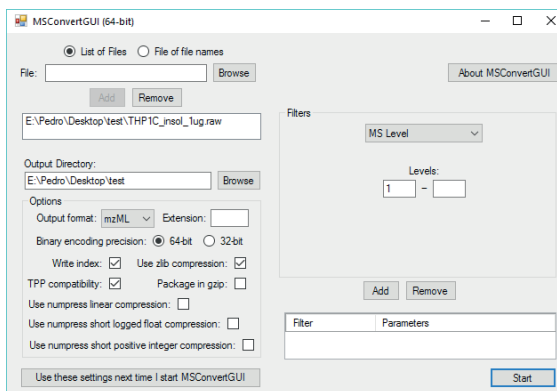
- se acopló a un sistema de nano-HPLC Ultimate 3000. Se utilizaron una precolumna (5 mm × 300 µm I.D.) y una columna analítica (150 mm × 75 µm I.D.) Se utilizaron columnas C18.
- El espectrómetro de masas se utilizó en el modo de adquisición dependiente de datos. Los diez picos más intensos se sometieron a HCD.

#### 4.2.1. Transformación de datos a mzML con MSConverter en ProteoWizard

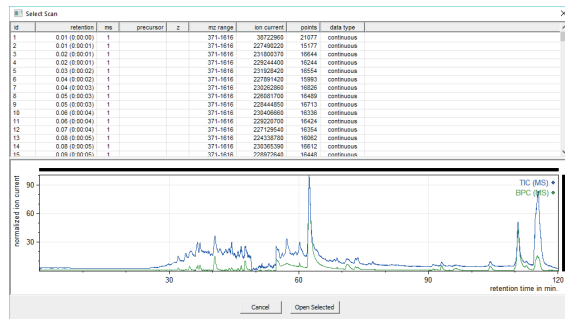
- Utilice MSConverter en ProteoWizard para convertir los datos de HPC-MS data adquiridos en la Orbitrap (archivo de datos BRUTOS) a un formato que se pueda leer con SearchGUI (archivo de datos MGF).
- Si quiere ver los datos en la aplicación mMass, aunque pueda leer el formato de archivo mgf que se utilice en SearchGUI, es mejor que convierta los datos BRUTOS (o en mgf) a formato mzML.



Cromatograma de datos BRUTOS



Aplicación MSConverter



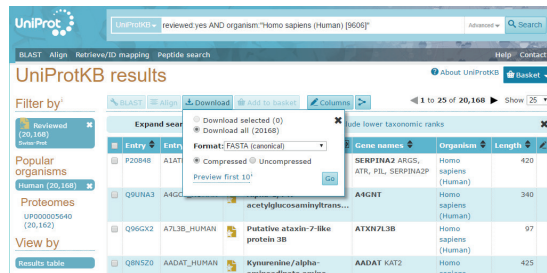
Datos observados en la aplicación mmass.

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

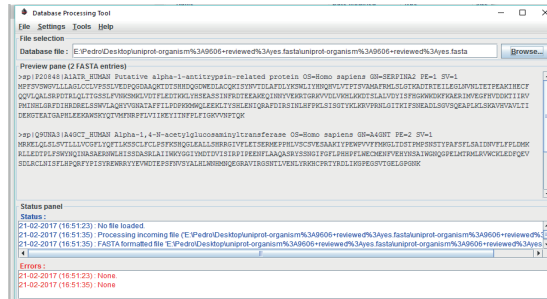
1. Ahora descargue el archivo FASTA del proteoma de Homo Sapiens. Esto se debe hacer en la página de taxonomía del sitio web de Uniprot y buscando Homo Sapiens.



2. Ahora debe descargar el archivo FASTA (de Swiss-Prot) Revisado.

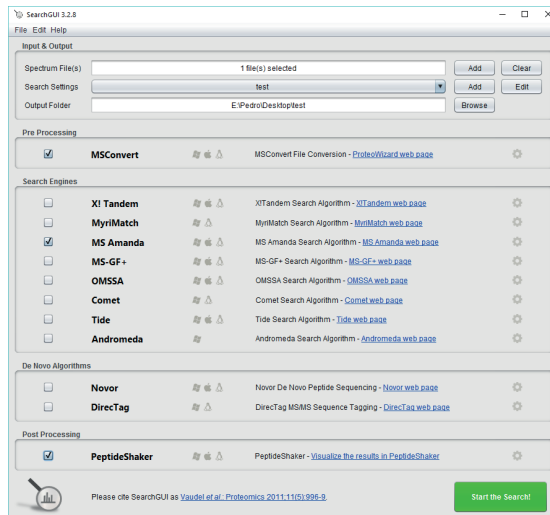


3. Puede ver y editar la información del archivo FASTA con la dataBase Processing Tool (herramienta de procesamiento de bases de datos) de Dbtoolkit.

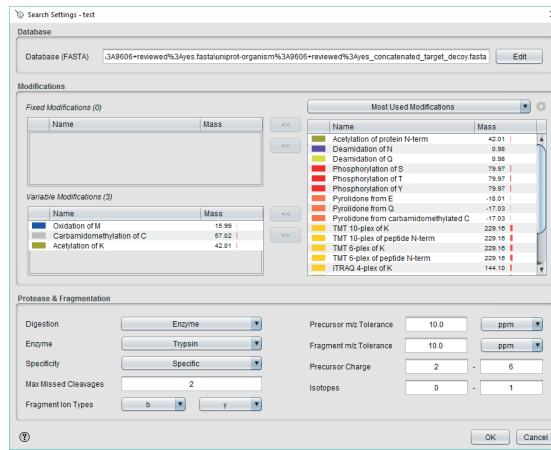


#### 4.2.3. Motores de búsqueda: SearchGUI

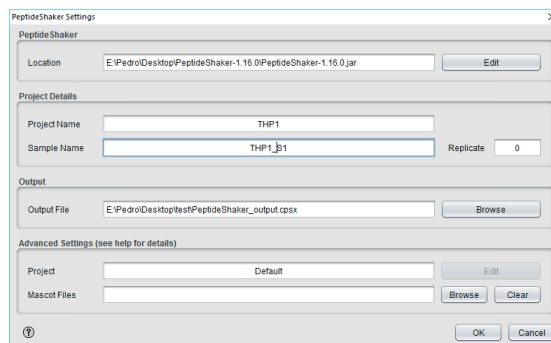
1. Abra SearchGUI. En el ejemplo anterior, el archivo seleccionado era el archivo bruto, por lo que también se seleccionó la conversión MS.
2. En los ajustes de la búsqueda, edite y cumplimente el formulario del modo que se muestra. Se escogieron modificaciones de variables (¿por qué?).
3. Puede configurar también el mezclador de péptidos para abrir el archivo de resultados, tal y como se muestra abajo.
4. SearchGUI preguntará si quiere crear un archivo fasta concatenated\_target\_decoy. Seleccione sí (¿por qué es esto importante?)



5. Seleccione las opciones de búsqueda y la base de datos FASTA.

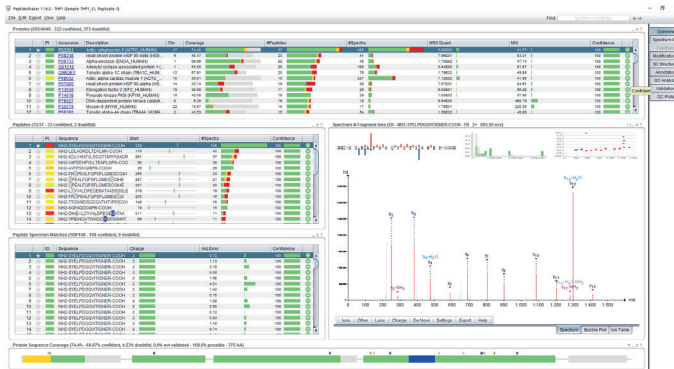


6. Seleccione las opciones de PeptideShaker.



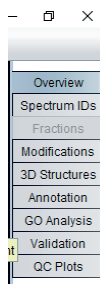
### 4.2.4. Generación y evaluación de resultados: PeptideShaker para la visualización de péptidos y proteínas, y validación. Análisis PTM

1. Después de que SearchGUI haya ejecutado la búsqueda (3 min aproximadamente, con un Intel I7-6700K de 16MB de RAM), abrirá los resultados en la plataforma de Peptideshaker.
2. Podrá ver aquí que se han identificado aproximadamente 700 proteínas, aunque 373 hayan sido clasificadas como dudosas (¿por qué?).
3. Además, podrá ver información sobre los péptidos identificados para cada proteína y los espectros de masas con el patrón de fragmentación anotado.



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

1. Explore las opciones avanzadas de análisis de datos del Peptideshaker abriendo la pestaña de modificaciones, y la pestaña «GO Analysis».



2. En la pestaña «Annotation» podrá anotar información para cada proteína.
3. No obstante, si quiere anotar múltiples proteínas, será necesario que exporte sus resultados (informe de proteínas por defecto) y que haga clic en el enlace web que hay al lado del recurso, y que siga las instrucciones que se ofrecen en la página web del recurso.

The screenshot shows the Peptideshaker software interface for protein annotation. The protein is PRQD1 (Accession: PRQD1, Description: Adh1, OYE036912.2, UCT2\_HUMAN). The interface includes a 'Protein Annotation - Help' section with instructions on how to use the tool. Below this, there are several search buttons for different databases and resources, such as UniProt, RefSeq, and NCBI. The 'GO Analysis' button is highlighted in yellow.



### **Agradecimientos:**

Gracias a la Fundação para a Ciência e a Tecnologia (FCT, Portugal), la Unión Europea, el Quadro de Referência Estratégica Nacional (QREN), el Programa Operacional Potencial Humano (POPH), el Fundo Europeu de Desenvolvimento Regional (FEDER) y el Programa Operacional Factores de Competitividade (COMPETE) por financiar la unidad de investigación QOPNA (Química Orgânica de Produtos Naturais e Agroalimentares), y a RNEM (REDE/1504/REM/2005) por la Red Nacional de Espectrometría de Masas de Portugal.

Este manuscrito está bajo la licencia Creative Commons Attribution-NonCommercial 3.0 CC BY-NC 3.0. Se atribuye a Pedro Domingues, Antonia García y Elżbieta Skrzydlewska como editores.

