

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 767 375**

21 Número de solicitud: 202030061

51 Int. Cl.:

**G16C 20/20** (2009.01)

**G01N 30/72** (2006.01)

12

SOLICITUD DE PATENTE

A1

22 Fecha de presentación:

**24.01.2020**

43 Fecha de publicación de la solicitud:

**17.06.2020**

71 Solicitantes:

**CONSORCIO CENTRO DE INVESTIGACIÓN  
BIOMÉDICA EN RED, M.P. (45.0%)  
C/ Monforte de Lemos, 3-5, Pabellón 11  
28029 MADRID ES;  
FUNDACIÓ INSTITUT D'INVESTIGACIÓ  
SANITÀRIA PERE VIRGILI (17.5%);  
UNIVERSITAT ROVIRA I VIRGILI (17.5%) y  
GINÉ BERTOMEU, Roger (20.0%)**

72 Inventor/es:

**GINÉ BERTOMEU, Roger;  
YANES TORRADO, Óscar y  
CAPELLADES TOMÁS, Jordi**

74 Agente/Representante:

**CARVAJAL Y URQUIJO, Isabel**

54 Título: **MÉTODO Y SISTEMA PARA LA IDENTIFICACIÓN DE COMPUESTOS EN MUESTRAS  
BIOLÓGICAS O AMBIENTALES COMPLEJAS**

57 Resumen:

Método y sistema para la identificación de compuestos en muestras biológicas o ambientales complejas. El método comprende: recibir (102) un espectro de masas (1) de una espectrometría de masas acoplada a una técnica de separación; para cada punto (2) del espectro de masas (1), anotar (106) en una base de datos de anotaciones (12) combinaciones de fórmulas y aductos cuya relación masa carga teórica  $(m/z)^T$  corresponde con la relación masa carga medida  $(m/z)$  del punto (2); para cada fórmula y aducto anotado, detectar (108) regiones de interés en un rango de tiempo de retención  $(RT_0-RT_1)$  según unos criterios de caracterización; generar (110) una lista de inclusión (14) con los rangos de tiempo de retención  $(RT_0-RT_1)$  y las relaciones de masa carga teóricas  $(m/z)^T$  de las fórmulas y aductos asociados a las regiones de interés; y enviar (112) la lista de inclusión a un espectrómetro de masas para la identificación de compuestos en la muestra mediante espectrometría de masas en tándem.

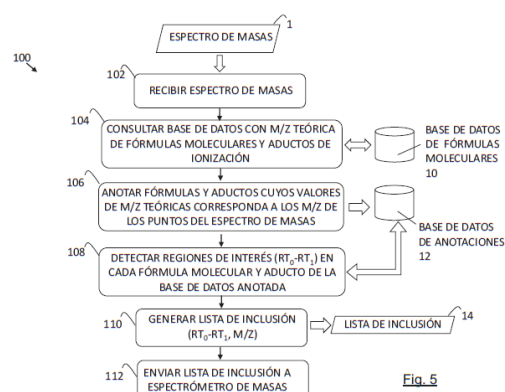


Fig. 5

ES 2 767 375 A1

**DESCRIPCIÓN****MÉTODO Y SISTEMA PARA LA IDENTIFICACIÓN DE COMPUESTOS EN MUESTRAS  
BIOLÓGICAS O AMBIENTALES COMPLEJAS**

5

**Campo de la invención**

La presente invención se engloba dentro del campo de la metabolómica, que se caracteriza por el análisis de metabolitos y pequeñas moléculas orgánicas en muestras biológicas o ambientales complejas, tales como plasma, orina, tejidos y aguas residuales.

10

**Antecedentes de la invención**

Los espectrómetros de masas modernos de muy alta resolución (>40.000 FWHM) y exactitud de masa (<1-5ppm), denominados HRMS (siglas de “High Resolution Mass Spectrometry”), realizan barridos de masas (en modo MS1 o “full scan”) con gran rapidez (en unos pocos milisegundos) con el objetivo de analizar iones originados por compuestos ionizados de una muestra biológica o ambiental compleja.

Cuando un espectrómetro de masas de alta resolución se acopla a una técnica de separación (conocido como “hyphenated MS”), por ejemplo de cromatografía líquida (LC-HRMS, “Liquid Chromatography-High Resolution Mass Spectrometry”) o de electroforesis capilar (CE-HRMS, “Capillary Electrophoresis-High Resolution Mass Spectrometry”), para experimentos de metabolómica no dirigida, la matriz de datos crudos puede contener decenas o centenares de miles de puntos (comúnmente conocidos en inglés como “scans” o “data points”) en el caso de muestras complejas.

25

Sin embargo, hasta el momento la anotación e identificación de metabolitos mediante LC-HRMS o CE-HRMS en estudios de metabolómica no dirigida es complicada, y el número de metabolitos identificados es bastante limitado. La presente invención propone un nuevo procedimiento que permite identificar muchos más (e incluso todos) de los compuestos ionizados de la muestra biológica, por lo que aumenta la cobertura de posibles biomarcadores detectados.

**Descripción de la invención**

La presente invención se refiere a un método para el análisis de datos de metabolómica no dirigida basada en acoplamiento de la espectrometría de masas con una técnica de

separación, por ejemplo de cromatografía líquida (LC-MS) o de electroforesis capilar (CE-MS). En la espectrometría de masas se pueden emplear diferentes métodos de ionización para producir iones, como por ejemplo ionización por electrospray (ESI, "Electrospray ionization") o ionización química a presión atmosférica (APCI, "Atmospheric pressure chemical ionization"). La muestra analizada puede ser una muestra biológica (plasma, tejidos, etc.) o una muestra ambiental compleja (e.g. aguas residuales).

De acuerdo a un primer aspecto de la presente invención, se presenta un método para la identificación de compuestos en muestras biológicas o ambientales complejas. El método comprende las siguientes etapas:

- Recibir un espectro de masas de un análisis de espectrometría de masas acoplado a una técnica de separación aplicado a una muestra, donde el espectro de masas comprende una pluralidad de puntos con información de tiempo de retención, relación masa carga medida e intensidad de la señal medida.
- Consultar una base de datos de fórmulas moleculares que incluye la relación masa carga teórica del ion molecular de una pluralidad de fórmulas moleculares y de aductos de ionización.
- Para cada punto del espectro de masas, anotar en una base de datos de anotaciones las combinaciones de fórmulas moleculares y aductos de ionización cuya relación masa carga teórica corresponde con la relación masa carga medida de dicho punto considerando un determinado error de masa, donde cada anotación incluye el tiempo de retención y la intensidad de la señal medida del punto.
- Para cada fórmula molecular y aducto de ionización anotada en la base de datos de anotaciones, detectar regiones de interés definidas en un rango de tiempo de retención donde los puntos anotados cumplen unos criterios de caracterización.
- Generar una lista de inclusión que incluye los rangos de tiempo de retención de las regiones de interés detectadas y las relaciones de masa carga teóricas de las fórmulas moleculares y aductos de ionización asociados a cada una de las regiones de interés.
- Enviar la lista de inclusión a un espectrómetro de masas para la identificación de compuestos en la muestra mediante espectrometría de masas en tándem.

El método puede comprender una etapa de detección en el espectro de masas de isotopólogos asociados a las fórmulas moleculares y/o aductos de ionización anotados. La

detección de isotopólogos comprende:

- Buscar, en el rango de tiempo de retención de cada región de interés, puntos del espectro de masas cuya relación masa carga medida corresponda, considerando un error de masa, con una relación masa carga teórica de un isotopólogo de la fórmula molecular y/o aducto de ionización asociada a la región de interés.
- Obtener la intensidad de la señal medida de los puntos encontrados.
- Calcular una intensidad teórica de los puntos encontrados a partir de la intensidad de la señal medida de los puntos de la región de interés correspondientes a la fórmula molecular y/o aducto de ionización.
- Comparar las intensidades medidas con las intensidades teóricas calculadas.
- Determinar la detección del isotopólogo en base a dicha comparación.

En una realización, la detección de las regiones de interés comprende determinar regiones candidatas, definidas en un rango de tiempo de retención con un número mínimo de puntos y/o una densidad mínima de puntos anotados; caracterizar las regiones candidatas (20), obteniendo unos parámetros de caracterización; y seleccionar como regiones de interés aquellas regiones candidatas cuyos parámetros de caracterización cumplen unos determinados criterios de caracterización.

Los criterios de caracterización empleados pueden ser muy diversos:

- Calcular una pendiente de una regresión lineal de los puntos anotados en las regiones candidatas, y comprobar que el valor absoluto de la pendiente calculada es superior a una pendiente umbral.
- Calcular una intensidad media y/o una intensidad máxima de la señal medida de los puntos anotados en las regiones candidatas, y comprobar que la intensidad media y/o la intensidad máxima calculada es superior a una intensidad media y/o máxima umbral.
- Calcular un rango de intensidad de la señal medida de los puntos anotados en las regiones candidatas, estando el rango de intensidad definido por una relación entre la intensidad máxima y la intensidad mínima en la región candidata, y comprobar que el rango de intensidad calculado es superior a un rango de intensidad umbral.
- Calcular una relación señal/ruido entre un nivel de intensidad asociado a los puntos anotados en la región candidata y un nivel de intensidad asociado a los puntos del espectro de masas localizados en un área circundante a la región

candidata, y comprobar que la relación señal/ruido calculada es superior a una relación señal/ruido umbral. El área circundante a la región candidata puede quedar definida por un espacio delimitado por:

- i. un rango de relación masa carga que incluye un rango de relación masa carga correspondiente a la región candidata, y
- ii. un rango de tiempo de retención que incluye el rango de tiempo de retención correspondiente a la región candidata.

El método puede comprender definir un conjunto de fórmulas moleculares en función de la muestra a analizar, definir aductos de ionización asociados a las fórmulas moleculares, y generar la base de datos de fórmulas moleculares incluyendo, para cada fórmula molecular y aducto de ionización asociado, la relación masa carga teórica.

El método puede comprender realizar un análisis de espectrometría de masas acoplado a una técnica de separación aplicado a la muestra para obtener el espectro de masas.

El método puede comprender realizar un análisis de espectrometría de masas en tándem utilizando la información incluida en la lista de inclusión para identificar compuestos en la muestra.

Un segundo aspecto de la presente invención se refiere a un sistema para la identificación de compuestos en muestras biológicas o ambientales complejas. El sistema comprende una unidad de control con medios de procesamiento de datos configurados para ejecutar las etapas del método previamente definido.

El sistema puede comprender un espectrómetro de masas encargado de realizar un análisis de espectrometría de masas acoplado a una técnica de separación en la muestra para obtener el espectro de masas.

El sistema puede comprender un espectrómetro de masas encargado de realizar un análisis de espectrometría de masas en tándem utilizando la información incluida en la lista de inclusión para identificar compuestos en la muestra.

La presente invención también se refiere a un producto de programa para la identificación de compuestos en muestras biológicas o ambientales complejas. El producto de programa

comprende instrucciones de programa para llevar a cabo el método previamente definido cuando el programa se ejecuta en un procesador. El producto de programa puede comprender al menos un medio de almacenamiento legible por computadora que almacena las instrucciones de programa.

5

### **Breve descripción de los dibujos**

A continuación se pasa a describir de manera muy breve una serie de dibujos que ayudan a comprender mejor la invención y que se relacionan expresamente con una realización de dicha invención que se presenta como un ejemplo no limitativo de ésta.

10

La Figura 1 representa un espectro de masas adquirido por un espectrómetro de masas acoplado a una técnica de separación.

15

Las Figuras 2A y 2B ilustran, según el estado del arte, la detección de región de interés y picos espectrales en el proceso de anotación del espectro de masas.

Las Figuras 3A, 3B y 3C representan, de acuerdo al estado del arte, la agrupación de picos espectrales en el proceso de anotación del espectro de masas.

20

Las Figuras 4A y 4B representan, respectivamente, las formas de elución del trifosfato de adenosina y del S-adenosil metionina.

25

La Figura 5 representa un diagrama de flujo de una realización del método la presente invención.

Las Figuras 6A y 6B representa el número de solapamientos de fórmulas y aductos anotados para un mismo punto del espectro de masas considerando un error de masa de 1 ppm y de 5 ppm, respectivamente, del espectrómetro de masas.

30

La Figura 7 ilustra un diagrama de flujo del proceso de detección de regiones de interés de acuerdo a una realización.

La Figura 8 muestra un ejemplo de determinación y caracterización de regiones candidatas.

35

La Figura 9 representa la caracterización de las regiones candidatas de acuerdo a diferentes

criterios.

La Figura 10 ilustra el área circundante a la región candidata empleada para determinar un criterio de caracterización de las regiones candidatas (relación señal/ruido).

5 La Figura 11 ilustra un diagrama de flujo del proceso de detección de isotopólogos de acuerdo a una realización de la presente invención.

10 Las Figuras 12A y 12B muestran dos ejemplos de los efectos de la resolución del espectrómetro de masas en la posible detección de isotopólogos.

La Figura 13 representa un ejemplo donde se aprecia el patrón de isotopólogos (M1, M2) real y teórico de una fórmula concreta (M0).

## 15 Descripción detallada de la invención

Los puntos 2 de un espectro de masas 1 adquirido en modo MS1 en un espectrómetro de masas acoplado a una técnica de separación (e.g., cromatografía líquida acoplada a una espectrometría de masas, LC-MS, o electroforesis capilar acoplada a una espectrometría de masas, CE-MS) contiene, tal y como se representa en la gráfica de la **Figura 1**, tres ejes de  
 20 información: masa/carga de los iones detectados ( $m/z$ ), intensidad (proporcional a la abundancia de los iones detectados), y tiempo de elución o tiempo de retención RT ("Retention Time"). Cada punto 2 ("scan" o "data point") del espectro de masas 1 contiene normalmente información en un rango de masa carga  $m/z$  bastante amplio (por ejemplo, desde  $m/z$  100 hasta 1.000) para un instante de tiempo determinado, pudiendo disponer de  
 25 hasta miles de medidas (en función de la resolución del equipo) de relaciones masa carga  $m/z$ .

En la actualidad, la anotación del espectro de masas 1 en modo MS1 (anotación MS1) sigue el siguiente esquema:

30 1) Se emplea un algoritmo (e.g., CentWave) para la detección de regiones de interés 3 (ROI, "Regions of Interest") en los datos sin procesar que aplica una transformación continua de wavelets y el ajuste de Gauss en el dominio de separación cromatográfico, o cualquier otra técnica de separación acoplada a HRMS (en el eje horizontal el tiempo de retención RT y en el eje vertical la intensidad de la señal  
 35 medida, como se representa en la **Figura 2A**), para detectar picos espectrales 4 en

todo el espectro de masas 1, para diferentes valores de relaciones masa carga ( $m/z$ ) y tiempo de retención (RT), como se muestra en la **Figura 2B**.

- 2) Posteriormente, otro algoritmo (e.g., CAMERA, CliqueMS) agrupa los picos espectrales 4 pertenecientes al mismo compuesto debido a la redundancia que hay por aductos e isótopos (**Figura 3A**). Los fragmentos de la fuente mostrados en la Figura 3A son principalmente debidos a la pérdida de un agua, es decir:  $[M-H_2O+H]^+$  en ionización positiva, o  $[M-H_2O-H]^-$  en ionización negativa. La agrupación de los picos espectrales 4 se puede realizar mediante una correlación de la forma del pico (**Figura 3B**), en la que se busca una correlación alta de la forma de los picos. Los picos espectrales 4 que muestran una correlación débil no se agrupan. La agrupación se puede realizar también mediante una correlación de la abundancia o intensidad de los picos utilizando diferentes muestras. En el ejemplo de la **Figura 3C** se aprecia un ratio casi constante entre la intensidad del pico de la relación masa/carga A y la intensidad del pico de la relación masa/carga C (coeficiente de determinación de la regresión lineal  $R^2=0.98$ ). De forma similar, existe una fuerte correlación entre la relación masa/carga B y la relación de masa/carga D ( $R^2=0.92$ ). Sin embargo, no existe ninguna correlación entre las relaciones masa/carga A y B ( $R^2=0.03$ ). Este procedimiento tiene sin embargo serias limitaciones cuando las formas de elución de los metabolitos no se ajustan a la función (e.g., gaussiana) que se pretende ajustar a los datos, lo cual ocurre por ejemplo con el trifosfato de adenosina (ATP, **Figura 4A**) o el S-adenosil metionina (SAM, **Figura 4B**).

Una vez completada la anotación MS1, para la caracterización o identificación de metabolitos se realiza una anotación MS2 utilizando espectrometría de masas en tándem o MS<sup>n</sup> ( $n \geq 2$ ). En la actualidad existen tres procedimientos para la identificación de metabolitos mediante LC-MS/CE-MS (o LC-HRMS/CE-HRMS) y MS<sup>n</sup> en metabolómica no dirigida:

- Lista de inclusión (MS/MS dirigido): Se analizan las muestras en modo MS<sup>1</sup> y se procesan los datos mediante uno o más programas de software que detectan y alinean picos (como se ha explicado en la Figura 3A). Siguiendo normalmente criterios de cambios estadísticos entre grupos o condiciones experimentales, una proporción de estos  $m/z$  son fragmentados mediante análisis MS<sup>2</sup> o MS<sup>n</sup> en un experimento posterior.



- Adquisición dependiente de datos (DDA): El espectrómetro de masas recopila datos de  $MS^1$  y  $MS^n$  en un mismo análisis metabolómico no dirigido. Un breve ciclo de trabajo de reconocimiento  $MS^1$  de los  $m/z$  que eluyen en ese momento sirve para controlar la intensidad de los  $m/z$  y para identificar/seleccionar posibles  $m/z$  a fragmentar. Luego, se aplica “n” ciclos de  $MS^2$  o  $MS^n$ , durante cada uno de los cuales se aísla un sólo  $m/z$  precursor, se fragmenta y se detectan sus fragmentos. Los precursores están fragmentados en orden decreciente de intensidad. Normalmente se utiliza una ventana de exclusión dinámica para asegurarse de que los  $m/z$  que han sido analizados recientemente por  $MS^2$  no se vuelvan a fragmentar constantemente si hay nuevos  $m/z$  disponibles.
- Adquisición independiente de datos (DIA): El espectrómetro de masas recopila datos de  $MS^1$  y  $MS^n$  en un mismo análisis metabolómico no dirigido. Es un método de determinación de estructuras moleculares en el que todos los iones dentro de un rango de  $m/z$  seleccionado se fragmentan y se detecta la mezcla de fragmentos. Los espectros de masas en  $MS^n$  se adquieren ya sea fragmentando todos los iones que ingresan al espectrómetro de masas en un momento dado (DIA de banda ancha) o aislando y fragmentando secuencialmente todos los iones en rangos de  $m/z$  (SWATH™, Sequential windowed acquisition of all theoretical mass spectra).

La presente invención consiste en un nuevo procedimiento para procesar datos crudos de un análisis LC-HRMS o CE-HRMS en modo  $MS^1$  y seleccionar relaciones masa/carga ( $m/z$ ) y rangos de tiempo de retención (RT) para la identificación de metabolitos en un posterior análisis realizado mediante espectrometría de masas en tándem o  $MS^n$  ( $n \geq 2$ ).

El método 100 de la presente invención comprende los pasos mostrados en el diagrama de flujo de la **Figura 5**.

En primer lugar, el método 100 comprende recibir 102 un espectro de masas 1 de un análisis LC-MS o CE-MS (adquirido en modo  $MS^1$ ) aplicado sobre una muestra biológica o ambiental. El espectro de masas 1 comprende una pluralidad de puntos 2 con información que incluye el tiempo de retención (RT), la relación masa carga medida ( $m/z$ ) y la intensidad de la señal medida.

A continuación, se accede o consulta 104 una base de datos de fórmulas moleculares 10 que incluye la relación masa carga teórica  $(m/z)^T$  del ion molecular de una pluralidad de fórmulas moleculares y de aductos de ionización asociados. En una realización, la base de datos de fórmulas moleculares 10 comprende un listado de fórmulas y un listado de aductos, con la masa carga teórica de cada fórmula y de cada aducto, de forma que la masa carga teórica de las combinaciones de fórmulas moleculares y aductos de ionización se puede calcular posteriormente a partir de la masa monoisotópica de la fórmula molecular más la diferencia de masa que aporta el aducto de ionización al cargarse en la fuente (e.g., H, Na, K). En otra realización, la base de datos de fórmulas moleculares 10 almacena directamente la masa carga teórica de las distintas combinaciones de fórmulas y aductos, de forma que no es necesario un cálculo posterior.

El contenido de la base de datos de fórmulas moleculares 10, o la información a la que se accede en la consulta 104, está preferentemente orientado a la muestra particular a analizar, basado en un gran universo o espacio de fórmulas moleculares relacionadas con la matriz que se va a analizar (suero, orina, células, muestras ambientales, etc.). En el caso de matrices biológicas de interés biomédico se puede emplear las fórmulas moleculares incluidas en la base de datos Human Metabolome Database (HMDB). Por ejemplo, se puede definir un conjunto de fórmulas moleculares en función de la muestra a analizar y aductos de ionización cuya asociación a las fórmulas moleculares sea conocida. Se puede considerar bases de datos que incluyan únicamente las fórmulas moleculares orientadas a la muestra en particular (por ejemplo con las fórmulas que se esperen encontrar en plasma sanguíneo), o bases de datos más amplias, como la base de datos HMDB que incluye información sobre más de 10.000 metabolitos que se encuentran en el cuerpo humano.

Una vez están definidas las fórmulas moleculares y sus aductos de ionización, se puede generar el contenido de la base de datos de fórmulas moleculares 10 incluyendo, para cada ion molecular de la fórmula molecular y para cada aducto de ionización asociado, la relación masa carga teórica  $(m/z)^T$ , la cual se puede obtener directamente de la fórmula molecular considerando los pesos atómicos correspondientes. El método puede comprender la etapa de generación de la base de datos de fórmulas moleculares 10. Alternativamente, la base de datos de fórmulas moleculares 10 puede haber sido ya creada previa a la implementación del método 100, de forma que el método 100 únicamente requiera acceder a una memoria (e.g., en un dispositivo local o en la nube) en la que se almacena la base de datos de fórmulas moleculares 10 previamente generada.

La construcción de la base de datos de fórmulas moleculares 10 puede comprender la generación de una tabla que contiene todas las relaciones masa carga teóricas  $(m/z)^T$  tras considerar los isotopólogos principales (e.g., M1, M2, M3) y los aductos conocidos tanto en ionización positiva como negativa (los fragmentos en la fuente se pueden considerar como un aducto en la lista de aductos) para cada fórmula molecular única considerada. La información contenida en la base de datos de fórmulas moleculares 10 puede por ejemplo estructurarse en forma de tabla, donde se incluya en cada fila una fórmula/aducto/isotopólogo diferente. La tabla puede estar ordenada por relación masa carga teórica  $(m/z)^T$ , la primera columna, como se representa en el siguiente ejemplo:

$m/z^T$	Fórmula	Aducto	Isotopólogo
376.2312	C <sub>21</sub> H <sub>27</sub> NO <sub>4</sub>	+NH <sub>4</sub>	M1

El procedimiento busca todos los valores teóricos de relación masa carga  $(m/z)^T$  en cada punto 2 del espectro de masas 1 de LC-MS o CE-MS, dentro de un error predefinido (normalmente entre 1 y 5 ppm). Alternativamente, se hace un barrido en los puntos 2 del espectro de masas 1 y se comprueba, para cada punto 2, si su relación masa carga medida  $(m/z)$  corresponde con alguna relación masa carga teórica  $(m/z)^T$  de la base de datos de fórmulas moleculares 10. Para facilitar la búsqueda, la base de datos de fórmulas moleculares 10 puede incluir los datos ordenados de menor a mayor relación masa carga teórica  $(m/z)^T$ .

Para cada punto 2 del espectro de masas 1, se anota 106 en una base de datos de anotaciones 12 las fórmulas moleculares y los aductos de ionización cuya relación masa carga teórica  $(m/z)^T$  corresponde con la relación masa carga medida  $(m/z)$  de dicho punto, considerando un determinado margen o error de masa (procedente de la exactitud de la medición o calibración del espectrómetro de masas). La base de datos de anotaciones 12 incluye, para cada fórmula molecular y aducto de ionización anotado, el tiempo de retención (RT) y la intensidad de la señal medida del punto asociado a la fórmula/aducto. La información contenida en la base de datos de anotaciones 12 puede estructurarse por ejemplo en forma de tabla, donde se incluya en cada fila una anotación diferente. Cada fila será por tanto una nueva anotación que incluirá la fórmula y/o aducto anotado, su correspondiente tiempo de retención (RT), la intensidad de la señal medida del punto 2 del espectro de masas 1 asociado y, opcionalmente, la relación masa carga medida  $(m/z)$ .

Fórmula	Aducto	RT	Intensidad	m/z medida
C21H27NO4	+NH4			375.2281

Las distintas anotaciones de tiempos de retención (RT) y de intensidad que se realicen en la base de datos de anotaciones 12 para una misma fórmula y aducto en diferentes filas de la tabla se pueden agrupar (e incluso representar en una gráfica, como la mostrada en la

5 Figura 8 para el aducto +NH4 de la fórmula C21H27NO4) para el posterior análisis de detección de regiones de interés.

Según el error de masa definido, existirá un mayor o un menor solapamiento de posibles fórmulas y/o aductos anotados para un mismo punto 2. En la gráfica de la **Figura 6A** se

10 representa en el eje horizontal el número de solapamientos producidos, considerando un error de masa de 1 ppm, para las distintas relaciones de masa carga (m/z) de los puntos del espectro de masas 1 (desde 0 solapamientos hasta 7 solapamientos), y en el eje vertical el número de ocurrencias para cada número de solapamiento diferente. Por ejemplo, la

15 relación masa/carga (m/z) medida en un punto 2 del espectro de masas 1 puede corresponder a dos fórmulas/aductos diferentes, considerando el error de masa de 1 ppm: el aducto de ionización negativa -H-NH3 de la fórmula molecular C6H9NO2, y el ion molecular negativo -H de la fórmula molecular C6H6O2. Se produce por ello un solapamiento entre dos fórmulas y/o aductos. Cuando el solapamiento se produce entre N fórmulas y/o aductos, se considera que existen N-1 solapamientos. De los casi 100.000 puntos 2 del espectro de

20 masas 1 en el ejemplo de la Figura 6A, en más de 40.000 puntos no hay solapamientos, en más de 20.000 puntos hay 1 solapamiento y en más de 10000 puntos ocurren dos solapamientos. A medida que se incrementa el error de masa, aumenta el solapamiento entre diferentes fórmulas-aductos posibles (en la **Figura 6B** se muestra el solapamiento con error de masa de 5 ppm).

25 A continuación, una vez realizada la anotación 106, se procede a analizar cada fórmula molecular y aducto de ionización de la base de datos de anotaciones 12, agrupando todas las anotaciones ocurridas para una misma fórmula/aducto (ver ejemplo de la Figura 8), con el objeto de detectar 108 regiones de interés definidas en un rango de tiempo de retención

30 (RT<sub>0</sub>-RT<sub>1</sub>) en el cual los puntos anotados cumplen unos determinados criterios de caracterización. Una misma fórmula/aducto de la base de datos de anotaciones 12 puede incluir una única región de interés o varias regiones de interés detectadas en diferentes rangos de tiempo de retención.

El método 100 implementa un algoritmo para encontrar regiones de interés basado en la comprobación de uno o varios criterios de caracterización, considerando primeramente un criterio de densidad mínima y/o número mínimo de puntos en la región de interés (lo cual determinará regiones candidatas), y considerando a continuación criterios adicionales, como una pendiente mínima de los puntos en la región interés o un ratio señal/ruido mínimo determinado. Las regiones de interés detectadas también se pueden comparar, de manera opcional pero recomendable, con un blanco de muestra para descartar falsos positivos o puntos exógenos a la muestra.

Por tanto, y a diferencia del estado del arte, la determinación de las regiones de interés no consiste en encontrar picos en el espectro de masas 1 ajustando un modelo (p.ej. gaussiano) a los datos. El enfoque del nuevo procedimiento es independiente de la forma y determinación de los picos espectrales 4, no siendo necesario realizar ningún tipo de correlación entre los picos espectrales (como se muestra en las Figuras 3B y 3C para el estado del arte). Esto hace que el procedimiento de la presente invención sea independiente de las condiciones cromatográficas.

La **Figura 7** muestra un diagrama de flujo del proceso de detección 108 de regiones de interés de acuerdo a una realización. La detección 108 de regiones de interés comprende determinar 122, para cada fórmula molecular y aducto de ionización de la base de datos de anotaciones 12, regiones candidatas 20 definidas en un rango de tiempo de retención ( $RT_{C0}$ - $RT_{C1}$ ), como las representas en el ejemplo de la **Figura 8**, con un número mínimo de puntos y/o una densidad mínima de puntos anotados en la región candidata 20. Este paso corresponde a realizar un filtrado por densidad de puntos, solo considerando posibles regiones de interés (i.e. regiones candidatas 20) a aquellas ventanas de tiempo que agrupen un número mínimo de puntos y/o una densidad mínima de puntos. En el ejemplo mostrado en la Figura 8 se representan los puntos 2 del espectro de masas 1 anotados en la base de datos de anotaciones 12 para el aducto  $[M+NH_4]^+$  de la fórmula molecular  $C_{21}H_{27}NO_4$ . También se muestran las regiones candidatas 20 que han superado el filtrado por densidad; por ejemplo, se seleccionan como regiones candidatas 20 los rangos de tiempo que recogen al menos cinco puntos 2 del espectro de masas 1 en un rango máximo de tiempo determinado.

A continuación, se caracterizan 124 las regiones candidatas 20, obteniendo unos parámetros de caracterización 22 de las regiones candidatas 20. Finalmente, se comparan 126 los parámetros de caracterización 22 obtenidos con unos criterios de caracterización, y se seleccionan 128 como regiones de interés aquellas regiones candidatas 20 cuyos parámetros de caracterización 22 cumplen unos determinados criterios de caracterización.

En la **Figura 9** se muestran distintas formas de caracterizar 124 a las regiones candidatas 20 y diferentes criterios de caracterización que se puede considerar deben cumplir las regiones candidatas 20. Por ejemplo, se puede considerar, entre otros, cualquier combinación de los siguientes criterios de caracterización:

- Una pendiente mínima de los puntos 2 en la región candidata 20: La caracterización de las regiones candidatas 20 puede comprender calcular 132 la pendiente ( $m$ ) de una regresión lineal 24 (ver Figura 8) de los puntos 2 anotados en las regiones candidatas 20. Los criterios de caracterización pueden incluir comprobar 142 que el valor absoluto de la pendiente calculada es superior a una pendiente mínima ( $m_{\min}$ ) o pendiente umbral.
- Una intensidad media y/o máxima de la señal medida en la región candidata: La caracterización de las regiones candidatas puede comprender calcular 134 una intensidad media ( $I_{\text{med}}$ ) y/o calcular 136 una intensidad máxima ( $I_{\text{max}}$ ) de la señal medida de los puntos 2 anotados en las regiones candidatas 20. Los criterios de caracterización puede incluir comprobar 144 que la intensidad media ( $I_{\text{med}}$ ) calculada es superior a una intensidad media umbral ( $I_{\text{med}}^{\text{TH}}$ ) y/o comprobar 146 que la intensidad máxima ( $I_{\text{max}}$ ) calculada es superior a una intensidad máxima umbral ( $I_{\text{max}}^{\text{TH}}$ ).
- Un rango de intensidad en la región candidata: La caracterización de las regiones candidatas puede comprender calcular 138 un rango de intensidad de la señal medida de los puntos anotados en las regiones candidatas, donde el rango de intensidad está definido por una relación entre la intensidad máxima y la intensidad mínima en la región candidata (e.g. una relación logarítmica entre el máximo valor de intensidad,  $I_{\text{max}}$ , y el mínimo valor de intensidad,  $I_{\text{min}}$ , de los puntos 2 anotados en la región candidata 20). Los criterios de caracterización pueden incluir comprobar 148 que el rango de intensidad calculado es superior a un rango de intensidad umbral.

- Una mínima relación señal/ruido (SNR): La caracterización de las regiones candidatas puede comprender calcular 140 una relación señal/ruido (SNR) entre un nivel de intensidad asociado a los puntos 2 anotados en la región candidata 20 y un nivel de intensidad asociado a los puntos 2 del espectro de masas 1 localizados en un área circundante a la región candidata 20. Los criterios de caracterización pueden incluir comprobar 150 que la relación señal/ruido (SNR) calculada es superior a una relación señal/ruido umbral ( $SNR^{TH}$ ). De acuerdo a la realización mostrada en la **Figura 10**, el área circundante 26 a la región candidata 20 puede quedar definida por un espacio delimitado por un rango de relación masa carga ( $m/z_{P0}-m/z_{P1}$ ) que incluye el rango de relación masa carga ( $m/z_{C0}-m/z_{C1}$ ) correspondiente a la región candidata 20, y por un rango de tiempo de retención ( $RT_{P0}-RT_{P1}$ ) que incluye el rango de tiempo de retención ( $RT_{C0}-RT_{C1}$ ) correspondiente a la región candidata 20, donde dicho espacio puede incluir o no a la propia región candidata 20 (en la Figura 10 el rango  $m/z_{C0}-m/z_{C1}$  de la región candidata no está a escala, se ha ampliado con fines ilustrativos; en la práctica el rango  $m/z_{P0}-m/z_{P1}$  es mucho mayor que el rango  $m/z_{C0}-m/z_{C1}$ , incluso hasta unas 100000 veces mayor). La región candidata se puede considerar que incluye un rango de relación masa carga ( $m/z_{C0}-m/z_{C1}$ ) ya que se considera un error de masa en la anotación en la base de datos de anotaciones 12.
- Una mínima y/o máxima amplitud del rango de tiempo de retención ( $RT_0-RT_1$ ) de las regiones de interés (i.e. mínima y/o máxima distancia de tiempo desde el inicio al final de la región ( $RT_0-RT_1$ )).

Sin embargo, es posible utilizar otros parámetros o criterios de caracterización diferentes. Además, los criterios de caracterización se pueden acoplar a técnicas de aprendizaje automático (redes neuronales artificiales, bosques aleatorios, etc.) para filtrar regiones candidatas 20 y generar una lista de inclusión más concreta a cambio de aplicar un sesgo asociado al propio método de aprendizaje.

En el ejemplo de la Figura 8, la región candidata de la izquierda no se selecciona como región de interés debido a que no cumple el criterio de una pendiente mínima ( $|m| < m_{min}$ ). La región candidata central tampoco se selecciona como región de interés porque la intensidad media ( $I_{med}$ ) de sus puntos 2 es inferior a una intensidad media umbral ( $I_{med}^{TH}$ ). La región candidata 20 de la derecha sí se selecciona 128 como región de interés 28 debido a que los parámetros de caracterización 22 cumplen los criterios de caracterización exigidos (e.g.,

$|m| > m_{\min}$ ;  $I_{\text{med}} > I_{\text{med}}^{\text{TH}}$ , etc.). En el caso representado, la región de interés 28 coincide con la región candidata ( $RT_{\text{CO}}=RT_0$ ,  $RT_{\text{C1}}=RT_1$ ). Sin embargo, la región de interés 28 finalmente considerada puede resultar de la agrupación de otras regiones solapadas (por ejemplo, agrupación de regiones candidatas o de otras regiones de interés solapados).

5

El método 100 continúa con la generación 110 de una lista de inclusión 14 anotada y muy precisa, con rangos de tiempo variable según el perfil de elución de cada  $m/z$ , para experimentos de MS/MS (o  $MS^n$ ) que facilita la identificación de metabolitos. La lista de inclusión 14 incluye los rangos de tiempo de retención ( $RT_0$ - $RT_1$ ) de las regiones de interés detectadas y las relaciones masa carga teórica  $(m/z)^T$  de las fórmulas moleculares y/o aductos de ionización asociados a cada una de las regiones de interés detectadas. Opcionalmente, la lista de inclusión puede incluir también las fórmulas moleculares y/o aductos de ionización asociados a cada una de las regiones de interés detectadas.

10

15

Finalmente, la lista de inclusión 14 se envía 112 a un espectrómetro de masas para, mediante un análisis de espectrometría de masas en tándem, realizar la identificación de metabolitos en la muestra utilizando los datos de la lista de inclusión 14. Opcionalmente, el método puede comprender realizar el análisis de espectrometría de masas en tándem utilizando la información incluida en la lista de inclusión para identificar metabolitos en la muestra. Los análisis de MS/MS son posteriores a los barridos de masas en modo MS1 realizado en el análisis LC-MS, requiriendo de una segunda inyección de la misma muestra ya que actualmente no existe una tecnología para acumular o guardar iones después de ser detectados en el MS1.

20

25

El nuevo procedimiento analiza los puntos del espectro de masas de una muestra biológica representativa, adquirida en modo MS1, para seleccionar aquellas relaciones masa carga  $m/z$  (y sus rangos de tiempo) que serán fragmentados en posteriores experimentos de  $MS^n$ . Un aspecto novedoso de la presente invención es la manera de seleccionar las relaciones masa carga  $(m/z)$  y los rangos de tiempo de retención para realizar el análisis  $MS^n$ , ya que al no basarse en la detección de picos es un método independiente del perfil de elución cromatográfica del compuesto, pudiendo detectar metabolitos con formas de elución no gaussianas o similares (como los de la Figuras 4A y 4B). Además, en el caso de que se envíe al espectrómetro de masas las fórmulas moleculares y/o aductos de ionización asociados a las regiones de interés detectadas, el espectrómetro de masas puede utilizar esta

30



información en los análisis posteriores a la fragmentación para identificar más rápidamente los compuestos, ya que parte de una lista determinada de fórmulas candidatas.

Además, la presente invención presenta una forma novedosa de detección de isotopólogos de fórmulas moleculares y/o aductos de ionización en el espectro de masas 1. La detección de isotopólogos se puede comprobar una vez se han detectado 108 las regiones de interés 28 de las fórmulas moleculares y aductos de ionización. La detección de isotopólogos 120 comprende, tal y como se representa en el diagrama de flujo de la **Figura 11**, buscar 162 en el rango de tiempo de retención ( $RT_0$ - $RT_1$ ) de cada región de interés 28 (o al menos en un intervalo de tiempo comprendido en dicho rango  $RT_0$ - $RT_1$ ), puntos 2 del espectro de masas 1 cuya relación masa carga medida ( $m/z$ ) corresponda, considerando un error de masa, con una relación masa carga teórica ( $m/z$ )<sup>T</sup> de un isotopólogo (e.g., M1) de la fórmula molecular y aducto de ionización (M0) asociado a la región de interés. De esta forma se comprueba que la relación masa carga  $m/z$  cuadre con la teórica del isotopólogo, considerando un error de masa del espectrómetro.

A continuación, se obtiene 164 la intensidad de la señal medida de cada uno de los puntos encontrados en la búsqueda 162. Se calcula 166 una intensidad teórica de los puntos encontrados en la búsqueda 162 a partir de la intensidad de los puntos de la región de interés (i.e. los puntos correspondientes a la fórmula/aducto principal M0), y en función del ratio teórico de abundancia del isotopólogo en cuestión (ya sea M1, M2, etc.) que se espere encontrar con respecto a la fórmula o aducto principal M0. Por ejemplo, si el ratio teórico de la abundancia de un isotopólogo M1 es de un 2.5% con respecto a la fórmula/aducto principal M0, la intensidad teórica del isotopólogo sería un 2.5% del nivel de intensidad de los puntos de la región de interés. Se comparan 168 las intensidades medidas con las intensidades teóricas calculadas y se determina 170, en base a dicha comparación, la detección o no del isotopólogo. En una realización se comprueba, para cada uno de los puntos encontrados, si la intensidad medida del punto corresponde con la intensidad teórica del isotopólogo, considerando un determinado margen de intensidad (para contemplar por ejemplo posibles errores de sensibilidad en la medición o divergencia con respecto al ratio de abundancia teórica del isotopólogo con respecto a la fórmula/aducto M0). Para calcular la intensidad teórica del isotopólogo se considera la intensidad del M0 ( $Int(M0)$ ) correspondiente (i.e. la intensidad de la señal medida del punto de la región de interés 28 en un instante de tiempo RT correspondiente -en el mismo scan-) y el ratio de abundancia teórica (*ratio*) del isotopólogo con respecto al M0. En la comparación 168 de las

intensidades medidas con las intensidades teóricas se considera un margen de intensidad; por ejemplo, se comprueba que la intensidad medida del isotópologo ( $Int(iso)$ ) esté incluida en un intervalo construido (en función de un valor  $k$ ) en torno al valor teórico ( $Int(M0)*ratio$ ) que le correspondería al isotópologo:

$$5 \quad Int(M0)*ratio*(1+k) > Int(iso) > Int(M0)*ratio*(1-k)$$

A continuación se puede realizar opcionalmente una comprobación adicional basada en la comparación por similitud coseno (cosine similarity), el cual se define como:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

10 Dicha comprobación se puede realizar de la siguiente forma:

- Se busca en la base de datos de anotaciones 12 las entradas que correspondan a las condiciones a comparar (e.g., M0 comparado con un isotópologo M1) en el intervalo de RT que corresponde a la región de interés que se está analizando correspondiente al M0.
- 15 • Se busca todas aquellas entradas de cada conjunto que compartan el tiempo de retención RT (es decir, que se haya encontrado entradas de las dos condiciones M0 y M1 en un mismo scan -i.e., mismo instante de tiempo RT-).
- Si hay suficientes entradas (e.g. más de 5, para evitar falsos positivos cuando N es pequeña), se calcula similitud coseno (sea  $I = \langle i_1, i_2, i_3 \dots i_N \rangle$  y  $J = \langle j_1, j_2, j_3 \dots j_N \rangle$  los vectores de las intensidades de las dos condiciones a comparar):

20

$$\text{Cos} = (i_1 j_1 + i_2 j_2 + \dots i_N j_N) / (\text{modulo}(I) * \text{modulo}(J))$$

- Si  $\text{Cos} > k$  (e.g.  $k=0.99$ ), entonces se determina que se ha encontrado un isotopologo y se registra cuál ha sido.

25 La búsqueda 162 de puntos correspondientes a un isotópologo de una fórmula y aducto determinada puede realizarse consultando la base de datos de anotaciones 12, la cual puede incluir anotaciones de los isotopólogos (M1, M2,...) además de las anotaciones de las fórmulas/aductos (M0). Para ello, cuando se realiza una anotación 106 de una fórmula/aducto (M0), se comprueba la existencia de un punto con relación masa carga correspondiente a un isotópologo (M1, M2,...) y una intensidad próxima a la teórica, y en  
30 ese caso se realiza la anotación del isotópologo. Alternativamente, se puede realizar la

búsqueda 162 de isotopólogos directamente en el espectro de masas 1 (ya que se conoce el instante de tiempo RT y la relación de masa carga donde hay que buscar).

La búsqueda de los isotopólogos cuya presencia o ausencia hay que determinar para cada  
 5 fórmula y/o aducto anotado puede venir determinada en la base de datos de fórmulas  
 moleculares 10, la cual puede incluir por ejemplo los isotopólogos a considerar para cada  
 fórmula y/o aducto (por ejemplo, los isotopólogos principales M1 y M2 de cada  
 fórmula/aducto M0) y su correspondiente relación masa/carga teórica  $(m/z)^T$ . La base de  
 10 datos de fórmulas moleculares 10 también puede incluir el ratio teórico de abundancia del  
 isotopólogo. En una realización, se determinan los isotopólogos que se pueden teóricamente  
 detectar en función de la resolución de masas del espectro en el rango de relación masa  
 carga  $m/z$  analizado, lo cual permite ajustar para cada M0 el espacio de isotopólogos que el  
 espectrómetro de masas puede llegar a detectar en función de la resolución del equipo. La  
 información relativa a los isotopólogos se puede incluir por ejemplo en una base de datos de  
 15 isotopólogos, en la cual se almacena la composición de los isotopólogos (M1, M2,...)  
 detectables con el espectrómetro de masas, la relación masa carga  $m/z$  con respecto al M0  
 y el ratio de abundancia.

Por tanto, el método permite calcular el patrón isotópico de cada fórmula y diferenciar qué  
 20 isotopólogos son detectables por el aparato dado el ratio de intensidad respecto al M0 y la  
 resolución del espectrómetro de masas. El procedimiento para determinar si los picos de los  
 isotopólogos calculados son separables depende del analizador de masas utilizado (tal y  
 como se explica por ejemplo en el documento "Orbitrap Mass Spectrometry", Zubarev et al.,  
 Analytical Chemistry 2013, 85 (11), págs. 5288-5296). En el caso de analizadores Orbitrap,  
 25 la resolución es inversamente proporcional a la raíz cuadrada de la  $m/z$ , y por tanto se  
 puede calcular matemáticamente. En el caso de analizadores FTICR, la resolución tiene una  
 escala inversa a la  $m/z$ , por lo que también se puede calcular matemáticamente. En cambio,  
 la resolución en analizadores TOF es independiente de la  $m/z$ , por lo que la resolución de  
 cada  $m/z$  se calcula mediante una curva de calibración.

30 En las Figuras 12A y 12B se muestra un ejemplo de patrones isotópicos (M0, M1 y M2) de la  
 fenilalanina (C<sub>9</sub>H<sub>11</sub>N<sub>5</sub>O<sub>2</sub>), en la que se explica el efecto de la resolución para distinguir  
 isotopólogos y cómo una mayor resolución permite distinguir otros isotopólogos que no sean  
 M1 y M2. La Figura 12A corresponde a una resolución 200000 (tipo Orbitrap) y la Figura 12B  
 35 a una resolución 60000 (tipo qTOF). En estas figuras se representan:

- Líneas verticales azules: representan las relaciones masa carga  $m/z$  teóricas (y su abundancia), de acuerdo con la abundancia relativa de cada isótopo natural de cada átomo.
  - Líneas curvas verdes: A consecuencia de que los equipos, según su resolución, no pueden distinguir perfectamente entre las relaciones masa carga  $m/z$  teóricas (en su caso, los detectados), lo que realmente se aprecia en el espectro de masas es una curva que los engloba (línea curva verde). Según la resolución, esa curva define mejor o peor las líneas verticales azules.
  - Líneas verticales rojas: son una simplificación (una suma) de la línea curva verde, una manera de evitar recoger todos los puntos de la línea curva verde, y en su lugar englobarlos en una única señal (llamada centroide). Este valor es una "media ponderada" de las relaciones masa carga  $m/z$  englobadas dentro de la curva verde, y su abundancia.
- En el ejemplo de la Figura 12A, la resolución 200000 (tipo Orbitrap) es suficiente para separar totalmente los isotopólogos M1 y M2. Sin embargo, para el caso mostrado en la Figura 12B (resolución 60000 qTOF), los isotopólogos M1 pueden separarse pero los isotopólogos M2 no se distinguen y no se pueden separar.
- La **Figura 13** representa un ejemplo real donde se aprecia el patrón de isotopólogos (M1, M2) de una fórmula M0 en particular, y como siguen aproximadamente el ratio de intensidad teórico calculado (en línea discontinua).

En caso de solapamiento de varias fórmulas-aductos para una determinada relación de masa carga ( $m/z$ ), el número de isotopólogos asociados a una misma fórmula que hayan sido detectados por el método puede ser utilizado para priorizar una fórmula candidata respecto a otra, aportando información relevante de qué compuesto se puede tratar antes de realizar siquiera la espectrometría de masas en tándem.

## REIVINDICACIONES

1. Un método para la identificación de compuestos en muestras biológicas o ambientales  
5 complejas, caracterizado por que comprende:

recibir (102) un espectro de masas (1) de un análisis de espectrometría de masas acoplado a una técnica de separación aplicado a una muestra, donde el espectro de masas (1) comprende una pluralidad de puntos (2) con información de tiempo de retención (RT), relación masa carga medida ( $m/z$ ) e intensidad de la señal medida;

10 consultar (104) una base de datos de fórmulas moleculares (10) que incluye la relación masa carga teórica ( $m/z$ )<sup>T</sup> del ion molecular de una pluralidad de fórmulas moleculares y de aductos de ionización;

para cada punto (2) del espectro de masas (1), anotar (106) en una base de datos de anotaciones (12) las combinaciones de fórmulas moleculares y aductos de ionización cuya  
15 relación masa carga teórica ( $m/z$ )<sup>T</sup> corresponde con la relación masa carga medida ( $m/z$ ) de dicho punto (2) considerando un determinado error de masa, donde cada anotación incluye el tiempo de retención (RT) y la intensidad de la señal medida del punto (2);

para cada fórmula molecular y aducto de ionización anotada en la base de datos de anotaciones (12), detectar (108) regiones de interés definidas en un rango de tiempo de  
20 retención (RT<sub>0</sub>-RT<sub>1</sub>) donde los puntos anotados cumplen unos criterios de caracterización;

generar (110) una lista de inclusión (14) que incluye los rangos de tiempo de retención (RT<sub>0</sub>-RT<sub>1</sub>) de las regiones de interés detectadas y las relaciones de masa carga teóricas ( $m/z$ )<sup>T</sup> de las fórmulas moleculares y aductos de ionización asociados a cada una de las regiones de interés; y

25 enviar (112) la lista de inclusión a un espectrómetro de masas para la identificación de compuestos en la muestra mediante espectrometría de masas en tándem.

2. El método de la reivindicación 1, que comprende detectar en el espectro de masas (1) isotopólogos asociados a las fórmulas moleculares y/o aductos de ionización anotados,  
30 donde la detección de isotopólogos comprende:

buscar (162), en el rango de tiempo de retención (RT<sub>0</sub>-RT<sub>1</sub>) de cada región de interés (28), puntos (2) del espectro de masas (1) cuya relación masa carga medida ( $m/z$ ) corresponda, considerando un error de masa, con una relación masa carga teórica ( $m/z$ )<sup>T</sup> de un isotopólogo de la fórmula molecular y/o aducto de ionización asociada a la región de  
35 interés (28);

obtener (164) la intensidad de la señal medida de los puntos encontrados;  
 calcular (166) una intensidad teórica de los puntos encontrados a partir de la  
 intensidad de la señal medida de los puntos de la región de interés (28) correspondientes a  
 la fórmula molecular y/o aducto de ionización;

- 5           comparar (168) las intensidades medidas con las intensidades teóricas calculadas;  
              determinar (170) la detección del isotopólogo en base a dicha comparación.

3. El método de cualquiera de las reivindicaciones anteriores, donde la detección (108) de  
 las regiones de interés comprende:

- 10           determinar (122) regiones candidatas (20) definidas en un rango de tiempo de  
 retención ( $RT_{C0}-RT_{C1}$ ) con un número mínimo de puntos y/o una densidad mínima de puntos  
 anotados;

             caracterizar (124) las regiones candidatas (20), obteniendo unos parámetros de  
 caracterización (22); y

- 15           seleccionar (128) como regiones de interés aquellas regiones candidatas (20) cuyos  
 parámetros de caracterización (22) cumplen unos determinados criterios de caracterización.

4. El método de la reivindicación 3, donde la caracterización (124) de las regiones  
 candidatas (20) comprende calcular (132) una pendiente ( $m$ ) de una regresión lineal (24) de  
 20           los puntos (2) anotados en las regiones candidatas (20);

y donde los criterios de caracterización comprenden comprobar (142) que el valor absoluto  
 de la pendiente ( $m$ ) calculada es superior a una pendiente umbral ( $m_{min}$ ).

5. El método de cualquiera de las reivindicaciones 3 a 4, donde la caracterización (124) de  
 25           las regiones candidatas (20) comprende calcular (134, 136) una intensidad media ( $I_{med}$ ) y/o  
 una intensidad máxima ( $I_{max}$ ) de la señal medida de los puntos (2) anotados en las regiones  
 candidatas (20);

y donde los criterios de caracterización comprenden comprobar (144, 146) que la intensidad  
 media ( $I_{med}$ ) y/o la intensidad máxima ( $I_{max}$ ) calculada es superior a una intensidad media  
 30           ( $I_{med}^{TH}$ ) y/o máxima ( $I_{max}^{TH}$ ) umbral.

6. El método de cualquiera de las reivindicaciones 3 a 5, donde la caracterización (124) de  
 las regiones candidatas (20) comprende calcular (138) un rango de intensidad de la señal  
 medida de los puntos (2) anotados en las regiones candidatas (20), estando el rango de  
 35           intensidad definido por una relación entre la intensidad máxima y la intensidad mínima en la

región candidata (20);

y donde los criterios de caracterización comprenden comprobar (148) que el rango de intensidad calculado es superior a un rango de intensidad umbral.

- 5 7. El método de cualquiera de las reivindicaciones 3 a 6, donde la caracterización (124) de las regiones candidatas (20) comprende calcular (140) una relación señal/ruido (SNR) entre un nivel de intensidad asociado a los puntos (2) anotados en la región candidata (20) y un nivel de intensidad asociado a los puntos (2) del espectro de masas (1) localizados en un área circundante (26) a la región candidata (20);
- 10 y donde los criterios de caracterización comprenden comprobar (150) que la relación señal/ruido (SNR) calculada es superior a una relación señal/ruido umbral ( $SNR^{TH}$ ).

8. El método de la reivindicación 7, donde el área circundante (26) a la región candidata (20) queda definida por un espacio delimitado por un rango de relación masa carga ( $m/z_{P0}-m/z_{P1}$ ) que incluye un rango de relación masa carga ( $m/z_{C0}-m/z_{C1}$ ) correspondiente a la región candidata (20), y por un rango de tiempo de retención ( $RT_{P0}-RT_{P1}$ ) que incluye el rango de tiempo de retención ( $RT_{C0}-RT_{C1}$ ) correspondiente a la región candidata (20).
- 15

9. El método de cualquiera de las reivindicaciones anteriores, que comprende:

- 20 - definir un conjunto de fórmulas moleculares en función de la muestra a analizar;  
 - definir aductos de ionización asociados a las fórmulas moleculares; y  
 - generar la base de datos de fórmulas moleculares (10) incluyendo, para cada fórmula molecular y aducto de ionización asociado, la relación masa carga teórica ( $m/z$ )<sup>T</sup>.

- 25 10. El método de cualquiera de las reivindicaciones anteriores, que comprende realizar un análisis de espectrometría de masas acoplado a una técnica de separación aplicado a la muestra para obtener el espectro de masas (1).

- 30 11. El método de cualquiera de las reivindicaciones anteriores, que comprende realizar un análisis de espectrometría de masas en tándem utilizando la información incluida en la lista de inclusión para identificar compuestos en la muestra.

- 35 12. Un sistema para la identificación de compuestos en muestras biológicas o ambientales complejas, caracterizado por que comprende una unidad de control con medios de procesamiento de datos configurados para ejecutar las etapas del método según cualquiera

de las reivindicaciones 1-11.

13. El sistema de la reivindicación 12, que comprende un espectrómetro de masas encargado de realizar un análisis de espectrometría de masas acoplado a una técnica de  
5 separación en la muestra para obtener el espectro de masas (1).

14. El sistema de cualquiera de las reivindicaciones 12 a 13, que comprende un espectrómetro de masas encargado de realizar un análisis de espectrometría de masas en  
10 tandem utilizando la información incluida en la lista de inclusión para identificar compuestos en la muestra.

15. Un producto de programa para la identificación de compuestos en muestras biológicas o ambientales complejas, que comprende instrucciones de programa para llevar a cabo el  
15 método definido en cualquiera de las reivindicaciones 1-11 cuando el programa se ejecuta en un procesador.

16. El producto de programa según la reivindicación 15, que comprende al menos un medio de almacenamiento legible por computadora que almacena las instrucciones de programa.



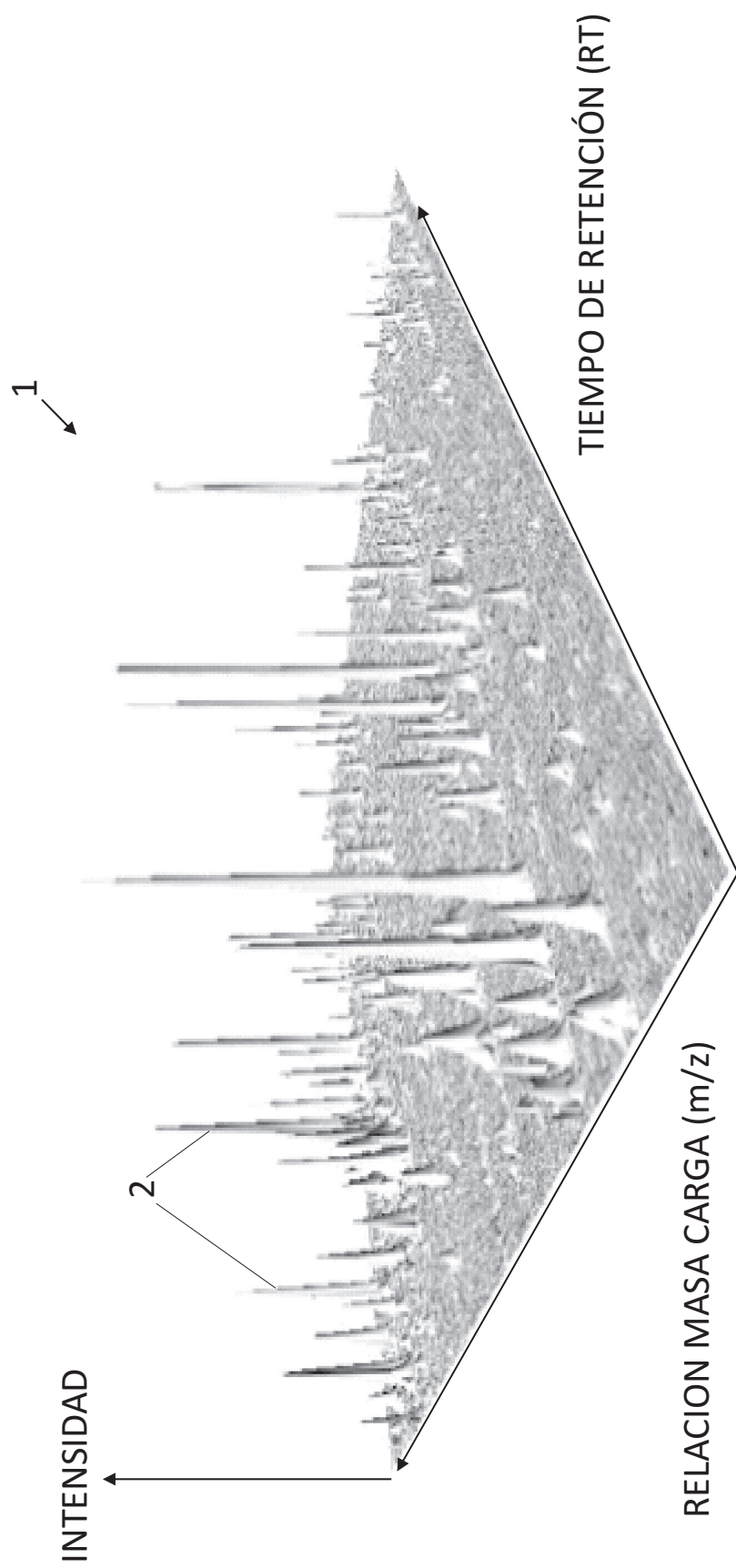


Fig. 1

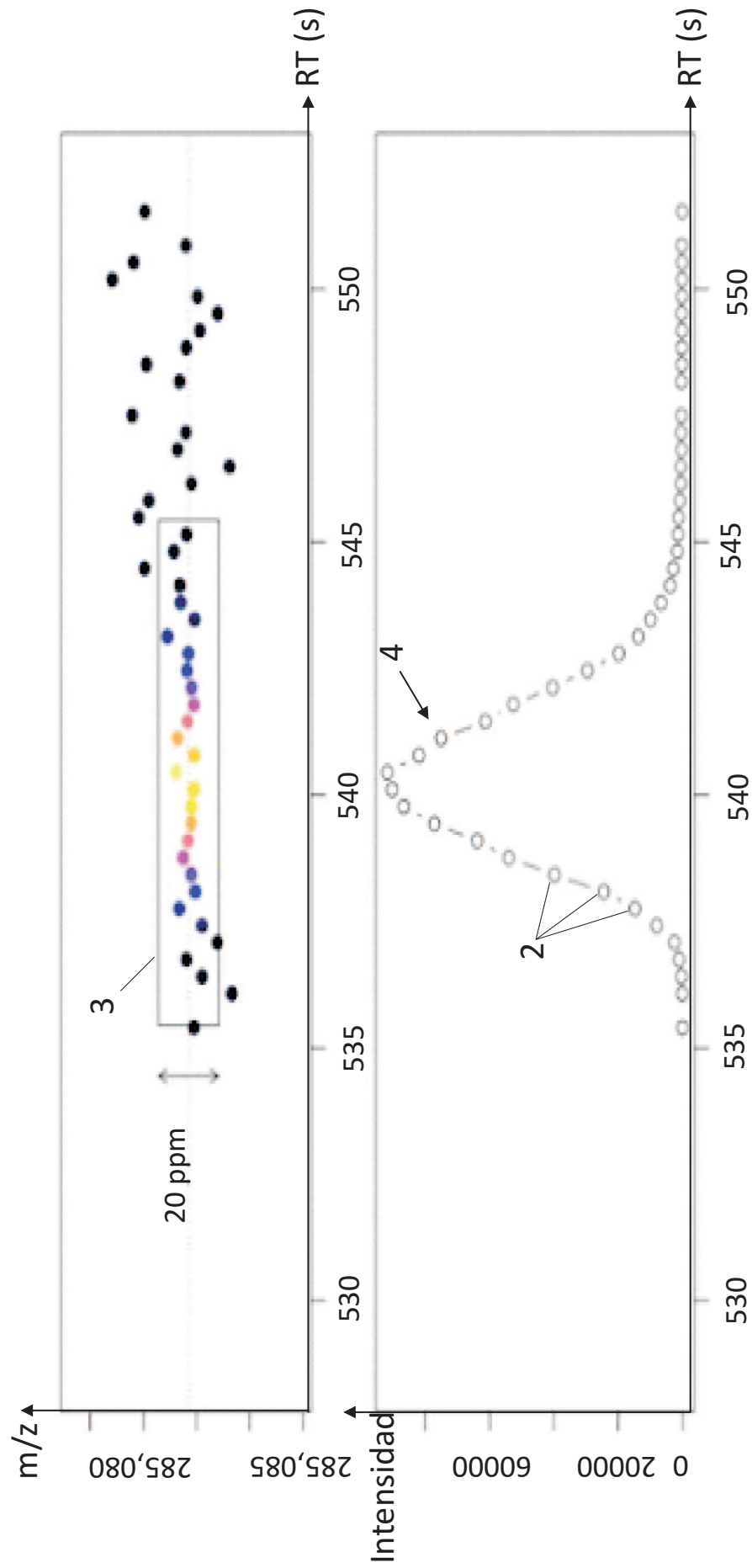


Fig. 2A

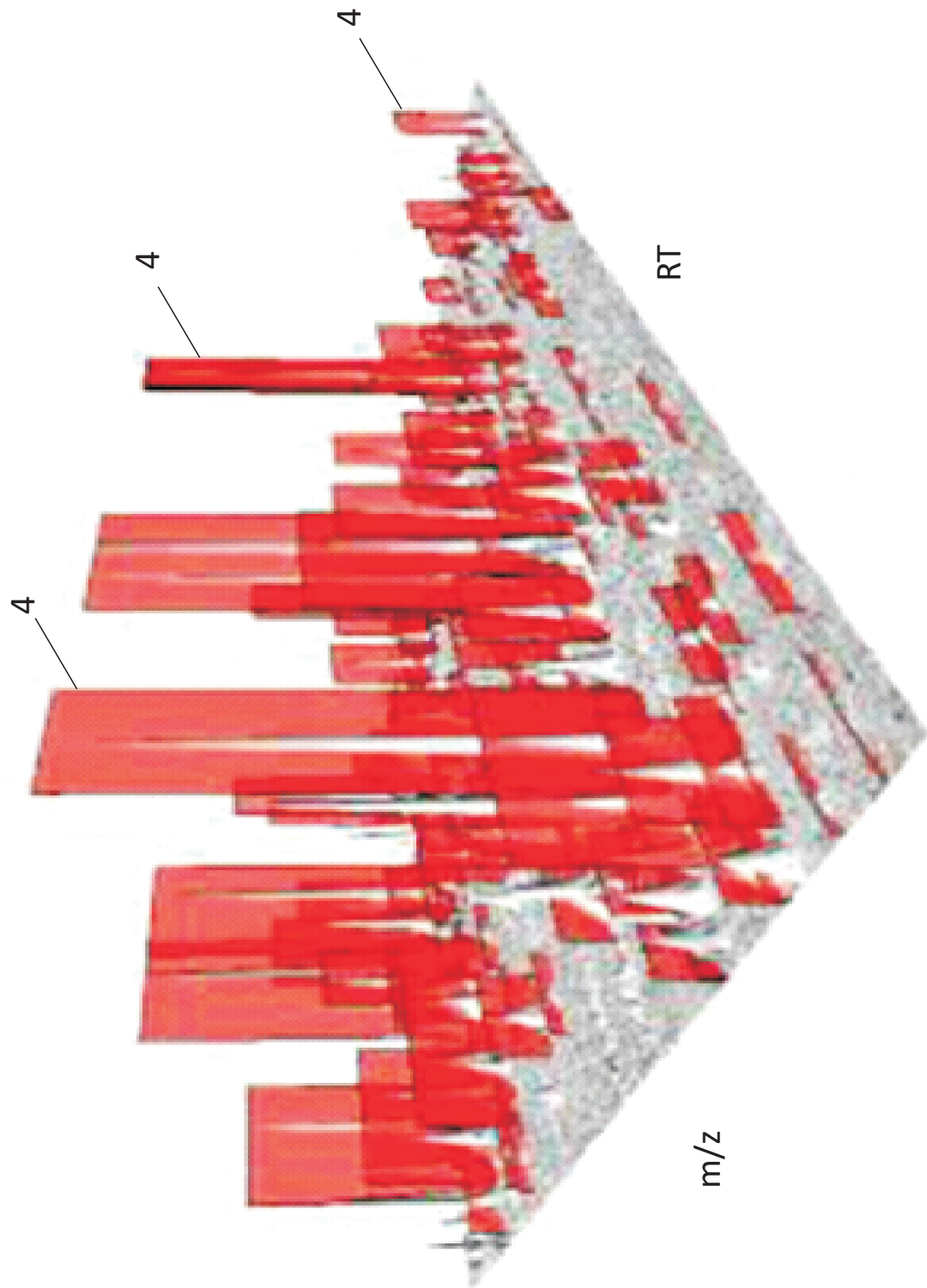


Fig. 2B

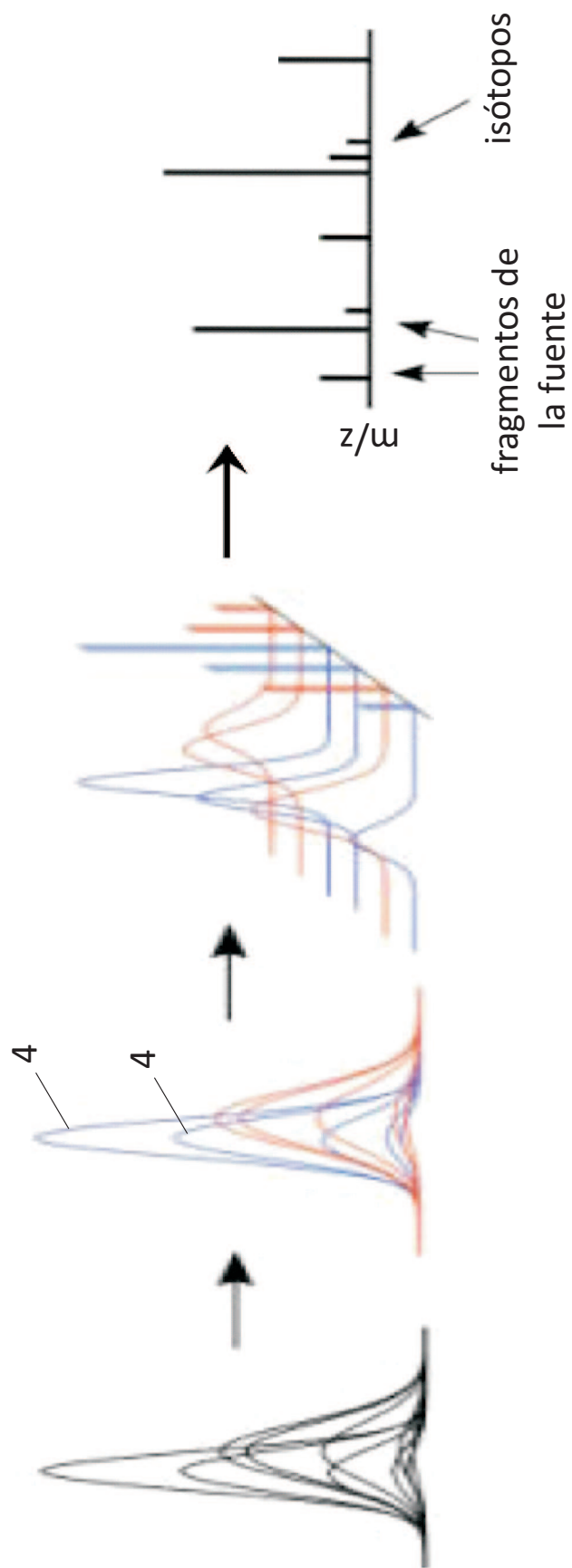


Fig. 3A

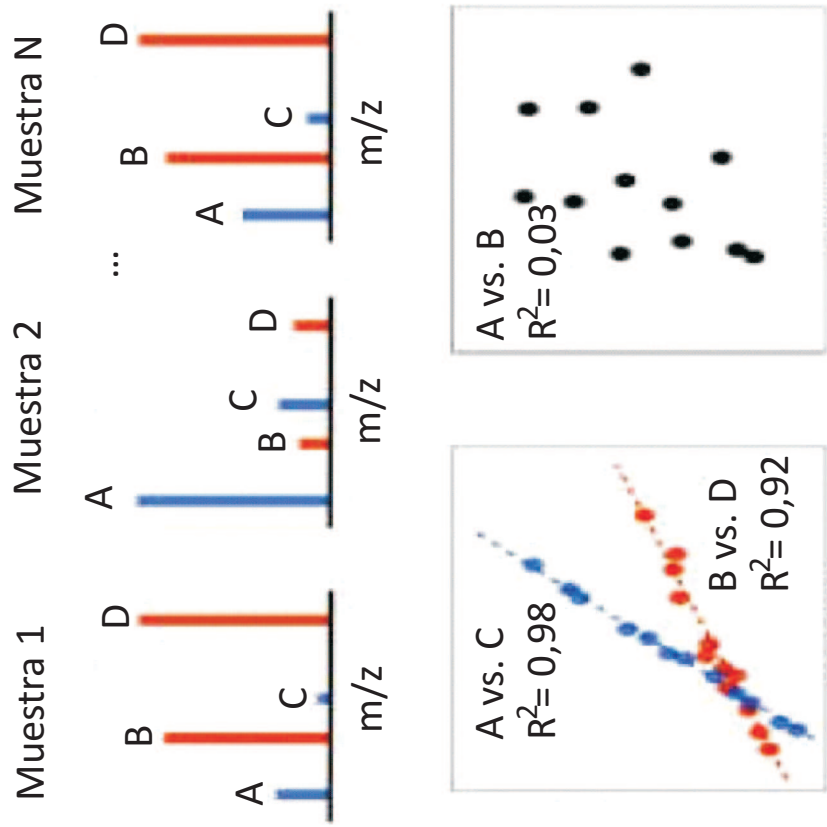


Fig. 3C

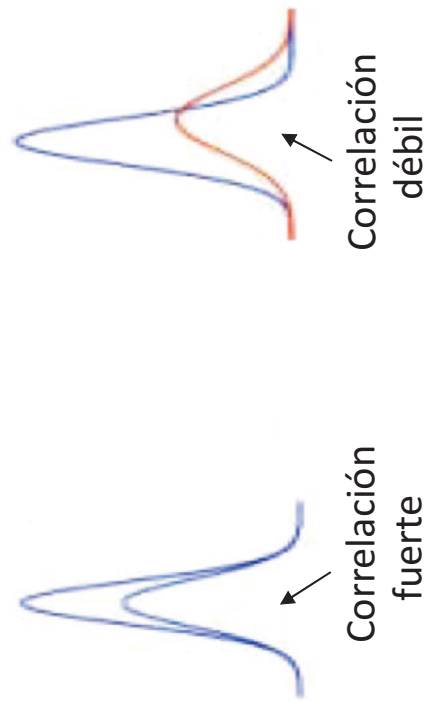


Fig. 3B

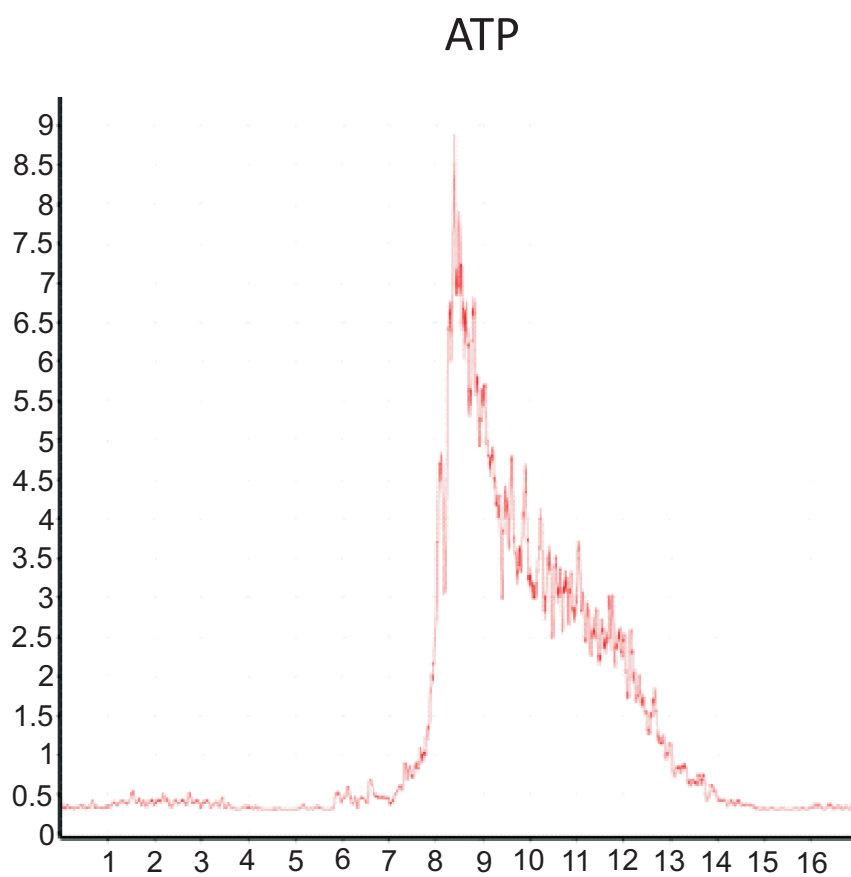


Fig. 4A

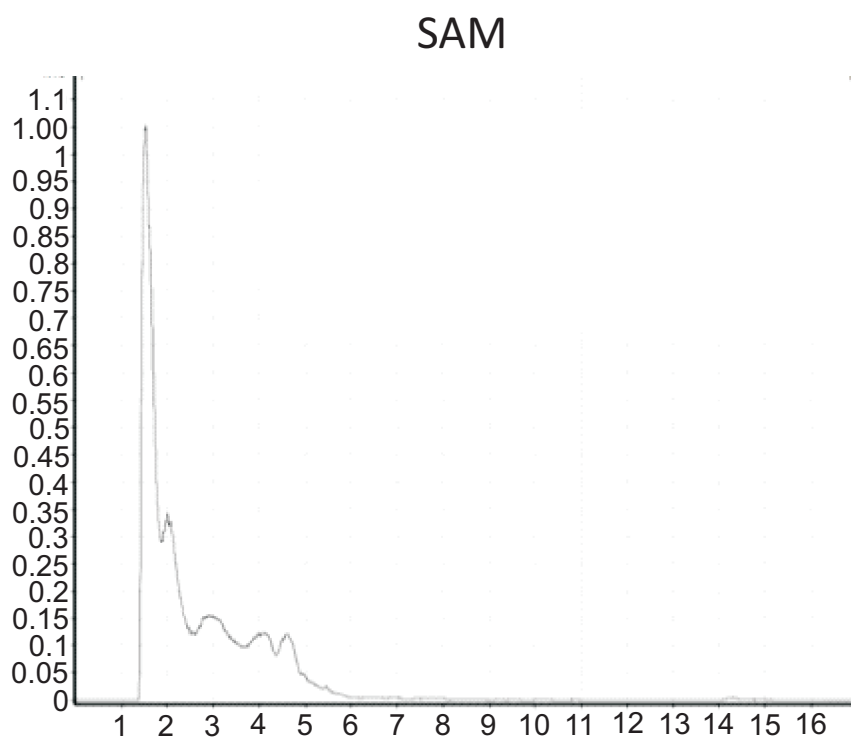


Fig. 4B

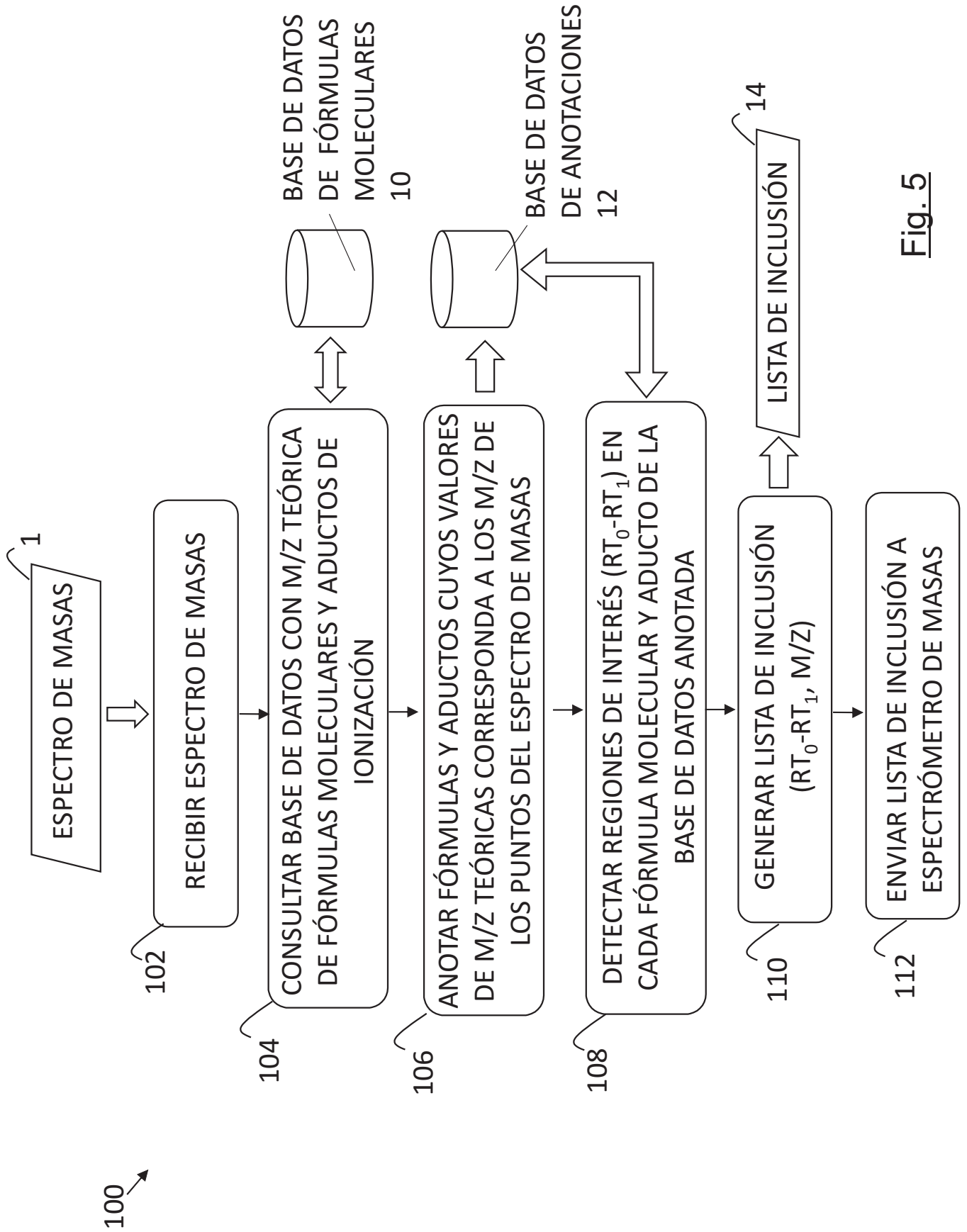


Fig. 5

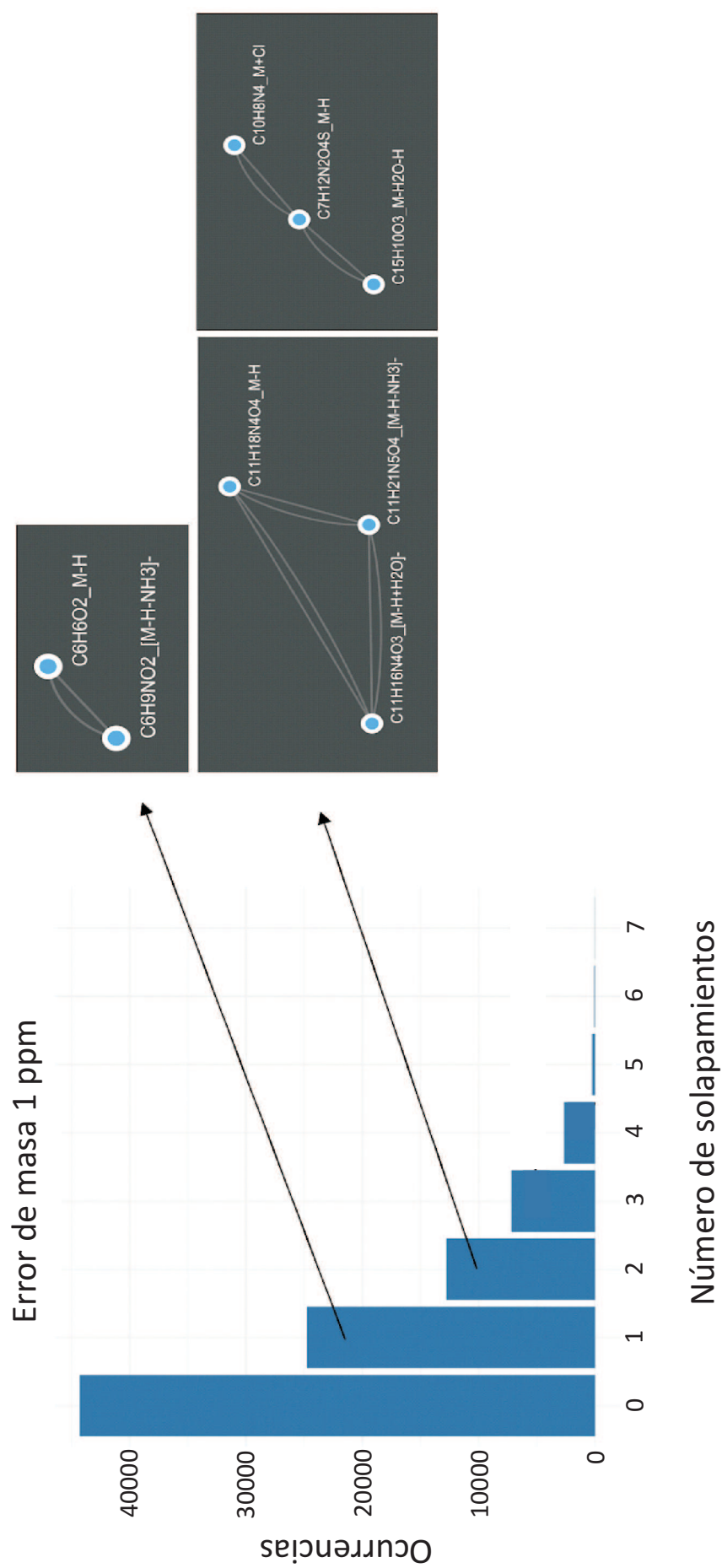


Fig. 6A



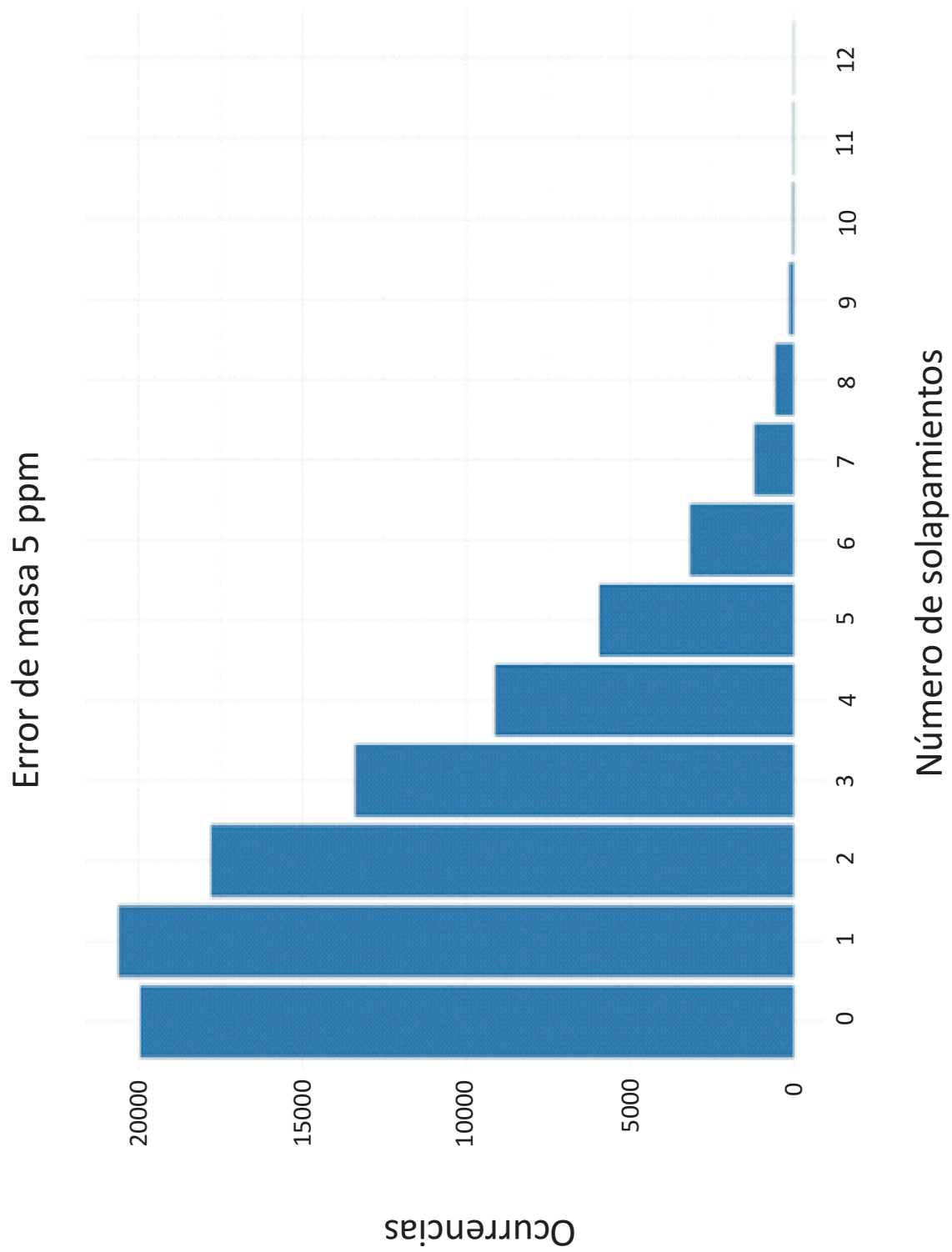


Fig. 6B

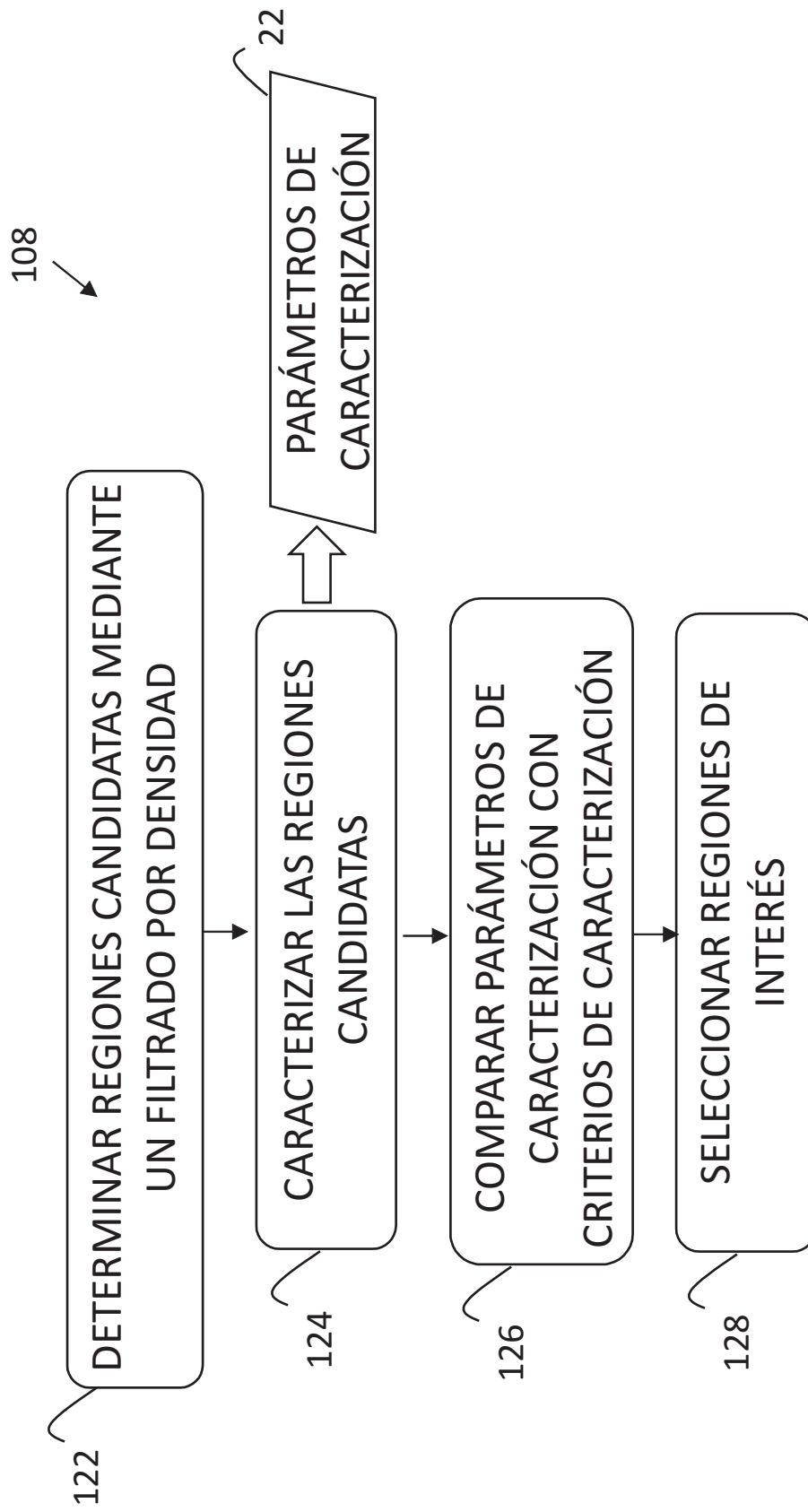


Fig. 7

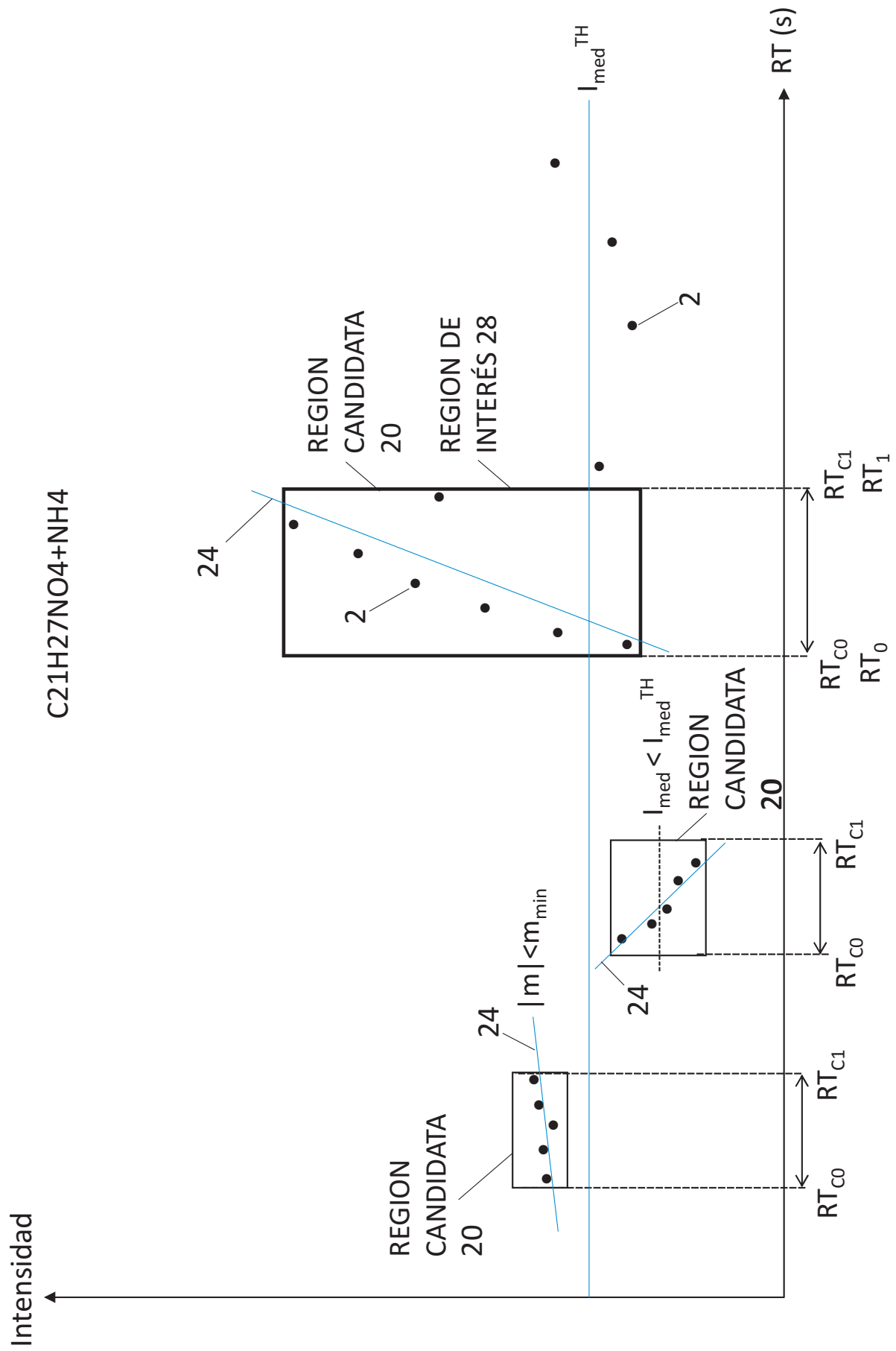
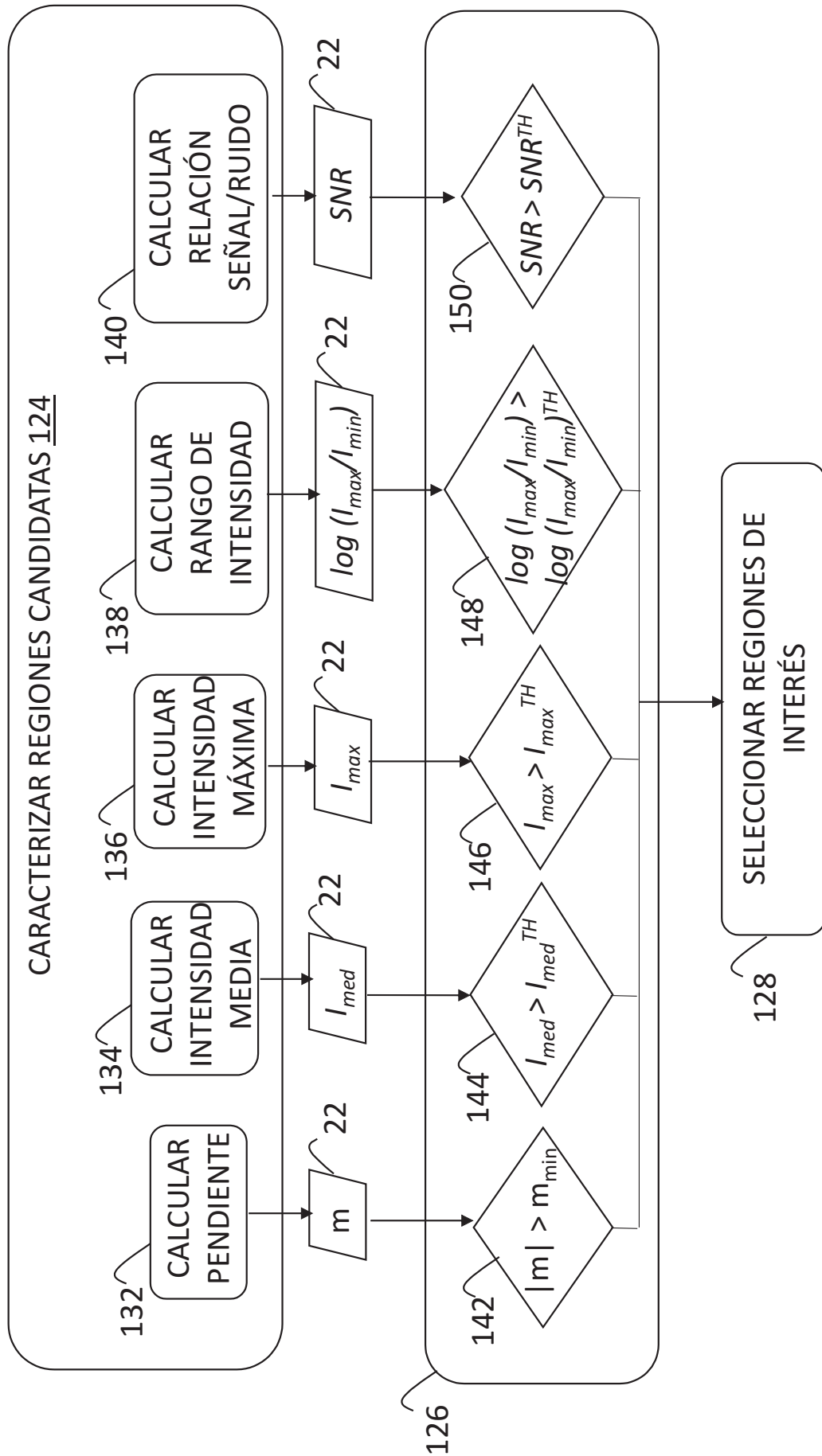


Fig. 8



**Fig. 9**

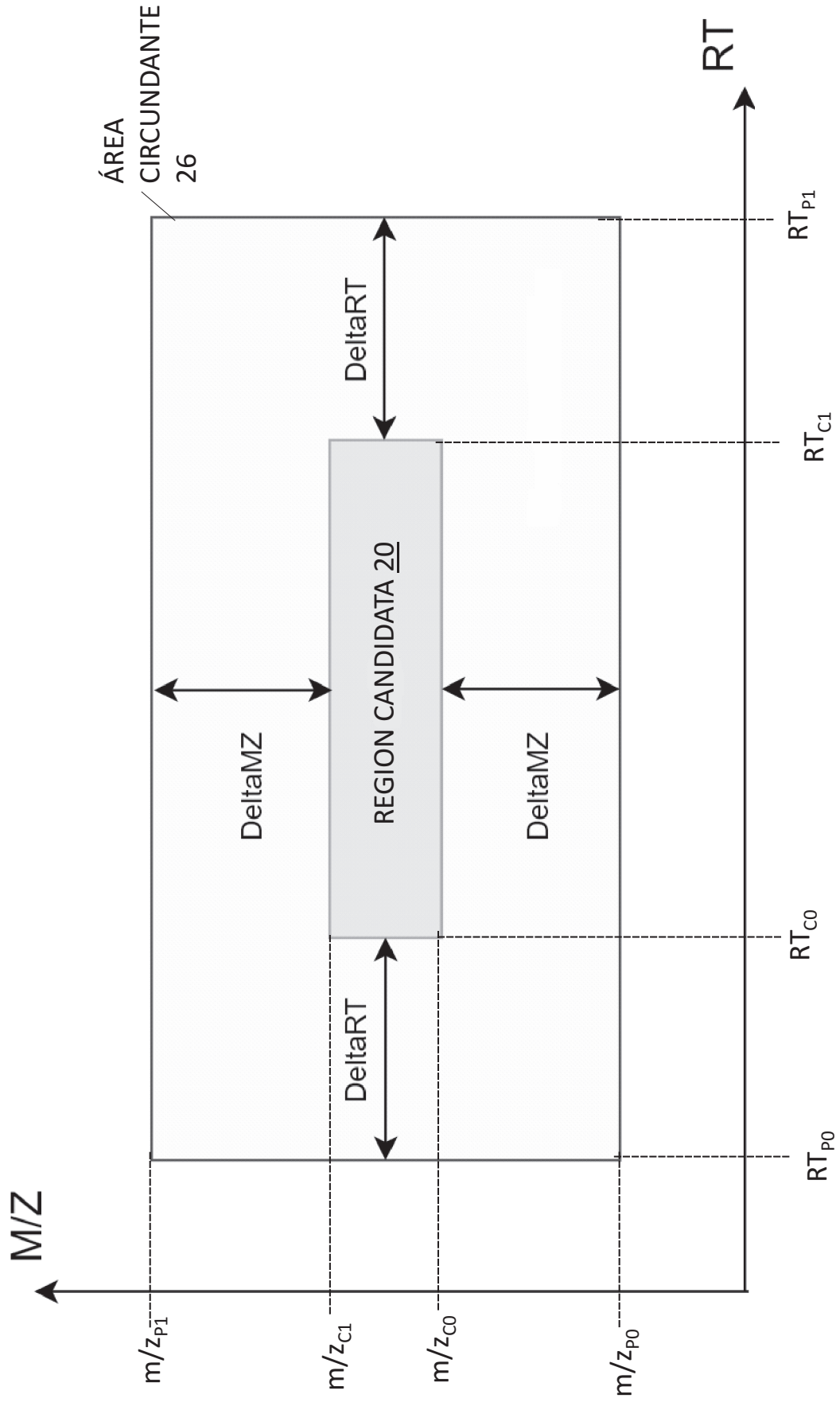


Fig. 10

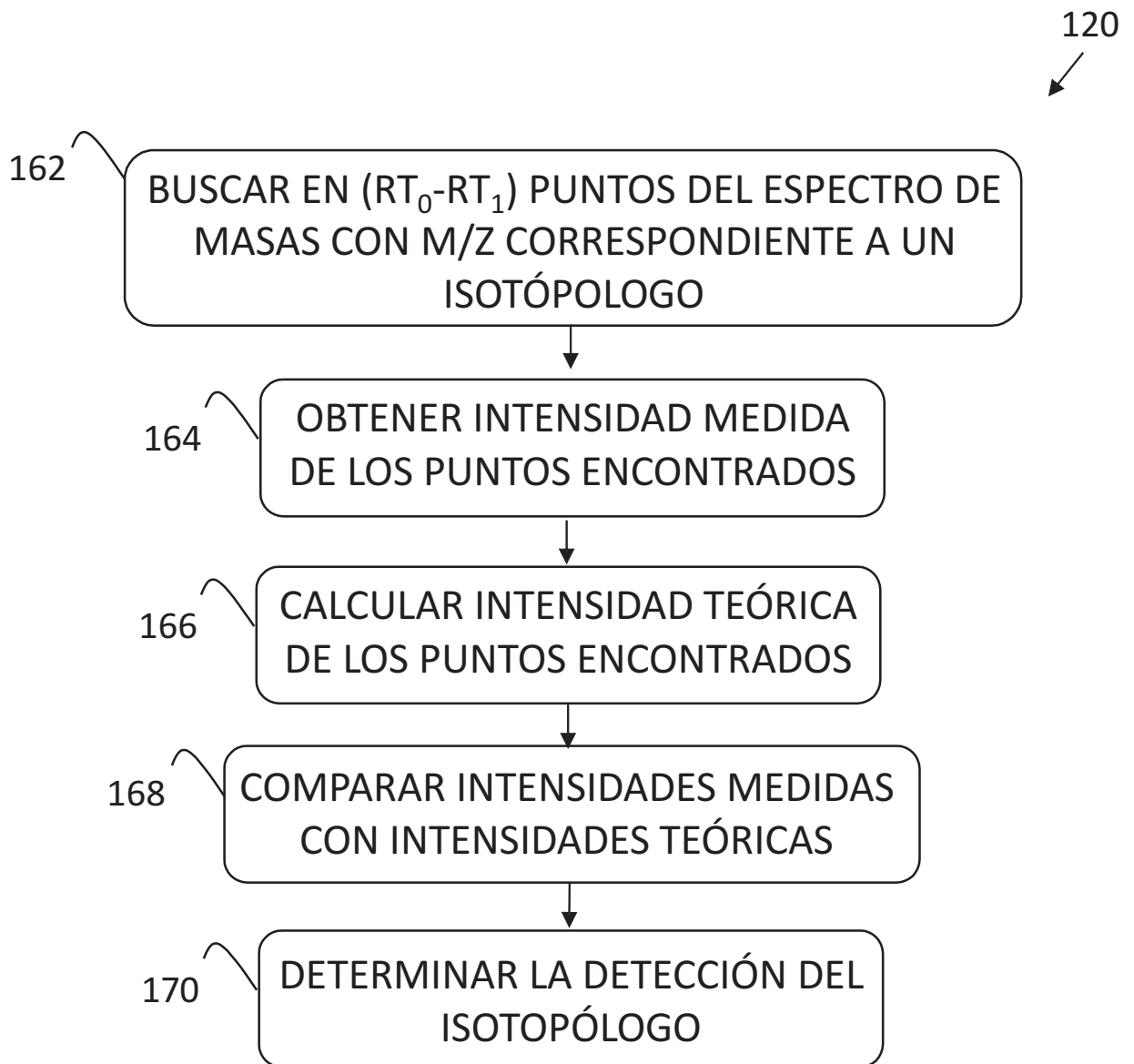


Fig. 11

# Resolución 200000 (tipo Orbitrap)

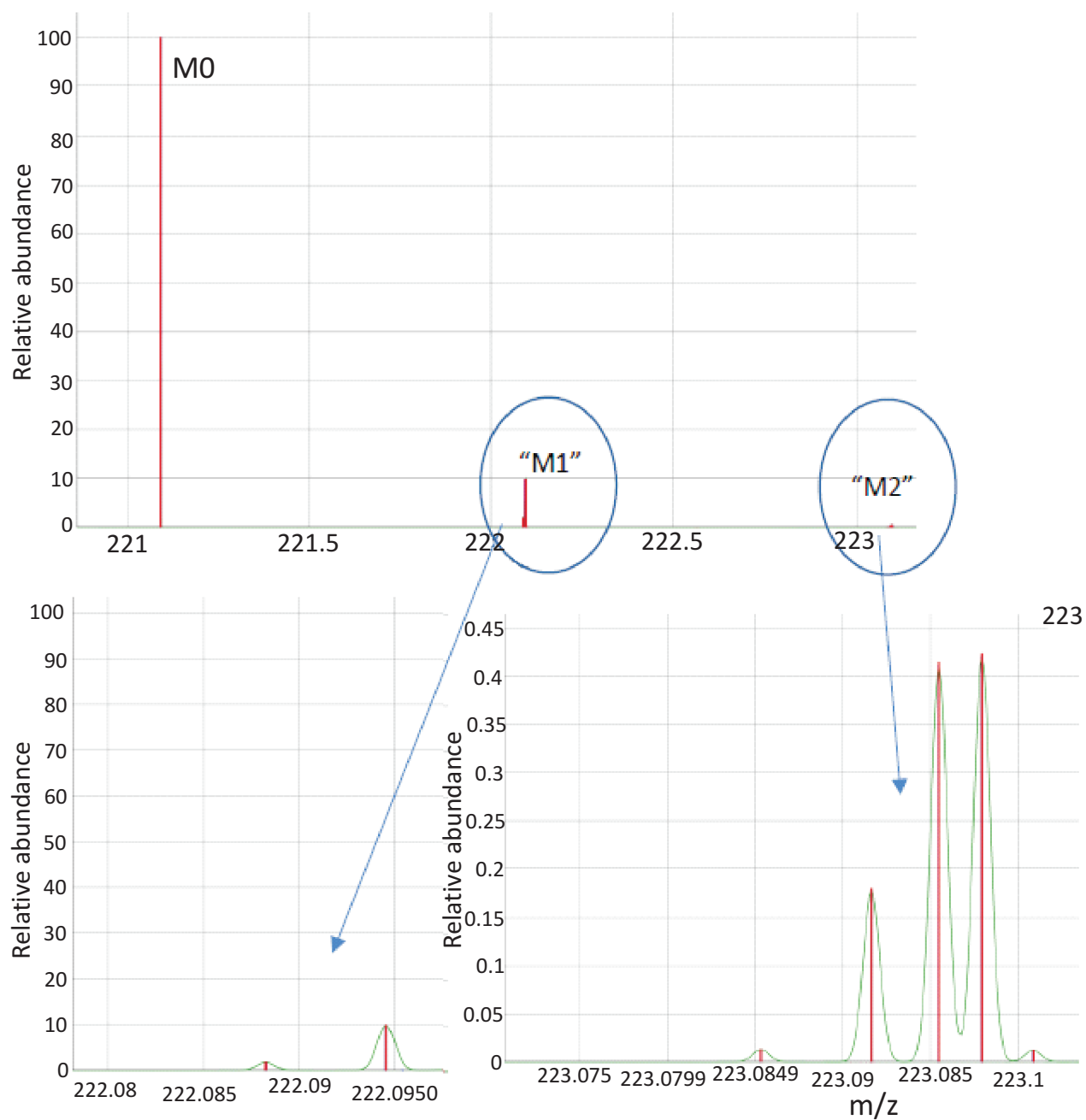


Fig. 12A

# Resolución 60000 (tipo qTOF)

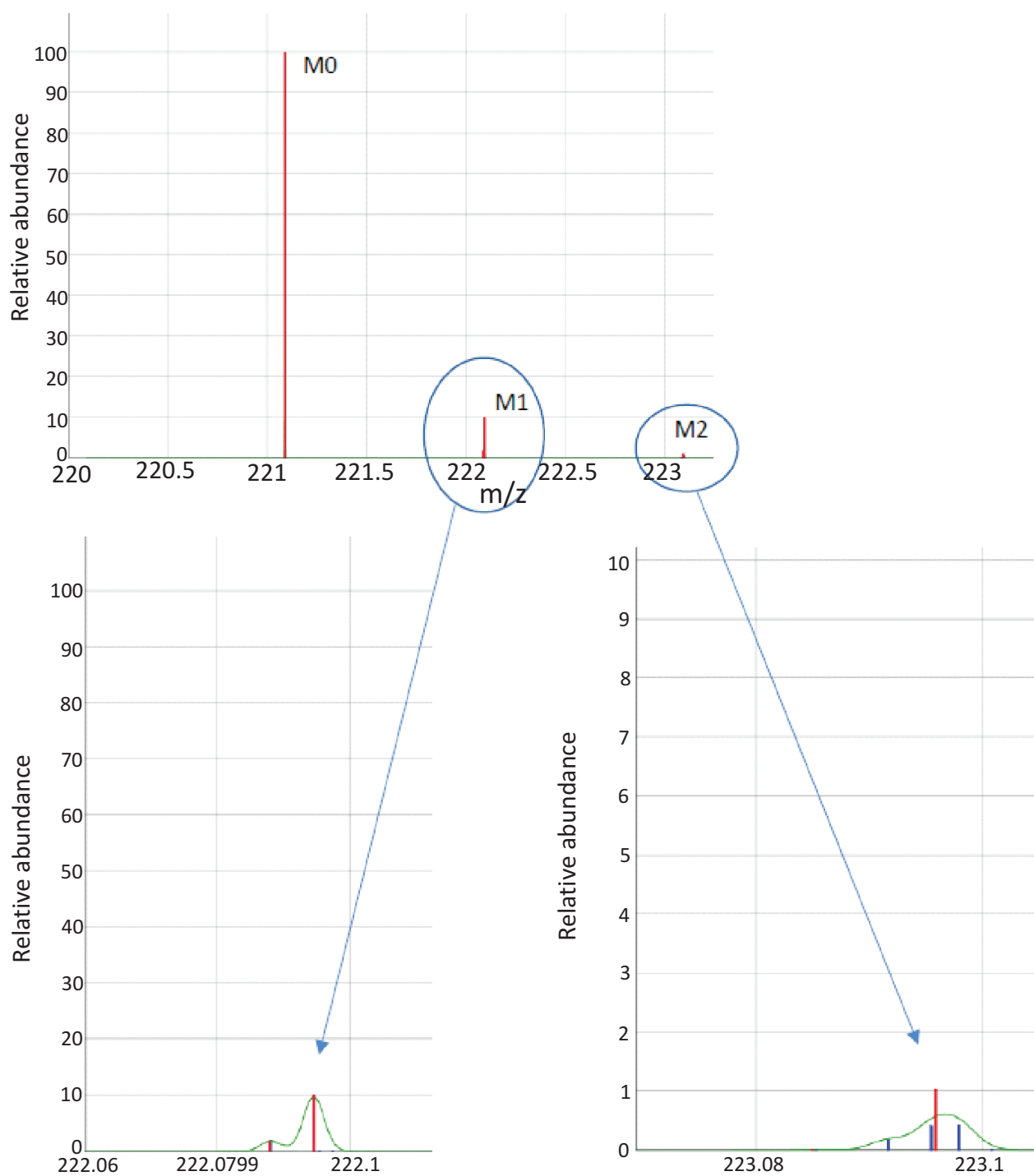
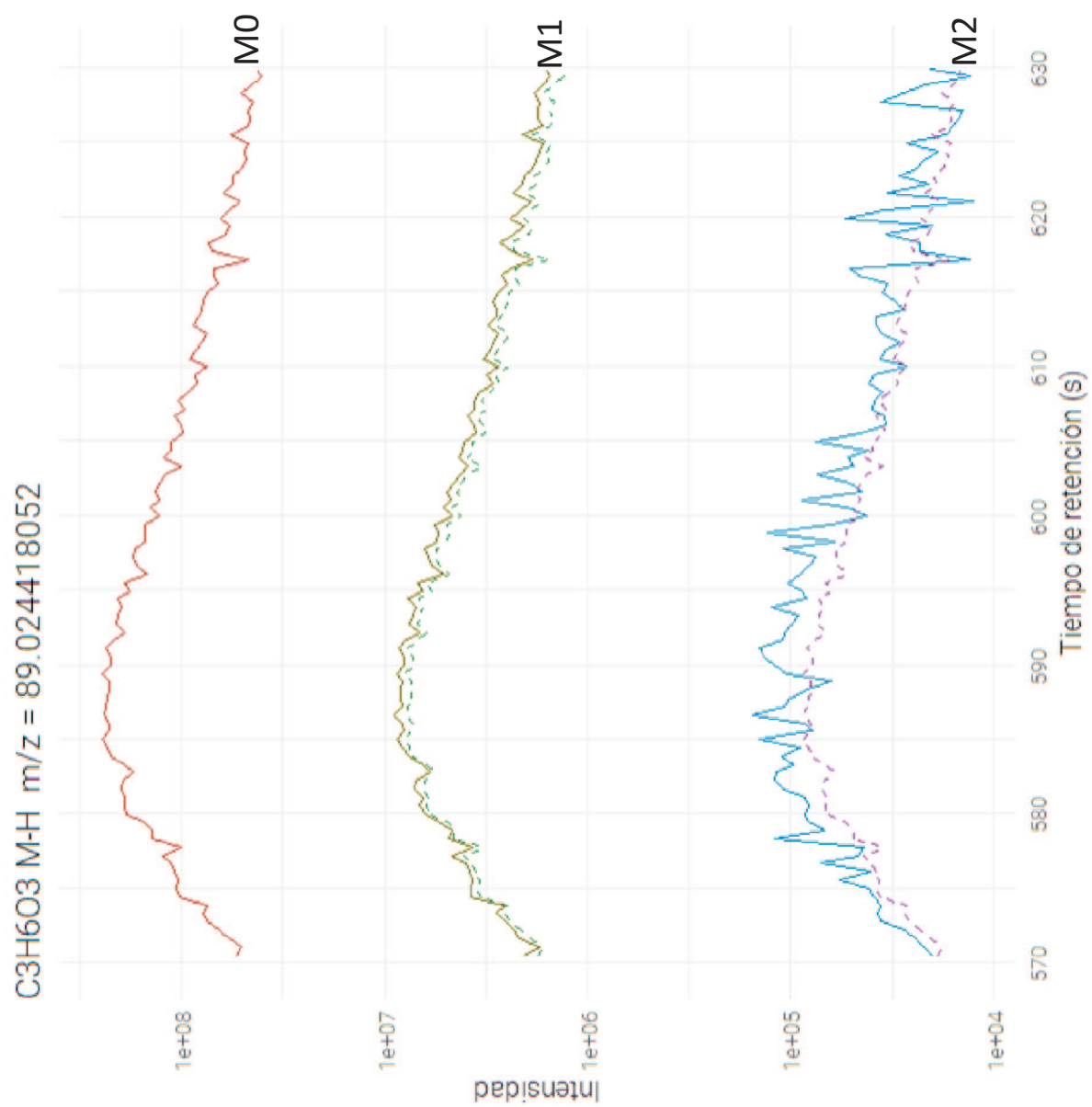


Fig. 12B



Fig. 13



- ②① N.º solicitud: 202030061  
②② Fecha de presentación de la solicitud: 24.01.2020  
②③ Fecha de prioridad:

## INFORME SOBRE EL ESTADO DE LA TECNICA

⑤① Int. Cl.: **G16C20/20** (2019.01)  
**G01N30/72** (2006.01)

### DOCUMENTOS RELEVANTES

Categoría	⑤⑥ Documentos citados	Reivindicaciones afectadas
A	PERIS-DIAZ MANUEL D et al. <b>R-MetaboList 2: A Flexible Tool for Metabolite Annotation from High-Resolution Data-Independent Acquisition Mass Spectrometry Analysis.</b> Metabolites SEP 2019, <b>31/08/2019</b> , Vol. 9, Nº 9, Article No.: 187, ISSN 2218-1989(electronic), <DOI: doi:10.3390/metabo9090187> figura 1, páginas 2 - 6;	1-16
A	SENAN ORIOL et al. <b>CliqueMS: a computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network.</b> Bioinformatics (Oxford) OCT 15 2019, <b>15/10/2019</b> , Vol. 35, Nº 20, Páginas 4089-4097, ISSN 1367-4803(print) ISSN 1460-2059(electronic), <DOI: doi:10.1093/bioinformatics/btz207> página 4090;	1-16
A	EDMANDS WILLIAM M B et al. <b>compMS2Miner: An Automatable Metabolite Identification, Visualization, and Data-Sharing R Package for High-Resolution LC-MS Data Sets.</b> Analytical chemistry United States, <b>04/04/2017</b> , Vol. 89, Nº 7, Páginas 3919 - 3928, ISSN 1520-6882 (Electronic), <DOI: doi:10.1021/acs.analchem.6b02394 pubmed:28225587>. páginas 6 - 8;	1-16
A	TIKUNOV Y M et al. <b>MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data.</b> Metabolomics, 20111015 Kluwer Academic Publishers-Plenum Publishers, NL., <b>15/10/2011</b> , Vol. 8, Nº 4, Páginas 714 - 718, ISSN 1573-3890, <DOI: doi:10.1007/s11306-011-0368-2>. figura 1, páginas 715 - 716;	1-16
A	ALONSO ARNALD et al. <b>AStream: an R package for annotating LC/MS metabolomic data.</b> Bioinformatics (Oxford) MAY 1 2011, <b>30/04/2011</b> , Vol. 27, Nº 9, Páginas 1339-1340, ISSN 1367-4803(print) ISSN 1367-4811(electronic), <DOI: doi:10.1093/bioinformatics/btr138>. páginas 1339 - 1340; figura 1,	1-16
A	TAUTENHAHN RALF et al. <b>Highly sensitive feature detection for high resolution LC/MS.</b> BMC BIOINFORMATICS, 20081128 BIOMED CENTRAL, LONDON, GB, <b>28/11/2008</b> , Vol. 9, Nº 1, Páginas 504, ISSN 1471-2105, <DOI: doi:10.1186/1471-2105-9-504>. figuras 1 - 2. páginas 3 - 8;	1-16
<p>Categoría de los documentos citados X: de particular relevancia Y: de particular relevancia combinado con otro/s de la misma categoría A: refleja el estado de la técnica</p> <p>O: referido a divulgación no escrita P: publicado entre la fecha de prioridad y la de presentación de la solicitud E: documento anterior, pero publicado después de la fecha de presentación de la solicitud</p>		
<p><b>El presente informe ha sido realizado</b> <input checked="" type="checkbox"/> para todas las reivindicaciones</p> <p><input type="checkbox"/> para las reivindicaciones nº:</p>		
<b>Fecha de realización del informe</b> 05.06.2020	<b>Examinador</b> M. López de Rego Lage	<b>Página</b> 1/3



OFICINA ESPAÑOLA  
DE PATENTES Y MARCAS  
ESPAÑA

- ②① N.º solicitud: 202030061  
②② Fecha de presentación de la solicitud: 24.01.2020  
③② Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤① Int. Cl.: **G16C20/20** (2019.01)  
**G01N30/72** (2006.01)

DOCUMENTOS RELEVANTES

Categoría	⑤⑥ Documentos citados	Reivindicaciones afectadas
A	US 2017338089 A1 (TAUTENHAHN RALF et al.) 23/11/2017, reivindicaciones 1-14;	1-16
A	US 2013131998 A1 (WRIGHT DAVID A) 23/05/2013, figura 3, reivindicaciones 1-7;	1-16
A	EP 2418481 A1 (SHIMADZU CORP) 15/02/2012, Figuras 3 - 14. Párrafos [36 - 62].	1-16

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones

☐ para las reivindicaciones nº:

Fecha de realización del informe  
05.06.2020

Examinador  
M. López de Rego Lage

Página  
2/3

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G16C, G01N, G06F

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, BIOSIS, INSPEC, EMBASE, MEDLINE