

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 539 710**

51 Int. Cl.:

**H01J 49/00** (2006.01)

**G06K 9/00** (2006.01)

**G06F 19/00** (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **02.11.2012 E 12191116 (8)**

97 Fecha y número de publicación de la concesión europea: **15.04.2015 EP 2590206**

54 Título: **Método y dispositivo para calcular las distribuciones de isótopos oleculares y para estimar la composición elemental de una molécula a partir de una distribución isotópica**

30 Prioridad:

**02.11.2011 EP 11187499**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**03.07.2015**

73 Titular/es:

**VITO NV (VLAAMSE INSTELLING VOOR  
TECHNOLOGISCH ONDERZOEK NV) (100.0%)  
Boeretang 200  
2400 Mol, BE**

72 Inventor/es:

**VALKENBORG, DIRK y  
DITWALD, PIOTR**

74 Agente/Representante:

**LAZCANO GAINZA, Jesús**

Observaciones :

**Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes**

ES 2 539 710 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

## Descripción

Método y dispositivo para calcular las distribuciones de isótopos moleculares y para estimar la composición elemental de una molécula a partir de una distribución isotópica

5

Campo técnico

La invención se refiere al campo técnico de la identificación de la composición elemental de las moléculas de polipéptidos y, más específicamente, de la identificación y cuantificación de los eventos de fosforilación de las moléculas de polipéptidos.

10

Antecedentes

La espectrometría de masa (MS) es una técnica analítica que se desarrolló en el siglo pasado y que mide la relación de la masa a la carga de las partículas cargadas. Se usa para determinar las masas de las partículas, para determinar la composición elemental de una muestra o molécula, y para elucidar las estructuras químicas de las moléculas, tales como los péptidos y otros compuestos químicos. El principio de la MS consiste en ionizar compuestos químicos para generar moléculas o fragmentos de moléculas cargados y medir sus relaciones de la masa a la carga. En un procedimiento típico de la MS (i) una muestra se carga en el instrumento de MS, y se somete a la vaporización, (ii) los componentes de la muestra se ionizan mediante uno de una variedad de métodos (por ejemplo, impactarlos con un haz de electrones), lo que resulta en la formación de partículas cargadas (iones), (iii) los iones se separan de acuerdo con su relación de la masa a la carga en un analizador mediante campos electromagnéticos, (iv) los iones se detectan, usualmente mediante un método cuantitativo, y (v) la señal del ion se procesa en los espectros de masa. Los instrumentos MS comprenden típicamente tres módulos: (a) una fuente de iones, que puede convertir las moléculas de la muestra en fase gaseosa en iones (o, en el caso de la ionización por electropulverización, puede mover los iones que existen en la solución hacia la fase gaseosa), (b) un analizador de masa, que ordena los iones por sus masas al aplicar campos electromagnéticos, y (c) un detector, que mide el valor de una cantidad indicadora y por lo tanto proporciona los datos para calcular las abundancias de cada ion presente. La técnica de MS tiene usos tanto cualitativos como cuantitativos. Estos incluyen identificar compuestos desconocidos, determinar la composición isotópica de los elementos en una molécula, y determinar la estructura de un compuesto al observar su fragmentación. Otros usos incluyen cuantificar la cantidad de un compuesto en una muestra o estudiar los fundamentos de la química de iones en fase gaseosa (la química de iones y neutros en un vacío). La MS se encuentra ahora en un uso muy común en los laboratorios analíticos que estudian las propiedades físicas, químicas, o biológicas de una gran variedad de compuestos.

15

20

25

30

35

La muestra que se introduce en un espectrómetro de masa puede consistir de una multitud de átomos o moléculas. Durante la etapa de ionización, los átomos y las moléculas pueden requerir una carga y por lo tanto pueden manipularse fácilmente por los campos eléctricos y magnéticos. La partícula cargada se acelera bajo la influencia de estos campos, inversamente proporcional a su relación de la masa a la carga  $m/e$ . Una muestra puede consistir en diferentes moléculas, con diferentes masas. Un espectrómetro de masa puede, por lo tanto, separar las moléculas que tienen masas que difieren por el ancho de la resolución de masa del espectrómetro de masa. Un resultado típico de un análisis de MS es un espectro de masa que muestra picos en, o alrededor de, ciertos valores para la masa, las alturas de los picos en ciertos valores de masa que son proporcionales a la cantidad de moléculas con esa masa que estaban presentes en la muestra. La muestra puede comprender además las mismas moléculas que se componen de átomos que tienen diferentes isótopos. Las propiedades electroquímicas de estas moléculas son las mismas, pero sus masas difieren en una cantidad que está próxima a un número entero de unidades de masa atómica (amu) o daltons (Da). Los diferentes isótopos de una molécula pueden separarse además en un espectrómetro de masa, ya que tienen una relación de la masa a la carga diferente. Puesto que una molécula puede comprender una multitud de átomos cuyos isótopos ocurren con sus abundancias elementales, cada molécula tendrá una distribución isotópica que en principio consiste en un conjunto de grupos de picos cuyas alturas dependen del contenido de átomos y de las abundancias de isótopos atómicos, estos grupos que se separan esencialmente en 1 Da. Debe ampliarse que la distribución isotópica de una molécula es una característica distintiva de esa molécula: cada molécula con un contenido atómico especificado, es decir, con una fórmula molecular especificada, tiene su distribución isotópica característica. Los diferentes isómeros de una molécula tienen, para todos los propósitos prácticos, la misma distribución isotópica. Cuando en un espectro de masa de una muestra, por ejemplo, obtenido a partir de un análisis de MS, se observa un conjunto de picos cuyas alturas relativas corresponden a las alturas relativas de la distribución isotópica de una molécula, puede deducirse la presencia de esta molécula en la muestra original. Las alturas absolutas de los picos son una medida cuantitativa de la cantidad de moléculas presentes en la muestra.

40

45

50

55

60

La distribución isotópica de una molécula puede calcularse a partir de la fórmula molecular y las bien conocidas abundancias elementales y masas de los átomos constituyentes. Las alturas de los picos en una distribución isotópica representan la probabilidad con la que los diferentes isótopos moleculares ocurren en la naturaleza. Como un ejemplo de una distribución isotópica de una molécula, o la huella isotópica de una molécula, puede mirarse al monóxido de

## ES 2 539 710 T3

5 carbono o CO. El carbono (C) tiene dos isótopos estables de origen natural,  $^{12}\text{C}$  y  $^{13}\text{C}$  con abundancias de 0.9893 y 0.0107 y con masas de 12 Da y 13.0033548378 Da respectivamente. El oxígeno (O) tiene tres isótopos estables de origen natural,  $^{16}\text{O}$ ,  $^{17}\text{O}$  y  $^{18}\text{O}$  con abundancias de 0.99757,  $3.8 \times 10^{-4}$  y  $2.05 \times 10^{-3}$  y con masas de 15.99491461956 Da, 16.99913170 Da y 17.9991610 Da respectivamente. La distribución isotópica comprende 6 picos distribuidos en 4 grupos. El primer pico es un pico monoisotópico cerca de la masa de 27.995 Da con una altura de 0.986896001. Esta altura representa la probabilidad con la que el isótopo molecular de CO con un átomo  $^{12}\text{C}$  y un átomo  $^{16}\text{O}$  ocurre en una amplia muestra de moléculas de CO. El segundo y el tercer pico de la distribución isotópica de CO se agrupan alrededor de 29 Da, cerca de las masas de 28.998 Da para  $^{13}\text{C}^{16}\text{O}$  y 28.999 Da para  $^{12}\text{C}^{17}\text{O}$  con las alturas de 0.010673999 y 0.000375934 respectivamente. El cuarto y el quinto pico se agrupan alrededor de 30 Da cerca de las masas de 29.999 Da para  $^{12}\text{C}^{18}\text{O}$  y 30.002 Da para  $^{13}\text{C}^{17}\text{O}$  con las alturas de  $0.002028065$  y  $4.06600 \times 10^{-6}$  respectivamente. El sexto y último pico es de nuevo un pico monoisotópico cerca de la masa de 31.003 Da con la altura de  $2.1935 \times 10^{-5}$ . Los números se resumen en la siguiente tabla:

15	Número de masa total	Masa	Abundancia
	28	27.995	0.986896001
	29	28.998	0.010673999
20	29	28.999	0.000375934
	30	29.999	0.002028065
	30	30.002	0.000004066
25	31	31.003	0.000021935

De esta tabla, es evidente que el grupo de picos corresponde al número de masa total, es decir, el número de nucleones presentes en la molécula. Obviamente, la masa depende principalmente de los isótopos específicos de los átomos que componen la molécula. También se desprende de este ejemplo que los espectrómetros de masa con una resolución de masa del orden de 0.002 amu o mayor, no serán capaces de distinguir todas las diferentes variantes isotópicas dentro de un grupo. En lugar de ello, tales espectrómetros de masa mostrarán un espectro de masa con un pico ampliado para cada grupo cuya superficie debajo del pico será proporcional a las probabilidades sumadas de las variantes isotópicas dentro de ese grupo, y que se centrará alrededor de una masa central que es el promedio ponderado de las masas de las variantes isotópicas constituyentes para un grupo, los factores de ponderación que son proporcionales a las abundancias relativas o probabilidades. Este tipo de distribución isotópica agrupada se llama la distribución isotópica agregada. La distribución isotópica agregada de una molécula consiste en un número de picos, cada pico que corresponde a un grupo de las variantes isotópicas de esa molécula con el mismo número de masa total, cada pico que se localiza en una masa central que es la masa promedio de las variantes isotópicas que contribuyen al grupo correspondiente ponderado por sus abundancias correspondientes, y la altura de cada uno de los picos que corresponde a la suma de las abundancias de las variantes isotópicas que contribuyen al grupo correspondiente.

Aunque el acceso a la espectrometría de masa (MS) de alta resolución, especialmente en el campo de la MS biomolecular, se hace disponible cada vez más fácilmente debido a los recientes avances en la tecnología de la MS, la información que acompaña a la distribución isotópica en los espectros de alta resolución no se usa en todo su potencial, principalmente debido a la falta de conocimiento y/o conciencia. Una de las principales dificultades al usar la MS en la MS biomolecular, es que la distribución isotópica de los polipéptidos, que pueden consistir de cientos de aminoácidos y miles de átomos, es muy difícil de calcular, y mucho menos de reconocer en un espectro de MS. Dicho cálculo puede requerir tiempos de cálculo que están más allá de las capacidades actuales cuando se usa un simple método combinatorio. Además, las técnicas anteriores para identificar una molécula a partir de una distribución isotópica pueden incluir típicamente una técnica de ensayo y error o de ajuste que requiere una multitud de tales cálculos.

El documento US7904253B2 describe un método para determinar la composición elemental de iones a partir de los datos del espectro de masa, que comprende las etapas de:

- obtener al menos una medición exacta de la masa a partir de los datos del espectro de masa;
- obtener una lista de búsqueda de composiciones elementales candidatas cuyas masas exactas caigan dentro de un intervalo dado de tolerancia de masa a partir de dicha masa exacta;
- informar una medida de probabilidad basada en un error de masa;
- calcular un patrón de isótopos para cada composición elemental candidata a partir de dicha lista de búsqueda;
- construir una matriz de componentes pico que incluya al menos uno de dicho patrón de isótopos y datos del espectro de masa;

- realizar una regresión en función de al menos uno de, el patrón de isótopos, los datos del espectro de masa, y la matriz de componentes pico;
- informar una segunda medida de probabilidad para al menos una composición elemental candidata basado en dicha regresión del patrón de isótopos; y
- 5 • combinar dichas dos medidas de probabilidad en una medida de probabilidad global a través del uso de multiplicaciones de probabilidad.

Una etapa limitante en este método es la etapa donde debe calcularse un patrón de isótopos para cada composición elemental candidata de la lista de búsqueda. Para las moléculas muy grandes, por ejemplo, del orden de 10000 Da, esta puede ser una etapa tediosa y consumidora de tiempo, incluso con la tecnología de computación actual. Alternativamente, el patrón de isótopos puede obtenerse a partir de una gran base de datos que contenga los patrones de isótopos relevantes. Sin embargo, dicha base de datos puede ser más grande que la que puede almacenarse en, por ejemplo, la memoria de una computadora, especialmente si las moléculas son grandes, o dicha base de datos puede estar incompleta. Además, dicha base de datos tendrá que calcularse al menos una vez. La presente invención proporciona un método para analizar al menos una parte de una distribución isotópica de una muestra mediante el uso, entre otros, de un método correcto y eficiente para calcular un patrón isotópico de un ion, por ejemplo, para su uso en el análisis de MS o para su almacenamiento en una base de datos. Más específicamente, la presente invención ofrece un método para analizar al menos una parte de una distribución isotópica mediante el uso de un método rápido y eficiente para calcular la distribución isotópica agregada de una molécula. Esto puede hacerse de manera recursiva. El método comienza con el cálculo de la masa central y de la probabilidad de una variante isotópica agregada de partida de una molécula que, por ejemplo, puede esperarse que esté presente en la muestra. Preferentemente, esta variante isotópica agregada de partida es una variante monoisotópica, por ejemplo, la variante isotópica más ligera, lo que significa que la masa central es simplemente la masa de la variante isotópica, y la probabilidad es el producto de las abundancias elementales de los átomos constituyentes. A partir del conocimiento de la probabilidad de, por ejemplo, la variante isotópica agregada más ligera, puede calcularse la probabilidad de una siguiente variante isotópica agregada con número de masa total diferente del número de masa total de la variante isotópica agregada de partida, por ejemplo, la segunda más ligera. A partir del conocimiento de las probabilidades de las variantes isotópicas agregadas de partida y siguiente, puede calcularse la probabilidad de una tercera variante isotópica agregada, etc. Además, la presente invención ofrece además la posibilidad de calcular la masa central de cada variante isotópica agregada de una molécula de manera recursiva similar.

Los documentos Rockwood '95 (Rockwood, Alan L., Rapid Commun. Mass Spectrom. 9:103-105, 1995) y Rockwood '96 (Rockwood, Alan L., Van Orden, Steven L., Anal. Chem. 68:2027-2030, 1996) describen un método para calcular la distribución isotópica agregada de una molécula al modelar el problema en términos de las transformadas de Fourier. Debido a que las transformadas discretas de Fourier pueden calcularse de manera muy eficiente, esta manera de ver el problema tiene implicaciones prácticas significativas. Específicamente, los documentos describen un algoritmo de ultraalta velocidad para calcular las distribuciones de isótopos a partir de fórmulas moleculares, masas isotópicas elementales y abundancias isotópicas elementales. Para un conjunto dado de datos de entrada (fórmula molecular, masas isotópicas elementales y abundancias isotópicas elementales), y bajo la suposición de un error de redondeo despreciable, el algoritmo produce rigurosamente distribuciones de isótopos cuyas medias y desviaciones estándar son "correctas" en el sentido de que un algoritmo libre de errores produciría una distribución que tiene la misma media y desviación estándar. Las alturas de los picos son también "correctas" en el sentido de que la altura de cada pico de isótopo nominal a partir del cálculo de ultraalta velocidad es igual a la del área integrada del pico de isótopo nominal correspondiente a partir de un cálculo exacto. Como consecuencia de estas propiedades, el algoritmo generalmente coloca los picos de isótopo dentro de milidaltos de sus masas centrales o centroides exactos. El método usa los métodos de la transformada de Fourier.

Aunque el método introducido en Rockwood '95 y '96 es rápido para calcular las distribuciones isotópicas y las distribuciones isotópicas agregadas, aún puede ser computacionalmente intensivo. Además, este método no puede invertirse directamente, es decir, cuando se presenta una distribución isotópica o una distribución isotópica agregada de una molécula, la técnica de la transformada de Fourier no es capaz de deducir la fórmula molecular directamente: sólo puede deducirla mediante una técnica de ensayo y error o de ajuste tal como la del documento US7904253B2 descrito en la presente anteriormente. El objetivo de esta invención es proporcionar un método que sea menos intensivo computacionalmente que el método de la transformada de Fourier. Además, el objetivo de esta invención es proporcionar un método que sea invertible directamente, es decir, que pueda deducir la fórmula molecular directamente a partir de una distribución isotópica agregada sin tener que recurrir a un método de ensayo y error complicado computacionalmente y lento.

Los métodos de la técnica anterior para calcular la distribución isotópica agregada de una molécula no son rápidos o no son suficientemente exactos, se necesita mucha memoria, se calculan demasiados detalles, pueden ocurrir problemas computacionales tales como el desbordamiento numérico e inexactitudes numéricas, etc., y esto especialmente para las moléculas grandes tales como los polipéptidos o los oligonucleótidos (ADN, ARN).

Aun existe en la técnica la necesidad de un método mejorado para analizar al menos una parte de una distribución isotópica de una muestra mediante el cálculo de la distribución agregada de isótopos de una molécula de manera más eficiente, estable, computacionalmente menos intensiva, y de un método mejorado para analizar al menos una parte de una distribución isotópica de una muestra mediante la que la masa central de una variante isotópica agregada se calcule de una manera mejorada. La presente invención proporciona tales métodos, mediante los cuales los problemas de la técnica anterior se superan debido a un método para calcular la probabilidad de la variante isotópica agregada a partir de las probabilidades de las variantes isotópicas agregadas calculadas o conocidas previamente. Además, las masas centrales se calculan en este mismo sentido y no implican una gran cantidad de tiempo de cálculo adicional.

En "An Efficient Method to Calculate the Aggregated Isotopic Distribution and Exact Center-Masses", J. Claesen, P. Dittwald, T. Burzykowski, D. Valkenburg, J. Am. Soc. Mass Spectrom. (2012), se presenta un método de cálculo y que usa eficientemente la memoria, con el cual pueden calcularse las probabilidades de ocurrencia y las masas centrales exactas de la distribución isotópica agregada de una molécula. El cálculo se basa en la composición atómica de la molécula y las abundancias naturales de los isótopos elementales en la materia terrestre normal. Para las moléculas grandes la forma isotópica puede obtenerse al expandir simbólicamente una función polinomial, que generalmente puede escribirse como:

$$Q(I; v, w, x, y, z) \equiv \sum_{j=0}^n q_j I^j$$

Donde  $n = v+w+x+2y+4z$  es una función de la composición atómica de la molécula. El coeficiente  $q_j$  representa la probabilidad de ocurrencia de la  $j$ -ésima variante isotópica agregada de la molécula y se expresa como:

$$q_j = \frac{1}{j} \sum_{i=1}^j q_i - 1 \cdot I^i$$

La complejidad de los cálculos depende principalmente del número de elementos químicos diferentes presentes en la molécula. No depende del número de átomos de cada elemento presente en la molécula, sino del número de las variantes isotópicas agregadas para las que se requieren cálculos. En la práctica los cálculos podrían detenerse cuando el valor de  $q_j$  caiga por debajo de un umbral particular o cuando se cubra un porcentaje preestablecido de la distribución isotópica.

Este método puede usarse para calcular la masa central  $m_j$  de las variantes isotópicas agregadas de la siguiente manera:

$$m_j = \frac{\sum_k m_{jk} p_{jk}}{\sum_k p_{jk}}$$

donde  $p_{jk}$  y  $m_{jk}$  denotan, respectivamente, la probabilidad de ocurrencia y la masa de la  $k$ -ésima variante isotópica que contribuye a la  $j$ -ésima variante agregada.

La distribución isotópica se calculó para la angiotensina II y la insulina bovina considerando todas las posibles variantes isotópicas, aunque mediante el uso de una implementación de la expansión multinomial. Para las moléculas grandes, las probabilidades de ocurrencia de las variantes monoisotópicas y las (varias) variantes isotópicas agregadas consecutivas pueden ser muy pequeñas. El método descrito anteriormente permite calcular las masas centrales exactas de las variantes agregadas. Proporciona los valores correctos de las probabilidades de ocurrencia de las diferentes variantes isotópicas agregadas y las masas centrales.

La presente invención proporciona un método para identificar la composición elemental de una molécula en una muestra al invertir las etapas para calcular una distribución isotópica agregada de una molécula, sin la necesidad de una técnica de ensayo y error o de ajuste, como se usa comúnmente en la técnica anterior. La presente invención proporciona además un método para identificar y cuantificar la presencia de elementos en una molécula, que no alteran la distribución isotópica de la molécula aunque contribuyen a la masa de la molécula. El fósforo es un elemento de este tipo.

La presente invención proporciona así un método para estimar la composición elemental basado en la distribución isotópica agregada observada en un espectro de masa, al invertir la transformación descrita en "An Efficient Method to Calculate the Aggregated Isotopic Distribution and Exact Center-Masses". La invención reivindica que las herramientas anteriores se han usado para detectar, por ejemplo, los péptidos fosforilados, o cualquier elemento monoisotópico en una molécula. Esto se logra al estimar la composición elemental a partir de la distribución isotópica agregada observada de una molécula. La composición elemental estimada se usa para calcular la masa monoisotópica de la molécula observada. La última etapa es una adición de las masas elementales monoisotópicas. La masa monoisotópica calculada basada en la composición elemental estimada puede compararse con la masa monoisotópica observada en el espectro de masa, por ejemplo, si la diferencia de masas es aproximadamente 31 Da, puede concluirse que el patrón de isótopos observado se origina a partir de una molécula fosforilada.

Las técnicas anteriores para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra pueden implicar etapas experimentales adicionales costosas y consumidoras de tiempo, tales como las realizadas en la MS en tándem. Estas etapas experimentales adicionales conducen a mayores tiempos de medición, costos adicionales, equipos más grandes, etc. Una técnica de ensayo y error o de ajuste es inefectiva en tiempo, especialmente para las moléculas grandes, para identificar la molécula a partir de una distribución de isótopos. La patente de Estados Unidos US7904253, por ejemplo, describe una técnica típica de ensayo y error para identificar la fórmula química de una molécula cuya distribución isotópica se ha medido en una pequeña ventana de masa.

La presente invención pretende resolver al menos algunos de los problemas mencionados anteriormente.

Aun existe en la técnica la necesidad de un método mejorado para estimar la información sobre la composición de una molécula a partir de una distribución isotópica. La presente invención proporciona dicho método al invertir el método mencionado anteriormente. Invertir el método permite la estimación de la fórmula química de una molécula, es decir, el número de átomos de cada especie a partir de una distribución isotópica. Además, el presente método no necesita conocer las masas absolutas, pero puede tratar con una parte de una distribución isotópica que muestra sólo los picos de las variantes isotópicas agregadas de una molécula, es decir, dadas las alturas de un conjunto de picos separados esencialmente en 1 Da, el presente método puede calcular la información sobre la composición estimada de la molécula responsable de estos picos cuando se trate de elementos atómicos que son no monoisotópicos en la naturaleza. Para identificar los elementos monoisotópicos en la fórmula química de una molécula, puede usarse el valor de la menor masa de la molécula, necesariamente monoisotópica, derivado de la información sobre la composición estimada de la molécula y comparar este con el valor medido para la masa del primer pico. La diferencia entre la masa observada en el espectro de masa y la masa calculada mediante el uso de la información sobre la composición estimada da la masa sumada de todos los elementos monoisotópicos presentes en la molécula que se observa.

La identificación es especialmente difícil para las moléculas grandes. Las proteínas y los polipéptidos son moléculas grandes de ese tipo, por ejemplo, en el intervalo de 50000 Da. Una modificación fundamental de los polipéptidos que regulan muchas actividades y funciones celulares es la fosforilación de los polipéptidos. La presente invención proporciona un método para identificar y detectar la presencia y cantidad de un polipéptido fosforilado a partir de un análisis espectrométrico de masa, basado en los métodos debatidos anteriormente. El fósforo P es un elemento monoisotópico, es decir, que sólo tiene un isótopo que es estable o al menos lo suficientemente estable como para encontrarse en sustancias de origen natural.

Resumen de la invención

La presente invención proporciona un método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, de manera que dicha molécula tiene una fórmula química  $\prod_{\alpha}(Z_{\alpha})n_{\alpha}$  con

- $\alpha$ , un índice que recorre un número de especies atómicas constituyentes esperadas de dicha molécula;
  - $Z_{\alpha}$ , la  $\alpha$ -ésima especie atómica esperada de la molécula; y
  - $n_{\alpha}$ , el número de átomos de la especie  $Z_{\alpha}$  en la molécula,
- que comprende las etapas de:
- obtener al menos una parte de una distribución isotópica a partir de dicha muestra;
  - obtener un conjunto de  $N_p$  alturas de los picos  $t_j$  con  $j = 1 \dots N_p$ , de al menos una parte de una distribución isotópica agregada a partir de dicha distribución isotópica;
  - calcular a partir de dicho conjunto de alturas de los picos un conjunto de  $N_p - 1$  alturas relativas de los picos  $\hat{t}_i$  con  $i = 1 \dots N_p - 1$ ;
  - obtener un valor de  $n_{\alpha}$  para al menos una especie atómica  $Z_{\alpha}$  a partir de dicho conjunto de alturas relativas de los picos,

caracterizado porque dicho valor de  $n_{\alpha}$  se obtiene al calcular una solución de un sistema de ecuaciones lineales  $\sum_{\alpha} E_{i\alpha} n_{\alpha} = F_i$ , preferentemente, esta solución puede redondearse a los enteros más próximos, de manera que el conjunto de números  $F_i$  comprende dicho conjunto de alturas relativas de los picos y los coeficientes  $E_{i\alpha}$  de dicho sistema lineal comprenden las potencias y/o las sumas de las potencias de las raíces  $r_{\alpha, i\alpha}$  con  $i_{\alpha} = 1 \dots N_{\alpha}$ , de una ecuación polinomial elemental de  $N_{\alpha}$ -ésimo orden cuyo coeficiente del término de  $m_{\alpha}$ -ésimo orden se determina por una abundancia

elemental  $P\left(\begin{smallmatrix} A_{\alpha} + m_{\alpha} \\ Z_{\alpha} \end{smallmatrix} Z_{\alpha}\right)$  del isótopo de la especie de átomo  $Z_{\alpha}$  con  $m_{\alpha}$  neutrones más que el isótopo estable más ligero de dicha especie de átomo que tiene número másico  $A_{\alpha}$ , de manera que, las masas en las que se observan las alturas de los picos  $t_j$  se separan esencialmente en 1 Da, de manera que dicho conjunto de  $N_p - 1$  alturas relativas de los picos  $\hat{t}_i$  con  $i = 1 \dots N_p - 1$  se calcula recursivamente a partir de dicho conjunto de  $N_p$  alturas de los picos  $t_j$  con  $j = 1 \dots N_p$ , mediante el uso de las etapas de:

- calcular  $\hat{t}_1 = -\frac{t_2}{t_1}$
- calcular  $\hat{t}_{i+1} = -\frac{1}{t_1} [(i+1)t_{i+2} + \sum_{j=1}^i \hat{t}_{i-j+1} t_{j+1}]$  para  $i = 2 \dots N_p - 1$ .

5 De esta manera, no es necesario que  $t_1$  se refiera a la altura del pico del isótopo más ligero. Debe quedar claro que cualquier altura del pico puede usarse como referencia, así como también cualquier combinación lineal de las alturas de los picos. En el caso de que se use otra altura del pico o combinación lineal de las alturas de los picos, los coeficientes  $E_{i\alpha}$  pueden adaptarse en consecuencia para reflejar el cambio en la altura del pico de referencia.

10 Los términos "alturas relativas de los picos" y "relaciones de las alturas de los picos", aquí y en todo este texto, se refieren a números adimensionales, que pueden expresarse como una relación entre las alturas de los picos. Debe quedar claro que el valor absoluto de la altura de un pico, tomado o derivado a partir de un espectro isotópico, en sí mismo, no contiene información sobre el número de especies atómicas específicas presentes dentro de una molécula. Para obtener información sobre este número, y por lo tanto sobre la fórmula química de la molécula, las cantidades pertinentes deben expresarse en función de las relaciones entre las alturas de los picos. De esta manera, puede  
15 hacerse una elección en cuanto a cuál altura del pico se usa como una referencia.

En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describió anteriormente, dicho conjunto de alturas de los picos  $t_j$  se ordena de acuerdo con la masa en la que se observa la altura de un pico. En una modalidad preferida, no se  
20 observa ningún pico en dicha distribución isotópica o dicha distribución isotópica agregada alrededor de una masa que es esencialmente 1 Da menor que la masa en la que se observa el primero de dicho conjunto de alturas de los picos  $t_j$ , es decir  $t_1$ .

En una modalidad preferida, dichos coeficientes  $E_{i\alpha}$  de dicho sistema lineal se determinan por la suma de la (-i)-ésima potencia de las  $N_\alpha$  raíces de la ecuación polinomial elemental de  $N_\alpha$ -ésimo orden cuyo coeficiente del término de  $m_\alpha$ -  
25 ésimo orden se determina por la abundancia elemental  $P(A_\alpha + m_\alpha Z_\alpha)$  del isótopo de la especie atómica  $Z_\alpha$  con  $m_\alpha$  neutrones más que el isótopo estable más ligero de dicha especie atómica que tiene un número másico  $A_\alpha$ , es decir  
 $E_{i\alpha} = \sum_{j=1}^{N_\alpha} r_{\alpha,j}^{-i}$  y de manera que dicho conjunto de números  $F_i$  se da por dicho conjunto de alturas relativas de los picos  $\hat{t}_i$ , es decir  $F_i = \hat{t}_i$ .

En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describió anteriormente, dicho número  $N_p - 1$  de alturas  
relativas de los picos es al menos dicho número de especies atómicas constituyentes esperadas de dicha molécula con dos o más isótopos estables. En una modalidad con mayor preferencia, dicho número  $N_p - 1$  de alturas relativas de los  
35 picos es al menos 5, es decir, dicho número  $N_p$  de alturas de los picos es al menos 6, y/o dicha molécula comprende como máximo 5 elementos atómicos, tales como C, H, N, O, S.

En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describió anteriormente, dicho número  $N_p - 1$  de alturas  
40 relativas de los picos es mayor que el número de especies atómicas constituyentes esperadas de dicha molécula con dos o más isótopos estables y los valores de  $n_\alpha$  para cada valor de  $\alpha$  se obtienen al resolver un sistema sobredeterminado de ecuaciones lineales  $\sum_\alpha E_{i\alpha} n_\alpha = F_i$ . Esta solución puede obtenerse mediante técnicas anteriores, tales como un algoritmo de resolución de mínimos cuadrados y un posterior redondeo de las soluciones resultantes a los enteros más próximos. La siguiente lista no limitativa proporciona aún otras técnicas que pueden usarse para  
45 resolver el sistema de ecuaciones mencionado anteriormente:

Métodos clásicos:

- Newton
- Levenberg-Marquardt
- 50 Cuasi-Newton
- Gradiente conjugado
- Funciones suaves no cuadráticas
- Métodos del punto interior
- Métodos del conjunto activo

55 Optimización discreta:

- Programación lineal con enteros
- Programación cuadrática
- Programación dinámica

Estadística robusta:

- 60 Mínimos cuadrados ponderados iterativamente

- 5 En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describió anteriormente, dicha parte de una distribución isotópica se proporciona mediante el análisis de una medición espectrométrica de la masa.
- 10 En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describió anteriormente, dicha muestra comprende moléculas de polipéptidos, moléculas de lípidos y/o de glicoles.
- 15 En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describió anteriormente, dichas especies atómicas esperadas comprenden carbono, hidrógeno, nitrógeno, oxígeno y/o azufre.
- 20 En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describió anteriormente, dicha molécula comprende al menos un átomo constituyente monoisotópico en su fórmula química, y comprende además las etapas de:
- obtener a partir de dicha distribución isotópica una masa medida, dentro de una precisión de  $\pm 0.5$  Da, en la que aparece un pico de isótopo;
  - calcular una masa esperada en la que aparece dicho pico de isótopo para una segunda molécula con una segunda fórmula química que es dicha fórmula química a partir de la cual se elimina el átomo constituyente monoisotópico;
  - comparar dicha masa medida con dicha masa esperada para deducir la presencia de dicha molécula que comprende al menos un átomo constituyente monoisotópico en dicha muestra.
- 25 En una modalidad preferida, al menos dicho átomo constituyente monoisotópico es el fósforo.
- 30 En un aspecto adicional, la presente invención proporciona un dispositivo capaz de identificar la presencia, la composición elemental y/o la cantidad de una molécula en una muestra mediante el uso de cualquiera de los métodos para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describió anteriormente.
- 35 En aún un aspecto adicional, la presente invención proporciona un dispositivo instalable en computadora que comprende una implementación de cualquiera de los métodos para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describió anteriormente.
- 40 En aún un aspecto adicional, la presente invención proporciona el uso de un método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describe en este documento, preferentemente, en una computadora y/o un espectrómetro de masa y/o un instrumento periférico de un espectrómetro de masa.
- Descripción detallada de la invención
- 45 A menos que se defina de cualquier otra manera, todos los términos usados en la descripción de la invención, que incluyen los términos técnicos y científicos, tienen el significado que se entiende comúnmente por un experto en la técnica a la que pertenece esta invención. Por medio de una orientación adicional, las definiciones de los términos se incluyen para apreciar mejor la enseñanza de la presente invención.
- 50 Como se usa en la presente, los siguientes términos tienen los siguientes significados:  
 "Un", "una" y "el/la" como se usa en la presente, se refieren tanto a los referentes en singular como en plural, a menos que el contexto indique claramente de otra manera. A manera de ejemplo, "un compartimento" se refiere a uno o a más de un compartimento.
- 55 "Aproximadamente", como se usa en la presente con referencia a un valor medible tal como un parámetro, una cantidad, una duración temporal, y similares, se entiende que abarca las variaciones de  $\pm 20$  % o menos, preferentemente,  $\pm 10$  % o menos, con mayor preferencia  $\pm 5$  % o menos, aún con mayor preferencia  $\pm 1$  % o menos, y aún con mayor preferencia  $\pm 0.1$  % o menos de, y a partir de, el valor especificado, en la medida en que tales variaciones sean adecuadas para llevarse a cabo en la invención descrita. Sin embargo, debe entenderse que el valor al que el modificador "aproximadamente" se refiere, en sí mismo, también se describe específicamente.
- 60 "Comprenden", "que comprende", y "comprende" y "que se comprende de" como se usa en la presente son sinónimos

de "incluyen", "que incluye", "incluye" o "contienen", "que contiene", "contiene", y son términos inclusivos o abiertos que especifican la presencia de lo que sigue, por ejemplo, un componente, y no excluyen o impiden la presencia de componentes, características, elementos, miembros, etapas, adicionales no mencionados, conocidos en la técnica o descritos en la misma.

5

La mención de intervalos numéricos mediante los puntos extremos incluye todos los números y fracciones subsumidas dentro de ese intervalo, así como también los puntos extremos mencionados.

10

La expresión "% en peso" (por ciento en peso), aquí y en toda la descripción a menos que se defina de otra manera, se refiere al peso relativo del componente respectivo basado en el peso total de la formulación.

A continuación, se dará una explicación de cómo el método de la presente invención para calcular una distribución isotópica agregada de una molécula puede derivarse de los principios básicos.

15

La distribución isotópica es un concepto importante, pero frecuentemente olvidado, en el campo de la espectrometría de masa. Es particularmente útil para la interpretación de los patrones complejos observados en los datos espectrales de masa. Pueden encontrarse aplicaciones tanto en la proteómica como en la metabolómica. Por ejemplo, una molécula de un péptido visualizada mediante MS debería exhibir una señal característica en la forma de una serie de picos regularmente espaciados. La envoltura de la serie de picos muestra un perfil específico relacionado con la distribución isotópica del péptido. El conocimiento previo acerca de la distribución isotópica puede usarse para deconvolucionar los patrones de los péptidos y para reducir la información espectral (ver, por ejemplo, D. Valkenborg y otros, *Rapid Communications In Mass Spectrometry* 21:3387-3391, 2007 y D. Valkenborg y otros, *Journal of the American Society for Mass Spectrometry* 19(5):703-712, 2008). Otra aplicación puede encontrarse en el campo de la metabolómica. Por ejemplo, comparar la distribución isotópica observada en un espectro de masa contra un conjunto de distribuciones isotópicas hipotéticas a partir de partes de moléculas con una masa similar a la molécula observada puede aumentar la confianza en la identificación.

20

25

La razón de que una molécula aparezca en un espectro de masa como diferentes picos se origina a partir del hecho de que una molécula puede existir como diferentes variantes con una masa diferente. Se denotarán estas variantes como variantes de isótopos o isotópicas. Por lo tanto, una molécula puede tener diferentes variantes de isótopos que dependen del número de diferentes isótopos elementales incorporados en su composición atómica. Las probabilidades de ocurrencia de estas variantes de isótopos se reflejan en el espectro de masa, debido a que la espectrometría de masa mide una población de iones de la molécula particular en lugar de un solo ion. En este documento, se presenta un método computacional y que usa eficientemente la memoria para calcular la probabilidad de ocurrencia y las masas centrales exactas de la distribución isotópica de las moléculas. El cálculo se basa en la composición atómica de la molécula y las abundancias elementales naturales de los isótopos elementales en la materia terrestre normal (ver, por ejemplo, en K.J.R. Rosman y P.D.P. Taylor, "Isotopic composition of the elements 1997", *Pure and Applied Chemistry* 70(1):217-235, 1988). Ya se han propuesto varios métodos para realizar los cálculos de la distribución isotópica. Estos métodos se han propuesto durante un intervalo de tiempo desde principios de los años sesenta, donde Biemann sugirió el procedimiento en etapas (K. Bieman, "Mass spectrometry, organic chemical application", McGraw-Hill, New York, 1962) durante finales de los años setenta donde Yamamoto y McCloskey (*Analytical Chemistry* 49:281, 1977) y Brownawell y Fillippo (*Journal of Chemical Education* 59(8):663-665, 1982) sostuvieron que, para las moléculas grandes, la distribución isotópica podría obtenerse fácilmente al expandir simbólicamente una función polinomial. Más tarde, en los años ochenta Yergey y otros (*International Journal of Mass Spectrometry and Ion Physics* 52:337-349, 1983 y *Analytical Chemistry* 55:353-356, 1983) generalizaron la expansión polinomial a una expansión multinomial. Una interesante visión general de los diferentes procedimientos para calcular las distribuciones isotópicas por Valkenborg y otros (*Mass Spectrometry Review*, 2011, DOI 10.1002/mas.20339) está disponible para una mayor comprensión de esta técnica.

35

40

45

50

55

En los años noventa, Rockwood y colaboradores propagaron el uso de la convolución para calcular la distribución isotópica (*Rapid Communication in Mass Spectrometry* 9:103-105, 1995). En su presentación de la expansión simbólica insertaron un nuevo indicador simbólico que les permitió relacionar la expansión polinomial a la convolución. También se usará esta notación, pero se cambiará la interpretación del indicador de manera que simbolice el contenido adicional de neutrones debido a la incorporación de isótopos elementales. Esto implica que las variantes de isótopos con el mismo contenido adicional de neutrones, pero con masas ligeramente diferentes (<50 ppm), se compilan en una variante de isótopos agregada. Hacerlo así presenta la información como una lista de masas (m, M) de las variantes de isótopos agregadas y sus probabilidades de ocurrencia correspondientes (p, q), similar en espíritu como en el método de Rockwood y otros (*Analytical Chemistry* 68:2027-2030, 1996).

60

Un elemento vital en el cálculo de las distribuciones isotópicas es la asignación de las masas centrales a las variantes de isótopos agregadas. Para este objetivo, la masa central puede calcularse como la suma ponderada por la probabilidad de las masas de las combinaciones de isótopos que contribuyen a esta variante, como se define por Roussis y Proulx (*Analytical Chemistry* 75(6): 1470-1482, 2003). Sin embargo, la exactitud de este cálculo de masa

depende del número de combinaciones de isótopos que se consideren. Rockwood y otros (Analytical Chemistry 68:2027-2030, 1996) resolvieron este problema mediante una transformación lineal basada en la masa promedio y la desviación estándar de la distribución isotópica para adquirir masas semiexactas. Más tarde, Rockwood y colegas se centraron en el cálculo exacto de la masa de una variante de isótopos agregada preseleccionada (Journal of the American Society for Mass Spectrometry 15:1221, 2004 e *ibidem* 17:415-419, 2006). Otra solución al problema mencionado anteriormente se propuso por Olson y Yergey en Journal of the American Society for Mass Spectrometry 20:295-302, 2009, donde propagaron la idea de usar isótopos ecuitransneutrónicos. Sin embargo, ellos aún observaron algunos errores en las asignaciones de masas de las variantes de isótopos agregadas. Para superar esta inexactitud estimaron el error y lo tuvieron en cuenta en el cálculo de las masas centrales. El método presentado en este documento es capaz de calcular las masas centrales exactas debido a que el mismo acumula la información de las masas a lo largo del cálculo recursivo de la distribución isotópica. Con el fin de demostrar su exactitud, se comparará la masa central calculada con el resultado obtenido mediante la expansión multinomial de Yergey (International Journal of Mass Spectrometry and Ion Physics 52:337-349, 1983) como se implementó en el paquete de software IsoPro® (disponible en <http://sites.google.com/site/isoproms/home>). Se usa el método de Yergey como punto de referencia porque la expansión multinomial refleja realmente el principio físico detrás de las distribuciones de isótopos.

Además, debe mencionarse que el método descrito en este documento es muy adecuado para las distribuciones de isótopos de las moléculas bioquímicas largas tales como los polipéptidos y los polinucleótidos, y que la mayoría de los ejemplos dados en este texto se limitan a los cálculos de la distribución de isótopos de las moléculas que sólo contienen carbono (C), nitrógeno (N), hidrógeno (H), oxígeno (O) y azufre (S). Los isótopos más abundantes (y más ligeros) para

los últimos elementos son  $^{12}_6\text{C}$ ,  $^1_1\text{H}$ ,  $^{14}_7\text{N}$ ,  $^{16}_8\text{O}$  y  $^{32}_{16}\text{S}$ . Una molécula compuesta sólo por los elementos anteriores se llama la variante monoisotópica. Adicionalmente, sólo se consideran los isótopos estables, es decir, los isótopos que se acaban de mencionar, junto con  $^{13}_6\text{C}$ ,  $^2_1\text{H}$ ,  $^{15}_7\text{N}$ ,  $^{17}_8\text{O}$ ,  $^{18}_8\text{O}$ ,  $^{33}_{16}\text{S}$ ,  $^{34}_{16}\text{S}$  y  $^{36}_{16}\text{S}$ . Extender los algoritmos presentados a las moléculas que contienen otros elementos poliisotópicos es sencillo.

Yamamoto y otros (Analytical Chemistry 49:281, 1977) y Brownawell y otros (Journal of Chemical Education 59(8):663-65, 1982) sostuvieron que, para las moléculas grandes, las formas isotópicas podrían obtenerse fácilmente al expandir simbólicamente una función polinomial. En el caso de las proteínas o los péptidos con una composición  $\text{C}_v\text{H}_w\text{N}_x\text{O}_y\text{S}_z$ , este polinomio toma la forma de

$$\begin{aligned} & (^{12}_6\text{C} + ^{13}_6\text{C})^v \times (^1_1\text{H} + ^2_1\text{H})^w \times (^{14}_7\text{N} + ^{15}_7\text{N})^x \\ & \times (^{16}_8\text{O} + ^{17}_8\text{O} + ^{18}_8\text{O})^y \times (^{32}_{16}\text{S} + ^{33}_{16}\text{S} + ^{34}_{16}\text{S} + ^{36}_{16}\text{S})^z. \end{aligned} \quad (\text{EL.1})$$

La expansión simbólica de la ecuación (EL.1) resulta en muchos términos de productos equivalentes, que corresponden a las moléculas con la misma masa. Recolectar estos términos de productos equivalentes y sustituir las probabilidades

de ocurrencia para  $^{12}_6\text{C}$ ,  $^{13}_6\text{C}$ , ...,  $^{36}_{16}\text{S}$  en cada término por separado, resulta en la prevalencia de las variantes de isótopos del péptido con su valor de masa correspondiente. La expansión polinomial simbólica se realizó explícitamente en computadoras. Aunque este método da las masas exactas e información sobre la estructura fina isotópica de la molécula, tiene propiedades de escalado desfavorables y un proceso de cálculo engorroso. Por lo tanto, se reivindica que la metodología anterior es adecuada para el cálculo de las distribuciones isotópicas sólo para las moléculas pequeñas. Por el contrario, el método que se propone en este documento se concibe predominantemente para calcular las variantes de isótopos agregadas de las moléculas grandes. Hacerlo así, ignora la estructura fina isotópica, pero esto no es un problema importante ya que para las moléculas grandes, como por ejemplo, las proteínas intactas, la resolución en la MS es una limitación para resolver la estructura fina de las variantes de isótopos agregadas. Sin embargo, para las moléculas grandes el cálculo de las masas centrales exactas se vuelve fundamental, lo cual se atiende por el método propuesto. Cuando se requiere la información acerca de la estructura fina isotópica, por ejemplo, debido a que se usa la espectrometría de masa de alta resolución, están disponibles otros métodos propuestos por Snider (Journal of the American Society for Mass Spectrometry 18:1511-1515, 2007) y Li y otros (Journal of the American Society for Mass Spectrometry 19:1867-1874, 2008 y Rapid Communications in Mass Spectrometry 24:2689-2696, 2010). Si la molécula no es demasiado grande, pueden usarse entonces el método anterior o la expansión multinomial de Yergey.

Con el fin de hacer una abstracción de las masas de los isótopos, se introduce una variable indicadora (polinomio en la variable  $l$ ) en la ecuación (EL.1), que proyecta el cálculo de la distribución isotópica en términos del contenido adicional de neutrones, es decir, la distribución isotópica agregada. Esto puede escribirse de la siguiente manera:

60

$$(P(^{12}_6C)I^0 + P(^{13}_6C)I^1)^v \times (P(^1_1H)I^0 + P(^2_1H)I^1)^w \times (P(^{14}_7N)I^0 + P(^{15}_7N)I^1)^x \\ \times (P(^{16}_8O)I^0 + P(^{17}_8O)I^1 + P(^{18}_8O)I^2)^y \times (P(^{32}_{16}S)I^0 + P(^{33}_{16}S)I^1 + P(^{34}_{16}S)I^2 + P(^{36}_{16}S)I^4)^z, \quad (EL.2)$$

donde los términos  $P(^{12}_6C)$ ,  $P(^1_1H)$ ,  $P(^{14}_7N)$ ,  $P(^{16}_8O)$  y  $P(^{32}_{16}S)$  representan las probabilidades de ocurrencia de las abundancias naturales de los elementos en la materia terrestre normal, tal como se muestra en la tabla (ELT.1). Se hace notar que la potencia del indicador simbólico  $I$  representa el contenido adicional de neutrones (o el desplazamiento de masa discreto) con respecto a la variante monoisotópica. El indicador  $I$  es el nuevo dispositivo de control de muestras en lugar de los símbolos en la expansión de la ecuación (EL.1) y lleva el control de las diferentes variantes de isótopos agregadas. Debe destacarse que la ecuación (EL.2) hace una abstracción de la información de las masas a medida que las variantes de isótopos agregadas se presentan por su recuento adicional de neutrones.

Tabla (ELT.1): Pesos atómicos estándares (IUPAC 1997, K.J.R. Rosman y P.D.P. Taylor, "Isotopic composition of the elements 1997", Pure and Applied Chemistry 70(1):217-235, 1988) para los isótopos de los elementos tal como existen de forma natural en el material terrestre normal.

Isótopo	Masa atómica (ma/u)	Abundancia natural (% átomo)
<b>Carbono</b>		
$^{12}_6C$	12.0000000000	98.93
$^{13}_6C$	13.0033548378	1.07
<b>Hidrógeno</b>		
$^1_1H$	1.0078250321	99.9885
$^2_1H$	2.0141017780	0.0115
<b>Nitrógeno</b>		
$^{14}_7N$	14.0030740052	99.632
$^{15}_7N$	15.0001088984	0.368
<b>Oxígeno</b>		
$^{16}_8O$	15.9949146	99.757
$^{17}_8O$	16.9991312	0.038
$^{18}_8O$	17.9991603	0.205
<b>Azufre</b>		
$^{32}_{16}S$	31.97207070	94.93
$^{33}_{16}S$	32.97145843	0.76
$^{34}_{16}S$	33.96786665	4.29
$^{36}_{16}S$	35.96708062	0.02

El polinomio en la ecuación (EL.2) puede representarse en varias formas. Por ejemplo, el polinomio en el lado izquierdo de la ecuación (EL.3) en su forma simple, puede escribirse como una suma de potencias de  $l$  en su forma expandida; la forma expandida puede verse en el lado derecho de la ecuación (EL.3).

5 
$$(al^0 + bl^1 + cl^2)^3 = q_0l^0 + q_1l^1 + q_2l^2 + q_3l^3 + q_4l^4 + q_5l^5 + q_6l^6 \quad \text{(EL.3)}$$

Es obvio que ambos términos son equivalentes. Ahora el truco será calcular los coeficientes  $q_0, \dots, q_6$ , que están en función de  $a, b$  y  $c$ . Sin embargo, incluso para este ejemplo aparentemente simple, el resultado puede ser ya bastante complejo, como se muestra en la ecuación (EL.4).

10 
$$\begin{aligned} q_0 &= a^3 & q_1 &= 3a^2b & q_2 &= 3a^2c + 3ab^2 \\ q_3 &= b^3 + 6abc & q_4 &= 3b^2c + 3ac^2 & q_5 &= 3bc^2 & q_6 &= c^3 \end{aligned} \quad \text{(EL.4)}$$

15

Se hace notar que la solución en la ecuación (EL.4) es idéntica a la distribución isotópica agregada de una molécula de ozono ( $O_3$ ). Por supuesto, los coeficientes  $a, b$  y  $c$  deben sustituirse por las probabilidades de ocurrencia de  $^{16}_8O, ^{17}_8O$  y  $^{18}_8O$  respectivamente.

20

Más generalmente, el polinomio en la ecuación (EL.2) puede expresarse como

25 
$$Q(l; v, w, x, y, z) = \sum_{i=0}^n q_i l^i \quad \text{(EL.5)}$$

donde  $n = v + w + x + 2y + 4z$  denota la mayor potencia, que es una función de la composición atómica de la molécula. Por lo tanto, el problema de calcular la distribución isotópica puede reformularse como el problema de encontrar los coeficientes  $q_i$  para  $i = 0 \dots n$  del polinomio expandido. En principio, los coeficientes  $q_i$  representan la probabilidad de ocurrencia de la  $i$ -ésima variante de isótopos agregada y puede obtenerse mediante

30

$$q_i = \sum_j p_{ij} \quad \text{(EL.6)}$$

35 donde  $p_{ij}$  denota las probabilidades de todas las  $j$  variantes de isótopos con  $i$  neutrones adicionales. Desafortunadamente, con el fin de calcular la suma en la ecuación (EL.6), se necesita conocer todas las posibles variantes de isótopos que contribuyen a la  $i$ -ésima variante agregada. Esto puede lograrse a través de la solución de una ecuación diofántica, que es un proceso engorroso y computacionalmente intensivo. Más adelante, se describirá cómo estos coeficientes pueden calcularse de una manera matemática elegante mediante el uso de las propiedades de los polinomios simétricos elementales y las sumas de las potencias de las raíces del polinomio de la ecuación (EL.2). El método evita la solución de la ecuación diofántica.

40

Basado en una adaptación del teorema de Newton-Girard y las fórmulas de Viète (ver, por ejemplo, I.G. Macdonald, "Symmetric functions and Hall polynomials", Clarendon Press, Oxford University Press, Oxford: New York, 1979), pueden expresarse los coeficientes  $q_i$  en términos de las sumas de las potencias de las raíces  $\rho_{-1}, \rho_{-2}, \dots$  del polinomio en la ecuación (EL.2) y el coeficiente  $q_0, \dots, q_{i-1}$ :

45

$$q_i = \sum_{j=0}^{i-1} C_{i,j} q_j \quad \text{(1)}$$

50 con

$$C_{i,j} = -\frac{1}{i} \rho_{-(i-j)} \quad \text{(2)}$$

55 y por lo tanto

60 
$$q_i = -\frac{1}{i} \sum_{j=0}^{i-1} \rho_{-(i-j)} q_j \quad \text{(EL.7)}$$

los términos  $q_i$  son los coeficientes del polinomio expandido y corresponden aquí a la probabilidad de ocurrencia de la

variante de isótopos agregada con  $i$  neutrones adicionales. Esta es una ecuación recursiva, que comienza con el coeficiente  $q_0$ . El coeficiente  $q_0$  puede calcularse fácilmente, ya que corresponde a la probabilidad de la variante monoisotópica. Como se señaló por Beynon, la probabilidad de que isótopos no pesados aparezcan en un péptido de composición  $C_vH_wN_xO_yS_z$  fue

5

$$q_0 = P(^{12}_6C) \times P(^1_1H) \times P(^{14}_7N) \times P(^{16}_8O) \times P(^{32}_{16}S) \quad (\text{EL.8})$$

En la siguiente etapa debe calcularse la suma de las potencias de las raíces  $\rho_1, \rho_2, \dots$  del polinomio (EL.2). Estos términos pueden calcularse con antelación mediante el uso de la transformación logarítmica para mejorar la estabilidad numérica y pueden almacenarse en la computadora para las etapas de cálculo consecutivas.

10

Sin embargo, puede reescribirse la relación recursiva de las ecuaciones (1), (2) y (EL.7) en forma de matriz como:

15

$$\begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} = \begin{bmatrix} C_{1,0} & 0 & \dots & 0 \\ C_{2,0} & C_{2,1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_{n,0} & C_{n,1} & \dots & C_{n,n-1} \end{bmatrix} \begin{bmatrix} q_0 \\ q_1 \\ \vdots \\ q_{n-1} \end{bmatrix} \quad (\text{M3})$$

20

donde la matriz de coeficientes es una matriz triangular inferior. Debe quedar claro para el experto que puede calcularse cualquier  $q_i$  con  $i > 0$  como una proporción del valor para  $q_0$  mediante el uso de, por ejemplo, un algoritmo de sustitución directa, sin la necesidad de tener que calcular todos los  $q_j$  intermedios con  $0 < j < i$ . Esto significa además que no es necesario empezar desde el valor de  $q_0$  para calcular todos los  $q_i$ , puede empezarse igualmente a partir de cualquier  $q_i$  con  $i > 0$ . Esto puede ilustrarse más fácilmente al reescribir la relación recursiva de las ecuaciones (1), (2), (EL.7) y (M3) como

25

$$Q \equiv \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ C_{1,0} & 0 & 0 & \dots & 0 \\ C_{2,0} & C_{2,1} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ C_{n,0} & C_{n,1} & \dots & C_{n,n-1} & 0 \end{bmatrix} \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} \quad (\text{M4})$$

30

35 que puede reescribirse como

40

$$\mathcal{Q}Q = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ C_{1,0} & -1 & 0 & \dots & 0 \\ C_{2,0} & C_{2,1} & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ C_{n,0} & C_{n,1} & \dots & C_{n,n-1} & -1 \end{bmatrix} \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} = 0 \quad (\text{M5})$$

45

La solución de este sistema de ecuaciones lineales se da por el espacio nulo de la matriz  $\mathcal{Q}$ . Si los números  $q_0, q_1, \dots, q_n$  se toman como los  $(n+1)$  números de coordenadas en un espacio  $(n+1)$ -dimensional, la solución de la ecuación (M5) se da por todos los vectores a lo largo de una sola dirección en este espacio de soluciones  $(n+1)$ -dimensional. Esto puede observarse a partir de la forma de la matriz  $\mathcal{Q}$  que claramente tiene  $n$  filas linealmente independientes y 1 fila dependiente: la primera con todos sus elementos iguales a cero. El espacio nulo de la matriz  $\mathcal{Q}$  es por lo tanto unidimensional. La dirección en el espacio  $(n+1)$  dimensional que define el espacio nulo se determina por un vector  $(n+1)$ -dimensional cuyos componentes son proporcionales a las probabilidades  $q_i, i = 0 \dots n$ . Por lo tanto, todas las probabilidades  $q_i, i = 0 \dots n$  pueden obtenerse una vez que se encuentre una solución a la ecuación (M5) y sólo se conozca una probabilidad, por ejemplo,  $q_0$  pero cualquier probabilidad lo hará. Una solución de la ecuación (M5) puede encontrarse mediante el uso de un gran número de técnicas conocidas en la técnica: puede usarse la eliminación de Gauss, pero puede dar lugar a problemas de exactitud numérica debido a los errores de redondeo para las matrices grandes; la descomposición en valores singulares (SVD) es una técnica actual que mantiene las inexactitudes numéricas bajo control y que puede usarse además para encontrar el núcleo o el espacio nulo de la matriz  $\mathcal{Q}$ ; la bien conocida descomposición QR puede usarse también, etc.

50

55

60

El debate anterior ilustra que no es necesario comenzar con la probabilidad de la variante isotópica más ligera  $q_0$ , cuando se calcula la distribución isotópica agregada de una molécula. Al realizar operaciones lineales bien conocidas en la matriz  $\mathcal{Q}$ , tales como tomar combinaciones lineales de filas o columnas, puede ponerse la matriz  $\mathcal{Q}$  en una forma tal

que permita empezar con cualquier  $q_i$ , por ejemplo,  $q_1$ , en cuyo caso la ecuación (M5) puede multiplicarse por la izquierda por la matriz invertible

$$\begin{bmatrix} 1 & -\frac{1}{c_{1,0}} & 0 & \dots & 0 \\ c_{1,0} & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \quad (M6)$$

que conduce a un conjunto de ecuaciones  $e^t Q = 0$ , con

$$e^t = \begin{bmatrix} -1 & \frac{1}{c_{1,0}} & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ c_{2,0} & c_{2,1} & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n,0} & c_{n,1} & \dots & c_{n,n-1} & -1 \end{bmatrix} \quad (M7)$$

Aquí, puede verse fácilmente que  $q_1$  es la primera probabilidad a partir de la cual pueden calcularse las otras probabilidades. Puede ser beneficioso no empezar con la probabilidad de la variante isotópica agregada más ligera como una probabilidad de partida  $q_0$ , especialmente cuando sólo se quiere obtener un conjunto de probabilidades de las variantes isotópicas agregadas que no incluya la variante monoisotópica más ligera. Por lo tanto, en una modalidad preferida, la variante isotópica agregada de partida puede comprender la variante isotópica más abundante y/o puede ser la variante isotópica agregada más abundante.

Cuando se quiere obtener sólo un subconjunto de las probabilidades  $q_i$ , no siempre es necesario calcular una solución completa de la ecuación (M5). El lector experto apreciará que sólo se requiere esa parte de una solución para la ecuación (M5) lo cual corresponde al subconjunto de probabilidades  $q_i$ .

Por ejemplo, para la molécula de propano  $C_3H_8$ , la variante isotópica agregada de partida es la variante monoisotópica más ligera para la que se calcula la probabilidad como  $q_0 = P(^{12}_6C)^3 \times P(^1_1H)^8$ . La siguiente variante de isótopos agregada se calcula como  $q_1 = q_0 \times \rho_{-1}$ , donde  $\rho_{-1}$  es la primera suma de las potencias 'negativas' del polinomio

$$(P(^{12}_6C)I^0 + P(^{13}_6C)I^1)^3 \times (P(^1_1H)I^0 + P(^2_1H)I^1)^8$$

que toma la forma

$$3 \times r_C^{-1} + 8 \times r_H^{-1} \quad (EL.9)$$

y

donde  $r_C$  y  $r_H$  son las raíces de los términos  $(P(^{12}_6C)I^0 + P(^{13}_6C)I^1)$  and  $(P(^1_1H)I^0 + P(^2_1H)I^1)$  respectivamente. Se hace notar que el índice de la suma de las potencias de las raíces  $\rho$  indica la potencia a la cual se elevan las raíces elementales individuales. Ya merece mencionar que la complejidad del cálculo no depende de los valores  $v, w, x, y, z$ , sino del número de variantes de isótopos agregadas requeridas, que es diferente de las técnicas anteriores, y que resulta en un cálculo y/o análisis mejorado, más rápido, más eficiente y más fiable de la distribución isotópica de una muestra.

Las raíces de los polinomios elementales son triviales y pueden almacenarse en la memoria de la computadora para su posterior uso. Para los elementos  $C, H$  y  $N$  las raíces se convierten en

$$r_C = -\frac{P(^{12}_6C)}{P(^{13}_6C)} \quad r_H = -\frac{P(^1_1H)}{P(^2_1H)} \quad r_N = -\frac{P(^{14}_7N)}{P(^{15}_7N)} \quad (EL.10)$$

Las raíces de 0 son el número complejo  $r_0$  y el conjugado del complejo  $r_0$  que toman la forma

$$r_0 = \frac{-P(^{17}_8O) + \sqrt{P(^{17}_8O)^2 - 4P(^{16}_8O)P(^{18}_8O)}}{2P(^{18}_8O)} \quad \bar{r}_0 = \frac{-P(^{17}_8O) - \sqrt{P(^{17}_8O)^2 - 4P(^{16}_8O)P(^{18}_8O)}}{2P(^{18}_8O)} \quad (EL.11)$$

Las raíces de S son las raíces de un polinomio de cuarto orden y son menos triviales, pero existe una solución de forma cerrada. La expresión parece no muy elegante, pero puede encontrarse mediante el uso de la solución de forma cerrada

o mediante algoritmos numéricos bien conocidos. Las raíces del polinomio de azufre tienen dos pares de raíces complejas y conjugadas, específicamente  $(r_{s,1}, r_{s,1})$  y  $(r_{s,2}, r_{s,2})$ .

5 La suma de las potencias de las raíces (EL.9) en el ejemplo anterior puede expresarse de manera más general. Para las proteínas y los péptidos esto se convierte en

$$\rho_{-d} = vr_C^{-d} + wr_H^{-d} + xr_N^{-d} + yr_O^{-d} + y\bar{r}_O^{-d} + zr_{S,1}^{-d} + z\bar{r}_{S,1}^{-d} + zr_{S,2}^{-d} + z\bar{r}_{S,2}^{-d} \quad (\text{EL.12})$$

La suma de las potencias de las raíces complejas  $r$  y conjugadas  $\bar{r}$  puede escribirse como

$$r^{-d} + \bar{r}^{-d} = 2|r|^{-d} \cos(-d\phi(r)) \quad (\text{EL.13})$$

10 donde  $|r|$  y  $\phi(r)$  indican el módulo y el argumento de las raíces  $r$  y  $\bar{r}$  respectivamente. Dada la formulación anterior (EL.13), la ecuación (EL.12) puede simplificarse al sustituir las raíces complejas y conjugadas del oxígeno y el azufre por su forma reducida. Para los elementos con más de cuatro variantes isotópicas, una solución de forma cerrada de las raíces es a menudo impracticable (teorema de Abel-Ruffini), por lo tanto, las raíces pueden calcularse mediante el uso de métodos numéricos de búsqueda de raíces, tales como el método de Newton-Raphson o de Dandelin-Graeffe. El valor de  $\rho_{-d}$  puede calcularse fácilmente mediante el uso de la vectorización y las fórmulas recursivas para los números

15 añadidos (por ejemplo,  $r_{S,1}^{-d} = r_{S,1}^{-1}r_{S,1}^{-(d-1)}$ , por lo que si ya se han calculado  $\rho_{-(d-1)}$  pueden usarse para calcular  $\rho_{-d}$ ). En el texto anterior, las potencias de la suma de potencias son negativas porque la teoría anterior calcula los coeficientes desde las variantes monoisotópicas más ligeras hasta las variantes más pesadas. De hecho, este esquema es el opuesto al presentado en el teorema de Newton-Girard, donde el cálculo comienza con la variante más pesada. Usualmente, la información acerca de las variantes agregadas más pesadas no es necesaria, pero si fuera necesaria el método algorítmico puede paralelizarse al realizar el cálculo desde la variante más ligera hasta la  $(\frac{n}{2})$ -ésima variante y desde la variante más pesada hasta la  $(\frac{n}{2})$ -ésima variante. En la última parte el cálculo se realiza tomando las potencias naturales en lugar de las potencias inversas como en la ecuación (EL. 7).

25 Adicionalmente, puede verse que la complejidad de los cálculos depende sólo del número de átomos diferentes (en el texto anterior C, H, N, O, S, pero obviamente esto puede extenderse a todos los átomos de la tabla periódica, preferentemente, todos los átomos de la tabla periódica que tienen más de un isótopo estable) y no depende de los valores de  $v, w, x, y, z$ . El método usa eficientemente la memoria ya que sólo dos variables, específicamente,  $q_i$  y  $\rho_{-d}$ , deben almacenarse por cada variante de isótopos agregada requerida. Por lo tanto, calcular las primeras 100 variantes de isótopos agregadas sólo requiere que se almacenen 201 números.

30 La velocidad de cálculo del método descrito es al menos comparable con, y en muchos casos mejor que, el método de Olson y Yergey (Journal of the American Society for Mass Spectrometry 20:295-302, 2009) o el método de Rockwood y otros (Analytical Chemistry 68:2027-2030, 1996). Sin embargo, una ventaja adicional de los métodos de la presente invención está en cómo se calculan las masas centrales de las variantes de isótopos agregadas. Por definición, la masa central  $\bar{m}_i$  de la  $i$ -ésima variante de isótopos se calcula como una suma ponderada por la probabilidad, como se debate por Proulx y Roussis (Analytical Chemistry 75(6):1470-1482, 2003):

$$\bar{m}_i = \frac{\sum_j m_{ij} p_{ij}}{\sum_j p_{ij}} \quad (\text{EL.14})$$

40 donde  $p_{ij}$  y  $m_{ij}$  indican la probabilidad de ocurrencia y la masa de la  $j$ -ésima variante de isótopos que contribuye a la  $i$ -ésima variante de isótopos agregada. Es obvio que los cálculos exactos de las masas centrales sólo pueden lograrse si todas las variantes de isótopos contribuyente se consideran, lo cual es difícil y frecuentemente impracticable debido a la explosión combinatoria del número de variantes de isótopos para las moléculas grandes (es decir, el requisito de la solución diofántica). Sin embargo, puede intentarse eludir este cálculo exhaustivo mediante el uso del teorema de Newton-Girard y las fórmulas de Viète como se propuso anteriormente. En este caso, se preferiría la solución del siguiente polinomio

$$U(I; v, w, x, y, z) = \sum_i (\sum_j m_{ij} p_{ij}) I^i \quad (\text{EL.15})$$

50 donde la suma de las masas ponderadas  $\sum_j m_{ij} p_{ij}$  de las variantes de isótopos con  $i$  neutrones adicionales son los objetos de interés.

Es importante darse cuenta de que el denominador de la ecuación (EL. 14) o, equivalentemente, la solución de la ecuación (EL.6), ya se obtuvo mediante el teorema de Newton-Girard. En principio, el algoritmo de Newton-Girard convierte los coeficientes del polinomio colapsado (EL.2), es decir, las probabilidades elementales  $P(^{13}_6\text{C}), \dots, P(^{36}_{16}\text{S})$ , en los coeficientes del polinomio expandido (EL.5). Debe notarse que esta transformación sólo puede lograrse debido a que la probabilidad de ocurrencia de una variante de isótopos particular es un producto de las probabilidades elementales que componen la variante de isótopos. Ahora en el caso de los cálculos de masa, la

implementación del principio de Newton-Girard es menos trivial. La masa de una variante de isótopos es aditiva en función de las masas elementales que componen la variante de isótopos. Por lo tanto, con el fin de obtener los términos en el numerador de la ecuación (EL.14) o equivalentemente, los coeficientes del polinomio (EL.15), debe encontrarse un método abreviado para escribir la adición de términos como una multiplicación de términos. Para ello se usa un principio básico del cálculo elemental:

$$K^a \times K^b = K^{a+b}. \quad (\text{EL.16})$$

Con el fin de obtener una fórmula en forma de polinomio (EL.15), se introducirá una variable indicadora adicional en el polinomio (EL.2):

$$\begin{aligned} & (P({}^{12}_6\text{C})K^{M({}^{12}_6\text{C})}I^0 + P({}^{13}_6\text{C})K^{M({}^{13}_6\text{C})}I^1)^v \\ & \times (P({}^1_1\text{H})K^{M({}^1_1\text{H})}I^0 + P({}^2_1\text{H})K^{M({}^2_1\text{H})}I^1)^w \\ & \times (P({}^{14}_7\text{N})K^{M({}^{14}_7\text{N})}I^0 + P({}^{15}_7\text{N})K^{M({}^{15}_7\text{N})}I^1)^x \\ & \times (P({}^{16}_8\text{O})K^{M({}^{16}_8\text{O})}I^0 + P({}^{17}_8\text{O})K^{M({}^{17}_8\text{O})}I^1 + P({}^{18}_8\text{O})K^{M({}^{18}_8\text{O})}I^2)^y \\ & \times (P({}^{32}_{16}\text{S})K^{M({}^{32}_{16}\text{S})}I^0 + P({}^{33}_{16}\text{S})K^{M({}^{33}_{16}\text{S})}I^1 + P({}^{34}_{16}\text{S})K^{M({}^{34}_{16}\text{S})}I^2 + P({}^{36}_{16}\text{S})K^{M({}^{36}_{16}\text{S})}I^4)^z, \end{aligned} \quad (\text{EL.17})$$

donde la variable indicadora  $K$  actúa como un dispositivo de seguimiento para las masas elementales  $M({}^{12}_6\text{C}), \dots, M({}^{36}_{16}\text{S})$ . El polinomio (EL.17) puede expresarse ahora como un producto de las probabilidades elementales y la variable indicadora  $K$  de una manera similar como en la parte anterior donde se calcularon las probabilidades de la variante de isótopos:

$$T(I, K; v, w, x, y, z) = \sum_i (\sum_j p_{ij} K^{m_{ij}}) I^i \quad (\text{EL.18})$$

En las siguientes dos etapas, se manipulará el polinomio en (EL.18) de manera que se obtendrá el polinomio en la ecuación (EL.15). En la primera etapa se diferenciará el polinomio en  $K$  para obtener:

$$\frac{d}{dx} T(I, K; v, w, x, y, z) = \sum_i (\sum_j m_{ij} p_{ij} K^{m_{ij}-1}) I^i \quad (\text{EL.19})$$

En la segunda etapa, hay que evaluar la variable indicadora en la ecuación (EL.19) para un valor de 1. De esta forma se obtiene el polinomio  $U(I; v, w, x, y, z)$  como se define en (EL.15).

Ahora al aplicar la diferenciación y la evaluación de  $K$  en 1 a la forma colapsada del polinomio (EL.17) se obtendrá

$$U(I; v, w, x, y, z) = vQ(I; v-1, w, x, y, z)(P({}^{12}_6\text{C})M({}^{12}_6\text{C})I^0 + P({}^{13}_6\text{C})M({}^{13}_6\text{C})I^1) \quad (\text{EL.20})$$

$$+ w(I; v, w-1, x, y, z)(P({}^1_1\text{H})M({}^1_1\text{H})I^0 + P({}^2_1\text{H})M({}^2_1\text{H})I^1) \quad (\text{EL.21})$$

$$+ x(I; v, w, x-1, y, z)(P({}^{14}_7\text{N})M({}^{14}_7\text{N})I^0 + P({}^{15}_7\text{N})M({}^{15}_7\text{N})I^1) \quad (\text{EL.22})$$

$$+ y(I; v, w, x, y-1, z)(P({}^{16}_8\text{O})M({}^{16}_8\text{O})I^0 + P({}^{17}_8\text{O})M({}^{17}_8\text{O})I^1 + P({}^{18}_8\text{O})M({}^{18}_8\text{O})I^2) \quad (\text{EL.23})$$

$$+ z(I; v, w, x, y, z-1) \times (P({}^{32}_{16}\text{S})M({}^{32}_{16}\text{S})I^0 + P({}^{33}_{16}\text{S})M({}^{33}_{16}\text{S})I^1 + P({}^{34}_{16}\text{S})M({}^{34}_{16}\text{S})I^2 + P({}^{36}_{16}\text{S})M({}^{36}_{16}\text{S})I^4). \quad (\text{EL.24})$$

Anteriormente ya se mostró cómo los coeficientes del polinomio expandido  $Q$  pueden calcularse fácilmente. Sólo se necesita modificar un poco el algoritmo para tener en cuenta la multiplicación adicional con el polinomio que contiene las masas elementales ponderadas. Este enfoque algorítmico dividirá los cálculos del polinomio  $U(I; v, w, x, y, z)$  en cinco partes. Esto significa que el método encuentra iterativamente los coeficientes de cada uno de los cinco polinomios (EL.20), (EL.21), (EL.22), (EL.23), (EL.24) a partir de la masa monoisotópica; y luego añade los coeficientes con el mismo orden. Las masas centrales finales pueden ahora obtenerse a partir de la ecuación (EL.14), donde los resultados obtenidos deben ponderarse por las probabilidades de ocurrencia de la variante de isótopos agregada correspondiente.

5 Para cada potencia de  $I$  (que representa el nucleón adicional), el método anterior resuelve implícitamente la ecuación diofántica y devuelve todas las combinaciones de isótopos para las cuales el recuento adicional de neutrones de las variantes de isótopos se suma a la potencia. Por otra parte, puede adaptarse el esquema de cálculo de manera que las operaciones matemáticas sobre el resultado de la ecuación diofántica se habiliten sin calcular explícitamente el resultado de la ecuación diofántica.

10 El experto apreciará que los procedimientos que se describen en este documento pueden aplicarse de manera más amplia a las moléculas que consisten de más átomos diferentes presentados previamente entonces. Más específicamente, los átomos constituyentes pueden ser cualquier combinación de hidrógeno, helio, litio, berilio, boro, carbono, nitrógeno, oxígeno, flúor, neón, sodio, magnesio, aluminio, silicio, fósforo, azufre, cloro, argón, potasio, calcio, escandio, titanio, vanadio, cromo, manganeso, hierro, cobalto, níquel, cobre, cinc, galio, germanio, arsénico, selenio, bromo, kriptón, rubidio, estroncio, itrio, circonio, niobio, molibdeno, tecnecio, rutenio, rodio, paladio, plata, cadmio, indio, estaño, antimonio, telurio, yodo, xenón, cesio, bario, lantano, cerio, praseodimio, neodimio, prometio, samario, europio, gadolinio, terbio, disprosio, holmio, erbio, tulio, iterbio, lutecio, hafnio, tantalio, tungsteno, renio, osmio, iridio, platino, oro, mercurio, talio, plomo, bismuto, polonio, ástato, radón, francio, radio, actinio, torio, protactinio, uranio. Más específicamente, la distribución isotópica agregada de dos moléculas que difieren en sus constituyentes sólo por un número de átomos monoisotópicos parecerá la misma, excepto que las distribuciones se desplazarán en la masa absoluta. Los átomos monoisotópicos, es decir, los átomos con sólo un isótopo estable, son berilio-9, flúor-19, sodio-23, aluminio-27, fósforo-31, escandio-45, vanadio-51, manganeso-55, cobalto-59, arsénico-75, rubidio-85, itrio-89, niobio-93, rodio-103, indio-113, yodo-127, cesio-133, lantano-139, praseodimio-141, europio-153, terbio-159, holmio-165, tulio-169, lutecio-175, renio 185, oro-197. Además, el enfoque específico que se ha postulado sugiere un posible cálculo invertido para obtener la información de la molécula basado en información de la distribución isotópica.

25 En una modalidad, la invención proporciona un método para analizar al menos una parte de una distribución isotópica de una muestra, de manera que dicha fórmula química se describe por  $\prod_{\alpha}(Z_{\alpha})n_{\alpha}$  con

- $\alpha$ , un índice que recorre todas las especies atómicas constituyentes de la molécula;
  - $Z_{\alpha}$ , la  $\alpha$ -ésima especie atómica de la molécula; y
  - 30 •  $n_{\alpha}$ , el número de átomos de la especie  $Z_{\alpha}$  en la molécula,
- y de manera que el método comprende además las etapas de:
- generar los datos que comprenden las masas promedio ponderadas  $p_i^{\alpha}$  de cada especie atómica  $Z_{\alpha}$  dentro de dicha  $i$ -ésima variante isotópica agregada de dicha molécula con un número de masa  $A_i$ ;
  - 35 • calcular una masa central  $M_i$  de dicha  $i$ -ésima variante isotópica agregada de dicha molécula a partir de dichas masas promedio ponderadas  $p_i^{\alpha}$  de cada especie atómica  $Z_{\alpha}$  y a partir de dicha probabilidad  $q_i$ .

En una modalidad preferida, dicha masa central  $M_i$  se calcula de acuerdo con:

$$M_i = \frac{1}{q_i} \sum_{\alpha} n_{\alpha} p_i^{\alpha} \quad (5)$$

40 Las masas promedio ponderadas  $p_i^{\alpha}$  son las contribuciones relativas de la especie atómica  $Z_{\alpha}$  a la masa de la  $i$ -ésima variante isotópica agregada, normalizadas a la probabilidad  $q_i$ . En otras palabras,  $p_i^{\alpha}$  es la contribución de masa de 1 átomo de la especie  $Z_{\alpha}$  a la masa  $M_i$  de la  $i$ -ésima variante isotópica agregada, ponderada por las probabilidades relativas con las que los diferentes isótopos de la especie atómica ocurren en la  $i$ -ésima variante isotópica agregada, estas probabilidades relativas que se normalizan a la probabilidad  $q_i$  con la que la  $i$ -ésima variante isotópica agregada ocurre en la distribución isotópica agregada. Un simple ejemplo puede explicar mejor el significado de  $p_i^{\alpha}$ . Se refiere el ejemplo 2 para esta explicación.

50 La generación de datos que comprenden las masas centrales de las variantes isotópicas agregadas de una molécula puede hacerse en el mismo sentido que la generación de las probabilidades con las que las variantes isotópicas agregadas de una molécula ocurren en una distribución isotópica agregada. La masa central  $M_i$  de la  $i$ -ésima variante isotópica agregada de una molécula se define por:

$$M_i = \frac{\sum_j M_{i,j} p_{i,j}}{\sum_j p_{i,j}} \quad (5a)$$

55 con  $M_{i,j}$  la masa de la  $j$ -ésima variante isotópica contribuyente a la  $i$ -ésima variante isotópica agregada de la molécula y  $p_{i,j}$  la probabilidad con la que una variante isotópica (se hace notar que esta no es una variante isotópica *agregada*) contribuye a la distribución isotópica de la molécula, y de manera que la suma en  $j$  recorra todas las variantes isotópicas que contribuyen a la  $i$ -ésima variante isotópica agregada. Se hace notar que el denominador en la ecuación (5a):  $\sum_j p_{i,j} = q_i$ , de manera que  $q_i$  es la probabilidad con la que la  $i$ -ésima variante isotópica agregada de dicha molécula con un

número de masa  $A_i$  ocurre en la distribución isotópica agregada de la molécula. Se hace notar que la ecuación (5a) es la misma que la ecuación (EL. 14) para  $M_i = \bar{m}_i$  and  $M_{i,j} = m_{i,j}$ .

5 En el caso de que la variante isotópica agregada de partida sea una variante monoisotópica, la masa central  $M_0$  de la variante isotópica agregada de partida de la molécula con un número de masa predeterminado  $A_0$  se da por la suma de las masas de esos isótopos de los átomos constituyentes que contribuyen a dicha variante isotópica agregada de partida. En el caso de que la variante isotópica agregada de partida sea la variante isotópica más ligera, la masa central  $M_0$  de la variante isotópica agregada de partida de la molécula con un número de masa predeterminado  $A_0$  se determina por la suma de las masas de los isótopos más ligeros de los átomos constituyentes. En el caso de que la variante isotópica agregada de partida sea la variante isotópica más pesada, la masa central  $M_0$  de la variante isotópica agregada de partida de la molécula con un número de masa predeterminado  $A_0$  se determina por la suma de las masas de los isótopos más pesados de los átomos constituyentes. Aunque es posible elegir cualquier variante isotópica agregada como una variante isotópica agregada de partida para la cual va a generarse la masa central  $M_0$ , es más fácil elegir una variante monoisotópica, ya que esto requiere menos tiempo de cálculo y/o de recuperación de datos. Se hace notar que en lo anterior, el índice "0" se usó para denotar la variante isotópica agregada de partida, y no necesariamente, por ejemplo, la más ligera.

En una modalidad en la que se calcula una masa central, al menos una masa promedio ponderada  $p_i^f$  se calcula tomando una combinación lineal de las masas promedio ponderadas  $p_j^f$  de la especie atómica  $Z_\alpha$  dentro de las variantes isotópicas agregada de dicha molécula con números de masa  $A_j$  diferentes de  $A_i$ , es decir  $p_i^\alpha = \sum_{j=0}^{i-1} D_{i,j}^\alpha p_j^\alpha$ , con  $D_{i,j}^\alpha$  los coeficientes de la combinación lineal. Los beneficios de este enfoque para calcular la masa central de la  $i$ -ésima variante isotópica agregada en este sentido es que no se necesita ni hay que calcular ninguna información sobre la variante isotópica independiente que contribuye a la  $i$ -ésima variante isotópica agregada, con el consiguiente ahorro en el tiempo de cálculo o de recuperación de datos en comparación con las técnicas anteriores. La masa promedio ponderada  $p_i^f$  para una especie atómica  $Z_\alpha$  y la  $i$ -ésima variante isotópica agregada puede calcularse a partir de las masas promedio ponderadas  $p_j^f$  de la misma especie atómica  $Z_\alpha$  y de otras variantes isotópicas agregadas con el índice  $j \neq i$ , que pueden obtenerse a partir de una base de datos, calculadas de acuerdo con una técnica anterior, calculadas a partir de los principios básicos, calculadas de acuerdo con cualquier técnica especificada en este documento o cualquier otra técnica. Los valores exactos de los coeficientes  $D_{i,j}^\alpha$  dependerán de los índices  $i, j, \alpha$  y de la fórmula química de la molécula, pero son únicos, es decir, si se conocen las masas promedio ponderadas  $p_j^f$  para las variantes isotópicas agregadas con los números de masa  $A_j \neq A_i$  y para la especie atómica  $Z_\alpha$ , los coeficientes  $D_{i,j}^\alpha$  para calcular la masa promedio ponderada  $p_i^f$  se determinan unívocamente. En una modalidad preferida, dicha masa promedio ponderada  $p_i^f$  se calcula recursivamente comenzando a partir de una masa promedio ponderada  $p_0^f$  de cada especie atómica  $Z_\alpha$  dentro de una variante isotópica agregada cero o de partida de dicha molécula con número de masa  $A_0$ . En tal modalidad, las masas centrales de una o más variantes isotópicas agregadas pueden calcularse de manera más precisa y rápida que con las técnicas anteriores. Además, calcular las masas promedio ponderadas recursivamente reduce la necesidad de consultar bases de datos, con la posible reducción del tiempo de cálculo. En una modalidad preferida, un coeficiente  $D_{i,j}^\alpha$  para un valor predeterminado de  $\alpha$  se obtiene mediante el uso de las etapas de:

- 40 • obtener para cada valor de  $\alpha$ , cada raíz  $r_{\alpha,i_\alpha}$ , con  $i_\alpha = 1 \dots N_\alpha$ , de una ecuación polinomial elemental de  $N_\alpha$ -ésimo orden cuyo coeficiente del término de  $m_\alpha$ -ésimo orden se determina por una abundancia elemental  $P^{(A_\alpha+m_\alpha Z_\alpha)}$  del isótopo de la especie atómica  $Z_\alpha$  con  $m_\alpha$  neutrones más que el isótopo estable más ligero de dicha especie atómica que tiene un número másico  $A_\alpha$ ;
- 45 • obtener para dicho valor predeterminado de  $\alpha$ , cada raíz  $s_{\alpha,i_\alpha}$ , con  $i_\alpha = 1 \dots N_\alpha$ , de una ecuación polinomial extendida de  $N_\alpha$ -ésimo orden cuyo coeficiente del  $m_\alpha$ -ésimo orden se determina por el producto de una abundancia elemental  $P^{(A_\alpha+m_\alpha Z_\alpha)}$  y la masa  $M^{(A_\alpha+m_\alpha Z_\alpha)}$  del isótopo de la especie atómica  $Z_\alpha$  con  $m_\alpha$  neutrones más que el isótopo estable más ligero de dicha especie atómica que tiene un número másico  $A_\alpha$ ;
- calcular dicho coeficiente  $D_{i,j}^\alpha$  a partir de dichas raíces  $r_{\alpha,i_\alpha}$  y  $s_{\alpha,i_\alpha}$ .

50 Las ecuaciones polinomiales elementales y las raíces  $r_{\alpha, i_\alpha}$  ya se han discutido anteriormente, por ejemplo, en las ecuaciones (EL.25), (3a), (3b) y (3c) y en el texto. Las ecuaciones polinomiales extendidas usan las masas de los isótopos de la especie atómica constituyente. Más particularmente, las ecuaciones polinomiales pueden escribirse como:

55

$$\sum_{m_\alpha=0}^{N_\alpha} M({}^{A_\alpha+m_\alpha}Z_\alpha)P({}^{A_\alpha+m_\alpha}Z_\alpha)I^{m_\alpha} = 0 \quad (6)$$

en el caso de que la fórmula química de la molécula se dé por  $\Pi_\alpha(Z_\alpha)_{n_\alpha}$  de manera que  $Z_1, Z_2, \dots, Z_y$  son especies atómicas constituyentes de la molécula, con  $y$  el número de diferentes especies atómicas que constituyen la molécula. Las ecuaciones polinomiales extendidas son así:

$$M({}^{A_1}Z_1)P({}^{A_1}Z_1)I^0 + M({}^{A_1+1}Z_1)P({}^{A_1+1}Z_1)I^1 + \dots + M({}^{A_1+N_1}Z_1)P({}^{A_1+N_1}Z_1)I^{N_1} = 0 \quad (7a)$$

$$M({}^{A_2}Z_2)P({}^{A_2}Z_2)I^0 + M({}^{A_2+1}Z_2)P({}^{A_2+1}Z_2)I^1 + \dots + M({}^{A_2+N_2}Z_2)P({}^{A_2+N_2}Z_2)I^{N_2} = 0 \quad (7b)$$

$$\dots$$

$$M({}^{A_y}Z_y)P({}^{A_y}Z_y)I^0 + M({}^{A_y+1}Z_y)P({}^{A_y+1}Z_y)I^1 + \dots + M({}^{A_y+N_y}Z_y)P({}^{A_y+N_y}Z_y)I^{N_y} = 0 \quad (7c)$$

Aquí,  $I$  es la variable desconocida que puede representar el número de neutrones excedentes en las variantes isotópicas en comparación con la invariante isotópica estable más ligera;  $A_1, A_2, \dots, A_y$  pueden ser números de masa de la variante isotópica estable más ligera de las especies de átomos  $Z_1, Z_2, \dots, Z_y$  respectivamente;  $N_1, N_2, \dots, N_y$  pueden representar el número máximo de neutrones excedentes que pueden añadirse a la variante isotópica más ligera de las especies atómicas  $Z_1, Z_2, \dots, Z_y$  respectivamente, para obtener la variante isotópica estable más pesada. Los coeficientes

$$M({}^{A_1+i}Z_1)P({}^{A_1+i}Z_1), M({}^{A_2+j}Z_2)P({}^{A_2+j}Z_2), \dots, M({}^{A_y+k}Z_y)P({}^{A_y+k}Z_y)$$

de las ecuaciones polinomiales son productos de las abundancias elementales y las masas de los isótopos de las especies de átomos  $Z_1, Z_2, \dots, Z_y$  con números de masa  $A_1 + i, A_2 + j, \dots, A_y + k$  respectivamente. La solución de las ecuaciones (7a)-(7c) son un conjunto de  $N_1 + N_2 + \dots + N_y$  raíces ( $s_{1,1}, s_{1,2}, \dots, s_{1,N_1}; s_{2,1}, s_{2,2}, \dots, s_{2,N_2}; \dots; s_{y,1}, s_{y,2}, \dots, s_{y,N_y}$ ) que pueden ser números complejos. En una modalidad más preferida, dicha variante isotópica agregada cero o de partida de dicha molécula es la variante isotópica más ligera, y dichos coeficientes  $D_{i,j}^\alpha$  se calculan de acuerdo con

$$D_{i,j}^\alpha = -\frac{1}{i} \left[ \sum_{\alpha'} (n_{\alpha'} \sum_{i_{\alpha'}} r_{\alpha', i_{\alpha'}}^{-(i-j)} - r_{\alpha', i_{\alpha'}}^{-(i-j)} + s_{\alpha', i_{\alpha'}}^{-(i-j)} \right] \quad (8)$$

donde  $\alpha'$  es un índice que recorre todas las especies atómicas que constituyen la molécula, es decir,  $\alpha' = 1 \dots y$ . En otra modalidad más preferida, dicha variante isotópica agregada cero de dicha molécula es la variante isotópica más pesada y los coeficientes  $D_{i,j}^\alpha$  se calculan como la suma de las potencias positivas de las raíces de las ecuaciones polinomiales elementales y extendidas.

En una modalidad, las masas centrales  $M_i$  para las  $\binom{N}{2}$  variantes isotópicas agregadas más ligeras se calculan recursivamente a partir de la variante isotópica agregada más ligera y las masas centrales  $M_i$  para las  $\binom{N}{2}$  variantes isotópicas agregadas más pesadas se calculan recursivamente a partir de la probabilidad de la variante isotópica agregada más pesada. Los coeficientes  $D_{i,j}^\alpha$  se calculan en este caso de acuerdo con la ecuación (8) para las  $\binom{N}{2}$  variantes isotópicas agregadas más ligeras como una suma de las potencias positivas de las raíces de las ecuaciones polinomiales elementales y extendidas para las  $\binom{N}{2}$  variantes isotópicas agregadas más pesadas.

En una modalidad preferida, las potencias de dichas raíces  $r_{\alpha', i_{\alpha'}}^{-(i-j)}$  y/o  $s_{\alpha', i_{\alpha'}}^{-(i-j)}$  en el cálculo de dichos coeficientes  $D_{i,j}^\alpha$  se calculan recursivamente, es decir, en un cálculo recursivo, la  $(i - j - 1)$ -ésima potencia negativa de dichas raíces ya se ha obtenido y puede almacenarse para su uso posterior cuando se calcula la  $(i - j)$ -ésima potencia negativa:

$$r_{\alpha', i_{\alpha'}}^{-(i-j)} = r_{\alpha', i_{\alpha'}}^{-(i-j-1)} r_{\alpha', i_{\alpha'}}^{-1} \quad y \quad s_{\alpha', i_{\alpha'}}^{-(i-j)} = s_{\alpha', i_{\alpha'}}^{-(i-j-1)} s_{\alpha', i_{\alpha'}}^{-1}$$

Es evidente que tal procedimiento reduce el tiempo de cálculo considerablemente para las potencias grandes.

En un aspecto adicional, la presente invención proporciona un dispositivo capaz de analizar al menos una parte de una distribución isotópica de una muestra y/o de una molécula en una muestra, mediante el uso de un método para analizar al menos una parte de una distribución isotópica de una muestra como se describe en este documento. Esto incluye, pero no se limita a, una computadora y un espectrómetro de masa.

En aún otro aspecto, la presente invención proporciona un dispositivo instalable en computadora que comprende una implementación de un método para analizar al menos una parte de una distribución isotópica de una muestra como se describe en este documento. Tal dispositivo instalable en computadora puede ser un disco duro, un disquete, una memoria USB u otro tipo de memoria flash, una conexión de internet a una base de datos o sitio web donde está disponible una implementación de tal método.

En aún otro aspecto, la presente invención proporciona un uso de un método para analizar al menos una parte de una distribución isotópica de una muestra como se describe en este documento, preferentemente, en una computadora y/o un espectrómetro de masa y/o un instrumento periférico de un espectrómetro de masa y/o cualquier otro aparato.

En un aspecto adicional, la presente invención proporciona un método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, de manera que dicha molécula tiene una fórmula química  $\Pi_{\alpha}(Z_{\alpha})_{n_{\alpha}}$  con

- $\alpha$ , un índice que recorre un número de especies atómicas constituyentes esperadas de dicha molécula;
  - $Z_{\alpha}$ , la  $\alpha$ -ésima especie atómica esperada de la molécula; y
  - $n_{\alpha}$ , el número de átomos de la especie  $Z_{\alpha}$  en la molécula,
- que comprende las etapas de:
- obtener al menos una parte de una distribución isotópica a partir de dicha muestra;
  - obtener un conjunto de  $N_p$  alturas de los picos  $t_j$  con  $j = 1 \dots N_p$ , de al menos una parte de una distribución isotópica agregada a partir de dicha distribución isotópica;
  - calcular a partir de dicho conjunto de alturas de los picos un conjunto de  $N_p - 1$  alturas relativas de los picos  $\hat{t}_i$  con  $i = 1 \dots N_p - 1$ ;
  - obtener un valor de  $n_{\alpha}$  para al menos una especie atómica  $Z_{\alpha}$  a partir de dicho conjunto de alturas relativas de los picos,

**caracterizado porque** dicho valor de  $n_{\alpha}$  se obtiene al calcular una solución de un sistema de ecuaciones lineales  $\sum_{\alpha} E_{\alpha} n_{\alpha} = F_i$ , preferentemente, esta solución puede redondearse a los enteros más próximos o la solución puede calcularse mediante el uso de un algoritmo de optimización discreta, de manera que el conjunto de números  $F_i$  comprende dicho conjunto de alturas relativas de los picos y los coeficientes  $E_{\alpha}$  de dicho sistema lineal comprenden las potencias y/o las sumas de las potencias de las raíces  $r_{\alpha, i_{\alpha}}$ , con  $i_{\alpha} = 1 \dots N_{\alpha}$ , de una ecuación polinomial elemental de  $N_{\alpha}$ -ésimo orden cuyo coeficiente del término de  $m_{\alpha}$ -ésimo orden se determina por una abundancia elemental  $P({}^{A_{\alpha}+m_{\alpha}}Z_{\alpha})$  del isótopo de la especie de átomo  $Z_{\alpha}$  con  $m_{\alpha}$  neutrones más que el isótopo estable más ligero de dicha especie de átomo que tiene un número másico  $A_{\alpha}$ . Debe quedar claro que obtener al menos un valor de  $n_{\alpha}$  para al menos una especie atómica  $Z_{\alpha}$  en la forma prescrita anteriormente reduce drásticamente el tiempo para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, ya que no se usa ningún método de ensayo y error o de ajuste como en la técnica anterior para obtener este  $n_{\alpha}$ , mediante el cual se hizo una estimación en cuanto al valor de  $n_{\alpha}$ , la distribución isotópica (agregada) para este valor estimado tuvo que calcularse o recuperarse de una base de datos y, posteriormente compararse con la distribución isotópica obtenida de la muestra, después de lo cual otra estimación para  $n_{\alpha}$  se probó y comprobó. Por lo tanto, el método de la presente invención es una clara mejora respecto a los métodos de la técnica anterior, por ejemplo, los métodos descritos en los documentos que se debatieron anteriormente.

En una modalidad, en el método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, dicho conjunto de alturas de los picos  $t_j$  se ordena de acuerdo con la masa en la que se observa una altura del pico. Esto hace que el cálculo de los coeficientes  $E_{\alpha}$  sea más fácil y más sencillo. En una modalidad preferida, no se observa ningún pico en dicha distribución isotópica o dicha distribución isotópica agregada alrededor de una masa que es esencialmente 1 Da menor que la masa en la que se observa el primero de dicho conjunto de alturas de los picos  $t_j$ , es decir  $t_1$ . De nuevo, esto hace que el cálculo de los coeficientes  $E_{\alpha}$  sea más fácil y más sencillo. En una modalidad preferida, las masas en las que se observan las alturas de los picos  $t_j$  se separan esencialmente en 1 Da. Esto también hace que el cálculo de los coeficientes  $E_{\alpha}$  sea más fácil y más sencillo.

En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, dicho conjunto de  $N_p - 1$  alturas relativas de los picos  $\hat{t}_i$  con  $i = 1 \dots N_p - 1$  se calcula recursivamente a partir de dicho conjunto de  $N_p$  alturas de los picos  $t_j$  con  $j = 1 \dots N_p$ , mediante el uso de las etapas de:

- calcular

$$\hat{t}_1 = -\frac{t_2}{t_1}; \quad (9a)$$

- calcular

$$\hat{t}_{i+1} = -\frac{1}{t_1} [(i+1)t_{i+2} + \sum_{j=1}^i \hat{t}_{i-j+1} t_{j+1}] \text{ for } i = 2 \dots N_p - 1. \quad (9b)$$

Como tal, puede calcularse fácil, rápida y exactamente las alturas de los picos normalizadas que son directamente comparables con las probabilidades  $q_j$  normalizadas a una probabilidad  $q_j$  correspondiente al pico  $t_1$ , es decir, el pico  $t_1$  se observa en el número de masa  $A_j$  al que corresponde una  $q_j$ . Además, debe notarse que la relación  $t_i/q_j$  es una

medida de la cantidad de la molécula en la muestra, es decir, a partir del valor de  $t_i/q_j$ , puede derivarse la cantidad de la molécula en la muestra. En una modalidad preferida, dichos coeficientes  $E_{i\alpha}$ , de dicho sistema lineal se dan por la suma de la (-i)-ésima potencia de las  $N_\alpha$  raíces de la ecuación polinomial elemental de  $N_\alpha$ -ésimo orden cuyo coeficiente del término de  $m_\alpha$ -ésimo orden se determina por la abundancia elemental  $P({}^{A_\alpha+m_\alpha}Z_\alpha Z_\alpha)$  del isótopo de la especie atómica  $Z_\alpha$  con  $m_\alpha$  neutrones más que el isótopo estable más ligero de dicha especie atómica que tiene un número másico  $A_\alpha$ , es

5 decir  $E_{i\alpha} = \sum_{i_\alpha=1}^{N_\alpha} r_{\alpha,i_\alpha}^{-i}$ , y de manera que dicho conjunto de números  $F_i$  se da por dicho conjunto de alturas relativas de los picos  $\hat{t}_i$ , es decir  $F_i = \hat{t}_i$ . Esta modalidad permite un cálculo de  $n_\alpha$  para al menos una y posiblemente todas las especies atómicas  $Z_\alpha$ , sin necesidad de consultar una base de datos exterior, es decir, todo lo que se necesita es el conjunto de alturas de los picos  $t_j$  con  $j = 1 \dots N_p$ , de al menos una parte de una distribución isotópica agregada a partir de una distribución isotópica obtenida de una muestra.

10 En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, dicho número  $N_p - 1$  de alturas relativas de los picos es al menos dicho número  $y$  de especies atómicas constituyentes esperadas de dicha molécula con dos o más isótopos estables. Como tal, se tiene la posibilidad de calcular los valores  $n_\alpha$  para todas las especies atómicas mediante el uso de las técnicas rápidas y exactas descritas en este documento.

15 En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, dicho número  $N_p - 1$  de las alturas relativas de los picos es mayor que el número de especies atómicas constituyentes esperadas de dicha molécula con dos o más isótopos estables y los valores de  $n_\alpha$  para cada valor de  $\alpha$  se obtienen al resolver un sistema sobredeterminado de ecuaciones lineales  $\sum_\alpha E_{i\alpha} n_\alpha = F_i$ . Esta solución puede obtenerse mediante cualquier técnica anterior, por ejemplo, mediante el uso de un algoritmo de resolución de mínimos cuadrados y, preferentemente, un posterior redondeo de las soluciones resultantes a los enteros más próximos. La última técnica ofrece la posibilidad de estimar el error en las soluciones obtenidas para los valores de  $n_\alpha$ .

20 En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, dicha parte de una distribución isotópica se proporciona mediante el análisis de una medición espectrométrica de masa. Las mediciones espectrométricas de masa son, por mucho, la forma más popular de obtener una distribución isotópica de una muestra. El método presentado puede usarse en el análisis de la composición de una muestra por medio de la espectrometría de masa. Dado que el método presentado es, en muchos casos, más rápido, más exacto y más fiable que las técnicas existentes para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, es una mejora importante con respecto a las técnicas anteriores, especialmente si la muestra consiste de moléculas grandes. El método puede servir además como una métrica adicional de la calidad, de manera que la estimación de la composición elemental puede compararse con el resultado de una búsqueda en base de datos mediante el uso de Sequest o Mascot para la identificación de péptidos. Las moléculas grandes se encuentran típicamente en las muestras biológicas. Por lo tanto, en una modalidad, dicha muestra comprende las moléculas de polipéptidos. Dado que estas moléculas de polipéptidos consisten, en gran parte, de un pequeño conjunto de elementos químicos y el método presentado para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra aún requiere una elección en cuanto a qué especies atómicas pueden encontrarse en la muestra, en una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, dichas especies atómicas esperadas comprenden carbono, hidrógeno, nitrógeno, oxígeno y/o azufre. En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, dichas especies atómicas esperadas comprenden hidrógeno, helio, litio, berilio, boro, carbono, nitrógeno, oxígeno, flúor, neón, sodio, magnesio, aluminio, silicio, fósforo, azufre, cloro, argón, potasio, calcio, escandio, titanio, vanadio, cromo, manganeso, hierro, cobalto, níquel, cobre, cinc, galio, germanio, arsénico, selenio, bromo, kriptón, rubidio, estroncio, itrio, circonio, niobio, molibdeno, tecnecio, rutenio, rodio, paladio, plata, cadmio, indio, estaño, antimonio, telurio, yodo, xenón, cesio, bario, lantano, cerio, praseodimio, neodimio, prometio, samario, europio, gadolinio, terbio, disprosio, holmio, erbio, tulio, iterbio, lutecio, hafnio, tantalio, tungsteno, renio, osmio, iridio, platino, oro, mercurio, talio, plomo, bismuto, polonio, ástato, radón, francio, radio, actinio, torio, protactinio y/o uranio.

55 En una modalidad del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, dicha molécula comprende al menos un átomo constituyente monoisotópico en su fórmula química, y el método comprende además una combinación de las etapas de:

- obtener a partir de dicha distribución isotópica una masa medida, dentro de una precisión de  $\pm 0.9$  Da,  $\pm 0.8$  Da,  $\pm 0.75$  Da,  $\pm 0.6$  Da,  $\pm 0.5$  Da,  $\pm 0.4$  Da,  $\pm 0.3$  Da,  $\pm 0.25$  Da,  $\pm 0.1$  Da, en la que aparece un pico de isótopo;

- calcular una masa esperada en la que aparece dicho pico de isótopo para una segunda molécula con una segunda fórmula química que es dicha fórmula química a partir de la cual se elimina el átomo constituyente monoisotópico;
- comparar dicha masa medida con dicha masa esperada para deducir la presencia de dicha molécula que comprende al menos un átomo constituyente monoisotópico en dicha muestra.

5

Una de las ventajas del presente método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, es que puede derivar los valores de  $n_\alpha$  para al menos uno y posiblemente cada valor de  $\alpha$  en la fórmula química de una molécula, en esta sólo a partir del tamaño del pico relativo de al menos una parte de una distribución isotópica agregada. No se requieren las masas exactas y/o centrales en las que se observan estos picos. Sin embargo, este método sólo puede identificar los valores de  $n_\alpha$  para las especies atómicas con al menos dos isótopos estables. Cuando una molécula comprende al menos una especie atómica monoisotópica, esta especie no altera la distribución isotópica o la distribución isotópica agregada con respecto a las alturas de los picos. Sólo resultará en un cambio de las masas en la que se observan los picos. Por lo tanto, cuando se emplea el método descrito aquí anteriormente, puede calcularse qué cantidad de la masa de una molécula de la que se mide u observa una distribución isotópica surge de las especies atómicas monoisotópicas que constituyen la molécula. Para ello, hay que comparar al menos una masa en la que se observa una variante isotópica agregada con la masa de esta variante isotópica agregada calculada a partir de la fórmula química de la molécula, como se deriva al usar un método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describe en este texto. De esta manera, la diferencia de masa puede atribuirse a las especies atómicas no monoisotópicas comprendidas en la molécula presente en la muestra. La diferencia de masa así obtenida puede compararse con las masas de las especies atómicas monoisotópicas. En una modalidad preferida del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, dicho al menos único átomo constituyente monoisotópico es el fósforo-31. En otra modalidad preferida, dicha diferencia de masa se atribuye a una combinación de los siguientes elementos monoisotópicos presentes en la molécula: berilio-9, flúor-19, sodio-23, aluminio-27, fósforo-31, escandio-45, vanadio-51, manganeso-55, cobalto-59, arsénico-75, rubidio-85, itrio-89, niobio-93, rodio-103, indio-113, yodo-127, cesio-133, lantano-139, praseodimio-141, europio-153, terbio-159, holmio-165, tulio-169, lutecio-175, renio-185, oro-197.

10

15

20

25

30

35

40

45

50

En un aspecto adicional, la presente invención proporciona un dispositivo capaz de identificar la presencia, composición elemental y/o cantidad de una molécula en una muestra mediante el uso del método como se describe anteriormente. Preferentemente, este dispositivo comprende una instalación espectrométrica de masa o es un espectrómetro de masa con aparato de análisis correspondiente. El dispositivo puede comprender además una computadora implementada con el método descrito anteriormente.

En aún otro aspecto, la presente invención proporciona un dispositivo instalable en computadora que comprende una implementación del método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describe en este documento. Tal dispositivo instalable en computadora puede ser un disco duro, un disquete, una memoria USB u otro tipo de memoria flash, una conexión de internet a una base de datos o sitio web donde está disponible una implementación de tal método.

En aún otro aspecto, la presente invención proporciona un uso de un método para identificar la composición elemental y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra como se describe en este documento, preferentemente, en una computadora y/o un espectrómetro de masa y/o un instrumento periférico de un espectrómetro de masa y/o cualquier otro aparato.

Se supone que la presente invención no se limita a cualquier modalidad descrita anteriormente y que algunas modificaciones pueden añadirse a la metodología y/o dispositivos presentados sin la revaluación de las reivindicaciones adjuntas. Por ejemplo, la presente invención se ha descrito en este documento principalmente en referencia a las moléculas de polipéptidos, pero está claro que la invención puede aplicarse a todo tipo de moléculas.

## EJEMPLOS

55

Ejemplo 1:

Para el monóxido de carbono, CO, puede elegirse  $A_0 = 28$ , es decir, la variante monoisotópica más ligera. Entonces fácilmente se calcula  $q_0 = P(^{12}\text{C})P(^{16}\text{O}) = 0.986896001$ . Para calcular la probabilidad  $q_1$  con la que ocurre la variante isotópica agregada de CO con número de masa  $A_1 = 29$ , puede usarse  $q_1 = \sum_{j=0}^1 C_{1,j} q_0^j$  con  $C_{1,0} = -\frac{p-1}{1}$

y  $\rho_{-1} = r_c^{-1} + r_o^{-1} + \bar{r}_o^{-1}$ , de manera que  $r_c = -\frac{P(^{12}C)}{P(^{13}C)}$ ,  $r_o = \frac{-P(^{16}O) + \sqrt{P(^{16}O)^2 - 4P(^{16}O)P(^{18}O)}}{2P(^{18}O)}$  y  $\bar{r}_o = \frac{-P(^{16}O) - \sqrt{P(^{16}O)^2 - 4P(^{16}O)P(^{18}O)}}{2P(^{18}O)}$  son las raíces de las ecuaciones  $P(^{12}C) + P(^{13}C)I = 0$  y  $P(^{16}O) + P(^{17}O)I + P(^{18}O)I^2 = 0$ . Esto conduce a  $\rho_{-1} = -0.011196653942$  y  $q_1 = 0.011049933$ . En las siguientes etapas, se obtiene  $\rho_{-2} = r_c^{-2} + r_o^{-2} + \bar{r}_o^{-2} = -0.003992862186$  y  $q_2 = 0.002032131$ , y  $q_2 = 0.002032131$ , y  $\rho_{-3} = r_c^{-3} + r_o^{-3} + \bar{r}_o^{-3} = 0.000001083120$  y  $q_3 = 0.000021935$ . Estos números se resumen en la siguiente tabla:

$i$	$A_i$	$\rho_{-i} = r_c^{-i} + r_o^{-i} + \bar{r}_o^{-i}$	$q_i$
0	28	-	0.986896001
1	29	-0.011196653942	0.011049933
2	30	-0.003992862186	0.002032131
3	31	0.000001083120	0.000021935

Para  $C_3H_8$ , puede construirse una tabla similar:

$i$	$A_i$	$\rho_{-i} = r_c^{-i} + r_o^{-i} + \bar{r}_o^{-i}$	$q_i$
0	44	-	0,967351820549
1	45	-0,033367290690	0,032277909396
2	46	0,000351045760	0,000368720815
3	47	-0,000003795683	0,000001547984
4	48	0,000000041053	0,000000001255
5	49	-0,000000000444	0,000000000000
6	50	0,000000000005	0,000000000000
...			

En esta tabla se ve que sólo los primeros pocos picos pueden observarse en una medición de MS.

Ejemplo 2: Ejemplo del cálculo de las masas centrales de las variantes isotópicas agregadas de CO.

Se proporciona una distribución isotópica de una muestra, y quiere analizarse la muestra para el CO. De esta manera, quiere compararse una parte de la distribución isotópica medida con la distribución isotópica agregada, también con respecto a las masas centrales. De esta manera, se necesita obtener esta distribución isotópica agregada y las masas centrales. Después, puede hacerse la comparación entre la distribución isotópica agregada y las masas centrales obtenidas, y la distribución isotópica agregada derivada de la distribución isotópica medida.

Anteriormente, ya se ha explicado la distribución isotópica de la molécula de CO.

Para la variante más ligera, que puede tomarse como la variante de partida o cero, se tiene:

- $A_0 = A(^{12}C) + A(^{16}O) = 28$ ,
- $M_0 = M(^{12}C) + M(^{16}O) = 27.995\text{Da}$ ,
- $q_0 = P(^{12}C)P(^{16}O) = 0.986896001$ .

Se hace notar que, dado que es una variante monoisotópica, la masa central es además la masa isotópica.

Obviamente, la masa central de una variante monoisotópica se calcula fácilmente mediante la suma de las masas de los isótopos constituyentes, es decir,  $M_0 = M(^{12}_6\text{C}) + M(^{16}_8\text{O}) = 27.995\text{Da}$ . Sin embargo, se ilustrará a continuación que el método descrito en este texto puede aplicarse además a las variantes monoisotópicas.

El isótopo  $^{12}_6\text{C}$  ocurre en la variante isotópica (agregada) cero del CO con probabilidad = 1. Por lo tanto, la masa promedio ponderada de la especie atómica carbono en la variante isotópica más ligera del CO se determina por

10 •  $p_0^C = P(^{12}_6\text{C})P(^{16}_8\text{O})M(^{12}_6\text{C}) = 11.842752012\text{Da},$

debido a que esta masa promedio ponderada debería normalizarse a la probabilidad  $q_0 = P(^{12}_6\text{C})P(^{16}_8\text{O})$ , es decir, si las masas de todos los isótopos atómicos fueran 1, la masa promedio ponderada  $p_0^C$  sería igual a la probabilidad  $q_0$ .

Un cálculo similar para el O conduce a

15 •  $p_0^O = P(^{12}_6\text{C})P(^{16}_8\text{O})M(^{16}_8\text{O}) = 15.785317274\text{Da}.$

La masa central para la variante isotópica agregada más ligera entonces se calcula fácilmente como

•  $M_0 = \frac{p_0^C + p_0^O}{q_0} = \frac{(^{12}_6\text{C})P(^{16}_8\text{O})M(^{12}_6\text{C}) + P(^{12}_6\text{C})P(^{16}_8\text{O})M(^{16}_8\text{O})}{P(^{12}_6\text{C})P(^{16}_8\text{O})} = M(^{12}_6\text{C}) + M(^{16}_8\text{O}) = 27.995\text{Da}.$

20 El cálculo de la masa central de la variante isotópica agregada con número de masa molecular 29 es un poco más complicado. Se tienen dos variantes isotópicas contribuyentes.  $^{13}\text{C}^{16}\text{O}$  con

• abundancia  $p_{1,0} = P(^{13}_6\text{C})P(^{16}_8\text{O}) = 0.010673999$

• masa  $m_{1,0} = M(^{13}_6\text{C}) + M(^{16}_8\text{O}) = 28.9982695\text{Da}$

y  $^{12}\text{C}^{17}\text{O}$  con

25 • abundancia  $p_{1,1} = P(^{12}_6\text{C})P(^{17}_8\text{O}) = 0.000375934$

$m_{1,1} = M(^{12}_6\text{C}) + M(^{17}_8\text{O}) = 28.9991317\text{Da}$

• masa

y abundancia total

•  $q_1 = p_{1,0} + p_{1,1} = 0.011049933.$

30 El isótopo  $^{13}_6\text{C}$  ocurre en la primera variante isotópica agregada del CO con una probabilidad directamente proporcional a la abundancia relativa  $p_{1,1}$  con la que la variante isotópica  $^{12}\text{C}^{17}\text{O}$  contribuye a la primera variante isotópica agregada. El isótopo  $^{12}_6\text{C}$  ocurre en la primera variante isotópica agregada del CO con una probabilidad directamente proporcional a la abundancia relativa  $p_{1,0}$  con la que la variante isotópica  $^{13}\text{C}^{16}\text{O}$  contribuye a la primera variante isotópica agregada. Por lo tanto, la masa promedio ponderada de la especie atómica carbono en la primera variante isotópica agregada del CO se da por

35 •  $p_1^C = p_{1,0}M(^{13}_6\text{C}) + p_{1,1}M(^{12}_6\text{C}) = P(^{13}_6\text{C})P(^{16}_8\text{O})M(^{13}_6\text{C}) + P(^{12}_6\text{C})P(^{17}_8\text{O})M(^{12}_6\text{C}) = 0.14331\text{Da}$

Un cálculo similar para el O conduce a

40 •  $p_1^O = p_{1,0}M(^{16}_8\text{O}) + p_{1,1}M(^{17}_8\text{O}) = P(^{13}_6\text{C})P(^{16}_8\text{O})M(^{16}_8\text{O}) + P(^{12}_6\text{C})P(^{17}_8\text{O})M(^{17}_8\text{O}) = 0.17712\text{Da}$

Se hace notar nuevamente, que si se ponen todas las masas de todos los isótopos iguales a uno, las masas promedio ponderadas tanto de la especie carbono como de la especie oxígeno serían iguales a  $q_1$ , es decir, las masas promedio ponderadas se normalizan a  $q_1$ .

45 La masa central para la primera variante isotópica agregada se calcula entonces como

$$\bullet M_1 = \frac{p_1^C + p_1^O}{q_1} = \frac{P(^{13}C)P(^{16}O)[M(^{13}C) + M(^{16}O)] + P(^{12}C)P(^{17}O)[M(^{12}C) + M(^{17}O)]}{P(^{13}C)P(^{16}O) + P(^{12}C)P(^{17}O)} = \frac{p_{1,0}m_{1,0} + p_{1,1}m_{1,1}}{q_1} =$$

28.99838888Da,

que claramente es un valor entre las masas de las dos variantes isotópicas, más próximo a la masa de la variante más prominente  $^{13}C^{16}O$ .

5 Para la segunda variante isotópica agregada, se obtienen las masas promedio ponderadas, por ejemplo, al tomar una combinación lineal de las masas promedio ponderadas ya calculadas  $p_2^{C,O} = \sum_{j=0}^1 D_{2,j}^{C,O} p_j^{C,O}$ , de manera que los coeficientes  $D_{2,j}^{C,O}$  se calculan de la siguiente manera:

10

$$\bullet D_{2,0}^C = -\frac{1}{2}[(r_C^{-2} + r_O^{-2} + \bar{r}_O^{-2}) - r_C^{-2} + s_C^{-2}]$$

$$\bullet D_{2,1}^C = -\frac{1}{2}[(r_C^{-1} + r_O^{-1} + \bar{r}_O^{-1}) - r_C^{-1} + s_C^{-1}]$$

15

$$\bullet D_{2,0}^O = -\frac{1}{2}[(r_C^{-2} + r_O^{-2} + \bar{r}_O^{-2}) - r_O^{-2} + s_O^{-2}]$$

$$\bullet D_{2,1}^O = -\frac{1}{2}[(r_C^{-1} + r_O^{-1} + \bar{r}_O^{-1}) - r_O^{-1} + s_O^{-1}]$$

Mediante el uso de las ecuaciones (EL. 10) y (EL.11) para las raíces  $r_{C,O}$  y expresiones similares (con el cambio de las abundancias isotópicas atómicas en (EL.10) y (EL.11), por el producto de las abundancias isotópicas atómicas y las masas isotópicas atómicas) para las raíces  $s_{C,O}$ , se obtiene

20

$$\bullet p_2^C = D_{2,0}^C p_0^C + D_{2,1}^C p_1^C = P(^{13}C)P(^{17}O)M(^{13}C) + P(^{12}C)P(^{18}O)M(^{12}C) = 0.024389652\text{Da},$$

$$\bullet p_2^O = D_{2,0}^O p_0^O + D_{2,1}^O p_1^O = P(^{13}C)P(^{16}O)M(^{17}O) + P(^{12}C)P(^{18}O)M(^{18}O) = 0.036572587\text{Da},$$

con la probabilidad

25

$$\bullet q_2 = P(^{13}C)P(^{17}O) + P(^{12}C)P(^{18}O) = 0.002032131.$$

Esto da para la masa central

$$\bullet M_2 = \frac{p_2^C + p_2^O}{q_2} = 29.9991678\text{Da}.$$

Para la tercera variante isotópica agregada, que es la más pesada y por lo tanto una variante monoisotópica, aún puede usarse el mismo procedimiento que el anterior, lo que conduce a  $p_3^{C,O} = \sum_{j=0}^2 D_{3,j}^{C,O} p_j^{C,O}$  con

30

$$\bullet D_{3,0}^C = -\frac{1}{3}[(r_C^{-3} + r_O^{-3} + \bar{r}_O^{-3}) - r_C^{-3} + s_C^{-3}]$$

35

$$\bullet D_{3,1}^C = -\frac{1}{3}[(r_C^{-2} + r_O^{-2} + \bar{r}_O^{-2}) - r_C^{-2} + s_C^{-2}]$$

$$\bullet D_{3,2}^C = -\frac{1}{3}[(r_C^{-1} + r_O^{-1} + \bar{r}_O^{-1}) - r_C^{-1} + s_C^{-1}]$$

40

$$\bullet D_{3,0}^O = -\frac{1}{3}[(r_C^{-3} + r_O^{-3} + \bar{r}_O^{-3}) - r_O^{-3} + s_O^{-3}]$$

$$\bullet D_{3,1}^O = -\frac{1}{3}[(r_C^{-2} + r_O^{-2} + \bar{r}_O^{-2}) - r_O^{-2} + s_O^{-2}]$$

45

$$\bullet D_{3,2}^O = -\frac{1}{3}[(r_C^{-1} + r_O^{-1} + \bar{r}_O^{-1}) - r_O^{-1} + s_O^{-1}]$$

Esto conduce a

$$\bullet p_3^C = D_{3,0}^C p_0^C + D_{3,1}^C p_1^C + D_{3,2}^C p_2^C = P(^{13}C)P(^{18}O)M(^{13}C) = 0.0002852286\text{Da},$$

$$\bullet p_3^O = D_{3,0}^O p_0^O + D_{3,1}^O p_1^O + D_{3,2}^O p_2^O = P(^{13}C)P(^{18}O)M(^{18}O) = 0.0003948116\text{Da},$$

50

con la probabilidad

- $q_3 = P(^{13}_6C)P(^{18}_8O) = 0.000021935.$

Esto da para la masa central

- $M_3 = \frac{p_3^C + p_3^O}{q_3} = 31.00252\text{Da},$

5 que puede observarse que es igual a la masa de la variante de isótopos más pesada  $^{13}_6C^{18}_8O$ :

$$M(^{13}_6C) + M(^{18}_8O).$$

Ejemplo 3: Ejemplo del cálculo de la fórmula química cuando se proporciona una distribución isotópica.

10 La figura 1 presenta la distribución isotópica agregada de una molécula en una muestra según lo observado mediante la espectrometría de masa, de manera que los picos de las variantes isotópicas agregadas se indican con puntos (1), y de manera que la distribución en los valores de m/z inferiores (2) puede atribuirse a las moléculas no fosforiladas, mientras que la distribución observada en los valores de m/z superiores (3) puede atribuirse a las moléculas fosforiladas.

15 La molécula bajo escrutinio en la figura 1 es la beta-caseína con una masa monoisotópica de 23568.3197 Da para las moléculas no fosforiladas. La molécula tiene una carga de z igual a 8. Por lo tanto, la masa monoisotópica observada en términos de m/z es igual a (23568.3197+8)/8. El cálculo invertido puede aplicarse para calcular la masa monoisotópica basado en las m/z y z observadas. Por ejemplo, el pico monoisotópico observado se observa en 2947.04779227045 m/z. La carga de las moléculas es 8, por lo tanto  $2947.04779227045 \times 8 = 23568.3823381636$  Da es la masa de la variante monoisotópica más ligera. La distribución alrededor de m/z = 2949 (2) se provoca por la variante no fosforilada de la betacaseína mientras que la distribución alrededor de m/z = 2959 (3) se provoca por la variante fosforilada. Una única fosforilación provoca un desplazamiento de masa de 80 dalton debido a la adición de  $\text{HPO}_3$ . A partir de la distribución observada (2) las alturas de los picos y sus rangos en la distribución agregada se extraen mediante el uso de un algoritmo de extracción de picos de acuerdo con la presente invención. Los picos detectados se indican por

25 puntos (1) y el resultado se enumera a continuación:

$$\begin{aligned} t_1 &= 100000 \\ t_2 &= 1303159.40498057 \\ 30 \quad t_3 &= 8575419.36617413 \\ &\dots \\ t_{50} &= 2.39573123747609 \end{aligned}$$

35 Estas alturas de los picos pueden transformarse a números tipo fracción mediante el uso de las ecuaciones (9a) y (9b) de la siguiente manera:

$$\begin{aligned} \hat{f}_1 &= -1/t_1 \times t_2 = -13.0315940498057 \\ \hat{f}_2 &= -1/t_1 \times (2xt_3 + \hat{f}_1xt_2) = -1.6859438445517 \\ 40 \quad \hat{f}_3 &= -1/t_1 \times (3xt_4 + \hat{f}_2xt_2 + \hat{f}_1xt_3) = 0.00585728955879807 \\ &\dots \\ \hat{f}_{49} &= -1/t_1 \times (49xt_{50} + \hat{f}_{48}xt_2 + \hat{f}_{47}xt_3 + \dots + \hat{f}_1xt_{49}) = 1.3669076173386e-008 \end{aligned}$$

45 Mientras tanto basado en la información del rango y las raíces de los cinco elementos supuestos (C, H, N, O, S) se construye una matriz **A** de 49x5. El ejemplo 1 explica cómo pueden calcularse las raíces. Las cinco columnas de la matriz **A** contienen una suma de las potencias de las raíces de C, H, N, O y S, respectivamente. La potencia de la suma de potencias se determina por el índice de la fila. Por ejemplo, matriz **A** =

50

$$\begin{aligned} \text{Fila 1: } & [r_C^{-1}, r_H^{-1}, r_N^{-1}, r_O^{-1} + \bar{r}_O^{-1}, r_{1,S}^{-1} + \bar{r}_{1,S}^{-1} + r_{2,S}^{-1} + \bar{r}_{2,S}^{-1}] \\ \text{Fila 2: } & [r_C^{-2}, r_H^{-2}, r_N^{-2}, r_O^{-2} + \bar{r}_O^{-2}, r_{1,S}^{-2} + \bar{r}_{1,S}^{-2} + r_{2,S}^{-2} + \bar{r}_{2,S}^{-2}] \\ \text{Fila 3: } & [r_C^{-3}, r_H^{-3}, r_N^{-3}, r_O^{-3} + \bar{r}_O^{-3}, r_{1,S}^{-3} + \bar{r}_{1,S}^{-3} + r_{2,S}^{-3} + \bar{r}_{2,S}^{-3}] \\ & \dots \\ \text{Fila 49: } & [r_C^{-49}, r_H^{-49}, r_N^{-49}, r_O^{-49} + \bar{r}_O^{-49}, r_{1,S}^{-49} + \bar{r}_{1,S}^{-49} + r_{2,S}^{-49} + \bar{r}_{2,S}^{-49}] \end{aligned}$$

55

La solución del sistema de ecuaciones,  $\hat{t} = \mathbf{A.M}$ , de manera que  $\mathbf{M}$  es una matriz columna de los valores  $n_{\alpha}$  para las especies atómicas  $\alpha = \text{C, H, N, O, S}$ , da una estimación de la composición elemental de la muestra analizada mediante la espectrometría de masa. En este caso se resuelve el sistema de ecuaciones con un método de mínimos cuadrados y redondeo consecutivo. Esto resulta en las siguientes estimaciones:

5  
 $n_{\text{C}} = 1080$   
 $n_{\text{H}} = 1699$   
 $n_{\text{N}} = 268$   
 $n_{\text{O}} = 310$   
 $n_{\text{S}} = 6$

10  
 Debe notarse que si se usara como referencia una altura de pico o una combinación lineal de las alturas de los picos diferente, es decir, cuando  $t_1$  se reemplace por otro número, la misma solución para los valores de  $n_{\alpha}$  se hubieran obtenido trivialmente mediante el uso de una matriz  $\mathbf{A}$  reescalada.

15  
 Basado en los resultados anteriores la masa monoisotópica se calcula como se explica en el ejemplo 2. Esto conduce a 23575.3745131315 Da. Debido a que la molécula tiene una carga de ocho, hay que restar este número de átomos de hidrógeno. Esto da 23567.3745131315 Da. Debe notarse que debido a la optimización simplista, la variación de la señal y el redondeo más próximo ha ocurrido una pequeña estimación incorrecta de un átomo de hidrógeno.

20  
 Ejemplo 4: Ejemplo de la observación de la presencia de un átomo de fósforo en una molécula de la que se proporciona la distribución isotópica.

25  
 El procedimiento explicado en el ejemplo 3 se repite para la distribución más pesada (3). Esto resulta en las estimaciones:

30  
 $n_{\text{C}} = 1080$   
 $n_{\text{H}} = 1699$   
 $n_{\text{N}} = 268$   
 $n_{\text{O}} = 313$   
 $n_{\text{S}} = 6$

35  
 La masa monoisotópica se calcula mediante el uso de los resultados anteriores como 23623.3592569315 Da. Después de la resta de los ocho átomos de hidrógeno, se obtiene: 23615.3592569315 Da. La masa monoisotópica de la distribución más pesada (3) se observa en 2957.04358337446 m/z o equivalentemente en una masa de 2957.04358337446x8-8 = 23648.3486669957 Da. La diferencia entre la masa observada y la masa estimada basada en la distribución observada es 23648.3486669957 - 23615.3592569315 = 32.9894100641977. Nuevamente ocurre un pequeño error en la estimación del número de átomos de hidrógeno. En este caso, faltaron dos átomos de hidrógeno. Sin embargo, la discrepancia entre la masa observada y estimada es demasiado grande para explicarse por errores de redondeo. Por lo tanto, se concluye que la diferencia de masa se debe a la presencia de un elemento monoisotópico. El elemento más próximo a 32.9894100641977 Da es el fósforo con una masa de 30.97376 Da.

40

Reivindicaciones

- 5 1. Método para identificar la composición elemental de una molécula y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra, de manera que dicha molécula tiene una fórmula química  $\Pi_{\alpha}(Z_{\alpha})n_{\alpha}$  con
- $\alpha$ , un índice que recorre un número de especies atómicas constituyentes esperadas de dicha molécula;
  - $Z_{\alpha}$ , la  $\alpha$ -ésima especie atómica esperada de la molécula; y
  - $n_{\alpha}$ , el número de átomos de la especie  $Z_{\alpha}$  en la molécula,
- que comprende las etapas de:
- 10 • obtener al menos una parte de una distribución isotópica a partir de dicha muestra;
- calcular un conjunto de  $N_p$  alturas de los picos  $t_j$  con  $j = 1 \dots N_p$ , de al menos una parte de una distribución isotópica agregada a partir de dicha distribución isotópica;
- calcular a partir de dicho conjunto de alturas de los picos un conjunto de  $N_p - 1$  relaciones de las alturas de los picos  $\hat{t}_i$  con  $i = 1 \dots N_p - 1$ ;
- 15 • obtener un valor de  $n_{\alpha}$  para al menos una especie atómica  $Z_{\alpha}$  a partir de dicho conjunto de alturas relativas de los picos,
- en donde dicho valor de  $n_{\alpha}$  se obtiene al calcular una solución de un sistema de ecuaciones lineales  $\sum_{\alpha} E_{\alpha} n_{\alpha} = F_i$ , de manera que el conjunto de números  $F_i$  comprende dicho conjunto de alturas relativas de los picos y los coeficientes  $E_{\alpha}$  de dicho sistema lineal comprenden las potencias y/o las sumas de las potencias de las raíces  $r_{\alpha, i_{\alpha}}$ , con  $i_{\alpha} = 1 \dots N_{\alpha}$ , de una ecuación polinomial elemental de  $N_{\alpha}$ -ésimo orden cuyo coeficiente del término de  $m_{\alpha}$ -ésimo orden se
- 20 determina por una abundancia elemental  $P^{(A_{\alpha} + m_{\alpha} Z_{\alpha})}$  del isótopo de la especie de átomo  $Z_{\alpha}$  con  $m_{\alpha}$  neutrones más que el isótopo estable más ligero de dicha especie de átomo que tiene el número másico  $A_{\alpha}$ , de manera que las masas en las que se observan las alturas de los picos  $t_j$  se separan esencialmente en 1 Da, de manera que dicho conjunto de  $N_p - 1$  relaciones de las alturas de los picos  $\hat{t}_i$  con  $i = 1 \dots N_p - 1$  se calcula recursivamente a partir de dicho conjunto de  $N_p$  alturas de los picos  $t_j$  con  $j = 1 \dots N_p$ , mediante el uso de las etapas de:
- Calcular  $\hat{t}_1 = -\frac{t_2}{t_1}$ ,
  - Calcular  $\hat{t}_{i+1} = -\frac{1}{t_1} [(i+1)t_{i+2} + \sum_{j=1}^i \hat{t}_{i-j+1} t_{j+1}]$  for  $i = 2 \dots N_p - 1$ .
- 30 2. Método para identificar la composición elemental de una molécula y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra de acuerdo con la reivindicación 1, de manera que dicho conjunto de alturas de los picos  $t_j$  se ordena de acuerdo con la masa en la que se observa una altura del pico, de manera que no se observa ningún pico en dicha distribución isotópica o dicha distribución isotópica agregada alrededor de una masa que es esencialmente 1 Da menor que la masa en la que se observa el primero de dicho conjunto de alturas de los picos  $t_i$ , es decir,  $t_1$ , de manera que dichos coeficientes  $E_{\alpha}$  de dicho sistema lineal se dan por la suma de la (-j)-ésima potencia de las  $N_{\alpha}$  raíces de la ecuación polinomial elemental de  $N_{\alpha}$ -ésimo orden cuyo coeficiente del
- 35 término de  $m_{\alpha}$ -ésimo orden se determina por la abundancia elemental  $P^{(A_{\alpha} + m_{\alpha} Z_{\alpha})}$  del isótopo de la especie atómica  $Z_{\alpha}$  con  $m_{\alpha}$  neutrones más que el isótopo estable más ligero de dicha especie atómica que tiene un número másico  $A_{\alpha}$ , es decir  $E_{i\alpha} = \sum_{i_{\alpha}=1}^{N_{\alpha}} r_{\alpha, i_{\alpha}}^{-i}$ , y de manera que dicho conjunto de números  $F_i$  se da por dicho conjunto de relaciones de las alturas de los picos  $\hat{t}_i$ , es decir,  $F_i = \hat{t}_i$ , y de manera que dicho número  $N_p - 1$  de las relaciones de las alturas de los picos es al menos dicho número de especies atómicas constituyentes esperadas de dicha molécula con dos o más isótopos estables.
- 40
- 45 3. Método para identificar la composición elemental de una molécula y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra de acuerdo con cualquiera de las reivindicaciones 1 a la 2, de manera que dichas especies atómicas esperadas comprenden carbono, hidrógeno, nitrógeno, oxígeno y/o azufre.
- 50 4. Método para identificar la composición elemental de una molécula y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra de acuerdo con cualquiera de las reivindicaciones 1 a la 3, de manera que dicha molécula comprende al menos un átomo constituyente monoisotópico en su fórmula química, que comprende además las etapas de:
- obtener a partir de dicha distribución isotópica una masa medida, dentro de una precisión de  $\pm 0.5$  Da, en la que aparece un pico de isótopo;
  - calcular una masa esperada en la que aparece dicho pico de isótopo para una segunda molécula con una segunda fórmula química que es dicha fórmula química a partir de la cual se elimina el átomo constituyente monoisotópico;
- 55

- comparar dicha masa medida con dicha masa esperada para deducir la presencia de dicha molécula que comprende al menos un átomo constituyente monoisotópico en dicha muestra.  
de manera que al menos dicho átomo constituyente monoisotópico es el fósforo.

- 5     **5.** Dispositivo adaptado para identificar la composición elemental de una molécula y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra mediante el uso del método de acuerdo con cualquiera de las reivindicaciones 1 a la 4.
- 10    **6.** Dispositivo de acuerdo con la reivindicación 5, que comprende una instalación espectrométrica de masa y una unidad de cálculo conectada funcionalmente a la instalación espectrométrica de masa, la unidad de cálculo dispuesta para identificar la composición elemental de una molécula y/o cuantificar la presencia de elementos monoisotópicos en una molécula en una muestra mediante el uso del método de acuerdo con cualquiera de las reivindicaciones 1 a la 4.
- 15    **7.** Dispositivo instalable en computadora que comprende una implementación del método de acuerdo con cualquiera de las reivindicaciones 1 a la 4.

