

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 755 734**

51 Int. Cl.:

H01J 49/00 (2006.01)

C12Q 1/04 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.11.2012 PCT/IB2012/056860**

87 Fecha y número de publicación internacional: **06.06.2013 WO13080170**

96 Fecha de presentación y número de la solicitud europea: **30.11.2012 E 12798884 (8)**

97 Fecha y número de publicación de la concesión europea: **02.10.2019 EP 2786397**

54 Título: **Procedimiento de identificación de microorganismos por espectrometría de masas**

30 Prioridad:

02.12.2011 EP 11306610

02.12.2011 US 201161566025 P

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

23.04.2020

73 Titular/es:

BIOMÉRIEUX INC. (100.0%)

100 Rodolphe Street

Durham, North Carolina 27712, US

72 Inventor/es:

STRUBEL, GRÉGORY;

ARSAC, MAUD;

DESSEREE, DENIS y

COTTE-PATTAT, PIERRE-JEAN

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 755 734 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

Procedimiento de identificación de microorganismos por espectrometría de masas

5 Campo de la invención

La invención se refiere al campo de la identificación de microorganismos, y especialmente de bacterias, mediante la espectrometría de masas.

10 Estado de la técnica

Se conoce utilizar la espectrometría de masas para identificar microorganismos, y más particularmente bacterias. Se prepara una muestra del microorganismo y después se adquiere y pre-trata un espectro de masas de la muestra: supresión del ruido del espectro (eliminación del ruido), filtración del ruido de fondo (atribuible a la sobrecarga del detector). Se detectan entonces picos importantes del espectro pretratado y la lista de picos así obtenida se "analiza" y "compara" con datos de una base de conocimientos construida a partir de listas de picos típicos de un microorganismo o de un grupo de microorganismos (cepa, género, familia, etc.) identificado.

Si el principio parece simple a priori, su realización es, sin embargo, delicada. En efecto, en primer lugar, la cantidad de información contenida en un espectro de masas, especialmente el número de picos, es muy importante, lo que necesita herramientas de cálculo muy potentes para crear una base de conocimientos sólida, así como para la realización de algoritmos de clasificación, de comparación y de decisión.

Después, existe una incertidumbre de medición importante, especialmente en lo que se refiere a la localización de los picos en el espectro. Se observa, en efecto, que de una medición a otra sobre un mismo espectrómetro, así como de un espectrómetro a otro, un pico que representa una molécula dada no tiene una posición fija en los espectros medidos, o como mínimo, el pico no está contenido en un intervalo. Así, un pico de un espectro adquirido, y que corresponde a una molécula de proteína dada, puede no identificarse como correspondiente a dicha molécula de proteína por el algoritmo de clasificación. Finalmente, esta incertidumbre no es constante en el rango de relaciones masa sobre carga y aumenta a medida que esta relación aumenta.

El documento US2004/0195500 describe un procedimiento de análisis de espectro de masas de material biológico que comprende una etapa de cuantificación logarítmica del intervalo m/z.

35 Presentación de la invención

El objetivo de la invención es proponer un procedimiento que permita la identificación sólida de los microorganismos por espectrometría de masas gracias a una reducción de la masa de informaciones a analizar y a una reducción del impacto de la ausencia de precisión en la localización de los picos del espectro de masas.

Para este fin, la invención tiene por objeto un procedimiento de identificación de un microorganismo por espectrometría de masas, que comprende:

- la adquisición de al menos un espectro de masas de dicho microorganismo;
- para cada espectro de masas adquirido:
 - la detección de picos del espectro en un rango predeterminado de masas;
 - la generación de una lista de picos que cataloga como mucho un pico en cada intervalo de una subdivisión predeterminada del rango de las relaciones masa sobre carga, aumentando la anchura de los intervalos de la subdivisión con la relación masa sobre carga según las relaciones:

$$L(b) = \exp\left(\frac{b - \beta}{\alpha}\right) \times \left(\exp\left(\frac{1}{\alpha}\right) - 1\right)$$

$$\alpha = \frac{b_{\min} - (b_{\max} + 1)}{\ln m_{\min} - \ln m_{\max}}$$

$$\beta = \frac{(b_{\max} + 1) \times \ln m_{\min} - b_{\min} \times \ln m_{\max}}{\ln m_{\min} - \ln m_{\max}}$$

en las que los intervalos de la subdivisión se referencian por unos números enteros superiores a 1 desde el número entero b_{\min} , para las relaciones masas sobre cargas más bajas en el rango, hasta el número entero b_{\max} , para las relaciones masas sobre cargas más importantes en el rango, $L(b)$ es la anchura del intervalo referenciado por los

números enteros b , m_{\min} es un umbral inferior del rango de las relaciones masa sobre carga, y m_{\max} es un umbral superior del rango de las relaciones masa sobre carga; y

- 5 ▪ el análisis de la o de las listas de picos obtenidos en función de la base de conocimientos de microorganismos y/o de tipos de microorganismos previamente identificados.

10 En otras palabras, el espacio continuo de las relaciones masa sobre carga, o Thompson, está cuantificado de manera logarítmica, y se retiene solo un único pico en cada intervalo de cuantificación si varios picos están presentes en este intervalo. Esto permite una reducción sustancial de la cantidad de datos a tratar. Además, la posición precisa de un pico está sustituida por la referencia del intervalo al cual el pico pertenece. Esto reduce la incertidumbre de medición que se refiere a la posición de picos puesto que ya no es necesario comparar una posición precisa con la base de conocimientos. Se compara más bien la pertenencia del pico medido a un intervalo. Finalmente, la progresión logarítmica de la anchura de los intervalos permite adaptarse al hecho de que el instrumento presenta una precisión

relativa constante:

$$p = \frac{\Delta\mu}{m} = \text{constante}$$

15 Según un modo de realización, el rango predeterminado de Thompsons está comprendido entre 3000 Thompsons y 17000 Thompsons. Los inventores han constatado, en efecto, que este rango era suficiente para la identificación de la mayoría de las bacterias y de las levaduras/mohos. Se observa especialmente que los picos localizados por debajo de 3000 Thompsons son comunes a muchos microorganismos y no son, por lo tanto, discriminantes.

20 Según la invención, el número de intervalos está comprendido entre 900 y 1500, especialmente entre 1200 y 1400. Los inventores han constatado que estos intervalos constituyen el compromiso óptimo entre la pérdida de informaciones inducida por la cuantificación del espacio de los Thompsons y la precisión ganada por la sustitución de la posición precisa de los picos por los intervalos.

25 Según un modo de realización, el pico mantenido en un intervalo de la subdivisión es el pico con la intensidad más fuerte. Sin embargo, son posibles otras elecciones. Por ejemplo, es posible elegir el valor medio o el valor mediano de las intensidades de los picos presentes en el intervalo.

30 Según un modo de realización, la espectrometría de masas es una espectrometría MALDI-TOF.

La invención tiene también como objetivo un dispositivo de identificación de un microorganismo por espectrometría de masas, que comprende:

- 35 ▪ un espectrómetro de masas apto para producir unos espectros de masas de microorganismos a identificar;
- una unidad de cálculo apta para identificar los microorganismos asociados a los espectros de masas producidos por el espectrómetro utilizando un procedimiento conforme a uno de los modos de realización anteriores.

40 Breve descripción de las figuras

La invención se entenderá mejor a partir de la lectura de la descripción siguiente, dada únicamente a título de ejemplo, y realizada en relación con los dibujos anexos, en los que:

- 45 ▪ la figura 1 es un organigrama del procedimiento según la invención; y
- la figura 2 es un trazado del número de picos eliminados en un espectro de masas en función del número de intervalos de la cuantificación según la invención.

50 Descripción detallada de la invención

Se describirá ahora, en referencia a la figura 1, un procedimiento según la invención de identificación de bacterias mediante una espectrometría de masas de tipo MALDI-TOF (acrónimo de "*Matrix-assisted laser desorption/ionization time of flight*").

55 El procedimiento empieza por la preparación, en 10, de una muestra de una bacteria a identificar, seguida de la adquisición, en 12, de uno o varios espectros de masas de la muestra preparada mediante un espectrómetro de masas de tipo MALDI-TOF. La espectrometría de masas MALDI-TOF es bien conocida en sí y no se detallará por lo tanto más en detalle a continuación. Se podrá, por ejemplo, referir al documento de Jackson O. Lay, "Maldi-tof spectrometry of bacteria", Mass Spectrometry Reviews, 2001, 20, 172-194.

60 El procedimiento se prosigue, en 14, por el pretratamiento de los espectros adquiridos, a fin, especialmente, de eliminar el ruido y alisar los espectros. Más particularmente, se retira la línea de base de los espectros, que representa el ruido de fondo del espectrómetro.

Se realiza entonces, en 16, una identificación de los picos presentes en los espectros adquiridos, por ejemplo mediante un algoritmo de detección de picos, por ejemplo basado en la detección de máximos locales. Se produce así una lista de los picos para cada espectro, que comprende la localización y la intensidad de los picos del espectro.

De manera ventajosa, los picos se identifican en el rango de Thompsons $[m_{\min}; m_{\max}]$ restringido, preferentemente el rango $[m_{\min}; m_{\max}] = [3000; 17000]$ Thompsons. En efecto, se ha observado que las informaciones suficientes para la identificación de los microorganismos se agrupan en este rango, y que por lo tanto no es necesario tener en cuenta un rango más amplio.

El procedimiento se prosigue, en 18, por una etapa de cuantificación, o "binning". Para este fin:

- el rango $[m_{\min}; m_{\max}]$ se subdivide en intervalos cuya anchura crece de manera logarítmica con los Thompsons según la relación:

$$L(b) = \exp\left(\frac{b - \beta}{\alpha}\right) \times \left(\exp\left(\frac{1}{\alpha}\right) - 1\right) \quad (1)$$

$$\alpha = \frac{b_{\min} - (b_{\max} + 1)}{\ln m_{\min} - \ln m_{\max}} \quad (2)$$

$$\beta = \frac{(b_{\max} + 1) \times \ln m_{\min} - b_{\min} \times \ln m_{\max}}{\ln m_{\min} - \ln m_{\max}} \quad (3)$$

en las que los intervalos de la subdivisión se referencian por números enteros superiores a 1 desde el número entero b_{\min} , por ejemplo igual a 1, al número entero b_{\max} , y $L(b)$ es la anchura del intervalo referenciado por el número entero b . El número entero b_{\min} corresponde al intervalo de las relaciones masas sobre cargas más bajas en el rango $[m_{\min}; m_{\max}]$, y el número entero b_{\max} corresponde al intervalo de las relaciones masas sobre cargas más importantes en el rango $[m_{\min}; m_{\max}]$. El eje de Thompsons se cuantifica, por lo tanto, según las relaciones:

$$b(m) = \lfloor \alpha \ln m + \beta \rfloor$$

en la que $\lfloor \cdot \rfloor$ es el símbolo del redondeado al número entero inferior;

- para cada intervalo que comprende varios picos, se conserva un solo pico, ventajosamente el pico que presenta la intensidad más fuerte. Se produce así un vector para cada espectro medido. Cada componente del vector corresponde a un intervalo de la cuantificación y tiene como valor la intensidad del pico conservado para este intervalo, significando el valor "0" que no se ha detectado ningún pico en este intervalo.

Por ejemplo, en la etapa 18 de la figura, en la figura 1, se ilustran tres listas de picos identificadas, a saber "lista 1", "lista 2" y "lista 3", que corresponden cada una a un espectro de masas medido. El espacio de los Thompsons se subdivide en 8 intervalos, de "bin1" a "bin8", de anchura creciente de manera logarítmica, y sólo se conserva en cada intervalo el pico de mayor intensidad. Así, para el intervalo "bin6", de la primera lista "lista 1", se elimina un pico. Para las listas "lista 1", "lista 2" y "lista 3", se obtiene por ejemplo la matriz siguiente, correspondiendo cada línea a una lista:

$$\begin{pmatrix} 980 & 0 & 98 & 0 & 1300 & 1556 & 400 & 2000 \\ 505 & 700 & 200 & 0 & 500 & 200 & 345 & 256 \\ 700 & 0 & 0 & 100 & 2340 & 1786 & 0 & 2507 \end{pmatrix}$$

Se muestra así que, con la ayuda de una cuantificación tal como se ha descrito anteriormente, se tiene en cuenta el incremento de la incertidumbre sobre la posición de los picos a medida que las masas aumentan. Especialmente, la subdivisión según la invención del eje de los Thompsons permite tener en cuenta una incertidumbre del tipo:

$$p = \frac{\Delta\mu}{m} \quad (4)$$

en la que p es la precisión sobre la localización de un pico, $\Delta\mu$ es la incertidumbre de medición sobre la posición de los picos del espectrómetro, y m es la posición real del pico. La cuantificación es por lo tanto una cuantificación adaptativa que tiene en cuenta el error de medición del espectrómetro de masas.

La sustitución de la localización medida de un pico por la referencia al intervalo al cual pertenece, es equivalente a alinear la posición del pico sobre el medio del intervalo. Se verifica que la subdivisión logarítmica según la invención permite reducir la incertidumbre según la relación (4). En efecto, se obtiene:

$$\frac{L(b)}{m_{bar}(b)} = \frac{2 \left(\exp\left\{\frac{1}{\alpha}\right\} - 1 \right)}{\left(\exp\left\{\frac{1}{\alpha}\right\} + 1 \right)} = \text{cte}$$

en la que $m_{bar}(b)$ es el centro del intervalo referenciado de referencia b .

La intensidad de un pico es muy variable de un espectro a otro y/o de un espectrómetro a otro. Debido a esta variabilidad, es muy difícil tener en cuenta los valores brutos de intensidad. De manera ventajosa, pero opcional, el procedimiento se prosigue por una etapa de discretización de las intensidades. Esta etapa puede, por ejemplo, consistir en una simple "binerización" (presencia/ausencia).

Así, cada línea de la matriz está "binarizada" y después normalizada, catalogando así la matriz para cada espectro adquirido la presencia o la ausencia de pico en los intervalos. Por ejemplo, la matriz anterior está binarizada en la matriz:

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}$$

Los inventores han observado además que la información pertinente para la identificación de una bacteria está contenida esencialmente en la ausencia y/o la presencia de picos, y que la información de intensidad es menos pertinente, especialmente debido a su fuerte variabilidad. Así, por ejemplo, es posible identificar unas bacterias en base a este tipo de listas con la ayuda de herramientas de clasificación habituales tales como la regresión logística, el análisis discriminante, los árboles de clasificación, los métodos LASSO, unos algoritmos de tipo SVM (acrónimo de la expresión anglosajona "*support vector machine*"). La matriz así binarizada puede utilizarse en todas las herramientas de clasificación conocidas.

El procedimiento se prosigue después, en 20, por el análisis de la matriz obtenida en la etapa anterior. Más particularmente, se utiliza un algoritmo de clasificación y de decisión 22 en función de una base de conocimientos 24 construida en función de listas de picos de microorganismos y/o tipos de microorganismos previamente identificados. Se identifican así uno o varios candidatos, o un tipo de microorganismos (familia, germen, especie, sub-especie) para la muestra analizada.

El procedimiento según la invención permite así volver a llevar una lista de picos de tamaño variable y a valores continuos sobre 2 ejes (m/z , intensidades) a un vector de tamaño fijo y razonable.

La base de conocimientos 24 está construida a partir de listas de pico producidas como se ha descrito anteriormente, y asociadas a microorganismos y/o tipos de microorganismos previamente identificados. Se comprenderá que la invención se aplica a cualquier tipo de algoritmo de clasificación y de bases de conocimientos. La cuantificación según la invención permite especialmente reducir la cantidad de datos, así como eliminar los problemas de precisión en la localización de los picos, y permite, por lo tanto, construir una base de conocimientos más sólida, y esto de manera más simple. La realización es mucho más simple que el cálculo de una distancia tolerante (por ejemplo) y permite una construcción casi automatizada de la base de conocimientos.

El número de intervalos se selecciona ventajosamente entre 900 y 1500, y preferentemente entre 1200 y 1400 para la identificación de microorganismos. Los inventores han constatado que estos intervalos constituyen el compromiso óptimo entre la pérdida de informaciones inducida por la cuantificación del espacio de los Thompsons y la precisión ganada por la sustitución de la posición precisa de los picos por los intervalos. Los inventores han procedido a unos ensayos y han modelizado, como se ilustra en la figura 2, el número de picos eliminados por la cuantificación en función del número de intervalos. Se observa especialmente que más allá de un cierto número de intervalos, la reducción de la cantidad de datos es mínima, y que por debajo de un cierto número, el número de picos eliminados crece de manera exponencial.

Se ha procedido a unos ensayos comparativos entre la cuantificación logarítmica de la invención y una cuantificación constante, es decir una cuantificación para la cual todos los intervalos son de anchura idéntica, y eso incluso para un mismo espectrómetro de masas y un algoritmo de clasificación y de decisión y una construcción de base de conocimientos idénticos. Estos ensayos se describen en la tabla siguiente. El error corresponde al error de identificación de los microorganismos.

Cuantificación	Número de intervalos	Error (en %)	Espacio memoria ocupado (en Mo)
----------------	----------------------	--------------	---------------------------------

ES 2 755 734 T3

logarítmica	300	10,33	120
logarítmica	600	6,25	240
logarítmica	800	5,3	320
logarítmica	1000	5,4	400
logarítmica	1300	5,0	520
logarítmica	1700	6,9	680
logarítmica	2300	8,52	920
logarítmica	4700	12,2	1880
constante	300	12,4	120
constante	600	8,75	240
constante	800	7,2	320
constante	1000	6,6	400
constante	1300	6,2	520
constante	1700	5,9	680
constante	2300	7,22	920
constante	4700	11,0	1880

5 Eligiendo 1000 intervalos de anchura constante, la anchura de los intervalos es igual a la resolución del espectrómetro de masas utilizado para los ensayos para una relación masa sobre carga igual a 17 000 Thompsons. Eligiendo 4700 intervalos de anchura constante, la anchura de los intervalos es igual a la resolución del espectrómetro para una relación masa sobre carga igual a 3000 Thompsons.

10 Eligiendo 1700 intervalos logarítmicos según las relaciones (1) a (3), con $b_{min} = 1$, la anchura de cada intervalo es igual a la precisión del espectrómetro para una relación masa sobre carga igual a la mitad del intervalo. Sin embargo, se observa que, de promedio, un número de 1300 intervalos permite obtener al mismo tiempo el porcentaje de error de identificación más bajo y el espacio ocupado más bajo. Especialmente, con respecto al número de 1700 intervalos, que parece a primera vista el más adecuado, se realiza una ganancia de 2 puntos de error (-28% de error), disminuyendo al mismo tiempo el espacio de memoria ocupado, como lo indica la tabla anterior. Por lo tanto, se prefiere el número de 1300 para la realización de la invención.

15 Se observa también que la cuantificación según la invención permite obtener un porcentaje de error máximo inferior en al menos 1 punto (-15% de error) al de una cuantificación constante, así como una huella memoria más reducida (-25%). Para unos números de intervalos bajos, la cuantificación según la invención da así mejores resultados que una cuantificación constante. Esto permite así mantener un número bajo de intervalos, incluso aumentando la resolución del espectrómetro de masas o el intervalo de Thompsons elegido $[m_{min}; m_{max}]$. Así, se observa que para un mismo porcentaje de error, por ejemplo aproximadamente el 6%, la cuantificación según la invención necesita únicamente 700 intervalos, mientras que la cuantificación constante necesita 1700.

REIVINDICACIONES

1. Procedimiento de identificación de un microorganismo por espectrometría de masa, que comprende:

- 5 ▪ la adquisición de al menos un espectro de masas de dicho microorganismo;
- para cada espectro de masas adquirido:
 - 10 ◦ la detección de los picos del espectro en un rango predeterminado de masas;
 - la generación de una lista de picos que cataloga como mucho un pico en cada intervalo de una subdivisión predeterminada del rango de las relaciones masa sobre carga, aumentando la anchura de los intervalos de la subdivisión con la relación masa sobre carga según las relaciones:

$$L(b) = \exp\left(\frac{b - \beta}{\alpha}\right) \times \left(\exp\left(\frac{1}{\alpha}\right) - 1\right)$$

$$\alpha = \frac{b_{\min} - (b_{\max} + 1)}{\ln m_{\min} - \ln m_{\max}}$$

$$\beta = \frac{(b_{\max} + 1) \times \ln m_{\min} - b_{\min} \times \ln m_{\max}}{\ln m_{\min} - \ln m_{\max}}$$

20 en las que los intervalos de la subdivisión se referencian por unos números enteros superiores a 1 desde el número entero b_{\min} , para las relaciones masas sobre cargas más bajas en el rango, hasta el número entero b_{\max} , para las relaciones masas sobre cargas más importantes en el rango, $L(b)$ es la anchura del intervalo referenciado por los números enteros b , m_{\min} es un umbral inferior del rango de las relaciones masa sobre carga, y m_{\max} es un umbral superior del rango de las relaciones masa sobre carga; y

- 25 ▪ el análisis de la o de las listas de picos obtenidos en función de la base de conocimientos de microorganismos y/o de tipos de microorganismos previamente identificados,

procedimiento en el que el número de intervalos está comprendido entre 900 y 1500.

30 2. Procedimiento según la reivindicación 1, en el que el intervalo predeterminado de las relaciones masa sobre carga está comprendido entre 3000 Thompsons y 17000 Thompsons.

3. Procedimiento según la reivindicación 1 o 2, en el que el número de intervalos está comprendido entre 1200 y 1400.

35 4. Procedimiento según una cualquiera de las reivindicaciones anteriores, en el que el pico mantenido en un intervalo de la subdivisión es el pico máximo.

5. Procedimiento según una cualquiera de las reivindicaciones anteriores, en el que la espectrometría de masas es una espectrometría MALDI-TOF.

40 6. Dispositivo de identificación de un microorganismo por espectrometría de masas, que comprende:

- un espectrómetro de masas apto para producir unos espectros de masas de microorganismos a identificar;
- 45 ▪ una unidad de cálculo apta para identificar los microorganismos asociados a los espectros de masas producidos por el espectrómetro utilizando un procedimiento conforme a una cualquiera de las reivindicaciones anteriores.

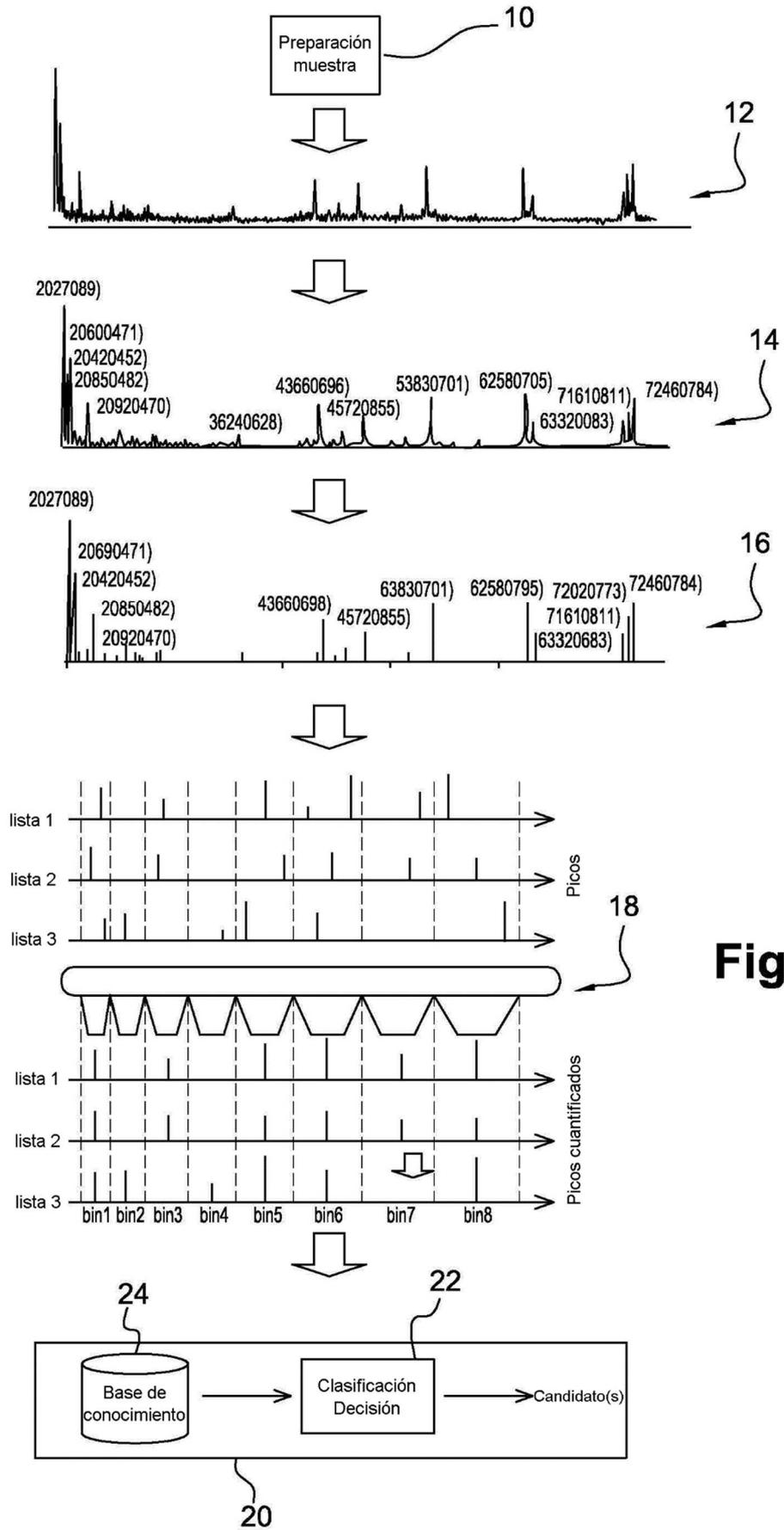


Fig. 1

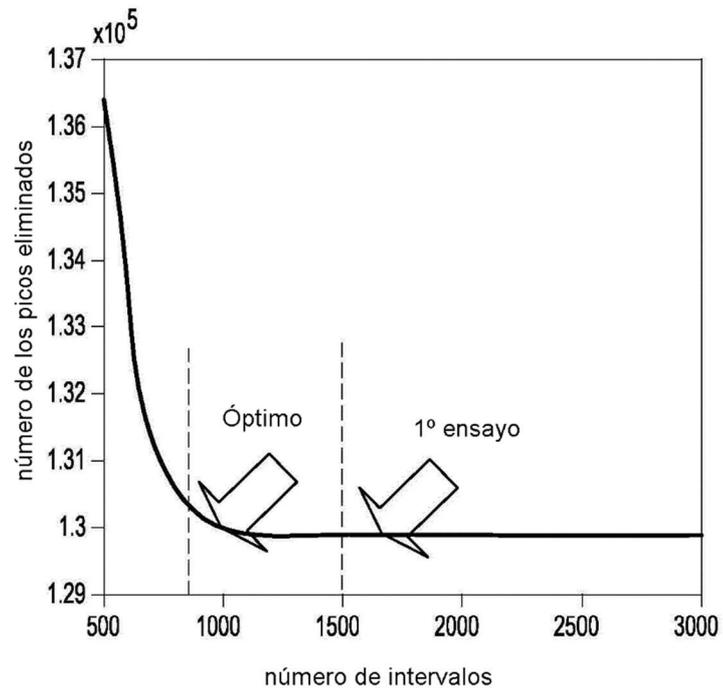


Fig. 2