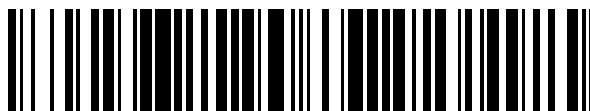


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 385 335**

51 Int. Cl.:  
**C12Q 1/68** (2006.01)  
**G01N 33/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **01270150 .4**  
96 Fecha de presentación: **19.10.2001**  
97 Número de publicación de la solicitud: **1358458**  
97 Fecha de publicación de la solicitud: **05.11.2003**

54 Título: **Marcaje de defecto de masa para la determinación de secuencias de oligómeros**

30 Prioridad:  
**19.10.2000 US 242165 P**  
**19.10.2000 US 242398 P**

45 Fecha de publicación de la mención BOPI:  
**23.07.2012**

45 Fecha de la publicación del folleto de la patente:  
**23.07.2012**

73 Titular/es:  
**TARGET DISCOVERY, INC.**  
**1539 INDUSTRIAL ROAD**  
**SAN CARLOS, CA 94070, US**

72 Inventor/es:  
**SCHNEIDER, Luke V.;**  
**HALL, Michael P. y**  
**PETESCH, Robert**

74 Agente/Representante:  
**Arias Sanz, Juan**

ES 2 385 335 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

Marcaje de defecto de masa para la determinación de secuencias de oligómeros

## Referencias cruzadas a solicitudes relacionadas

5 Esta solicitud reivindica el beneficio de la solicitud de patente estadounidense provisional con n.º de serie 60/242.165, presentada el 19 de octubre de 2000, titulada "Métodos para determinar secuencias terminales de péptidos y proteínas" y de la solicitud de patente estadounidense provisional con n.º serie 60/242.98, presentada el 19 de octubre de 2000, titulada "Métodos para determinar secuencias terminales de péptidos y proteínas", expediente del apoderado n.º 05265.P001.

## Antecedentes de la invención

10 Muchas moléculas se fragmentan por medios químicos, eléctricos (haz de electrones o colisiones inducidas por el campo con moléculas de gas neutro) u ópticos (láseres de excímero) en espectrómetros de masas de manera que las masas de los fragmentos iónicos marcados resultantes pueden usarse para identificar o reconstruir la molécula original. En otros casos, las moléculas pueden coeluirse de procedimientos de separación para distinguirse  
15 moléculas específicas en una mezcla, para ayudar en la identificación de los iones o fragmentos iónicos marcados resultantes con respecto a otro ruido químico en el espectro de masas. Normalmente, este marcador consiste en elementos, o isótopos de elementos, ya contenidos en la molécula original. De esta manera, dos o más picos de abundancias relativas predeterminadas pueden encontrarse en el espectro de masas y usarse para confirmar la identidad de los fragmentos marcados. Sin embargo, cuando el marcador contiene elementos (o isótopos de estos  
20 elementos) ya contenidos en la molécula original o en otros iones generados a partir de o que contaminan de otra manera la matriz de la muestra, uno o más de los picos de fragmentos marcados puede solaparse con otros picos de iones no marcados en el espectro, confundiendo la identificación de los iones marcados.

25 Históricamente, se han usado de manera extensa técnicas tales como degradación de Edman para la secuenciación de proteínas. Véanse, Stark, en: *Methods in Enzymology*, 25:103-120 (1972); Niall, en: *Methods in Enzymology*, 27:942-1011 (1973); Gray, en: *Methods in Enzymology*, 25:121-137 (1972); Schroeder, en: *Methods in Enzymology*, 25:138-143 (1972); Creighton, *Proteins: Structures and Molecular Principles* (W. H. Freeman, NY, 1984); Niederwieser, en: *Methods in Enzymology*, 25:60-99 (1972); y Thiede, *et al.* *FEBS Lett.*, 357:65-69 (1995). Sin embargo, la secuenciación mediante métodos de espectrometría de masas (EM) de disociación inducida por colisión (secuenciación por EM/EM) ha evolucionado rápidamente y ha demostrado ser más rápida y requerir menos  
30 proteína que las técnicas de Edman. Véanse, Shevchenko, A., *et al.*, *Proc. Natl. Acad. Sci. (EE.UU)*, 93:14440-14445 (1996); Wilm, *et al.*, *Nature*, 379:466-469 (1996); Mark, J., "Protein structure and identification with MS/MS", artículo presentado en la serie de seminarios PE/Sciex, *Protein Characterization and Proteomics: Automated high throughput technologies for drug discovery*, Foster City, CA (marzo, 1998); y Bieman, *Methods in Enzymology*, 193:455-479 (1990).

35 La secuenciación por EM se logra o bien usando voltajes superiores en la zona de ionización de la EM para fragmentar aleatoriamente un único péptido aislado a partir de una digestión de proteínas, o más normalmente mediante EM en tándem usando disociación inducida por colisión en la trampa iónica. Véase, Bieman, *ibid.* Pueden usarse varias técnicas para seleccionar el fragmento de péptido usado para secuenciación por EM/EM, incluyendo la acumulación del ion del fragmento de péptido original en la unidad de EM de cuadrupolo (véanse, Mark, J. *ibid.*;  
40 Mann, M., artículo presentado en la conferencia de IBC Proteomics, Boston, MA (10-11 de noviembre de 1997); y Bieman, *Methods in Enzymology*, 193:455-479 (1990)), separación electroforética capilar acoplada a detección por EM ES-TOF (véanse, Aebersold, R. "Proteome analysis: Biological assay or data archive?", artículo presentado en la conferencia de IBC Proteomics, Coronado, CA (11-12 de junio de 1998) y Smith, *et al.*, en: *CRC Handbook of Capillary Electrophoresis: A Practical Approach*, cap. 8, págs. 185-206 (CRC Press, Boca Raton, FL, 1994)), u otras  
45 separaciones por cromatografía de líquidos (Niall, H. D., en: *Methods in Enzymology*, 27:942-1011 (1973) y Creighton, T. E., *Proteins: Structures and Molecular Principles* (W. H. Freeman, NY, 1984)). La secuencia de aminoácidos del péptido se deduce a partir de las diferencias de peso molecular observadas en el patrón de fragmentación de EM resultante del péptido usando las masas publicadas asociadas con residuos de aminoácidos individuales en la EM (Biemann, K., en: *Methods in Enzymology*, 193:888 (1990), y se ha codificado en un algoritmo  
50 de secuenciación de péptidos semi-autónomo (Hines, *et al.*, *J Am Soc Mass Spectrom*, 3:326-336 (1992)).

Por ejemplo, en el espectro de masas de un péptido de 1425,7 Da (HSDAVFTDNYTR) aislado en un experimento de EM/EM adquirido en modo de ion positivo, la diferencia entre el péptido completo de 1425,7 Da y el siguiente fragmento de masa más grande ( $y_{11}$ , 1288,7 Da) es de 137 Da. Esto corresponde a la masa esperada de un residuo de histidina N-terminal que se escinde en el enlace amida. Para este péptido, la secuenciación completa es posible  
55 como resultado de la generación de iones de fragmentos de alta abundancia que corresponden a la escisión del péptido en casi cada residuo a lo largo de la estructura principal del péptido. En la secuencia de péptido mencionada anteriormente, la generación de un conjunto esencialmente completo de iones de fragmentos cargados positivamente que incluye cualquier extremo del péptido es un resultado de la basicidad de los residuos tanto N-como C-terminales. Cuando un residuo básico se ubica en el extremo N-terminal y/o C-terminal, la mayoría de los

iones producidos en el espectro de disociación inducida por colisión (CID) contendrán ese residuo (véanse, Zaia, J., en: *Protein and Peptide Analysis by Mass Spectrometry*, J.R. Chapman, ed., págs. 29-41, Humana Press, Totowa, NJ, 1996; y Johnson, R.S., *et al.*, *Mass Spectrom. Ion Processes*, 86:137-154 (1988)) puesto que la carga positiva se ubica generalmente en el sitio básico. Normalmente, la presencia de un residuo básico simplifica el espectro resultante, puesto que un sitio básico dirige la fragmentación hacia una serie limitada de iones hijo específicos. Los péptidos que carecen de residuos básicos tienden a fragmentarse en una mezcla más compleja de iones de fragmentos que hace que la determinación de secuencia sea más difícil.

La secuenciación de ácido nucleico se ha realizado históricamente a través de la síntesis de fragmentos de ácido nucleico que contienen números aleatorios de bases copiadas de una secuencia de ácido nucleico original, tal como los métodos definidos por Sanger y Colson, *Proc. Natl. Acad. Sci. (EE.UU)*, 74:5463-5467 (1977); y Maxam y Gilbert *Methods in Enzymology*, 65:499-560 (1980). Una variación en el método descrito por Sanger y Colson usa un método de la reacción en cadena de la polimerasa (PCR) incompleta para sintetizar el marcador de masa molecular de fragmentos de ADN (véase, Nakamaye *et al.*, *Nuc. Acids Res.*, 16(21):9947-9959 (1988)). Se han desarrollado métodos de espectrometría de masas para la identificación y separación multiplexada y más rápida de los marcadores de tamaño molecular de ADN, tal como describe Koster, patentes estadounidenses n.ºs 5.691.141 y 6.194.144; Monforte *et al.* patente estadounidense n.º 5.700.642, y Butler, *et al.* patente estadounidense n.º 6.090.558. En estos métodos, los fragmentos de ácido nucleico se introducen simultáneamente en el espectrómetro de masas y la secuencia o el número de "repeticiones en tándem cortas" se deducen a partir de las diferencias de masa entre elementos individuales del marcador de masa molecular de fragmentos de masa sintetizados. Tal como se describe por Koster, patente estadounidense n.º 6.194.144, es tanto posible como deseable secuenciar varios ácidos nucleicos simultáneamente en paralelo marcando de manera diferencial los fragmentos de ácido nucleico sintetizados a partir de moldes originales de ácido nucleico únicos con diferentes etiquetas de masas suficientemente únicas. Incluso usando marcadores de masa única, debe tenerse cuidado para evitar la subfragmentación de los elementos del marcador de masa molecular de secuencias durante la ionización o transmisión de iones en el espectrómetro de masas, y para purificar los fragmentos de ácido nucleico de otros ácidos nucleicos extraños y contaminantes de la matriz que confunden de manera que pueda obtenerse una secuencia inequívoca a partir del espectro de masas resultante.

También se han descrito métodos de secuenciación de polisacáridos que utilizan métodos de etiquetado de masa en el espectrómetro de masas por Rademacher *et al.* patente estadounidense n.º 5.100.778 y Parekh y Prime patente estadounidense n.º 5.667.984. En estos métodos, se une una etiqueta de masa única a una muestra de polisacárido purificado, que posteriormente se divide en alícuotas que se someten a diferentes regímenes de escisión enzimática y/o quimiolítica para producir una serie de fragmentos de oligosacárido marcados derivados del polisacárido original. Estos fragmentos se introducen simultáneamente en un espectrómetro de masas y se determina la secuencia de azúcares contenida en el polisacárido original a partir del marcador de masa molecular resultante generado en el espectro de masas a partir de los fragmentos de oligosacárido marcados aleatoriamente. Se reconoce que puede obtenerse un aumento de rendimiento procesando varias muestras diferentes simultáneamente en paralelo a través del uso de diferentes etiquetas de masa unidas a cada muestra original de polisacárido purificado único. De nuevo, debe tenerse cuidado con las muestras de oligosacárido para evitar la subfragmentación en el espectro de masas y para purificar los fragmentos marcados de contaminantes de oligosacárido no marcados para evitar ambigüedades de secuenciación.

La identificación de la composición de ácidos grasos y colocación en lípidos puede ser un indicador importante del estado de una célula. Por ejemplo, Oliver y Stringer, *Appl. Environ. Microbiol.*, 4:461 (1984) y Hood *et al.*, *Appl. Environ. Microbiol.*, 52:788 (1986) notifican ambos una pérdida del 99,8% de los fosfolípidos en privación de *Vibrio* sp. Cronan, *J. Bacteriol.*, 95:2054 (1968) encontró que el 50% del contenido en fosfatidilglicerol de *Escherichia coli* K-12 se convertía en cardiolipina en el plazo de dos horas del inicio de la privación de fosfato y que la composición de ácidos grasos también cambiaba significativamente. La composición de lípidos de la membrana celular también es de interés médico debido a sus posibles papeles en la captación de metabolitos y fármacos, el anclaje de proteínas transmembrana, el reconocimiento viral de superficies celulares, la proliferación y metástasis tumorales y la enfermedad arterial.

Se han descrito enfoques de etiqueta de masa similares para la identificación de componentes individuales de bibliotecas químicas sintetizadas de manera combinatoria por Sugarman *et al.* patente estadounidense n.º 6.056.926 y Brenner *et al.* *Proc. Natl. Acad. Sci. (EE.UU)*, 89:5381-5383 (1992), en las que se sintetiza un marcador de etiqueta de masa única simultáneamente con el compuesto químico de interés sobre una superficie sólida y se usa después para identificar las diversas etapas de procesamiento aplicadas a la superficie sólida. Este marcador de masa puede identificarse tras la escisión de la superficie sólida mediante espectrometría de masas. La limitación en el tamaño de la biblioteca que puede producirse mediante enfoques combinatorios es el número de marcadores de masa única que pueden generarse y la capacidad para discriminar estos marcadores de los compuestos de interés.

Ness *et al.* patente estadounidense n.º 6.027.890, Schmidt *et al.* documento WO 99/32501 y Aebersold *et al.* documento WO 00/11208 describen todos métodos para marcar de manera diferencial moléculas biológicas obtenidas de diferentes fuentes con una etiqueta de masa diferente para cada fuente. Entonces, las muestras pueden combinarse, tras el marcaje, y procesarse juntas a través de reacciones de separación o enriquecimiento de afinidad, de manera que se garantiza que los compuestos individuales de cada muestra van a tratarse de manera

idéntica en la mezcla. Entonces, se determinan las concentraciones relativas de compuestos biológicos marcados de manera diferencial individuales mediante las abundancias relativas de las etiquetas de masa individuales en el espectro de masas. Las limitaciones en estos métodos son que los marcadores de masa empleados deben comportarse prácticamente de manera idéntica con respecto a cualquier procesamiento de la mezcla de muestras y la ionización y el transporte de los iones resultantes en el espectrómetro de masas. Por este motivo, se eligen normalmente marcadores que son análogos químicos (por ejemplo, análogos de isótopos estables o son derivados sencillos entre sí). Una limitación de estos métodos es el número de derivados de etiqueta de masa que pueden mezclarse para un único análisis paralelo, que está limitado por el número de derivados de etiqueta de masa que pueden sintetizarse con eficacias de transmisión e ionización y comportamientos de separación casi idénticos. Otra limitación de estos métodos es la capacidad para distinguir los marcadores escindidos o moléculas marcadas de masa de contaminantes de la matriz y biomoléculas no marcadas que también pueden estar presentes en la muestra introducida en el espectrómetro de masas. Esta última limitación a menudo significa que la muestra marcada debe purificarse exhaustivamente antes de análisis del espectro de masas y que la subfragmentación de las moléculas marcadas en el espectrómetro de masas debe evitarse.

Schmidt *et al.* documento WO99/32501 (1 de julio de 1999) describen el uso de flúor en lugar de hidrógeno como elemento de defecto de masa distinguible en marcadores de masa escindibles. La base para este trabajo es la diferencia de masa monoisotópica de 0,009422 amu entre estos dos elementos. Sin embargo, esta es una diferencia de masa muy pequeña, que sólo puede resolverse con espectrómetros de masas de resolución de masas muy alta y a los intervalos de masas más bajos en estos espectrómetros de masas. La resolución de los espectrómetros de masas depende del intervalo de masas y se indica normalmente en partes por millón. Por ejemplo, los detectores de tiempo de vuelo típicos comunes en la industria tienen una resolución de masas de aproximadamente 10 amu a una masa de 1 millón de amu (10 ppm). Por tanto, la diferencia de masas comparativamente pequeña entre F y H es imposible de resolver por encima de una masa de aproximadamente 940 amu, y desde una perspectiva práctica a una  $m/z$  incluso inferior.

Schmidt *et al.* observan además que el defecto de masa de hidrocarburos perfluorados puede distinguirse de hidrocarburos simples. Por ejemplo, la masa monoisotópica de una etiqueta de arilo polifluorado con una estequiometría máxima de  $C_6F_5$  es exactamente de 166,992015 amu. La masa monoisotópica del hidrocarburo más cercano es de 167,179975, que corresponde a un estequiometría de  $C_{12}H_{23}$  y una diferencia de masa fácilmente resoluble de aproximadamente 1125 ppm. La masa de la etiqueta alifática polifluorada mínima es de 68,995209 amu, que corresponde a una estequiometría de  $CF_3$ . La masa de hidrocarburo monoisotópico más cercano a esto es de 69,070425, correspondiente a una estequiometría de  $C_5H_9$  y una diferencia de 1089 ppm.

Sin embargo, para moléculas orgánicas que incluyen heteroátomos, tales como N y O, que son típicos en moléculas biológicas, el defecto de masa de flúor no se distingue tan fácilmente. Por ejemplo, cualquier molécula que contiene una estequiometría de  $C_3HO_2$  tendrá una masa monoisotópica que es sólo 35 ppm diferente de la de  $CF_3$ , haciéndola casi indistinguible incluso a 69 amu. De manera similar, cualquier molécula que contiene una estequiometría monoisotópica de  $C_7H_3O_5$  es sólo 36 ppm diferente de  $C_6F_5$  a 167 amu.

Cuando se incluyen los isótopos estables de C, N y O en los cálculos, el defecto de masa de  $C_6F_5$  se reduce a un indistinguible 1,4 ppm cuando se compara con una molécula que contiene una estequiometría de  $[^{12}C]_4[^{13}C]_2[^{15}N]_3[^{16}O]_2$ . De manera similar, el defecto de masa para  $CF_3$  reduce a un mero 29 ppm en comparación con una molécula que contiene una estequiometría de  $[^{12}C]_2[^{13}C][^{16}O]_2$ . Puesto que la masa global de la etiqueta aumenta más allá de 200 amu, el defecto de masa introducido incluso con múltiples fluoros rápidamente se vuelve indistinguible entre los defectos de los otros heteroátomos e isótopos estables. Añadir incluso más fluoros a la molécula a menudo no es práctico debido a restricciones de solubilidad.

El problema general de la deconvolución de picos individuales de interés a partir de datos de espectros de masas complejos se ha descrito previamente para mezclas complejas de moléculas pequeñas, (véase Stein, S. E., "An integrated method for spectrum extraction and compound identification from GC/MS Data", *J Am Soc Mass Spect*, 10:770-781 (1999) y Mallard, G.W. y J. Reed, "Automated Mass Spectral Deconvolution & Identification System, AMDIS-User Guide" (US Department of Commerce, Gaithersburg, MD, 1997)) particularmente cuando se acopla a métodos de separación de resolución temporal (por ejemplo, CG/EM y CL/EM). Sin embargo, estas técnicas no se han aplicado a espectros de fragmentación de biopolímeros (por ejemplo, proteína, ácido nucleico y polisacárido) para el fin de determinación de la secuencia. De hecho, estos métodos normalmente intentan identificar especies químicas intactas y generalmente buscan evitar las condiciones de fragmentación en la EM. Tampoco se han acoplado a la identificación de iones de biomoléculas marcados que contienen etiquetas de masa única.

Extendiendo el concepto de simplificar el espectro CID de un péptido incluyendo un resto de concentración de carga en cualquier extremo terminal del péptido, otros han demostrado que unir una carga positiva dura al extremo N-terminal dirige la producción de una serie completa de iones de fragmentos N-terminales a partir de un péptido original en experimentos de CID independientemente de la presencia o ausencia de un residuo básico en el extremo N-terminal. Véanse, Johnson, R.S., *et al.*, *Mass Spectrom. Ion Processes*, 86:137-154 (1988); Vath, J.E., *et al.*, *Fresenius Z Anal. Chem.*, 331:248-252 (1988); Stults, J.T., *et al.*, *Anal. Chem.*, 65:1703-1708 (1993); Zaia, J., *et al.*, *J. Am. Soc. Mass Spectrom.*, 6:423-436 (1995); Wagner, D.S., *et al.*, *Biol. Mass Spectrom.*, 20:419-425 (1991); y Huang, Z.-H., *et al.*, *Anal. Biochem.*, 268:305-317 (1999). En teoría, todos los iones de fragmentos se producen mediante

fragmentación de carga remota que está dirigida por el grupo de carga fija. Véase, Tomer, K.B., *et al.*, J. Am. Chem. Soc., 105:5487-5488 (1983).

Los péptidos se han marcado con varias clases de grupos de carga fija, incluyendo derivados de dimetilalquilamonio, piridinio sustituido, fosfonio cuaternario y sulfonio. Las características de marcadores útiles incluyen facilidad de síntesis, aumento en la eficacia de ionización de péptidos marcados y formación de un péptido marcado de una serie de iones de fragmentos específicos con fragmentación de marcador desfavorable mínima. Zaia (en: Protein and Peptide Analysis by Mass Spectrometry, J.R. Chapman, ed., págs. 29-41, Humana Press, Totowa, NJ, 1996) notificó que los marcadores que satisfacen estos criterios incluyen los de la clase de dimetilalquilamonio y derivados de fosfonio cuaternario. Además, se ha notificado que derivados de piridinio sustituido son útiles en CID de alta energía. Véanse, Bures, E.J., *et al.*, Anal. Biochem., 224:364-372 (1995) y Aebersold, R., *et al.*, en: Protein Science, págs. 494-503 (Cambridge University Press, 1992).

A pesar de algún progreso en la metodología analítica, la identificación de proteínas sigue siendo un cuello de botella importante en el campo de la proteómica. Por ejemplo, puede requerir hasta 18 horas generar una etiqueta de secuencia de proteína de suficiente longitud para permitir la identificación de una única proteína purificada a partir de su secuencia genómica pronosticada (véase, Shevchenko, A., *et al.*, Proc. Natl. Acad. Sci. (EE.UU), 93:14440-14445 (1996)). Además, aunque puede lograrse la identificación de proteínas inequívoca generando una etiqueta de secuencia de proteína (PST, véanse Clauser, K.R., *et al.*, Proc. Natl. Acad. Sci. (EE.UU), 92:5072-5076 (1995) y Li, G., M., *et al.*, Electrophoresis, 18:391-402 (1997)), limitaciones en la eficacia de ionización de péptidos y proteínas más grandes restringen la sensibilidad de detección intrínseca de las técnicas de EM e inhiben el uso de EM para la identificación de proteínas de baja abundancia. Además, limitaciones en la precisión de masa de los detectores de tiempo de vuelo (TOF) también pueden restringir la utilidad de los métodos actualmente utilizados de secuenciación por EM/EM, que requieren que las proteínas se digieran por medios proteolíticos y/o quimiolíticos para dar péptidos más manejables (véanse Ambler, R. P., en: Methods in Enzymology, 25:143-154 (1972) y Gross, E., en: Methods in Enzymol., 11:238-255 (1967) antes de la secuenciación. Además, los algoritmos de secuenciación del marcador de masa molecular de EM previamente descritos no funcionan en proteínas porque la abundancia de fragmentos de péptidos generados durante CID de tales moléculas grandes y la incapacidad para identificar un ion original apropiado para iniciar la secuencia ocultan eficazmente el marcador de masa molecular.

Se han propuesto dos estrategias básicas para la identificación por EM de proteínas tras su separación de una mezcla de proteínas: 1) determinación del perfil de huella peptídica ("determinación de la huella de EM") (véanse, James, P., *et al.*, Biochem. Biophys. Res. Commun., 195:58-64 (1993) y Yates, J.R., *et al.*, Anal. Biochem., 214:397-408 (1993)); y 2) secuenciación de uno o más dominios peptídicos mediante EM/EM ("secuenciación por EM/EM") (véanse Mann, M., artículo presentado en la conferencia de IBC Proteomics, Boston, MA (10-11 de noviembre de 1997); Wilm, M., *et al.*, Nature, 379:466-469 (1996); y Chait, B.T., *et al.*, Science, 262:89-92 (1993)). La determinación de la huella de EM se logra midiendo con precisión las masas de varios péptidos generados por una digestión proteolítica de la proteína intacta y buscando en una base de datos una proteína conocida con esa huella peptídica. La secuenciación por EM/EM implica la determinación real de una o más PST de la proteína mediante la generación de iones de fragmentación específicos de la secuencia en el cuadrupolo de un instrumento de EM/EM.

Clauser *et al.*, Proc. Natl. Acad. Sci. (EE.UU), 92:5072-5076 (1995) ha sugerido que las proteínas sólo pueden identificarse inequívocamente a través de la determinación de PST que permiten la referencia a las secuencias teóricas determinadas a partir de las bases de datos genómicas. Li *et al.*, Electrophoresis, 18:391-402 (1997) parecen haber demostrado esta afirmación al encontrar que la identificación fiable de proteínas individuales mediante la determinación de la huella de EM degeneraba a medida que el tamaño de la base de datos de masa de péptidos teóricos comparativa aumentaba. Li *et al.*, *ibid.*, también notificaron que sólo pudieron obtener mapas de péptidos para las proteínas de mayor abundancia en el gel debido a limitaciones de sensibilidad de la EM, aún cuando se demostró que su metodología MALDI de desorción por láser asistido por matriz mejora la sensibilidad de detección con respecto a métodos anteriormente notificados. Claramente, técnicas de secuenciación de proteínas rentables y rápidas mejorarán la velocidad y disminuirán el coste de la investigación en proteómica. De manera similar, tal como describe Koster, la preparación y purificación de ácidos nucleicos antes de la secuenciación, incluso mediante espectrómetros de masas, aumenta el tiempo y coste de la secuenciación de ácidos nucleicos. La mejora de la capacidad de discriminación del espectrómetro de masas, de manera que múltiples proteínas, ácidos nucleicos, polisacáridos u otras secuencias pueden determinarse en paralelo o iones específicos pueden diferenciarse mejor de material orgánico no marcado, tiene utilidad considerable con respecto a métodos existentes.

Miyagi *et al.* (Rapid Comm. Mass Spectr., 1998, 12:603-608) dan a conocer un método para secuenciar una parte terminal de un oligómero que incluye derivatización específica del grupo amino N-terminal usando éster de N-hidroxisuccinimida del ácido 5-bromonicotínico.

Hobba *et al.* (J Biol. Chem., 1996, 271:30529-30536) dan a conocer un método de análisis de la función y estructura del sitio de unión del factor de crecimiento similar a la insulina (IGF) de la proteína 2 de unión al factor de crecimiento similar a la insulina bovina (bIG-FBP-2) mediante yodación química.

## Sumario de la invención

La presente invención es tal como se expone en las reivindicaciones.

La presente invención proporciona la aplicación del marcaje de defecto de masa a una amplia variedad de moléculas. Debido a que los métodos de la invención pueden aplicarse durante la disociación o bien "en fuente" o bien inducida por colisión del oligómero en un cuadrupolo, el método elimina preferentemente la necesidad de la síntesis química de fragmentos de oligómero (por ejemplo, digestión quimiolítica o enzimática, o síntesis de fragmentos de secuenciación por PCR o Sanger). Por tanto, los presentes métodos proporcionan tiempos de secuenciación de oligómeros que se reducen significativamente con respecto a los tiempos que pueden obtenerse usando otros métodos. El método también puede aplicarse como mejora con respecto a enfoques de secuenciación de oligómeros más convencionales, tales como secuenciación por EM/EM de péptidos, secuenciación por PCR o Sanger mediante espectrometría de masas (tal como describen Koster y Butler *et al.*), y secuenciación de polisacáridos tal como describen Rademacher *et al.* y Parekh *et al.* Al permitir que un mayor número de muestras se procesen simultáneamente en paralelo. El aumento de la capacidad para distinguir un mayor número de etiquetas en un único espectro de masas permite sintetizar o examinar bibliotecas combinatorias más grandes. Además, debido a que los oligómeros que están secuenciándose están altamente fragmentados usando los presentes métodos, la eficacia de ionización y la volatilidad de los fragmentos resultantes son más altas que las del oligómero original, conduciendo así a una sensibilidad de detección que se mejora con respecto a otros métodos.

En un aspecto, la presente descripción proporciona un método para secuenciar una parte terminal de un oligómero o polímero, que comprende:

(a) poner en contacto dicho oligómero con un resto de marcaje para unir covalentemente un marcador al extremo terminal del oligómero y formar un oligómero marcado, comprendiendo el resto de marcaje al menos un elemento que tiene un número atómico de desde 17 hasta 77, con la condición de que dicho elemento o elementos sea(n) distinto(s) de azufre o fósforo; y el oligómero no contenga elementos en su estructura que tienen un número atómico de entre 17 y 77 con la excepción de azufre o fósforo,

(b) fragmentar el oligómero marcado usando un método de fragmentación enzimático, quimiolítico o de espectrometría de masas para producir fragmentos de oligómero marcado; y

(c) analizar los fragmentos de oligómero marcado usando un algoritmo o método de secuenciación de espectrometría de masas para determinar la secuencia de al menos dos elementos monoméricos cerca del marcador.

En un aspecto de la presente descripción, en el que el oligómero es una proteína, un péptido, o un ácido nucleico, el método comprende además:

(d) identificar la proteína o el gen usando la secuencia de los al menos dos residuos del extremo terminal para buscar secuencias pronosticadas a partir de una base de datos de datos de secuencias génicas.

En otro aspecto, la presente descripción proporciona un método para secuenciar una parte de un oligómero en una mezcla de oligómeros, comprendiendo el método:

(a) poner en contacto la mezcla de oligómeros con un resto de marcaje para unir covalentemente un marcador a un extremo terminal del oligómero y formar una mezcla de oligómeros marcados, comprendiendo el resto de marcaje al menos un elemento que tiene un número atómico de desde 17 hasta 77, con la condición de que dicho elemento sea distinto de azufre o fósforo; y el oligómero no contenga elementos en su estructura que tienen un número atómico de entre 17 y 77 con la excepción de azufre o fósforo,

(b) separar los oligómeros marcados individuales en la mezcla de oligómeros; y

(c) analizar los oligómeros marcados de la etapa (b) mediante un método de espectrometría de masas para determinar la secuencia de al menos dos residuos terminales.

En un aspecto de la presente descripción, en el que el oligómero es una proteína, el método comprende además:

(d) identificar la proteína usando la secuencia de al menos dos residuos del extremo C-terminal o dos residuos del extremo N-terminal en combinación con una coordenada de separación de la proteína marcada y la ubicación en el extremo terminal de la proteína de la secuencia para buscar secuencias de proteína pronosticadas a partir de una base de datos de datos de secuencias génicas.

Aún en otro aspecto, la presente descripción proporciona un método para el análisis de la función y estructura de un oligómero que tiene una pluralidad de monómeros, comprendiendo el método:

(a) poner en contacto el oligómero con un reactivo de marcaje de defecto de masa para marcar de manera diferencial monómeros expuestos y monómeros no expuestos y producir un oligómero marcado de manera diferencial, en el que el reactivo de marcaje de defecto de masa comprende al menos un elemento que tiene un número atómico de desde 17 hasta 77 que es distinto de azufre o fósforo; y el oligómero no contiene elementos en su estructura que tengan un número atómico de entre 17 y 77 con la excepción de azufre o fósforo,

(b) analizar el oligómero marcado de manera diferencial mediante un método de espectrometría de masas para determinar las secuencias del oligómero que están expuestas en la estructura tridimensional y secuencias del oligómero que no están expuestas en la estructura tridimensional.

5 Aún en otro aspecto de la descripción, el método puede aplicarse a la determinación cuantitativa de biomoléculas marcadas a partir de mezclas, comprendiendo el método:

10 (a) poner en contacto las biomoléculas obtenidas a partir de una muestra con un resto de marcaje para unir covalentemente un marcador a las biomoléculas y formar una mezcla de biomoléculas marcadas, poner en contacto un conjunto similar de biomoléculas obtenidas a partir de al menos otra muestra con un resto de marcaje para unir covalentemente un marcador a las biomoléculas y formar una biomolécula marcada, comprendiendo el primer resto de marcaje al menos un elemento que tiene un número atómico de desde 17 hasta 77, y comprendiendo cada resto de marcaje sucesivo al menos un elemento adicional que tiene un número atómico de desde 17 hasta 77; con la condición de que dichos elementos son distintos de azufre o fósforo; y la biomolécula no contiene elementos en su estructura que tengan un número atómico de entre 17 y 77 con la excepción de azufre o fósforo;

(b) mezclar las muestras de biomoléculas marcadas de manera diferencial de cada fuente;

15 (c) separar opcionalmente las biomoléculas por afinidad u otros medios; y

(c) analizar las biomoléculas marcadas de manera diferencial mediante un método de espectrometría de masas para determinar cantidades relativas de las biomoléculas marcadas individuales o marcadores escindidos de dichas biomoléculas marcadas.

20 Aún en otro aspecto de la descripción, el método puede usarse para codificar bibliotecas combinatorias, comprendiendo el método:

(a) poner en contacto la superficie sintética con un monómero de etiqueta para identificar el proceso químico al que la superficie sintética está (o estará) expuesta y formar una superficie sintética marcada, comprendiendo el monómero de etiqueta al menos un elemento que tiene un número atómico de desde 17 hasta 77,

25 (b) opcionalmente poner en contacto la superficie sintética con monómeros de etiqueta sucesivos para identificar etapas de procesos químicos posteriores a los que la superficie sintética está (o estará) expuesta y formar una superficie sintética sucesivamente marcada, comprendiendo los monómeros de etiqueta al menos un elemento que tiene un número atómico de desde 17 hasta 77,

(c) escindir la(s) etiqueta(s) de monómero de la superficie sintética o bien como una única molécula con múltiples etiquetas, o como un conjunto de etiquetas individuales.

30 (d) analizar la composición de la(s) etiqueta(s) mediante un método de espectrometría de masas para determinar la historia de procesamiento químico de la superficie.

En cada uno de los métodos anteriores, el uso de un algoritmo sólido para la secuenciación de oligómeros marcados terminalmente de los fragmentos de oligómero proporciona ventajas con respecto a los enfoques de algoritmo de secuenciación por EM/EM convencionales. Una ventaja particular es, por ejemplo, la capacidad para secuenciar parcialmente oligómeros de longitud completa (por ejemplo, proteínas o ácidos nucleicos tales como ADN y ARN) sin la necesidad de digestión previa para dar fragmentos más pequeños si se usan técnicas de fragmentación de espectrometría de masas para generar los fragmentos. Otra ventaja es la capacidad para descartar automáticamente el ruido químico (por ejemplo, moléculas u oligómeros no marcados y contaminantes de la matriz orgánica) en el espectro de masas. Otra ventaja es que el método se autoinicia basándose en el defecto de masa conocido del marcador y no requiere ningún conocimiento de la composición o el tamaño del ion original para determinar la secuencia. Otra ventaja es que el método puede ser altamente automatizado. Todavía otra ventaja es que resultan menos ambigüedades de secuencia debido a la precisión de masa absoluta mejorada obtenida trabajando en el extremo bajo del espectro de masas. Aún otra ventaja es que resultan una mejor eficacia de ionización y una sensibilidad de detección correspondiente del uso de condiciones de ionización más energética y la introducción de una carga ionizable o dura en los fragmentos a través de la adición del marcador. Aún otra ventaja de introducir una carga a través del marcador es la capacidad de determinar secuencias de oligómeros parciales a partir de regiones de un oligómero que puede no contener residuos ionizables. Es obvio para los expertos en la técnica que los métodos de la presente invención pueden aplicarse para el análisis de secuencias de cualquier polímero orgánico o cuantificación relativa de cualquier molécula orgánica entre dos o más muestras con la condición de que pueda unirse un marcador de defecto de masa a las moléculas o polímero orgánico.

#### Breve descripción de los dibujos

La figura 1 muestra el espectro de masas de glicógeno fosforilasa tomado a condiciones de alta fragmentación de potencial de boquilla de 325V en un espectrómetro de masas PE Biosystems Mariner.

La figura 2 muestra el patrón de picos periódicos observado a aproximadamente una separación de 1 amu a lo largo

de varias secciones de 20 amu del espectro de masas de glicógeno fosforilasa (figura 1).

La figura 3 muestra la superposición de todos los picos de EM en una separación de 1 amu para datos de fragmentación de glicógeno fosforilasa a 325V en el intervalo de 50-550 amu.

5 La figura 4 ilustra la energía de unión nuclear para los elementos de la tabla periódica como función del número de masa de isótopos.

La figura 5 ilustra las estructuras de varios restos de marcaje que tienen al menos un sustituyente de bromo.

La figura 6 ilustra las estructuras de varios restos de marcaje que tienen al menos un sustituyente de yodo.

La figura 7 proporciona la estructura de un oligosacárido de tipo con alto contenido en manosa que puede secuenciarse usando los métodos proporcionados en el ejemplo 1.

10 La figura 8A muestra el espectro de defecto de masa deconvolucionado para un digesto de oligosacárido de tipo con alto contenido en manosa marcado con el marcador 1. La figura 8B muestra el espectro de defecto de masa deconvolucionado para un digesto de oligosacárido de tipo con alto contenido en manosa marcado con el marcador 2. La figura 8C muestra el espectro de defecto de masa deconvolucionado para un digesto de oligosacárido de tipo con alto contenido en manosa marcado con el marcador 3.

15 La figura 9 muestra el espectro de defecto de masa deconvolucionado para un lípido marcado con el marcador 1 y el marcador 2 (véase el ejemplo 2).

La figura 10 ilustra una estructura general para un marcador de defecto de masa fotoescindible en el que Br es el elemento de defecto de masa que está unido a través del aminoácido (R) al resto del marcador (o etiqueta).

20 La figura 11A muestra el espectro de masas deconvolucionado del defecto de masa fotoaclorado. La figura 11B muestra el espectro resuelto de isótopo de bromo de etiqueta de la etiqueta de defectos de masa monoisotópicos fotoaclorados. La figura 11C muestra el espectro de masa resuelto por isótopos.

La figura 12A muestra la resolución de fragmentos de ion b de otro ruido químico en el espectro de masas. La figura 12B muestra la resolución de fragmentos de ion a de otro ruido químico en el espectro de masas. La figura 12C muestra la resolución de fragmentos de ion d de otro ruido químico en el espectro de masas.

25 La figura 13A muestra los dobletes de pares de isótopos de bromo que se desvían del ruido periódico que corresponde a masas del ion b de carga única del marcador. La figura 13B muestra los dobletes de pares de isótopos de bromo que se desvían del ruido periódico que corresponde a masas del ion b1 de carga única del extremo N-terminal de mioglobina marcada.

30 La figura 14A y la figura 14B muestran los dobletes de pares de isótopos de bromo que se desvían de las especies periódicas correspondientes al ion a1 de carga única.

La figura 15A y la figura 15B muestran los dobletes de pares de isótopos de bromo que se desvían de las especies periódicas correspondientes al ion d<sub>a2</sub> de carga única.

La figura 16 muestra el espectro de masas sin procesar, de referencia y de factor  $\beta$  de mioglobina marcada con 5-Br-3-PAA fragmentada en fuente en un espectrómetro de masas ESI-TOF.

35 La figura 17 muestra el resultado de usar código de "secuenciador" para determinar la secuencia de los primeros cinco residuos en mioglobina marcada con 5-Br-3-PAA (SEQ ID NOS:1 y 2).

40 La figura 18 ilustra una fórmula general para un marcador de defecto de masa que contiene una combinación de grupos ionizables ( $A_1 \dots A_n$ ), elementos de defecto de masa ( $B_1 \dots B_n$ ) y un resto reactivo de anhídrido succínico central (SA) (figura 18a) así como un esquema sintético global para un marcador de defecto de masa  $\{(A_1 \dots A_n)-(B_1 \dots B_n)-SA\}$  (figura 18b).

La figura 19 ilustra una técnica de secuenciación a modo de ejemplo usando los métodos descritos por Sanger en combinación con la estrategia de marcaje proporcionada en el presente documento.

La figura 20 ilustra bases marcadas que pueden usarse en los métodos de secuenciación proporcionados en el presente documento.

45 La figura 21 ilustra un espectro de masas de ddA\*/ddG\* (SEQ ID NO:5) (véase el ejemplo 18).

La figura 22 ilustra un espectro de masas de ddT\*/ddC\* (SEQ ID NOS:5 y 6) (véase el ejemplo 18).

## Descripción de la invención

### Definiciones



A menos que se definan de otra manera, todos los términos científicos y técnicos usados en el presente documento tienen generalmente el mismo significado que entiende comúnmente un experto en la técnica a la que esta invención pertenece. Generalmente, la nomenclatura usada en el presente documento y los procedimientos de laboratorio en biología molecular, química orgánica y química de proteínas descritos a continuación son los bien conocidos y comúnmente empleados en la técnica. Se usan técnicas convencionales para síntesis de péptidos. Generalmente, se realizan reacciones enzimáticas y etapas de purificación según las especificaciones del fabricante. Las técnicas y los procedimientos se realizan generalmente según métodos convencionales en la técnica y diversas referencias generales (véase de manera general, Sambrook *et al.* MOLECULAR CLONING: A LABORATORY MANUAL, 2ª ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., y Methods in Enzymology, Biemann, ed. 193:295-305, 351-360 y 455-479 (1993)), que se proporcionan en todo este documento. La nomenclatura usada en el presente documento y los procedimientos en análisis estadístico y matemático, química analítica y síntesis orgánica descritos a continuación son los conocidos y empleados en la técnica. Se usan técnicas convencionales, o modificaciones de las mismas, para síntesis químicas y análisis químicos.

La expresión “defecto de masa” o “marcador de defecto de masa” se refiere a una parte de un marcador o el marcador completo que proporciona una masa suficiente y distinta para que se identifique fácilmente en el espectro de masas de la muestra. Por consiguiente, el defecto de masa es normalmente un elemento que tiene un número atómico de desde 17 hasta 77 y más específicamente entre 35 y 63, que es distinto de azufre o fósforo. Los marcadores de defecto de masa más eficaces para su uso con compuestos químicos orgánicos típicos (incluso compuestos químicos orgánicos que contienen heteroátomos del grupo 1 y el grupo 2), tales como biomoléculas, incorporan uno o más elementos que tienen un número atómico de 35 a 63. Ejemplos de los defectos de masa más preferidos son los elementos bromo, yodo, europio e itrio.

Tal como se usa en el presente documento, el término “oligómero” se refiere a cualquier polímero de residuos en el que los residuos son similares, aunque normalmente no idénticos. Generalmente, un oligómero pretende incluir los polímeros que se producen de manera natural tales como proteínas, oligonucleótidos, ácidos nucleicos, oligosacáridos, polisacáridos y lípidos, y similares. Oligómero también puede referirse a polímeros de radicales libres, de condensación, aniónicos o catiónicos de origen sintético, tales como pero sin limitarse a: acrilatos, metacrilatos, nailones, poliésteres, poliimidas, cauchos de nitrilo, poliolefinas y copolímeros aleatorios o de bloque de diferentes monómeros en estas clases de polímeros sintéticos. El oligómero que se somete a los métodos analíticos descritos en el presente documento tendrá varios residuos que son típicos de su número que se produce de manera natural. Por ejemplo, un oligómero que es un oligonucleótido puede tener cientos e incluso miles de residuos. De manera similar, una proteína tendrá generalmente cien o más residuos (aunque la secuenciación de fragmentos más pequeños, por ejemplo, péptidos, también es útil). Un oligosacárido tendrá normalmente de desde 3 hasta 100 residuos de azúcar. Un lípido tendrá normalmente 2 ó 3 residuos de ácido graso.

Tal como se usa en el presente documento, los términos proteína, péptido y polipéptido se refieren a un polímero de residuos de aminoácido. Los términos también se aplican a polímeros de aminoácido en los que uno o más aminoácidos son análogos químicos de aminoácidos que se producen de manera natural correspondientes, incluyendo aminoácidos que se modifican mediante procesos postraduccionales (por ejemplo, glicosilación y fosforilación).

“Proteína”, tal como se usa en el presente documento, significa cualquier proteína, incluyendo, pero sin limitarse a péptidos, enzimas, glicoproteínas, hormonas, receptores, antígenos, anticuerpos, factores de crecimiento, etc., sin limitación. Las proteínas actualmente preferidas incluyen las compuestas por al menos 25 residuos de aminoácido, más preferiblemente al menos 35 residuos de aminoácido y todavía más preferiblemente al menos 50 residuos de aminoácido.

“Péptido” se refiere a un polímero en el que los monómeros son aminoácidos y se unen entre sí a través de enlaces amida, alternativamente denominado polipéptido. Cuando los aminoácidos son  $\alpha$ -aminoácidos, puede usarse o bien el isómero óptico L o bien el isómero óptico D. Adicionalmente, también se incluyen aminoácidos no naturales, por ejemplo,  $\beta$ -alanina, fenilglicina y homoarginina. Los aminoácidos pueden ser o bien el isómero D o bien el L. Se prefieren generalmente los isómeros L. Para una revisión general, véase, Spatola, A. F., en CHEMISTRY AND BIOCHEMISTRY OF AMINO ACIDS, PEPTIDES AND PROTEINS, B. Weinstein, eds., Marcel Dekker, Nueva York, p. 267 (1983).

“Etiqueta de secuenciación de proteína”, tal como se usa en el presente documento, se refiere a una serie contigua de al menos dos aminoácidos que representan una secuencia parcial de una proteína. Una PST preferida incluye un marcador de la invención o un fragmento de un marcador de la invención o un derivado ionizado de un marcador de la invención.

El término “energía de unión nuclear” se refiere a la disparidad de masas entre las masas nuclear real y calculada de los elementos. Se define como el equivalente de masa (según la teoría de la relatividad) de la energía necesaria para separar un núcleo en sus nucleones aislados constituyentes. Véase Bueche, F., “Principles of Physics” (McGraw-Hill, NY, 1977).

El término “deconvolución” define ampliamente algoritmos y procedimientos matemáticos para recuperar información

de interés a partir de datos que contienen ruido tanto aleatorio como periódico, o que se ha ocultado de otra manera mediante la interacción con métodos de recogida físicos o electrónicos.

El término "alquilo" se usa en el presente documento para referirse a un radical hidrocarbonado monovalente ramificado o no ramificado, saturado o insaturado, que tiene generalmente desde aproximadamente 1-30 carbonos y preferiblemente, desde 4-20 carbonos y más preferiblemente desde 6-18 carbonos. Cuando el grupo alquilo tiene desde 1-6 átomos de carbono, se denomina "alquilo inferior". Los radicales alquilo adecuados incluyen, por ejemplo, estructuras que contienen uno o más grupos metileno y/o metino. Las estructuras ramificadas tienen un motivo de ramificación similar a i-propilo, t-butilo, i-butilo, 2-etilpropilo, etc. Tal como se usa en el presente documento, el término abarca "alquilos sustituidos" y "alquilo cíclicos".

"Alquilo sustituido" se refiere a alquilo tal como se describió que incluye uno o más sustituyentes tales como, por ejemplo, alquilo inferior, arilo, acilo, halógeno (es decir, haloalquilo, por ejemplo,  $CF_3$ ), hidroxilo, amino, alcoxilo, alquilamino, acilamino, tioamido, aciloxilo, ariloxilo, ariloxialquilo, mercapto, tia, aza, oxo, hidrocarburos cíclicos tanto saturados como insaturados, heterociclos y similares. Estos grupos pueden unirse a cualquier carbono o sustituyente del resto alquilo. Adicionalmente, estos grupos pueden ser colgantes de, o integrales a, la cadena de alquilo.

El término "arilo" se usa en el presente documento para referirse a un sustituyente aromático, que puede ser un único anillo aromático o múltiples anillos aromáticos que están condensados entre sí, unidos covalentemente o unidos a grupo común tal como un resto metileno o etileno. El grupo de unión común también puede ser un carbonilo como en benzofenona. El/los anillo(s) aromático(s) pueden incluir fenilo, naftilo, bifenilo, difenilmetilo y benzofenona entre otros. El término "arilo" abarca "arilalquilo" y "arilo sustituido".

"Arilo sustituido" se refiere a arilo tal como se describió que incluye uno o más grupos funcionales tales como alquilo inferior, acilo, halógeno, alquilhalos (por ejemplo  $CF_3$ ), hidroxilo, amino, alcoxilo, alquilamino, acilamino, aciloxilo, fenoxilo, mercapto e hidrocarburos cíclicos tanto saturados como insaturados que están condensados con el/los anillo(s) aromático(s), unidos covalentemente o unidos a un grupo común tal como un resto metileno o etileno. El grupo de unión también puede ser un carbonilo tal como en ciclohexilfenilcetona. El término "arilo sustituido" abarca "arilalquilo sustituido".

El término "arilaquilo" se usa en el presente documento para referirse a un subconjunto de "arilo" en el que el grupo arilo está unido a otro grupo mediante un grupo alquilo tal como se define en el presente documento.

"Arilalquilo sustituido" define un subconjunto de "arilo sustituido" en el que el grupo arilo sustituido está unido a otro grupo mediante un grupo alquilo tal como se define en el presente documento.

El término "acilo" se usa para describir un sustituyente de cetona,  $-C(O)R$ , en el que R es alquilo o alquilo sustituido, arilo o arilo sustituido tal como se define en el presente documento.

El término "halógeno" se usa en el presente documento para referirse a átomos de flúor, bromo, cloro y yodo.

La expresión "serie de lantánido" se refiere a los elementos en la tabla periódica con números atómicos de entre 57 y 71.

El término "hidroxilo" se usa en el presente documento para referirse al grupo  $-OH$ .

El término "amino" se usa para designar  $-NRR'$ , en el que R y R' son independientemente H, alquilo, arilo o análogos sustituidos de los mismos. "Amino" abarca "alquilamino" que indica aminas secundarias y terciarias y "acilamino" que describe el grupo  $RC(O)NR'$ .

El término "alcoxilo" se usa en el presente documento para referirse al grupo  $-OR$ , en el que R es alquilo, o un análogo sustituido del mismo. Los radicales alcoxilo adecuados incluyen, por ejemplo, metoxilo, etoxilo, t-butoxilo, etc.

Tal como se usa en el presente documento, el término "ariloxilo" indica grupos aromáticos que están unidos a otro grupo directamente a través de un átomo de oxígeno. Este término abarca restos "ariloxilo sustituido" en los que el grupo aromático está sustituido tal como se describió anteriormente para "arilo sustituido". Los restos ariloxilo a modo de ejemplo incluyen fenoxilo, fenoxilo sustituido, benciloxilo, fenetiloxilo, etc.

Tal como se usa en el presente documento "ariloxialquilo" define grupos aromáticos unidos a través de un átomo de oxígeno a un grupo alquilo, tal como se define en el presente documento. El término "ariloxialquilo" abarca restos "ariloxialquilo sustituido" en los que el grupo aromático está sustituido tal como se describe para "arilo sustituido".

Tal como se usa en el presente documento, el término "mercapto" define restos de la estructura general  $-S-R$  en la que R es H, alquilo, arilo o heterocíclico tal como se describe en el presente documento.

La expresión "hidrocarburo cíclico saturado" indica grupos tales como el ciclopropilo, ciclobutilo, ciclopentilo, etc., y análogos sustituidos de estas estructuras. Estos hidrocarburos cíclicos pueden ser estructuras de un único anillo o de anillos múltiples.

La expresión “hidrocarburo cíclico insaturado” se usa para describir un grupo no aromático monovalente con al menos un doble enlace, tal como ciclopenteno, ciclohexeno, etc. y análogos sustituidos de los mismos. Estos hidrocarburos cíclicos pueden ser estructuras de un único anillo o de anillos múltiples.

5 El término “heteroarilo” tal como se usa en el presente documento se refiere a anillos aromáticos en los que uno o más átomos de carbono del/de los anillo(s) aromático(s) se sustituye(n) por un heteroátomo tal como nitrógeno, oxígeno o azufre. Heteroarilo se refiere a estructuras que pueden ser un único anillo aromático, múltiples anillo(s) aromático(s), o uno o más anillo(s) aromático(s) acoplados a uno o más anillo(s) no aromático(s). En estructuras que tienen múltiples anillos, los anillos pueden estar condensados entre sí, unidos covalentemente o unidos a un grupo común tal como un resto metileno o etileno. El grupo de unión común puede ser también un carbonilo como en fenilpiridilcetona. Tal como se usa en el presente documento, anillos tales como tiofeno, piridina, isoxazol, ftalimida, pirazol, indol, furano, etc. o análogos benzo-condensados de estos anillos se definen mediante el término “heteroarilo”.

“Heteroarilalquilo” define un subconjunto de “heteroarilo” en el que un grupo alquilo, tal como se define en el presente documento, une el grupo heteroarilo a otro grupo.

15 “Heteroarilo sustituido” se refiere a heteroarilo tal como se describió en el que el núcleo de heteroarilo está sustituido con uno o más grupos funcionales tales como alquilo inferior, acilo, halógeno, haloalquilo (por ejemplo CF<sub>3</sub>), hidroxilo, amino, alcoxilo, alquilamino, acilamino, aciloxilo, mercapto, etc. Por tanto, análogos sustituidos de anillos heteroaromáticos tales como tiofeno, piridina, isoxazol, ftalimida, pirazol, indol, furano, etc. o análogos benzo-condensados de estos anillos se definen mediante el término “heteroarilo sustituido”.

20 “Heteroarilalquilo sustituido” se refiere a un subconjunto de “heteroarilo sustituido” tal como se describió anteriormente en el que un grupo alquilo, tal como se define en el presente documento, une el grupo heteroarilo a otro grupo.

El término “heterociclo” se usa en el presente documento para describir un grupo no aromático saturado o insaturado monovalente que tiene un único anillo o múltiples anillos condensados de desde 1-12 átomos de carbono y de desde 1-4 heteroátomos seleccionados de nitrógeno, azufre u oxígeno dentro del anillo. Tales heterociclos son, por ejemplo, tetrahidrofurano, morfolina, piperidina, pirrolidina, etc.

25 La expresión “heterociclo sustituido” tal como se usa en el presente documento describe un subconjunto de “heterociclo” en el que el núcleo de heterociclo está sustituido con uno o más grupos funcionales tales como alquilo inferior, acilo, halógeno, haloalquilo (por ejemplo CF<sub>3</sub>), hidroxilo, amino, alcoxilo, alquilamino, acilamino, aciloxilo, mercapto, etc.

30 El término “heterocicloalquilo” define un subconjunto de “heterociclo” en el que un grupo alquilo, tal como se define en el presente documento, une el grupo heterocíclico a otro grupo.

El término “quelato” se refiere a la unión fuertemente asociativa de un elemento metálico o ion metálico a una molécula sustancialmente orgánica a través de medios no covalentes. Éstas se conocen alternativamente como moléculas organometálicas.

### 35 General

La presente invención reside en un método de espectrometría de masas para una discriminación mejorada de moléculas o fragmentos de moléculas marcadas o no marcadas en el espectrómetro de masas. Este método puede usarse para la determinación de la secuencia de oligómeros y para el aumento de la complejidad combinatoria que puede discriminarse en el espectro de masas. El presente método se pone en práctica marcando el extremo terminal de una molécula o un oligómero con un reactivo de marcaje que incorpora un defecto de masa, y discriminando las moléculas marcadas con el defecto de masa resultantes de otras moléculas no marcadas o fragmentos de oligómeros no marcados en el espectro de masas.

45 En una realización, pueden secuenciarse los oligómeros marcados con el oligómero marcado intacto fragmentado en o bien la zona de ionización de un espectrómetro de masas (por ejemplo, fragmentación en fuente) o en la celda de colisión de un instrumento de EM/EM, y usando un algoritmo matemático para determinar la secuencia terminal del oligómero a partir del extremo marcado. En otra realización, pueden sintetizarse oligómeros marcados a partir de un molde original o digerirse química o enzimáticamente para formar fragmentos que comprenden un marcador de masa molecular de secuenciación de fragmentos marcados que se identifican de manera algorítmica en el espectro de masas del defecto de masa diferencial del marcador. Los oligómeros marcados y los fragmentos de oligómero se diferencian de fragmentos y oligómeros no marcados por sus firmas de masa única en el espectro de masas resultante y se deconvolucionan de fragmentos de oligómeros no marcados y picos asociados con la matriz de ionización y oligómeros contaminantes y fragmentos por su abundancia relativa y/o firmas de masa única (debido al defecto de masa). Un sistema de clasificación acumulativo se usa mediante el algoritmo para fortalecer la certeza de la secuencia determinada en residuos sucesivos del marcador de masa molecular. En algunas realizaciones, este proceso se lleva a cabo en menos de 1 min. para una proteína marcada purificada, produciendo un método de 500 a 1000 veces más rápido que las técnicas de secuenciación de proteína por EM/EM actuales.

En una realización, los oligómeros marcados están altamente fragmentados en la EM por disociación inducida por colisión (CID). La CID puede llevarse a cabo en la zona de ionización (por ejemplo, en fuente) o en una celda de colisión a través de impacto de alta energía con gases que no son de oligómeros introducidos a la zona de colisión. Marcadores preferidos conducen a un aumento de la eficacia de ionización y volatilidad potenciada de los iones de fragmentos de oligómeros marcados resultantes, en relación con el oligómero original, mejorando así la sensibilidad de detección global. Los marcadores preferidos confieren una firma de masa única a los fragmentos a los que están unidos. En una realización particularmente preferida, la firma de masa única puede consistir en uno o más elementos incorporados en el marcador que contiene una energía de unión nuclear que sustancialmente difiere de la de los elementos asociados con los residuos de oligómeros (por ejemplo, C, H, O, N, P y S). En otra realización, una mezcla de versiones isotópicamente distintas de un marcador puede usarse simultáneamente con la abundancia relativa de los pares isotópicos resultantes para deconvolucionar picos de interés en el espectro de masas. En otra realización, pueden usarse análogos de marcador que difieren por adición de una o más unidades de metilo o metileno y/o análogos isotópicamente puros (por ejemplo, D frente a H o Cl) para distinguir de manera única picos de interés en el espectro de masas. Todavía en otra realización, los picos asociados con oligómeros, fragmentos o iones marcados pueden deconvolucionarse de oligómeros o fragmentos no marcados mediante su desplazamiento de masa. La secuencia del oligómero o la etiqueta de secuencia se construye preferiblemente a partir del extremo de bajo peso molecular del espectro de masas, proporcionando ventajas con respecto a métodos anteriores, tales como mayor precisión de masa absoluta y secuenciación más fácil. En el caso de proteínas, esta ventaja incluirá resolución de residuos Q y K, a partir de los fragmentos de péptido marcados resultantes.

La selección de un marcador apropiado para esta técnica requiere la consideración de varios criterios. En primer lugar, el marcador es preferiblemente lo suficientemente robusto para sobrevivir a las condiciones de fragmentación de la EM. En segundo lugar, el marcador preferiblemente crea también una firma de masa/carga única (m/z) (por ejemplo, un defecto de masa) que puede distinguirse de cualquier fragmento de oligómero no marcado generado a partir de cortes internos de la estructura principal del oligómero o de otras moléculas orgánicas no marcadas que pueden estar presentes en la muestra. En tercer lugar, el marcador también puede portar un grupo ionizable o permanentemente ionizado para garantizar que la fragmentación produce iones de abundancia alta que incluyen incluso residuos terminales no cargados.

A diferencia de la utilidad limitada de F como elemento de defecto de masa (Schmidt *et al.* documento WO 99/32501 (1 de julio de 1999)), la presente invención usa elementos de defecto de masa que presentan una diferencia de masa mucho mayor y por tanto utilidad más amplia. Por ejemplo, una única sustitución de yodo en un grupo arilo crea un defecto de masa de 0,1033 amu, una mejora de más de 5 veces con respecto a la de 5 sustituciones de F en arilo. Un único I en un anillo de arilo (C<sub>6</sub>H<sub>4</sub>I) presenta una masa monoisotópica de 202,935777 amu. Esto es 192 ppm diferente de la combinación más cercana de la molécula orgánica que contiene heteroátomo e isótopo estable ([<sup>12</sup>C]<sub>9</sub>[<sup>15</sup>N]<sub>1</sub>[<sup>16</sup>O]<sub>5</sub>) a 202,974687 amu. Por tanto, una única sustitución de cualquiera de los elementos que presentan un defecto de masa similar al de I (es decir, números atómicos de entre 35 y 63) producirá un defecto de masa discernible (en un nivel de 10 ppm) para una masa total de 3,891 amu para cualquier combinación de heteroátomos orgánicos. Dos elementos de este tipo presentarán un defecto de masa discernible para una masa total de 7,782 amu. Tres elementos de este tipo presentarán un defecto de masa discernible para una masa total de 11,673 amu. Alternativamente, las adiciones única, doble y triple de I (o un elemento de defecto de masa equivalente) pueden discriminarse entre sí para una masa total de 4,970 amu en un espectro de masas con resolución de masa de 10 ppm.

La presente invención incorpora un algoritmo robusto para la identificación de moléculas o fragmentos marcados con defecto de masa y la determinación de la secuencia de oligómero a partir de fragmentos de oligómero marcado en el espectro de masas. Este algoritmo busca los datos del espectro para todas las secuencias de oligómero posibles empezando sólo a partir de la masa del marcador, que se conoce. El algoritmo usa tanto la razón de masa con respecto a carga de los fragmentos de oligómero marcado como la abundancia relativa de los picos de EM resultantes para clasificar todas las secuencias de oligómero posibles. Se usa clasificación acumulativa (proyección futura) para eliminar secuencias a medida que se encuentran números sucesivos de determinados residuos en el espectro de masas. En una realización preferida, el ruido químico se deconvoluciona selectivamente del espectro de masas antes de la aplicación del algoritmo de secuenciación. A diferencia de algoritmos de secuenciación previos, el algoritmo actual es robusto debido a que puede implementarse sin intervención humana o bien para definir un ion original o de partida, o bien para identificar o calificar picos de secuencia prospectivos en el espectro de masas. En otra realización, las posibilidades de secuencias clasificadas más altas pueden calificarse además por su existencia en una base de datos de posibles secuencias de oligómero pronosticadas a partir de los datos de secuencia, particularmente una limitada al organismo del que se obtuvo el oligómero. En otra realización, las posibilidades de secuencias clasificadas más altas pueden calificarse además por las coordenadas de separación del oligómero original (por ejemplo, punto isoeléctrico y peso molecular de una proteína) y/o su composición de monómeros.

La presente invención incorpora uno o más elementos en el marcador que tienen una energía de unión nuclear (a menudo denominada defecto de masa) que desplaza la masa del marcador a una posición de masa única en el espectro que ninguna otra combinación estequiométrica de los otros elementos puede tener. De esta manera, los fragmentos marcados se distinguen más fácilmente del ruido químico y pueden detectarse con más precisión, cuando están presentes en abundancias relativas inferiores, y cuando están presentes en mezclas de muestras más complejas. Además, el método puede usarse para ayudar a identificar fragmentos marcados de abundancia inferior

producidos mediante diversos métodos de ionización (por ejemplo, iones d, y w producidos por fragmentación de proteínas y péptidos).

5 El uso de defectos de masa también puede aplicarse a la cuantificación de las abundancias relativas de la misma molécula obtenida a partir de dos o más fuentes en un espectrómetro de masas (véase, por ejemplo, los documentos WO 00/11208, EP1042345A1 y EP979305A1). Usando esta metodología particular, puede unirse un marcador a un oligómero que difiere de los otros marcadores por la sustitución de un elemento por un isótopo estable de ese elemento añadido a las moléculas a partir de cada fuente. Las fuentes se mezclan posteriormente al marcaje y la abundancia relativa de moléculas o los marcadores de cada fuente se cuantifica en el espectro de masas. Los diferentes isótopos se usan para diferencia de manera única los picos que surgen para la misma molécula a partir de cada fuente. La modificación de este método para incorporar uno o más elementos de defecto de masa en el marcador puede mejorar esta cuantificación debido a que los marcadores o moléculas marcadas resultantes se desplazarán de cualquier ruido químico en el espectro de masas resultante.

10 La invención puede usarse conjuntamente con métodos de secuenciación de proteínas, tales como secuenciación de marcador de masa molecular invertida (véase, publicación de PCT WO 00/11208) y otros métodos de secuenciación, cuantificación e identificación de proteínas por EM, tal como se explican resumidamente en la patente estadounidense n.º 6.027.890 y las publicaciones de PCT WO 99/32501 y WO 00/11208. El marcaje de defecto de masa tal como se describe en el presente documento también puede aplicarse a métodos de secuenciación de ADN por EM explicados resumidamente en las patentes estadounidenses n.ºs 5.700.642, 5.691.141, 6.090.558 y 6.194.144. Todavía adicionalmente, el marcaje de defecto de masa tal como se describe en el presente documento puede usarse para determinar la secuencia de polisacáridos (tal como el patrón de glicosilación de una proteína). Véanse métodos generales proporcionados en las patentes estadounidenses n.ºs 5.100.778 y 5.667.984.

15 Más ampliamente, el método puede usarse para mejorar la identificación (determinación de la secuencia) o cuantificación de cualquier polímero a partir de diferentes fuentes, ya sean naturales o sintéticas, siempre que un marcador de defecto de masa pueda unirse covalentemente a un extremo terminal del polímero.

20 La invención también puede usarse para la identificación estructural o la cuantificación relativa de especies químicas no poliméricas a partir de diferentes fuentes, siempre que puedan unirse covalentemente marcadores a estas moléculas. Los ejemplos incluyen: análisis de aminoácidos diferencial (tejidos con enfermedad frente a sanos); análisis de nucleótidos diferencial; análisis de sacáridos diferencial; análisis de ácidos grasos diferencial y determinación de la estructura de ácidos grasos ramificados e insaturados; determinación estructural y análisis de lípidos; y aplicaciones de control de calidad de nutrientes, y etiquetas de biblioteca combinatoria (tal como se describe en la patente estadounidense n.º 6.056.926 y por Brenner, S. y R. A. Lerner, "Encoded combinatorial chemistry", Proc. Natl. Acad. Sci. (EE.UU), 89:5381-5383 (1992)).

25 Volviendo en primer lugar al marcaje de defecto de masa de ácidos nucleicos (por ejemplo, ADN o ARN), cada una de las patentes estadounidenses n.ºs 6.090.558 y 6.194.144 describe cómo puede secuenciarse ADN a partir de fragmentos sintetizados que incorporan un marcador de masa única en la secuencia del cebador. En cambio, la presente invención proporciona que el marcaje se lleve a cabo usando sólo marcadores que tienen un defecto de masa, para distinguir los fragmentos marcados de fragmento no marcado y proporcionar un método más robusto, aunque sensible. Otra ventaja del uso de marcadores de defecto de masa es el aumento del número de diferentes ácidos nucleicos que pueden secuenciarse en paralelo de esta manera. Las ventajas del marcaje de defecto de masa (en lugar de un procedimiento de marcaje más general) no se dieron a conocer en el trabajo anterior.

30 De manera similar, los documentos WO 00/11208, EP1042345A1, EP979305A1 y la patente estadounidense n.º 6.027.890 describen el uso de marcadores de masa única para análisis diferencial y cuantificación de moléculas de proteína y ADN entre diferentes fuentes. Sin embargo, cada una de estas referencias no anticipa ni identifica las ventajas de incorporar un elemento de defecto de masa en el marcador de masa única.

35 Volviendo a continuación al marcaje de oligosacáridos, el documento EP 698218B1 describe el uso de hidratos de carbono marcados y su uso en ensayos y las patentes estadounidenses n.ºs 5.100.778 y 5.667.984 describen el uso de marcadores de masa para determinar la secuencia de oligosacáridos mediante EM. Mientras que las técnicas dadas a conocer en esos documentos podrían aplicarse al marcaje con etiquetas de masa única, la incorporación de un defecto de masa en el marcador para los fines de desplazar los picos de EM a regiones que no interfieran de los espectros no se da a conocer ni se aprecia. Por tanto, la aplicación de la metodología del marcaje de defecto de masa descrita en el presente documento proporciona métodos para identificar la secuencia de azúcares de un hidrato de carbono complejo marcando el hidrato de carbono tal como se describe en la técnica anterior (con modificación adecuada para la incorporación de un defecto de masa en el marcador) o mediante cualquier otro método disponible por los expertos en la técnica e identificando los fragmentos marcados de defecto de masa en el espectrómetro de masas. La estructura del hidrato de carbono puede determinarse por completo o en parte mediante adición de masa a partir de los fragmentos marcados más pequeños similares a los métodos de secuenciación de proteínas por EM/EM y de ADN descritos anteriormente. De nuevo, la incorporación de un elemento de defecto de masa en el marcador tiene utilidad para aislar los fragmentos marcados del ruido químico.

Volviendo a continuación a los lípidos, la composición y secuencia de ácidos grasos puede determinarse marcando los ácidos grasos digeridos enzimáticamente a partir de la estructura principal de glicidol, en la que se aplican marcadores de defecto de masa diferentes como etiquetas para digestiones enzimáticas diferentes que son específicas de secuencia.

5 En cada una de las aplicaciones observadas, los aminoácidos, lípidos y nucleótidos pueden derivarse mediante métodos generalmente disponibles para los expertos en la técnica. Si se usan marcadores isotópicamente distintos para una derivatización de este tipo de estas moléculas obtenidas o extraídas de diferentes muestras, entonces el análisis de cuantificación diferencial puede realizarse mediante EM. Sin embargo, en cada caso, la incorporación de un elemento de defecto de masa en el marcador mejora la capacidad para aislar las moléculas marcadas de otro ruido químico en el espectro y obtener mediciones de abundancia relativa más precisas. Sin embargo, no se anticipa en la técnica anterior la incorporación de números diferentes de elementos de defecto de masa en los marcadores para aumentar el número de muestras que pueden discriminarse simultáneamente en el espectro de masas resultante. Esta metodología también puede aplicarse para mejorar el aislamiento y la identificación de metabolitos en muestras biológicas (véase, por ejemplo, el documento estadounidense con n.º serie 09/553.424, presentado el 15 19 de abril de 2000), en el que una mezcla de metabolitos isotópicamente enriquecida obtenidos a partir de una fuente se derivatizan posteriormente con un marcador que contiene un defecto de masa para facilitar la identificación y cuantificación del metabolito isotópicamente enriquecido de la forma no enriquecida.

Además de la secuenciación e identificación de oligómeros, el marcaje de defecto de masa puede usarse para explorar con sondas la estructura y función de macromoléculas biológicamente activas (por ejemplo, oligómeros tales como proteínas, ácidos nucleicos y oligosacáridos).

La metodología del intercambio de deuterio (véase, Andersen, *et al.*, J. Biol. Chem. 276(17):14204-11 (2001)) se ha usado para explorar con sondas la estructura secundaria de proteínas y de orden superior y regiones implicadas en la unión a ligando. Los restos que están expuestos al disolvente y no están enterrados o escondidos por ligandos unidos intercambiarán hidrógeno por deuterio a una velocidad mucho más rápida en presencia de agua deuterada. La proteólisis posterior de la proteína y el análisis de los espectros de masas de los fragmentos proteolíticos deuterados y no deuterados pueden producir información sobre qué restos están implicados en elementos estructurales de orden superior específicos o en epítomos de unión.

Se proporcionan métodos mejorados en el presente documento, en los que se usan elementos de defecto de masa para marcar un oligómero u otra macromolécula, en lugar de deuterio. Usando moléculas pequeñas que incorporan elementos con defectos de masa que pueden seleccionar como diana grupos reactivos específicos y analizando patrones de fragmentación de, por ejemplo, muestras de proteínas proteolizadas o intactas, puede obtenerse información sobre la estructura o función buscando productos que están marcados de manera única o múltiple o no están marcados con el marcador de defecto de masa. Esta información se obtiene más fácil e inequívocamente mediante la reducción del ruido químico que proporciona el marcador de defecto de masa. Específicamente, puede exponerse una proteína activa a un marcador de defecto de masa tal como gas de bromo o yodo, que selecciona como diana residuos de tirosina de la proteína. Los residuos de tirosina se marcan de manera diferencial dependiendo de sus lugares geométricos (es decir, superficie frente a enterrado) y su participación en la unión a ligando. La proteína puede fragmentarse, con o sin proteólisis previa, y explorarse con sondas el patrón de marcaje de tirosina fácilmente en el espectrómetro de masas buscando picos que surgen de la incorporación de átomos de bromo o yodo.

Otra área en la que los marcadores de defecto de masa tienen uso beneficioso es en el análisis combinatorio de tanto moléculas pequeñas como macromoléculas que ya no contienen elementos con defectos de masa (por ejemplo, la mayoría de materiales biológicamente derivados). En esta aplicación, una mezcla compleja de entidades (por ejemplo, proteínas y péptidos, incluyendo anticuerpos y enzimas, polisacáridos, polinucleótidos, productos farmacéuticos o catalizadores) generados como una biblioteca combinatoria puede explorarse con sondas para determinar la actividad e identificarse incorporando elementos de etiquetado tal como se describe en la patente estadounidense n.º 6.056.926 y por Brenner, S. y R. A. Lerner, Encoded combinatorial chemistry, Proc. Natl. Acad. Sci., 89:5381-5383 (1992). Aumentando el número de etiquetas, y usando etiquetas que incorporan un elemento de defecto de masa, puede evaluarse una biblioteca combinatoria más grande. Las entidades que tienen características de unión deseadas presentarán un desplazamiento en masa igual al marcador de defecto de masa. Incluso en una mezcla muy compleja, es sencillo identificar los picos desplazados como resultado del defecto de masa.

## Descripción de las realizaciones

### *Secuenciación las partes terminales de oligómeros*

En vista de lo anterior, la presente invención proporciona en un aspecto un método para secuenciar una parte terminal de un oligómero tal como se define en la reivindicación 1.

En este aspecto de la invención, el oligómero puede obtenerse a partir de esencialmente cualquier fuente (por ejemplo, muestras tisulares, biopsias, extractos celulares y similares). Preferiblemente, el oligómero se aísla y purifica para que esté libre de componentes de interferencia. El oligómero aislado puede ponerse en contacto con un

- resto de marcaje para unir covalentemente un marcador al extremo terminal del oligómero para formar un oligómero marcado, adecuado para el análisis mediante métodos de fragmentación por espectrometría de masas. Pueden llevarse a cabo condiciones y métodos específicos para el marcaje de un oligómero según métodos establecidos dependiendo de los grupos funcionales presentes en el oligómero y los grupos reactivos presentes en el agente de marcaje. Normalmente, puede realizarse unión covalente a través de, por ejemplo, la formación de enlaces amida, éster, urea, tiourea, disulfuro, fosfodiéster, sulfonato, imina e hidrazida entre el marcador y el oligómero. Una discusión más completa de la química de unión se proporciona en secciones a continuación.
- 5
- La fragmentación del oligómero marcado resultante puede llevarse a cabo mediante métodos de espectrometría de masas, enzimáticos o quimiolíticos. Alternativamente, pueden generarse fragmentos mediante replicación incompleta de la molécula original (por ejemplo, secuenciación de ácidos nucleicos).
- 10
- Tras la fragmentación de la muestra de oligómero marcado, los fragmentos se analizan usando un método de espectrometría de masas para determinar la secuencia de al menos dos, más preferiblemente tres, todavía más preferiblemente cuatro, cinco o seis residuos terminales. En algunas realizaciones, se determina la secuencia de siete o más residuos terminales. Un método de espectrometría de masas preferido se describe en los ejemplos a continuación.
- 15
- En un grupo de realizaciones preferidas, el resto de marcaje comprende al menos un elemento que tiene un número atómico de 35 a 63, todavía más preferiblemente de 39 a 58. En un grupo de realizaciones particularmente preferidas, el resto comprende al menos un elemento seleccionado de bromo, yodo, europio e itrio. Todavía más preferiblemente, el resto de marcaje tiene desde uno hasta tres átomos de bromo o yodo.
- 20
- Los oligómeros preferidos que van a secuenciarse en este aspecto de la invención incluyen una proteína (o péptido), un oligonucleótido, un oligosacárido y un lípido. Más preferiblemente, el oligómero es una proteína (o péptido) o un oligosacárido.
- En un aspecto relacionado, la invención proporciona un método para secuenciar una parte de un oligómero en una mezcla de oligómeros tal como se define en la reivindicación 14.
- 25
- La mezcla de oligómeros en este aspecto de la invención puede ser una mezcla de proteínas, una mezcla de oligonucleótidos, una mezcla de oligosacáridos, una mezcla de lípidos y similares, y normalmente se obtiene a partir de un fuente biológica, tal como un lisado celular. Alternativamente, las muestras pueden obtenerse de tejidos animales (con enfermedad o sanos), extractos vegetales, fuentes bacterianas, fuentes virales y similares. Preferiblemente, la mezcla de oligómeros se ha purificado para reducir la cantidad de componentes que interfieren potencialmente usando métodos conocidos en la técnica.
- 30
- El marcaje de los oligómeros en la mezcla de oligómeros puede llevarse a cabo generalmente tal como se describió anteriormente para un único oligómero, usando los marcadores descritos en más detalle a continuación. La separación de los oligómeros marcados en la mezcla de oligómeros puede llevarse a cabo usando métodos tales como electroforesis capilar, cromatografía de líquidos de alta resolución (HPLC), electroforesis en gel (en todas sus formas), cromatografía (por ejemplo, de exclusión molecular, intercambio iónico, etc.), o cromatografía de gases, para nombrar algunos. Entonces, los oligómeros marcados individuales separados obtenidos a partir del procedimiento de separación pueden analizarse usando los métodos de espectrometría de masas descritos en el presente documento para determinar la secuencia de al menos dos residuos terminales. Preferiblemente, el método de espectrometría de masas es EM ESI-TOF.
- 35
- En un grupo de realizaciones preferidas, el resto de marcaje comprende al menos un elemento que tiene un número atómico de 35 a 63, todavía más preferiblemente de 39 a 58. En un grupo de realizaciones particularmente preferidas, el resto comprende al menos un elemento seleccionado de bromo, yodo, europio e itrio. Todavía más preferiblemente, el resto de marcaje tiene desde uno hasta tres átomos de bromo o yodo.
- 40
- En un aspecto relacionado, la invención proporciona un método para secuenciar la parte terminal de un oligómero tal como se define en la reivindicación 32.
- 45
- Aún en otro aspecto relacionado, la presente invención proporciona un método de secuenciación de una parte de un oligómero tal como se define en la reivindicación 38.
- En determinadas realizaciones preferidas, los oligómeros marcados fragmentados se reubican y someten a etapas adicionales de marcaje y fragmentación.
- 50
- Todavía en otro aspecto relacionado, la descripción proporciona un método para secuenciar una parte de un oligómero, que comprende:
- (a) preparar varias alícuotas de una muestra de oligómero
- (b) poner en contacto el oligómero en cada alícuota con un resto de marcaje para unir covalentemente un marcador a un extremo terminal del oligómero y formar una mezcla de oligómeros marcados, en la que cada alícuota está

marcada con un marcador que contiene un número diferente de elementos que tienen un número atómico de desde 17 hasta 77, con la condición de que dichos elementos sean distintos de azufre o fósforo,

(c) fragmentar el oligómero marcado en cada alícuota mediante un método quimiolítico o enzimático diferente,

(d) mezclar los productos de reacción de cada alícuota

5 (e) analizar la mezcla de oligómeros marcados fragmentados usando un método de fragmentación de espectrometría de masas para determinar la secuencia de al menos dos residuos terminales.

En una variación relacionada, la descripción proporciona un método para analizar simultáneamente productos de reacción múltiples para determinar secuencias de una muestra de oligómero, comprendiendo el método:

(a) preparar varias alícuotas de una muestra de oligómero;

10 (b) fragmentar el oligómero en cada alícuota mediante un método quimiolítico o enzimático diferente;

(c) poner en contacto los productos de reacción en cada alícuota con un resto de marcaje para unir covalentemente un marcador a un extremo terminal de los productos de reacción para formar una mezcla de producto de reacción marcados, en la que cada alícuota está marcada con un marcador que contiene un número diferente de elementos que tienen un número atómico de desde 17 hasta 77, con la condición de que dichos elementos sean distintos de azufre o fósforo;

15 (d) mezclar las mezclas de productos de reacción marcados de cada alícuota; y

(e) analizar las mezclas de productos de reacción marcados combinados de la etapa (d) usando un método de espectrometría de masas para determinar la secuencia de al menos dos monómeros en el oligómero.

20 En otra variación, la descripción proporciona un método para secuenciar una pluralidad de oligonucleótidos en una muestra, comprendiendo el método:

(a) marcar cebadores sintéticos poniendo en contacto cada cebador con un resto de marcaje para unir covalentemente un marcador a un extremo terminal de cada cebador para formar una mezcla de cebadores marcados, en la que cada cebador está marcado con un marcador que contiene un número diferente de elementos que tienen un número atómico de desde 17 hasta 77, con la condición de que los elementos sean distintos de azufre o fósforo;

25 (b) añadir un molde de muestra a la mezcla de cebadores marcados;

(c) sintetizar una pluralidad de oligómeros marcados del molde de muestra mediante métodos químicos o enzimáticos, y

30 (d) analizar los oligómeros marcados sintetizados en la etapa (c) usando un método de espectrometría de masas para determinar la secuencia del molde de muestra.

En cada uno de estos últimos aspectos y variaciones, los marcadores preferidos son los que contienen elementos que tienen números atómicos de desde 35 hasta 63, más preferiblemente, de 39 a 58 tal como se describió anteriormente para aspectos anteriores de la invención. Otras realizaciones preferidas son también las mismas que se han descrito anteriormente.

35 *Proteínas marcadas*

El marcaje de proteínas con diversos agentes en un medio de disolvente acuoso/orgánico mezclado o acuoso se conoce en la técnica y una amplia gama de técnicas y reactivos de marcaje útiles en la puesta en práctica de la presente invención están fácilmente disponibles para los expertos en la técnica. Véanse, por ejemplo, Means *et al.*, CHEMICAL MODIFICATION OF PROTEINS, Holden-Day, San Francisco, 1971; Feeney *et al.*, MODIFICATION OF

40 PROTEINS: FOOD, NUTRITIONAL AND PHARMACOLOGICAL ASPECTS, Advances in Chemistry Series, vol. 198, American Chemical Society, Washington, D.C., 1982; Feeney *et al.*, FOOD PROTEINS: IMPROVEMENT THROUGH CHEMICAL AND ENZYMATIC MODIFICATION, Advances in Chemistry Series, vol. 160, American Chemical Society, Washington, D.C., 1977; y Hermanson, BIOCONJUGATE TECHNIQUES, Academic Press, San Diego, 1996.

45 El marcaje puede realizarse y determinarse las PST a partir del extremo o bien N- o bien C-terminal de la proteína. Aproximadamente el 59-90% de las proteínas eucariotas están acetiladas en el extremo N-terminal (véase, Creighton, T. E., Proteins: Structures and Molecular Principles (W. H. Freeman, NY, 1984) y por tanto son resistentes al marcaje del extremo N-terminal. Sin embargo, el grupo N-acetil natural de tales proteínas puede usarse a veces como marcador para los fines de esta invención, pero sólo cuando uno o más de los aminoácidos dentro de los 4

50 residuos del extremo N-terminal es ionizable (por ejemplo, es un residuo de lisina, arginina, histidina, ácido aspártico o ácido glutámico) o puede derivatizarse para que sea ionizable (por ejemplo, residuos de tirosina, serina y cisteína).



Por consiguiente, se proporcionan estrategias para marcar extremos o bien N- o bien C-terminales para proporcionar el mayor grado de capacidad de secuenciación para cualquier proteína dada. Una vez que se selecciona un marcador, puede modificarse un algoritmo de deconvolución para buscar masas que corresponden a cualquier residuo modificado.

### 5 Características de los espectros de fragmentación

10 El espectro de masas (figura 1) es básicamente el número de iones (recuentos) que inciden en una placa detectora. El tiempo en el que los iones inciden en la placa detectora determina la razón de masa con respecto a carga ( $m/z$ ) del ion que incide en la placa. La placa detectora se calibra con moléculas de  $m/z$  conocida. A cada periodo de tiempo de exploración en la placa detectora se le asigna entonces un valor de  $m/z$  promedio y recoge iones con

15 Varias características del espectro de masas de proteínas fragmentadas pueden inhibir la capacidad para identificar o clasificar apropiadamente la secuencia de proteína verdadera, dependiendo de la intensidad de señal relativa de los péptidos marcados y se deconvolucionan mediante el algoritmo de la invención. Definiéndose la intensidad de señal relativa como la abundancia del ion del fragmento de péptido marcado en relación con la abundancia de otros iones y el ruido en el espectro de masas. La primera característica es que la incidencia en detector en primer lugar  
20 puede cargar el detector provocando un desplazamiento en el nivel de referencia y aumentando la abundancia en la región por encima de aproximadamente 500 amu. Esto se observa como un desplazamiento del nivel de referencia aparente en el espectro de masas (figura 1). Los estados de carga múltiple de la proteína original también pueden contribuir a las variaciones del nivel de referencia local de la misma manera en las posiciones de  $m/z$  por encima de aproximadamente 1000 amu. Esto se observa más claramente en la figura 1 en las posiciones de  $m/z$  por encima de  
25 aproximadamente 2000 amu.

La segunda característica observada es (figura 2) que las condiciones de alta fragmentación (por ejemplo, potenciales de boquilla altos para fragmentación en fuente) dan como resultado un aumento de la abundancia de iones de fragmentos en posiciones de masa periódica con respecto a carga en el espectrómetro de masas. En una  
30 escala de calibración de masa de  $^{12}\text{C}$  definida como 12.000000, estos fragmentos de proteína forman un patrón característico de picos separados aproximadamente 1 amu. En condiciones de fragmentación altamente eficaces un pico aparece a casi cada 1 amu de separación en el espectro de masas. Se observa que la separación de pico a pico promedio varía ligeramente con la proteína particular que está fragmentándose. Se cree que esto se debe a ligeras diferencias en la composición elemental de la proteína o de los fragmentos representados por los picos a cada amu.

35 En condiciones de alta fragmentación prácticamente todos los picos en el espectro de masas se superponen en este patrón de casi 1 amu (figura 3). Es esta observación la que permite los aspectos clave de la presente invención. En primer lugar, puesto que la mayoría de los picos se superponen en este patrón (o un estado de carga múltiple análogo de este patrón) es posible distinguir fácilmente los picos señal de los fragmentos marcados que se encuentran fuera de esta separación periódica, tales como fragmentos marcados en los que el marcador contiene uno o más elementos con una energía de unión nuclear inusual. En segundo lugar, la periodicidad permite la  
40 determinación de mínimos y máximos locales en el espectro de masas, de manera que el espectro puede corregirse para ruido local, permitiendo una mejor determinación de la abundancia real de recuentos en cada posición de masa con respecto a carga en el espectro de masas. En tercer lugar, puede determinarse una forma de pico característica o promedio para el ruido espectral no deseado en condiciones de alta fragmentación y este ruido se deconvoluciona o se resta del resto del espectro de masas, reduciendo así su contribución al algoritmo de clasificación y mejorando la confianza de la determinación de la secuencia producida por el algoritmo de la invención. Un experto en la técnica apreciará que también pueden encontrarse otros patrones de periodicidad más grandes en los datos y aplicarse de manera similar para ayudar en la deconvolución de la secuencia además de este patrón principal mostrado.

### Marcadores

50 Tal como se indicó anteriormente, las siguientes consideraciones son relevantes para la selección de un agente de marcaje:

- i) la masa del marcador es preferiblemente única y preferiblemente desplaza las masas del fragmento a regiones del espectro con fondo bajo;
- 55 ii) el marcador contiene preferiblemente cargas positivas o negativas fijas para dirigir la fragmentación de carga remota en el extremo N o C terminal;
- iii) el marcador es preferiblemente robusto en las condiciones de fragmentación y no experimenta fragmentación desfavorable;

iv) la química de marcaje es preferiblemente eficaz en un intervalo de condiciones, particularmente condiciones de desnaturalización, marcando de ese modo de manera reproducible y uniforme los extremos N- o C-terminales;

v) la proteína marcada se mantiene preferiblemente soluble en el sistema tampón de EM de elección;

vi) el marcador aumenta preferiblemente la eficacia de ionización del oligómero, o al menos no la suprime; y

5 vii) el marcador puede contener una mezcla de dos o más especies isotópicamente distintas para generar un patrón de espectrometría de masa única en cada posición de fragmento marcado.

10 En vista de los criterios de selección del marcador, restos de marcaje preferidos son los que tienen un componente de potenciación de la detección, un componente de firma de masa iónica y un grupo funcional reactivo, preferiblemente uno que reacciona con el extremo C- o N-terminal de una proteína. El grupo reactivo puede estar unido directamente a cualquiera o ambos de los otros dos componentes de marcador.

15 En una realización, se usan marcadores por parejas para aumentar adicionalmente la capacidad para identificar el marcador de masa molecular de otros picos en el espectro de masas. El uso de marcadores de isótopos mezclados es particularmente adecuado para la deconvolución adicional de los picos de fragmentos marcados, puesto que pares de isótopos abundantes sólo existirán para fragmentos marcados en el espectro de masas y los isótopos normalmente presentan eficacias de fragmentación e ionización similares. También pueden usarse análogos de un marcador que difieren en uno o más grupos metilo o metileno, o estado de carga. Incluso dos moléculas químicamente distintas pueden usarse en situaciones de marcaje doble para potenciar la identificación del marcador de masa molecular de fragmentos marcados. En una realización, una única muestra se marca simultáneamente con marcadores doble y se genera el espectro de masas combinado. En una realización preferida, se marcan independientemente muestras por duplicado y se mezclan en proporciones aproximadamente similares antes de la fragmentación en la EM. Una ventaja de esta última realización reside en la minimización de la posibilidad de dilución de la señal cuando también se marcan residuos laterales. En otra realización, se marcan muestras por duplicado con marcadores separados, se fragmentan por separado en la EM y los espectros de masas se añaden juntos para formar un espectro marcado doble virtual.

20 25 Todavía en otra realización, el grupo funcional reactivo está separado de uno o ambos de los componentes de potenciación de la detección y el componente de firma de masa iónica por un ligador. El ligador se diseña preferiblemente de manera que sea químicamente estable e inerte, y de manera que permita una separación eficaz del grupo reactivo y al menos uno de los otros dos componentes de la etiqueta. Dentro de una realización preferida de la invención, el ligador está compuesto por una cadena hidrocarbonada o, lo más preferiblemente, por una cadena hidrocarbonada unida a un anillo de arilo o heteroarilo y preferiblemente proporciona una separación adicional entre el grupo ionizable y el grupo de unión.

30 35 Tal como entenderá un experto habitual en la técnica, puede utilizarse una variedad de cadenas hidrocarbonadas y cadenas hidrocarbonadas modificadas dentro de la presente invención. Pueden encontrarse cadenas hidrocarbonadas preferidas que están unidas al anillo de fenilo en la familia de alcanos, oscilando ligadores particularmente preferidos entre 2 átomos de carbono y aproximadamente 20 átomos de carbono de longitud. Dentro de una realización preferida de la invención, el ligador es un fenetilo, amida alifática o grupo amina secundaria.

#### Componentes de potenciación de la detección

40 45 Un componente de potenciación de la detección, tal como se usa en el presente documento, se refiere a una parte del resto de marcaje que facilita la detección de los fragmentos de proteínas en el espectrómetro de masas. Por consiguiente, los componentes de potenciación de la detección pueden proporcionar una especie iónica cargada positivamente en condiciones de fragmentación en una cámara de ionización del espectrómetro de masas, o el componente puede proporcionar una especie iónica cargada negativamente en condiciones de fragmentación en una cámara de ionización del espectrómetro de masas. Para muchos de los componentes de potenciación de la detección, la cantidad de especies ionizadas presentes dependerá del medio usado para solubilizar la proteína. Los componentes de potenciación de la detección preferidos (es decir, especies que pueden generar una carga positiva o negativa) pueden clasificarse en tres categorías: 1) componentes que portan una carga "dura", 2) componentes que portan una carga "blanda" y 3) componentes que no proporcionan carga pero están en proximidad estrecha a residuos de proteína que portan una carga "blanda".

50 Los componentes que portan una carga "dura" son disposiciones de átomos que están sustancialmente ionizados en todas las condiciones, independientemente del pH del medio. Los componentes de potenciación de la detección cargados positivamente "duros" incluyen, pero sin limitarse a, grupos tetraalquil o tetraarilamonio, grupos tetraalquil o tetraarilfosfonio y grupos heterociclilo y heteroarilo N-alquilados o N-acilados (por ejemplo, piridinio). Los componentes de potenciación de la detección cargados negativamente "duros" incluyen, pero sin limitarse a, grupos borato de tetraalquilo o tetraacilo.

55 Los componentes que portan una carga "blanda" son disposiciones de átomos que están ionizados a un pH por encima o por debajo de su pKa, respectivamente (es decir, bases y ácidos). Dentro del contexto de la presente invención, las cargas positivas "blandas" incluyen las bases con un pKa de más de 8, preferiblemente más de 10, y

los más preferiblemente más de 12. Dentro del contexto de la presente invención, cargas negativas “blandas” incluyen los ácidos con un pKa de menos de 4,5, y preferiblemente menos de 2, y lo más preferiblemente menos de 1. En los extremos de pKa, las cargas “blandas” se aproximan a la clasificación como cargas “duras”. Los componentes de potenciación de la detección cargados positivamente “blandos” incluyen, pero sin limitarse a, grupos alquil o arilamónio 1º, 2º y 3º, grupos heteroarilo o heterociclilo sustituidos y no sustituidos (por ejemplo, piridinio), grupos imina o base de Schiff de alquilo o arilo, y grupos guanidino. Los componentes de potenciación de la detección cargados negativamente “blandos” incluyen, pero sin limitarse a, grupos carboxilato de alquilo o arilo, grupos sulfonato de alquilo o arilo, y grupos fosfato o fosfonato de alquilo o arilo.

Para ambos grupos cargados “duros” y “blandos”, tal como entenderá un experto habitual en la técnica, los grupos estarán acompañados por contraiones de carga opuesta. Por ejemplo, dentro de diversas realizaciones, los contraiones para grupos cargados positivamente incluyen oxianiones de ácidos orgánicos de alquilo inferior (por ejemplo, acetato), ácidos orgánicos halogenados (por ejemplo, trifluoroacetato) y organosulfonatos (por ejemplo, sulfonato de N-morfolinoetano). Los contraiones para grupos negativamente cargados incluyen, por ejemplo, cationes de amonio, cationes de alquil o arilamónio y cationes de alquil o arilsulfonio.

Pueden usarse componentes que son neutros pero están en proximidad estrecha a residuos de proteína que portan una carga “blanda” (por ejemplo, lisina, histidina, arginina, ácido glutámico o ácido aspártico) como componentes de potenciación de la detección. En este caso, el marcador porta grupos no ionizados o ionizables, y la potenciación de la detección se proporciona mediante un residuo de proteína cercano que porta carga. Dentro del contexto de la presente invención, proximidad estrecha se define como dentro de aproximadamente 4 residuos desde el extremo terminal marcado de la proteína, y más preferiblemente dentro de aproximadamente 2 residuos del extremo terminal marcado de la proteína.

Los componentes de potenciación de la detección del marcador también pueden estar cargados de manera múltiple o pueden cargarse de manera múltiple. Por ejemplo, un marcador con múltiples cargas negativas puede incorporar una o más especies cargadas de manera única (por ejemplo carboxilato) o puede incorporar una o más especies cargadas de manera múltiple (por ejemplo, fosfato). En un ejemplo representativo de esta realización de la invención, una especie que porta múltiples carboxilatos, tal como, por ejemplo un agente quelante de poliaminocarboxilato (por ejemplo, EDTP, DTPA) está unido a la proteína. Métodos de unión de poliaminocarboxilatos a proteínas y otras especies se conocen bien en la técnica. Véase, por ejemplo, Meares *et al.*, “Properties of In Vivo Chelate-Tagged Proteins and Polypeptides”. En, MODIFICATION OF PROTEINS: FOOD, NUTRITIONAL, AND PHARMACOLOGICAL ASPECTS;” Feeney, *et al.*, Eds., American Chemical Society, Washington, D.C., 1982, págs. 370-387; Kasina *et al.*, Bioconjugate Chem., 9: 108-117 (1998); Song *et al.*, Bioconjugate Chem., 8: 249-255 (1997).

De una manera similar, pueden adquirirse o prepararse marcadores que tienen cargas positivas múltiples usando métodos accesibles para los expertos en la técnica. Por ejemplo, un resto de marcaje que porta dos cargas positivas puede prepararse rápida y fácilmente a partir de una diamina (por ejemplo, etilendiamina). En una ruta sintética representativa, la diamina está monoprotectada usando métodos conocidos en la técnica y el resto amina no protegido se dialquila posteriormente con una especie que porta una o más cargas positivas (por ejemplo, (bromuro de 2-bromoetil)trimetilamónio) (Aldrich)). La desprotección usando métodos reconocidos en la técnica proporciona una especie de marcaje reactivo que porta al menos dos cargas positivas. Muchas rutas sintéticas sencillas de este tipo para especies de marcaje cargadas de manera múltiple resultarán evidentes para un experto en la técnica.

#### Componente de firma de masa iónica

El componente de firma de masa iónica es la parte del resto de marcaje que presenta preferiblemente una firma de masa iónica única en análisis de espectrometría de masas. El componente de firma de masa iónica incluye restos que no se ionizan eficazmente en las condiciones en las que las proteínas se ionizan (por ejemplo, compuestos de carbono aromáticos) así como moléculas que se ionizan fácilmente en las condiciones de ionización de proteínas para generar especies iónicas cargadas de manera múltiple. Ambos tipos de entidades químicas pueden usarse para desplazar la firma de masa iónica de los aminoácidos y péptidos unidos al marcador (tras la fragmentación de la proteína marcada) en el espectro de masas. Como resultado, los aminoácidos y péptidos marcados se distinguen fácilmente de los aminoácidos y péptidos no marcados por su patrón de masa iónica en el espectro de masas resultante. En una realización preferida, el componente de firma de masa iónica confiere una masa a un fragmento de proteína producido durante la fragmentación por espectrometría de masas que no coincide con la masa de residuo para ninguno de los 20 aminoácidos naturales.

En la realización más preferida, el componente de firma de masa iónica es cualquier elemento que presenta una energía de unión nuclear diferente de los constituyentes principales de proteínas. Los constituyentes principales de proteínas son: C, H, N, O y S. Definiendo energías de unión nuclear en cuanto al patrón de masa de  $^{12}\text{C} = 12,000000$  (figura 4), elementos preferidos con firmas de masa iónica única son aquellos elementos en la tabla periódica con números atómicos entre 17 (Cl) y 77 (Ir). Los elementos particularmente preferidos para su uso como componentes de firma de masa iónica del marcador incluyen elementos con números atómicos entre 35 (Br) y 63 (Eu). Los elementos más preferidos para su uso como componentes de firma de masa iónica son aquéllos con números atómicos entre 39 (Y) y 58 (Ce). Br y Eu son también componentes particularmente preferidos del

5 marcador puesto que cada uno de estos elementos presenta dos isótopos estables de proporciones aproximadamente iguales y energías de unión nuclear que difieren significativamente del patrón de picos periódico observado para proteínas fragmentadas en el espectrómetro de masas. Los elementos I y Y son también componentes de firma de masa iónica particularmente preferidos puesto que presentan la diferencia más grande en energía de unión nuclear del pico de fragmento de proteína periódico en el espectro de masas y debido a que se incorporan fácilmente en los marcadores. Los metales de transición están dentro de las listas preferidas y más preferidas de elementos de firma de masa iónica única y pueden incorporarse a menudo en los marcadores como quelatos, similares a los quelatos de Y conocidos (Lewis, *et al.*, *Bioconjugate Chemistry* 5:565-576 (1994)) y Eu (Zarling, *et al.*, patente estadounidense n.º 5.891.656).

10 En otra realización, puede crearse un componente de firma de masa iónica única usando un marcador cargado de manera múltiple. Un marcador cargado de manera múltiple de este tipo puede incorporar un elemento con una energía de unión nuclear diferente o puede consistir únicamente en elementos similares en energías de unión nuclear a la de los constituyentes de proteína principales. Estados de carga de este tipo pueden formarse con cargas “duras” o “blandas” o una combinación de cargas “duras” y “blandas” incorporadas en el marcador. Se prefieren estados de carga “dura” múltiples entre 2 y 4. Un estado de carga “dura” múltiple de 3 es lo más preferido cuando el marcador consiste únicamente en elementos con energías de unión nuclear similares a C, H, N, O y S. Un estado de carga “dura” múltiple de 2 es lo más preferido cuando el marcador contiene al menos un elemento que muestra una energía de unión nuclear diferente de C, H, N, O y S.

20 Tal como entenderá un experto en la técnica, pueden surgir picos de espectros de masas falsos no sólo de la fragmentación de aminoácidos y péptidos no marcados sino también de impurezas en la muestra y/o matriz. Con el fin de aumentar adicionalmente la singularidad de la firma de masa iónica del marcador y poder identificar picos de fragmentos marcados deseados con respecto al “ruido”, es preferible desplazar los fragmentos marcados a regiones de menos ruido espectral optimizando la masa del marcador. Por ejemplo, se prefiere que la masa del marcador genere un ion mayor que 100 amu y menor que 700 amu. El límite superior en la masa del marcador se determina mediante la resolución de masa del espectrómetro de masas usado. Esto puede hacerse aumentando el peso molecular de un marcador de bajo peso molecular o aumentando el número de cargas en un marcador de alto peso molecular.

30 Un método alternativo para proporcionar una firma de masa más única a un resto de marcaje es incorporar isótopos estables en el marcador (véase, por ejemplo, Gygi *et al.*, *Nature Biotechnol.* 17: 994-999 (1999)). Por ejemplo, incorporando ocho átomos de deuterio en un resto de marcaje y marcando la proteína con una mezcla 50:50 del marcador deuterado y no deuterado, los fragmentos cargados de manera única resultantes que incluyen el marcador se identifican fácilmente como dobletes igualmente intensos; uno a la masa correspondiente a las especies con el marcador no deuterado y el otro a la masa correspondiente a las especies con el marcador deuterado con una separación de 8 amu. En una realización preferida, la diferencia de masa es mayor que aproximadamente 1 amu en el estado de carga única. En la realización más preferida, la diferencia de masa es de desde aproximadamente 4 hasta aproximadamente 10 amu en el estado de carga única. Se prefiere la incorporación de múltiples isótopos de elementos que presentan energías de unión nuclear significativamente diferentes de C, H, N, O y S. Los elementos Br y Eu son los más preferidos debido a que presentan dos abundancias isotópicas naturales de aproximadamente 50:50.

40 Otro método para proporcionar una firma de masa más única a un resto de marcaje es incorporar una mezcla de sustituciones de alquilo y/o arilo sobre el marcador, de manera que el correspondiente conjunto de picos de fragmentos puede reconocerse fácilmente en el espectro de masas. Por ejemplo, la proteína puede marcarse con una mezcla de un marcador que contiene un grupo trimetilamonio y el mismo marcador que contiene un grupo dimetiletilamonio en lugar del grupo trimetilamonio. Este resto de marcaje produce dos picos de iones de fragmentos para cada aminoácido en la secuencia que difiere en 14 amu entre sí. Resultará evidente para los expertos en la técnica que pueden derivarse muchas combinaciones de este tipo.

#### Grupos reactivos

50 Un tercer componente del resto de marcaje es un grupo funcional que es reactivo con el grupo amino del extremo N-terminal, el grupo amino del extremo C-terminal u otro constituyente del aminoácido de los extremos N- o C-terminales.

55 El grupo funcional reactivo puede ubicarse en cualquier posición en la etiqueta. Por ejemplo, el grupo reactivo puede ubicarse en un núcleo de arilo o en una cadena, tal como una cadena de alquilo, unida a un núcleo de arilo. Cuando el grupo reactivo está unido a un alquilo, o cadena de alquilo sustituida anclada a un núcleo de arilo, el grupo reactivo se ubica preferiblemente en una posición terminal de una cadena de alquilo. Clases y grupos reactivos de reacciones útiles en la práctica de la presente invención son generalmente aquéllos que se conocen bien en la técnica de química de bioconjugados. Clases actualmente favorecidas de reacciones son aquéllas que se realizan en condiciones relativamente suaves en un medio de disolvente acuoso/orgánico mezclado o acuoso.

Las químicas particularmente preferidas que seleccionan como diana los grupos amino primarios en proteínas (incluyendo el extremo N-terminal) incluyen, por ejemplo: fluoruros de arilo (véanse, Sanger, F., *Biochem. J.*, 39:507

- (1945); Creighton, T. E., *Proteins: Structures and Molecular Principles* (W. H. Freeman, NY, 1984); Niedewieser, A., en: *Methods in Enzymology*, 25:60-99 (1972); y Hirs, C.H.W., *et al.*, *Arch. Biochem. Biophys.*, 111:209-222 (1965), cloruros de sulfonilo (Gray, W. R., en: *Methods in Enzymology*, 25:121-137 (1972)), cianatos (Stark, G. R., en: *Methods in Enzymology*, 25:103-120 (1972)), isocianatos (Niall, H. D., en: *Methods in Enzymology*, 27:942-1011 (1973)), imidoésteres (Galella, G., *et al.*, *Can. J. Biochem.*, 60:71-80 (1982)), ésteres N-hidroxisuccinimídicos (Lomant, A.J., *et al.*, *J. Mol. Biol.*, 104:243-261 (1976)), O-acilisoureas (Lomant, A.J., *et al.*, *J. Mol. Biol.*, 104:243-261 (1976)), clorocarbonatos y carboniloazidas (Solomons, T.W.G, *Organic Chemistry* (John Wiley & Sons, NY, 1976)), aldehídos (Novotny *et al.*, *Anal. Chem.*, 63:408 (1991) y Novotny *et al.*, *J. Chromatography*, 499:579 (1990)), y haluros de alquilo y alquenos activados (Wagner, D.S., *et al.*, *Biol Mass Spectrometry*, 20:419-425 (1991)). Ejemplos preferidos de constituyentes químicos que reaccionan con los grupos carboxilo de proteínas son haluros de bencilo (Solomons, T.W.G, *Organic Chemistry* (John Wiley & Sons, NY, 1976); Merrifield, B., *Science*, 232: 341-347 (1986); y Horton, H. R., *et al.*, *Methods in Enzymology*, 25:468 (1972)) y carbodiimida (Yamada, H., *et al.*, *Biochem.*, 20:4836-4842)), particularmente si se estabiliza usando N-hidroxisuccinimida (véase, Grabarek, Z., *et al.*, *Anal. Biochem.* 185:131-135 (1990)). Ambos de estos enfoques de marcaje de carboxilo se espera que marquen residuos de aminoácido que contienen carboxilo (por ejemplo, aspartato y glutamato) junto con el del extremo C-terminal. Estas y otras reacciones útiles se discuten en, por ejemplo, March, *ADVANCED ORGANIC CHEMISTRY*, 3ª ed., John Wiley & Sons, Nueva York, 1985; Hermanson, *BIOCONJUGATE TECHNIQUES*, Academic Press, San Diego, 1996; y Feeney *et al.*, *MODIFICATION OF PROTEINS*; *Advances in Chemistry Series*, vol. 198, American Chemical Society, Washington, D.C., 1982.
- Los grupos funcionales reactivos pueden elegirse de manera que no participan en, o interfieren con, las reacciones necesarias para ensamblar la etiqueta. Alternativamente, puede impedirse que un grupo funcional reactivo participe en la reacción mediante la presencia de un grupo protector. Los expertos en la técnica entienden cómo proteger un grupo funcional particular de manera que no interfiera con un conjunto escogido de condiciones de reacción. Para ejemplos de grupos protectores útiles, véase, por ejemplo, Greene *et al.*, *PROTECTIVE GROUPS IN ORGANIC SYNTHESIS*, John Wiley & Sons, Nueva York, 1991.

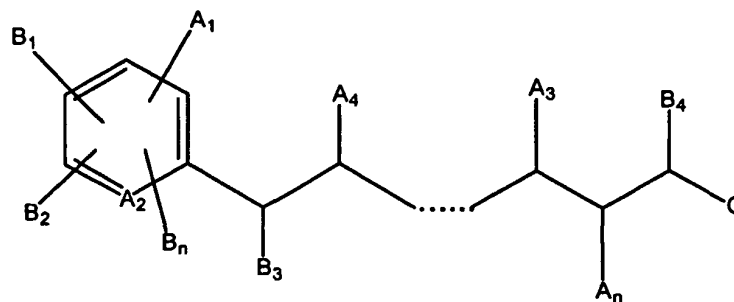
Las figuras 5 y 6 proporcionan las estructuras de una variedad de restos de marcaje que tienen uno o más átomos de bromo (figura 5) o yodo (figura 6). Los marcadores adecuados incluyen ácido 5-bromonicotínico, ácido 6-bromo-2-hidroxi-quinolin-4-carboxílico (BHQC), ácido 6-bromopicolínico, 4-bromobenzaldehído, así como otros ácidos arilcarboxílicos sustituidos con bromo disponibles comercialmente. Todavía otros marcadores pueden prepararse adecuadamente mediante bromación o yodación de un ácido arilcarboxílico o una forma protegida del mismo. Los expertos en la técnica conocen bien métodos para la bromación y yodación de compuestos aromáticos y pueden encontrarse referencias para su preparación en textos generales tales como March, *ADVANCED ORGANIC CHEMISTRY*, 4ª ed., John Wiley & Sons. Un experto en la técnica apreciará además que los residuos de ácido carboxílico normalmente se activan (por ejemplo, como un éster de NHS) antes de ponerse en contacto con una proteína. Adicionalmente, el examen de los residuos de marcaje indica que la mayoría tienen un átomo de nitrógeno ionizable (presente como un grupo amonio a un pH por debajo del  $pK_a$  del ion amonio).

La tabla 1 proporciona una descripción no limitativa de restos de marcaje útiles en los marcadores de la presente.

Tabla 1

Marcador de defecto de masa genérico

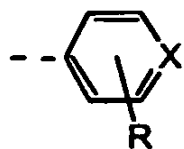
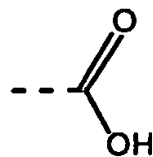
- Los restos A portan carga (positiva o negativa) para la ionización por EM.
- Los restos B son elementos de defecto de masa.
- Los restos C son grupos reactivos para la unión a biomoléculas.
- Los restos A, B y C se ubican en una variedad de estructuras aromáticas/alifáticas



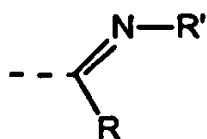
1A. Restos A<sub>n</sub> a modo de ejemplo



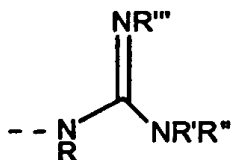
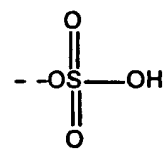
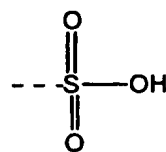
R,R'= H, alquilo, aromático



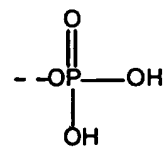
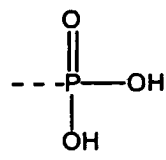
R=H, alquilo, aromático, OH, OR', NO<sub>2</sub>, Cl, F...  
X=N u otro heterociclo básico



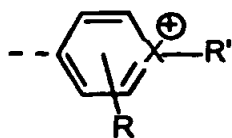
R,R'=H, alquilo, aromático



R,R',R'',R'''=H, alquilo, aromático



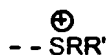
R,R'=H, alquilo, aromático  
R,R',R''=alquilo, aromático



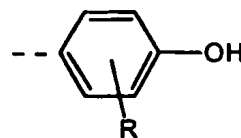
R=H, alquilo, aromático, OH, OR'', NO<sub>2</sub>, Cl, F...  
R'=alquilo, aromático  
X=heterociclo que contiene N, S, O



R,R',R''=alquilo, aromático



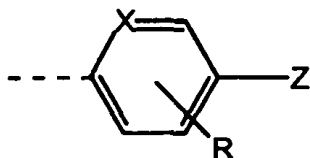
R,R'=alquilo, aromático



R,R'=alquilo, aromático

R=H, alquilo, aromático, NO<sub>2</sub>, Cl, F...

1B. Restos B<sub>n</sub> a modo de ejemplo:

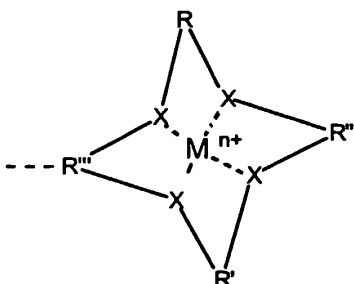


(Z=Br, I)  
(R=H, alquilo, aromático, OH, OR', Cl, F...)  
(X=C, N, S,



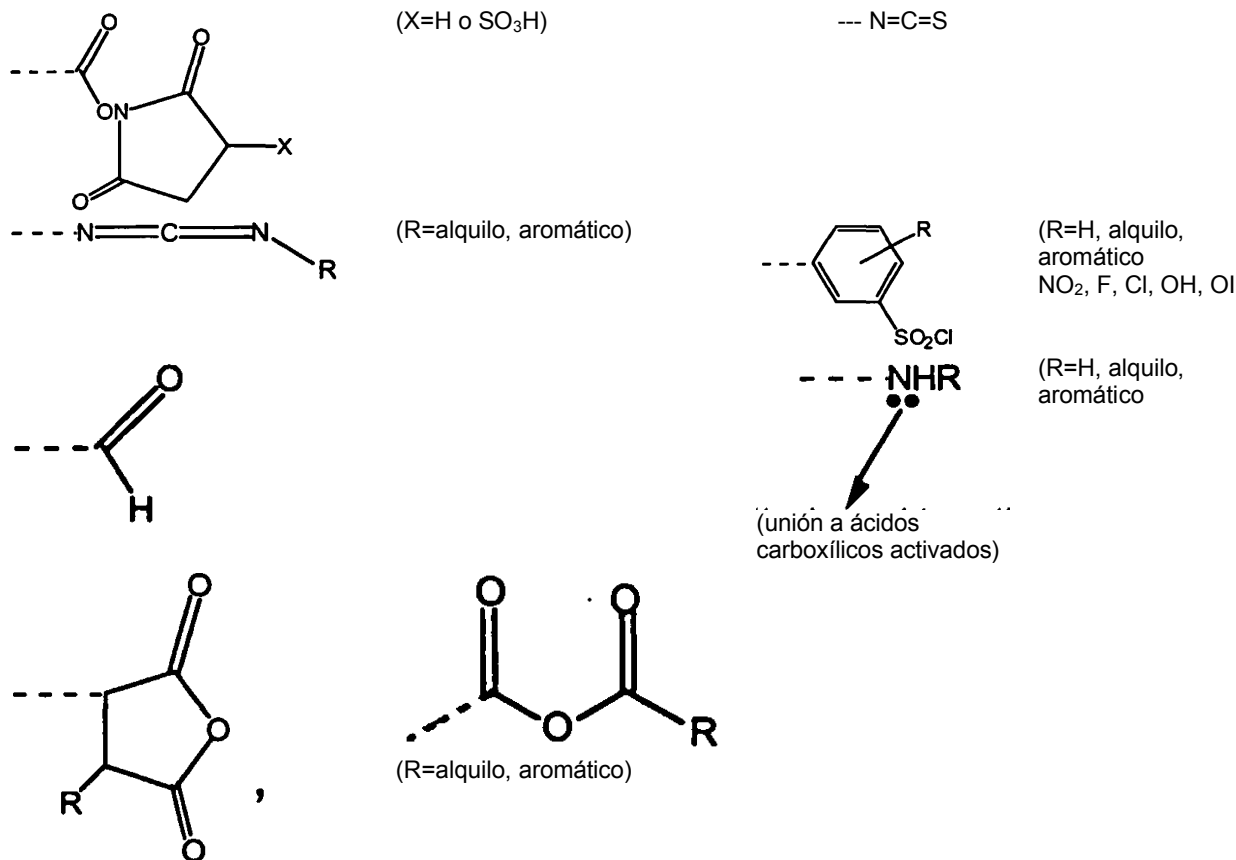
--CZ<sub>3</sub>

(Z=Br, L)



quelatos de metal  
(M=Y, La, Ce, Eu...; -5)  
(X=N, O, S)  
(R,R',R'',R'''=alquilo, aromático)  
(Nota: La estructura quelante puede ser geometría de cualquier tamaño con cualquier número de átomos de coordinación)

1C. Restos C a modo de ejemplo:



Síntesis

Una vez que el grupo reactivo, ligador, y los grupos ionizables o componentes de firma de masa iónica se han seleccionado, el compuesto final se sintetiza utilizando reacciones de química orgánica convencionales. Un compuesto preferido para su uso dentro de la presente invención es el éster de NHS del ácido 5-bromo-3-piridilacético (5-Br-3-PAA) o un agente análogo.

Con la selección de un resto de marcaje adecuado, las condiciones para unir el marcador a la proteína deben garantizar que los extremos N- o C-terminales de la proteína están marcados uniformemente y que la proteína marcada se mantiene soluble en sistemas de tampón de EM apropiados. Normalmente, el marcaje se llevará a cabo en condiciones desnaturalizantes (por ejemplo, tensioactivos o urea 8 M). Los tensioactivos y la urea suprimen ambos la ionización por EM y también deben emplearse métodos que proporcionan limpieza y transferencia rápida de la muestra de proteína marcada a un tampón de EM adecuado.

Restos detectables

En otra realización preferida, la proteína se marca con un resto que potencia su detectabilidad en, por ejemplo, procedimientos de purificación y separación de proteínas (por ejemplo, electroforesis). El resto detectable puede detectarse mediante, por ejemplo, espectroscopía (por ejemplo, UV/Visible, de fluorescencia, resonancia de espín electrónico (ESR), resonancia magnética nuclear (RMN) y similares), detección de isótopos radioactivos, etc. Cuando la proteína se detecta mediante UV/Vis, es deseable generalmente unir un marcador cromóforo a la proteína (por ejemplo, fenilo, naftilo, etc.). De manera similar, para la detección mediante espectroscopía de fluorescencia, un fluoróforo se une preferiblemente a la proteína. Por ejemplo, Quantum Dye™ es un quelato de Eu fluorescente y éster succinimidílico de 5-carboxi-2',4',5',7'-tetrabromosulfonafluoresceína es un fluoróforo que contiene bromo de reactivo N-terminal, (disponible comercialmente de Research Organics, n.º de catálogo 0723Q y Molecular Probes, n.º de catálogo C-6166, respectivamente). Para ESR, el resto detectable puede ser un radical libre, tal como un resto que incluye un grupo nitróxido. Cuando la proteína se detecta mediante un método de RMN, el resto detectable puede enriquecerse con núcleos accesibles para NMR, tales como flúor, <sup>13</sup>C y similares.

En una realización actualmente preferida, el resto detectable es un fluoróforo. Muchos marcadores fluorescentes reactivos están disponibles comercialmente de, por ejemplo, la compañía SIGMA chemical (Saint Louis, MO), Molecular Probes (Eugene, OR), R&D systems (Minneapolis, MN), Pharmacia LKB Biotechnology (Piscataway, NJ),

CLONTECH Laboratories, Inc. (Palo Alto, CA), Chem Genes Corp., Aldrich Chemical Company (Milwaukee, WI), Glen Research, Inc., GIBCO BRL Life Technologies, Inc. (Gaithersburg, MD), Fluka Chemica-Biochemika Analytika (Fluka Chemie AG, Buchs, Suiza) y PE-Applied Biosystems (Foster City, CA), así como muchas otras fuentes comerciales conocidas por un experto en la técnica. Además, los expertos en la técnica reconocerán cómo seleccionar un fluoróforo apropiado para una aplicación particular y, si no está fácilmente disponible comercialmente, podrán sintetizarse el fluoróforo necesario *de novo* o modificar sintéticamente compuestos fluorescentes disponibles comercialmente para llegar al marcador fluorescente deseado.

Existe una gran cantidad de guías prácticas disponibles en la literatura para seleccionar un fluoróforo apropiado para una etiqueta particular, tal como se muestra a modo de ejemplo mediante las siguientes referencias: Pesce *et al.*, Eds., FLUORESCENCE SPECTROSCOPY (Marcel Dekker, Nueva York, 1971); White *et al.*, FLUORESCENCE ANALYSIS: A PRACTICAL APPROACH (Marcel Dekker, Nueva York, 1970); y similares. La bibliografía también incluye referencias que proporcionan listas exhaustivas de moléculas fluorescentes y cromógenas y sus propiedades ópticas relevantes, para escoger pares de indicador-extintor (véanse, por ejemplo, Berlman, HANDBOOK OF FLUORESCENCE SPECTRA OF AROMATIC MOLECULES, 2ª Edición (Academic Press, Nueva York, 1971); Griffiths, COLOUR AND CONSTITUTION OF ORGANIC MOLECULES (Academic Press, Nueva York, 1976); Bishop, Ed., INDICATORS (Pergamon Press, Oxford, 1972); Haugland, HANDBOOK OF FLUORESCENT PROBES AND RESEARCH CHEMICALS (Molecular Probes, Eugene, 1992) Pringsheim, FLUORESCENCE AND PHOSPHORESCENCE (Interscience Publishers, Nueva York, 1949); y similares. Además, hay guías extensas en la bibliografía para derivatizar moléculas indicadoras y extintoras para su unión covalente mediante grupos reactivos fácilmente disponibles que pueden añadirse a una molécula.

La diversidad y utilidad de las químicas disponibles para conjugar fluoróforos con otras moléculas y superficies se muestra a modo de ejemplo mediante el extenso cuerpo de bibliografía sobre la preparación de ácidos nucleicos derivatizados con fluoróforos. Véase, por ejemplo, Haugland (citado anteriormente); Ullman *et al.*, patente estadounidense n.º 3.996.345; Khanna *et al.*, patente estadounidense n.º 4.351.760. Por tanto, está muy dentro las capacidades de los expertos en la técnica escoger un par de intercambio de energía para una aplicación particular y conjugar los miembros de este par con una molécula de sonda, tal como, por ejemplo, un material bioactivo molecular pequeño, ácido nucleico, péptido u otro polímero.

Además de fluoróforos que se unen directamente a una proteína, los fluoróforos también pueden unirse por medios indirectos. En una realización a modo de ejemplo, una molécula de ligando (por ejemplo, biotina) se une covalentemente de manera preferible a la proteína. El ligando se une entonces a otra molécula (por ejemplo, estreptavidina), que o bien puede detectarse de manera inherente o bien está unida covalentemente a un sistema señal, tal como un compuesto fluorescente de la invención, o una enzima que produce un compuesto fluorescente mediante conversión de un compuesto no fluorescente. Enzimas útiles de interés como marcadores incluyen, por ejemplo, hidrolasas, particularmente fosfatasa, esterasas y glicosidasas, u oxidasas, particularmente peroxidasas. Los compuestos fluorescentes incluyen fluoresceína y sus derivados, rodamina y sus derivados, dansilo, umbeliferona, etc., tal como se comentó anteriormente. Para una revisión de diversos sistemas de producción de señal o marcaje que pueden usarse, véase la patente estadounidense n.º 4.391.904.

Puede hacerse que restos detectables fluorescentes florezcan excitándolos de cualquier manera conocida en la técnica, incluyendo, por ejemplo, con energía electroquímica o luz (véase, por ejemplo, Kulmala *et al.*, Analytica Chimica Acta 386: 1 (1999)). Los expertos en la técnica conocen bien medios de detección de marcadores fluorescentes. Por tanto, por ejemplo, pueden detectarse marcadores fluorescentes excitando el fluoróforo con la longitud de onda de luz apropiada y detectando la fluorescencia resultante. La fluorescencia puede detectarse visualmente, por medio de película fotográfica, mediante el uso detectores electrónicos tales como dispositivos acoplados por carga (CCD) o fotomultiplicadores y similares. De manera similar, los marcadores enzimáticos pueden detectarse proporcionando los sustratos apropiados para la enzima y detectando el producto de reacción resultante.

Cuantas menos sean las etapas del procedimiento entre cualquier técnica de separación y el método de secuenciación por EM, más rápido pueden identificarse las proteínas, y menor es el coste de investigación proteómica. Los tampones de electroforesis típicos (por ejemplo, Hochstrasser *et al.* Anal Biochem., 173:424 (1988) y O'Farrel, J Biol. Chem., 250:4007 (1975)) contienen componentes (por ejemplo, tampones de tris(hidroximetil)aminometano y dodecil sulfato de sodio, que suprimen la ionización de proteínas en el espectrómetro de masas. Estos componentes pueden sustituirse por otros componentes más volátiles (por ejemplo, tampones de morfolinoalquilsulfonato y tensioactivos de efemeral) que no suprimen la ionización en la EM. En otra realización, las muestras se diluyen con tampón bicarbonato de amonio o acetato de amonio para proporcionar una fuente de protones volátiles para el espectrómetro de masas. Wilm, M. *et al.*, Anal. Chem., 68:1-8 (1996). En otra realización, se realiza un intercambio de tampón a través de diálisis de flujo tangencial o cromatográfico a medida que la muestra se transporta desde la salida del procedimiento de separación hasta la entrada de la EM.

#### Procedimiento de marcaje

En algunos casos, las sales (por ejemplo, TRIS y SDS) y urea presentes en los tampones de electroforesis pueden suprimir la ionización de las proteínas marcadas y pueden generar iones de masa/carga pequeñas que complican potencialmente el análisis de secuencia. Por consiguiente, pueden emplearse procedimientos de diálisis por



centrifugación para intercambiar rápidamente los sistemas tampón antes del análisis de EM. Alternativamente, pueden usarse columnas de desalinización (por ejemplo, la punta ZipTip™ vendida por Millipore) para la limpieza de la muestra y el intercambio de tampón. Las muestras desalinizadas pueden resuspenderse en bicarbonato de amonio 0,1 M tal como se describe por Wilm y Mann (véase, Wilm, *et al.*, *ibid.*) con adición mínima de metanol, o en tampón acetato de amonio 0,01 M (con ácido fórmico al 0,1%) con adición mínima de acetonitrilo tal como se describe por Mark (véase "Protein structure and identification with MS/MS," artículo presentado en la serie de seminarios PE/Sciex, Protein Characterization and Proteomics: Automated high throughput technologies for drug discovery, Foster City, CA (marzo de 1998)).

Las tasas de acoplamiento del compuesto pueden someterse a prueba para garantizar que el compuesto es adecuado para secuenciar polipéptidos. En general, cuanto más rápida sea la tasa de acoplamiento más preferido será el compuesto. Tasas de acoplamiento de entre 2 y 10 minutos a de 50°C a 70°C son particularmente preferidas. De manera similar, también se prefieren tasas de reacción rápidas, porque la exposición a la mezcla de reacción durante un período de tiempo prolongado podría hidrolizar los enlaces peptídicos, o conducir a reacciones secundarias irreproducibles e ineficaces con los residuos polipeptídicos, que podrían complicar la deconvolución del espectro de masas.

En otra realización preferida, uno o más de los componentes de una mezcla de proteínas se une de manera reversible a un soporte sólido antes de que el marcador se una a un polipéptido. Pueden usarse diversos materiales como soportes sólidos, incluyendo, por ejemplo, numerosas resinas, membranas o papeles. Estos soportes pueden derivatizarse adicionalmente para incorporar una funcionalidad escindible. Varios grupos escindibles que pueden usarse para este fin incluyen enlaces disulfuros (-S-S-), glicoles (-CH[OH]-CH[OH]-), azo (-N=N-), sulfona (-SO<sub>2</sub>-), y éster (-COO-) (véase, Tae, *Methods in Enzymology*, 91:580 (1983)). Soportes que se prefieren particularmente incluyen membranas tales como Sequelon™ (Milligen/ Biosearch, Burlington, Mass.). Materiales representativos para la construcción de estos soportes incluyen, entre otros, poliestireno, vidrio poroso, poli(fluoruro de vinilideno) y poli(acrilamida). En particular, los soportes de poliestireno incluyen, entre otros: (1) un (2-aminoetil)aminometil-poliestireno (véase, Laursen, *J. Am. Chem. Soc.* 88: 5344 (1966)); (2) un poliestireno similar al de número (1) con un grupo arilamino (véase, Laursen, *Eur. J. Biochem.* 20: 89 (1971)); (3) amino-poliestireno (véase, Laursen *et al.*, *FEBS Lett.* 21: 67 (1972)); y (4) trietilentetramina-poliestireno (véase, Horn *et al.*, *FEBS Lett.* 36:285 (1973)). Los soportes de vidrio poroso incluyen: (1) vidrio de 3-aminopropilo (véase, Wachter *et al.*, *FEBS Lett.* 35: 97 (1973)); y vidrio de (2)N-(2-aminoetil)-3-aminopropilo (véase, Bridgen, *FEBS Lett.* 50: 159 (1975)). La reacción de estos soportes de vidrio poroso derivatizado con diisotiocianato de p-fenileno conduce a vidrios de isotiocianato activados (véase, Wachter *et al.*, citado anteriormente). Soportes a base de poli(acrilamida) también son útiles, incluyendo una β-alanilhexametildiamina-polidimetilacrilamida reticulada (véase, Atherton *et al.*, *FEBS Lett.* 64: 173 (1976)), y una N-aminoetil-poli(acrilamida) (véase, Cavadore *et al.*, *FEBS Lett.* 66: 155 (1976)).

Un experto habitual en la técnica utilizará fácilmente química apropiada para acoplar el polipéptido a los soportes sólidos descritos anteriormente (véase, generalmente Machleidt y Wachter, *Methods in Enzymology*: [29] *New Supports in Solid-Phase Sequencing* 263-277 (1974)). Métodos de acoplamiento y soportes preferidos incluyen el uso de papel de fibra de vidrio de aminofenilo con acoplamiento de EDC (véase, Aebersold *et al.*, *Anal. Biochem.* 187: 56-65 (1990)); filtros de vidrio DITC (véase, Aebersold *et al.*, *Biochem.* 27: 6860-6867 (1988) y la membrana de poli(fluoruro de vinilideno) (PVDF) (Immobilon P™, Milligen/ Biosearch, Burlington, Mass.), junto con el procedimiento químico SequeNet™ (véase, Pappin *et al.*, *CURRENT RESEARCH IN PROTEIN CHEMISTRY*, Villafranca J. (ed.), págs. 191-202, Academic Press, San Diego, 1990)).

En la práctica de la presente invención, la unión del polipéptido al soporte sólido puede producirse mediante interacción o bien covalente o bien no covalente entre el polipéptido y el soporte sólido. Para la unión no covalente del polipéptido al soporte sólido, el soporte sólido se escoge de manera que el polipéptido se une al soporte sólido mediante interacciones no covalentes. Por ejemplo, un soporte sólido de fibra de vidrio puede recubrirse con Polybrene, una sal de amonio cuaternario polimérica (véase, Tarr *et al.*, *Anal. Biochem.*, 84:622 (1978)), para proporcionar una superficie de soporte sólido que se unirá de manera no covalente al polipéptido. Otras fases sólidas de adsorción adecuadas están disponibles comercialmente. Por ejemplo, los polipéptidos en disolución pueden inmovilizarse sobre polímeros sintéticos tales como poli(fluoruro de vinilideno) (PVDF, Immobilon, Millipore Corp., Bedford, Mass.) o PVDF recubierto con una superficie catiónica (Immobilon CD, Millipore Corp., Bedford, Mass.). Estos soportes pueden usarse con o sin Polybrene. Alternativamente, pueden prepararse muestras de polipéptidos para secuenciarse mediante extracción del polipéptido directamente de poli(acrilamida) mediante un procedimiento denominado electrotransferencia. El procedimiento de electrotransferencia elimina el aislamiento del polipéptido de otros péptidos que pueden estar presentes en disolución. Membranas de electrotransferencia adecuadas incluyen Immobilon e Immobilon CD (Millipore Corp., Bedford, Mass.).

Más recientemente, se han desarrollado métodos automatizados que permiten que se realicen procedimientos químicos sobre polipéptidos inmovilizados sobre soportes sólidos mediante interacción hidrófoba, no covalente. En este enfoque, las muestras en tampones acuosos, que pueden contener sales y agentes desnaturizantes, se cargan a presión en columnas que contienen un soporte sólido. Entonces, el polipéptido unido se enjuaga a presión para eliminar los componentes de interferencia, dejando el polipéptido unido listo para el marcaje (véase, Hewlett-Packard Product Brochure 23-5091-5168E (noviembre de 1992) y Horn, patente estadounidense n.º 5.918.273 (29 de junio de 1999)).

El polipéptido unido se hace reaccionar en condiciones y durante un tiempo suficiente para que produzca el acoplamiento entre los aminoácidos terminales del polipéptido y el resto de marcaje. Las propiedades físicas del soporte pueden seleccionarse para optimizar las condiciones de reacción para un resto de marcaje específico. Preferiblemente, el acoplamiento con los grupos amino del polipéptido se produce en condiciones básicas, por ejemplo, en presencia de una base orgánica tal como trimetilamina, o N-etilmorfolina. En una realización preferida, se permite que el marcador reaccione con el péptido unido en presencia de N-etilmorfolina al 5% en metanol:agua (75:25 v/v). Debido al modo de unión, el exceso de reactivo, la base de acoplamiento y los subproductos de reacción pueden eliminarse mediante disolventes de lavado muy polares antes de la eliminación y secuenciación del polipéptido marcado mediante espectrometría de masas. Diversos reactivos son adecuados como disolventes de lavado, incluyendo, por ejemplo, metanol, agua, mezclas de metanol y agua, o acetona.

Cuando la reacción de marcaje se realiza completamente en fase de disolución, la mezcla de reacción se somete preferiblemente a un ciclo de purificación, tal como diálisis, cromatografía de permeación en gel y similares.

#### **Secuenciación de una parte de una proteína**

En otro aspecto, la presente descripción proporciona un método para secuenciar una parte de una proteína en una mezcla de proteínas, comprendiendo el método:

(a) poner en contacto la mezcla de proteínas con un resto de marcaje en el extremo N-terminal o C-terminal para unir covalentemente un marcador al extremo C- o N-terminal de la proteína y formar una mezcla de proteínas marcadas, comprendiendo el resto de marcaje en el extremo N-terminal o C-terminal al menos un elemento que tiene un número atómico de desde 17 hasta 77, más preferiblemente de desde 35 hasta 63, con la condición de que dicho elemento sea distinto de azufre;

(b) separar las proteínas marcadas individualmente en dicha mezcla de proteínas; y

(c) analizar las proteínas marcadas de la etapa (b) mediante un método de espectrometría de masas para determinar la secuencia de al menos dos residuos en el extremo C-terminal o dos en el extremo N-terminal.

En un grupo de realizaciones, el método comprende además:

(d) identificar la proteína usando la secuencia de al menos dos residuos en el extremo C-terminal o dos residuos en el extremo N-terminal en combinación con una coordenada de separación de la proteína marcada y la ubicación en el extremo terminal de la proteína de la secuencia para buscar secuencias de proteína pronosticadas a partir de una base de datos de datos de secuencias génicas.

Ejemplos preferidos son los que se han descrito anteriormente para otros aspectos de la invención.

#### **Separación**

En una realización preferida, el procedimiento de etiquetado se realiza sobre una mezcla de proteínas. Tras el procedimiento de etiquetado, la mezcla de proteínas se somete a un procedimiento de separación, que preferiblemente permite la separación de la mezcla de proteínas en fracciones diferenciadas. Cada fracción sólo está enriquecida preferiblemente de manera sustancial en sólo una proteína marcada de la mezcla de proteínas.

Los métodos de la presente invención se utilizan con el fin de determinar la secuencia de un polipéptido. Dentro de las realizaciones preferidas de la invención, el polipéptido marcado es "sustancialmente puro," lo que significa que el polipéptido es aproximadamente homogéneo en un 80%, y preferiblemente homogéneo en aproximadamente el 99% o más. Pueden utilizarse muchos métodos bien conocidos por los expertos habituales en la técnica para purificar el polipéptido antes de determinar su secuencia de aminoácidos. Los ejemplos representativos incluyen HPLC, cromatografía de líquidos a alta presión en fase inversa (RP-HPLC), electroforesis en gel, cromatografía, o cualquiera de varios métodos de purificación de péptidos (véase, generalmente la serie de volúmenes titulados METHODS IN PROTEIN SEQUENCE ANALYSIS).

Incluso más preferido es el uso de electroforesis capilar y particularmente, electroforesis capilar multidimensional, tal como la descrita en la solicitud de patente estadounidense en tramitación junto con la presente de titularidad compartida n.º de serie 09/513,486, titulada "Protein Separation via Multidimensional Electrophoresis," y presentada el 25 de febrero de 2000 (patente estadounidense n.º 6.818.112).

Aunque los polipéptidos marcados sustancialmente puros se utilizan preferiblemente dentro de los métodos descritos en el presente documento, también es posible determinar la secuencia de mezclas de polipéptidos que contienen el mismo marcador. En resumen, en una realización, se utiliza un algoritmo con el fin de determinar todas las secuencias hipotéticas con una masa calculada igual a la masa observada de uno de los péptidos en la mezcla. Véase, Johnson *et al.*, Protein Science 1:1083-1091 (1992). Entonces, se asignan índices de mérito a estas secuencias según lo bien que represente cada una de ellas los iones de fragmentos en el espectro de masas en tándem del péptido que utiliza tales algoritmos, pudiendo determinarse la secuencia de polipéptidos dentro de la mezcla fácilmente. Las mezclas de oligómeros analizados simultáneamente se marcan de manera preferente con

marcadores que incorporan diferentes números de elementos de defecto de masa.

Tal como se describió anteriormente, los métodos en el presente documento son particularmente útiles para identificar proteínas de una muestra de tejido sano o enfermo. En un grupo de realizaciones, los métodos se aplican a tanto una mezcla de proteínas de una muestra de tejido sano como a una mezcla de proteínas de una muestra de tejido enfermo. Por consiguiente, las mezclas de proteínas usadas en este aspecto de la invención pueden obtenerse esencialmente de cualquier fuente. Se conocen bien métodos de aislamiento de proteínas de las muestras de tejidos.

Dentro de la presente invención, se secuencia el polipéptido con un aminoácido terminal derivatizado mediante un espectrómetro de masas. Pueden usarse diversos espectrómetros de masas dentro de la presente invención. Los ejemplos representativos incluyen, espectrómetros de masas de triple cuadrupolo, instrumentos de sector magnético (espectrómetro de masas en tándem magnético, JEOL, Peabody, Mass.); espectrómetros de masas de pulverización de iones, Bruins *et al.*, Anal. Chem. 59: 2642-2647 (1987); espectrómetros de masas de electrospray, Fenn *et al.*, Science 246: 64-71 (1989); espectrómetros de masas con detectores de tiempo de vuelo y desorción mediante láser, Karas *et al.*, Anal. Chem. 60: 2299-2301 (1988), y un espectrómetro de masas con resonancia ciclotrónica de iones por transformada de Fourier (Extrel Corp., Pittsburgh, Mass.). Dentro de una realización preferida, se utiliza el espectrómetro de masas de electrospray (modelo Mariner™, PE Biosystems, Foster City, California) para fragmentar el polipéptido terminal derivatizado, y se usa un detector de tiempo de vuelo con precisión de masa mejor de 50 ppm para determinar la secuencia de las masas de los fragmentos marcados.

Un experto en la técnica apreciará que la información de secuencia obtenida usando los métodos de la invención puede combinarse con otras características de la proteína en análisis para reducir incluso adicionalmente el posible número de identidades de la proteína. Por tanto, en una realización preferida, el método de la invención combina información de una etiqueta de secuencia de proteína con una o más otras características de la proteína para identificar la proteína. Los datos que son útiles para complementar los datos de la secuencia incluyen, pero no se limitan a, composición de aminoácidos, el número e identidad de residuos específicos (por ejemplo cisteína), información de escisión, masa del péptido quimiolítica y/o proteolítica (por ejemplo, tríptica), ubicación subcelular y coordenadas de separación (por ejemplo, tiempo de retención, pl, coordenadas de electroforesis 2-D, etc.). Otras formas de datos característicos de una proteína o clase de proteínas particular que puede combinarse con información de las PST de la invención para identificar una proteína serán evidentes para los expertos en la técnica. A medida que el cuerpo de los datos característicos de una proteína particular se hace más completo, las proteínas en análisis pueden identificarse usando etiquetas de secuencia de proteínas más cortas.

Por tanto, en una realización preferida adicional, se usa información con respecto a una o más características de una proteína combinada con información de una PST de aproximadamente 4 aminoácidos de longitud, más preferiblemente de aproximadamente 3 aminoácidos de longitud, más preferiblemente aún, de aproximadamente 2 aminoácidos de longitud, para identificar la proteína.

#### Algoritmo de secuenciación

La presente invención incluirá, en algunas realizaciones, el uso de un algoritmo matemático para determinar la etiqueta de secuencia de proteína directamente de espectros de masas de proteínas marcadas fragmentadas. El algoritmo puede usarse para determinar una etiqueta de secuencia de proteína de cualquier extremo terminal de la proteína, siempre que un esté unido marcador de etiqueta de masa única al extremo terminal que está secuenciándose. Los espectros de masas de partida para su uso en el algoritmo pueden producirse mediante cualquier espectrómetro de masas en el que puede fragmentarse un péptido o proteína marcada. Se prefieren los espectros de masas con detectores de tiempo de vuelo debido a su precisión de masa mejorada con respecto a otros sistemas de detección por espectrómetro de masas. Sin embargo, pueden usarse otros sistemas de detección por espectrómetro de masas menos precisos, particularmente si se usa un patrón de masa interno, tal como un marcador fragmentado sin péptido unido, para mejorar la precisión de masa del espectro de masas resultante. Puede realizarse la fragmentación de proteínas o bien mediante CID en la célula de colisión de un espectrómetro de masas en tándem o bien mediante fragmentación en fuente en una fuente de electrospray o ionización MALDI.

El algoritmo requiere el uso tanto de la posición de la masa con respecto a la carga de una señal como de su abundancia relativa. En una realización, la abundancia relativa de la señal se compara con la de la masa inmediatamente adyacente con respecto a las posiciones de carga y se usa para cuantificar la probabilidad relativa de que un pico esté presente en la posición de la masa con respecto a la carga de interés. En esta realización, se comparan las probabilidades relativas de que un pico esté presente entre todas las secuencias en competencia. En otra realización, se compara directamente la señal en cada posición de masa con respecto a la carga de interés con la de en las posiciones de la masa con respecto a la carga de todas las secuencias en competencia. Este último método se describe adicionalmente para claridad. Un experto en la técnica que apreciará que este método puede adaptarse de muchas maneras para proporcionar un sistema similar para clasificar las secuencias en competencia basándose en la abundancia relativa de la señal en las posiciones de la masa con respecto a la carga correlacionadas con cada secuencia en competencia.

El algoritmo consiste además en un sistema de clasificación de secuencia acumulativa, en la que la abundancia

relativa de los iones que se pronostica que resultan de cada secuencia posible se combina por producto o suma con las abundancias relativas de los iones que se pronostica que resultan de residuos posteriores (ecuación 1). De esta manera, pueden eliminarse las diferencias específicas de secuencia en la eficacia de fragmentación o ionización y la matriz adventicia o los picos de ruido solapantes que confunden la asignación de secuencia correcta en cada posición de residuo en la cadena polipeptídica. La probabilidad de una asignación de secuencia errónea en cualquier posición de residuo dado que se propaga hacia adelante a posiciones de residuo posteriores es menor que la asociada con la secuencia verdadera. El intervalo global para cada secuencia j posible puede determinarse por:

$$R_{j,n} = \prod_{i=1}^n p_{i,j} \quad \text{o} \quad \sum_{i=1}^n p_{i,j} \quad (1)$$

en la que  $R_{j,n}$  es la clasificación acumulativa dada a cualquier secuencia j dada en la longitud de residuo n, y  $p_{i,j}$  es la clasificación relativa asignado a la secuencia entre sus iguales j en la longitud de residuo i. Es evidente para los expertos en la técnica que pueden usarse muchos métodos para asignar una clasificación relativa (p) para cada secuencia j en cualquier longitud de residuo i, consecuente con la comparación de las abundancias relativas de las señales en cada posición de masa con respecto a carga en competencia (citado anteriormente). En una realización preferida, la clasificación relativa (p) de posibilidades de secuencia en competencia en cada longitud de residuo (i) puede determinarse mediante la autoescala de las posibilidades. En una variación particular de este método, la clasificación (p) puede asignarse basándose en una distribución de probabilidad demostrada o supuesta, tal como la distribución de probabilidad (Gaussiana) normal o la distribución de probabilidad (Poisson) logarítmica normal, de manera que la clasificación relativa para cada secuencia variará entre 0 y 1. Por ejemplo,

$$p_i = \text{NORMDIST} \left[ \frac{(C_{i,j} - \bar{C}_i)}{\sigma_i} \right] \quad (2)$$

en la que;

$$\bar{C}_i = \frac{\sum_{j=1}^{19^i} C_{i,j}}{19^i} \quad (3)$$

y

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{19^i} C_{i,j}^2 - \frac{\left(\sum_{j=1}^{19^i} C_{i,j}\right)^2}{19^i}}{(19^i - 1)}} \quad (4)$$

Un experto en la técnica apreciará que la señal ( $C_{i,j}$ ) correspondiente a cualquier secuencia j que contiene residuos de aminoácido i puede determinarse mediante cualquier método que relacione esta señal con la abundancia de señal relativa en el espectro de masas. La fragmentación inducida por colisión en el espectrómetro de masas puede dar como resultado la producción de más de un tipo de ion. Los métodos CID en un espectrómetro de masas en tándem comúnmente dan como resultado tipos de ion a, b y c del extremo N-terminal e iones x, y, y z del extremo C-terminal. Además, el marcador y determinados residuos de aminoácido pueden contener cargas "suaves" que pueden conducir a la producción de fragmentos de péptido marcados en más de una posición de masa con respecto a la carga en el espectro, dependiendo del número de tales cargas "suaves". En una variación del método, las señales asociadas con cada tipo de ion y estado de carga posible pueden combinarse para producir una señal acumulativa asociada con cualquier secuencia j dada:

$$C_{i,j} = \sum_{k=\text{estados de carga min}}^{\text{estados de carga máx}} \sum_{l=1}^{\text{tipos de ion máx}} c_{i,j,k,l} \quad (5)$$

en la que c se determina calculando el (m/z) del cada tipo de ion (k) y el estado de carga (l) y buscando "look up" los recuentos correspondientes (c<sub>i,j,k,l</sub>) en los datos del espectro de masas.

$$c_{i,j,k,l} = \text{LOOKUP}[(m/z)_{i,j,k,l}] \quad (6)$$

5 Los cálculos de razón de masa con respecto a carga para cualquier longitud de residuo i, secuencia j, estado de carga k, y tipo de ion 1, se determinarán a partir de la estequiometría y estados de carga posibles de los aminoácidos y cualquier marcador unido en la secuencia mediante métodos descritos previamente (véase, Methods in Enzymology, Biemann, ed. 193:295-305, 351-360 y 455-479 (1993)).

10 Pueden realizarse varias variaciones al método de secuenciación básico descrito. Por ejemplo, en una realización preferida, el número de estados de carga y tipos de ion que se usan para la determinación de la señal total asociada a cualquier secuencia dada puede restringirse a subconjuntos particulares que se encuentra empíricamente que están asociados lo más a menudo con el método de fragmentación. La fragmentación por CID en un espectrómetro de masas en tándem proporciona preferentemente iones b e iones y en la mayor abundancia e iones c y x en la menor abundancia. Se encuentra que la fragmentación en fuente sólo proporciona iones a, b e y en abundancia significativa. En estos casos, el algoritmo puede adaptarse preferentemente para ignorar iones c y x o iones c, x, y z. 15 La abundancia de iones también parece disminuir para los posibles estados de carga más altos de fragmentos de péptido tanto en CID como en fragmentación en fuente. Este fenómeno también puede ser específico de secuencia, teniendo la arginina y otras especies de carga "suave" imino una mayor probabilidad de retener una carga que otras aminas (por ejemplo, residuos de lisina o histidina). En otra variación, las posiciones de masa con respecto a carga asociadas con números más altos de estados de carga pueden ignorarse en una base específica de secuencia cuando se determina la señal total asociada con cualquier secuencia j. 20

En una variación, pueden incorporarse múltiples marcadores (tanto isotópicos como no isotópicos) en el algoritmo usando un enfoque de secuenciación dual. En este enfoque se definen dos tablas de residuos, uno para cada tipo de marcador (cualquier residuo marcado). Entonces, se aplica el algoritmo de secuenciación usando cada tabla de residuo independientemente, de manera que los recuentos asociados al primer marcador (c<sub>i,j,k,l</sub>) se determinan 25 independientemente de los de un segundo marcador (d<sub>i,j,k,l</sub>).

$$c_{i,j,k,l} = \text{LOOKUP}[(m/z)_{i,j,k,l} | \text{marcador 1}] \quad (7)$$

$$d_{i,j,k,l} = \text{LOOKUP}[(m/z)_{i,j,k,l} | \text{marcador 2}] \quad (8)$$

Todas las ecuaciones 1-6 se aplican tanto a c como a d, y se puede definir:

$$q_i = \text{NORMDIST} \left[ \frac{(D_{i,j} - \bar{D}_i)}{\sigma_i} \right] \quad (9)$$

$$\bar{D}_i = \frac{\sum_{j=1}^{19^i} D_{i,j}}{19^i} \quad (10)$$

$$\sigma_i^{\text{marcador2}} = \sqrt{\frac{\sum_{j=1}^{19^i} D_{i,j}^2 - \frac{\left(\sum_{j=1}^{19^i} D_{i,j}\right)^2}{19^i}}{(19^i - 1)}} \quad (11)$$

$$D_{i,j} = \sum_{\substack{k=\text{estados} \\ \text{de carga} \\ \text{mín}}}^{\text{estados} \\ \text{de carga} \\ \text{máx}} \sum_{l=1}^{\text{tipos de} \\ \text{ion} \\ \text{máx}} d_{i,j,k,l} \quad (12)$$

Multiplicando la probabilidad relativa de cada secuencia  $j$  obtenida con cada marcador, puede obtenerse entonces una clasificación compuesta para la secuencia.

$$R_{j,n} = \prod_{i=1}^n (p_i q_i) \quad \text{or} \quad \sum_{i=1}^n (p_i q_i) \quad (9)$$

5 Esta variación puede ampliarse además a más de un marcador. Los archivos del espectrómetro de masas usados en este enfoque de marcaje múltiple pueden crearse mediante fragmentación simultánea de una muestra de proteína que contiene una mezcla conocida de dos o más marcadores. Además, los datos del espectrómetro de masas de fragmentaciones de proteína de único marcador separadas pueden añadirse juntos para crear un archivo de espectrómetro de masas de múltiples marcadores virtual para el análisis mediante este método. Esta variación puede usarse con cualquier tipo de estrategia de marcaje múltiple (citado anteriormente).

10 En otra realización preferida para marcadores isotópicos, o bien abundancias isotópicas naturales o bien con múltiples marcadores de abundancias isotópicas relativas conocidas, el algoritmo puede adaptarse para calificar o clasificar los picos de secuencias en competencia mediante su conformidad con las abundancias esperadas de la serie isotópica. Por ejemplo, cuando se emplean dos marcadores isotópicamente distintos de una abundancia relativa conocida,  $\beta$ , puede determinarse la razón de masa con respecto a carga de cada secuencia para ambos isótopos marcados, determinándose los valores de recuento correspondientes a partir de los datos del espectro de masas, y determinándose un intervalo o probabilidad de coincidir con la abundancia esperada ( $\beta$ ).

#### **Oligómeros marcados**

20 Aunque la invención se ha descrito anteriormente con referencia a proteínas marcadas, un experto en la técnica reconocerá que los marcadores y métodos de marcaje usados se pueden adaptar a la preparación de otros oligómeros marcados (por ejemplo, oligonucleótidos marcados, oligosacáridos marcados y similares).

25 Los ácidos nucleicos pueden secuenciarse según la invención a través de la modificación de los métodos descritos por Butler *et al.*, patente estadounidense n.º 6.090.558; Moforte *et al.*, patente estadounidense n.º 5.700.642; y Koster, patentes estadounidenses n.ºs 6.194.144 y 5.691.141. Preferiblemente, los métodos en el presente documento usan la secuenciación de Sanger o cebadores de la reacción en cadena de la polimerasa con marcadores de “defecto de masa” o bien en la secuencia del cebador, en la secuencia de elongación, o bien en la secuencia de terminación. Los marcadores de “defecto de masa” son aquellos marcadores que incorporan uno o más elementos que tienen un número atómico de desde 17 hasta 77, más preferiblemente de desde 35 hasta 63, con la condición de que dichos elementos sean distintos de azufre o fósforo. Preferiblemente, se marca o bien el cebador o bien el extremo terminal para controlar el número de elementos con “defecto de masa” incorporados en los oligómeros de ácido nucleico así producidos. Alternativamente, los extremos 3' o 5' de un ácido nucleico pueden marcarse con un marcador de “defecto de masa” y el ácido nucleico fragmentado por el método descrito por Maxam y Gilbert, Proc. Natl. Acad. Sci. (EE.UU.) 74:560-564 (1977) para proporcionar una serie de fragmentos marcados.

35 Los oligosacáridos pueden secuenciarse según la invención a través de la modificación de los métodos descritos por Parekh *et al.*, patente estadounidense n.º 5.667.984 y Rademacher *et al.*, patente estadounidense n.º 5.100.778. Preferiblemente, se usan marcadores de “defecto de masa” reactivos con el extremo de azúcar reductor. Más preferiblemente, tales marcadores incluirían funcionalidades de piridilamino en los que el anillo de piridinilo está sustituido con desde uno hasta tres Br y/o I. El marcaje puede realizarse antes de la hidrólisis enzimática selectiva del oligosacárido o la hidrólisis posenzimática. Preferiblemente, un oligosacárido se marca con un marcador de “defecto de masa” antes de la hidrólisis enzimática y un marcador diferente de “defecto de masa” tras la hidrólisis enzimática para diferenciar el azúcar reductor terminal original.

#### **Etiquetas de masa para análisis simultáneo**

45 Aunque la invención se ha descrito anteriormente con referencia a oligómeros, un experto en la técnica reconocerá que los marcadores y métodos de marcaje usados pueden adaptarse a la preparación de “etiquetas” de masa única para otras muestras (por ejemplo, etiquetas para bibliotecas químicas combinatorias, etiquetas para metabolitos obtenidos de diferentes muestras, y similares). Tales métodos permiten el análisis y comparación simultáneos de varias muestras mediante espectrometría de masas.

**Métodos para análisis de función y estructura**

En aún otro aspecto, la presente invención proporciona un método para el análisis de función y estructura de un oligómero que tiene una pluralidad de residuos según la reivindicación 23.

5 Al igual que con los métodos anteriores, en relación con la secuenciación, este aspecto para la invención puede aplicarse esencialmente a cualquier oligómero (por ejemplo, proteínas, ácidos nucleicos, oligosacáridos) para los que se busca algún indicio de función y/o estructura tridimensional de residuos expuestos. El método encuentra su uso particular en el análisis de proteínas.

10 Además, como anteriormente, el reactivo de marcaje de defecto de masa contiene normalmente al menos un elemento que tiene un número atómico de desde 17 hasta 77, pero exclusivo de azufre o fósforo. Para aquellas aplicaciones en las que se desea la información de la estructura de la proteína, el reactivo de marcaje de defecto de masa puede contener un átomo de fósforo puesto que este elemento no está presente en las proteínas (con la excepción, por supuesto, de las proteínas fosforiladas). Preferiblemente, el reactivo de marcaje de defecto de masa comprende un elemento que tiene un número atómico de desde 35 hasta 63, más preferiblemente de 39 a 58, exclusivo de fósforo o azufre.

15 En un grupo de realizaciones particularmente preferidas, el oligómero es una proteína y el reactivo de marcaje comprende un átomo o bien de bromo o bien de yodo. Más preferiblemente, el reactivo de marcaje es gas de bromo que tiene puede marcar, por ejemplo, residuos de tirosina expuestos en una proteína plegada.

En otras realizaciones preferidas, el método de espectrometría de masas empleado usa EM ESI-TOF.

20 Otras realizaciones y condiciones preferidas para llevar a cabo este aspecto de la invención son, tal como se describió anteriormente, para el marcaje y secuenciación de oligómeros en general, y proteínas, en particular.

**Métodos para comparar las abundancias relativas de analitos**

En aún otro aspecto, la presente invención proporciona un método para comparar las abundancias relativas de analitos de dos o más muestras según la reivindicación 45.

25 Los restos de marcaje preferidos son los descritos a lo largo de la memoria descriptiva, más preferiblemente los que tienen elementos con números atómicos de desde 35 hasta 63, todavía más preferiblemente desde 39 hasta 58. Las muestras pueden ser esencialmente cualquier muestra biológica de tejido animal, plantas o extractos vegetales, muestras bacterianas, muestras virales y similares.

**Métodos para el marcaje de componentes de una biblioteca química**

30 Otra aplicación de los marcadores de defecto de masa es como etiquetas para bibliotecas químicas combinatorias. Un uso a modo de ejemplo sería la creación de una etiqueta de defecto de masa simultáneamente con la síntesis de una biblioteca combinatoria. Una síntesis típica implica la construcción de la biblioteca combinatoria unida a una superficie de pocillo o perla añadiendo en secuencia diferentes combinaciones o productos químicos a los pocillos. Tal como describieron Brenner y Lerner (Proc. Natl. Acad. Sci (EE.UU), 89:5381-5383 (1992)) y Sugarman *et al.* (patente estadounidense n.º 6.056.926, concedida el 2 de mayo de 2000), también es posible sintetizar  
35 simultáneamente una etiqueta de masa específica para las etapas de tratamiento o composición química del producto químico específico durante la síntesis de la biblioteca química. Etiquetas de este tipo pueden crearse mediante la adición en serie de diferentes grupos reactivos sobre una etiqueta de raíz. Una etiqueta de raíz de este tipo podría ser una amina primaria. En este caso, podría añadirse un grupo reactivo a la raíz para crear una amina secundaria, en la que la masa del grupo reactivo podría cambiarse para representar la identidad química específica del producto químico o la etapa de procesamiento a la que se expuso la entidad química de la biblioteca. En una  
40 etapa posterior, podría añadirse un segundo grupo reactivo a la amina secundaria en la etiqueta de raíz para crear una amina terciaria. Esta masa de este segundo grupo representaría una segunda etapa de procesamiento químico o el cambio en identidad del producto químico específico en la biblioteca combinatoria. Un enfoque similar, que permite más combinaciones de etiqueta de raíz, sería la adición en serie de diversas unidades de monómeros a una  
45 etiqueta de raíz oligomérica (por ejemplo, síntesis peptídico de Merrifield).

Sin embargo, este enfoque está limitado por el número de posibles combinaciones de etiqueta de manera que las masas aditivas de adiciones de grupo reactivo posteriores no se solapen cuando se analiza la etiqueta final mediante espectrometría de masas. La presente descripción evita este problema añadiendo elementos de defecto de masa adicionales a la etiqueta de raíz creciente en cada etapa. Un problema adicional resuelto por esta memoria  
50 descriptiva es la capacidad para distinguir la etiqueta creada para identificar el producto químico específico en una biblioteca combinatoria de la propia masa del producto químico o de la masa de cualquier otro componente que puede estar presente durante un ensayo o selección de actividad del producto químico específico de la biblioteca. El uso de los defectos de masa en la etiqueta permite que la etiqueta se identifique de manera única en un espectro de masas de o bien el producto químico o bien la biblioteca de productos químicos y cualquier contaminante de matriz  
55 que se origina a partir del ensayo de selección.

En vista de lo anterior, la presente descripción proporciona métodos para usar marcadores de defecto de masa para indicar la secuencia sintética a la que un miembro de la biblioteca química se ha expuesto durante el curso de la síntesis.

5 En particular, la presente descripción proporciona un método para etiquetar los elementos de bibliotecas químicas, durante o bien la síntesis o bien la selección, que comprende;

(a) poner en contacto una etiqueta de raíz con un resto de marcaje para unir covalentemente un marcador a la etiqueta de raíz y formar una etiqueta marcada, teniendo dicho resto de marcaje un elemento con un número atómico de desde 17 hasta 77, con la condición de que dicho elemento sea distinto de azufre o fósforo;

10 (b) opcionalmente, poner en contacto una raíz con restos de marcaje adicionales para unir covalentemente marcadores adicionales a la etiqueta de raíz y formar una etiqueta marcada de manera múltiple, teniendo dicho resto de marcaje un elemento con un número atómico de desde 17 hasta 77, con la condición de que dicho elemento sea distinto de azufre o fósforo; y

15 (c) analizar la etiqueta marcada mediante métodos de espectrometría de masas para determinar tanto su masa como el número de elementos con un número atómico de desde 17 hasta 77, de manera que la masa y el número de elementos identifique los procesos químicos a los que el producto químico específico de la biblioteca se ha expuesto y la identidad del producto químico de la biblioteca.

Se describen elementos preferidos (por ejemplo, elementos de defecto de masa) para otros aspectos de la invención anteriores.

20 Los métodos de la presente invención se ilustran adicionalmente mediante los ejemplos siguientes. Estos ejemplos se facilitan para ilustrar, pero no limitar la invención reivindicada.

### Ejemplos

#### Ejemplo 1

25 En este ejemplo, se marca y se secuencia un oligosacárido de tipo con alto contenido en manosa (figura 7). El oligosacárido se marca usando métodos similares a los descritos en Parekh, *et al.*, patente estadounidense n.º 5.667.984. En resumen, un marcador de defecto de masa (2-amino-6-yodopiridina (marcador 1)) se une covalentemente al extremo terminal reductor del oligosacárido en presencia de cianoborohidruro de sodio ( $\text{NaBH}_3\text{CN}$ ). Esto incorpora un único elemento de defecto de masa (yodo) en el oligosacárido original. La adición del elemento de defecto de masa permite que los fragmentos de oligosacárido marcados se distingan de fragmentos no marcados e iones de matriz en el espectro de masas.

30 Entonces, el oligosacárido conjugado con el marcador 1 se reparte en alícuotas a tubos de reacción que contienen diferentes sacarosas (véanse las tablas 1.1 y 1.2) en tampones de reacción apropiados. Se deja que avancen las reacciones hasta su finalización y los productos de reacción resultantes se conjugan en los extremos reductores recién formados de los fragmentos mediante la reacción con marcadores de defecto de masa mostrados para cada enzima (véase la tabla 1.2), de nuevo en presencia de cianoborohidruro de sodio. Cada uno de los marcadores 2 y 3  
35 contiene diferentes números de elementos de defecto de masa, dejando que los fragmentos de digestión se distingan del fragmento terminal del oligosacárido original.

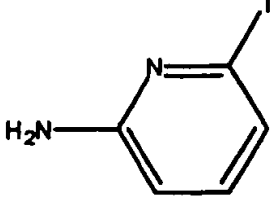
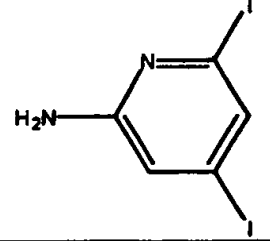
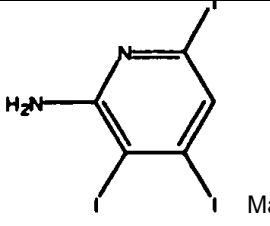
Tabla 1.1.

#### Enzimas oligosacaridasas

Enzima n.º	Especie	Enzima
1	<i>Aspergillus saitoi</i>	$\alpha$ -manosidasa I
2	<i>Jack bean</i>	$\alpha$ -manosidasa
3	<i>Achatina saitoi</i>	$\alpha$ -manosidasa II
4	<i>Jack bean</i>	$\beta$ -hexosaminidasa
5	<i>Prevotella sp.</i>	$\beta$ -hexosaminidasa
6	<i>Achatina fulica</i>	$\beta$ -manosidasa
7	<i>Streptococcus pneumoniae</i>	N-acetil $\beta$ -hexosaminidasa
8	<i>Helix pomatia</i>	$\beta$ -manosidasa



**Tabla 1.2.**  
**Combinaciones de reacción y marcador**

Enzima*	Acción	Marcador de defecto de masa usado
Ninguna	Ninguna	 <p>Marcador 1</p>
1	Escinde manosas unidas en 1 $\alpha$ 2 en cualquier sitio	 <p>Marcador 2</p>
3	Escinde manosas unidas en 1 $\alpha$ 3,6 en cualquier sitio  Escinde manosas unidas en 1 $\alpha$ 3 cuando se unen a un azúcar ramificado	 <p>Marcador 3</p>

\*El número de enzima corresponde a la descripción en la tabla 1.1

5 Una alícuota de la mezcla de reacción conjugada con el marcador 3 (es decir, digerida con la enzima n.º 3) se digiere además con la enzima 1. Los extremos terminales reductores del azúcar generados mediante esta reacción se conjugan posteriormente con el marcador 2 tal como se describió anteriormente.

10 Entonces, se mezclan las alícuotas de todas estas reacciones, se acidifican mediante la adición de una mezcla al 50% v/v de ácido acético al 2% en metanol y se someten a análisis del espectro de masas. Debido a la baja estabilidad del conjugado de acetal en disoluciones ácidas, se realiza el análisis del espectro de masas inmediatamente tras la acidificación. Alternativamente, una serie de marcadores diferentes que incorporan una carga dura (por ejemplo, una serie de N-alkil-yodo-piridinio) puede someterse a análisis del espectro de masas sin acidificación. El espectro de masas resultante se deconvoluciona para eliminar todo ruido químico que no contenga un pico marcado de defecto de masa mediante los métodos de esta invención. El espectro de defecto de masa deconvolucionado resultante se busca entonces mediante algoritmo mediante los métodos de esta invención prediciendo todas las posibles secuencias de oligosacáridos que podrían unirse a cada marcador de defecto de masa usado.

20 El algoritmo de búsqueda calcula la masa para cada combinación de ramificación de hexosa (Hex) y N-acetilaminohexosa (HexNAC). Cada unidad monomérica de Hex añade una unidad de masa monoisotópica de 179,055565 amu al peso de la masa del fragmento estimado. Cada unidad monomérica de HexNAC añade una masa monoisotópica de 220,082114 amu a la masa del fragmento estimado. Hay una pérdida neta de (n-1) veces 17,00274 amu para cada azúcar (n) contenido en el fragmento. La composición de oligosacáridos de los picos correspondientes a los criterios de búsqueda para los marcadores 1, 2, y 3 se muestran en las figuras 8A, 8B y 8C, respectivamente. El número de hexosas y N-acetilaminohexosas correspondientes a estos picos se muestran en la

25

**Tabla 1.3**

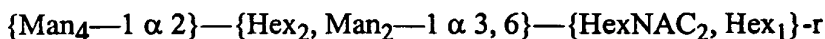
**Número y tipo de hexosas correspondientes a los picos de la figura 1 (A, B y C)**

Pico	Composición	
	HexNAC	Hex

A	2	1
B	2	5
C	2	9
D		1
E		1
F		2
G		3

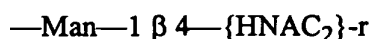
5 El marcador de masa molecular formado a partir de los fragmentos conjugados con el marcador 1 indica que los azúcares más exteriores son hexosas. El fragmento de masa superior conjugado con el marcador 1 corresponde al oligosacárido original. Como resultado, la diferencia de masas de cuatro hexosas entre el primer fragmento conjugado con el marcador 1 y el compuesto original indica la presencia de cuatro  $\alpha$ -manosas puesto que tanto la enzima 1 como la enzima 3 sólo escinden  $\alpha$ -manosas. Puesto que el pico D es la única coincidencia del conjugado con el marcador 2 en la figura 8B, cuatro de los azúcares más exteriores desde el extremo terminal reductor deben ser manos unidas en  $1 \alpha 2$  y no pueden haber manos unidas en  $1 \alpha 2$  internas.

10 El siguiente fragmento en el marcador de masa molecular del marcador 1 (pico A) difiere en 4 hexosas adicionales del fragmento anterior. Esto debe corresponder a una muestra digerida con la enzima 3. Los únicos fragmentos conjugados con el marcador 3 correspondientes (figura 8C) son E (un fragmento de 1 hexosa), F (un fragmento de 2 hexosas) y G (un fragmento de 3 hexosas). Puesto que los picos F y G hacen un total de 5 hexosas, al menos uno de estos fragmentos debe contener una manosa unida en  $1 \alpha 2$ . Puesto que la enzima 3 sólo escinde enlaces  $1 \alpha 3$  y  $1 \alpha 6$ , por tanto, debe haber al menos dos manos unidas en  $1 \alpha 3$  y/o  $1 \alpha 6$  separadas en la estructura y estas manos deben ser interiores a las 4 manos unidas en  $1 \alpha 2$ . De esta información puede deducirse la siguiente secuencia parcial:



en la que r indica el extremo reductor del oligosacárido.

20 Este proceso se repite con diferentes enzimas de la tabla 1.1 hasta que se determina la secuencia completa. Por ejemplo, la digestión con la enzima 3 seguida por la enzima 8 permite la determinación de que la secuencia inicial es:



La secuencia completa del extremo reductor del oligosacárido se determina mediante la reacción con la enzima 3 seguida por la enzima 7.

## 25 Ejemplo 2

En este ejemplo, se usa un marcador de defecto de masa para la identificación de la composición de ácidos grasos y su disposición en lípidos, o "secuenciación de lípidos." Este ejemplo utiliza fosfatidilcolina; sin embargo, un experto en la técnica apreciará que estos métodos en combinación con métodos de separación alternativos, selecciones por manchas y con lipasa pueden aplicarse a cualquiera de los lípidos saponificables tal como define Lehninger (véase, BIOCHEMISTRY (Worth, NY, 1975)).

30 Se prepara extracto de lípidos mediante extracción con éter de un sedimento de células *E. coli* K-12 usando el método de Hanson y Phillips (véase, MANUAL OF METHODS FOR GENERAL BACTERIOLOGY, pág. 328, Amer. Soc. Microbiol., Washington, DC, 1981). Se eliminó el éter del extracto mediante evaporación y se resuspendió el sedimento de lípidos en un sistema de disolvente metanol:cloroformo:ácido fórmico 65:25:5 (que contiene hidroxitolueno butilado al 0,1% para inhibir la oxidación). La mitad del volumen se sometió a ensayo por manchas en cada uno de los dos carriles de una placa HL de sílice delimitada (Altech, Deerfield, IL) y se dejó secar. Se separaron los lípidos usando un sistema de disolvente descrito por Waters y Huestis, AMPHIPATHIC INTERACTIONS WITH ERYTHROCYTES AND PLATELETS, DOCTORAL DISSERTATION (Stanford University, Stanford, CA, Departamento de Química, 1992). Este procedimiento separa los lípidos por grupos de cabeza. Se retiró un carril y se expuso al vapor de yodo para determinar las posiciones relativas de cada una de las fracciones de lípidos. Se desechó la matriz de sílice de la región en el carril no desarrollado correspondiente a la mancha de fosfatidilcolina y se colocó en un tubo de microcentrífuga.

Se resuspendió el sedimento de sílice en 100  $\mu$ l de tampón de reacción de fosfolipasa (100  $\mu$ l) tal como describe

Cottrell, METH. ENZYMOLOGY, 71:698 (1981) y se mezcló con vórtex vigorosamente. Se retiró una alícuota (50  $\mu$ l) de la suspensión de sílice a un segundo tubo de microcentrífuga. Se trató la primera alícuota mediante la adición de 1 U.I. de fosfolipasa A2 de *Apis mellifera* (Sigma-Aldrich, St. Louis, MO), que hidroliza selectivamente los ácidos grasos C2. Se trató la segunda alícuota mediante la adición de 1 U.I. de Novozyme 871 (Sigma-Aldrich, St. Louis, MO), que hidroliza selectivamente los ácidos grasos C3 de los fosfoglicéridos. Se incubaron ambas mezclas de reacción a temperatura ambiente durante la noche.

Se evaporaron las mezclas de reacción hasta sequedad a vacío, y se resuspendieron en aproximadamente 25  $\mu$ l de diclorometano. Se añadió el marcador 1 de defecto de masa (2-amino-5-yodo-piridina) (20  $\mu$ l de una disolución 1 M en diclorometano) a la mezcla de reacción de fosforilasa A2. Se añadió el marcador 2 de defecto de masa (2-amino-3,5-diyodo-piridina) (20  $\mu$ l de una disolución 1 M en diclorometano) a la mezcla de reacción de Novozyme 871. Entonces, se añadió una alícuota (20  $\mu$ l de una disolución 1 M de 1,3-diciclohexilcarbodiimida) a ambos tubos y se incubaron durante 2 horas. La carbodiimida catalizó la conjugación de los ácidos grasos liberados por la enzima con los marcadores de defecto de masa. Se acidificaron las mezclas de reacción mediante la adición de ácido fórmico al 1% (v/v) y se mezclan inmediatamente antes del análisis de espectrometría de masas mediante micropulverización en un EM ABI Mariner.

Se deconvolucionó el ruido químico del espectro de masas resultante mediante los algoritmos de la presente invención, dejando los espectros de masas deconvolucionados mostrados en la figura 9. Se determinaron las identidades y abundancias relativas de los diversos ácidos grasos en C2 y C3 en la estructura principal lipídica de fosfatidilcolina mediante adición de masa a cada marcador. Las longitudes de las colas de ácido graso natural se producen en múltiples unidades de o bien  $-\text{CH}_2\text{CH}_2-$  (28,031300 amu) o bien  $-\text{CH}=\text{CH}-$  (26,015650). La masa de un H (1,007825 amu) se añade a cada longitud de cadena pronosticada para completar la estequiometría del grupo metilo terminal. Los ácidos grasos ramificados no pueden distinguirse de análogos de cadena sencilla porque la pérdida de un hidrógeno de la masa en un punto de ramificación se recupera por el H extra necesario para completar la estequiometría en el extremo terminal de la nueva ramificación.

La abundancia relativa de los diversos ácidos grasos en la posición C2 puede estimarse a partir de las alturas de pico monoisotópico para los diversos picos conjugados con el marcador 1 ( $A_1 \rightarrow F_1$ , figura 9). La abundancia relativa de los diversos ácidos grasos en la posición C3 de fosfatidilcolina puede estimarse a partir de las alturas de pico monoisotópico para los diversos picos conjugados con el marcador 2 ( $A_2 \rightarrow F_2$ , figura 9). Por tanto, en la tabla 2.1 se muestra la secuencia promedio de la fosfatidilcolina de *E. coli*.

Puede obtenerse la resolución de la secuencia lipídica adicional a través del uso de una segunda dimensión de cromatografía en capa fina u otro método de separación en el que se use la hidrofobicidad de los ácidos grasos para resolver los lípidos (véase, por ejemplo, Morris, J. Lipid Res., 7:717-732 (1966)).

Tabla 2.1

Composición de la secuencia de fosfatidilcolina de *E. coli*

Pico	Ácido graso	Abundancia aproximada (%)	
		C3 (marcador 2)	C2 (marcador 1)
A	n-dodecanoico	20	10
B	n-tetradecanoico	20	30
C	palmitoléico	-	2
D	n-hexadecanoico	37	35
E	oléico	-	2
F	n-octadecanoico	22	20

### Ejemplo 3

Este ejemplo describe la preparación de marcadores de defecto de masa fotoescindibles que tienen sustituyentes de bromo o yodo. Estos marcadores son útiles para cuantificar las abundancias relativas de biomoléculas (por ejemplo, ácidos nucleicos, proteínas, o metabolitos) que pueden mostrar de otro modo eficacias de ionización o detección bajas en el espectrómetro de masas. El marcador de defecto de masa sirve como un marcador sustituto para su biomolécula conjugada en el espectrómetro de masas. Las variaciones de la química del extremo terminal proporcionan medios para la unión a biomoléculas que contienen amina primaria, sulfhidrilo y ácido carboxílico. La inclusión del elemento de defecto de masa en el marcador permite al marcador resolverse inequívocamente del

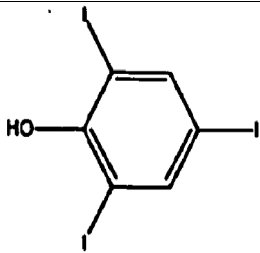
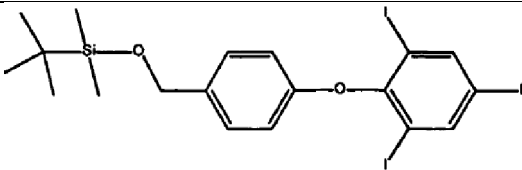
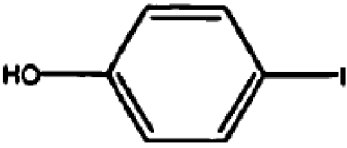
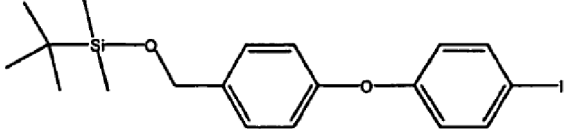
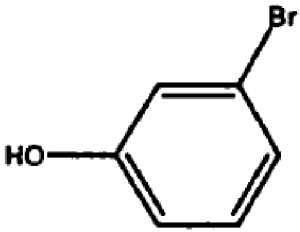
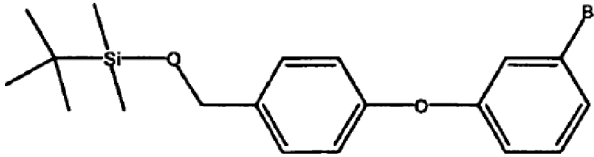
ruido químico solapante que puede estar presente en la muestra y dos muestras entre sí cuando se incorporan diferentes números de elementos de defecto de masa en dos marcadores (véase también el ejemplo 1).

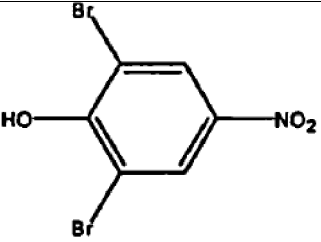
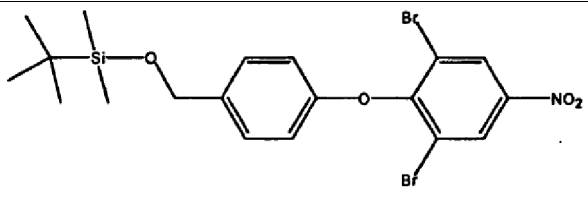
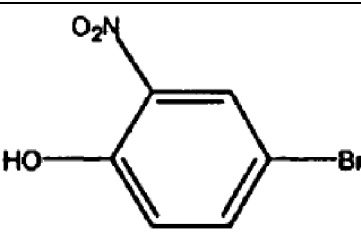
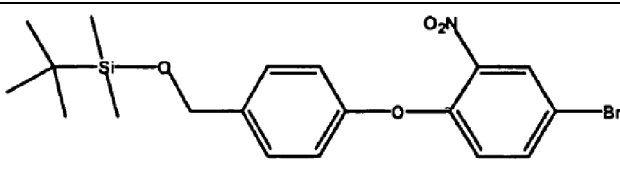
5 En resumen, se mezcla éter de 4-(*tert*-butildimetilsilil)-fenilborato (FT106), preparado tal como describe Schmidt *et al.*, documento WO 99/32501 (1 de julio de 1999) con uno de los bromo- o yodo-fenoles disponibles comercialmente correspondientes mostrados en la tabla 3.1 para formar los precursores de marcador de defecto de masa bromado o yodado correspondientes usando métodos tal como se describe en Schmidt *et al.*, documento WO 99/32501 (1 de julio de 1999). Los enlaces aril éter adicionales pueden insertarse entre FT106 y el grupo arilo que contiene el defecto de masa terminal a través de la incorporación de una hidroquinona o 4,4'-dihidroxidifenil éter disponibles comercialmente. Para llevar a cabo esto, el fenol terminal se reactiva usando un extremo terminal de ácido fenilborónico mediante el mismo método usado para crear FT106. De manera similar, pueden crearse aril éteres ramificados mediante la adición y reactivación del 1,2,4-benzenotriol disponible comercialmente.

15 El grupo protector *tert*-butil-dimetilsilano del precursor de marcador de defecto de masa (de MDP1 a MDP5, tabla 3.1) se elimina con un exceso molar de fluoruro de trimetilsulfonio en cloruro de metileno u otro medio adecuado conocido generalmente en la técnica. El fenol desprotegido correspondiente se acopla además a un ligador de amino bloqueado apropiadamente (véase, por ejemplo el documento GB 98/15163.2 (13 de julio de 1998)) que se convierte posteriormente en la amina primaria tal como describe Schmidt *et al. ibid.* La amina puede hacerse reaccionar adicionalmente con cualquier fenil-vinilsulfona apropiada. Los ejemplos de fenil-vinilsulfonas apropiadas incluyen, pero sin limitarse a, las que tienen una sustitución de amina primaria bloqueada (o un grupo nitro que puede reducirse posteriormente a una anilina), ácido carboxílico (por ejemplo, éster de trifluoroacetato), o tiol (por ejemplo, un enlace disulfuro) en el anillo de fenilo. Entonces, se hace reaccionar el 2º grupo amino del ligador con anhídrido trifluoroacético o cloruro de metanosulfonio para dar el marcador fotoescindible. Finalmente, se elimina el agente bloqueante mediante métodos generalmente reconocidos en la técnica y se conjuga el marcador de masa fotoescindible con la molécula o macromolécula a través del grupo amina, ácido carboxílico, o tiol libres mediante cualquier método de conjugación, generalmente reconocido, adecuado para proporcionar una molécula conjugada con el marcador de defecto de masa fotoescindible.

Tabla 3.1

Bromo- y yodo-fenoles disponibles comercialmente

Fenol sustituido	Código	Precursor de marcador de defecto de masa
 <p>2,4,6-Triyodo-fenol</p>	MDP1	 <p><i>tert</i>-Butil-dimetil-[4-(2,4,6-triyodo-fenoxi)-benciloxi]-silano</p>
 <p>4-Yodo-fenol</p>	MDP2	 <p><i>tert</i>-Butil-[4-(4-yodo-fenoxi)-benciloxi]-dimetil-silano</p>
 <p>3-Bromo-fenol</p>	MDP3	 <p>[4-(3-Bromo-fenoxi)-benciloxi]-<i>tert</i>-butil-dimetil-silano</p>

 <p>2,6-Dibromo-4-nitrofenol</p>	MDP4	 <p>[4-nitro-(2,6-dibromo-fenoxi)-benciloxi]-terc-butil-dimetilsilano</p>
 <p>4-Bromo-2-nitrofenol</p>	MDP5	 <p>[2-nitro-(4-bromo-fenoxi)-benciloxi]-terc-butil-dimetilsilano</p>

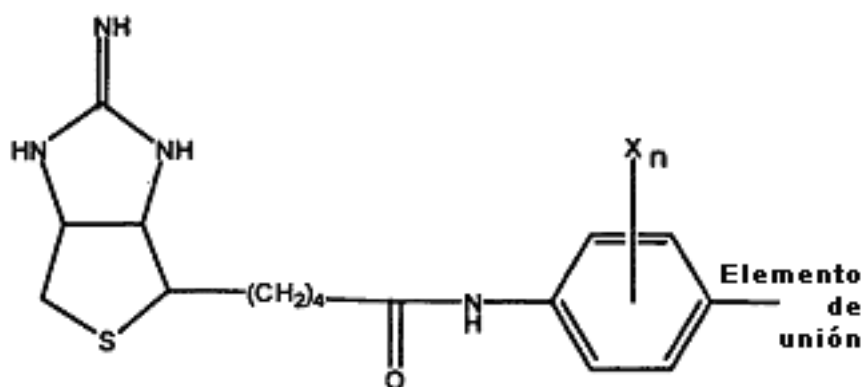
#### Ejemplo 4

5 Este ejemplo ilustra el uso de marcadores de masa acoplados por afinidad para el análisis rápido y cuantitativo de compuestos con marcador de defecto de masa purificados por afinidad obtenidos a partir de diferentes muestras. En este ejemplo, se usan proteínas, pero un experto en la técnica apreciará que este método puede aplicarse al análisis para la comparación de cualquier molécula copurificada a partir de diferentes muestras.

10 La preparación del marcador comienza con cualquier bromuro o yoduro de arilo heterobifuncional adecuado (tal como los ejemplos disponibles comercialmente mostrados en la tabla 4.1). MDP4 y MDP5 (tabla 3.1) también son útiles. Cada una de estas anilinas puede hacerse reaccionar con un exceso estequiométrico de un éster de N-hidroxisuccinimida (NHS) de un reactivo de afinidad, tales como las moléculas de biotina o NHS-iminobiotina disponibles comercialmente en acetonitrilo anhidro. Se incuba la mezcla de reacción durante al menos 2 h antes de la adición de agua para hidrolizar cualquier éster de NHS sin reaccionar. Se evapora el disolvente hasta sequedad.

15 Entonces el grupo nitro se reduce a una amina primaria usando métodos convencionales, tales como HCl diluido con  $\text{SnCl}_2$  añadido como catalizador. El producto de reacción (véase el compuesto 4.1 a continuación) se purifica mediante cromatografía de afinidad y se evapora hasta sequedad. Entonces, el grupo amino (producido mediante la reducción del grupo nitro) se hace reaccionar con otro agente de reticulación adecuado (por ejemplo, anhídrido yodoacético) o puede usarse directamente para la unión a las moléculas diana que contienen ácido carboxílico usando química de carbodiimida. También pueden usarse químicas de unión alternativas adecuadas para la reacción con aminas primarias.

20 Opcionalmente, el grupo amino (o segundo extremo terminal de anilina) puede extenderse mediante la reacción con polietilenglicoles hidrogenados y perdeuterados, tal como describe Aebersold *et al.* (documento WO 00/11208 (2 de marzo de 2000)) para producir una serie de etiquetas de defecto de masa distintas isotópicamente para marcaje diferencial. De manera similar, pueden usarse materiales de partida de bromuro o yoduro de arilo isotópicamente puros para generar etiquetas de afinidad acopladas a isótopo directamente.



4.1

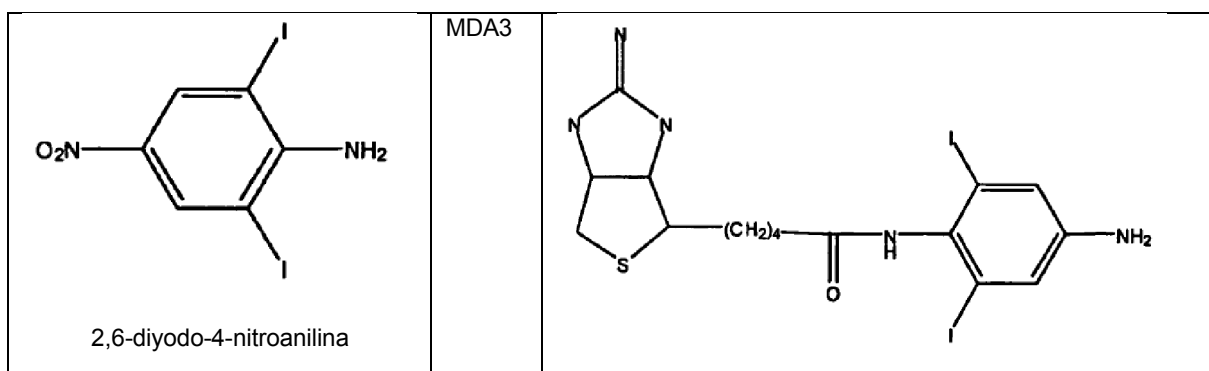
5

El compuesto 4,1 ilustra una etiqueta de afinidad de iminobiotina marcada con defecto de masa en el que X representa un elemento de defecto de masa (por ejemplo, bromo o yodo) y n representa el número de elementos de defecto de masa. El ligador es cualquier elemento químico de unión que pueda usarse para unir la etiqueta acoplada por afinidad de defecto de masa a una molécula diana. Los ejemplos incluyen anilina (que puede unirse a ácidos carboxílicos a través de química de carbodiimida), yodoacetamida (formada por la reacción de anilina con anhídrido yodoacético), o azida formada por la reacción con ácido nítrico.

Tabla 4.1

## Ejemplos de marcadores de defecto de masa acoplados por afinidad

Bromuro o yoduro de arilo heterobifuncional	Código	Marcador de defecto de masa acoplado por afinidad
<p>2-bromo-4-nitroanilina</p>	MDA1	
<p>2,6-dibromo-4-nitroanilina</p>	MDA2	



Para ilustrar el uso de tales etiquetas, se obtienen muestras de plasma sanguíneo (1 ml) de cada uno de dos pacientes y se colocan en tubos de microcentrifuga separados. Cada tubo se trata tal como sigue. Se precipitan las macromoléculas mediante la adición de ácido tricloroacético hasta una concentración final de aproximadamente el 10% p/v y se incuban los tubos en hielo durante 20 min. Se sedimenta el precipitado mediante centrifugación (14.000 g) y se retira el sobrenadante. Se seca el sedimento a vacío. Se resuspende el sedimento seco en 100 microlitros de un tampón de digestión triptica adecuado que contiene 100 U.I. de tripsina y clorhidrato de tris(2-carboxietil)fosfina al 0,1 % p/v . Se incuba la disolución durante la noche a 37°C.

Se preparan alícuotas isotópicamente puras de MDA1 (marcador 1 por afinidad de defecto de masa) con un ligador de yodoacetamida. Se añade una alícuota (50 microlitros) de la digestión triptica de la muestra 1 a un tubo de microcentrifuga que contiene 10 mg de [<sup>79</sup>Br]-MDA1. Se añade una alícuota de 50 microlitros similar de la digestión triptica de la muestra 2 a un tubo de microcentrifuga que contiene 10 mg de [<sup>81</sup>Br]-MDA1. Se incuban ambos tubos durante 3 h antes de mezclar los contenidos entre sí. Se purifican las moléculas marcadas por afinidad mediante cromatografía a través de una columna de afinidad de estreptavidina-agarosa (Sigma-Aldrich, St. Louis, MO) siguiendo el procedimiento recomendado por el fabricante. Se analiza la mezcla de péptido etiquetado recuperado mediante el espectrómetro de masas con los picos de defecto de masa deconvolucionados del ruido químico generado a partir de péptidos no marcados mediante los métodos de la presente invención. Se cuantificaron todos los pares de picos isotópicamente distintos restantes para determinar su abundancia relativa.

### Ejemplo 5

Este ejemplo ilustra el uso de etiquetas de masa fotoescindibles en métodos de secuenciación.

Usando los métodos de la presente invención, se incorporaron específicamente elementos de defecto de masa tales como bromo y europio en el componente de ajuste de intervalo de peso de un componente fotoescindible descrito por Ness *et al.*, patente estadounidense n.º 6.027.890. El defecto de masa proporcionado por estos elementos permite que fragmentos que contienen los marcadores de defecto de masa se deconvolucionen del ruido químico generado a partir de otras moléculas orgánicas que pueden estar presentes en la muestra. Además, este ejemplo muestra cómo el uso de algoritmos de deconvolución de pares de picos, descrito en el presente documento, permite la calificación de picos de señal baja en el espectro cuando se usan elementos de defecto de masa con abundancias naturales altas de isótopos estables.

La síntesis continuó tal como se describe en el ejemplo 5 de Nessel *al. ibid*, con la excepción que los compuestos R<sub>1-36</sub> añadidos en la etapa H consisten en derivados de bromofenilamida de aminoácidos con longitudes de cadena variables. Los derivados de bromofenilamida se preparan tal como sigue:

Se disuelven aproximadamente 5 g de ácido 3-bromobenzoico y 5 g de 1,3-diciclohexilcarbodiimida en 100 ml de tolueno seco. Se reparten en alícuotas aproximadamente 10 ml de esta disolución en cada uno de 10 viales de reacción. A cada alícuota de 10 ml, se añade una cantidad estequiométrica de uno de los ésteres *terc*-butílicos de los aminoácidos en la tabla 5.1 con respecto al ácido bromobenzoico. Se añade un éster *terc*-butílico de aminoácido diferente (preparado mediante métodos convencionales) a cada tubo. Se deja que avance la reacción durante la noche a temperatura ambiente y se elimina el éster *terc*-butílico sin reaccionar mediante la adición de ácido trifluoroacético. Entonces, se elimina el disolvente por evaporación y se purifican los derivados de bromofenilamida mediante HPLC de fase inversa preparativa usando cromatografía de fase inversa con elución por gradiente.

Se disuelven y se someten a cromatografía los derivados de bromofenilamida usando una fase estacionaria C<sub>8</sub> o C<sub>18</sub> de marca YMC (dimensiones ~ 25 cm x 6 mm de D.I., 5-15 μm, 120-150 Å) y una fase móvil en gradiente que consiste inicialmente en una mezcla de acetonitrilo y/o metanol con agua en una razón de 50/50; el analista ajusta la velocidad de flujo y el gradiente para el derivado de bromofenilamida específico. La fase acuosa puede modificarse opcionalmente para contener acetato de amonio, dietilamina, trietilamina, o hidróxido de amonio 0,1 molar para ayudar en la solubilidad del analito en la fase móvil en casos en los que se ha producido formación de cola o ensanchamiento de picos extremas. La parte orgánica puede modificarse opcionalmente en concentración mediante

5 la adición del 1-10% (en volumen) de alcohol isopropílico, alcohol diisopropílico o tetrahidrofurano para efectuar cambios en la selectividad entre los constituyentes en la mezcla de analitos y permitir el aislamiento del material de marcador de bromofenilamida deseado de sus impurezas. Se implementa el gradiente cambiando la concentración de disolvente total desde ~50% de parte orgánica (en volumen) hasta aproximadamente el 90-100% de parte orgánica en el periodo de 10 a 20 minutos. Se realizan el refino de los constituyentes de la fase móvil, la determinación de la velocidad de flujo, las concentraciones de disolvente inicial o final y la velocidad de gradiente para cada derivado tal como realizaría normalmente un experto en la técnica. Se combinan y evaporan fracciones aisladas del material de bromofenilamida deseado antes de la incorporación en la etiqueta de masa.

10 Este procedimiento genera una serie de marcadores con la composición general mostrada en la figura 10, que puede hacerse reaccionar con cualquier amina primaria que contenga molécula(s) diana a través del resto de ácido bloqueado con tetrafluorofenilo.

Tabla 5.1

Aminoácidos para su uso en la preparación de componentes de peso variable del grupo VI para etiquetas de masa

Glicina	Alanina	Ácido 4-aminobutírico
Leucina o Isoleucina	Ácido 2-amino-octanoico	Asparagina
Glutamina	Fenilalanina	Ácido 2-aminoheptanoico



ácido 2-Aminodecanoico	Valina	Ácido 2-aminopelargónico
------------------------	--------	--------------------------

### Ejemplo 6

Este ejemplo ilustra el uso de un marcador de defecto de masa fotoescindible generado en el ejemplo 5, anterior, en la determinación de la secuencia de bradiquinina.

5 En este ejemplo, el ácido 3-bromobenzoico y el marcador de etiqueta de masa conjugado con alanina se unen al extremo N-terminal del péptido bradiquinina usando métodos generalmente reconocidos en la técnica. Se diluye el péptido marcado hasta aproximadamente 1 ng por microlitro en una disolución de acetonitrilo:agua:triethylamina 50:50:1 en volumen. Se inyectó la disolución a aproximadamente 1 microlitro por minuto en un espectrómetro de masas ESI-TOF Mariner de Applied Biosystems equipado con un cabezal de micropulverizador convencional y ejecutado en modo de ion negativo. Se optimizaron los parámetros del pulverizador y el espectrómetro de masas para la abundancia relativa superior del estado de carga 3<sup>-</sup> del oligonucleótido dT<sub>6</sub> que podría conseguirse con una resolución de pico mayor de 5000. Un láser de colorante de onda estacionaria bombeado con Ar (Coherent), que se ajustó a 350 nm, se dirigió al hueco entre la punta del pulverizador y la boquilla del espectrómetro de masas, de manera que la pulverización de la muestra estaría completamente expuesta a la luz del láser para escindir la etiqueta de masa.

Se analizó la muestra marcada con etiqueta de masa acumulando 30 exploraciones de 3 segundos de duración. Se deconvolucionó el ruido químico en el espectro de masas usando los algoritmos de la presente invención, dejando los picos del marcador de defecto de masa (figura 11 A).

20 Se calificaron además estos picos deconvolucionados mediante las abundancias relativas de sus pares de isótopo usando el algoritmo:

$$\beta = \left( \frac{\left( \text{Recuentos}_{[{}^{79}\text{Br}]} + \text{Recuentos}_{[{}^{81}\text{Br}]} \right)}{2} \right) \left[ 1 - \frac{\left| \text{Recuentos}_{[{}^{79}\text{Br}]} - \text{Recuentos}_{[{}^{81}\text{Br}]} \right|}{\left( \text{Recuentos}_{[{}^{79}\text{Br}]} + \text{Recuentos}_{[{}^{81}\text{Br}]} \right)} \right]$$

25 Se substituyó la abundancia relativa del pico de masa inferior con el factor  $\beta$  de este cálculo. Los espectros de masas calificados por pico y deconvolucionados resultantes de la región de la etiqueta de masa se muestran en la figura 11B. Finalmente, se deconvolucionó adicionalmente la serie de isótopos en el espectro del factor  $\beta$  (figura 11C) a un único pico monoisotópico usando algoritmos generalmente conocidos en la técnica tal como se implementaron en el software BioSpec Data Explorer (versión 4.0, Applied Biosystems, Framingham, MA).

### Ejemplo 7

Este ejemplo ilustra la conjugación de un marcador de defecto de masa, éster de N-hidroxisuccinimida (NHS) del ácido 5-bromonicotínico, para montar la apomioglobina (Myo).

30 Se adquirieron Myo (calidad para secuenciación) (n.º de cat. A8673), ácido 5-bromonicotínico (5-BrNA) (n.º de cat. 228435), dodecilsulfato de sodio (SDS) (n.º de cat. L6026) y urea (n.º de cat. U0631) de Sigma-Aldrich y se usaron tal como se suministraron. Se adquirieron dimetilsulfóxido anhidro (DMSO) (n.º de cat. 20864), clorhidrato de 1-etil-3-(3-dimetilaminopropil)-carbodiimida (EDC) (n.º de cat. 22980)) y NHS (n.º de cat. 24500) de Pierce y se usaron tal como se suministraron.

35 Se preparó el éster de NHS de 5-BrNA *in situ* disolviendo 20,8 mg de 5-BrNA, 52,7 mg de NHS y 154,1 mg de EDC en 0,657 ml de DMSO. Se sonicó brevemente la muestra en un sonicador de baño para disolver rápidamente todos los sólidos. Se incubó la mezcla durante la noche a 4°C. El análisis del espectro de masas de la mezcla resultante indicó una conversión del 93% de 5-BrNA en éster de NHS (NHS-5-BrNA) mediante adición convencional.

40 Se desnaturizó Myo calentando a 95°C durante 20 min. a una concentración de 5,35 mg/ml en disolución acuosa de SDS al 5% (p/v). Tras enfriar a temperatura ambiente, se diluyó Myo hasta 1,07 mg/ml en tampón fosfato de sodio 80 mM, pH 7,0, que contenía concentraciones finales de SDS al 1% (p/v) y urea 6,4 M. Se marcó Myo con NHS-5-BrNA añadiendo 0,353 ml (50  $\mu$ mol) de NHS-5-BrNA preparado tal como se describió anteriormente a 2 ml (2,14 mg) de la mioglobina desnaturizada. Se incubó la muestra durante la noche a temperatura ambiente en la oscuridad. Entonces, se dializó extensamente la muestra con ácido acético acuoso al 50% (v/v) para eliminar la urea y SDS, que tiene un efecto perjudicial sobre el análisis del espectro de masas por electrospray. La pérdida de proteína fue evidente durante la diálisis extensa pero no se cuantificó. Tras la diálisis final, se secó la muestra hasta su finalización en un dispositivo Speedvac (Savant).

### Ejemplo 8

Este ejemplo ilustra la generación de especies de iones de fragmentos del espectro de masas de secuenciación a partir de mioglobina marcada con 5-BrNA (preparada tal como se describe en el ejemplo 7) mediante IMLS que están desviados del ruido químico periódico.

5 Se preparó una muestra para espectrometría de masas disolviendo mioglobina marcada con 5-BrNA seca (ejemplo 7) en 0,1 ml de una disolución acuosa de acetonitrilo al 50% que contiene ácido acético al 1% en volumen. Se sometió la proteína marcada a fragmentación en fuente en un espectrómetro de masas por electrospray con detectores de tiempo de vuelo (Mariner™, PE Biosystems, Inc.) tal como describen Schneider *et al.* (documento WO 00/63683, 26 de octubre 2000). Se optimizaron los parámetros del espectrómetro de masas y se calibró el instrumento inmediatamente antes de inyectar la muestra según las instrucciones del fabricante. Se infundió la muestra de manera continua mediante un capilar con D.I. de 50  $\mu\text{m}$  en la fuente de electrospray a una velocidad de 1  $\mu\text{l}/\text{min}$ . Se fijó el potencial de la boquilla a 300 V para inducir fragmentación en fuente. Se acumularon los espectros y se resumieron durante 345 s en el intervalo de 50-2000 unidades de masa con respecto a carga.

15 El examen de los datos del espectro de masas sin procesar mostró clara evidencia del ion de tipo b cargado de manera única del propio marcador (masa monoisotópica de 183,94) que se desvía  $\sim 0,15$  amu a la izquierda de los picos que son parte del ruido químico periódico que aparece en un periodo de  $\sim 1$  amu (véase la figura 12A). Se corrobora la identidad de este pico por la aparición de un segundo pico (185,94) que está  $\sim 2$  amu antes del primer pico, que corresponde al ion de fragmento marcado que incorpora el isótopo de masa superior de bromo ( $^{81}\text{Br}$ ). Las intensidades relativas de estos dos picos son casi equivalentes, reflejando la abundancia natural  $\sim 1:1$  de isótopos de bromo.

20 Esto ilustra la utilidad de generar iones de fragmentos específicos de marcador que incorporan elementos de defecto de masa (por ejemplo, bromo) que pueden resolverse del ruido químico generado a partir de proteínas (que están compuestas por elementos que no muestran defectos de masas fuertes) durante IMLS.

25 Se examinaron los datos del espectro para determinar la evidencia de picos desviados por defecto de masa que corresponden a iones de fragmentos del extremo N-terminal de la mioglobina. El doblete de ion  $a_1$  cargado de manera única (glicina) es evidente en 212,97 y 214,96 m/z (véase la figura 12B). Además, un doblete correspondiente a las masas calculadas del ion  $d_2$  (glicina-leucina) (284,05 y 286,05 m/z) es evidente (véase la figura 12C). Por tanto, se generan los iones de secuenciación. La abundancia generalmente baja de picos de iones de secuenciación observados con este marcador es un resultado de la alta intensidad del ion generado del propio marcador que está altamente estabilizado por la conjugación del carbonilo del marcador con el anillo piridililo (véase la figura 12A). La generación de esta especie altamente conjugada conducirá a escisión preferente del enlace de amida del marcador con respecto a la estructura principal de amida de la proteína, lo que conduce a una pérdida de iones de secuenciación significativa. Como resultado, sería preferible separar el carbonilo del marcador del anillo aromático mediante uno o más grupos metileno para hacer la unión de amida del marcador de energía de unión similar a la de la estructura principal de amida de la proteína.

### 35 Ejemplo 9

Este ejemplo ilustra la conjugación de un marcador de defecto de masa, éster de N-hidroxisuccinimida (NHS) del ácido 5-bromo-3-piridilacético (5-Br-3-PAA), para montar la apomioglobina (Myo).

40 Se adquirió 5-Br-3-PAA (n.º de cat. 13579) de Lancaster Synthesis y se usó tal como se suministró. Se adquirieron Myo (calidad para secuenciación) (n.º de cat. A8673), dodecilsulfato de sodio (SDS) (n.º de cat. L6026) y urea (n.º de cat. U0631) de Sigma-Aldrich y se usaron tal como se suministraron. Se adquirieron dimetilsulfóxido anhidro (DMSO) (n.º de cat. 20864), clorhidrato de 1-etil-3-(3-dimetilaminopropil)-carbodiimida (EDC) (n.º de cat. 22980) y NHS (n.º de cat. 24500) de Pierce y se usaron tal como se suministraron.

45 Se preparó el éster de NHS de 5-Br-3-PAA (NHS-5-Br-3-PAA) *in situ* disolviendo 12,7 mg de 5-Br-3-PAA, 7,4 mg de NHS y 12,5 mg de EDC en 0,235 ml de DMSO. Se incubó la mezcla durante 24 h a temperatura ambiente en la oscuridad. El análisis del espectro de masas de la mezcla resultante indicó que una conversión del 53% de 5-Br-3-PAA mediante adición convencional. Puesto que la conversión no estaba cerca a su finalización, se añadieron NHS adicional (7,2 mg) y EDC (7,5 mg) y se incubaron durante otras 24 h. El análisis del espectro de masas de la mezcla resultante tras este segundo periodo de incubación indicó una conversión del 93% del material de partida.

50 Se desnaturalizó Myo calentando 1,89 mg en 0,54 ml de disolución acuosa de SDS al 5% (p/v) a 95°C durante 20 min. Tras enfriar hasta temperatura ambiente, se añadieron a la muestra 1,89 ml de urea 9 M en tampón fosfato de sodio 20 mM, pH 7,0. Se añadió NHS-5-Br-3-PAA (0,24 ml,  $\sim 19$  mM de concentración final) a la mioglobina desnaturalizada. Se incubó la muestra durante la noche a temperatura ambiente en la oscuridad. Se dializó por centrifugación la mezcla de reacción frente a tampón Tris 25 mM, pH 8,3 que contiene SDS al 0,1% (p/v) para eliminar subproductos de reacción de NHS-5-Br-3-PAA y urea. La fracción retenida final ( $\sim 0,6$  ml) que contiene la mioglobina marcada se sometió a un procedimiento de extracción con cloroformo para eliminar el SDS unido (Puchades *et al.* (1999), Rap. Comm. Mass. Spec. 13, 344-349). Se añaden a la muestra, 2,4 ml de metanol, 0,6 ml de cloroformo y 1,8 ml de agua. Se mezcló la muestra invirtiendo el tubo una vez. Se centrifugó la muestra (3743 g, 20 min., temperatura ambiente) para ayudar en la separación de fases, y se desechó gran parte de la fase superior.

Se añadió metanol (1,8 ml) a la fase inferior restante y a la proteína que se había precipitado en la superficie de contacto. Se mezcló con vórtex vigorosamente el tubo y se sedimentó la proteína precipitada mediante centrifugación (3743 g, 40 min., temperatura ambiente). Se decantó y se desechó el sobrenadante y se secó el sedimento de proteína residual con una corriente de gas nitrógeno. Se resuspendió la Myo marcada secada en 0,4 ml de disolución acuosa de ácido acético al 10% (v/v). Se midió la concentración de proteína (2,6 mg/ml) mediante ensayo de BCA usando BSA como patrón.

### **Ejemplo 10**

Este ejemplo ilustra la generación de especies de iones de fragmentos del espectro de defecto de masas de secuenciación a partir de mioglobina marcada con 5-Br-3-PPA (preparada en el ejemplo 9) mediante IMLS que están desviados del ruido químico periódico.

Se preparó una muestra para espectrometría de masas disolviendo mioglobina marcada con 5-Br-3-PAA (150 µg) en 0,5 ml de una disolución acuosa de acetonitrilo al 50% que contiene ácido acético al 1,2% en volumen. Se sometió la proteína marcada a fragmentación en fuente en un espectrómetro de masas por electrospray con detectores de tiempo de vuelo (Mariner™, PE Biosystems, Inc.) tal como describen Schneider *et al.* (documento WO 00/63683, 26 de octubre de 2000). Se optimizaron los parámetros del espectrómetro de masas y se calibró el instrumento inmediatamente antes de inyectar la muestra según las recomendaciones del fabricante. Se infundió la muestra de manera continua mediante un capilar con D.I. de 50 µm en la fuente de electrospray a una velocidad de 1,2 µl/min. Se fijó el potencial de la boquilla a 225 V para inducir la fragmentación en fuente. Se acumularon los espectros y se resumieron durante 180 s en el intervalo de 50-2000 unidades de masa con respecto a carga.

El examen de los datos del espectro de masas sin procesar muestra clara evidencia de los iones b<sub>1</sub>-b<sub>6</sub> cargados de manera única derivados del extremo N-terminal de la mioglobina marcada (figuras 13A-B, figuras 14A y B y figuras 15 A y B). Los picos de doblete que corresponden a iones b que incorporan átomos o bien <sup>79</sup>Br o bien <sup>81</sup>Br (reflejando la abundancia natural de bromo ~50:50) se desvían por su defecto de masa (~0,15) a la izquierda del ruido químico que es evidente en una frecuencia periódica de ~1 amu. Por tanto, a diferencia de los iones de secuenciación limitada que se produjeron por la mioglobina marcada con 5-BrNA (comentada en los ejemplos 1 y 2), esta mioglobina marcada, con una unidad de metileno intermedia entre el anillo de piridina y el carbonilo del marcador, produce un marcador de masa molecular completo de iones b a través de los primeros seis residuos N-terminales. Esto es lo más probablemente un resultado de que el enlace de amida de 5-Br-3-PAA es más similar a los enlaces de amida de la estructura principal de la proteína en lo que se refiere a la capacidad de escisión en comparación con la de 5-BrNA, que parece ser más sensible a la escisión.

### **Ejemplo 11**

Este ejemplo describe un método para IMLS de apomioglobina marcada con 4-bromobenzaldehído.

El marcaje N-terminal de proteínas con aldehídos seguido por la estabilización con agentes reductores da como resultado marcadores con uniones de amina secundaria. Esta unión es más estable desde el punto de vista energético que la correspondiente amida. Por tanto, durante IMLS, el marcador debe mostrar fragmentación mínima en comparación con los enlaces amida del péptido y por tanto debe producir iones de fragmentos de interés más abundantes. Además, la generación de un sitio básico directamente a partir de la química proporciona un resto de carga "suave" que hace innecesaria la incorporación de un resto cargado en el marcador antes de la unión.

El 4-bromobenzaldehído está disponible de Sigma-Aldrich. Puede añadirse el aldehído a una disolución tamponada acuosa de mioglobina desnaturalizada de una manera similar a las condiciones descritas en el ejemplo 10 para el marcaje de mioglobina con el éster de NHS de ácido 5-bromonicotínico. Tras la finalización de la reacción, la mioglobina marcada se estabiliza mediante reducción de las iminas generadas con cianoborohidruro de sodio. Este agente reductor reduce selectivamente los enlaces dobles de tipo imina y no reduce otros grupos funcionales comúnmente encontrados en péptidos y proteínas. La unión química del marcador resultante es una amina secundaria. Entonces, la proteína marcada se purifica mediante diálisis o filtración en gel. La proteína marcada se disuelve en un sistema de tampón compatible con EM y se realiza IMLS de la manera tal como se describe en el ejemplo 10.

### **Ejemplo 12**

Este ejemplo ilustra IMLS de ubiquitina marcada con ácido 5-bromonicotínico.

La ubiquitina se marca con ácido 5-bromonicotínico según el procedimiento descrito en el ejemplo 7 para la apomioglobina. Se somete la ubiquitina marcada a fragmentación en fuente en el espectrómetro de masas tal como se describe en el ejemplo 8, y se analizan los espectros para determinar la presencia de iones del fragmento N-terminal que contienen el marcador.

### **Ejemplo 13**

Este ejemplo ilustra IMLS de apomioglobina marcada con ácido 6-bromo-2-hidroxi-quinolina-4-carboxílico (BHQC)

5 Se prepara el éster de NHS de BHQC según el procedimiento descrito en el ejemplo 7 para preparar el éster de NHS de ácido 5-bromonicotínico. La apomioglobina se marca con BHQC usando un procedimiento similar al descrito en el ejemplo 7. Se somete la apomioglobina marcada a fragmentación en fuente en el espectrómetro de masas tal como se describe en el ejemplo 10. Se analizan los datos del espectro de masas para determinar iones de fragmentos de extremo N-terminal marcados.

#### Ejemplo 14

Este ejemplo ilustra IMLS de ubiquitina marcada con ácido 6-bromo-2-hidroxiquinolina-4-carboxílico (BHQC).

10 La ubiquitina puede marcarse con BHQC según el procedimiento descrito en el ejemplo 13 para apomioglobina. Entonces la ubiquitina marcada se somete a fragmentación en fuente en el espectrómetro de masas tal como se describe en el ejemplo 10. Se analiza posteriormente los datos del espectro de masas para determinar iones de fragmentos de extremo N-terminal marcados.

#### Ejemplo 15

15 Este ejemplo ilustra el uso de los algoritmos de secuenciación y deconvolución automatizados de esta invención para encontrar la secuencia N-terminal de mioglobina marcada con 5-Br-3-PAA fragmentada en fuente en un espectrómetro de masas ESI-TOF tal como se describe en el ejemplo 5.

20 Se exportan los datos sin procesar usados para generar el espectro de masas del ejemplo 5 en formato ASCII del sistema de adquisición de datos. El periodo natural del ruido químico se determina a partir de estos datos sin procesar usando el código "deconvolver" mostrado en el apéndice y se determina que es 1,000575 amu. Usando este periodo natural, el espectro se convierte en referencia (archivo de salida \*.bs1) para corregir el error del instrumento, que siempre es positivo en EM (figura 16). Convertir en referencia significa que el valor de datos mínimo en cada bloque de datos de 1,000575 amu se ajusta a cero restando de cada punto de datos en el bloque de datos. Posteriormente, se procesa el archivo de datos de referencia con el "factor beta" como manera para calificar los picos de defecto de masa (que contienen Br), que siempre deben tener un pico [<sup>81</sup>Br] coincidente de 1,997954 amu antes del pico [<sup>79</sup>Br] (figura 16). Entonces, se procesa el archivo \*.bfc resultante a través del código de "secuenciador" (véase, solicitud en tramitación junto con la presente presentada el 19 de octubre de 2001, titulada "Methods for determining protein and peptide terminal sequences" expediente del apoderado 05265.P001 (publicación PCT WO 02/061661)), siendo la secuencia de mioglobina N-terminal verdadera (5-Br-3-PAA-GLSDGE; SEQ ID NO:1) la disolución clasificada más alta en los primeros cuatro residuos. En este ejemplo el código de "secuenciador" fue una búsqueda limitada para el primer estado de carga de los iones b.

30 Cuando se ejecuta el código de "secuenciador" para determinar la secuencia de los primeros cinco residuos, la secuencia GLSDW (SEQ ID NO:8), que da una masa teórica de 756,1993 se solapa (figura 17) con el pico correspondiente a la posición de defecto de masa del sexto residuo de la secuencia verdadera (GLSDGE (SEQ ID NO:9) a 756,1840). Esto da como resultado GLSDW (SEQ ID NO:8) que es la secuencia clasificada más alta en cinco residuos. Sin embargo, cuando el "secuenciador" se ejecuta a través de seis residuos, la secuencia verdadera GLSDGE (SEQID NO:9) se convierte de nuevo en la clasificada más alta debido a que GLSDW (SEQID NO:8) no propaga una secuencia en competencia en el sexto residuo. Esto muestra la ventaja de un algoritmo de probabilidad acumulativa.

#### Ejemplo 16

40 Este ejemplo ilustra la síntesis de un marcador de defecto de masa genérico que incorpora un elemento de defecto de masa de esta invención (es decir, bromo), un grupo ionizable (es decir, piridilo) y un resto de unión de anhídrido succínico para una unión al extremo N-terminal u otro grupo amino primario o secundario deseado de un polipéptido u otra especie. Se ha demostrado que el anhídrido succínico, y aparentemente sus derivados, reaccionan con eficacia casi cuantitativa hacia grupos amino del polipéptido (véase, Munchbach *et al.*, Anal. Chem. 72: 4047-4057 (2000)). Puede sintetizarse fácilmente otra especie alifática/aromática comparable que contiene cualquier combinación de grupos ionizables (A<sub>1</sub>...A<sub>n</sub>), elementos de defecto de masa (B<sub>1</sub>...B<sub>n</sub>) y un resto reactivo de anhídrido succínico de núcleo (SA) (véase la figura 18a).

50 Como estrategia a modo de ejemplo, la figura 18b explica un esquema sintético global para un marcador de defecto de masa {(A<sub>1</sub>...A<sub>n</sub>)-(B<sub>1</sub>...B<sub>n</sub>)-SA}. En este esquema, se convierte ácido 5-bromo-3-piridil acético (Lancaster, n.º de cat.13579) inicialmente en etil éster mediante la reacción con etanol en presencia de un catalizador ácido con eliminación de agua. Entonces, el éster resultante se bromina mediante la reacción con bromo elemental en una disolución básica de etóxido de sodio en etanol. Entonces, el carbono bromado se hace reaccionar selectivamente en un disolvente orgánico anhidro tal como tetrahidrofurano con di-(bromoacetaldehído dimetilacetal)cuprato de litio que se prepara mediante la reacción de bromoacetaldehído dimetilacetal disponible comercialmente (Aldrich, n.º de cat.242500) con litio para formar la especie de organolitio que se convierte en el cuprato mediante la reacción con Cu(II). El producto resultante se trata con ácido acuoso para eliminar el resto de acetal e hidrolizar el éster de nuevo en el ácido libre. El aldehído liberado se oxida en el ácido carboxílico correspondiente mediante agentes de oxidación convencionales (por ejemplo, Ag<sup>+</sup>), y la síntesis se completa mediante ciclación y deshidratación de los dos grupos de ácido carboxílico generados para formar el derivado de anhídrido succínico deseado.

Ejemplo 17

Este ejemplo ilustra el uso de marcadores de defecto de masa en aplicaciones de secuenciación de ADN. El esquema presentado (figura 19) representa una técnica de secuenciación a modo de ejemplo usando el método de Sanger; sin embargo, podría aplicarse metodología similar a otras estrategias de secuenciación de ADN tales como Maxam-Gilbert o PCR u otras estrategias conocidas por lo expertos en la técnica.

En resumen, un plásmido de M13 que porta una secuencia de ADN desconocida clonada (por ejemplo, d(GTTACAGGAAAT) (SEQ ID NO: 10) se hibrida inicialmente con un cebador de origen de replicación de M13 (3<sup>1</sup>d(AGTCACGACGACGTTGT)rA-5' (SEQ ID NO:4) que está marcado en el extremo 5' con rA para hacer que el cebador sea selectivamente escindible por la ARNasa (Integrated ADN Technologies, Inc., Coralville, Iowa). El volumen de reacción se divide por la mitad y se transfiere a dos tubos. En un tubo, se añaden polimerasa, dNTP, dGTP y ddATP\* con marcador de defecto de masa (véase la figura 20a) y ddGTP\* (véase la figura 20b). Al otro tubo, se añaden polimerasa, dNTP y ddTTP\* con marcador de defecto de masa (véase la figura 20c) y ddCTP\* (véase la figura 20d). Los ddNTP modificados mostrados en la figura 20 son a modo de ejemplo y se preparan según procedimientos convencionales (véase, Kricka, L.J., NONISOTOPIC DNA PROBE TECHNIQUES, Academic Press, New York (1992); Keller, G.H. y Manak, M.M., DNA PROBES, Stockton, New York (1989)). Son útiles otros muchos ddNTP modificados, que contienen bases de purina y pirimidina derivatizadas con restos de marcador de defecto de masa y se separan mediante una gran variedad de agentes de reticulación con diferentes longitudes y/o composiciones. La replicación de ADN y la extensión de cadena se inician mediante la incubación a 37°C. Se producen marcadores de masa molecular mediante terminación de cadena con los ddNTP. Una etapa de desnaturalización y escisión con ARNasa al final de la reacción elimina el producto terminado de cadena del molde y libera el cebador que puede eliminarse selectivamente mediante hibridación. Se disuelven los fragmentos de ADN en un tampón compatible con el espectrómetro de masas y se hacen pasar por un espectrómetro de masas ESI-TOF en modo de ion negativo. Los picos correspondientes a una serie de iones cargados de manera múltiple para cada fragmento se deconvolucionan usando algoritmos convencionales suministrados por el fabricante del instrumento (Applied Biosystems) para generar espectros que contienen sólo las masas de carga cero. Los espectros de carga cero se ubican en el centroide posteriormente usando también los algoritmos del proveedor del instrumento.

Se analizan los datos del espectro de masas tal como sigue. El espectro de la muestra que contiene ddA\* y ddG\* se deconvolucionan y se elimina el ruido químico, dejando sólo picos que han incorporado átomos de bromo o yodo (figura 21). El espectro de la muestra que contiene ddT\* y ddC\* se trata de manera similar (figura 22). Mirando a ambos espectros deconvolucionados, se encuentra el fragmento de masa más alto (4114,733) en el espectro ddA\*/ddG\* (figura 21). Este fragmento también puede identificarse como el fragmento que contiene un elemento de masa de yodo puesto que no hay par isotópico; por tanto, el último nucleótido en la secuencia "desconocida" es A. El fragmento de masa con la siguiente masa más baja es un doblete a 3695,611 y 3697,609 que se encuentra en el espectro ddT\*/ddC\* (figura 22). El doblete indica la incorporación de un átomo de bromo, y, por tanto, el siguiente nucleótido en la secuencia es T. Este procedimiento se repite hasta que se encuentra el último pico, en este caso, un pico de singlete a 748,1850 en el espectro de ddT\*/ddC\* correspondiente, por tanto, a C. Por tanto, se determina la secuencia ATTTCTGTAAAC (SEQ ID NO:11), y cuando se invierte y se sustituyen los complementos del nucleótido, se determina la secuencia "desconocida" GTTACAGGAAAT (SEQ ID NO:10).

En este ejemplo, se secuencia un segmento de ADN de ~4000 MW que está dentro las especificaciones para esta invención. Puesto que la capacidad para distinguir especies de defecto de masa que incorporan un átomo de defecto de masa se distorsiona en masas por encima de 5000, pueden secuenciarse segmentos de ADN más grandes que el ejemplo presentado aquí o bien usando más elementos de defecto de masa en los ddNTP de terminación, o bien, alternativamente, usando el método del "cebador rodante" (*rolling primer*). Con el método del "cebador rodante", se obtiene un segmento más corto del ADN deseado que va a secuenciarse usando el procedimiento anterior, y se prepara un nuevo cebador a partir de esta secuencia deducida para continuar la secuenciación a lo largo de la hebra de ADN más larga. Al final, los fragmentos más cortos pueden colocarse de extremo a extremo para revelar la secuencia del ADN desconocido.

Ejemplo 18

En este ejemplo se usa el marcador de defecto de masa (5-Br-3-PAA) para secuenciar ubiquitina bovina (Sigma-Aldrich). Se marcó la ubiquitina mediante el mismo procedimiento descrito en el ejemplo 7 para la mioglobina, excepto en que la etapa de marcaje de la proteína se realizó en acetonitrilo al 100%. Se preparó la muestra de ubiquitina marcada y se introdujo a un espectrómetro de masas ESI-TOF tal como se describe en el ejemplo 8. Se deconvolucionó y se secuenció el espectro de masas resultante tal como se describe en el ejemplo 10.

Se determinó correctamente la secuencia N-terminal de la ubiquitina verdadera (MQIFVK (SEQ ID NO: 12), obtenida del GenBank) cuando el "secuenciador" se ejecutó con dos, tres y cuatro residuos. La secuencia correcta se clasificó la segunda de 19 posibilidades en competencia en el primer residuo. También se clasificó segunda la secuencia correcta (para MQIFR; SEQ ID NO:13) en el quinto residuo.

## REIVINDICACIONES

1. Método para secuenciar una parte terminal de un oligómero, que comprende:
  - 5 (a) poner en contacto dicho oligómero con un resto de marcaje de defecto de masa para unir covalentemente el resto de marcaje de defecto de masa al extremo terminal del oligómero y formar un oligómero marcado, comprendiendo dicho resto de marcaje de defecto de masa al menos un elemento que tiene un número atómico de desde 17 hasta 77;
  - (b) fragmentar dicho oligómero marcado usando un método de fragmentación enzimático, quimiolítico o de espectrometría de masas para producir fragmentos de oligómero marcado; y
  - 10 (c) analizar dichos fragmentos de oligómero marcado usando un método de fragmentación de espectrometría de masas para determinar la secuencia de al menos dos residuos terminales basándose en una energía de unión nuclear de dicho al menos un elemento que confiere una masa única a dichos fragmentos de oligómero marcado.
2. Método según la reivindicación 1, en el que dicho resto de marcaje comprende al menos un elemento de número atómico de 35 a 63.
- 15 3. Método según la reivindicación 2, en el que dicho resto de marcaje comprende al menos un elemento de número atómico de 39 a 58.
4. Método según la reivindicación 2, en el que dicho resto de marcaje comprende al menos un elemento seleccionado del grupo que consiste en bromo, yodo, europio e itrio.
5. Método según la reivindicación 4, en el que dicho elemento es europio.
- 20 6. Método según la reivindicación 4, en el que dicho elemento es itrio.
7. Método según la reivindicación 4, en el que dicho elemento es bromo.
8. Método según la reivindicación 4, en el que dicho elemento es yodo.
9. Método según la reivindicación 1, en el que dicho oligómero se selecciona del grupo que consiste en una proteína, un oligonucleótido, un oligosacárido y un lípido.
- 25 10. Método según la reivindicación 9, en el que dicho oligómero es un oligonucleótido.
11. Método según la reivindicación 9, en el que dicha secuencia es de al menos tres residuos.
12. Método según la reivindicación 9, en el que dicha secuencia es de al menos cuatro residuos.
13. Método según la reivindicación 1, en el que varios oligómeros, cada uno marcado con un número diferente de elementos de defecto de masa, se mezclan antes de dicha etapa de fragmentación o análisis.
- 30 14. Método para secuenciar una parte de un oligómero en una mezcla de oligómeros, comprendiendo dicho método:
  - 35 (a) poner en contacto dicha mezcla de oligómeros con un resto de marcaje de defecto de masa para unir covalentemente el resto de marcaje de defecto de masa al extremo terminal de dicho oligómero y formar una mezcla de oligómeros marcados, comprendiendo dicho resto de defecto de masa al menos un elemento que tiene un número atómico de desde 17 hasta 77;
  - (b) separar oligómeros marcados individuales en dicha mezcla de oligómeros marcados; y
  - 40 (c) analizar dichos oligómeros marcados individuales de la etapa (b) mediante un método de espectrometría de masas para determinar la secuencia de al menos dos residuos terminales basándose en una energía de unión nuclear de dicho al menos un elemento que confiere una masa única a dichos oligómeros marcados individuales.
15. Método según la reivindicación 14, en el que dicho elemento tiene un número atómico de desde 35 hasta 63.
16. Método según la reivindicación 14, en el que dicho elemento tiene un número atómico de desde 39 hasta 58.
- 45 17. Método según la reivindicación 14, en el que dicho elemento se selecciona del grupo que consiste en bromo, yodo, europio e itrio.

18. Método según la reivindicación 14, que comprende además una etapa antes de la etapa (a) de aislar un grupo de oligómeros de una muestra biológica.
19. Método según la reivindicación 18, en el que dicha muestra biológica es de una muestra de tejido con enfermedad.
- 5 20. Método según la reivindicación 18, en el que dicha muestra biológica es de una muestra de tejido sano.
21. Método según la reivindicación 14, en el que dicha separación se realiza mediante al menos un método de electroforesis capilar de la mezcla de oligómeros marcados.
22. Método según la reivindicación 14, en el que dicho método de espectrometría de masas usa EM ESI-TOF.
- 10 23. Método para el análisis de la función y estructura de un oligómero que tiene una pluralidad de residuos, comprendiendo dicho método:
- (a) poner en contacto dicho oligómero con un reactivo de marcaje de defecto de masa para marcar de manera diferencial residuos expuestos y residuos no expuestos y producir un oligómero marcado de manera diferencial, en el que dicho reactivo de marcaje de defecto de masa comprende al menos un elemento que tiene un número atómico de desde 17 hasta 77; y
- 15 (b) analizar dicho oligómero marcado de manera diferencial mediante un método de espectrometría de masas para determinar secuencias de dicho oligómero que están expuestas en la estructura tridimensional y secuencias de dicho oligómero que no están expuestas en la estructura tridimensional basándose en una energía de unión nuclear de dicho al menos un elemento que confiere una masa única a dicho oligómero marcado de manera diferencial.
- 20 24. Método según la reivindicación 23, en el que dicho oligómero es una proteína, un ácido nucleico o un oligosacárido.
25. Método según la reivindicación 23, en el que dicho reactivo de marcaje de defecto de masa comprende al menos un elemento de número atómico de 35 a 63.
- 25 26. Método según la reivindicación 23, en el que dicho reactivo de marcaje de defecto de masa es bromo y dicho oligómero es una proteína.
27. Método según la reivindicación 23, en el que dicho reactivo de marcaje de defecto de masa comprende al menos un elemento de número atómico de 39 a 58.
28. Método según la reivindicación 23, en el que dicho oligómero marcado de manera diferencial se fragmenta mediante métodos quimiolíticos o enzimáticos antes de la etapa (b).
- 30 29. Método según la reivindicación 23, en el que dicho oligómero es una proteína, dicho defecto de masa es bromo o yodo y dichos residuos expuestos comprenden una parte de los residuos de tirosina presentes en dicha proteína.
30. Método según la reivindicación 23, en el que dicho método de espectrometría de masas usa EM ESI-TOF.
31. Método según la reivindicación 29, en el que dicho método de espectrometría de masas usa EM ESI-TOF.
- 35 32. Método para secuenciar la parte terminal de un oligómero, que comprende:
- (a) poner en contacto una primera muestra de dicho oligómero con un resto de marcaje de defecto de masa para unir covalentemente el resto de marcaje de defecto de masa al extremo terminal del oligómero y formar un oligómero marcado, teniendo dicho resto de marcaje de defecto de masa un elemento con un número atómico de desde 17 hasta 77;
- 40 (b) poner en contacto una segunda muestra de dicho oligómero con un resto de marcaje de defecto de masa para unir covalentemente el resto de marcaje de defecto de masa al extremo terminal del oligómero y formar un oligómero marcado, teniendo dicho resto de marcaje de defecto de masa dos elementos con un número atómico de desde 17 hasta 77;
- 45 (c) opcionalmente, repetir la etapa (b) desde una hasta tres veces con muestras adicionales, en el que los restos de marcaje de defecto de masa tienen tres, cuatro o cinco elementos, respectivamente, con un número atómico de desde 17 hasta 77;
- (d) mezclar los oligómeros marcados de las etapas (a) a la (c);
- (e) fragmentar dichos oligómeros marcados usando un método de fragmentación enzimático, quimiolítico o de espectrometría de masas para producir fragmentos de oligómero marcado; y

- (f) analizar dichos fragmentos de oligómero marcado usando un método de fragmentación de espectrometría de masas para determinar la secuencia de al menos dos residuos terminales basándose en una energía de unión nuclear de dicho un elemento, dichos dos elementos, o dichos tres, cuatro o cinco elementos que confieren una masa única a dichos fragmentos de oligómero marcado.
- 5 33. Método según la reivindicación 32, en el que cada uno de dichos elementos tiene un número atómico de desde 35 hasta 63.
34. Método según la reivindicación 32, en el que cada uno de dichos elementos tiene un número atómico de desde 39 hasta 58.
- 10 35. Método según la reivindicación 32, en el que cada uno de dichos elementos se selecciona del grupo que consiste en bromo, yodo, europio e itrio y dicho oligómero es una proteína.
36. Método según la reivindicación 32, en el que cada uno de dichos elementos se selecciona del grupo que consiste en bromo, yodo, europio e itrio y dicho oligómero es un oligonucleótido.
37. Método según la reivindicación 32, en el que cada uno de dichos elementos se selecciona del grupo que consiste en bromo, yodo, europio e itrio y dicho oligómero es un oligosacárido.
- 15 38. Método para secuenciar una parte de un oligómero, que comprende:
- a) fragmentar alícuotas de dicho oligómero usando uno o más métodos de fragmentación quimiolíticos o enzimáticos específicos para producir fragmentos de oligómero, en el que se aplica un método de fragmentación diferente a cada alícuota;
- 20 (b) poner en contacto una primera alícuota de fragmentos de oligómero con un primer resto de marcaje de defecto de masa para unir covalentemente dicho primer resto de marcaje de defecto de masa al extremo terminal de los fragmentos de oligómero y formar fragmentos de oligómero marcado, teniendo dicho primer resto de marcaje de defecto de masa un elemento con un número atómico de desde 17 hasta 77;
- 25 (c) opcionalmente poner en contacto las otras alícuotas de fragmentos de oligómero con otros restos de marcaje de defecto de masa distintos para unir covalentemente dichos restos de marcaje de defecto de masa distintos a los extremos terminales de los fragmentos de oligómero y formar fragmentos de oligómero marcado, teniendo dicho resto de marcaje de defecto de masa distinto dos o más elementos con un número atómico de desde 17 hasta 77;
- (d) opcionalmente mezclar las alícuotas de fragmentos de oligómero marcado; y
- 30 (e) analizar dichos fragmentos de oligómero marcado usando un método de fragmentación de espectrometría de masas para determinar la secuencia de al menos dos residuos de dicho oligómero basándose en una energía de unión nuclear de dicho un elemento o dichos dos o más elementos que confieren una masa única a dichos fragmentos de oligómero marcado.
39. Método según la reivindicación 38, en el que dicho oligómero es un lípido.
40. Método según la reivindicación 38, en el que dicho oligómero es una proteína.
- 35 41. Método según la reivindicación 38, en el que dicho oligómero es un ácido nucleico.
42. Método según la reivindicación 38, en el que dicho oligómero es un oligosacárido.
43. Método según la reivindicación 38, en el que dichos elementos tienen un número atómico de desde 35 hasta 63.
- 40 44. Método según la reivindicación 43, en el que dichos elementos tienen un número atómico de desde 39 hasta 58.
- 45 45. Método para comparar las abundancias relativas de analitos a partir de dos o más muestras, que comprende:
- (a) poner en contacto los analitos de la primera muestra con un resto de marcaje de defecto de masa para unir covalentemente el resto de marcaje de defecto de masa a los analitos y formar analitos marcados, teniendo dicho resto de marcaje de defecto de masa un elemento con un número atómico de desde 17 hasta 77;
- 50 (b) poner en contacto los analitos de muestras posteriores con restos de marcaje de defecto de masa para unir covalentemente los restos de marcaje de defecto de masa a los analitos en cada muestra, en el que los restos de marcaje de defecto de masa usados para cada muestra posterior contienen un elemento adicional con un número atómico de desde 17 hasta 77;



(c) mezclar las alícuotas de analitos marcados; y

(d) analizar dichos fragmentos de oligómero marcado usando un método de fragmentación de espectrometría de masas para determinar las abundancias relativas de uno o más de los analitos entre las muestras basándose en una energía de unión nuclear de dicho un elemento o dichos elementos adicionales que confieren una masa única a dichos fragmentos de oligómero marcado.

- 5
46. Método según la reivindicación 45, en el que dichos elementos tienen un número atómico de desde 35 hasta 63.
47. Método según la reivindicación 45, en el que dichos elementos tienen un número atómico de desde 39 hasta 58.
- 10 48. Método según la reivindicación 45, en el que al menos una parte de dicho resto de marcaje de la etapa (a) es un isótopo estable de dicho resto de marcaje de la etapa (b).

Figura 1

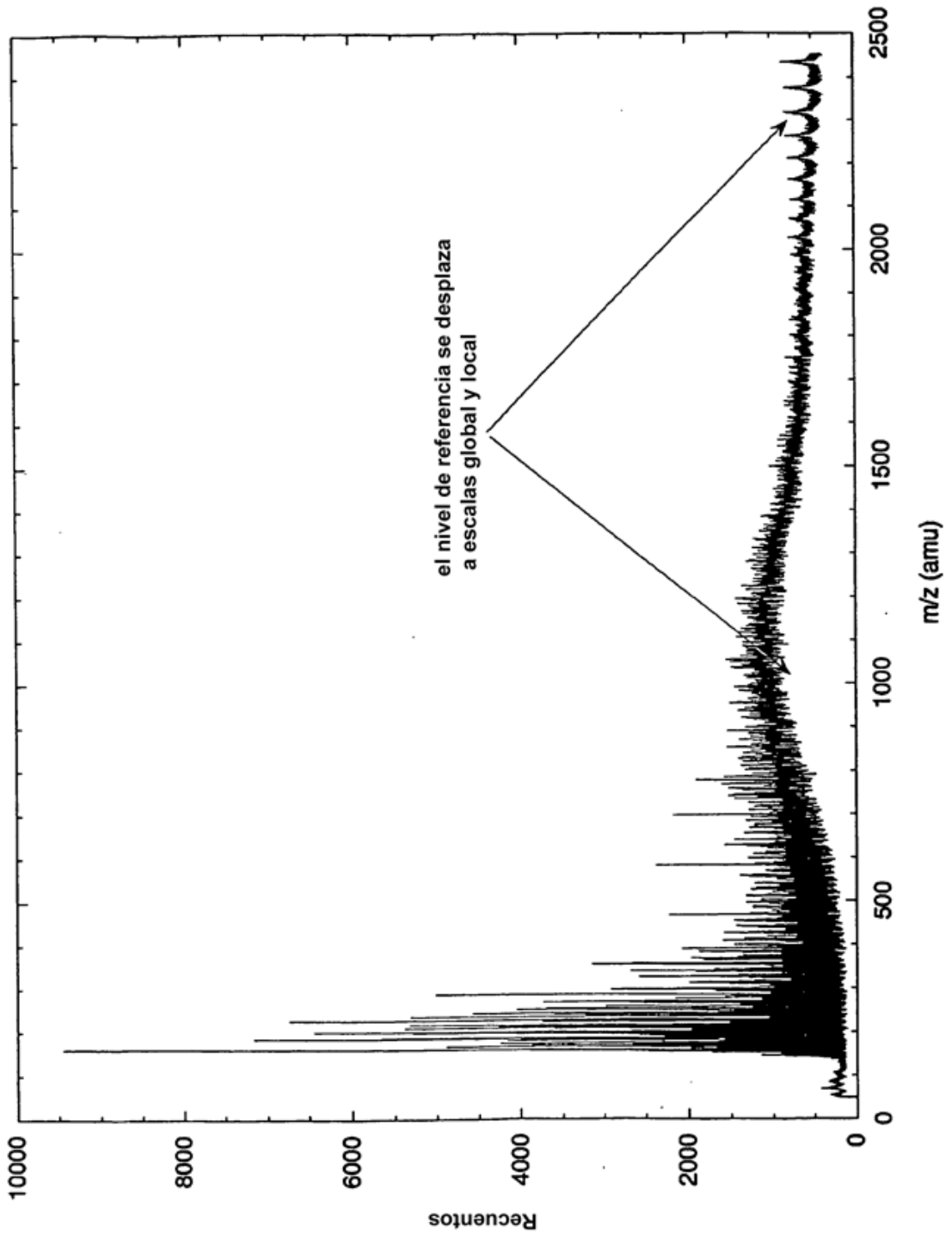


Figura 2

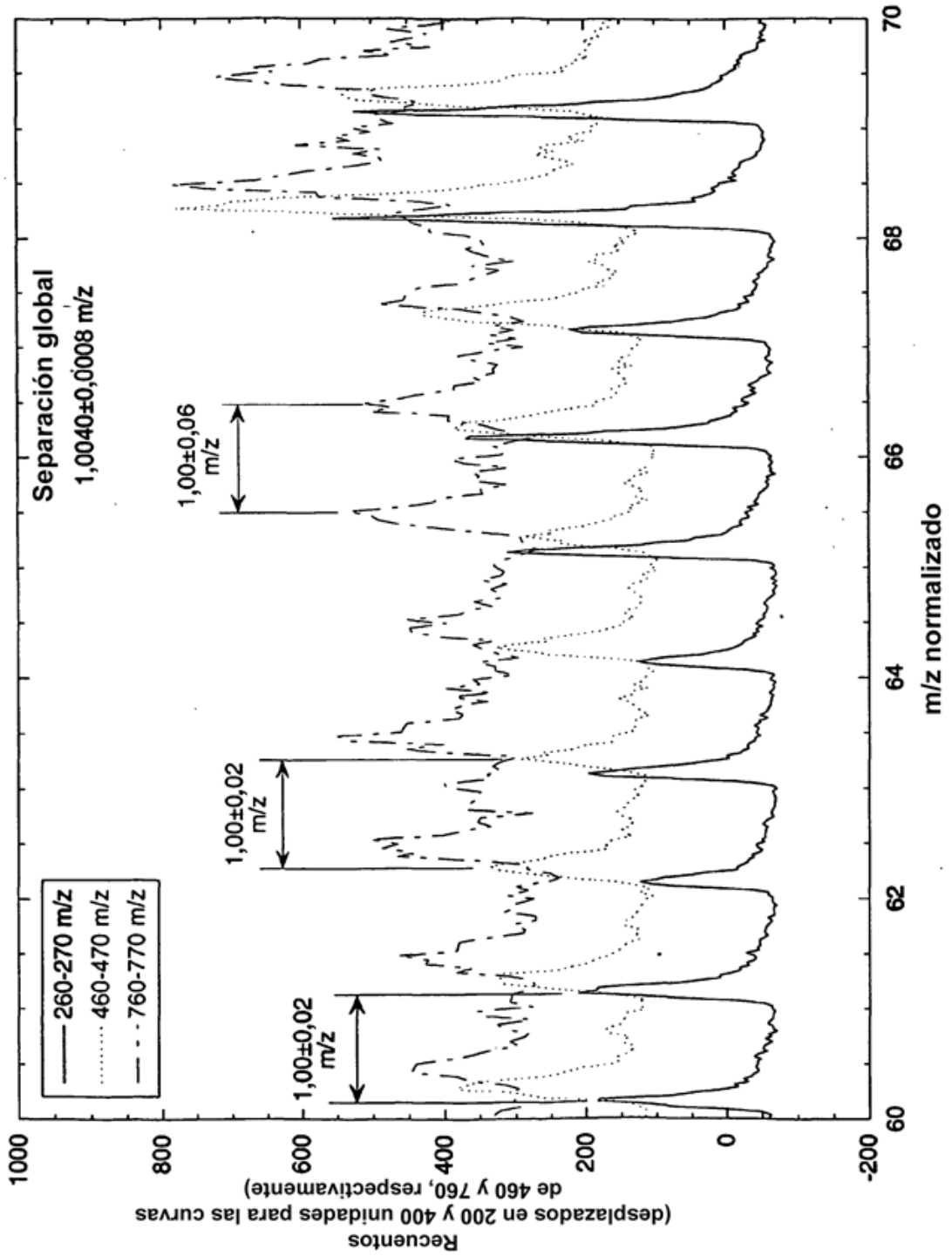


Figura 3

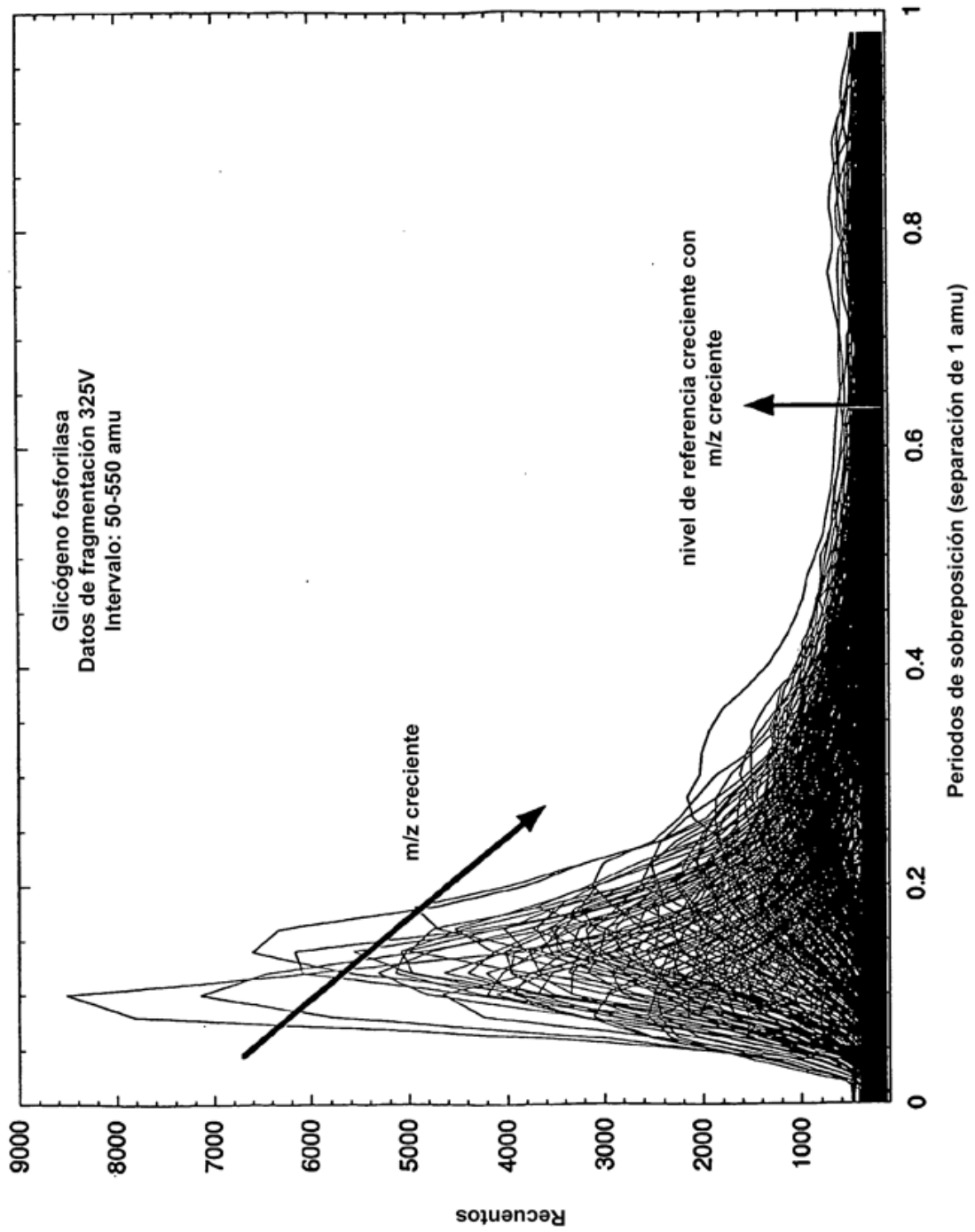


Figura 4

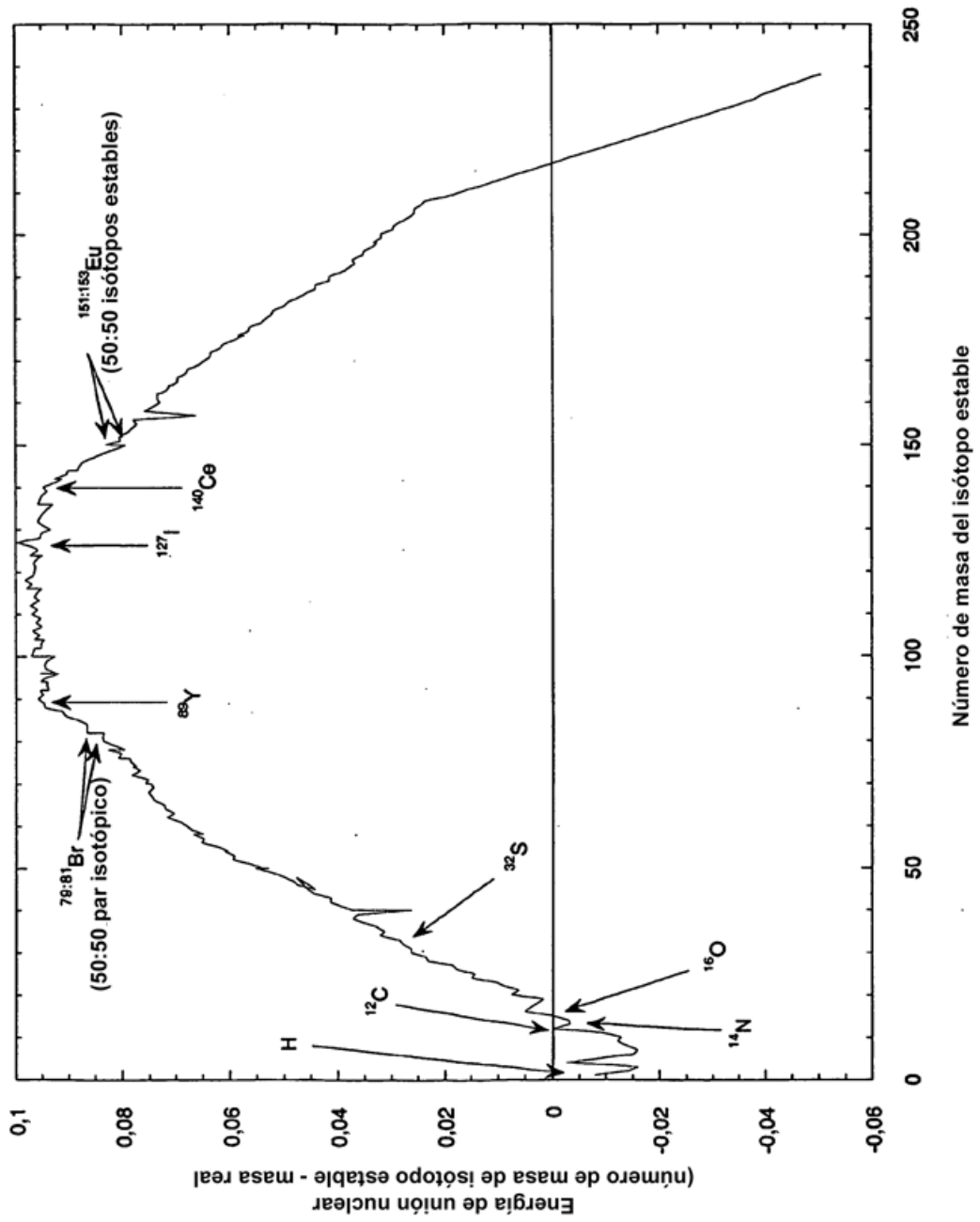


Figura 5

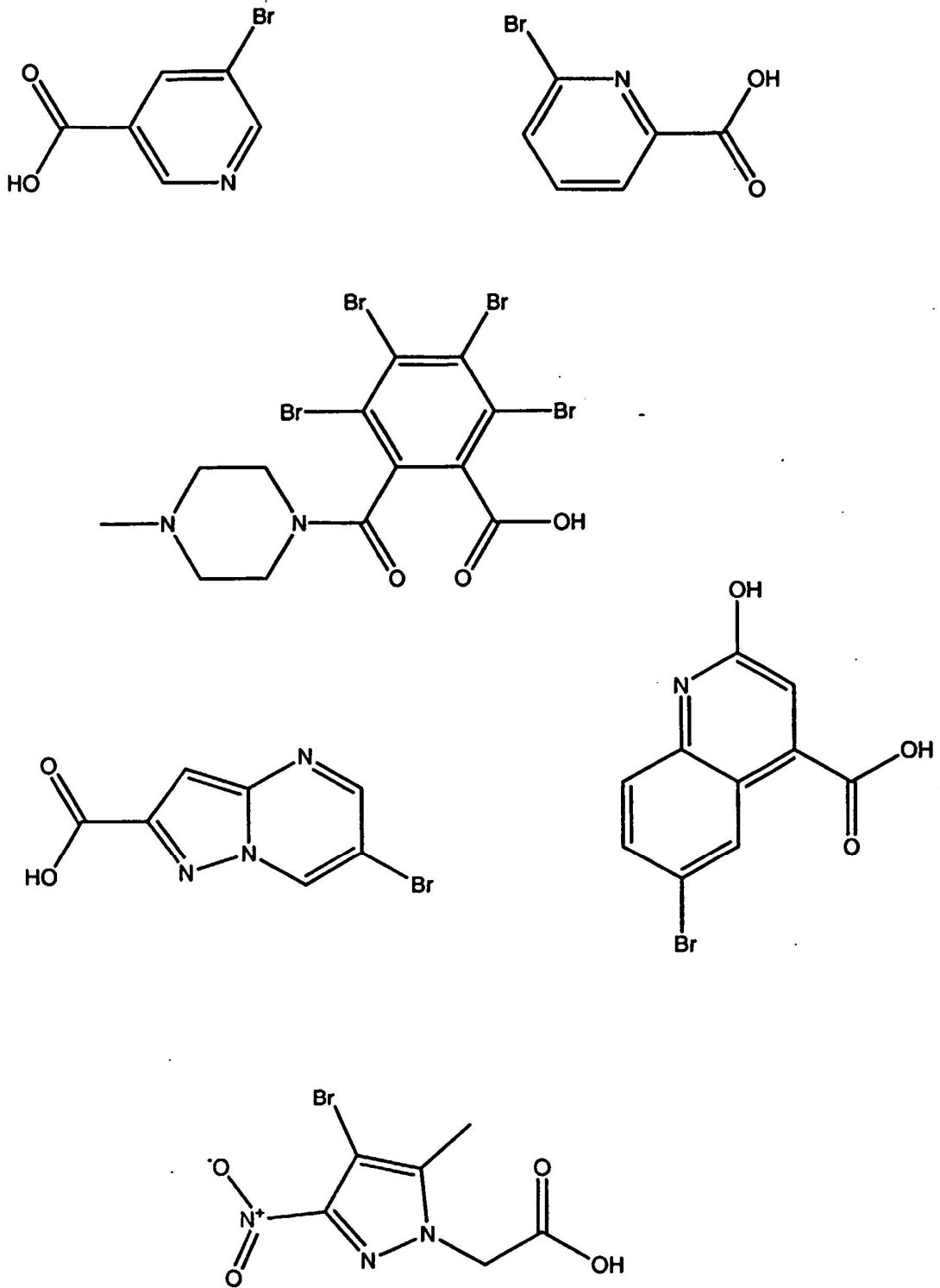
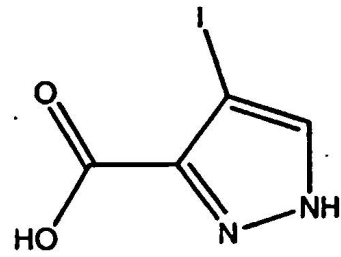
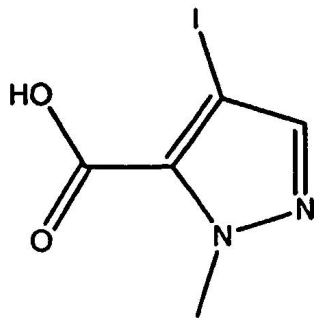
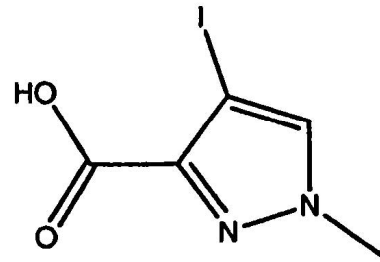
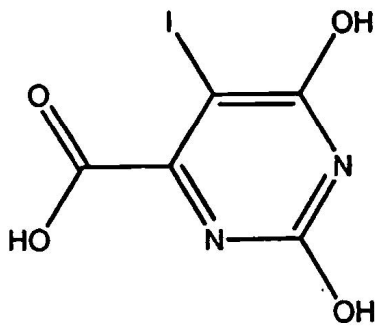


Figura 6



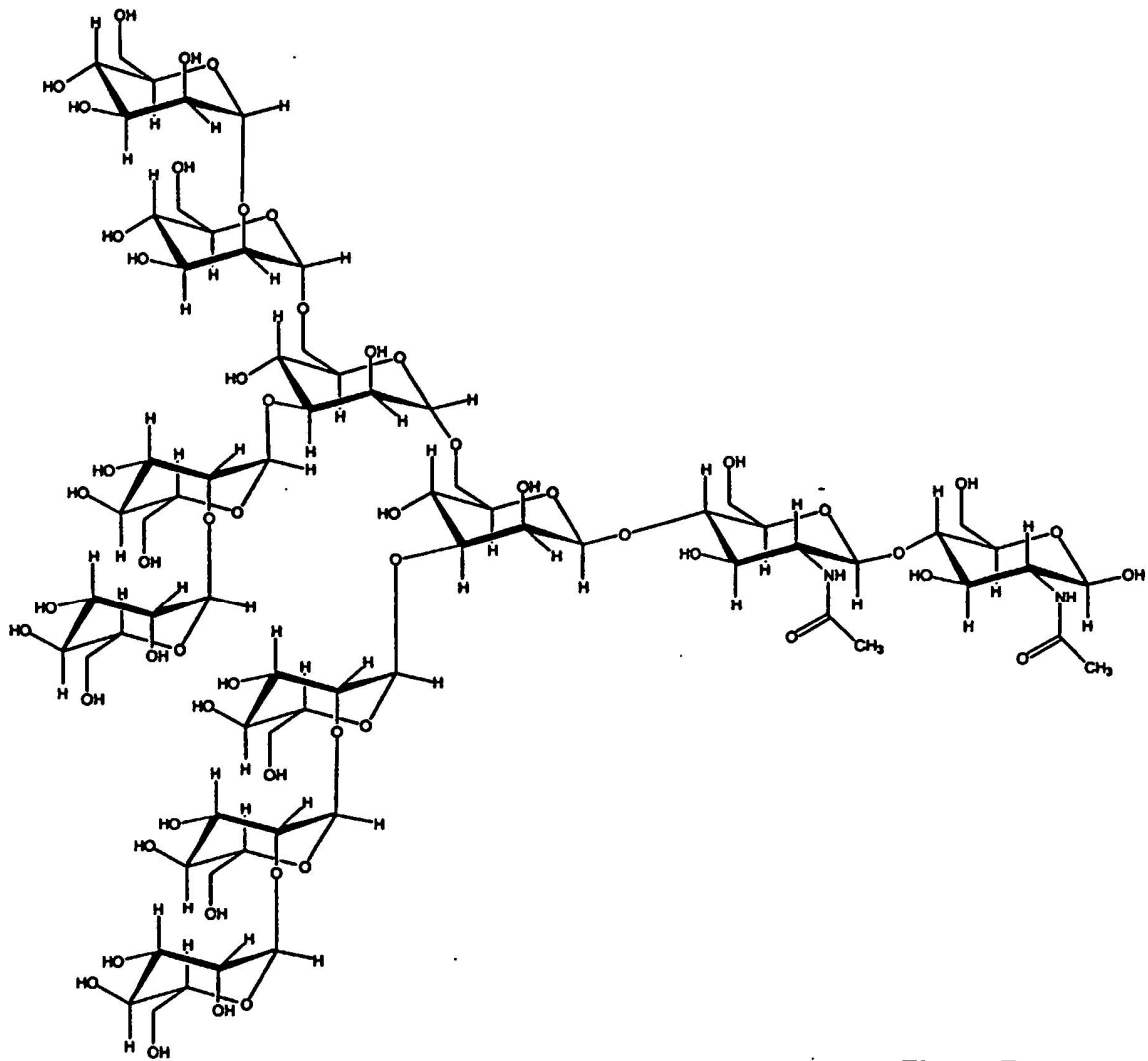


Figura 7



Figura 8A

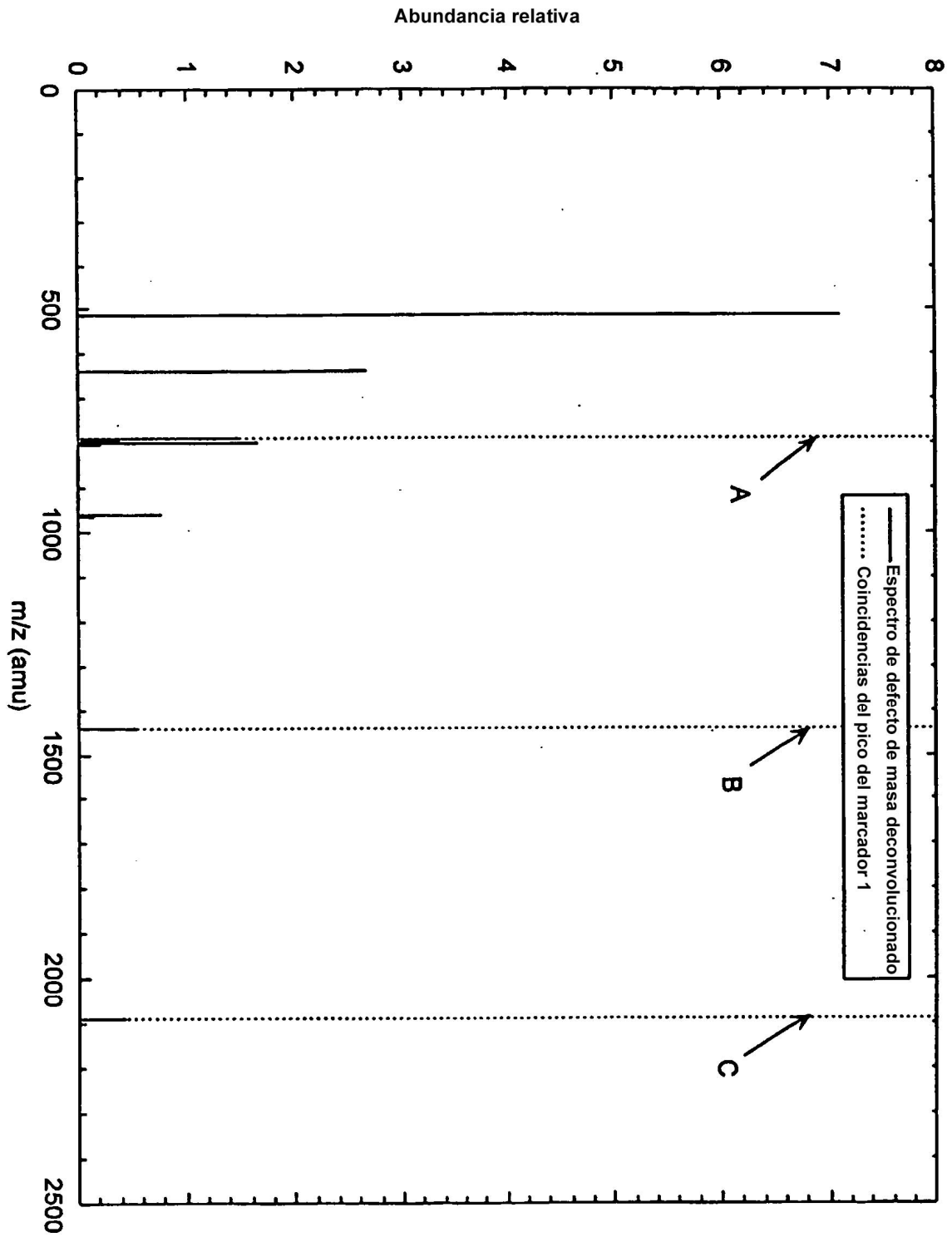


Figura 8B

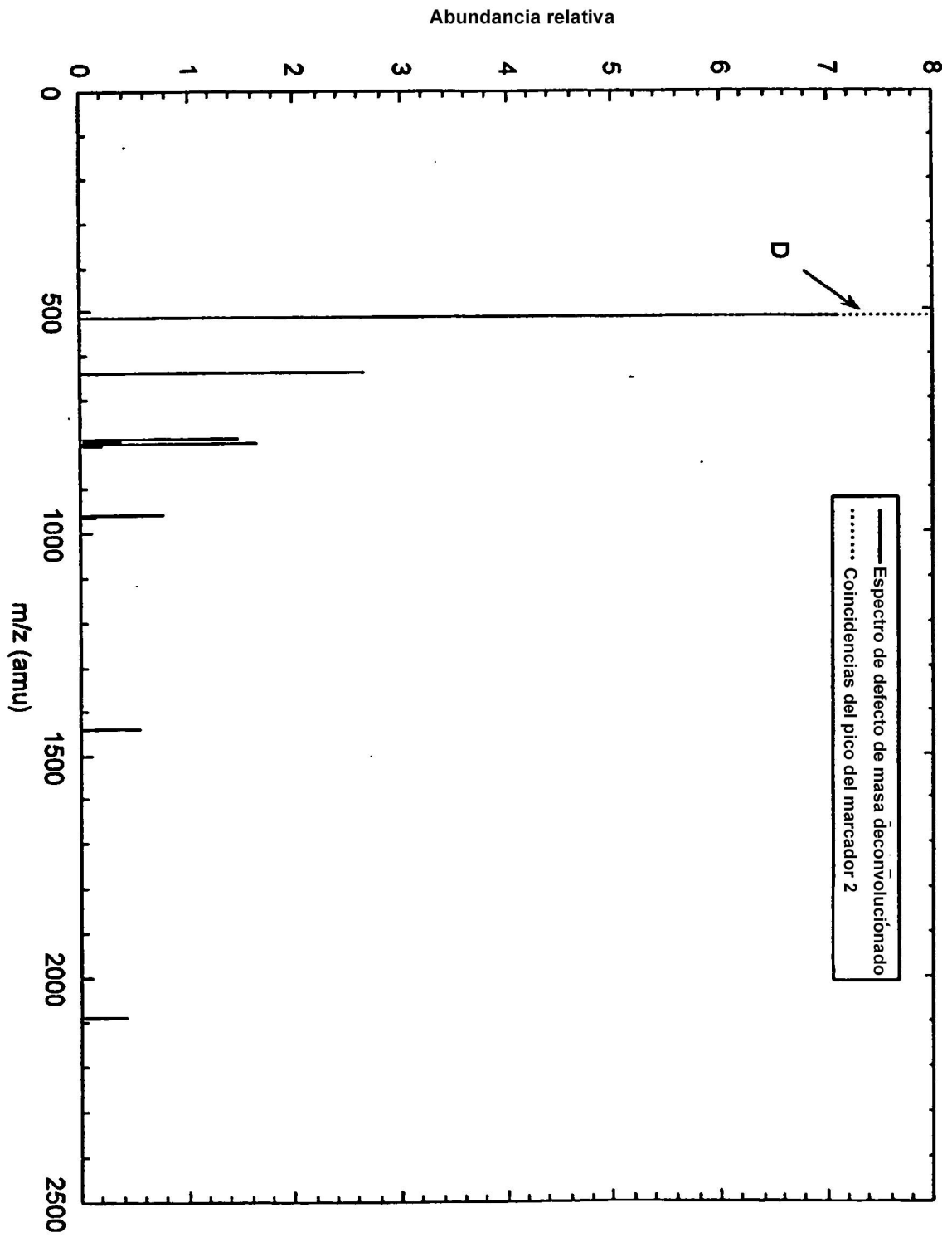


Figura 8C

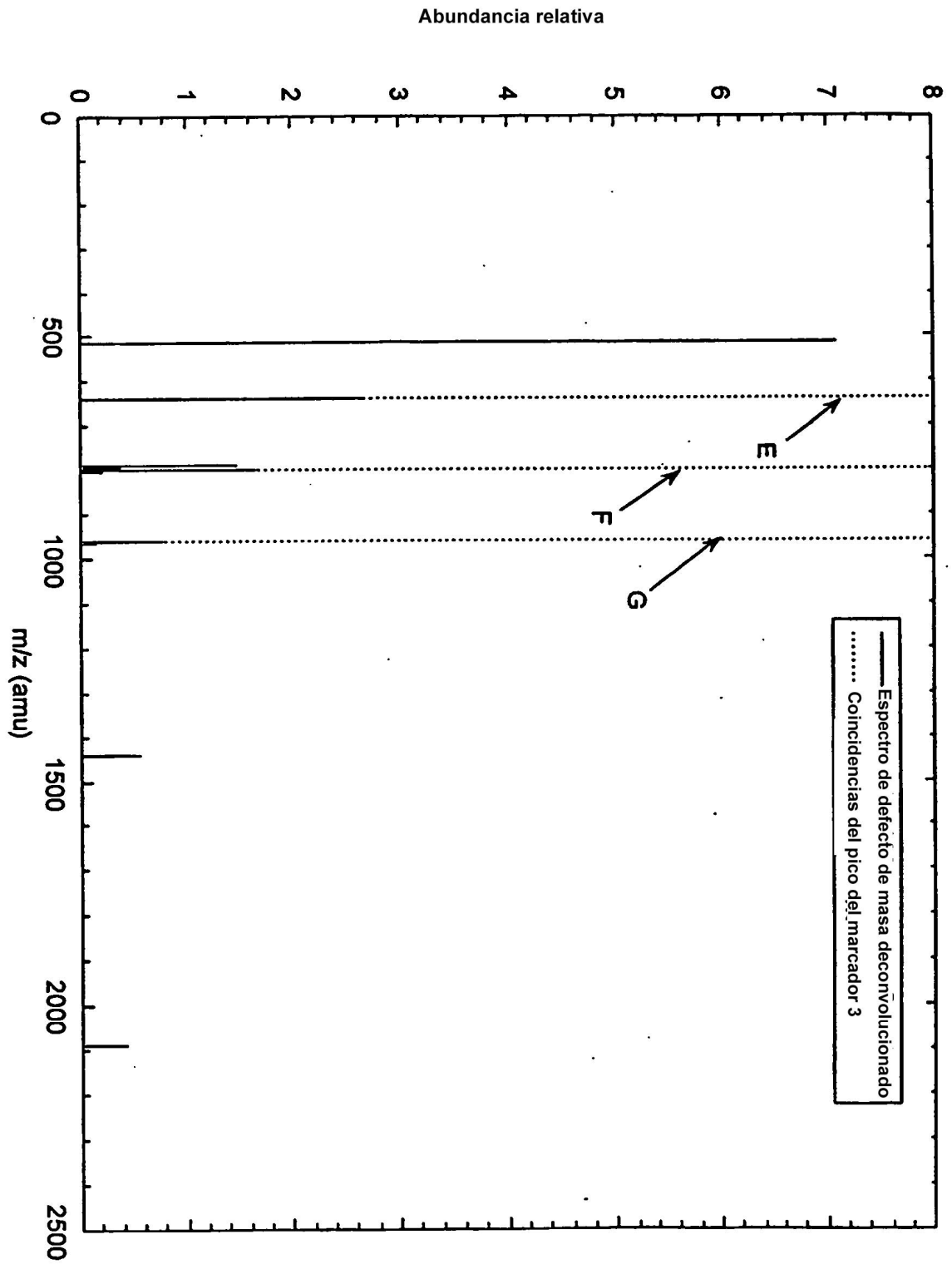


Figura 9

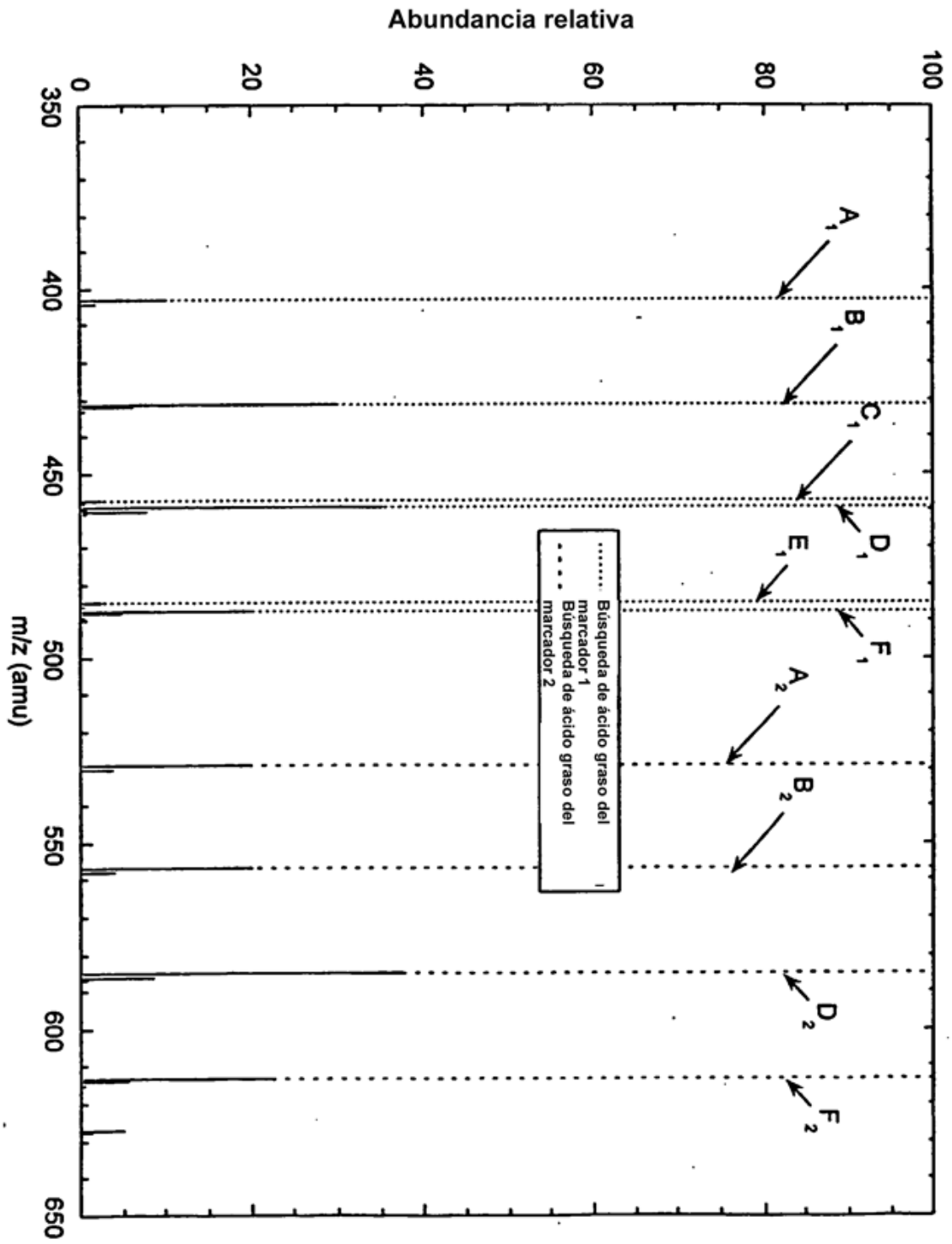
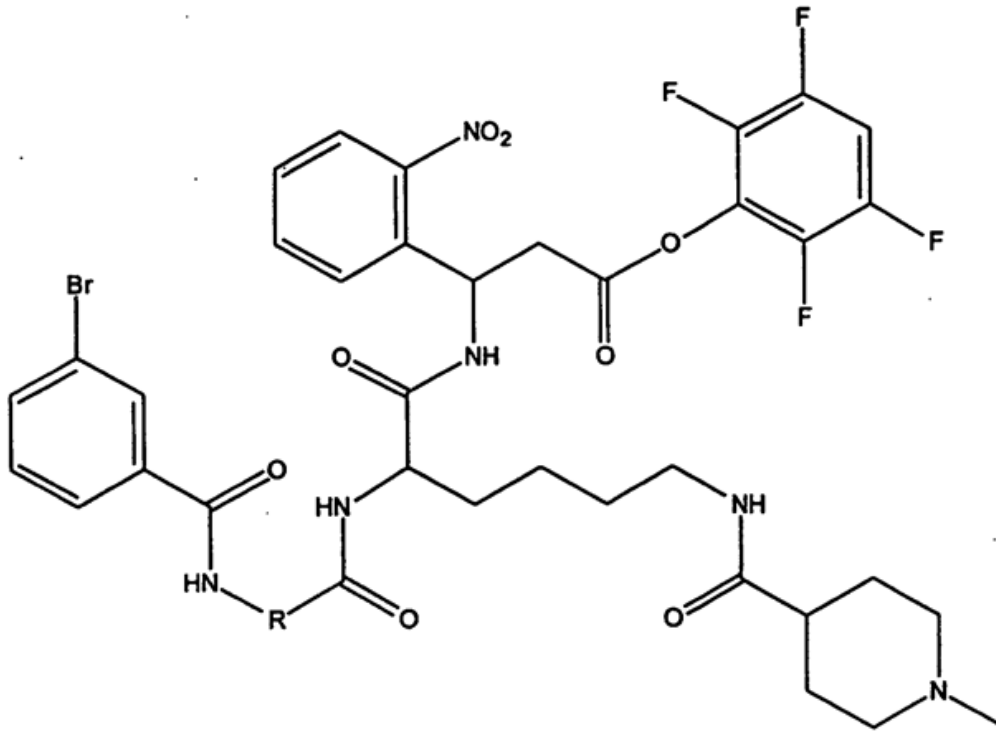


Figura 10



La estructura general de la etiqueta de defecto de masa fotoescindible en la que Br es el elemento de defecto de masa que está unido a través del aminoácido (R) al resto de la etiqueta.

Figura 11A

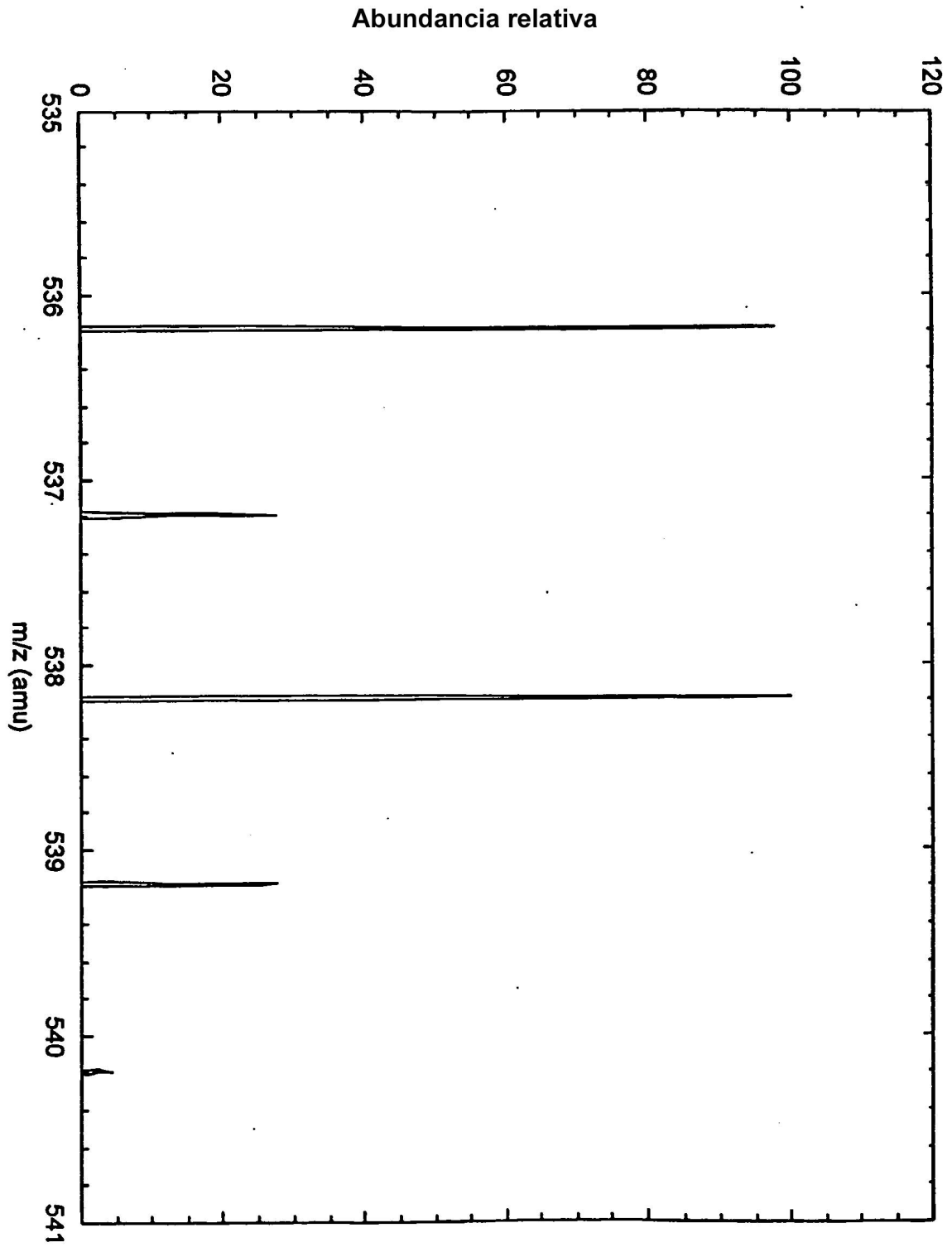


Figura 11B

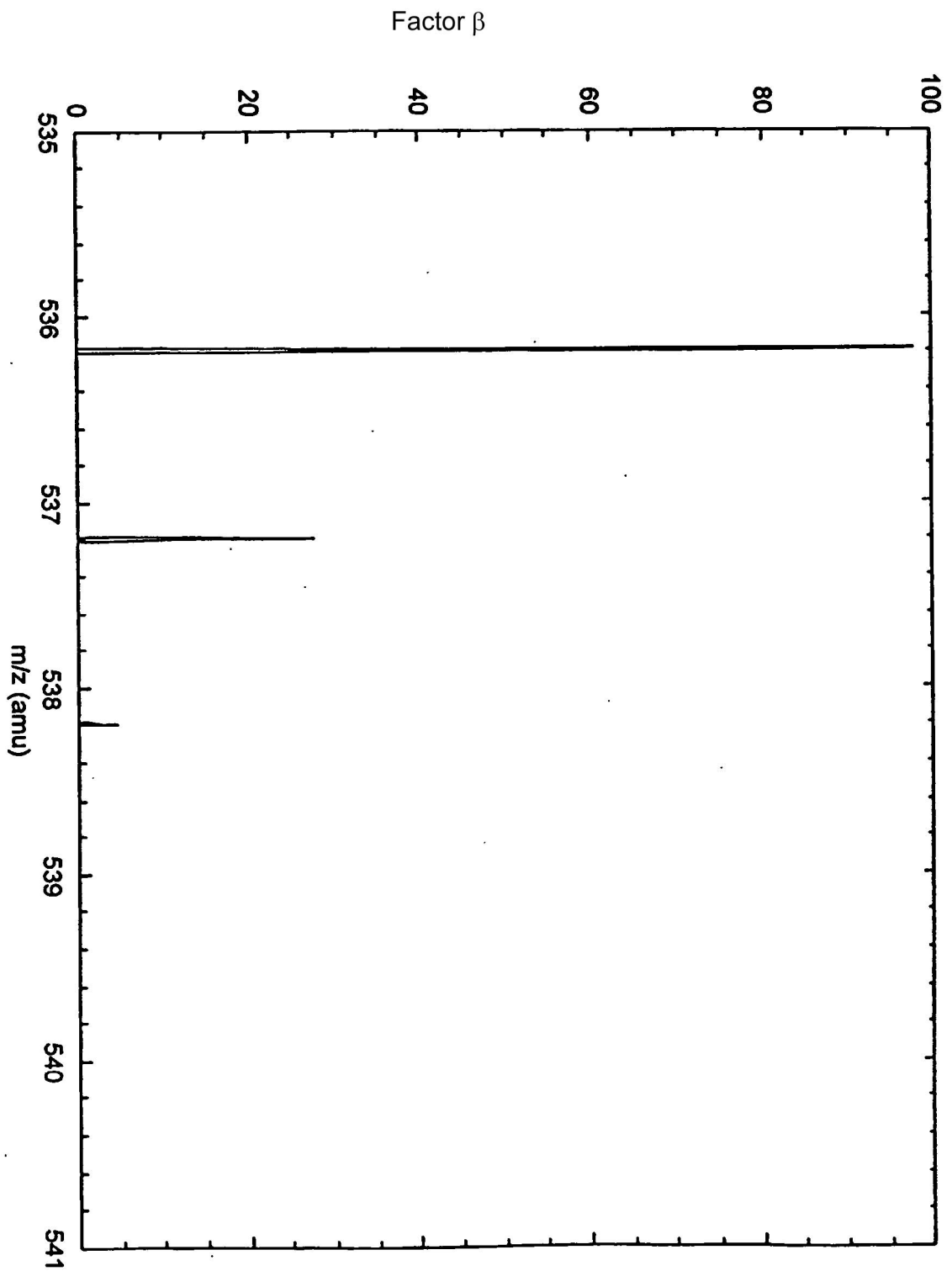


Figura 11C

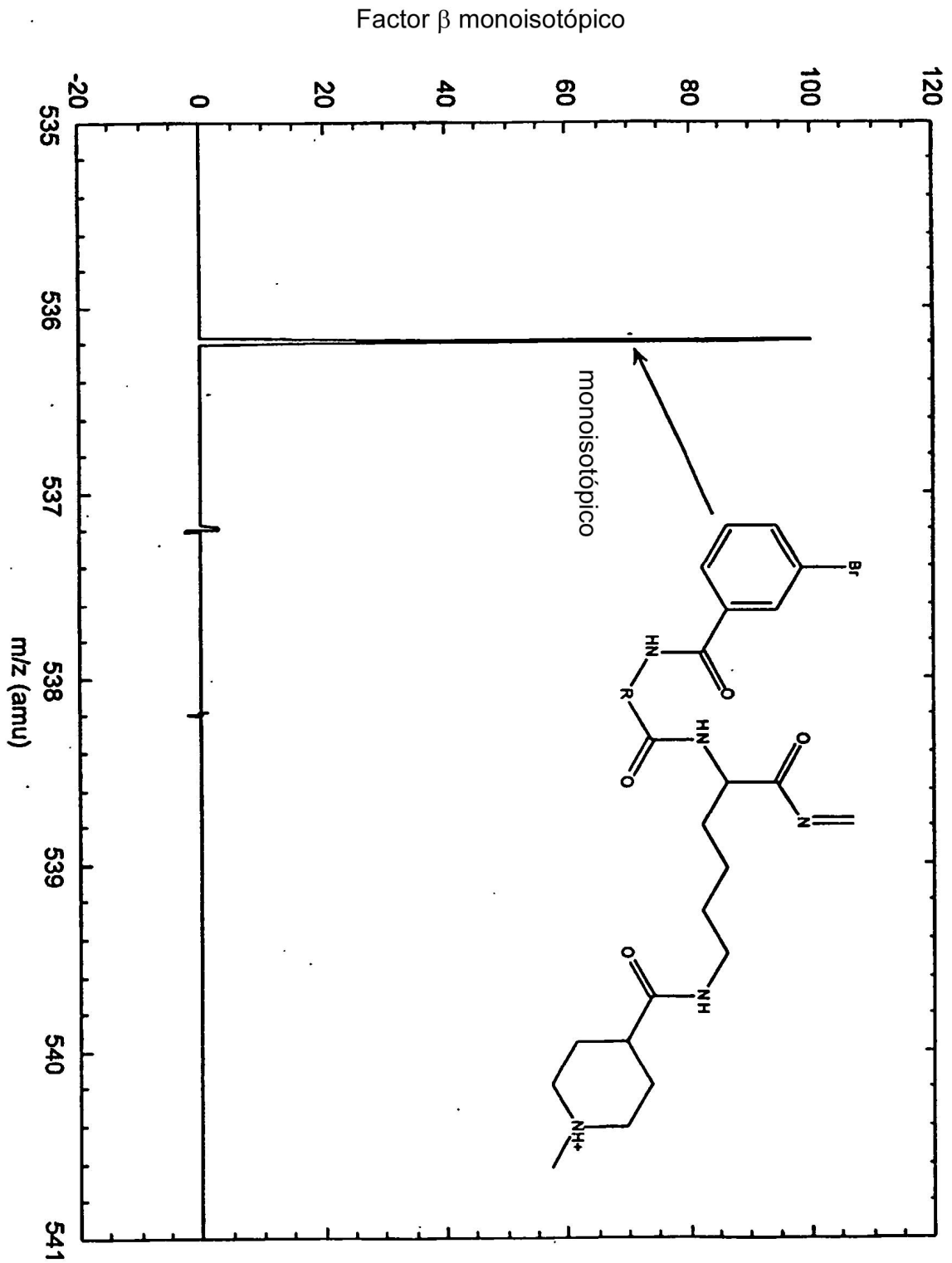




Figura 12A

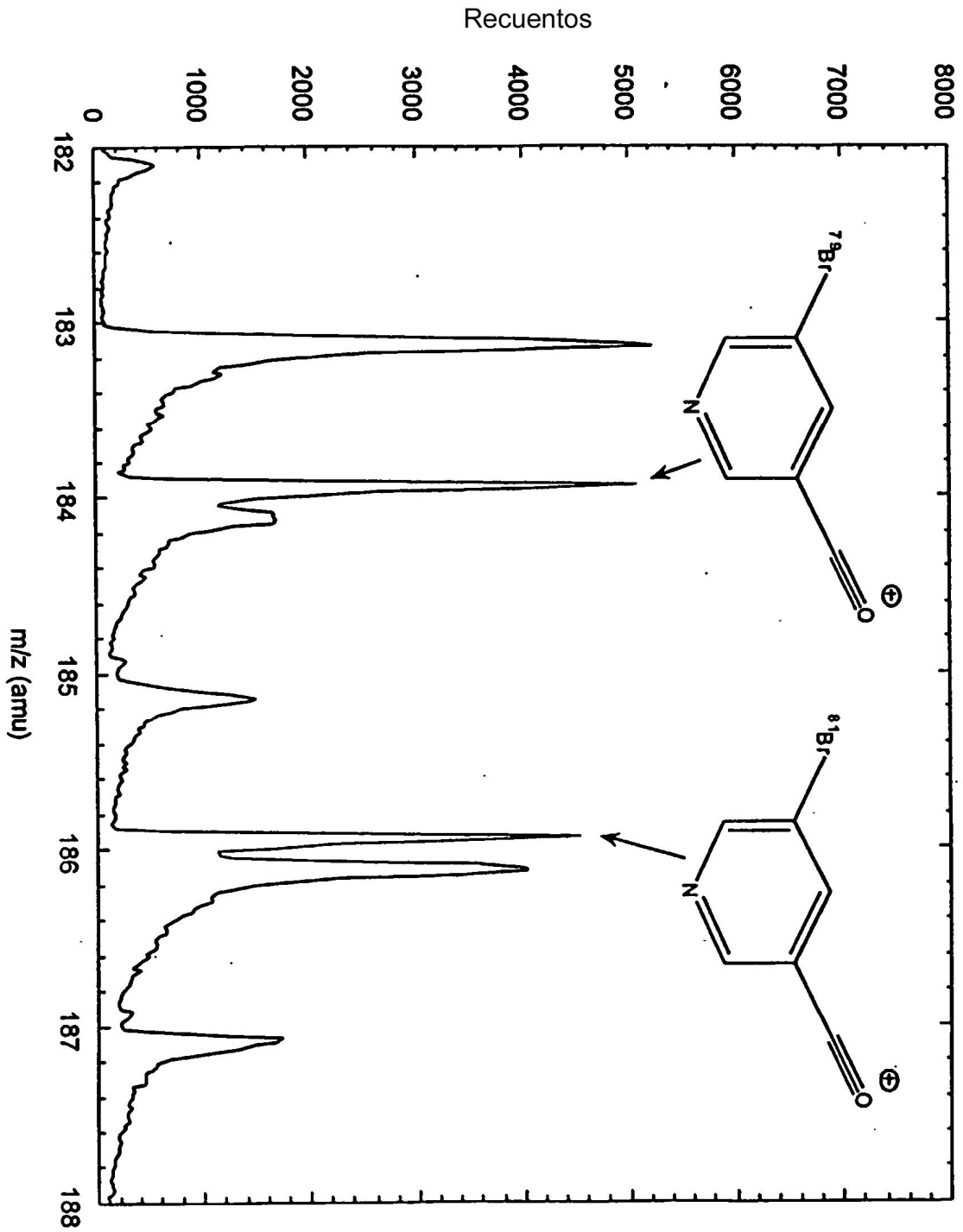


Figura 12B

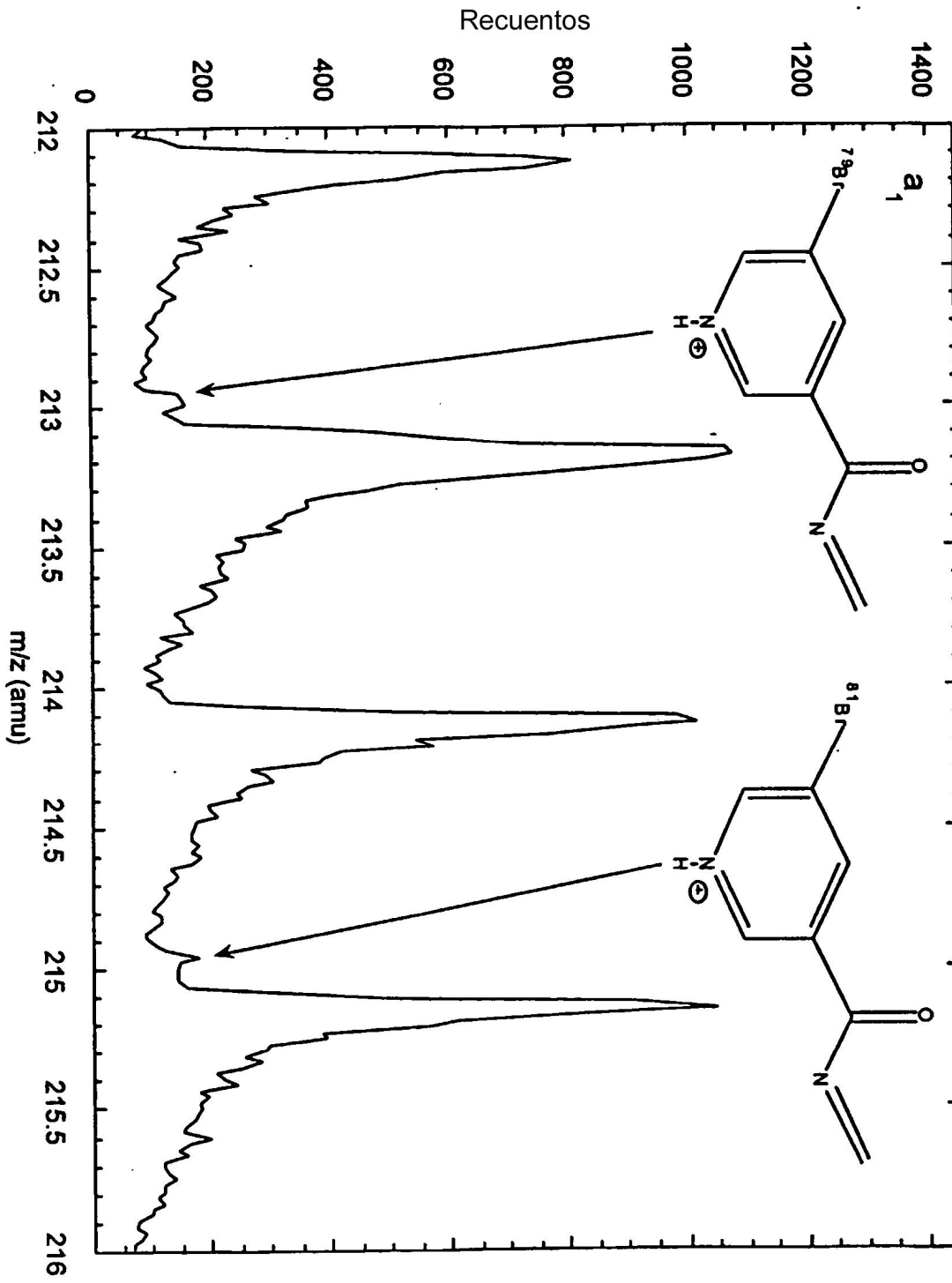


Figura 12C

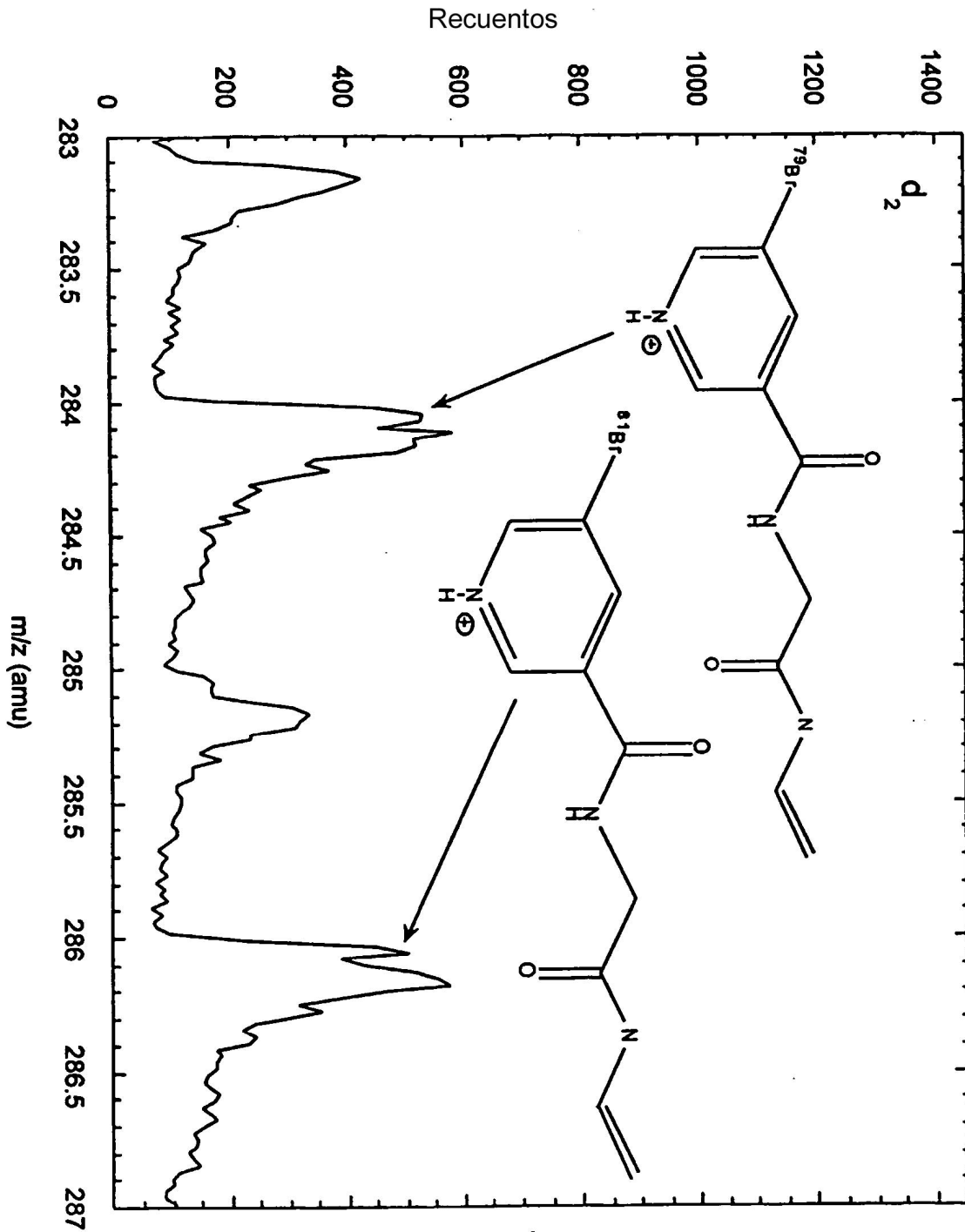


FIGURA 13A

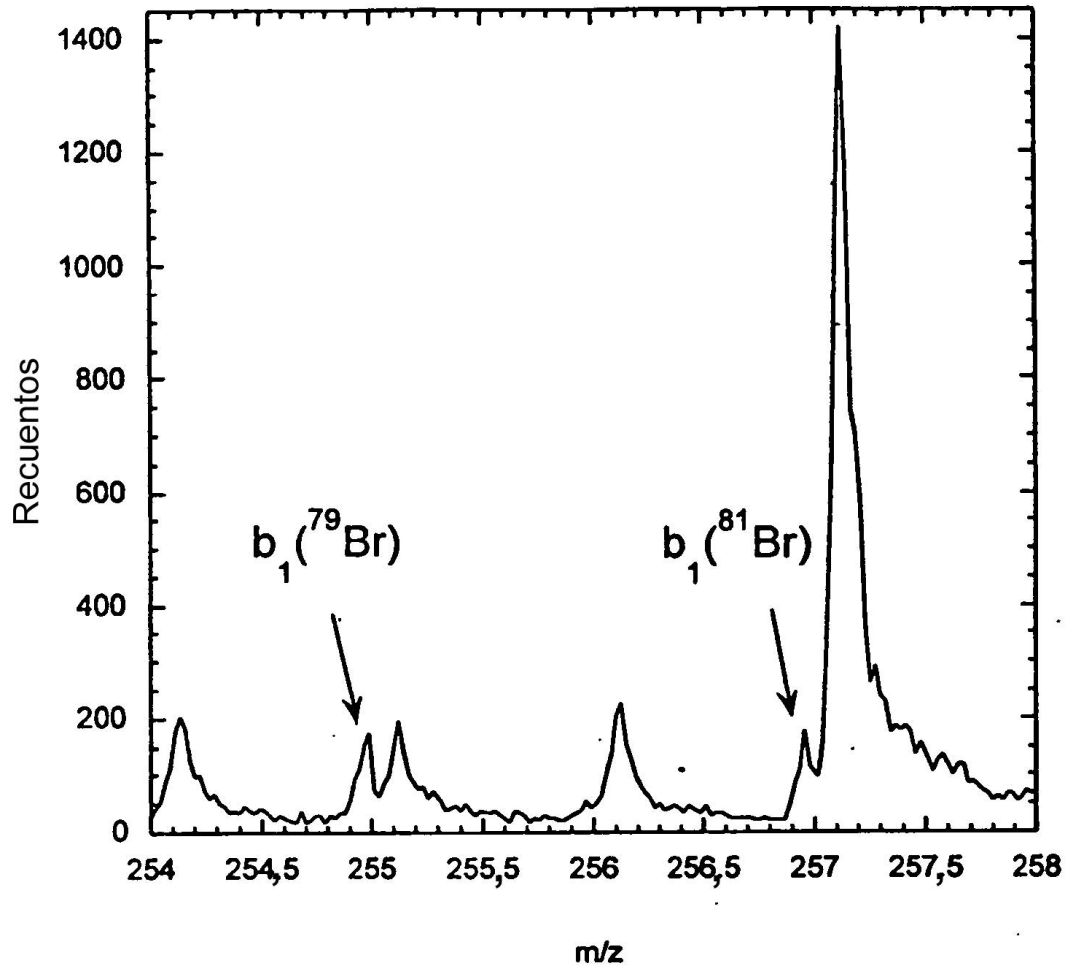


FIGURA 13B

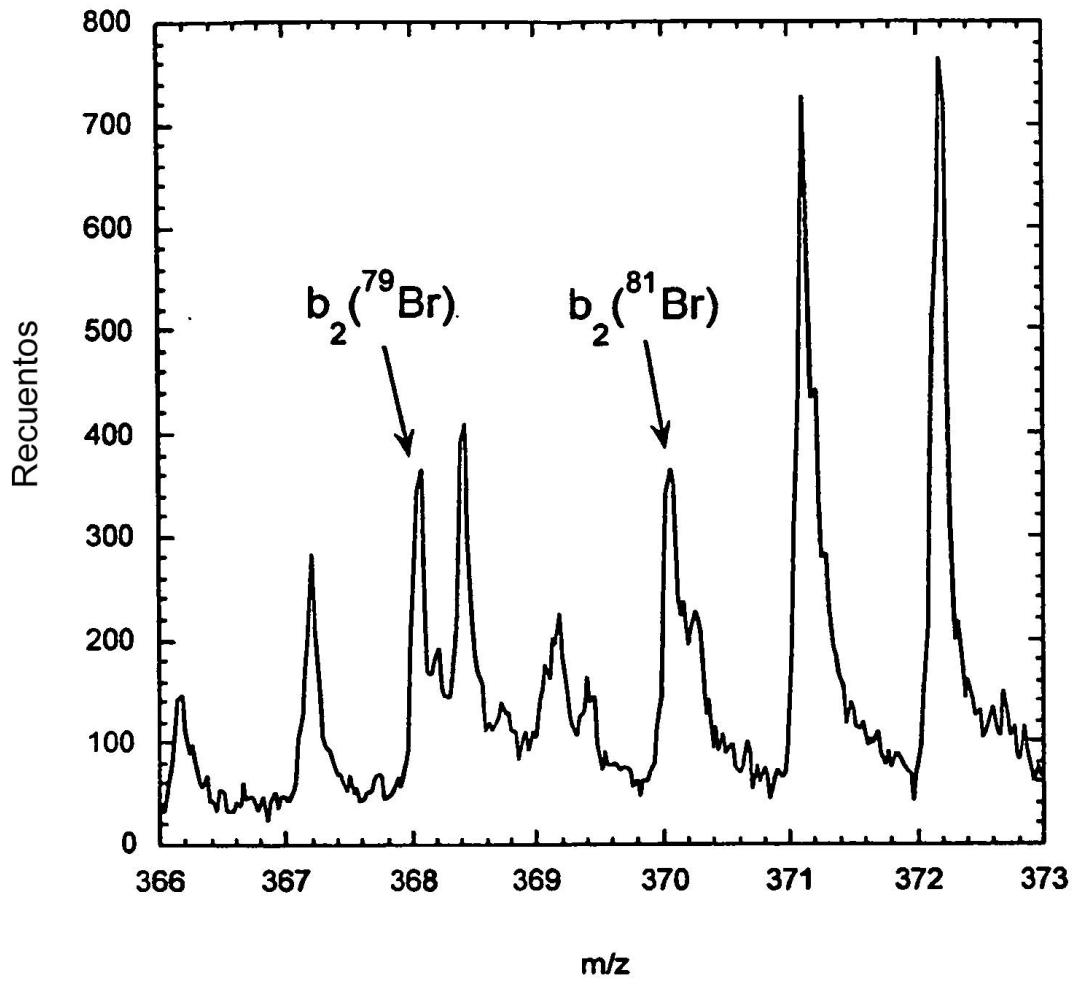


FIGURA 14A

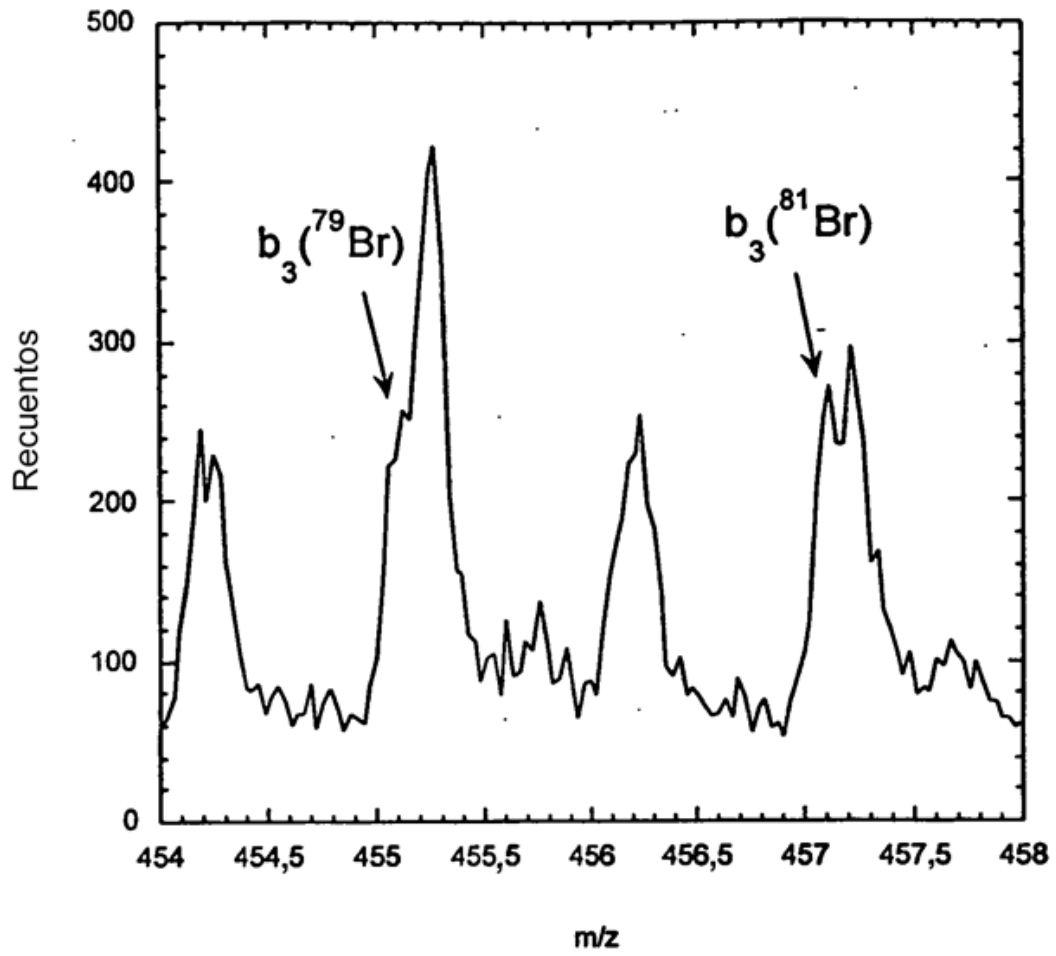


FIGURA 14B

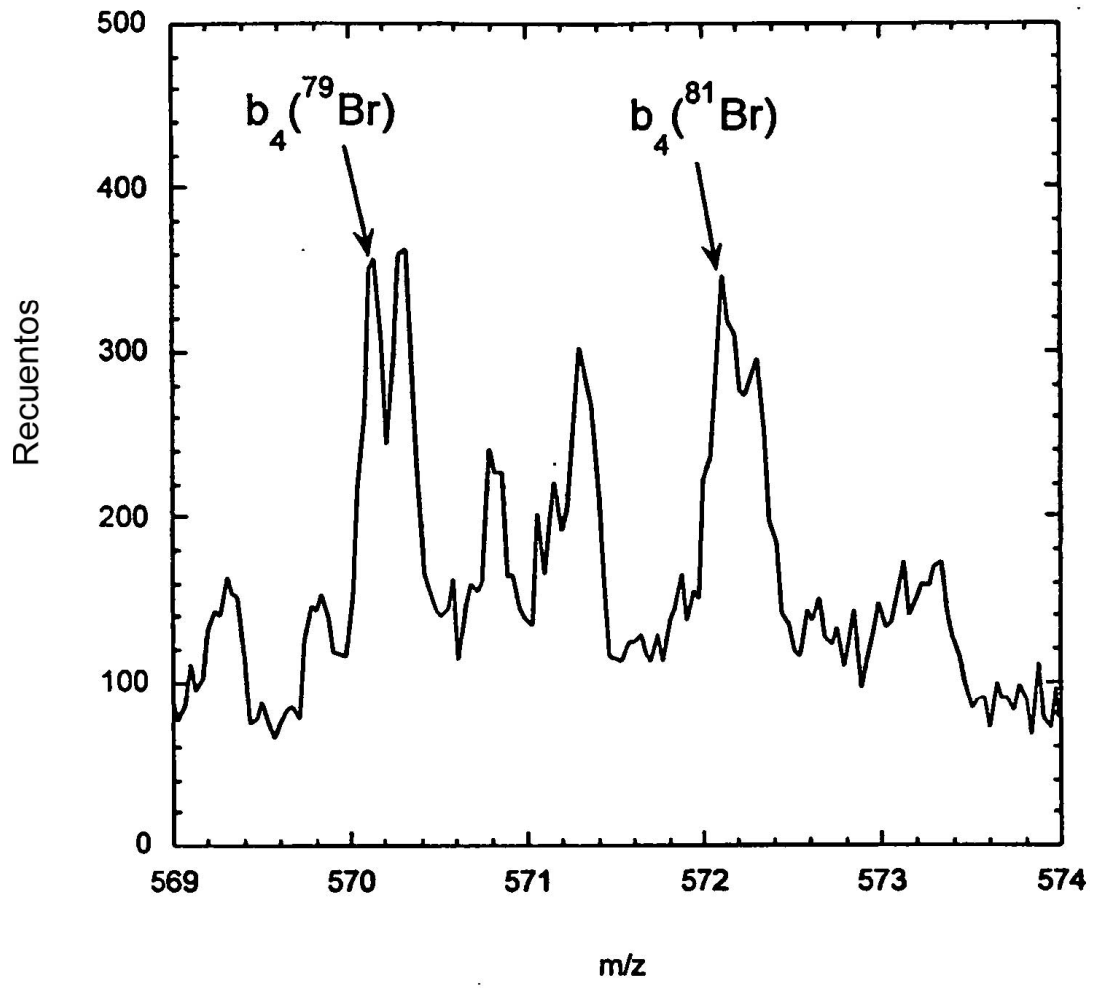


FIGURA 15A

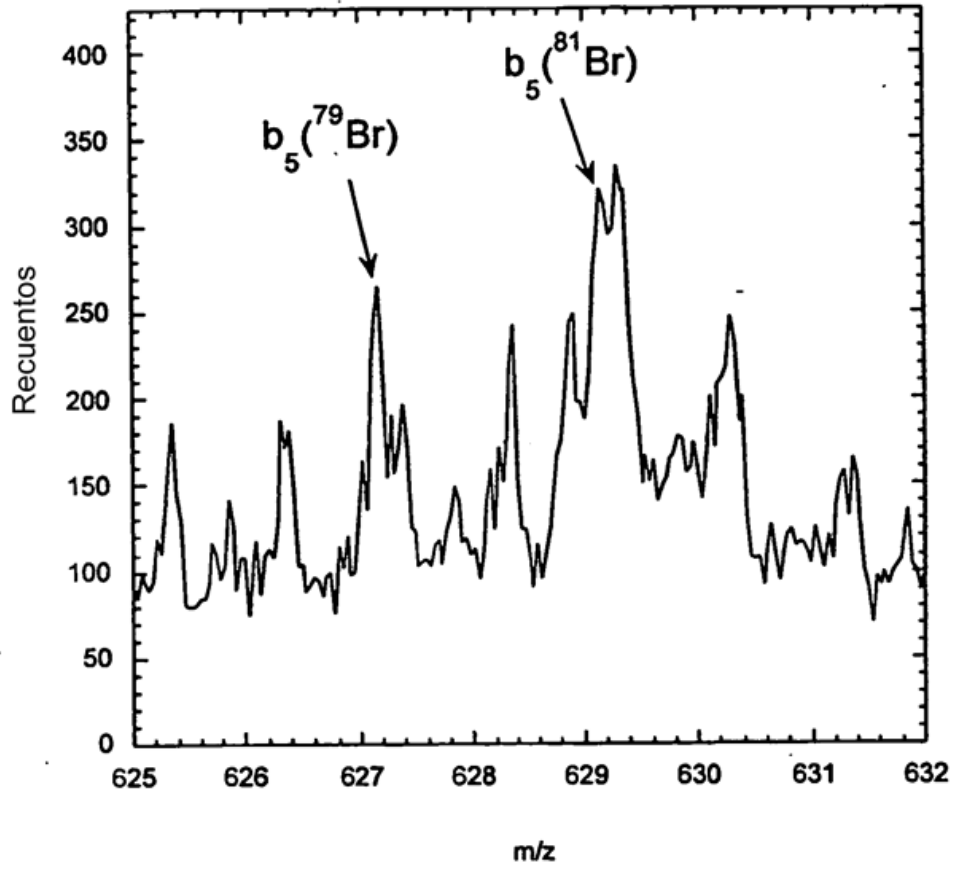




FIGURA 15B

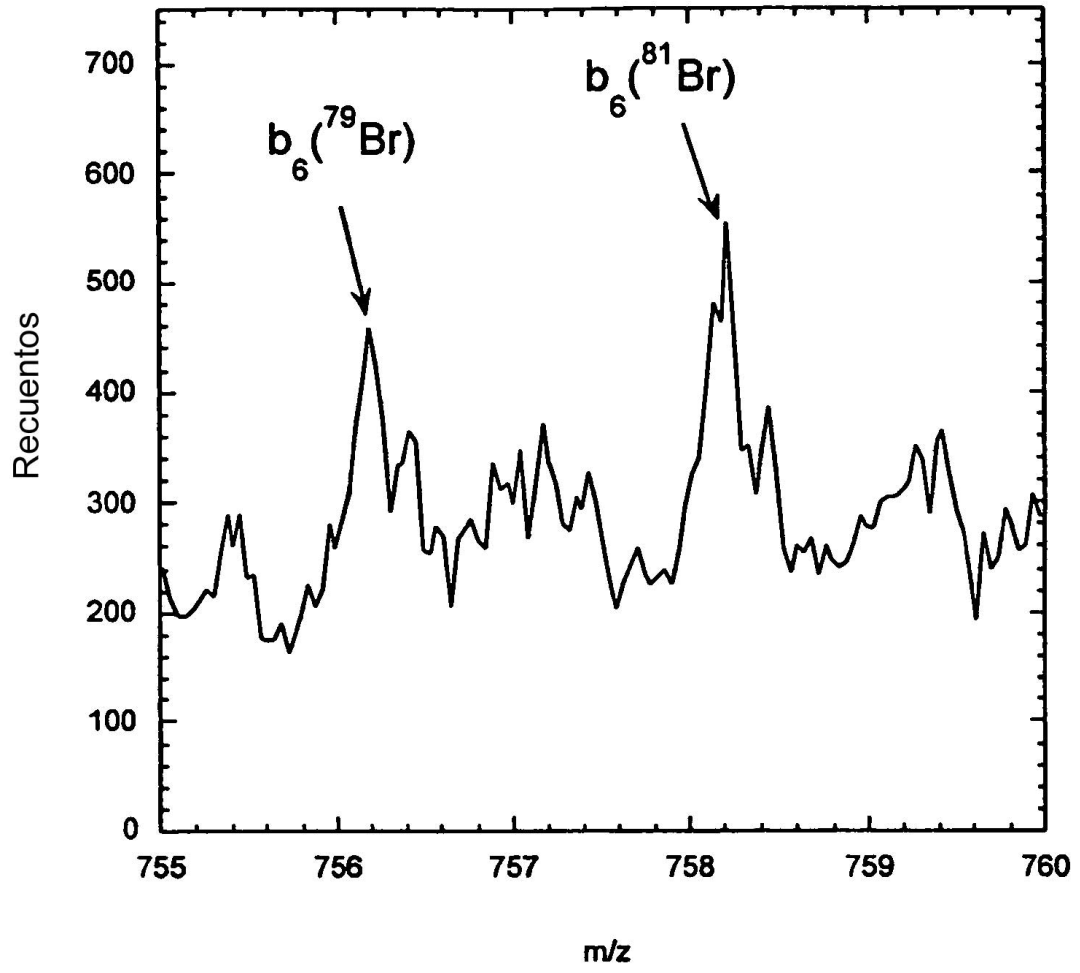


Figura 16

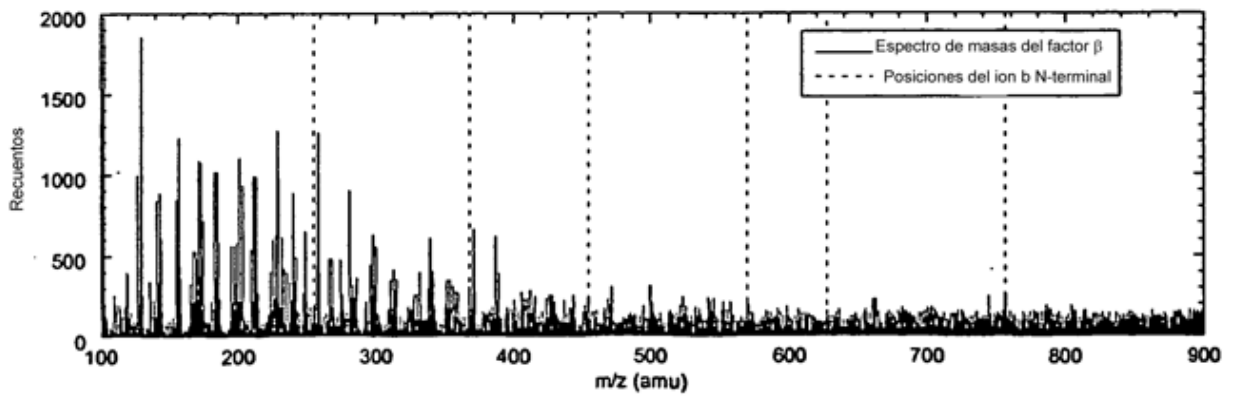
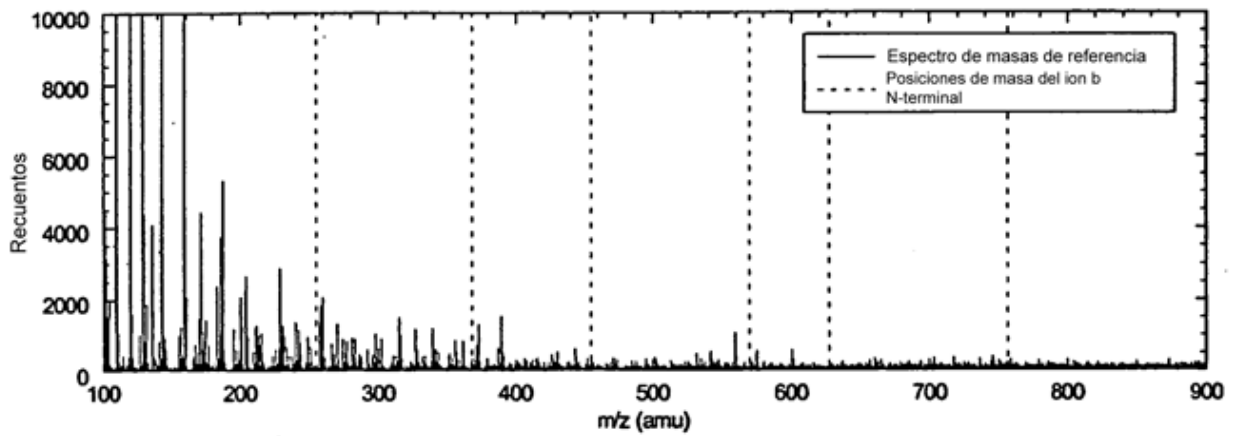
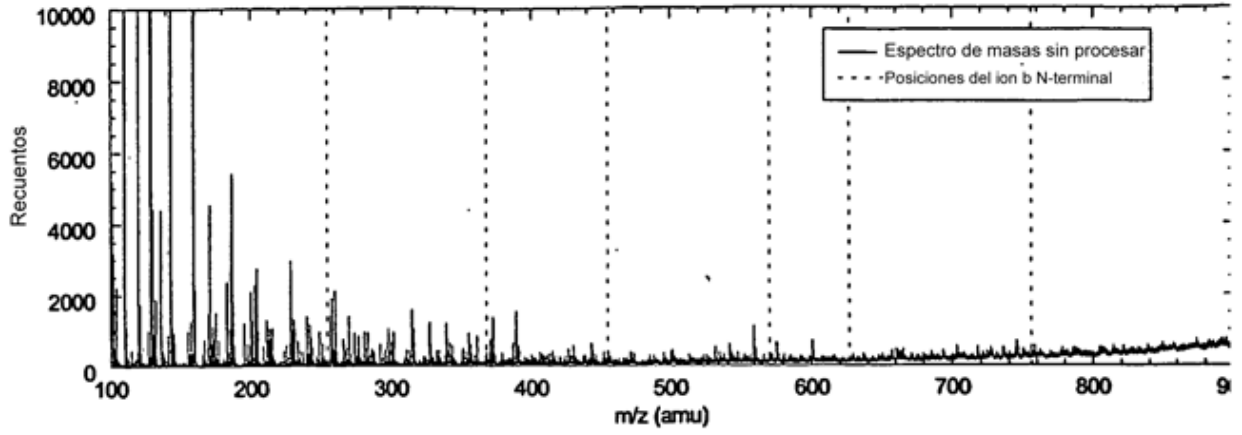


Figura 17

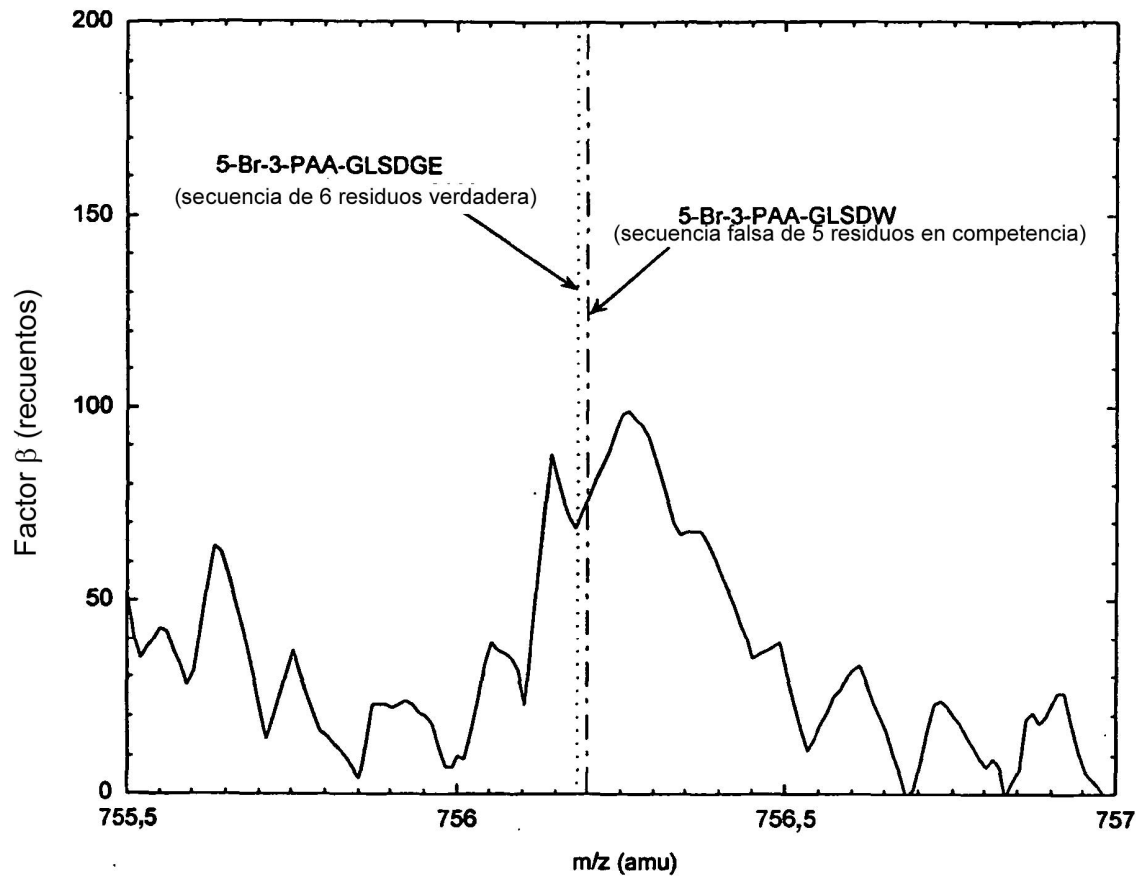


Figura 18A

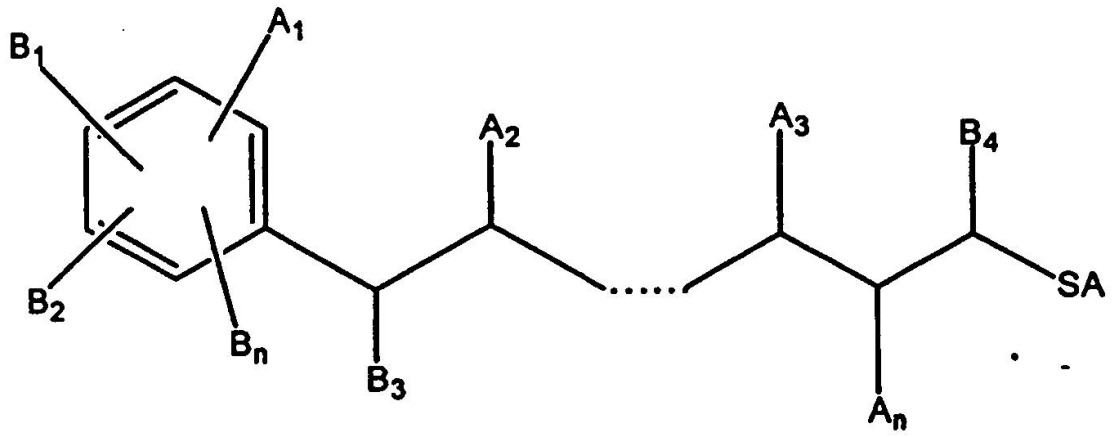


Figura 18B

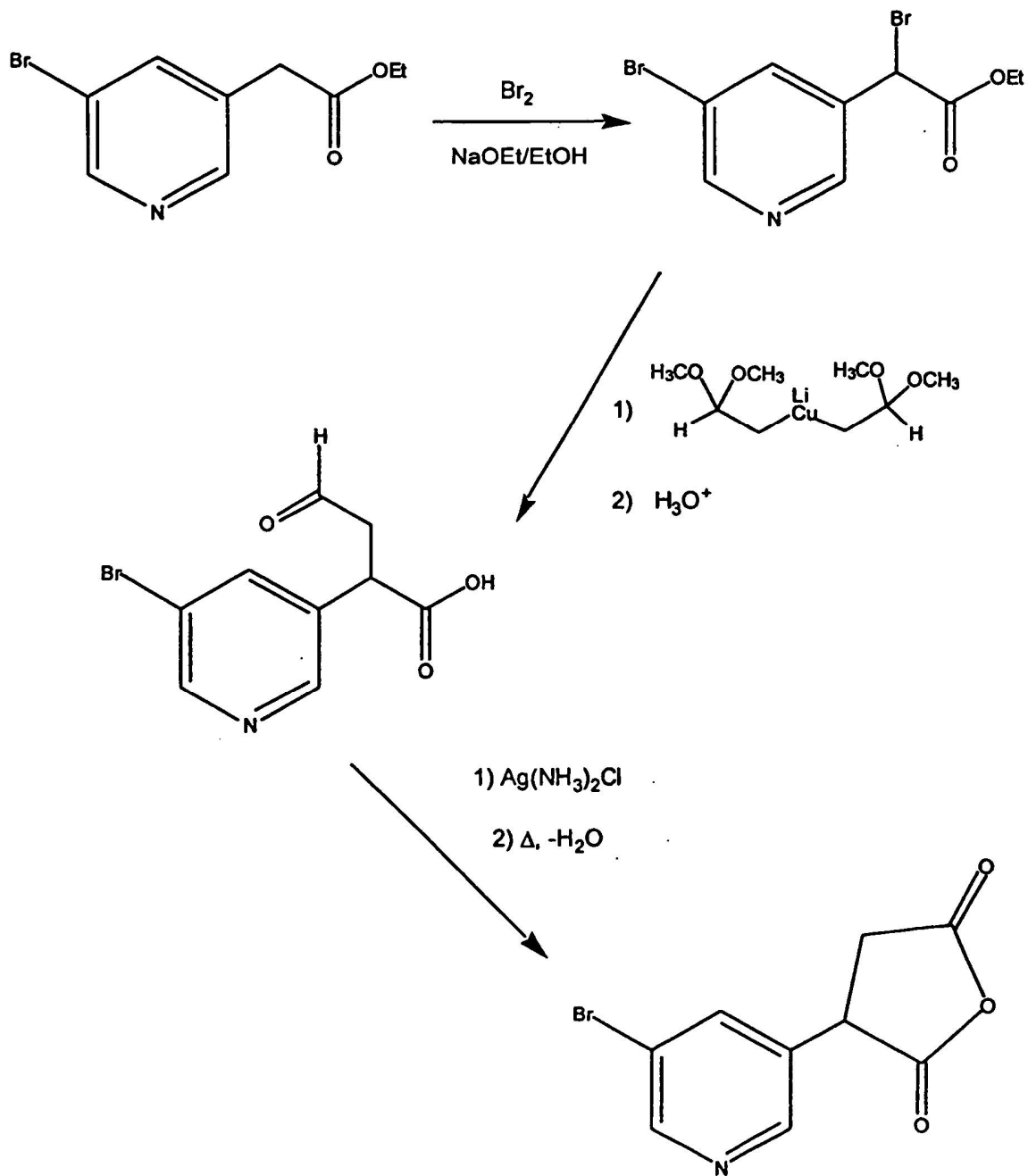


Figura 19

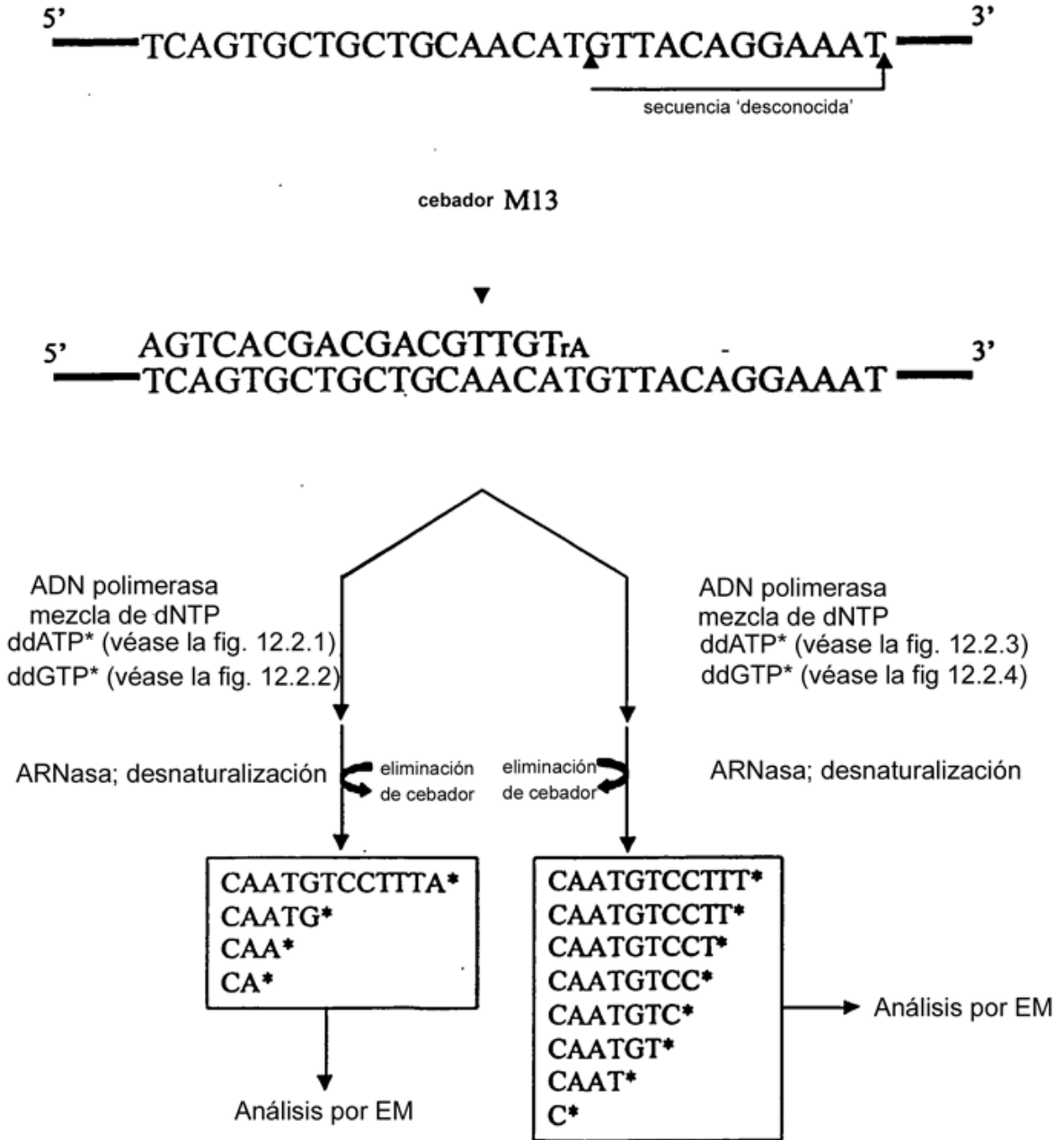


Figura 20A

ddATP\*

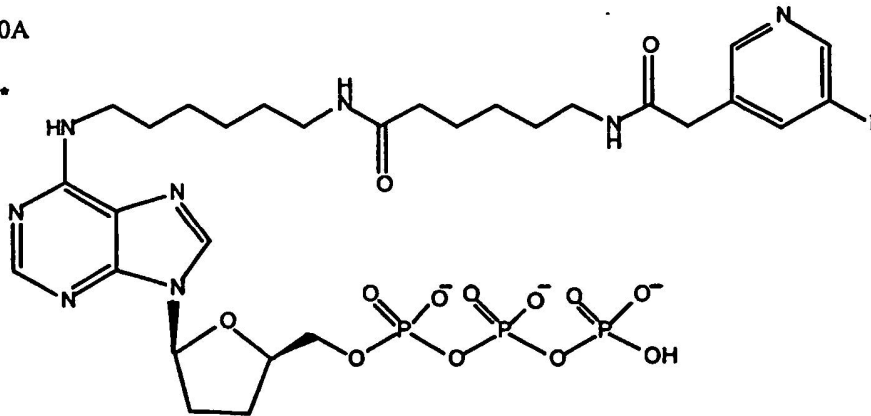


Figura 20B

ddGTP\*

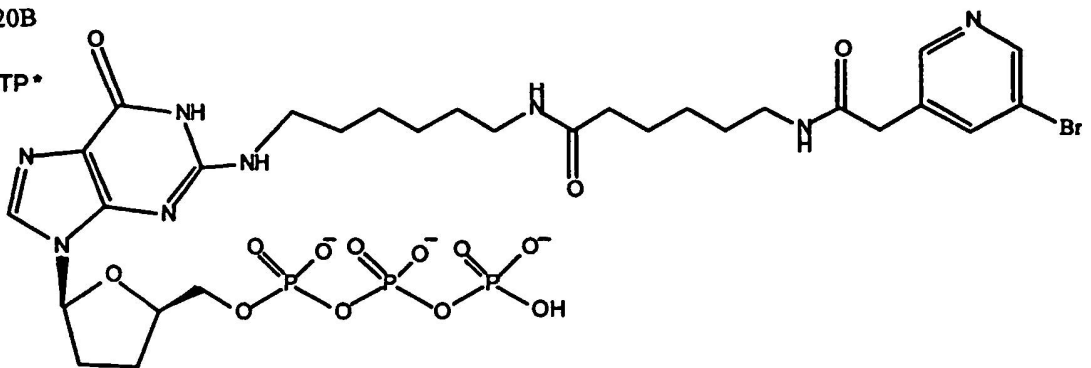


Figura 20C

ddTTP\*

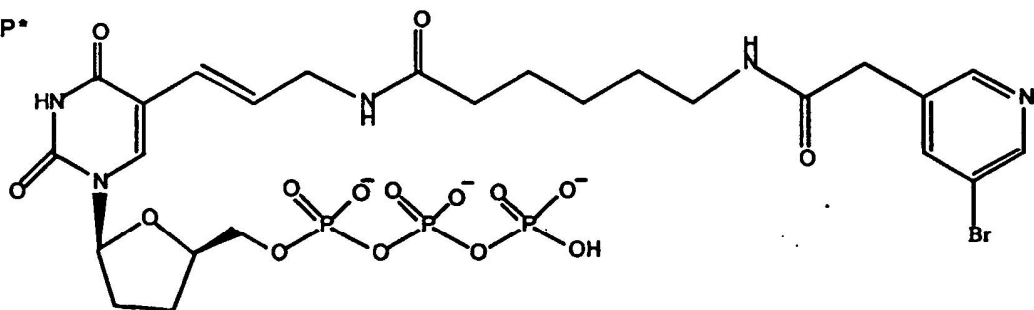


Figura 20D

ddCTP\*

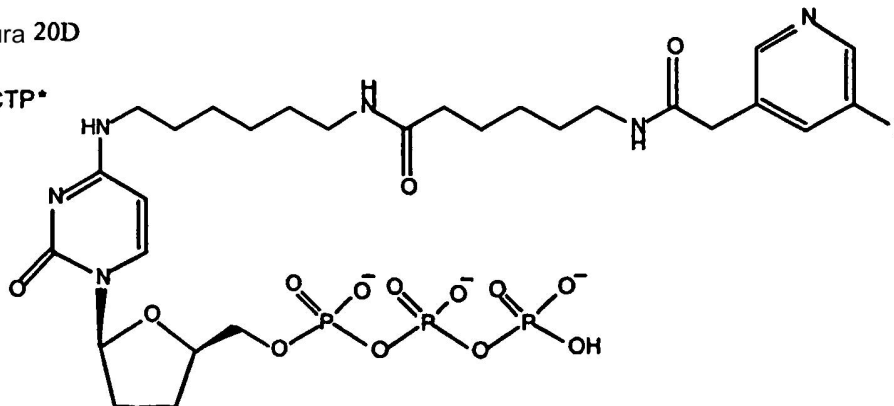


Figura 21

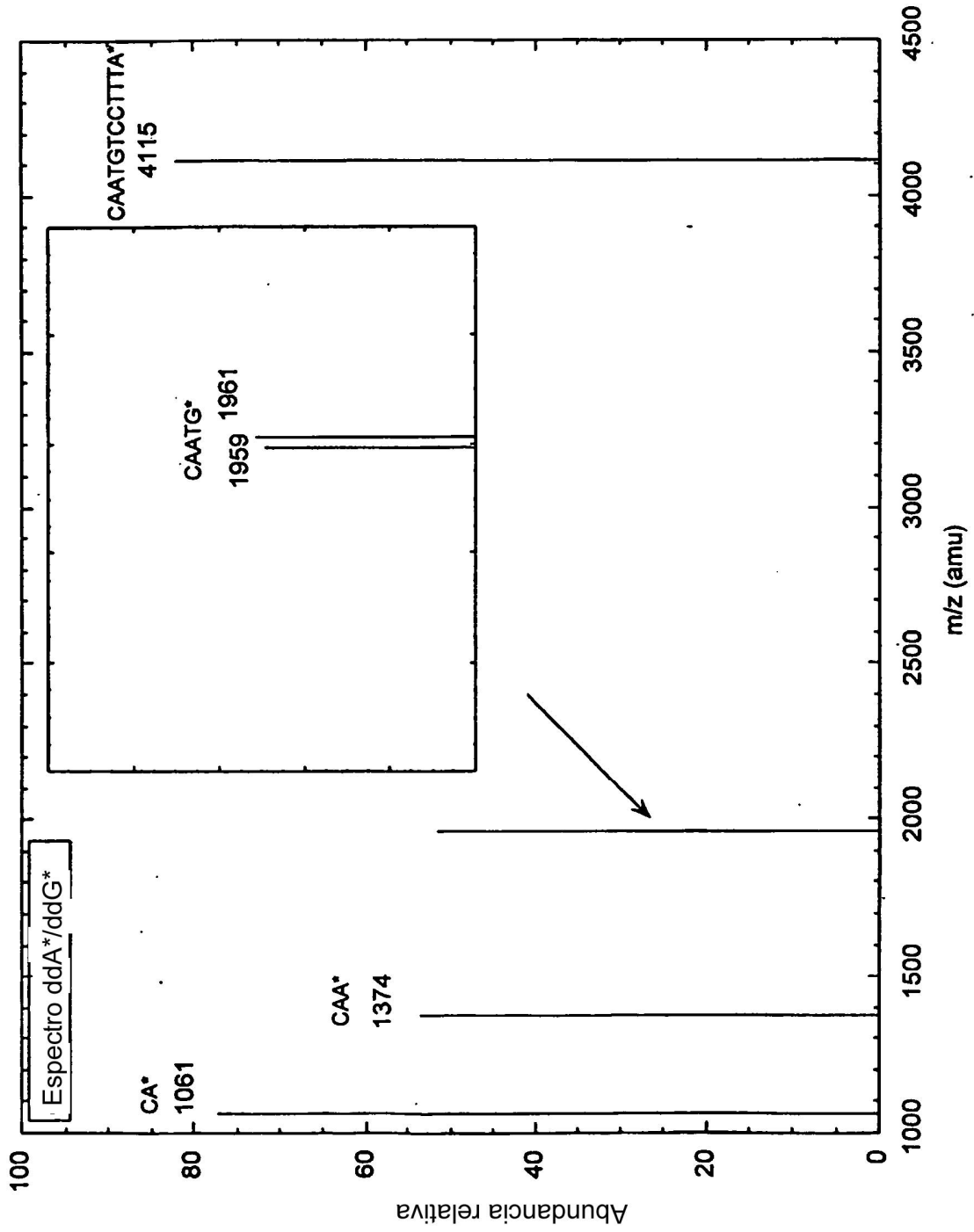




Figura 22

