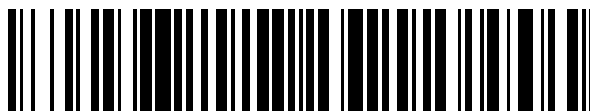


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 448 426**

51 Int. Cl.:

C12N 15/11 (2006.01)

G01N 33/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.10.2008** **E 08837175 (2)**

97 Fecha y número de publicación de la concesión europea: **20.11.2013** **EP 2209893**

54 Título: **Uso de aptámeros en proteómica**

30 Prioridad:

12.10.2007 EP 07020049

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

13.03.2014

73 Titular/es:

**PRNOTA NV (100.0%)
VIB BIO-INCUBATOR TECHNOLOGIEPARK 4
9052 ZWIJNAARDE GHENT, BE**

72 Inventor/es:

**BROWN, CLIVE GAVIN;
KAS, KOEN y
EYCKERMAN, SVEN AGNES JAN**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 448 426 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

Uso de aptámeros en proteómica

La presente invención se refiere a la identificación y cuantificación de proteínas en una mezcla compleja. En particular, la invención se refiere al uso de aptámeros como etiquetas o representantes para proteínas y cantidades de proteínas en una mezcla compleja. En otro aspecto, la invención se refiere al uso de métodos de secuenciación de polinucleótidos de la siguiente generación en proteómica.

El proteoma se describe habitualmente como el complemento completo de proteínas que se encuentra en un sistema biológico tal como, p. ej., una célula, tejido, fluido corporal, órgano u organismo. El estudio de proteínas que se producen de forma natural se denomina generalmente 'proteómica' y abarca el estudio del proteoma expresado en instantes particulares y/o bajo condiciones internas o externas de interés. Las estrategias de la proteómica tienen generalmente como objetivo el análisis global del proteoma y requieren que grandes números de proteínas, p. ej., cientos o miles, puedan ser resueltos, identificados y cuantificados de forma rutinaria a partir de una sola muestra o de múltiples muestras.

Entre las promesas de la proteómica se encuentra su capacidad de reconocer nuevos biomarcadores, es decir, proteínas en calidad de indicadores biológicos que señalizan un estado fisiológico alterado, por ejemplo debido a una enfermedad o intervención terapéutica. El descubrimiento de biomarcadores implica habitualmente comparar proteomas expresados en distintos estados fisiológicos e identificar proteínas cuyos niveles de aparición o expresión difieran de forma consistente entre los estados fisiológicos (Schrattenholz A, Groebe K. Electrophoresis. (2007) Junio 28(12) 1970-9).

Las proteínas en la sangre son una diana particular para la identificación de marcadores de estados patológicos y tratamientos farmacológicos. Se asumen ampliamente que las cantidades y/o la conformación de proteínas en la sangre debería estar estadísticamente relacionada con estos estados de una manera que sopesa la variabilidad natural intrínseca. La sangre y otros fluidos corporales son una diana particular, ya que bañan a los tejidos afectados, transportan proteínas vitales y pueden obtenerse para el ensayo utilizando procesos relativamente económicos y directos durante una consulta médica.

Sin embargo, las proteínas en la sangre tienen un intervalo muy amplio de concentraciones, contabilizando un número pequeño de proteínas más del 99,9% de todas las proteínas y ocupando el resto una distribución desde picogramo a miligramo por mililitro (Qian W.J. et al., Mol. Cell. Prot. (2006) 5(10) 1727-1744). Debido a las limitaciones de las técnicas proteómicas existentes, este intervalo de abundancia sigue siendo una hipótesis. Los científicos de la proteómica han empleado una diversidad de métodos para alcanzar este intervalo, con el objetivo de minimizar también la alteración de las abundancias relativas de las proteínas. A menudo, esto requiere la exclusión de la elevada abundancia de proteínas mediante purificación selectiva. También se han realizado intentos para reducir la complejidad de los péptidos obtenidos, centrándose en subconjuntos de todos los péptidos en la muestra. Estos procesos son largos y todavía ha de demostrarse la reproducibilidad entre muestras, réplicas, máquinas y laboratorios de una manera que sea un requisito previo para el descubrimiento estadístico de proteínas biomarcadoras.

La proteómica basada en la espectrometría de masas (MS) convencional, habitualmente utilizada en el descubrimiento de biomarcadores, procede separando muestra biológicas para aislar proteínas individuales de la mezcla sometida a investigación. Más recientemente, esto ha avanzado de geles en 2D a cromatografía líquida de alta resolución (HPLC) basada en columna multidimensional. Las proteínas pueden disgregarse en subunidades más cortas o péptidos. Los péptidos aislados son luego alimentados a un espectrómetro de masas, el cual ioniza los péptidos y los disgrega adicionalmente, proporcionando una escalera de mediciones de masa/carga. Estas mediciones y sus abundancias también se pueden cuantificar bajo una diversidad de esquemas, habitualmente relacionados con un cierto control. La escalera de espectros resultante es interpretada después frente a secuencias de péptidos conocidas o en ciego a partir de datos en bruto, y la información sobre la masa y las secuencias obtenidas se utiliza para la búsqueda de bases de datos de secuencias para identificar las proteínas a partir de las cuales se originaron los péptidos respectivos.

Sin embargo, la proteólisis de muestras biológicas complejas produce habitualmente cientos de miles de péptidos que pueden superar la capacidad de resolución de sistemas cromatográficos y de MS conocidos, provocando una resolución incompleta y una identificación perjudicada de los péptidos constituyentes. Típicamente, en proteómicas basadas en MS, tanto como un 80% de los espectros derivados de la muestra no pueden ser re-interpretados de forma fiable o consistente en péptidos discriminatorios o, por lo tanto, proteínas. Su comportamiento en la fragmentación y abundancia en el proceso MS también puede depender del contexto, complicando adicionalmente la reproducibilidad (Liu H, Sadygov RG, Yates JR 3º, Anal. Chem. (2004) Jul 15 76(14) 4193-201).

Un método para permitir el análisis proteómico de muestras biológicas consiste en reducir la complejidad de mezclas de péptidos generadas por la separación de este tipo de muestras, antes de someter dichas mezclas de péptidos a etapas de resolución e identificación aguas abajo tales como separación cromatográfica y/o espectrometría de masas (MS). De manera ideal, la reducción de la complejidad de las mezclas de proteínas y péptidos disminuirá el número medio de péptidos distintos presentes por proteína individual de la muestra, pero maximizará la fracción de proteínas de la muestra realmente representada en la mezcla de péptidos.

El uso de sangre (suero o plasma) se ve oscurecido adicionalmente por el procesamiento biológico de proteínas de una diversidad de modos que confunden la verificación basada en MS (Qian W.J. et al., *Mol. Cell. Prot.* (2006) 5(10) 1727-1744). Estudios recientes han mostrado intentos incontroladamente contradictorios de identificar y recontar proteínas a partir de muestras clínicas. El propio acto de aislamiento, fragmentación y medición de proteínas en la proteómica basada en MS convencional altera su abundancia relativa y constitución química.

El resultado del análisis basado en proteómica es una "lista de éxitos" de proteínas que están significativamente correlacionadas en las muestras sometidas a ensayo. Típicamente, la lista es una selección de proteínas de importancia estadística variable. Habitualmente se realiza un estudio biológicamente orientado de la lista y se hace una elección racional sobre la importancia biológica potencial de cada uno de los miembros. El proceso de alcanzar una lista de biomarcadores hipotéticos o de proteínas de importancia supuestamente estadística se denomina generalmente "Descubrimiento".

Los esfuerzos están entonces enfocados hacia la verificación y validación de un pequeño número de "éxitos" elegidos. Esto implica la confirmación de las abundancias de proteínas medidas en una población más amplia de muestras clínicas, siendo el objetivo demostrar que las proteínas descubiertas y elegidas son genuinas y no falsas positivas y que son específicas para la enfermedad o estado médico. A este proceso de confirmar la importancia cuantitativa de una medición proteómica y de un supuesto biomarcador en una población más generalizada se le alude habitualmente como 'Validación' (Zolg, *Mol. Cell. Prot.* (2006) 5(10), 1720-1726). A menudo esto requiere el uso de una tecnología alternativa a la fase de descubrimiento.

Típicamente, esto se consigue utilizando anticuerpos dirigidos contra las supuestas proteínas purificadas. Estos anticuerpos pueden luego utilizarse en un ensayo basado en ELISA contra cientos de muestras, re-aplicarse la estadística relevante y establecerse la validez de la proteína en calidad de biomarcador. La intención final de este proceso es también que el anticuerpo se vuelva parte de un análisis clínico. Sin embargo, la producción de los anticuerpos lleva mucho tiempo y es costosa, y no siempre es posible dirigir un anticuerpo específico a una proteína dada. Una complicación adicional es la presencia de variantes alternativas de la proteína o isoformas que puedan tener incrustaciones de moléculas de azúcar fijadas y otras modificaciones sobre su superficie. Las isoformas pueden no ser evidentes en la fase de Descubrimiento de la identificación basada en MS y, por tanto, pueden complicar la importancia estadística de las mediciones, así como el desarrollo de anticuerpos específicos (Zolg, *Mol. Cell. Prot.* (2006) 5(10), 1720-1726; Rifai et al., *Nature Biotech* (2006) 24(8) 971-83).

Tanto el Descubrimiento como la Validación tiene altas tasas de fracaso y son ejercicios que llevan mucho tiempo y costosos. Habitualmente, el éxito y el fracaso sólo se pueden juzgar después de la fase de Validación después de mucho tiempo y de un coste financiero. Por lo tanto, los esfuerzos recientes se han dirigido al refinamiento de estrategias basadas en MS para el descubrimiento y sobre la generación a granel de anticuerpos en un esfuerzo de acelerar tanto el Descubrimiento y la Validación como de derivar un producto clínico (Anderson y Hunter, *Mol. Cell Prot.* (2006) 5(4) 573-588; Zangar et al., *Expert Review of Proteomics* (2006) 3(1) 37-44).

Por consiguiente, existe una clara necesidad de nuevos métodos para interrogar proteínas, particularmente mezclas complejas de proteínas, derivadas de fluidos corporales. Específicamente, existe la necesidad de una tecnología para identificar y medir la abundancia de proteínas, especialmente la abundancia de múltiples proteínas en una mezcla compleja, que supere los inconvenientes técnicos y las limitaciones científicas y de costes que obstaculiza la actual tecnología. En particular, se necesitan herramientas para reemplazar la generación costosa y a veces poco fiable y el uso de anticuerpos para identificar y validar proteínas diana. Este tipo de herramientas también se puede utilizar para suplantar las tecnologías basadas en Descubrimiento existentes tal como MS, con la condición de que se pueda generar rápidamente y de forma suficientemente económica una diversidad suficiente de alternativas de anticuerpos para sondear una gama completa de proteínas nativas en una muestra biológica.

Una aspiración clara para una herramienta de este tipo es una manipulación mínima de la mezcla, de modo que en la muestra se pueda obtener una representación verdadera de la proteína en su configuración nativa, junto con cualquier modificación y variaciones naturales. Una herramienta de este tipo también debería ser altamente reproducible y de forma fiable.

Los aptámeros son polímeros cortos, habitualmente ácidos nucleicos (ADN, ARN, APN) que constituyen formas tridimensionales bien definidas, permitiéndoles unir moléculas diana de una manera que sea conceptualmente

similar a anticuerpos. Los aptámeros combinan las características óptimas de moléculas pequeñas y anticuerpos, incluidas la alta especificidad y afinidad, estabilidad química, baja inmunogenicidad y la capacidad de fijar como objetivo interacciones proteína-proteína. Además de una alta especificidad, los aptámeros tienen afinidades muy elevadas para sus dianas. Típicamente, los aptámeros generados contra proteínas tienen afinidades en el intervalo picomolar a nanomolar bajo. En contraposición a anticuerpos monoclonales, los aptámeros son sintetizados químicamente más que expresados biológicamente, ofreciendo una importante ventaja de costes.

Mientras que los aptámeros proporcionan una alternativa útil y eficaz frente a los anticuerpos, sigue existiendo el problema de cómo cuantificar proteínas a través de aptámeros. Un problema particular lo suponen mezclas complejas de proteínas y la amplia gama dinámica de proteínas presentes en muestras biológicas. Típicamente, los aptámeros se cuantifican utilizando radiomarcadores. Aun cuando es aceptable la cuantificación de radiomarcadores, y lo ha sido durante un cierto número de años, sigue existiendo la necesidad de mejorar y acelerar la cuantificación, así como de mejorar la precisión y reproducibilidad.

Por tanto, en su sentido más amplio, la presente invención se refiere a un método para medir la cantidad de al menos una molécula en una muestra biológica, comprendiendo el método:

- a) combinar la muestra con uno o más aptámeros y permitir que una o más moléculas en la muestra se unan al o a los aptámeros;
- b) separar las moléculas unidas de las no unidas; y
- c) cuantificar la o las moléculas unidas al o a cada uno de los aptámeros,

en donde la cuantificación de la o las moléculas unidas se lleva a cabo secuenciando al menos parte del o de cada uno de los aptámeros.

Expresado de otra manera, la invención reside en el uso de un aptámero en calidad de una etiqueta para cuantificar una proteína, en donde la secuencia del aptámero es la etiqueta. Se ha apreciado que un aptámero tiene una única secuencia que puede ser leída, por ejemplo, tal como un código de barras. De acuerdo con la presente invención, este único código de barras es el que es la etiqueta para cada una de las proteínas a las que se une el código de barras (aptámero). Al leer el código de barras (secuenciar el aptámero), se obvia la necesidad de etiquetas y/o marcadores adicionales.

En contraposición con los métodos existentes, el uso de la secuencia del aptámero como una etiqueta permite la identificación y cuantificación de una proteína utilizando sólo una única entidad molecular. De hecho, la secuenciación de un aptámero permite el recuento de los casos o apariciones de cada una de las distintas secuencias y, así, permite la cuantificación directa de la proteína a la que se ha unido cada uno de los aptámeros. En otras palabras, cada uno de los aptámeros es una representación directa de una proteína dentro de una muestra o mezcla de proteínas cuando se une un aptámero individual y una proteína individual. Se apreciará que, por medio del método de la presente invención, se minimiza la manipulación de la proteína o muestra original, permitiendo con ello la interrogación de una proteína o mezcla de proteínas en un estado lo más natural posible. De este modo se puede obtener una imagen verdadera de una proteína o población de proteínas.

En una realización preferida, la secuenciación de los aptámeros se consigue sin amplificación, así un recuento absoluto o digital de los aptámeros contando la aparición de cada una de las especies en un dispositivo de secuenciación que separa cada una de las moléculas individuales antes de la secuenciación, por ejemplo un secuenciador de ADN de la siguiente generación. Se obtiene un recuento directo de la aparición de cada una de las secuencias, lo cual, a su vez, se iguala a un recuento absoluto de las proteínas a las que se ha unido el aptámero. Se puede representar un subconjunto de las secuencias en el dispositivo de recuento, lo cual sería una muestra representativa de la aparición de las secuencias dadas en la población de moléculas sometida a estudio. De nuevo, esto está en contraposición con los actuales métodos de cuantificación de aptámeros, los cuales requieren típicamente una etiqueta unida o una amplificación de la molécula del aptámero, proporcionando una estimación análoga de su abundancia relativa (tal como se describe en la Publicación de Patente Nº US 2007-0166741). Esto se debe a las limitaciones de utilizar marcadores y etiquetas adicionales mediante los cuales se identifican los aptámeros que tiene proteínas unidas. Como resultado, la cuantificación es proporcional a la abundancia real, es decir, no se pueden resolver una señal análoga ni apariciones absolutas o individuales.

Una ventaja adicional de utilizar la secuencia del aptámero como etiqueta es que las secuencias proporcionan una enorme gama de etiquetas posibles. Por ejemplo, una secuencia de cuatro bases ofrece 256 combinaciones diferentes y, por tanto, 256 identificadores o etiquetas únicos posibles. Así, aun cuando se puede obtener la secuencia completa del aptámero, el número de nucleótidos secuenciados necesita sólo ser el suficiente para

identificar un aptámero particular. En otras palabras, toda la secuencia del aptámero es la etiqueta, o la etiqueta es parte de la secuencia del aptámero.

A pesar de que el método de la presente invención se puede utilizar para cuantificar una sola proteína en una muestra con una secuencia del aptámero, una ventaja adicional de la presente invención es que la gama de posibles etiquetas proporcionadas por secuencias únicas permite cuantificar muchas proteínas al mismo tiempo y en la misma muestra. Así, se puede utilizar un panel o banco de secuencias de aptámeros conocidas para cuantificar múltiples proteínas en una mezcla compleja y, de este modo, la invención permite la interrogación de muestras biológicas complejas. Además de ello, puede dirigirse un panel o banco de aptámeros a proteínas específicas, en particular proteínas que se sabe están presentes en una muestra en bajas cantidades. Ya que es raramente posible identificar y cuantificar proteínas en baja abundancia en una muestra utilizando métodos conocidos, la presente invención proporciona una solución para interrogar estas pequeñas y, hasta ahora, elusivas poblaciones. El banco o panel debería contener, de manera ideal, un exceso de especies de aptámeros para el sondeo de muestras.

La invención puede analizar una sola proteína (p. ej., proteína escindida en gel), y es particularmente adecuada para analizar mezclas de proteínas, incluidas mezclas de proteínas complejas. La expresión “mezcla de proteínas” se refiere generalmente a una mezcla de dos o más proteínas diferentes, p. ej., una composición que comprende dos o más proteínas diferentes.

En realizaciones preferidas, una mezcla de proteínas a analizar puede incluir más de aproximadamente 10, preferiblemente más de aproximadamente 50, incluso más preferiblemente más de aproximadamente 100, todavía más preferiblemente más de aproximadamente 500 proteínas diferentes tal como, p. ej., más de aproximadamente 1000 o más de aproximadamente 5000 proteínas diferentes. Una mezcla de proteínas compleja ilustrativa puede implicar, sin limitación, todas o una fracción de proteínas presentes en una muestra biológica o parte de la misma.

La expresión “muestra biológica” o el término “muestra”, tal como se utiliza en esta memoria, se refiere generalmente a material, en una forma no purificada o purificada, obtenido a partir de una fuente biológica. A modo de ejemplo y no de limitación, las muestras se pueden obtener de: virus de hospedantes procarióticos o eucarióticos; células procarióticas, p. ej., bacterias o arqueas, p. ej. procariotas que viven libremente o colonias o bio-películas que comprenden procariotas; células eucarióticas u organelas de las mismas, incluidas células eucarióticas obtenidas in vivo o in situ o cultivadas in vitro; tejidos u organismos eucarióticos, p. ej. muestras con contenido en células o exentas de células de tejidos u organismos eucarióticos; los eucariotas pueden comprender protistas, p. ej. protozoos o algas, hongos, p. ej. levaduras o mohos, plantas y animales, p. ej. mamíferos, seres humanos o mamíferos no humanos. Por lo tanto, “muestra biológica” puede abarcar, por ejemplo, una célula, tejido, organismo o extractos de los mismos. Una muestra biológica puede separarse preferiblemente de su fuente biológica, p. ej. de un animal tal como un mamífero, ser humano o mamífero no humano, por métodos adecuados tales como, sin limitación, recogida o extracción de orina, saliva, esputo, semen, leche, mucus, sudor, heces, etc., extracción de sangre, fluido cerebro-espinal, fluido intersticial, fluido óptico (vítreo) o fluido sinovial, o mediante biopsia del tejido, resección, etc. Una muestra biológica puede subdividirse, además, para aislar o enriquecer partes de la misma a utilizar para obtener proteínas para análisis en la invención. A modo de ejemplo y no de limitación, se pueden separar uno de otro diversos tipos de tejidos, se pueden aislar de una muestra tipos de células o fenotipos de células específicos, p. ej. utilizando la clasificación FACS, batea de anticuerpos, disección de captura por láser, etc.; las células se pueden separar del fluido intersticial, por ejemplo las células de la sangre se pueden separar del plasma sanguíneo o del suero; o similares. La muestra se puede aplicar al método de la invención directamente o se puede procesar, extraer o purificar en grados variables antes de ser utilizada.

La muestra se puede derivar de un sujeto sano o de un sujeto que padece una afección, trastorno, enfermedad o infección. Por ejemplo, sin limitación, el sujeto puede ser un animal sano, p. ej. un ser humano o un mamífero no humano, o un animal, p. ej. un ser humano o un mamífero no humano que tiene cáncer, una enfermedad inflamatoria, enfermedad autoinmune, enfermedad metabólica, enfermedad del SNC, enfermedad ocular, enfermedad cardíaca, enfermedad pulmonar, enfermedad hepática, enfermedad gastrointestinal, enfermedad neurodegenerativa, enfermedad genética, enfermedad infecciosa o infección viral, u otro u otros males.

Preferiblemente, se pueden tratar mezclas de proteínas derivadas de muestras biológicas para agotar de las mismas proteínas altamente abundantes, con el fin de aumentar la sensibilidad y el comportamiento de análisis del proteoma. A modo de ejemplo, muestras de mamíferos tales como muestras de suero o plasma humano pueden incluir abundantes proteínas, entre otras, albúmina, IgG, antitripsina, IgA, transferrina, haptoglobina y fibrinógeno, que pueden preferiblemente ser así agotadas de las muestras. Se conocen métodos y sistemas para la separación de proteínas abundantes tales como, p. ej., agotamiento por inmuno-afinidad y con frecuencia comercialmente disponible, p. ej. Sistema de Eliminación por Afinidad Múltiple (MARS-7, MARS-14 – siglas en inglés) de Agilent

Technologies (Santa Clara, California).

Aun cuando la presente invención tiene aplicación particular en la cuantificación de proteínas, se ha de apreciar que la invención también tiene aplicación para la cuantificación de otras moléculas tales como metabolitos y para la identificación de potenciales productos terapéuticos de moléculas pequeñas y biológicos.

- 5 En una realización, la presente invención utiliza las técnicas de secuenciación de polinucleótidos de siguiente generación para identificar y cuantificar etiquetas de aptámeros. De este modo, la presente invención proporciona un método más sensible y eficaz de cuantificación, en comparación con los métodos actualmente disponibles, y acopla la amplia gama dinámica, especificidad y sensibilidad ofrecida por esta metodología genómica en el sector de la proteómica.
- 10 Técnicas de secuenciación de polinucleótidos de siguiente generación requieren la separación de moléculas individuales de ADN o ARN u otras secuencias de polinucleótidos sobre una superficie tal como una perla o chip, creando con ello una disposición ordenada de secuencias de moléculas sencillas. La disposición ordenada tiene una densidad de superficie que permite resolver individualmente a cada una de las moléculas, por ejemplo mediante microscopía óptica. La secuenciación de las moléculas de polinucleótidos en la disposición ordenada
- 15 permite el recuento "digital", es decir absoluto, de las secuencias y, de esta forma, la cuantificación directa de las secuencias presentes en la disposición ordenada. En algunas tecnologías, las secuencias pueden ser amplificadas clonalmente una vez dispuestas ordenadamente para reforzar y/o clarificar la señal procedente de cada una de las secuencias. No obstante, la cuantificación se obtiene mediante el recuento de la aparición de secuencias en la disposición ordenada y no la señal producida a partir de cada uno de los amplicones. Ejemplos de técnicas de
- 20 secuenciación y cuantificación adecuadas pueden encontrarse en publicaciones tales como el documento WO 00/006770 y Branton et al (Nature Biotechnology (2008) 26, 1146-1153).

Por consiguiente, en una realización preferida, la cuantificación de las secuencias de aptámeros unidos se lleva a cabo utilizando una disposición ordenada de moléculas sencillas de las secuencias unidas. Alternativamente, las secuencias de aptámeros pueden ser cuantificadas sobre una disposición ordenada clonal, en la que cada una de

25 las secuencias se amplifica después de haber sido dispuesta ordenadamente sobre una superficie.

Debido a la alta especificidad de un aptámero para su proteína diana, se puede utilizar un aptámero en calidad de representante para una proteína dentro de una mezcla. Después de la unión a su proteína y de la subsiguiente elución, la cuantificación de un aptámero a través del recuento de la frecuencia de su secuencia en un dispositivo de secuenciación paralelo permitirá la cuantificación de una proteína particular dentro de la muestra. Esto se hará,

30 mediante implicación o representación, de una manera similar al uso de anticuerpos como representantes de proteínas. De hecho, la distribución de las secuencias o etiquetas de aptámeros será representativa de la distribución de las correspondientes proteínas en la población o muestra biológica de la que se obtuvo la proteína.

Además de ello, el recuento de moléculas clonales en disposiciones ordenadas proporciona un número absoluto para la cantidad o proporción de cada una de las etiquetas de aptámeros. A esto se denomina "recuento digital" y difiere de los métodos de cuantificación de ADN/ARN existentes que se basan en una medición de análogos

35 indirecta de parámetros tales como fluorescencia de un colorante fijado. De este modo, no se requiere manipulación alguna de las proteínas o de la mezcla de proteínas, excepto la exposición de los aptámeros bajo condiciones adecuadas. Además, se piensa que el recuento digital es superior, ya que evita muchas de las ambigüedades y problemas de ruido de señales de una cuantificación análoga (Smyth GK, Bioinformatics (2007) Sep 19).

40 En una realización específica y tal como se muestra en la Figura 1, se rastrea un banco de aptámeros y se selecciona frente a conjuntos in vitro de proteínas plegadas derivadas de material biológico, p. ej. suero. Se apreciará que el banco de etiquetas de aptámeros puede ser generado por cualquier método, incluido SELEX y métodos descritos en la solicitud en tramitación con la presente número EP 07020629.7

Aptámeros que no unen proteínas se separan o eluyen. Para conseguir esto, están disponibles diversas opciones conocidas. Por ejemplo, una muestra de proteína sencilla o una muestra de proteína compleja se puede inmovilizar sobre un soporte sólido. Se puede realizar un lavado riguroso para separar aptámeros no unidos o débilmente

45 unidos. Otra opción es el uso de reticulación reversible del aptámero sobre las dianas de proteínas (foto-aptámeros). Estrategias no SELEX, basadas en electroforesis capilar en equilibrio de mezclas en equilibrio (ECEEM) o la variante no en equilibrio (NECEEM) también se pueden emplear rápidamente para seleccionar

50 aptámeros "inteligentes" con propiedades de unión definidas (Drabovich et al, Anal. Chem. (2006) 78, 3171-3178).

Los aptámeros unidos remanentes se separan de sus hospedantes proteínicos y pasan por un secuenciador de generación siguiente en donde son secuenciados y contados.

Como un ejemplo específico de secuenciación de siguiente generación, los aptámeros unidos se disponen al azar

sobre una superficie tal como una perla o un chip. Los aptámeros se amplifican opcionalmente de forma cíclica, proporcionando grupos de moléculas de cadena sencilla clonales en coordenadas x e y discretas sobre la superficie o sobre una perla individual. Después de haber unido las secuencias a la disposición ordenada, se añade un cebador a cada una de las secuencias. Un secuenciador de ADN ejecuta entonces una química escalonada que
 5 consiste en reactivos que permiten la determinación de una base en cada una de las secuencias complementarias por ciclo. De este modo, se constituye una secuencia complementaria que refleja cada una de las secuencias de aptámeros unidas. Un sistema de iluminación y de formación de imágenes permite fotografiar este proceso y, repitiendo el proceso varias veces, se pueden obtener las secuencias de los aptámeros originales. Típicamente, tecnológicas de este tipo son capaces de secuenciar más de 40-100 millones de tramos de ADN hasta 50 pares de
 10 bases de longitud, y estos números están aumentando rápidamente. Este proceso discurre actualmente en menos de 1 a 3 días desde la preparación de la muestra hasta el resultado de la secuenciación, y estos periodos de tiempo están acortándose.

Se apreciará que se pueden utilizar otros métodos que caigan dentro de la categoría de “secuenciación de polinucleótidos de siguiente generación”, y la presente invención no está limitada al ejemplo específico
 15 proporcionado. “Secuenciación de polinucleótidos de siguiente generación” es una expresión generalmente acuñada para describir plataformas de secuenciación de ADN/ARN que han aparecido desde 2004. Su característica común es que utilizan una química de secuenciación diferente de la vieja tecnología que se basaba en la secuenciación de “Sanger”. Las nuevas plataformas utilizan nuevas químicas, y generalmente tienen un rendimiento muy elevado, y costes mucho menores. Esto se ha conseguido a través de la capacidad de paralelizar las reacciones de secuenciación. Las nuevas plataformas prosiguen típicamente mediante síntesis o extensión de
 20 la cadena (construcción), a diferencia del método de Sanger que trabaja recortando bases (degrada). Además, las nuevas plataformas operan sobre pequeños números de moléculas clonales, a diferencia del método de Sanger que tiene una relación de medición de molécula a base del secuenciador de ADN muy grande.

También se apreciará que aun cuando se hace referencia a la secuenciación de ADN, la tecnología de
 25 secuenciación del ADN también se puede utilizar con secuencias de ARN o ANP (ácido nucleico peptídico). Así, referencias a los secuenciadores o a la secuenciación de siguiente generación abarca ADN, ARN y ANP, así como todas las otras variantes químicas de polímeros basados en ácidos nucleicos y sus análogos que son adecuadas para uso en el método de la presente invención.

Aun cuando una realización preferida de la invención utiliza una disposición ordenada bidimensional de las
 30 secuencias de aptámeros, la invención abarca la secuenciación de siguiente generación en disolución y en tres dimensiones. Para esta última, las etiquetas se pueden concatenar para aumentar la gama dinámica de la disposición ordenada.

Una ventaja significativa de la presente invención es que se pueden utilizar múltiples aptámeros dirigidos a
 35 diferentes proteínas para interrogar a una mezcla de proteínas complejas en un ensayo o un experimento, permitiendo con ello la cuantificación de muchas proteínas en una muestra en paralelo. Una multiplexación de este tipo es claramente susceptible a ensayos de alto rendimiento tales como secuenciación de siguiente generación. Se piensa que el poder de múltiples proteínas de actuar como biomarcadores específicos y sensibles, cuando se utilizan en conjunción, superan en mucho el uso de proteínas sencillas, dado que un cierto número de biomarcadores actúa típicamente en concierto en un estado patológico más que individualmente.

De manera ideal, las secuencias de aptámeros del banco comprenden una agrupación de secuencias de aptámeros
 40 que son específicas para una proteína sencilla. Alternativamente, la agrupación comprende secuencias de aptámeros específicas para más de una proteína. Por ejemplo, el banco puede contener secuencias de aptámeros a1, a2 y a3 que se sabe son específicas para la proteína A. Alternativamente, el banco puede contener secuencias de aptámeros a1, a2 y a3, así como b1 y b2 específicas para la proteína B, c1, etc. Aun cuando las secuencias de
 45 aptámeros a1, a2 y a3 pueden ser específicas para la proteína A, la invención abarca el escenario, por ejemplo, cuando la secuencia de aptámeros a2 es también específica para la proteína B. En otras palabras, el banco de aptámeros puede contener secuencias que sean discretamente específicas para una proteína, así como secuencias que se unan a más de una proteína. Este último escenario puede ser útil cuando se buscan familias y grupos relacionados de proteínas dentro de una mezcla.

Aun cuando se prefiere conocer la identidad de la proteína o molécula a la que el o cada uno de los aptámeros se
 50 une, dicha característica no es esencial. Por ejemplo, se puede conocer la clase de proteínas a la que se une un aptámero, pero los miembros específicos de la clase pueden ser alterados en un estado patológico. La identidad de proteínas se puede dilucidar una vez que se haya completado la cuantificación. Por ejemplo, se pueden analizar uno o más aptámeros específicos frente a una mezcla de proteínas. El o los aptámeros se unen específicamente a
 55 la proteína de interés, la cual puede ser luego aislada y purificada a partir de la mezcla. Una ventaja significativa de la presente invención es que la identificación real del biomarcador de proteínas no necesita ser establecido hasta el

final del proceso de interrogación. Así, durante la interrogación no se requiere la manipulación de la muestra de proteínas.

Se apreciará que la secuenciación de polipéptidos se puede llevar a cabo por cualquier método conocido. Por ejemplo, un aptámero de interés se puede inmovilizar sobre una columna, una mezcla de proteínas se puede hacer fluir a través de la columna y la proteína unida al aptámero sobre la columna se puede analizar utilizando una proteómica basada en MS convencional.

En otra realización, los aptámeros portan etiquetas de secuencias únicas. De este modo, una vez que se han identificado los aptámeros como unión a una o más proteínas de interés, las etiquetas pueden ser separadas de los aptámeros, y las etiquetas son luego secuenciadas y contadas. De este modo, las etiquetas se utilizan como una dirección para un aptámero y proporcionan un método alternativo de proporcionar una cuantificación.

Se apreciará que el método se puede utilizar para proporcionar una diversidad de soluciones y respuestas. En particular, el método se puede utilizar para interrogar a una muestra biológica para establecer si determinadas proteínas o moléculas están supra- o sub-reguladas en un estado patológico cuando se comparan con un estado sano. Dado que se pueden utilizar paneles de aptámeros al mismo tiempo, la invención permite la interrogación de un montón de biomarcadores potenciales y/o conocidos al mismo tiempo. El uso de secuencias de aptámeros como etiquetas, más que etiquetas indirectas y marcadores tales como radiomarcadores y anticuerpos, permite la identificación de moléculas que están supra- y sub-reguladas, así como alteraciones en las modificaciones post-traducción, cambios conformacionales, apariciones de variantes y similares. De este modo, el sondeo de posibles biomarcadores que se pueden identificar es bastante más amplio que el que es posible si se utilizan las técnicas actualmente disponibles.

De hecho, zonas particulares del proteoma se pueden interrogar selectivamente utilizando un conjunto enfocado de aptámeros. Además de ello, dado que las secuencias de los aptámeros ofrecen dicha amplia gama de posibles etiquetas, se puede interrogar al mismo tiempo un cierto número de zonas discretas del proteoma.

En una realización adicional, un panel o agrupación de aptámeros puede ser enfrentado a una muestra compleja. La fracción unida puede ser entonces eluida y re-sondeada frente a una segunda muestra diferente. Los aptámeros libres, no unidos, reflejarán entonces la diferencia en la abundancia de proteínas. De este modo, se pueden medir cambios en la abundancia de proteínas, así como la aparición de cualesquiera proteínas que aparezcan en la muestra como resultado de un estado patológico o post-tratamiento, por ejemplo. En estas realizaciones, la mezcla de proteínas compleja se deriva típicamente de un fluido corporal tal como sangre, plasma o suero.

Se apreciará que la invención también se puede utilizar para validar biomarcadores. Específicamente, se puede rastrear un panel o colección de distintos aptámeros frente a una muestra de proteínas procedente de uno o más individuos sanos y se puede comparar con muestras de proteínas tomadas de uno o más individuos que tienen una enfermedad de interés.

Dado que muchas proteínas se pueden cuantificar al mismo tiempo, la invención es altamente adaptable a análisis de alto rendimiento, reduciendo con ello significativamente el coste y el tiempo típicamente requerido para la validación del biomarcador.

Una ventaja de la presente invención con respecto a la validación de un biomarcador de proteínas es que en lugar de anticuerpos se pueden utilizar aptámeros. La generación de anticuerpos específicos es costosa y no siempre va acompañada de éxito. Por otra parte, habitualmente es posible general un aptámero selectivo a un proceso dado, y el proceso es bastante más rápido y económico. En uso, uno o más aptámeros específicos para una proteína particular de interés se someten a ensayo frente a una mezcla de proteínas complejas. La ausencia de la unión de aptámeros en una muestra sana en comparación con la unión en una muestra enferma (o viceversa) se puede dilucidar a través del recuento de los aptámeros. De nuevo, no es necesaria manipulación alguna de la muestra de proteínas y no se requiere una identificación de las proteínas per se. Además, múltiples aptámeros se pueden analizar frente a un montón de proteínas de interés en un experimento.

Al permitir interrogar un complemento completo de proteínas, el método de la presente invención también se puede utilizar para formar una imagen de la conformación y abundancia de proteínas en individuos. Es decir, para proporcionar un perfil o mapa de una constitución de proteínas de un paciente. Por ejemplo, se puede utilizar una imagen del antígeno específico de próstata (PSA) para establecer un pronóstico y diagnóstico de cáncer de próstata, así como para proporcionar un indicio del progreso de la enfermedad.

En otro ejemplo, si se encuentra que un paciente tiene una cierta variante proteínica, se puede utilizar la información sobre el perfil de proteínas del paciente para el pronóstico de o la susceptibilidad a una enfermedad, y se puede utilizar para ayudar al desarrollo de regímenes de tratamiento individualizados para el paciente. Por

ejemplo, personas que portan una mutación específica en su gen CCR5 producen un receptor CCR5 variante. Estos receptores variantes no se unen al HIV, y de esta forma una persona que porte el receptor variante tienen una posibilidad menor de convertirse en VIH+ y progresar al SIDA como resultado de ello.

5 La invención también se puede utilizar para determinar la especificidad de un aptámero para una molécula sencilla tal como una proteína, estableciendo la varianza entre réplicas de la misma muestra, así como la varianza entre muestras obtenidas de diferentes individuos.

10 El uso de la invención reside también en el diagnóstico. Específicamente, uno o más aptámeros que se sabe identifican biomarcadores validados se pueden utilizar simplemente para interrogar a un fluido corporal tal como sangre o suero, tomado de un individuo. Dado que la unión de un aptámero indica la presencia de una proteína, se puede utilizar un panel de aptámeros para identificar una proteína sencilla o un conjunto de proteínas biomarcadoras que se sabe están alteradas en un estado anormal. Por ejemplo, la determinación del perfilado puede utilizarse para evaluar cantidades de proteínas específicas frente a una línea base conocida. Ejemplos de usos de este tipo serían el uso como un test de embarazo o un test para niveles de PSA. En otro ejemplo, la sepsis es un estado anormal en la que se sabe que muchos factores están supra- o sub-regulados de acuerdo con el tipo de sepsis y la fase de infección.

15 Un diagnóstico de este tipo puede estar en forma de un kit que incluye un panel de aptámeros, cuya unión indica la presencia de biomarcadores que se sabe que son específicos para un estado anormal particular en una muestra biológica derivada de un individuo.

20 En otra realización, la invención se puede utilizar para determinar el progreso de la enfermedad. Por ejemplo, el modelo de biomarcador en la sepsis cambia a lo largo del curso de la infección. La identificación de biomarcadores como indicadores del progreso de la sepsis ayudaría grandemente a establecer regímenes de tratamiento apropiados.

La invención permite también la determinación del perfil del paciente que, a su vez, puede utilizar para dirigir regímenes de tratamiento y un posible tratamiento profiláctico.

25 Así, la invención proporciona una plataforma sencilla sobre la que se puede conseguir una diversidad de resultados.

Todavía en un aspecto adicional, la presente invención abarca el uso de disposiciones ordenadas de moléculas clonales o sencillas para la secuenciación y el recuento de aptámeros.

30 Los autores de la invención han reconocido que los aptámeros pueden ser adaptados de manera ideal para potenciar su compatibilidad con plataformas de secuenciación de ADN de siguiente generación. Por lo tanto, el aptámero puede ser potenciado para permitir la secuenciación e identificación del aptámero utilizando plataformas tecnológicas de este tipo. Así, desde otro aspecto, la presente invención abarca un aptámero posibilitado para la secuenciación que comprende una secuencia de aptámero y una secuencia de adaptador. La secuencia de adaptador comprende una secuencia que posibilita la fijación del aptámero a una superficie para posibilitar la secuenciación del aptámero.

35 El aptámero posibilitado para la secuenciación puede comprender adicionalmente un cebador de secuenciación para secuenciar el aptámero.

Construcciones de este tipo pueden incluir más de un adaptador de secuenciación y de fijación. Por ejemplo, la secuenciación de aptámeros puede estar emparezada entre pares de adaptadores de secuenciación y fijación reasociados a cada uno de los extremos de la secuencia del aptámero.

40 El aptámero posibilitado para la secuenciación también puede comprender un marcador para permitir la visualización del aptámero. Se apreciará que se puede utilizar cualquier marcador adecuado, incluidos marcadores fluorescentes y radiomarcadores.

Aptámeros posibilitados para la secuenciación de la presente invención se pueden utilizar en los métodos, usos y kits de diagnóstico descritos anteriormente.

45 La invención se describirá ahora por medio de ejemplos no limitantes, en que:

La Figura 1 ilustra el método de la presente invención, en el que un banco de secuencias de aptámeros se mezcla con proteoma del suero. La fracción aptámero/proteína unida se eluye antes de ser secuenciada y se cuenta en un secuenciador de segunda generación. El resultado del secuenciador es el número de cada una de las secuencias presentes en la fracción unida.

La Figura 2 ilustra el uso del método para el descubrimiento de biomarcadores, en el que proteomas procedentes de diferentes pacientes se comparan con una línea base o control.

La Figura 3 ilustra una realización alternativa de la presente invención, en la que un banco de aptámeros se mezcla con una primera muestra de proteoma del suero. Después de la cuantificación de la fracción unida, la población de aptámeros que se une a proteínas de interés dentro de la muestra de suero se amplifica y mezcla con una segunda muestra de proteoma del suero. Luego se realiza una comparación de los respectivos resultados de las cuantificaciones primera y segunda.

La Figura 4 es un cromatograma que muestra los diferentes constituyentes de la mezcla obtenidos después 30' de incubación de volúmenes iguales de IgE 150 nM y ProNuc1FR 100 nm.

La Figura 5 muestra cromatogramas CE de IgE 0/20/40....2000 nM incubada con una concentración constante (100 nM) de ProNuc1F (desde la base a la parte superior).

La Figura 6 es una gráfica que muestra la fracción de ProNuc1F unido representada frente a la concentración cambiante de proteínas.

La Figura 7 es una representación esquemática del protocolo de preparación de muestras adaptado para la secuenciación de aptámeros de la siguiente generación.

La Figura 8 es una gráfica que muestra los recuentos absolutos de las secuencias de aptámeros de IgE, recuperadas utilizando dos métodos de análisis diferentes.

La Figura 9 es una gráfica que muestra la fracción de las secuencias de aptámeros de IgE presentes en una mezcla de secuencias contada por la secuenciación de siguiente generación representada frente al número de secuencias, dispersada en una mezcla de secuencias irrelevantes (PhiX).

Ejemplo 1

Este ejemplo está ilustrado en la Figura 1 que muestra la cuantificación de proteínas en una muestra de suero, utilizando un banco de etiquetas de aptámeros. Un banco diverso de aptámeros conocidos se rastrea frente a una mezcla de proteínas derivadas del suero. La mezcla de proteínas está inmovilizada sobre un soporte sólido, y se realiza un lavado riguroso para separar los aptámeros no unidos y débilmente unidos.

Los aptámeros unidos remanentes se separan entonces de sus hospedantes proteínicos y se secuencian y cuentan en un secuenciador de siguiente generación. Después de la disposición ordenada, las secuencias de aptámeros unidos se pueden amplificar por métodos estándar tales como PCR, para aumentar la claridad de la señal frente al ruido de fondo en el secuenciador.

El resultado del secuenciador será una cuantificación absoluta de las proteínas presentes en la muestra biológica, tal como se representa por el banco de aptámeros.

Ejemplo 2

Para uso en un descubrimiento de biomarcador, y tal como se muestra en la Figura 2, el banco de aptámeros se rastrea frente a muestras derivadas de diferentes individuos. Se requerirá también un número adecuado de muestras control o sanas conocidas para establecer una línea base o condición sana. También se rastreará un cierto número de muestras que se sabe representan un estado patológico. Una comparación de las dos poblaciones, la sana frente a la enferma, permitirá también la identificación de aptámeros que muestran alternancias significativas entre las dos poblaciones. Una vez que se ha dilucidado la secuencia del o de cada uno de los aptámeros, se puede entonces identificar la proteína a la que se unen los aptámeros. La identificación de la proteína puede ser mediante técnicas de secuenciación de siguiente generación, o métodos basados en proteómica tales como cromatografía y MS.

Ejemplo 3

Este Ejemplo describe un método alternativo de la presente invención y se ilustra en la Figura 3.

Un banco de aptámeros se rastrea frente a una primera muestra biológica. Al igual que con el Ejemplo 1 o el Ejemplo 2, los aptámeros no unidos se separan y los aptámeros unidos remanentes se liberan de sus proteínas hospedantes y se secuencian y recuentan en un secuenciador de siguiente generación.

La población de secuencias que se unen a proteínas en la primera muestra se rastrea luego frente a una segunda muestra biológica. Tal como antes, los aptámeros no unidos se desechan, y los aptámeros remanentes se

secuencian y recuentan. Si las dos muestras se derivan de sujetos sanos, cualquier variación entre las dos muestras se puede atribuir a la varianza dentro la población normal. De manera similar, cualquier varianza entre el resultado de las dos muestras derivadas de pacientes que se sabe tienen una enfermedad particular, también se puede atribuir a varianza. Sin embargo, cualquier varianza significativa entre sujetos sanos y enfermos puede atribuirse a un cambio en la población de proteínas como resultado de la enfermedad de interés. Las secuencias de aptámeros en las que se encuentran cambios significativos se pueden luego traducir para identificar la o las proteínas a la o a las que se une el o los aptámeros.

Ejemplo 4

En este ejemplo, el banco de aptámeros comprende al menos una agrupación de aptámeros específica para una proteína. Igualmente, el panel puede contener un cierto número de grupos de aptámeros, en que cada uno de los grupos es específico para una proteína diferente.

La mezcla de proteínas puede ser simplemente una muestra biológica tal como sangre o suero. Alternativamente, la mezcla de proteínas se puede obtener enriqueciendo la muestra biológica para la fracción de proteínas. En este ejemplo, ha de tenerse cuidado de minimizar la alteración y la desnaturalización de la población de proteínas.

El banco de secuencias de aptámeros seleccionadas, conocidas, se rastrea frente a la mezcla de proteínas. En un ejemplo, la mezcla de proteínas está unida a una columna, y se deja que el panel de aptámeros fluya a través de la columna. Los aptámeros que no se unen se retiran. Los aptámeros que se unen se separan de sus hospedantes proteínicos y se secuencian y recuentan en un secuenciador de siguiente generación.

Si la proteína de interés es de hecho un biomarcador para un estado patológico, los aptámeros se encontrarán en la fracción unida procedente de la muestra enferma y no en una muestra derivada de un individuo sano. Alternativamente, el papel de una proteína como un biomarcador puede manifestarse por supra- o sub-regulación de la proteína en un individuo enfermo cuando se compara con un sujeto sano.

Ejemplo 5

En este ejemplo, el banco de aptámeros contiene una o más secuencias que se sabe que son específicas para una o más proteínas. La identidad de la o de cada una de las proteínas de interés también es conocida. El banco puede comprender un solo grupo específico para una proteína sencilla, o puede comprender muchos grupos específicos para una disposición ordenada de proteínas.

El banco de aptámeros se rastrea frente a una muestra biológica derivada de un individuo, tal como sangre. El banco de aptámeros puede mantenerse sobre un soporte, y la muestra biológica se hace pasar por encima de los aptámeros. Cualesquiera proteínas no unidas se separan y se conservan los complejos de aptámero:proteína. Los aptámeros retenidos se liberan de sus hospedantes proteínicos y se secuencian y recuentan en un secuenciador de siguiente generación.

La presencia y cantidad, o la ausencia de aptámeros en el secuenciador de siguiente generador puede utilizarse para diagnosticar una enfermedad o problema particular.

Ejemplo 6

Se diseñó un nuevo aptámero oligonucleótido, de modo que el aptámero era compatible con la plataforma de tecnología de secuenciación de ADN de siguiente generación que se utilizó.

El aptámero tenía una secuencia de oligonucleótidos con tres regiones distintas: a) una región de aptámero funcional, b) una región de adaptador y c) un marcador.

La región de aptámero en este estudio se basó en un aptámero bien estudiado (Wiegand et al Journal of Immunology (1996) 157 221-230) que tiene una elevada afinidad para la inmunoglobulina E humana.

La región de adaptador hizo compatible la construcción con los procesos de preparación de la muestra de un secuenciador de siguiente generación Illumina GA1 y se diseñó de acuerdo con los protocolos del fabricante.

El marcador era 6-carboxifluoresceína fluorescente (FAM), cuya inclusión permitió la visualización de la construcción.

Las construcciones se obtuvieron mediante síntesis de ADN en fase sólida estándar y se purificaron mediante electroforesis en gel de poliacrilamida para asegurar bajas tasas de error.

A la construcción con el marcador FAM se aludirá aquí en lo que sigue como ProNuc1F y sin el marcador como

ProNuc1.

Las construcciones se utilizaron para demostrar que:

i) secuencias de oligonucleótidos de cadena sencilla derivadas de aptámeros de este tipo son compatibles con procesos de preparación de muestras, utilizados para las aplicaciones de la secuencia de ADN de siguiente generación;

ii) las secuencias se pueden secuenciar utilizando la tecnología de secuenciación Illumina GA1, es decir, se puede recuperar ("leer") la identidad de la ProNuc1; y

iii) el recuento de secuencias ProNuc1 completas, siguiendo un experimento de secuenciación Illumina GA1 estándar, se correlaciona directamente con el número de secuencias ProNuc1 preparadas, dispersadas en una muestra, confirmando con ello que la secuenciación de siguiente generación se puede aplicar a la proteómica de acuerdo con la presente invención.

Ejemplo 7

Se utilizó la electroforesis capilar (CE) para confirmar que la secuencia de oligonucleótidos exhibe una afinidad para proteínas distintas.

Preparación de reactivos y disoluciones

Se preparó una disolución patrón 10 nM de ProNucF1 a partir de una fuente de aptámeros seca, diluyendo el material de ADN en tampón TGK (tampón tris(hidroxiamino)metano-glicina-potasio), pH 8,4. Después de la dilución, la disolución patrón del aptámero se incubó a 75°C durante 10 minutos y subsiguientemente se enfrió y almacenó en hielo para asegurar la ausencia de multímeros. El proceso de calentamiento también asegura la formación de la infraestructura secundaria del aptámero.

Partiendo de una disolución patrón 5500 nM concentrada de proteína, se preparó una serie de dilución de IgE en tampón TGK (Tabla 1):

Tabla 1: Series de dilución de IgE

Disolución	Concentración (nM)
1	20
2	40
3	60
4	100
5	150
6	200
7	300
8	1000
9	2000

Las muestras para CE, que demuestran la formación de completo aptámero-proteína, se prepararon incubando 5 µl de la disolución patrón ProNucF1 con 5 µl de disolución de proteínas. De este modo se preparó un total de seis réplicas para cada uno de los niveles de concentración de proteínas. Todas las muestras se incubaron a la temperatura ambiente durante un mínimo de 30 minutos durante no más de 40 minutos.

Aparato

Las separaciones se consiguieron en un sistema Beckman P/ACE 2200 CE-LIF (Beckman-Coulter, Fullerton, CA, EE.UU.). Las separaciones se realizaron utilizando capilares de sílice condensada (5 µm de ID, 360 µm de OD) con una longitud total de 40,2 cm, detector a 30 cm. Los capilares se trataron previamente bombeando NaOH 1 M, H₂O desionizada y tampón a través del capilar durante 10 min en cada caso. Entre cada una de las separaciones electroforéticas, el capilar se lavó con base, agua y tampón de nuevo, para separar cualquier residuo de las paredes del capilar. Las muestras se inyectaron utilizando presión durante 4 s a 1 psi (0,07 kg/cm²). El detector LIF empleaba la línea de 488 nm de un láser de iones Ar de 3 mW (Beckman-Coulter) para la excitación, y la emisión se recogió a través de un filtro de 520 + 10 nm. Los datos se registraron y analizaron con el software P/ACE

(Beckman-Coulter).

Unión de proteínas-aptámero

- 5 Muestras a cada uno de los niveles de concentración de proteínas se hicieron pasar sobre la columna en un modo normal utilizando las condiciones descritas (Tabla 2), primeramente para asegurar que de hecho se formaba un complejo proteína- aptámero después de la incubación de cada una de las muestras.

Tabla 2: Condiciones de separación en modo normal

Acción	Duración (s)
Lavar 20 psi (1,4 kg/cm ²) – NaOH 0,1 M	120
Lavar 20 psi (1,4 kg/cm ²) – Tampón	120
Inyectar 1 psi (0,07 kg/cm ²) – Muestra	4
Inyectar 0,1 psi (0,07 kg/cm ²) – Agua	1
Separar 20 kV	760
Lavar 10 psi (0,7 kg/cm ²) – Tampón	60

Resultados

- 10 Como un ejemplo, en la Figura 4 se muestra un cromatograma para la inyección en modo normal de una muestra que contenía 150 nM de IgE y 100 nM de ProNuc1F.

Con el fin de aplicar la formación del complejo aptámero-proteína para fines de cuantificación, la fracción de aptámero unido ha de reflejar la serie de dilución de proteínas. Para demostrar esto, se incubó una cantidad constante de ProNuc1F con diferentes concentraciones de proteínas (véase antes).

- 15 A partir de la Figura 5 resulta claro que con concentraciones crecientes de proteínas (carril inferior “0”; carril superior 5000 nM) aumenta el área pico del completo aptámero-proteína. Simultáneamente, disminuye la señal de ProNuc1F libre hasta el punto en el que no se detecta aptámero libre alguno (carril superior). Esto demuestra que la adición de una secuencia de adaptador específica para el proceso de secuenciación de siguiente generación a una secuencia de aptámero conocida contra IgE no perjudica las capacidades del aptámero de ProNuc1F – ProNuc1F no retiene su elevada y selectiva afinidad hacia IgE.

- 20 La Figura 6 muestra la información en la Figura 5 trasladada a una gráfica en donde la “fracción ProNuc1F unida” se representa frente a la concentración de proteína. A partir de esta gráfica se puede observar que el aptámero ProNuc1F se puede utilizar para “leer” la información de proteínas cuantitativa. A partir de esta (y de otras) gráficas es también posible deducir las constantes de disociación en equilibrio K_d , que proporcionan una perspectiva en la afinidad de ProNuc1F para IgE, es decir entre 76 y 92 nM.

25 Ejemplo 8

- 30 Se llevaron a cabo experimentos para demostrar que las características de unión ilustradas en Ejemplo previo son aplicables en un contexto cuantitativo, es decir, las diferencias en la concentración de proteínas se correlacionan con la fracción de aptámero unido (complejo de aptámero-proteína). La selección de estos complejos de aptámero-proteína se consiguió por medio de electroforesis capilar en equilibrio de mezclas en equilibrio (NECEEM) tal como se desarrolla por Krylov et al (Journal of the American Chemical Society (2002) 124 13674-13575; Analytical Chemistry (2003) 75 1382-1386).

Método

Se desarrolló un protocolo de preparación de muestras adaptado para permitir la secuenciación del aptámero de IgE funcional que contenía la secuencia de secuenciación añadida (ProNuc1).

- 35 La Figura 7 esboza las diferentes etapas en el protocolo de preparación de la muestra. Después de la reasociación del cebador 1 en el sitio del cebador 1 de secuenciación, se formó una molécula de doble hebra mediante extensión Taq polimerasa. Un adaptador se ligó a la molécula de doble hebra generando la molécula completa.

La secuencia de aptámero de IgE utilizada en este estudio era un oligonucleótido de cadena sencilla con un sitio de cebador 1 de secuenciación en su extremo 3'. Con el fin de producir la hebra complementaria, un cebador con cola

se reasoció con el cebador 1 y la extensión se realizó utilizando Taq polimerasa.

La receta para esta reacción era:

5	ProNuc1 10 μ M	2 μ l
	Cebador 1 10 μ M	5 μ l
	dNTPs 2,5 mM	4 μ l
	tampón 10x Taq polimerasa	5 μ l
	MgCl ₂ 50 mM	2 μ l
10	Taq polimerasa	0,5 μ l
	agua	31,5 μ l

La reacción se mezcló y calentó hasta 94°C durante 2 minutos, seguido de enfriamiento controlado hasta 60°C e incubación a 72°C durante 10 minutos. Los productos de la reacción se limpiaron utilizando una columna MinElute de Qiagen siguiendo las instrucciones del fabricante, y eluyendo en 10 μ l de tampón EB.

20 μ l de cada una de las soluciones 100 μ M de adaptadores ProAd1 y ProAd2 se reasociaron juntos en un volumen total de 50 μ l de Tris-HCl 10 mM, pH 8,3 calentando hasta 94°C durante 2 minutos y enfriando de manera controlada hasta la temperatura ambiente. ProAd1 y ProAd2 son oligonucleótidos sintetizados que se derivan del kit Illumina GA1.

La mezcla utilizada era:

20	Aptámero Ds	10 μ l
	Tampón 2x ADN ligasa de Illumina	25 μ l
	Mezcla oligo de adaptador reasociada	10 μ l
	ADN ligasa de Illumina	5 μ l

Las reacciones se mezclaron e incubaron durante 15 minutos a la temperatura ambiente antes de la electroforesis en gel en agarosa al 2%. La banda apropiadamente dimensionada se escindió del gel y el ADN se extrajo y amplificó siguiendo el protocolo PCR de enriquecimiento Illumina estándar.

Amplicones de la PCR se cuantificaron utilizando un bioanalizador Agilent 2100 y las reacciones de secuenciación se prepararon añadiendo estos aptámeros amplificados a una cantidad constante de PhiX, con la intención de generar los números de racimo/tramos de Illumina GA1 dados en la Tabla 3 para producir series de dilución 2x. El carril 4 es PhiX sin aptámero añadido y sirve como un carril control.

Para evitar dudas, PhiX se refiere al genoma circular del bacteriófago PhiX174 del ADN de doble cadena que consiste en 5386 nucleótidos. En este estudio, el genoma se procesó de acuerdo con el protocolo de secuenciación de Illumina GA1 para servir como normalización para la carga de secuencia en los diferentes carriles.

Se preparó una celda de flujo de extremo sencillo y se secuenció en un Illumina AG1 de acuerdo con las instrucciones del fabricante.

Tabla 3: Números de secuencia de ProNuc1 y PhiX reales utilizados por tramos para los ocho carriles diferentes en la celda de flujo de Illumina GA1. Un carril consiste en alrededor de 300 tramos

Carril	ProNuc1/tramo	ProNuc1/carril	PhiX
1	5000	1500000	20.000
2	2500	750000	20.000
3	1250	375000	20.000
4	0	0	20.000
5	625	187500	20.000
6	312	93600	20.000
7	156	46800	20.000
8	78	23400	20.000

Resultados:

- Los datos de la secuenciación se procesaron de acuerdo con las instrucciones del fabricante. Para el análisis de los datos se aplicaron diferentes estrategias para evaluar las lecturas de las secuencias obtenidas en los diferentes carriles: a) una búsqueda para los emparejamientos exactos con la secuencia del aptámero (exacto), b) una búsqueda que permite hasta 3 errores de amplificación/secuenciación (Agreg). Es claro que se pueden también aplicar otros métodos de procesamiento. La Figura 8 representa los números absolutos de aptámeros de IgE contados en los diferentes carriles buscando emparejamientos exactos/emparejamientos con hasta tres errores recuperados en los diferentes carriles, es decir, un recuento de la secuencia de aptámeros exacta y un recuento de la secuencia de aptámeros tolerante al error (Agreg). El valor en el eje X corresponde a las lecturas de los carriles 8 hasta 1 (se excluye el carril control 4; véase la Tabla 3).
- En contraposición, la Figura 9 representa la fracción de aptámeros de IgE contados buscando emparejamientos exactos en las secuencias totales obtenidas frente al número de aptámeros dispersados en la mezcla de secuencias PhiX. En ambos casos se demuestra una correlación lineal que confirma el uso de la secuenciación de la siguiente generación para cuantificar aptámeros.
- Los datos anteriores demuestran que, cuando se sigue un protocolo de secuenciación Illumina GA1 estándar, el recuento de secuencias ProNuc1 completas, se correlaciona directamente con el número de secuencias ProNuc1 preparadas dispersadas en la muestra, confirmando con ello que se puede utilizar la secuenciación de siguiente generación para la cuantificación de acuerdo con la presente invención.
- Además de ello, los datos demuestran que las secuencias de ProNuc1 preparadas se pueden secuenciar utilizando la tecnología de secuenciación de siguiente generación. En otras palabras, se puede recuperar ("leer") la identidad de la ProNuc1.

REIVINDICACIONES

1. Un método para medir la cantidad de al menos una molécula en una muestra biológica, comprendiendo el método:
 - a) combinar la muestra con uno o más aptámeros y permitir que una o más moléculas en la muestra se unan al o a los aptámeros;
 - b) separar las moléculas unidas de las no unidas; y
 - c) cuantificar la o las moléculas unidas al o a cada uno de los aptámeros,en donde la cuantificación de la o las moléculas unidas se lleva a cabo secuenciando al menos parte del o de cada uno de los aptámeros.
2. Un método de acuerdo con la reivindicación 1, en el que la al menos una molécula es una proteína.
3. Un método de acuerdo con la reivindicación 1 o la reivindicación 2, en el que se conoce la identidad de la molécula.
4. Un método de acuerdo con la reivindicación 1 o la reivindicación 2, en el que se desconoce la identidad de la molécula.
5. Un método de acuerdo con la reivindicación 4, en el que dicho método comprende, además, determinar la identidad de la molécula.
6. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que se conoce la secuencia del o de cada uno de los aptámeros.
7. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 6, en el que cada una de las secuencias de los aptámeros porta una etiqueta única.
8. Un método de acuerdo con la reivindicación 7, en el que la etiqueta es la secuencia del aptámero.
9. Un método de acuerdo con la reivindicación 7 o la reivindicación 8, en el que la etiqueta es parte de la secuencia del aptámero.
10. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 9, en el que la secuenciación se lleva a cabo en una disposición ordenada de una molécula sencilla o en una disposición ordenada de una molécula sencilla clonal.
11. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 10, en el que el método comprende, además, separar el o los aptámeros unidos a la o a cada una de las moléculas y disponer ordenadamente el o los aptámeros sobre una superficie.
12. Un método de acuerdo con la reivindicación 11, en el que el método comprende, además, amplificar los aptámeros dispuestos ordenadamente.
13. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 12, en el que el uno o más aptámeros comprenden diferentes secuencias que se unen a la misma molécula.
14. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 12, en el que el uno o más aptámeros comprenden paneles de secuencias de aptámeros, uniéndose cada uno de los paneles a una molécula diferente.
15. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 14, en el que los aptámeros se derivan de polinucleótidos.
16. Un método de acuerdo con la reivindicación 15, en el que los aptámeros son polinucleótidos y tienen entre aproximadamente 30 y aproximadamente 60 bases.
17. Un método de acuerdo con la reivindicación 16, en el que los aptámeros tienen entre aproximadamente 40 bases.
18. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 17, en el que la muestra biológica es un fluido corporal.

19. Un método de acuerdo con la reivindicación 18, en el que el fluido corporal es sangre o se deriva de sangre.
20. Un método de acuerdo con la reivindicación 19, en el que el fluido corporal es suero o plasma.
21. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 20, en el que el método comprende, además:
- 5 d) combinar una segunda muestra biológica con la fracción de aptámero unida, obtenida en c);
 - e) separar moléculas unidas de las no unidas;
 - f) cuantificar la o las moléculas unidas a aptámeros; y
 - g) comparar las cantidades obtenidas en c) con las obtenidas en f),
- 10 en donde la cuantificación de aptámeros unidos a una o más moléculas en la segunda muestra se lleva a cabo secuenciando al menos parte de cada uno de los aptámeros.
22. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 21, en el que el método comprende, además comparar la cantidad del o de cada uno de los aptámeros frente a una cantidad control o de línea base.
23. Un método de acuerdo con la reivindicación 21 o la reivindicación 22, en el que la segunda muestra se obtiene de un individuo del que se sabe que se encuentra en un estado patológico.
- 15 24. Un método de acuerdo con la reivindicación 21 o la reivindicación 22, en el que la segunda muestra se obtiene de un individuo del que se sabe que se encuentra en un estado sano.
25. Un método de acuerdo con la reivindicación 21 o la reivindicación 22, en el que la segunda muestra se obtiene de un individuo después de tratamiento con fármacos.
- 20 26. Uso de un método según una cualquiera de las reivindicaciones 1 a 25, para identificar uno o más biomarcadores.
27. Uso de un método según una cualquiera de las reivindicaciones 1 a 25, para validar uno o más biomarcadores.
28. Uso de un método según una cualquiera de las reivindicaciones 1 a 25 como un método de diagnóstico.

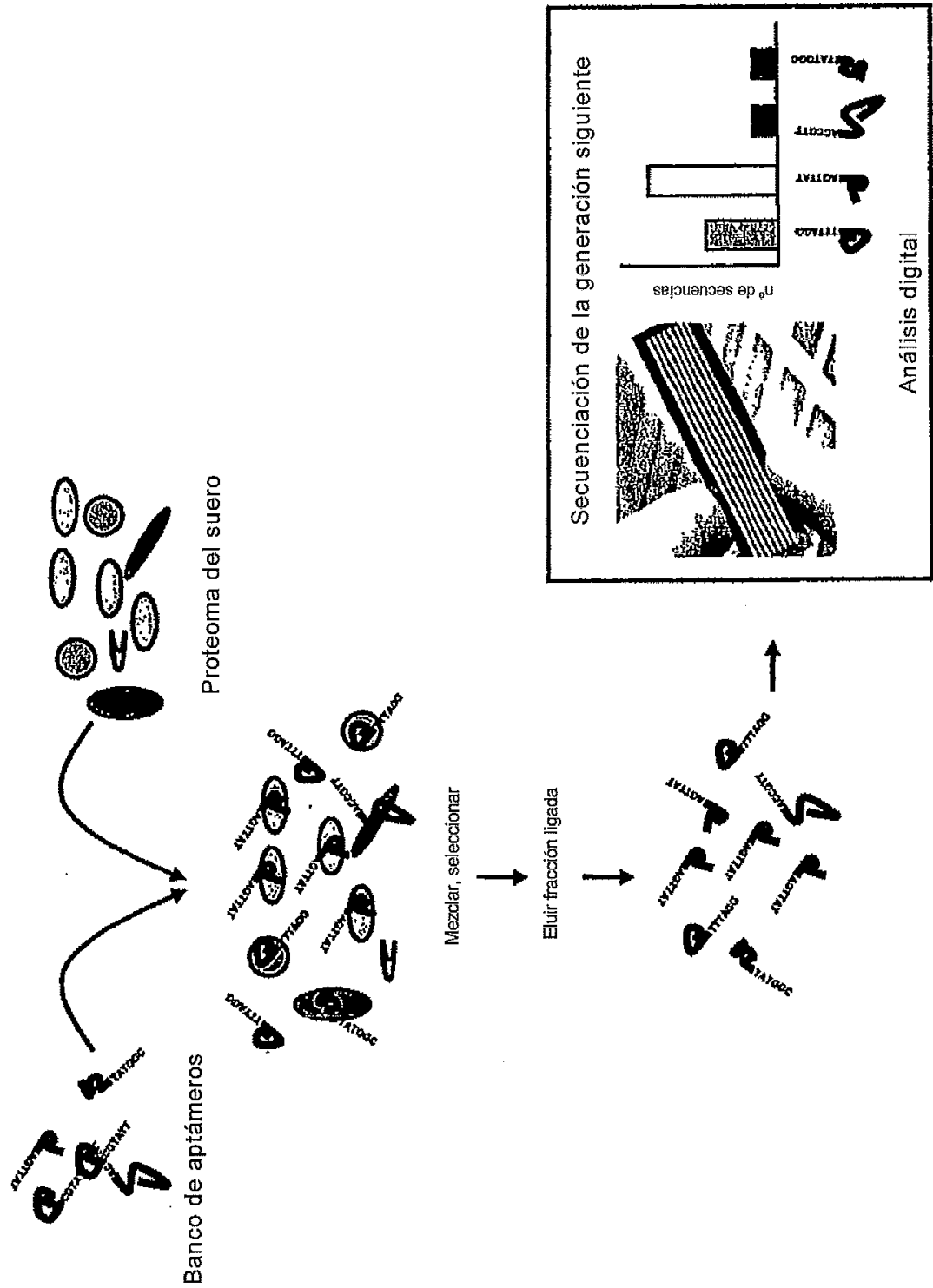


Figura 1

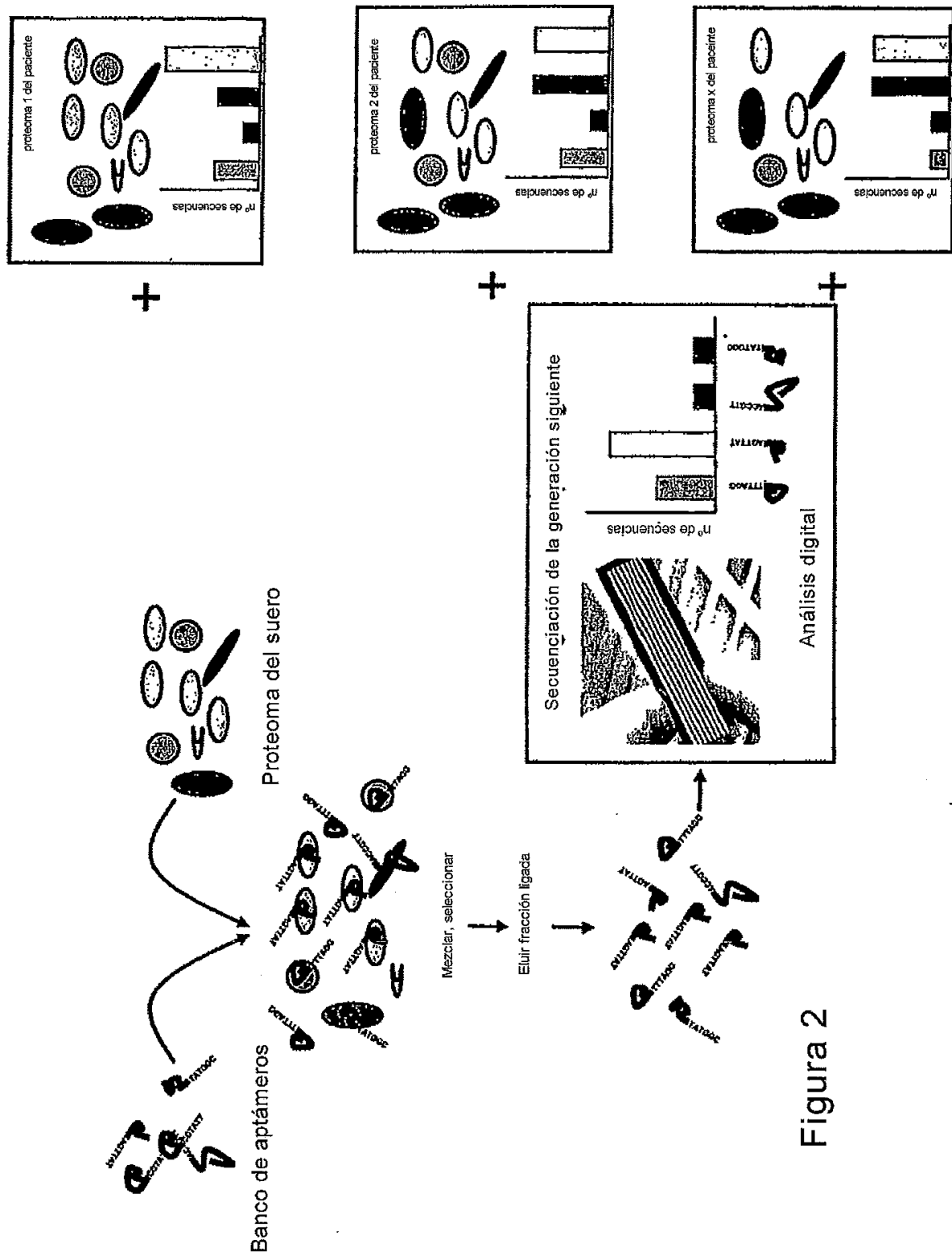


Figura 2

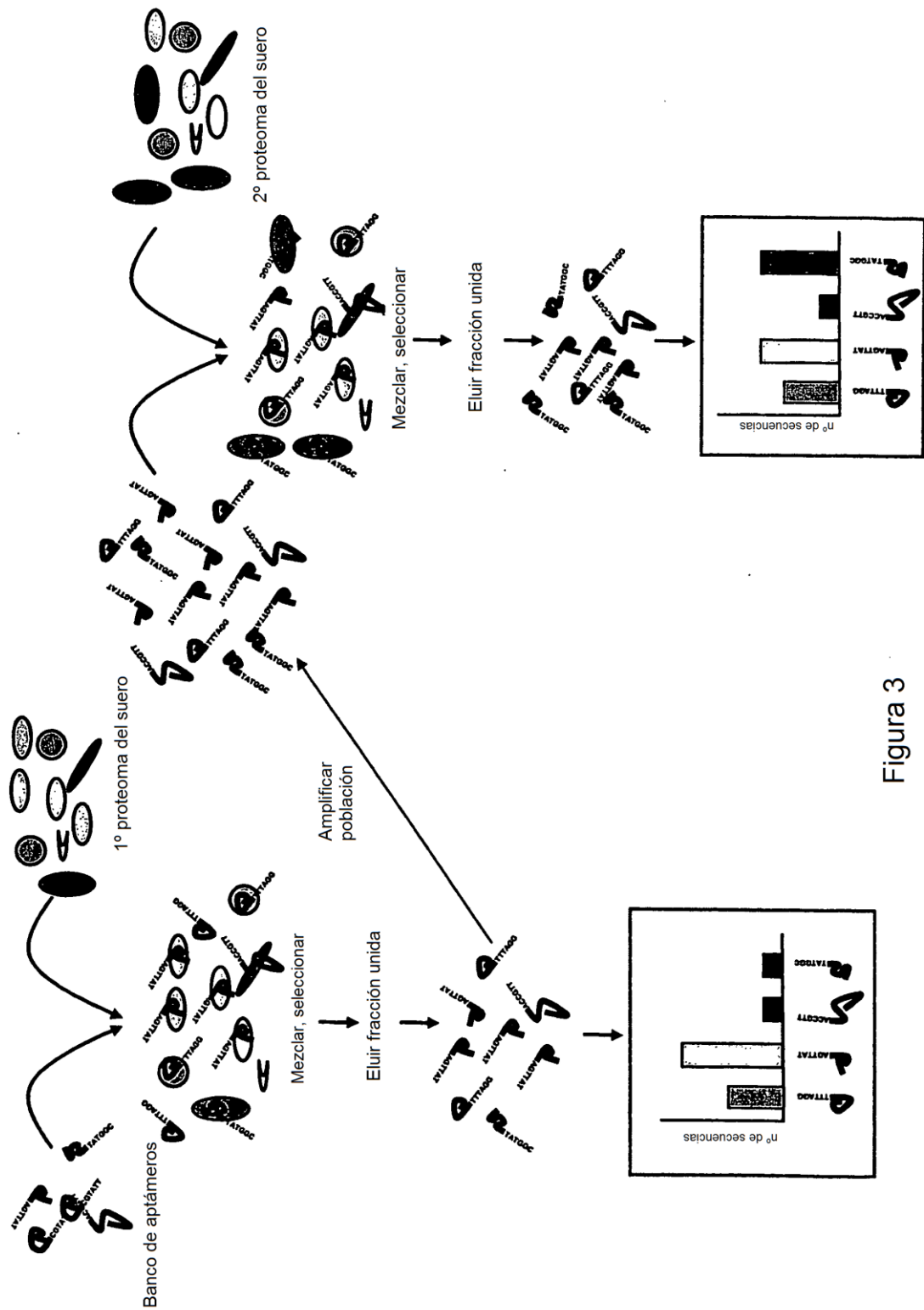


Figura 3

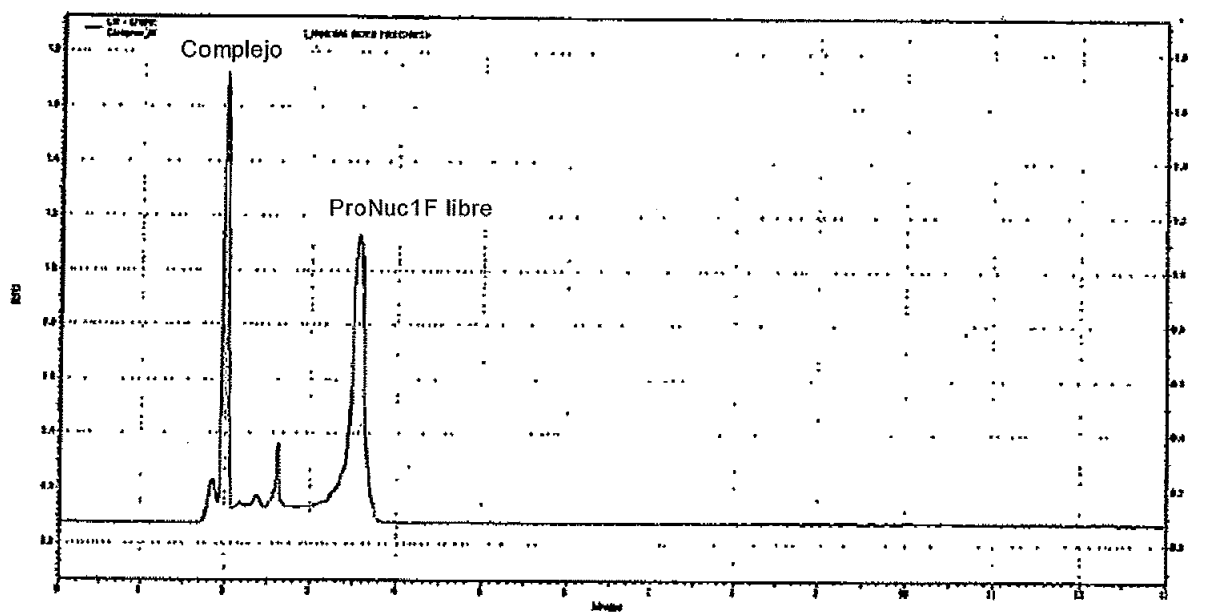


Figura 4

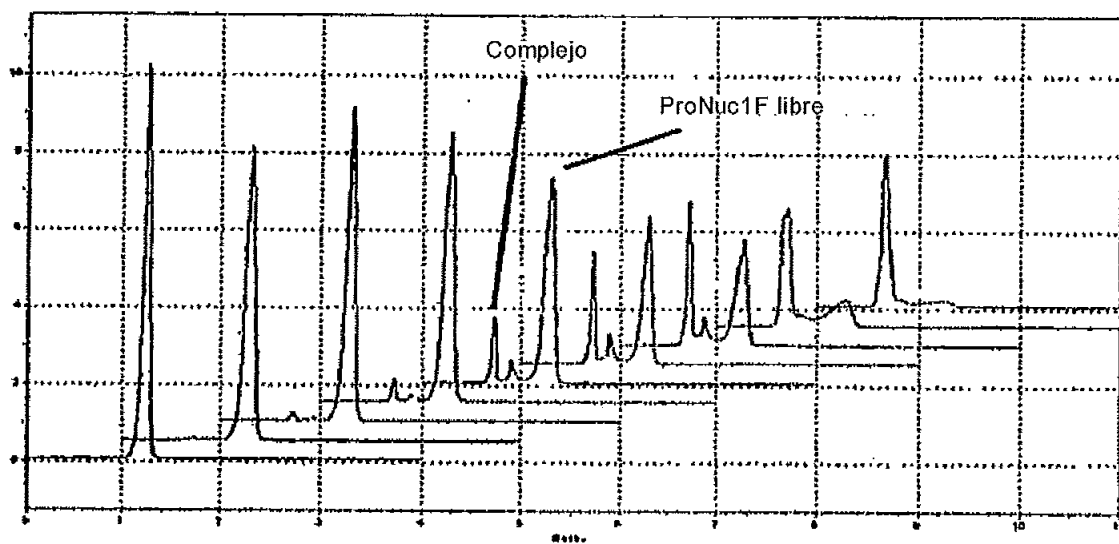


Figura 5

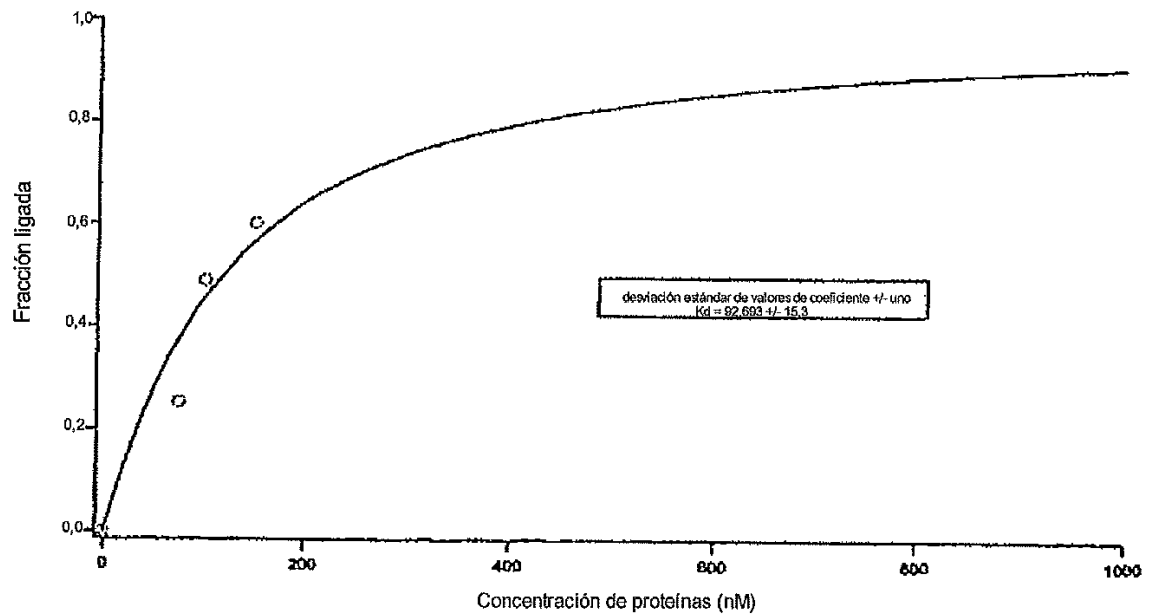


Figura 6

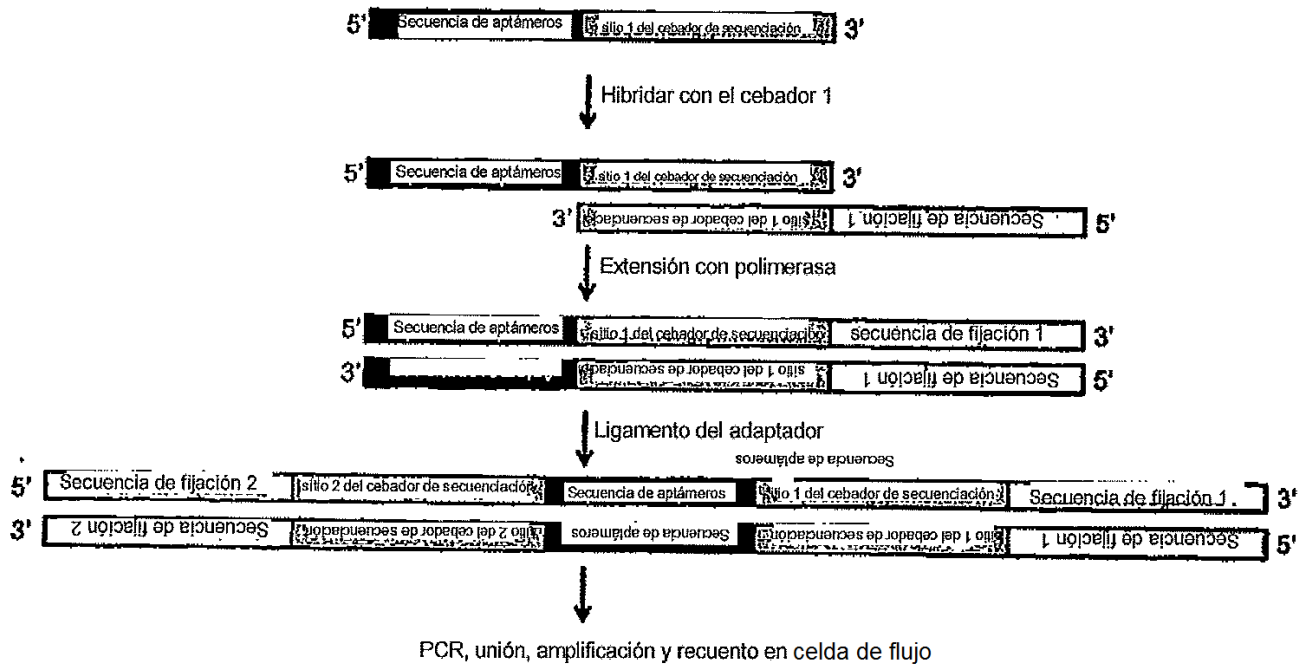


Figura 7

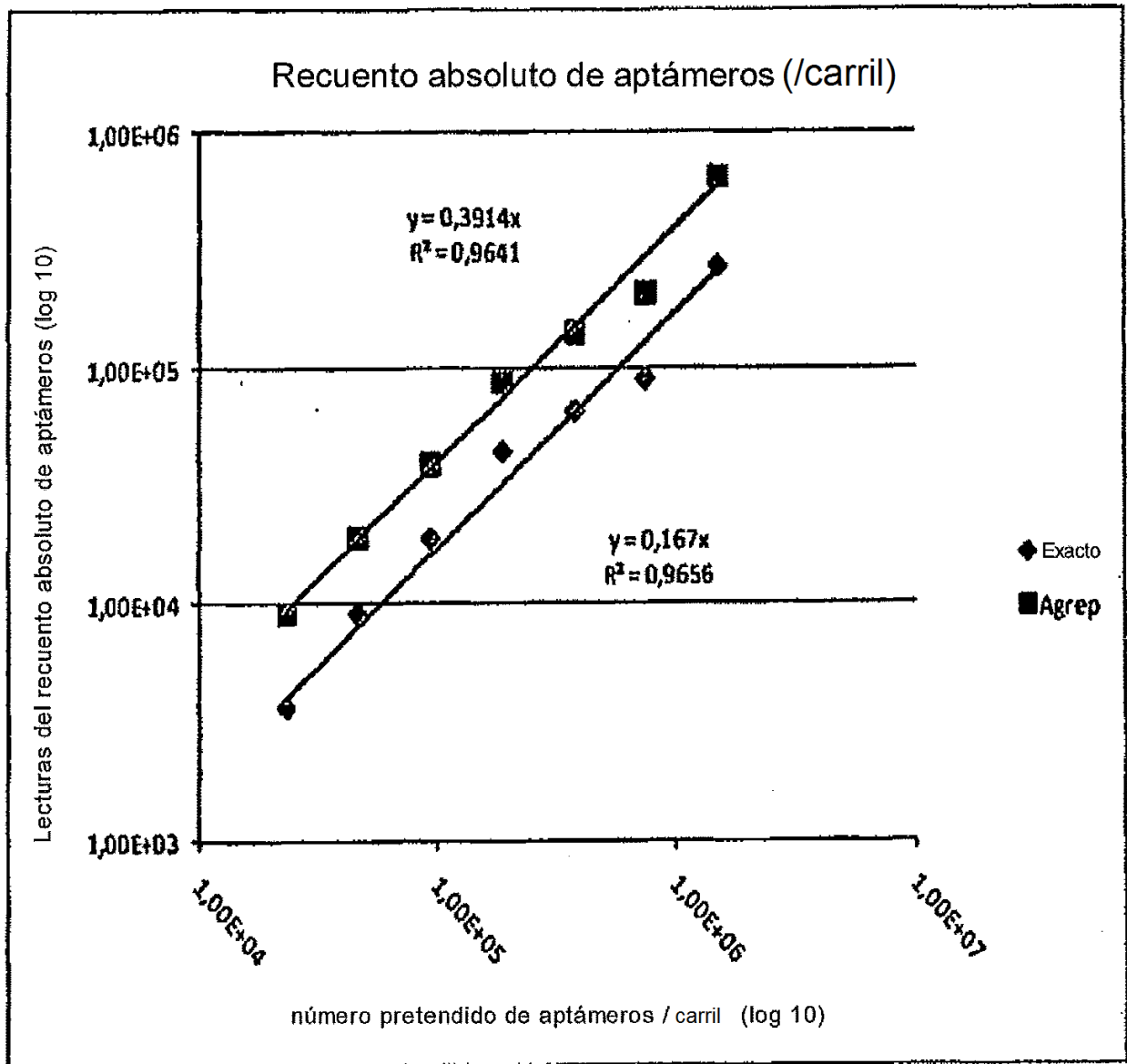


Figura 8

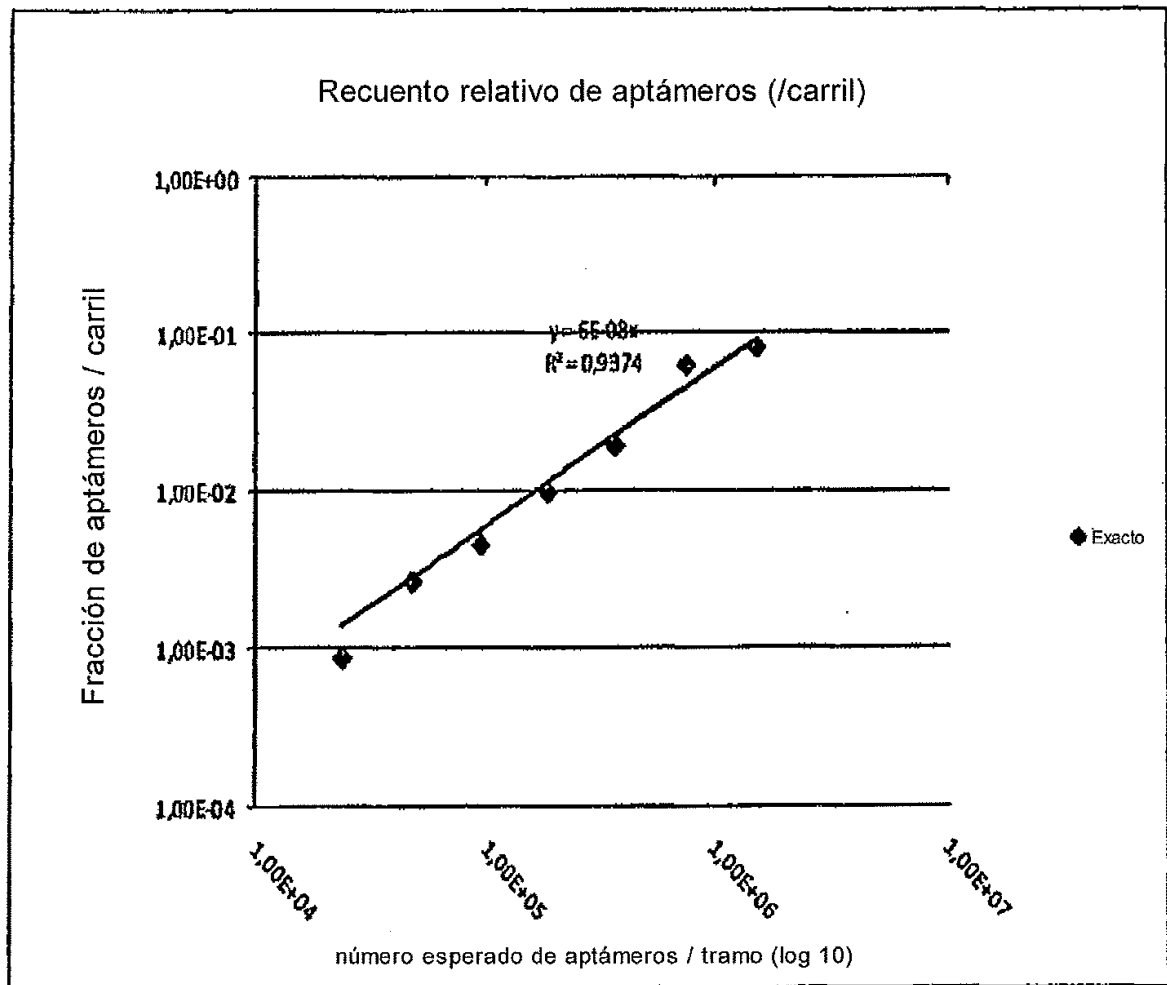


Figura 9