

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 613 745**

51 Int. Cl.:

C12Q 1/37 (2006.01)

G01N 33/487 (2006.01)

G01N 33/58 (2006.01)

G01N 33/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.07.2013 PCT/NL2013/050537**

87 Fecha y número de publicación internacional: **23.01.2014 WO2014014347**

96 Fecha de presentación y número de la solicitud europea: **15.07.2013 E 13759843 (9)**

97 Fecha y número de publicación de la concesión europea: **07.12.2016 EP 2872898**

54 Título: **Secuenciación de proteínas de una sola molécula**

30 Prioridad:

16.07.2012 NL 2009191

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

25.05.2017

73 Titular/es:

**TECHNISCHE UNIVERSITEIT DELFT (100.0%)
Stevinweg 1
2628 CN Delft, NL**

72 Inventor/es:

**JOO, CHIRLMIN;
DEKKER, CORNELIS;
VAN GINKEL, HENDRIKA GEERTRUIDA
THEODORA MARIA y
MEYER, ANNE SARA**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 613 745 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

Secuenciación de proteínas de una sola molécula

Campo de la invención

5 La invención se refiere a un método (de una sola molécula) para determinar el tipo de una proteína (mediante secuenciación), así como a un dispositivo que puede usarse para dicho método.

Antecedentes de la invención

10 Métodos para análisis de proteínas de una sola molécula son conocidos en la técnica. El documento WO2010065531, por ejemplo, describe que dichos métodos pueden usarse para el descubrimiento de nuevos biomarcadores, cuantificación y cribado de alto rendimiento. Se indica que péptidos unidos a la superficie son capaces de ser secuenciados directamente usando una degradación de Edman modificada seguida por detección, p. ej., detección de anticuerpos marcados. El cribado de alto rendimiento se habilita usando combinados de moléculas (p. ej., anticuerpos marcados) para identificar y cuantificar analitos de proteínas individuales en una muestra biológica.

15 Además, el documento WO2010144151 describe composiciones, métodos y sistemas para realizar análisis en tiempo real, de una sola molécula, de reacciones analíticas en las que se está produciendo síntesis de proteínas. La capacidad de analizar dichas reacciones proporciona una oportunidad de estudiar esas reacciones así como de potencialmente identificar factores y/o estrategias para influir en dichas reacciones, p. ej., para mejorar, inhibir, o afectar de otro modo a dichas reacciones incluyendo, aunque sin limitarse a, afectar a la tasa, procesividad, fidelidad, duración y similares, de la reacción. Este documento describe especialmente un método de determinación de una secuencia de aminoácidos codificada por una molécula de ARNm diana, que comprende: a) proporcionar una mezcla de reacción que comprende la molécula de ARNm diana, un complejo ribosómico que comprende tMet-ARNt, tMet en el sitio P, y una pluralidad de tipos de aminoacil-ARNt marcados, libres en solución, en donde el ribosoma y/o la molécula de ARNm diana está inmovilizada/o sobre un soporte, de modo que un volumen de observación contenga no más de un ribosoma y/o molécula de ARNm, y en donde además el complejo ribosómico no comprenda un marcador detectable o un grupo de extinción; b) iniciar una traducción procesiva de la molécula de ARNm por el complejo ribosómico; c) durante dicha traducción procesiva, detectar secuencial y ópticamente la asociación del complejo ribosómico con al menos un primer aminoacil-ARNt marcado y un segundo aminoacil-ARNt marcado, donde dicha asociación da como resultado una incorporación de un primer aminoácido del primer aminoacil-ARNt marcado y un segundo aminoácido del segundo aminoacil-ARNt marcado en una cadena polipeptídica naciente; y d) identificar el primer aminoácido y el segundo aminoácido, determinando de este modo una secuencia de aminoácidos codificada por la molécula de ARNm diana.

Sumario de la invención

35 La invención se describe según las reivindicaciones adjuntas, en detalle de la siguiente manera. Las proteínas son la base de la vida, dado que son los engranajes de funcionamiento en todas las formas de vida. Para entender los fenómenos biológicos, se requiere tener un conocimiento exhaustivo de las proteínas implicadas. La secuenciación de proteínas, que determina la secuencia de aminoácidos de una proteína, se usa para obtener un perfil de poblaciones de proteínas, de líneas celulares a tejidos celulares a organismos individuales. Desde la primera secuenciación de proteínas de insulina en los años 1950, la tecnología de secuenciación ha evolucionado de forma constante para abrir la era de la proteómica, el mapeo exhaustivo de las proteínas celulares.

40 La secuenciación de proteínas moderna se basa principalmente en técnicas de espectrometría de masas (ESI, MALDI, etc.). Cada una tiene sus propias ventajas y desventajas, pero todas ellas comparten las mismas limitaciones. En primer lugar, pueden analizar solamente fragmentos de proteínas (de aproximadamente 10-20 aminoácidos). Cuando se examinan proteínas de longitud completa (normalmente de varios cientos de aminoácidos de longitud), una complicación computacional impide una predicción de secuencias precisa. En segundo lugar, a menudo no consiguen reconocer especies secundarias integradas entre otras especies dominantes, dado que la predicción de secuencias se realiza mediante análisis de picos espectrales complejos. Dado que muchas proteínas celulares existen en baja abundancia, esto hace difícil obtener información proteómica a gran escala.

50 En la secuencia de ADN, se afrontan desafíos similares, pero se superan cuando se amplifican muestras de ADN hasta que se consigue una elevada relación de señal con respecto a ruido. A diferencia del ADN, no existe ninguna maquinaria natural que pueda amplificar proteínas. En esta memoria se pretende desarrollar un método completamente novedoso que pueda cuantificar proteínas celulares con una precisión tan elevada como para técnicas a gran escala, y usar cantidades de muestra tan pequeñas como una sola célula.

55 Por lo tanto, es un aspecto de la descripción proporcionar un método (de secuenciación) (de una sola molécula) alternativo para determinar el tipo de una proteína y/o un dispositivo alternativo para determinar el tipo de una proteína, especialmente adecuado para uso en dicho método (de una sola molécula) alternativo, método y/o dispositivo que preferiblemente obvian al menos parcialmente, además, una o más de las limitaciones descritas anteriormente.

Por lo tanto, se propone un novedoso método de secuenciación usando una técnica de fluorescencia de una sola molécula. Este nuevo enfoque explorará proteínas, molécula a molécula, sin tomar solamente su promedio; por lo tanto, puede abarcar proteínas completas a pesar de la compleja naturaleza y el amplio intervalo dinámico de proteínas celulares. A diferencia de la secuenciación basada en espectrometría de masas, este enfoque leerá la secuencia de proteínas de longitud completa, lo que hará a la predicción de secuencias menos propensa al error. La detección de una sola molécula es tan sensible que este enfoque puede requerir solamente una pequeña cantidad de muestra (no más de 1 fmol) para el análisis de proteínas celulares. Esto creará la oportunidad de análisis de célula única. Estas ventajas contrastan con las limitaciones de la espectrometría de masas, lo que normalmente requiere 10^3 - 10^5 veces más proteínas para el análisis.

Este nuevo método puede estar marcado por tres conceptos novedosos: (1) identificación genética ("*fingerprinting*"), (2) manipulación basada en enzimas, y (3) secuenciación en tiempo real. Dado que el análisis de los inventores sugiere que la secuencia de una proteína (ya puede predecirse con la lectura de solamente dos tipos de aminoácidos (véase acerca del poder de predicción más adelante), las proteínas pueden identificarse sondeando dos aminoácidos diferentes solamente, tales como residuos de cisteína y lisina solamente. Para controlar el proceso de secuenciación con precisión nanométrica, puede aplicarse una proteína chaperona (p. ej., ClpXP), aunque también pueden aplicarse otras máquinas transportadoras moleculares adecuadas. Usando microscopía de fluorescencia de una sola molécula, se vigilarán sustratos de secuenciación individuales que están siendo sondeados mediante, p. ej., proteínas ClpXP individuales en tiempo real. Por lo tanto, la técnica novedosa es especialmente adecuada para la secuenciación de proteínas que tienen al menos 300 aminoácidos, aún más especialmente al menos 600 aminoácidos. Por lo tanto, incluso con tan solo aproximadamente 300 aminoácidos puede realizarse secuenciación (aunque en algunos casos puede ser posible un número aún menor; véase también más adelante).

Se aplica una técnica de nanoporos.

La descripción proporciona un método para analizar una proteína en un líquido (que comprende la proteína), especialmente para determinar el tipo de una proteína en un líquido que comprende la proteína, comprendiendo el método (a) funcionalizar una proteína con marcadores de aminoácidos, especialmente al menos 2, tal como 2-8, especialmente 2-4, como solamente 2, tipos de, marcadores de aminoácidos, que son selectivos para especialmente al menos 2, tal como 2-8, especialmente 2-4, como solamente 2, tipos de, aminoácidos de proteína predefinidos, especialmente solamente dos marcadores de aminoácidos selectivos para solamente dos aminoácidos, especialmente los aminoácidos C y K, (b) guiar en la fase líquida la proteína funcionalizada con una máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada a través de un área de detección de un detector, configurado para detectar una señal en función de los marcadores de los aminoácidos marcados (cuando se realiza el guiado, la proteína funcionalizada es guiada a través de un área de detección del detector (con la máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada)); (c) determinar, a partir de la señal detectada, una secuencia de los aminoácidos predefinidos; y opcionalmente (d) comparar la secuencia de los aminoácidos predefinidos con la presencia de dicha secuencia en una base de datos de proteínas y determinar el tipo de proteína en el líquido. La expresión "marcador de aminoácido" se refiere a un marcador para un aminoácido. En la presente memoria, en lugar de "marcador de aminoácido" también se aplica el término "marcador". Estos marcadores son especialmente colorantes, véase también más adelante.

Además, la descripción proporciona un dispositivo para determinar el tipo de proteína en un líquido, comprendiendo el dispositivo (a) una máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada configurada para guiar a una proteína que está funcionalizada con marcadores (es decir algunos aminoácidos (predeterminados) están funcionalizados con marcadores) a través de un área de detección de un detector, (b) dicho detector, configurado para detectar una señal en función de los marcadores de los aminoácidos marcados, (c) una unidad procesadora, configurada para identificar, a partir de la señal detectada, una secuencia de aminoácidos de la proteína funcionalizada, en donde (d) la unidad procesadora está opcionalmente configurada, además, para comparar la secuencia de aminoácidos identificada con la presencia de dicha secuencia en una base de datos de proteínas y para (basándose en la comparación) identificar el tipo de proteína. Por lo tanto, el dispositivo puede comprender una máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada configurada para guiar una proteína que está funcionalizada con marcadores de aminoácidos, que son selectivos para (especialmente 2-4 tipos de) aminoácidos de proteína predefinidos, a través de un área de detección de un detector.

Este revolucionario enfoque de una sola molécula proporcionará una novedosa herramienta de secuenciación profunda para análisis de proteínas. Se esperan diversas aplicaciones en biología, biotecnología y ciencias médicas. Cuando esta técnica se desarrolle a una herramienta de mesa en el futuro, se prevé que los investigadores médicos serán capaces de dilucidar perfiles de expresión de proteínas rastreando variaciones entre individuos, entre diferentes tejidos, y en diferentes condiciones medioambientales. Esta novedosa técnica de secuenciación puede cambiar el paradigma de la proteómica y puede convertirse en una herramienta de diagnóstico universal.

Un desafío inmediato para secuenciación de proteínas de una sola molécula es que las proteínas están compuestas por 20 aminoácidos diferentes. A diferencia de la secuenciación de ADN, que distingue solamente cuatro nucleótidos diferentes (A, G, C y T) y requiere solamente cuatro fluoróforos, la secuenciación de proteínas completas demanda

20 marcadores fluorescentes. Sin embargo, es prácticamente imposible encontrar 20 fluoróforos cuyos espectros no se solapen entre sí.

De este modo, se proporcionará un nuevo concepto, "identificación genética de proteínas", abordando la secuenciación de proteínas desde un ángulo muy diferente a la secuenciación de ADN. En primer lugar, hay que preguntarse cuánta información se necesita para la identificación de proteínas. La secuenciación de ADN requiere una lectura precisa de cada nucleótido; en caso contrario, la información obtenida carece de sentido debido a deleciones, inserciones y mutaciones. La secuenciación de proteínas, por otro lado, no requiere auténtica lectura de todos los aminoácidos. En su lugar, con referencia a bases de datos genómicas y proteómicas públicas, la secuenciación puede reducirse a un problema de identificar proteínas entre un combinado de poblaciones de proteínas (por ejemplo, ~20.000 especies en células humanas) en cada organismo.

Supongamos que solamente se leen dos tipos de aminoácidos. Con esta información de dos elementos, ¿cuántos aminoácidos se deben leer en serie para identificar una proteína? Una estimación matemática, $2^{14} < 20.000 < 2^{15}$, sugiere que solamente se necesitan leer 15 aminoácidos o más. Para convertir la información de 20 elementos de péptidos de proteínas en 2 elementos, se seleccionarán como diana los dos aminoácidos altamente nucleófilos que pueden marcarse tanto eficiente como específicamente - lisina (Lys, K) y cisteína (Cys, C). Tal como se presente en la figura 1a (véase también más adelante), una lectura de aminoácidos K y C se registrará como '...CKCKCKCKCKCK...' y estos datos se usarán para predecir la identidad de proteínas. K se observa 1 de cada 20 aminoácidos de promedio; y C, 1 de cada 40. Una proteína típica tiene ~400 aminoácidos de longitud, y de este modo contiene 30 residuos de K y C de promedio. Este número está muy por encima del número mínimo, 15 aminoácidos, que se requiere (en general) para identificación con referencia a una base de datos de proteínas (véase, sin embargo, además también más adelante).

Mientras que la predicción anterior se basa en una simple estimación matemática, una estimación del poder de predicción práctico de la identificación genética se llevó a cabo usando la base de datos de proteínas humanas Uniprot (www.uniprot.org). Se llevó a cabo un análisis de identificación genética computacional usando un proteoma completo y revisado del organismo humano (N.º 9606). Toda la información posicional procedente de la base de datos se descargó y cada secuencia de proteínas completa se redujo a las secuencias de los aminoácidos C y K solamente (una base de datos de C-K). Dada una secuencia simulada, el programa de los inventores la comparaba con todas las secuencias en la base de datos de C-K y sugería la mejor o mejores coincidencias. La fidelidad de predicción se define como la inversa del número de las mejores coincidencias. Para estos análisis computacionales, se adoptó tanto un algoritmo de coincidencia punto a punto (que se basa en el análisis de la correlación) como el algoritmo de Smith-Waterman (Smith, Temple F.; y Waterman, Michael S. (1981). "Identification of Common Molecular Subsequences". *Journal of Molecular Biology* 147: 195-197).

El análisis de los inventores demuestra que el poder de predicción alcanza un nivel satisfactorio cuando se leen aproximadamente 17 aminoácidos o más (figura 1b). Este número es ligeramente mayor que los 15 aminoácidos que se estimaron, pero mucho más pequeño que el número promedio de residuos de K y C, 30, por proteína. Por lo tanto, el método implica especialmente detectar la presencia de al menos 15, especialmente al menos 17, de los aminoácidos de proteína predefinidos en la proteína a identificar. Esto puede aplicarse especialmente a proteínas humanas. Para proteínas no humanas (y/o para proteínas no animales), puede ser que menos de al menos 15 aminoácidos amino puedan estar marcados, y seguir obteniendo una buena predictibilidad. Además, se observa que, en caso de que la técnica se use para determinar el tipo de proteína entre un número limitado de proteínas predeterminadas, también el número de al menos 15 aminoácidos puede ser menor.

Nótese que la frase "detectar la presencia de al menos 15, especialmente al menos 17, de los aminoácidos de proteína predefinidos en la proteína" no excluye al método de secuenciar y medir, además, más de solamente 15, o solamente 17 de las proteínas predeterminadas. No se excluye medir más de, p. ej., (en total) 15 aminoácidos K y C, tales como más de 20, como más de 40 aminoácidos.

Además, basándose en el principio anterior, se sugiere marcar 2-8 aminoácidos diferentes, especialmente solamente 2-4, tal como incluso solamente 2 aminoácidos diferentes, tales como especialmente Lys (K) y Cys (C). En la presente memoria, la expresión "aminoácido de proteína" y expresiones similares se usan para indicar que la invención se refiere a aminoácidos que es conocido que están disponibles en procedimientos. En lugar de esta expresión, también el término corto "aminoácidos" se aplica en la presente memoria. El término "proteína" se refiere especialmente a proteínas de origen natural, tales como proteínas humanas, animales o vegetales.

Se observa que errores de medición intrínsecos y el marcado del fluoróforo incompleto pueden interferir en la predicción, dado que conducirán a aparentes deleciones, inserciones, o intercambios de señales de Lys y Cys. El análisis de error indica que la identificación de proteínas es tolerable al -10 % de error en el marcado y la medición por fluorescencia. Por ejemplo, el análisis de los inventores (figura 4a) muestra que la fidelidad de predicción (PF) no cae significativamente incluso cuando hay un número de errores de intercambio (NSE) presentes (es decir el orden de C y K se intercambia durante la exploración) dada una secuencia CK de 40 de longitud. Se realizó este análisis de error usando tanto un algoritmo de coincidencia punto a punto como el algoritmo de Smith-Waterman.

Para leer residuos C y K de una proteína, se necesita un nano-aparato específico que despliegue la proteína desde

una conformación tridimensional complicada hasta una unidimensional lineal y la explore con precisión nanométrica. Se propone un enfoque en tiempo real basado en enzimas, usando una herramienta para leer rápidamente (tal como 0,1-60 aminoácidos/s) sin que esté implicada ninguna reacción química. En esta perspectiva, se pretende usar una proteína que existe de forma natural que se transloca a lo largo de un sustrato de proteínas con interacción estrecha. La secuenciación de ADN en tiempo real usa una enzima (ADN polimerasa) que se une a DNA y explora la cadena de ADN, acción que proporciona más información de forma natural sobre la secuencia de ADN.

ClpXP (un complejo de proteínas de ClpX y ClpP) es un complejo de chaperona que despliega un sustrato de proteínas y se transloca a lo largo de éste (figura 2a). Es una enzima procesiva y rápida que se mantiene en contacto estrecho con su sustrato. Una única ClpXP puede translocarse a lo largo de monómeros concatenados (ocho monómeros de 20 kD, >1000 aminoácidos de longitud). ClpXP puede translocarse a hasta 60 aminoácidos por segundo (Maillard, R.A., Chistol, G., Sen, M., Righini, M., Tan, J., Kaiser, C.M., Hodges, C, Martin, A., y Bustamante, C. (2011). ClpX(P) genera fuerza mecánica para desplegar y translocar sus sustratos de proteínas. *Cell* 145, 459-469). ClpX tiene un canal de anchura nanométrica, y ClpP tiene poros de anchura nanométrica, cuyas dimensiones proporcionan la capacidad de manipular un sustrato de secuenciación con precisión nanométrica. Las propiedades bioquímicas de ClpXP tales como especificidad del sustrato y resistencia a modificaciones de proteínas están bien caracterizadas. ClpXP puede translocar y degradar sustratos modificados tales como sustratos desnaturalizados químicamente, sustratos marcados con colorantes, y proteínas reticuladas químicamente. ClpXP parece altamente indiscriminado en el procesamiento del sustrato. Se ha descubierto que la modificación de sustratos de secuenciación no interferirá en la actividad de ClpXP. La secuenciación de una sola molécula también puede requerir modificación de ClpP o ClpX. La mutación y el marcado con colorante de ClpP y ClpX parecen ser factibles.

Sin embargo, también pueden aplicarse otras proteasas dependientes de ATP. En una realización, la máquina transportadora molecular es una máquina transportadora molecular seleccionada entre el grupo que consiste en entre otros una ClpXP, una ClpAP, una ClpCP, una ClpEP, una ClpYQ (una HslUV), una ClpB, una Lon, una FtsH, una PAN de arqueas y una máquina transportadora molecular basada en el proteasoma (véase también Kirstein, J., Molliere, N., Dougan, D.A., y Turgay, K., *Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases*, *Nature Reviews - Biology*, 7 de agosto de 2009, 589-599), especialmente una máquina transportadora molecular basada en ClpXP. Sin embargo, también pueden aplicarse otras máquinas transportadoras moleculares (proteasa dependiente de ATP). Estas proteasas dependientes de ATP comparten una estructura básica y un mecanismo de acción comunes, en los que la enzima AAA+ puede seleccionar sustratos de proteínas y se transloca a lo largo de ellos de forma procesiva, mientras que la proteasa puede degradar los sustratos de proteínas. En una realización específica, se aplica ClpXP. En otra realización más, se aplica ClpAP.

Usando ClpXP como sonda de exploración, pueden aplicarse técnicas de fluorescencia de la siguiente manera. En primer lugar se necesita marcar sustratos de secuenciación con colorantes. Se seleccionan, aminoácidos altamente nucleófilos, lisina (K) y cisteína (C). El grupo amino de K se conjugará con colorante éster de NHS y el grupo tiol de C con colorante maleimida. Las dos reacciones son ortogonales entre sí, lo que impide el marcado cruzado. Estas reacciones químicas se producen de forma tan eficiente, alcanzando ~100 % de rendimiento en una condición de reacción general (una concentración micromolar de colorantes reactivos y varias horas de incubación), que varios kits de marcado están disponibles en el mercado. Esto hace al procedimiento de marcado eficaz en tiempo y coste. Para garantizar el completo marcado de residuos C y K, se expondrán aminoácidos internos a través de desnaturalización de proteínas. En lugar o además de lisina y cisteína, también pueden seleccionarse serina, treonina, tirosina, y aminoácidos modificados de forma postraduccional. Por lo tanto, especialmente dos o más de lisina, cisteína, serina, treonina, tirosina, y aminoácidos modificados de forma postraduccional pueden seleccionarse para ser marcados, especialmente dos o más de lisina, cisteína, serina, treonina, y tirosina. En una descripción, solamente se marcan cisteína y lisina.

Entre varios métodos de desnaturalización, puede emplearse desnaturalización mediada por SDS (dodecilsulfato sódico) e inducida por calor, que es eficaz para alterar estructuras secundaria y terciaria de proteínas. Para romper puentes disulfuro, pueden usarse reactivos reductores fuertes (tales como β -mercaptoetanol), en esta condición de desnaturalización rigurosa. En la posterior etapa de marcado con colorante, reactivos reductores se eliminarán preferiblemente dado que pueden interferir en el marcado (con cisteína). Esta eliminación puede llevarse a cabo p. ej., con un procedimiento de precipitación de proteínas general (con acetona, sulfato de amonio, o polietiliminina). Después del marcado con colorante, puede llevarse a cabo un intercambio de tampón (a través de diálisis o una columna de exclusión por tamaño) o una ronda adicional de precipitación de proteínas para eliminar colorantes en exceso. Estas etapas de purificación pueden garantizar especialmente que las reacciones de secuenciación mediadas por enzimas discurren en condiciones biológicamente óptimas.

Como resultado del marcado, se entremezclarán colorantes a lo largo de un sustrato de secuenciación sobre, en general, unos pocos nanómetros de distancia. Para resolverlos en orden, se necesita un método de imaginología de resolución nanométrica. En este caso se presenta una regla nanométrica, FRET (transferencia de energía de resonancia por fluorescencia entre fluoróforos donador y aceptor) (Roy, R., Hohng, S., y Ha, T. (2008). *A practical guide to single-molecule FRET*. *Nat Methods* 5, 507-516). Más adelante, también se presenta un filtro de nanoporos, que puede usarse para resolver el orden de los marcadores. Con residuos K y C marcados con dos colores diferentes de colorantes aceptores respectivamente (Cy5 y Cy7), las moléculasceptoras pueden sondearse mediante exploración con una molécula donadora de Cy3 y midiendo FRET de señales de fluorescencia de Cy5 y

Cy7 con Cy3. Otros colorantes aceptores tales como Cy3.5 y Cy5.5 (para otros aminoácidos) pueden sondearse de manera análoga. Por lo tanto, los marcadores pueden comprender un fluoróforo orgánico seleccionado p. ej., entre uno o más de la familia de cianina, la familia Alexa, la familia Atto, la familia Dy, y la familia de rodamina, aunque otros fluoróforos no están excluidos.

5 Tal como se ilustra en la figura 2b, ClpX (un anillo hexamérico, mostrado como una elipse) tiene un puerto de entrada en la parte superior, y ClpP (un dímero de anillos heptaméricos) tiene un supuesto puerto de salida en la parte inferior. Estas dos proteínas, unidas estrechamente, son de aproximadamente 16 nm de altura y de aproximadamente 9-15 nm de anchura. Cuando un sustrato de secuenciación se acopla sobre el puerto de entrada de ClpX mediante su marcador, el sustrato se transloca al interior del canal estrecho (1 nm de ancho) de ClpX y se
10 vuelve desplegado. La proteína desplegada es transferida a su proteína socia, la proteasa de ClpP, que a continuación escinde el péptido translocado en fragmentos pequeños.

ClpXP reconoce sustratos que muestran cierta marca específica solamente. Una marca bien conocida, λ O (TNTAKILNFGR) (Farell, CM., Baker, T.A. y Sauer, R.T., *Altered specificity of a AAA+ protease*, *Molecular Cell* 25, 161-166, 12 de enero de 2007, 161-166-), puede ligarse en el extremo N de sustratos de secuenciación. Puede
15 usarse química de EDC para conjugar el extremo C (grupo carboxilo) del péptido λ O sintético con el extremo N (grupo amino) de los sustratos de secuenciación. Dado que el grupo amino de Lys también es dirigido por química de EDC, este ligamiento requiere consideraciones cuidadosas. En primer lugar, dado que el valor pK del grupo amino alfa (pKa = 8,9) es menor que el del grupo amino épsilon de Lys (pKa = 10,5), la química de EDC puede llevarse a cabo a un pH de 6,5 - 8,5. En segundo lugar, la reacción de ligamiento puede llevarse a cabo después del
20 marcado de Lys a pH 9,0-11,0, para minimizar cualquier ligamiento inespecífico. En lugar de la marca λ O N-terminal (TNTAKILNFGR), también puede aplicarse una marca C-terminal ssrA (AANDENYALAA), o cualquier otra marca N- o C-terminal (véase también Flynn J.M., Neher, S.B., Kim, Y, Sauer, R.T., Baker, A.T., *Proteomic Discovery of Cellular Substrates of the ClpXP Protease Reveals Five Classes of ClpX-Recognition Signals*, *Molecular Cell*, Vol.11, marzo de 2003, 671- 683). Por lo tanto, el método comprende además marcar la proteína con una marca (o
25 marcador) que es reconocible por la máquina transportadora molecular.

El radio de la cámara de ClpP es ~5 nm. Esta dimensión es óptima para aprovechar la sensibilidad de FRET, dado que el par Cy3-Cy5 es el más sensible a 6 nm y el Cy3-Cy7 a 4 nm. Si se coloca el donador (Cy3) en una cámara de ClpP, se puede usar esta distancia óptima para medir FRET entre el donador (Cy3) y los aceptores (Cy5 y Cy7) de un fragmento peptídico. Además, de esta manera puede evitarse FRET inespecífica entre el donador y un
30 combinado de aceptores cerca del puerto de entrada, dado que la distancia entre el puerto de entrada y la cámara de ClpP es 7,5-12,5 nm, mayor que distancias FRET normales (que son del orden de 4-6 nm). Como alternativa, ClpX puede marcarse con un fluoróforo donador de Cy3. Además, Cy3, Cy5 y Cy7 se refieren a colorantes de cianina, conocidos en la técnica, tal como p. ej., se describen en Lee et al, 2010 (Lee, J., Lee, S., Ragunathan, K., Joo, C, Ha, T., y Hohng, S. (2010). *Single-molecule four-color FRET*. *Angew. Chem Int. Ed. Engl.* 49, 9922-9925).
35 Sin embargo, también pueden aplicarse otros marcadores, seleccionados entre una o más de la familia de cianina, la familia Alexa, la familia Atto, la familia Dy, y la familia de rodamina.

Se producirá FRET desde el donador con cualesquiera moléculasceptoras dentro de una cámara de ClpP. De este modo, para obtener rastros temporales de FRET de alta calidad, es esencial tener los menos colorantes posibles dentro de una cámara en un momento. Para conseguir esto, la técnica de secuenciación utiliza la característica
40 única de ClpP. ClpP escinde una proteína en fragmentos y libera cada fragmento fuera de la cámara. Especialmente cuando se ralentiza la velocidad de translocación de ClpX lo suficiente (tal como reduciendo la cantidad de energía de ATP disponible) (Martin, A., Baker, T.A., Sauer, R.T., *Protein unfolding by a AAA+ protease is dependent on AT-hydrolysis rates and substrate energy landscapes*, *Nature Structural & Molecular Biology*, volumen 15, no. 2, febrero de 2008, 139-145; Shin, Y, Davids, J.H., Brau, R.R., Martin, A., Kenniston, J.A., Baker, Sauer, R.T., Lang, M.J.,
45 *Single-molecule denaturation and degradation of proteins by the AAA+ ClpXP protease*, *PNAS*, 17 de noviembre de 2009, vol. 106, no. 46, 19340-19345), se producirá solamente una reacción de escisión que se produce dentro de la cámara de ClpP cada vez, que es seguida por la liberación regida por difusión del fragmento único. Este esquema de translocación controlada mantendrá el número de fragmentos dentro de una cámara inferior a uno de promedio, lo que permitirá interpretar un rastro temporal de FRET con ambigüedad mínima. De forma relacionada, este ensayo
50 garantizará también que el orden de la liberación de fragmentos sigue el mismo orden de la secuencia de proteína original, lo que minimiza cualesquiera errores de intercambio en la lectura (C \leftarrow \rightarrow K).

El marcado de ClpX o ClpP con donador y biotina puede llevarse a cabo conjugando el grupo tiol de cisteína con colorante maleimida-Cy3 y maleimida-biotina, respectivamente. Tanto ClpX como ClpP de *E. coli* tienen dos residuos
55 cisteína por monómero. Dado que ambos aminoácidos no son conservados entre especies bacterianas, se pueden inactivar ambos de ellos e introducir nuevos residuos de cisteína donde deben posicionarse Cy3 y biotina. Las estructuras cristalinas de la proteína ClpX y ClpP se utilizarán cuando se asigne la posición de la mutación puntual de cisteína.

Una cámara de ClpP está compuesta por 14 monómeros de ClpP, lo que plantea dos problemas prácticos. En primer lugar, debido a esta estructura oligomérica, la cámara puede estar marcada con más de una molécula
60 donadora. Dado que solamente pueden analizarse señalesceptoras, los múltiples colorantes donadores no interferirán en la medición de los inventores. De hecho, se puede añadir intencionadamente más de una molécula

donadora para prolongar el tiempo de observación total. En segundo lugar, debido a la simetría puntual de la cámara de ClpP, el posicionamiento aleatorio de una molécula donadora puede dar como resultado que ésta se coloque en un lado no deseado, es decir en la interfaz entre ClpX y ClpP en lugar de adyacente al puerto de salida. Este problema puede resolverse generando una cámara de ClpP asimétrica, en la que uno de los anillos está compuesto por ClpP mutante que no puede interactuar con ClpX. Se marcará con colorante solamente esta ClpP mutante antes de ensamblar la cámara. Puede aplicarse el procedimiento tal como se describe por Maglica et al. (Maglica, Z., Kolygo, K., y Weber-Ban, E., Optimal efficiency of ClpAP and ClpXP chaperone-proteases is achieved by architectural symmetry, *Structure* 17, 2009, 508-516.). Por lo tanto, especialmente la máquina transportadora molecular es una máquina transportadora molecular basada en ClpXP, en donde la máquina transportadora molecular basada en ClpXP comprende una cámara de ClpP asimétrica de monómeros de ClpP y al menos un monómero de ClpP mutante, en la que el al menos un monómero de ClpP mutante no puede acoplarse a ClpX, y en donde esta al menos una ClpP mutante está marcada con donador fluorescente (antes de ensamblar la cámara). Se observa que, en el caso de que ClpX esté marcada con un fluoróforo donador, no es necesario crear una proteína ClpP asimétrica. Por lo tanto, en otra divulgación, la máquina transportadora molecular es una máquina transportadora molecular basada en ClpXP, en donde la ClpX está marcada con un fluoróforo donador.

Por lo tanto, más en general, la máquina transportadora molecular puede comprender un donador, especialmente un donador fluorescente (fluoróforo donador), donador que puede estar especialmente configurado para sondear un marcador de aminoácido. Esto puede incluir una ClpXP marcada con donador fluorescente o una ClpAP marcada con donador fluorescente, etc. ClpX o ClpA, etc., puede marcarse. Sin embargo, opcionalmente además de o como alternativa, ClpP puede marcarse con el donador fluorescente (fluoróforo donador). Opcionalmente, pueden aplicarse dos o más marcadores.

Para observar eventos de secuenciación con fluorescencia de una sola molécula, se pueden inmovilizar proteínas de ClpXP sobre una superficie de cuarzo y se obtendrán imágenes con microscopía TIRF (fluorescencia por reflexión interna total). La inmovilización puede llevarse a cabo marcando proteínas ClpX o ClpP con biotina y presentándolas sobre una superficie de cuarzo estratificada con estreptavidina (figuras 3a-3c). Las reacciones de secuenciación comenzarán cuando sustratos de secuencia marcados con colorante se introducen en una cámara de una sola molécula, especialmente mediante flujo laminar, y se acoplan sobre las proteínas ClpXP inmovilizadas mediante difusión dentro de la cámara.

Trayectorias temporales de FRET consecuentes pueden obtenerse a resolución temporal elevada (hasta 10 milisegundos) con una cámara CCD del estado de la técnica (tal como p. ej., Andor, iXon, CCD con multiplicación de electrones). Para la mejor relación de señal con respecto a ruido, un trio de colorantes de cianina (Cy3, Cy5 y Cy7) (Lee, J., Lee, S., Ragunathan, K., Joo, C, Ha, T., y Hohng, S. (2010). Single-molecule four-color FRET, *Angew. Chem Int. Ed. Engl.* 49, 9922-9925) pueden usarse como pares de FRET. Para foto-parpadeo mínimo y foto-blanqueo lento, pueden usarse un sistema secuestrador de oxígeno (glucosa oxidasa y catalasa) y un extintor en estado de triplete (Trolox). Tal como se prefiere para prevenir adsorción inespecífica de proteínas sobre una superficie, una superficie de cuarzo se revestirá con polímero (PEG, polietilenglicol). Sin embargo, también pueden aplicarse otras superficies. Por ejemplo, pueden aplicarse superficies revestidas con albúmina de suero bovino o caseína. En el caso en el que en lugar de una señal óptica, se evalúe una señal eléctrica (como en el caso de un sistema de nanoporos), la superficie también puede estar revestida con proteínas o bicapas de lípidos.

La translocación de ClpX debe ser suficientemente rápida para que la secuenciación se complete antes de que la moléculas donadoras se foto-blanqueen. Un reciente estudio de una sola molécula sugiere que la velocidad de ClpX es de 60 aminoácidos por segundo. Esta tasa se convierte en 6,7 segundos por secuenciación (de un sustrato de proteínas de tamaño promedio, ~400 aminoácidos). Esta ventana temporal está bien dentro de la escala temporal de observación de Cy3, normalmente unos pocos minutos cuando se usa una resolución temporal de 0,1 s. Por otro lado, la translocación de ClpX debe ser suficientemente lenta para estadísticas de fotón adecuadas y análisis fiable de rastros temporales. Cuando se necesita ralentizar el proceso, se necesita solamente rebajar la concentración de ATP (de milimolar a micromolar) debido a que la translocación de ClpX es dependiente de energía (ATP).

Por lo tanto, en una descripción, la proteína funcionalizada se transloca con la máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada con una velocidad de translocación a través del área de detección del detector, en donde la velocidad de translocación se selecciona entre el intervalo de 0,1-60 aminoácidos por segundo. Especialmente, la velocidad de translocación se controla controlando una concentración de ATP en el líquido.

Por lo tanto, en una realización, con una máquina transportadora molecular (basada en proteasa dependiente de ATP) inmovilizada, la proteína funcionalizada (en la fase líquida) es guiada a través de un área de detección de un detector.

Debido a la baja afinidad de asociación de los monómeros de ClpX, se desea usar una forma hexamérica enlazada artificialmente de ClpX para un estudio de una sola molécula. Se ha expresado y se ha purificado esta forma hexamérica de ClpX. ClpP, por otro lado, tiene una elevada afinidad de asociación, y de este modo se ha expresado ClpP como monómeros.

5 ClpX y ClpP pueden formar un complejo 1:2 o 2:2 (ClpX:ClpP). Para la secuenciación de proteínas, se puede usar especialmente el complejo 1:2, para evitar el procesamiento de múltiples sustratos simultáneamente mediante el mismo complejo. Esta estequiometría puede alcanzarse de forma fiable usando una relación superestequiométrica, tal como relación al menos 3:1 ratio entre concentraciones de tetradecámero de ClpP y hexadecámero de ClpX en mezcla.

10 Las proteínas desplegadas tienden a agregarse. Para impedir que esto ocurra con sustratos de secuenciación, los sustratos marcados desplegados pueden guardarse en un desnaturalizante. Nótese que este desnaturalizante puede no interferir en la reacción de secuenciación debido a que los sustratos desplegados se introducirán en una cámara de secuenciación a concentración ~1 nM mediante dilución en un tampón fisiológico, y esto dará como resultado la baja concentración de desnaturalizante. Para minimizar adicionalmente cualquier agregación, se pueden diluir rápidamente sustratos de secuenciación inmediatamente antes de una reacción de secuenciación, después de un procedimiento optimizado por Meyer (Meyer, A.S., Gillespie, J.R., Walther, D, Millet, IS., Doniach, S., y Frydman, J. (2003). Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. Cell 113, 369-381).

20 Por la presente se proporciona una demostración del proceso de exploración basado en ClpXP usando la técnica FRET de una sola molécula (figura 4b). Se generó una forma hexamérica enlazada artificialmente de ClpX y se biotiniló su extremo C-terminal. Se inmovilizó esta proteína de nanocanal, en complejo con la proteasa ClpP, sobre una superficie de cuarzo usando conjugación de estreptavidina-biotina. En esta cámara de muestra, se añaden péptidos que contenía un residuo de K y uno de C. El residuo de K se marcó con un fluoróforo Cy3 que contenía un grupo éster de NHS (esfera verde en la figura), y el residuo de C se marcó con un fluoróforo Cy5 que contenía un grupo mono-maleimida (esfera roja). El péptido contenía la marca ssrA que es reconocida por ClpX con especificidad elevada.

25 Se obtuvieron imágenes del proceso de translocación del péptido a través del nanocanal de ClpX usando un microscopio de reflexión interna total construido en laboratorio y se registraron señales de fluorescencia usando una cámara CCD con multiplicación de electrones. Un súbito incremento de la señal de fluorescencia (tiempo a 25,5 s en la figura Y) indica el acoplamiento de un péptido a un escáner de ClpXP inmovilizado. FRET entre Cy3 y Cy5 es inicialmente eficiente debido a la estructura plegada del péptido designado, tal como se muestra mediante la elevada fluorescencia del aceptor (el rastro temporal del donador es una línea verde, y el aceptor en rojo). Cuando el péptido es arrastrado por ClpX y es translocado a través del nanocanal de la ClpX mediante hidrólisis de ATP, se estira linealmente, la distancia entre Cy3 y Cy5 se vuelve mayor, y la eficiencia de FRET se vuelve menor (tiempo a 26,2 s). Este proceso de translocación viene seguido por el replegamiento del péptido dentro de la cámara de ClpP (tiempo a 26,5 s) y sus eventos de escisión y posterior disociación (tiempo a 27,0 y 27,6 s).

30 Por lo tanto, tal como se describió anteriormente, la divulgación puede proporcionar, en una realización, un método para analizar una proteína en un líquido que comprende la proteína, especialmente para determinar el tipo de una proteína en un líquido que comprende la proteína, comprendiendo el método (a) funcionalizar una proteína con 2-4 tipos de marcadores de aminoácidos, que son selectivos para 2-4 tipos de aminoácidos de proteína predefinidos, especialmente solamente dos marcadores de aminoácidos selectivos para solamente dos aminoácidos, especialmente los aminoácidos C y K, (b) guiar en la fase líquida la proteína funcionalizada con una máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada a través de un área de detección de un detector, configurado para detectar una señal en función de los marcadores de los aminoácidos marcados; y (c) determinar, a partir de la señal detectada, una secuencia de los aminoácidos predefinidos. Especialmente, este método de análisis implica, además, (d) comparar la secuencia de los aminoácidos predefinidos con la presencia de dicha secuencia en una base de datos de proteínas y determinar el tipo de proteína en el líquido. De esta manera, puede determinarse el tipo de proteína. Dicha base de datos puede ser una base de datos remota. Por ejemplo, la secuencia descubierta puede compararse con datos de Internet sobre aminoácidos secuenciados conocidos.

40 Especialmente, el método es un método *ex vivo*. El líquido mencionado anteriormente puede ser un líquido corporal, pero también puede ser un líquido corporal diluido. Además, también son concebibles otros líquidos, tales como extractos de células y orgánulos de bacterias, arqueas, eucariotas. Especialmente, el líquido es un líquido acuoso.

45 Como también se indicó anteriormente, la descripción también proporciona un dispositivo para determinar el tipo de proteína en un líquido, comprendiendo el dispositivo (a) una máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada, configurada para guiar una proteína que está funcionalizada con marcadores a través de un área de detección de un detector, (b) dicho detector, configurado para detectar una señal en función de los marcadores de los aminoácidos marcados, (c) una unidad procesadora, configurada para identificar a partir de la señal detectada una secuencia de aminoácidos de la proteína funcionalizada, y opcionalmente configurada, además, para comparar la secuencia de aminoácidos identificada con la presencia de dicha secuencia en una base de datos de proteínas y para identificar el tipo de proteína.

5 Especialmente, la unidad procesadora puede estar configurada para comparar la secuencia de los aminoácidos de proteína predefinidos de la proteína con la presencia de dicha secuencia en una base de datos de proteínas y determinar el tipo de proteína en el líquido, en donde la proteína está funcionalizada con 2-8, tal como especialmente solamente 2-4 tipos de marcadores de aminoácidos, que son selectivos para 2-8, especialmente 2-4 tipos de aminoácidos de proteína predefinidos. Por lo tanto, la unidad procesadora puede estar configurada para determinar, basándose en solamente 2-8, especialmente 2-4 tipos de aminoácidos marcados, es decir 2-8, especialmente 2-4 tipos de diferentes marcadores selectivos para aminoácidos, el tipo de proteína cuando los aminoácidos de proteína respectivos (o al menos parte del número total de los mismos) están marcados con estos marcadores selectivos. Incluso solamente 2 tipos de marcadores, p. ej., para lisina y cisteína, pueden ser suficientes.

10 El marcador, para uso en el método, puede usarse para métodos de análisis basados en fluorescencia o el marcador puede usarse para un método de análisis basado en señales electrónicas (véase a continuación cuando se describe en más detalle el método de nanoporos). En principio, los aceptores fluorescentes tal como se describió anteriormente, también pueden aplicarse en el método de nanoporos, aunque también pueden aplicarse otros marcadores en el último método, tales como perlas de oro, puntos cuánticos (“*quantum dots*”), y otras nanopartículas en estado sólido. Por lo tanto, los marcadores comprenden aceptores fluorescentes, en donde un donador fluorescente, configurado para formar temporalmente un par donador-aceptor con uno de los aceptores fluorescentes, está configurado dentro del área de detección, y en donde el detector comprende un microscopio de fluorescencia que incluye un microscopio de fluorescencia por reflexión interna total (TIRF), un microscopio de fluorescencia confocal, o un microscopio de fluorescencia a base de guía de ondas en modo cero. Por lo tanto, especialmente el detector comprende un microscopio de fluorescencia que incluye un microscopio de fluorescencia por reflexión interna total (TIRF), un microscopio de fluorescencia confocal, o un microscopio de fluorescencia a base de guía de ondas en modo cero. En una descripción específica, máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada comprende un donador fluorescente unido a ella. Aún más especialmente, véase también anteriormente, la máquina transportadora molecular es una máquina transportadora molecular basada en ClpXP, en donde la máquina transportadora molecular basada en ClpXP comprende una cámara de ClpP asimétrica de monómeros de ClpP y al menos un monómero de ClpP mutante, en la que al menos un monómero de ClpP mutante no puede interactuar con ClpX, y en donde este al menos una ClpP mutante está marcado con donador fluorescente, o en donde la máquina transportadora molecular basada en ClpXP comprende una proteína ClpX que comprende un marcador de donador fluorescente por hexámero. Como se mencionó anteriormente, pueden aplicarse marcadores p. ej., seleccionados entre una o más de la familia de cianina, la familia Alexa, la familia Atto, la familia Dy y la familia de rodamina, etc.

35 Sin embargo, en otra descripción más, el método comprende guiar la proteína funcionalizada con la máquina transportadora molecular a través de un nanoporo de un filtro que comprende nanoporos que tiene un lado nano-transportador y un lado opuesto, en donde el detector comprende una unidad detectora configurada para medir un parámetro eléctrico entre el lado nano-transportador y el lado opuesto del filtro que comprende nanoporos, y en donde el parámetro eléctrico se selecciona entre el grupo que consiste en una diferencia de potencial, una corriente y la resistencia. Por lo tanto, el dispositivo, como se describe en la presente memoria, puede comprender además un filtro que comprende nanoporos que tiene un lado nano-transportador y un lado opuesto y en donde el detector está configurado para medir un parámetro eléctrico entre el lado nano-transportador y el lado opuesto del filtro que comprende nanoporos, en donde el parámetro eléctrico se selecciona entre el grupo que consiste en una diferencia de potencial, una corriente y la resistencia, en donde el dispositivo está configurado además para guiar a la proteína que está funcionalizada con marcadores a través del nanoporo durante el uso del dispositivo. Los poros de los nanoporos pueden tener diámetros en el intervalo de 0,1-10 nm. La anchura del filtro, es decir la longitud del canal del nanoporo puede estar en el intervalo de 0,3-10 nm.

45 En lugar o además de determinar el orden de los aminoácidos (marcados), el intervalo de tiempo entre los aminoácidos (marcados) (también) puede determinarse. A partir de esta información (y la velocidad de secuenciación de la máquina transportadora molecular), la distancia entre aminoácidos (marcados) adyacentes puede determinarse. Esta información puede usarse (además) para determinar el tipo de proteína (en el líquido). Por lo tanto, a partir de la señal del sensor, también puede determinarse un parámetro de distancia entre AA marcados adyacentes. Este parámetro de distancia también puede aplicarse para determinar adicionalmente el tipo de proteína (en el líquido).

55 En la presente memoria, se aplica especialmente un sistema de transporte molecular para translocar un sustrato de secuenciación con precisión nanométrica y también para ralentizar todo el proceso de translocación. La comunidad de fluorescencia y nanoporos ha estado esperando una estrategia de secuenciación de proteínas de una sola molécula durante décadas. A pesar de toda la bibliografía existente sobre sistemas de transporte molecular y el desarrollo de nuevas técnicas de una sola molécula, nadie de la comunidad ha conseguido ningún diseño similar al nuestro. Por ejemplo, en la técnica anterior no se conoce o se sugiere ningún marcado con colorante de ClpXP u otros transportadores.

60 El término “sustancialmente” en la presente memoria, tal como en “sustancialmente toda la emisión” o en “consiste sustancialmente”, será entendido por el experto en la materia. El término “sustancialmente” también puede incluir realizaciones con “totalmente”, “completamente”, “todo”, etc. Por lo tanto, en realizaciones el adjetivo sustancialmente también puede eliminarse. Donde sea aplicable, el término “sustancialmente” también puede

referirse al 90 % o superior, tal como el 95 % o superior, especialmente el 99 % o superior, aún más especialmente el 99,5 % o superior, incluyendo el 100 %. El término "comprenden" incluye también realizaciones en donde el término "comprende" significa "consiste en".

5 Además, los términos primero, segundo, tercero y similares en la descripción y en las reivindicaciones, se usan para distinguir entre elementos similares y no necesariamente para describir un orden secuencial o cronológico. Debe entenderse que los términos usados de este modo son intercambiables en circunstancias apropiadas y que las realizaciones de la descripción descritas en la presente memoria son capaces de funcionamiento en secuencias diferentes a las descritas o ilustradas en la presente memoria.

10 Los dispositivos o aparatos en la presente memoria se describen, entre otros, durante el funcionamiento. Como quedará claro para el experto en la materia, la invención no está limitada a métodos de funcionamiento o dispositivos en funcionamiento.

15 El uso del verbo "comprender" y sus conjugaciones no excluye la presencia de elementos o etapas diferentes de las indicadas en una reivindicación. El artículo "un" o "una" precediendo a un elemento no excluye la presencia de una pluralidad de dichos elementos. La invención puede implementarse por medio de hardware que comprende varios elementos distintos, y por medio de un ordenador programado adecuadamente. En la reivindicación del dispositivo que enumera varios medios, varios de estos medios pueden estar materializados por uno y el mismo artículo de hardware. El mero hecho de que ciertas medidas se enumeren en reivindicaciones dependientes mutuamente diferentes no indica que una combinación de estas medidas no pueda usarse para su conveniencia.

20 La invención se aplica, además, a un aparato o dispositivo que comprende uno o más de los elementos caracterizantes descritos en la descripción y/o mostrados en los dibujos adjuntos. La invención se refiere, además, a un método o proceso que comprende uno o más de los elementos caracterizantes descritos en la descripción y/o mostrados en los dibujos adjuntos.

Los diversos aspectos descritos en esta patente pueden combinarse para proporcionar ventajas adicionales. Además, algunos de los elementos pueden formar la base para una o más solicitudes divisionales.

25 **Breve descripción de los dibujos**

A continuación se describirán realizaciones de la invención, a modo de ejemplo solamente, con referencia a los dibujos esquemáticos adjuntos en los que símbolos de referencia correspondientes indican partes correspondientes, y en los que:

Las figuras, 1a-1b representan algunos aspectos de un principio de la invención;

30 Las figuras 2a-2b representan esquemáticamente algunos aspectos de una realización del transportador molecular basado en proteasa dependiente de ATP, que en este caso es una ClpXP;

Las figuras 3a-3c representan esquemáticamente algunas realizaciones y variantes del método y dispositivo de la invención; y

Las figuras 4a y 4b muestran algunas simulaciones y resultados de experimentación.

35 Los dibujos no son necesariamente a escala.

Descripción detallada de las realizaciones

40 La figura 1a representa esquemáticamente una proteína, tal como una enzima. La proteína se indica con la referencia 100. La proteína 100 consiste esencialmente en una cadena de aminoácidos 110. Los aminoácidos 110 que están marcados, se indican con las referencias 111. La referencia L1 indica un primer marcador y la referencia L2 se refiere a un segundo marcador (es decir aminoácidos con el marcador L1 y L2, respectivamente). Como se indicó anteriormente, especialmente dos aminoácidos pueden estar marcados, tal como lisina (a partir de 1 de 20 aminoácidos) y cisteína (aproximadamente 1 de 40 aminoácidos). Como se indicó anteriormente, una proteína puede identificarse a partir del orden de, p. ej., residuos de C (Cys) y K (Lys) solamente.

45 La figura 1b muestra un cálculo sobre el poder de predicción del método sugerido. La fidelidad de predicción es proporcionar al número de residuos de C y K en un sustrato de secuenciación. La identificación genética se vuelve fiable cuando el número es mayor de 15, tal como al menos 16, como especialmente al menos 17. Por encima de 25, el poder de predicción es del 100 %. Nótese que el número de al menos 15 se refiere al número de diferentes aminoácidos marcados con diferentes marcadores. Por lo tanto, 6 aminoácidos C marcados y 12 aminoácidos K marcados hacen 18 aminoácidos marcados, lo que dará un elevado poder predictivo. Por lo tanto, después de medir
50 al menos del orden de 15-25 aminoácidos marcados, la determinación puede ser concluyente.

La figura 2a representa esquemáticamente la enzima ClpXP 200, con ClpX (ATPasa), estando indicada con la referencia 210 y con ClpP (proteasa) estando indicada con la referencia 220. Los monómeros individuales se indican con las referencias 211 (ClpX) y 221 (ClpP), respectivamente. La figura 2b muestra una vista de sección transversal

de ClpXP. Un sustrato de secuenciación 100 marcado con Cy5 y Cy7, es decir la proteína 100 (que tiene aminoácidos marcados, en este caso de nuevo marcado con los marcadores L1 y L2), es arrastrado al interior del canal estrecho 212 de ClpX 210. El sustrato estirado es entregado a ClpP 220 donde es digerido a un fragmento pequeño (~7 aminoácidos). Habrá FRET entre Cy3 (donador, indicado con la referencia 50) cerca del puerto de salida de ClpP y Cy5 y Cy7 en un fragmento dentro de la cámara de ClpP (véase a continuación).

La altura de ClpX 210 se indica con la referencia H1 (aproximadamente 7,5 nm); la anchura se indica con la referencia d2 (aproximadamente 15 nm). El diámetro del canal del canal 212 se indica con d1 (aproximadamente 1 nm). La altura de ClpP 220 se indica con H2 (aproximadamente 9 nm); la anchura (no indicada) es también de aproximadamente 9 nm; la anchura de la cámara 223 se indica con el diámetro d3 (que es de aproximadamente 5 nm).

La figura 3a a la izquierda muestra una realización de inmovilización de la enzima ClpXP 200 en una superficie 300, tal como cuarzo. La inmovilización se lleva a cabo, p. ej., marcando proteínas (enzimas) ClpP 200 con biotina 302 y presentándolas sobre una superficie estratificada con estreptavidina, tal como cuarzo. La estreptavidina se indica con la referencia 303. La superficie del sustrato puede revestirse previamente con un polímero, indicado con la referencia 301, tal como PEG (polietilenglicol). En el lado izquierdo de la figura 3a, se indica esquemáticamente el proceso (con la flecha). PEG puede impedir la adsorción inespecífica de proteínas (ClpXP y sustratos de secuenciación) a una superficie. Si PEG, ClpXP puede perder su función, y las señales procedentes de sustratos de secuenciación pueden aparecer de forma inespecífica en una pantalla CCD. PEG también puede proporcionar biotina a la que se une estreptavidina.

La figura 3b representa esquemáticamente un dispositivo (de detección) basado en FRET para uso, p. ej., en el método de la invención. El dispositivo se indica con la referencia 400. Las referencias 431 y 432 indicaban láseres, respectivamente, tales como un láser de 532 nm y 633 nm, respectivamente. La referencia 434 es un espejo, la referencia 433 es un espejo transmisivo (tal como un espejo dicróico). La referencia 433 puede estar compuesta por más de un espejo. La referencia 435 es un espejo, y la referencia 436 es una lente, con p. ej., una distancia focal de 100 mm, la referencia 440 indica un microscopio TIRF (fluorescencia por reflexión interna total) (véase el aumento). La referencia 441 indica una hendidura y la referencia 442 indica una lente, con p. ej., una distancia focal de 100 mm, la referencia 443 indica un espejo dicróico (especialmente para el haz donador), y la referencia 444 indica un espejo (especialmente para el haz aceptor). La referencia 445 es de nuevo una lente, con una distancia focal de p. ej., 150 mm; la referencia 446 indica un espejo, y la referencia 447 indica un detector, tal como un EM-CCD (CCD con multiplicación de electrones). La referencia 402 indica un prisma de Pellin-Broca y la referencia 300 indica la superficie, en este caso una placa de cuarzo. Sobre la superficie 300, con respecto a la figura 3a descrita, está presente ClpXP 200 inmovilizada. El campo evanescente se indica con la referencia 401. La referencia 404 indica una cubierta transparente, tal como un cubreobjetos de vidrio. La referencia 405 indica un medio transparente, tal como agua. La referencia 406 indica un filtro de paso de banda y la referencia 407 indica una lente del objetivo (tal como 60x agua, NA 1.2). La referencia 480 indica un procesador y la referencia 490 indica una biblioteca opcional (que puede ser remota, p. ej., base de datos de internet con secuencias de aminoácidos de proteína).

La figura 3c representa esquemáticamente una realización alternativa del dispositivo 1, con una placa 472 con un nanoporo 461. La placa puede ser, p. ej., de los siguientes materiales, nitruro de silicio ("SiN"), Si/SiO₂, Al₂O₃, polímero, grafeno, BN. Una señal eléctrica se mide entre un lado nano-transportador 472 y un lado opuesto 471. En este caso, el detector 440 comprende una unidad detectora configurada para medir un parámetro eléctrico entre el lado nano-transportador y el lado opuesto del filtro que comprende nanoporos. El parámetro eléctrico se selecciona entre el grupo que consiste en una diferencia de potencial, una corriente y la resistencia. Cuando la proteína 200 pasa a través del nanoporo 461 (desde el lado opuesto hasta el lado nano-transportador 462), debido a la presencia de los marcadores, la señal eléctrica cambiará. El cambio de señal dependerá del tipo de marcador. Nótese que, en este caso los marcadores no son necesariamente luminiscentes. Además, nótese que la máquina transportadora molecular está configurada para arrastrar la proteína 100 a través del nanoporo 461.

Por ejemplo, el análisis (figura 4a) muestra que la fidelidad de predicción (PF) no cae de forma significativa incluso cuando hay un número de errores de intercambio (NSE) presentes (es decir el orden de C y K se intercambia durante la exploración) dada una secuencia CK de 40 de longitud. Este análisis de error se realizó usando tanto un algoritmo de coincidencia punto por punto como el algoritmo de Smith-Waterman.

Por la presente se proporciona una demostración del proceso de exploración basado en ClpXP usando la técnica FRET de una sola molécula (figura 4b). Se generó una forma hexamérica enlazada artificialmente de ClpX y se biotiniló su extremo C-terminal. Se inmovilizó esta proteína de nanocanal, en complejo con la proteasa ClpP, sobre una superficie de cuarzo usando conjugación de estreptavidina-biotina. En esta cámara de muestra, se añadieron péptidos que contenían un residuo de K y uno de C. El residuo de K se marcó con un fluoróforo Cy3 que contenía un grupo éster de NHS (esfera verde en la figura), y el residuo de C se marcó con un fluoróforo Cy5 que contenía un grupo mono-maleimida (esfera roja). El péptido contenía la marca ssrA que es reconocida por ClpX con alta especificidad. Las referencias RN, TN, DN y RE indican reconocimiento, translocación, degradación y liberación, respectivamente. En el gráfico, I en el eje y indica intensidad (en unidades arbitrarias), y t en el eje x es tiempo (en segundos).

REIVINDICACIONES

1. Un método para determinar el tipo de una proteína en un líquido que comprende la proteína, comprendiendo el método:
- 5 a) funcionalizar una proteína con al menos 2 tipos de marcadores de aminoácidos, que son selectivos para 2 tipos de aminoácidos de proteína predefinidos,
 b) guiar en la fase líquida la proteína funcionalizada con una máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada, a través de un área de detección de un detector, configurado para detectar una señal en función de los marcadores de los aminoácidos marcados;
 10 c) determinar, a partir de la señal detectada, una secuencia de los aminoácidos de proteína predefinidos;
 d) comparar la secuencia de los aminoácidos de proteína predefinidos con la presencia de dicha secuencia en una base de datos de proteínas y determinar el tipo de proteína en el líquido.
2. El método según la reivindicación 1, en donde la máquina transportadora molecular es una máquina transportadora molecular seleccionada entre el grupo de proteasas dependientes de ATP que consisten en una ClpXP, una ClpAP, una ClpCP, una ClpEP, una ClpYQ, una ClpB, una Lon, una FtsH, una PAN de arqueas, y una máquina transportadora molecular basada en el proteasoma, especialmente en donde la máquina transportadora molecular es una máquina transportadora molecular basada en ClpXP.
3. El método según una cualquiera de las reivindicaciones precedentes, en donde solamente están marcadas cisteína y lisina y en donde la máquina transportadora molecular comprende un donador configurado para sondear un marcador de aminoácidos.
- 20 4. El método según una cualquiera de las reivindicaciones precedentes, en donde los marcadores comprenden aceptores fluorescentes, en donde un donador fluorescente, configurado para formar temporalmente un par donador-aceptor con uno de los aceptores fluorescentes, está configurado dentro del área de detección, y en donde el detector comprende un microscopio de fluorescencia que incluye un microscopio de fluorescencia por reflexión interna total (TIRF), un microscopio de fluorescencia confocal, o un microscopio de fluorescencia a base de guía de ondas en modo cero; y en donde los marcadores comprenden un fluoróforo orgánico seleccionado entre uno o más de la familia de cianina, la familia Alexa, la familia Atto, la familia Dy y la familia de rodamina.
- 25 5. El método según la reivindicación 6, en donde la máquina transportadora molecular es una máquina transportadora molecular basada en ClpXP, en donde (i) especialmente la máquina transportadora molecular basada en ClpXP comprende una cámara de ClpP asimétrica de monómeros de ClpP y al menos un monómero de ClpP mutante, en el que al menos un monómero de ClpP mutante no puede acoplarse a ClpX, y en donde esta al menos una ClpP mutante está marcada con un donador fluorescente o en donde (ii) especialmente, la máquina transportadora molecular basada en ClpXP comprende una ClpX marcada con un donador fluorescente en complejo con una proteína ClpP no marcada.
- 30 6. El método según una cualquiera de las reivindicaciones 1-5, en donde el método comprende guiar la proteína funcionalizada con la máquina transportadora molecular a través de un filtro que comprende nanoporos que tiene un lado nano-transportador y un lado opuesto, en donde el detector comprende una unidad detectora configurada para medir un parámetro eléctrico entre el lado nano-transportador y el lado opuesto del filtro que comprende nanoporos, y en donde el parámetro eléctrico se selecciona entre el grupo que consiste en una diferencia de potencial, una corriente y la resistencia.
- 35 7. El método según una cualquiera de las reivindicaciones precedentes, que comprende además marcar la proteína con una marca que es reconocible por la máquina transportadora molecular.
- 40 8. El método según una cualquiera de las reivindicaciones precedentes, en donde la proteína funcionalizada es translocada con la máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada con una velocidad de translocación a través del área de detección del detector, en donde la velocidad de translocación se selecciona entre el intervalo de 0,1-60 aminoácidos por segundo, especialmente en donde la velocidad de translocación se controla controlando una concentración de ATP en el líquido.
- 45 9. El método según una cualquiera de las reivindicaciones precedentes, que comprende detectar la presencia de al menos 15, especialmente al menos 17, de los aminoácidos de proteína predefinidos en la proteína a identificar, que comprende funcionalizar la proteína dos 2-4 tipos de marcadores de aminoácidos, que son selectivos para 2-4 tipos de aminoácidos de proteína predefinidos.
- 50 10. Un dispositivo para determinar el tipo de proteína en un líquido, comprendiendo el dispositivo:
- 55 a) una máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada configurada para guiar una proteína que está funcionalizada con marcadores de aminoácidos, que son selectivas para al menos 2 tipos de aminoácidos de proteína predefinidos, a través de un área de detección de un detector,

b) dicho detector, configurado para detectar una señal en función de los marcadores de los aminoácidos marcados,

5 c) una unidad procesadora, configurada para identificar, a partir de la señal detectora, una secuencia de aminoácidos de la proteína funcionalizada, en donde la unidad procesadora está configurada, además, para comparar la secuencia de aminoácidos identificada con la presencia de dicha secuencia en una base de datos de proteínas y para identificar el tipo de proteína.

10 11. El dispositivo según la reivindicación 17, en donde la máquina transportadora molecular es una máquina transportadora molecular seleccionada entre el grupo de proteasas dependientes de ATP que consisten, entre otras, en una ClpXP, una ClpAP, una ClpCP, una ClpEP, una ClpYQ, una ClpB, una Lon, una FtsH, una PAN de arqueas y una máquina transportadora molecular basada en el proteasoma, especialmente en donde la máquina transportadora molecular es una máquina transportadora molecular basada en ClpXP.

15 12. El dispositivo según una cualquiera de las reivindicaciones 17-19, en donde el detector comprende un microscopio de fluorescencia que incluye un microscopio de fluorescencia por reflexión interna total (TIRF), un microscopio de fluorescencia confocal, o un microscopio de fluorescencia a base de guía de ondas en modo cero, y en donde especialmente la máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada comprende un donador fluorescente unido a ella.

20 13. El dispositivo según una cualquiera de las reivindicaciones 17-21, en donde la máquina transportadora molecular es una máquina transportadora molecular basada en ClpXP, en donde la máquina transportadora molecular basada en ClpXP comprende una cámara de ClpP asimétrica de monómeros de ClpP y al menos un monómero de ClpP mutante, en el que el al menos un monómero de ClpP mutante no puede acoplarse a ClpX, y en donde esta al menos una ClpP mutante está marcada con un donador fluorescente, especialmente en donde la máquina transportadora molecular es una máquina transportadora molecular basada en ClpXP, en donde la máquina transportadora molecular basada en ClpXP comprende una ClpX marcada con un donador fluorescente en complejo con una proteína ClpP.

25 14. El dispositivo según una cualquiera de las reivindicaciones 17-23, en donde el dispositivo comprende además un filtro que comprende nanoporos que tiene un lado nano-transportador y un lado opuesto y en donde el detector está configurado para medir un parámetro eléctrico entre el lado nano-transportador y el lado opuesto del filtro que comprende nanoporos, en donde el parámetro eléctrico se selecciona entre el grupo que consiste en una diferencia de potencial, una corriente y la resistencia, en donde el dispositivo está configurado además para guiar a la proteína que está funcionalizada con marcadores a través del nanoporo durante el uso del dispositivo.

30 15. El dispositivo según una cualquiera de las reivindicaciones 17-24, en donde una máquina transportadora molecular basada en proteasa dependiente de ATP inmovilizada está configurada para guiar una proteína que está funcionalizada con marcadores de aminoácidos, que son selectivos para 2-4 tipos de aminoácidos de proteína predefinidos, a través de un área de detección de un detector, en donde el detector está configurado para comparar la secuencia de los aminoácidos de proteína predefinidos de la proteína con la presencia de dicha secuencia en una base de datos de proteínas y determinar el tipo de proteína en el líquido, en donde la proteína está funcionalizada con solamente 2-4 tipos de marcadores de aminoácidos, que son selectivos para 2-4 tipos de aminoácidos de proteína predefinidos.

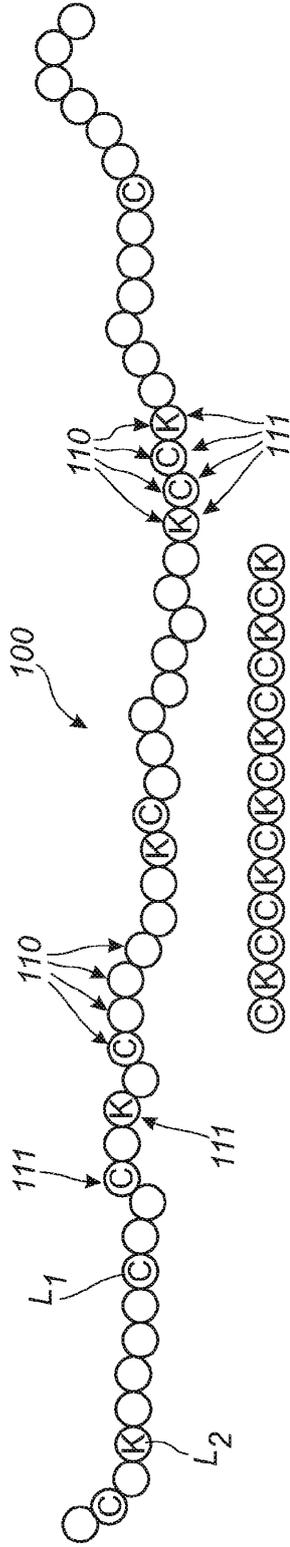


FIG. 1a

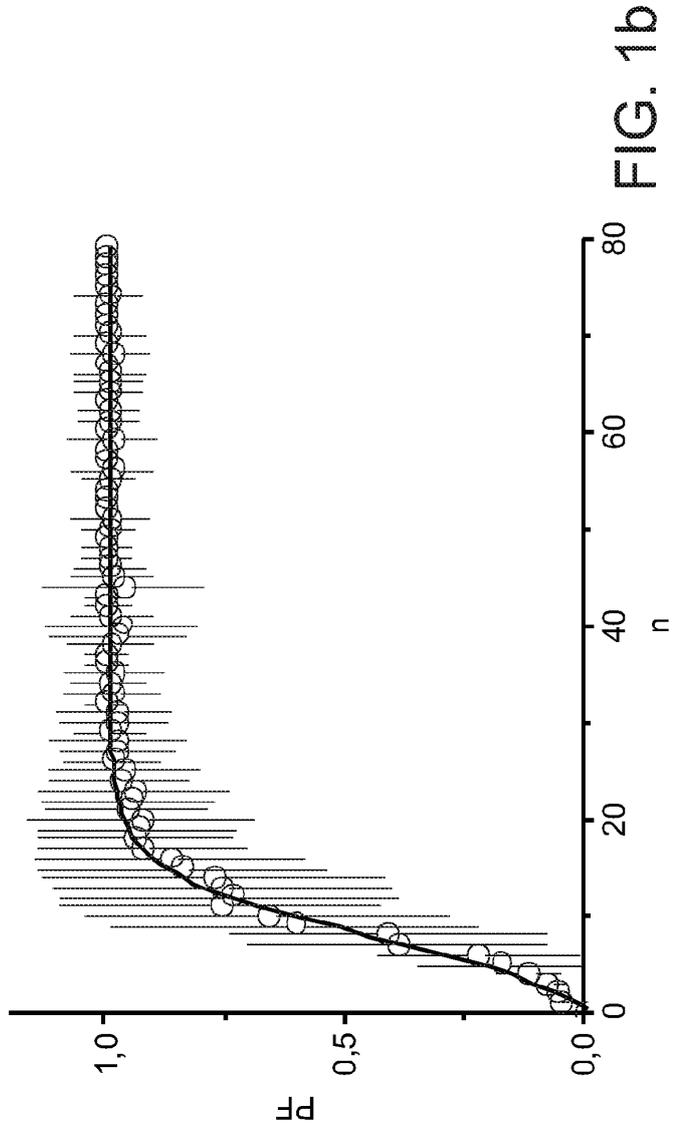


FIG. 1b

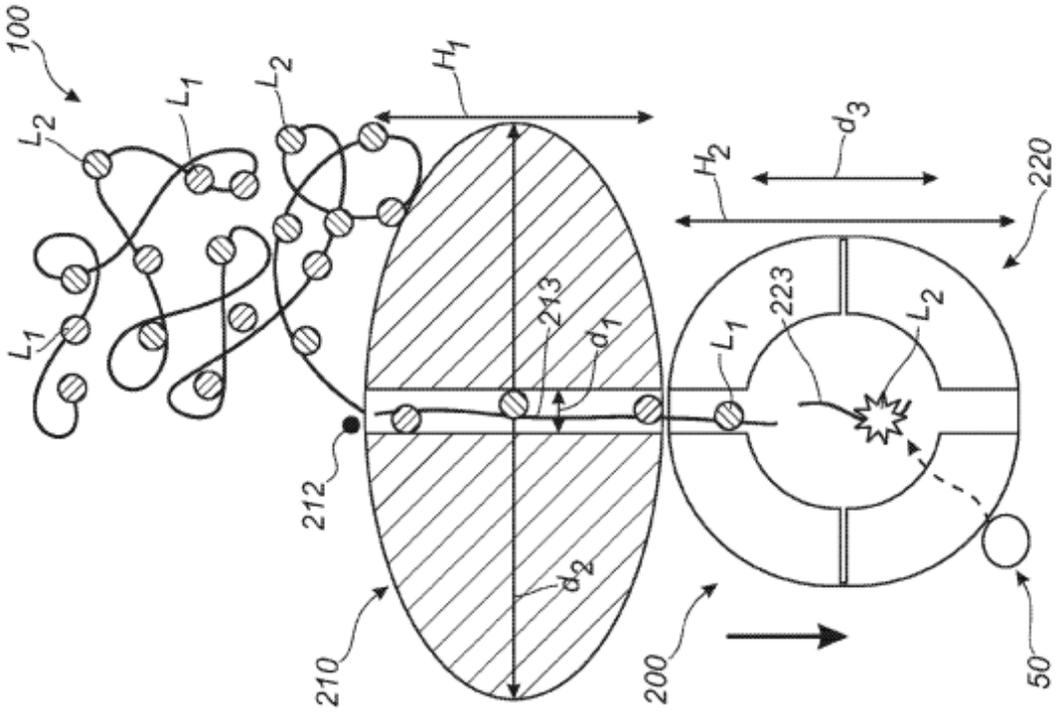


FIG. 2b

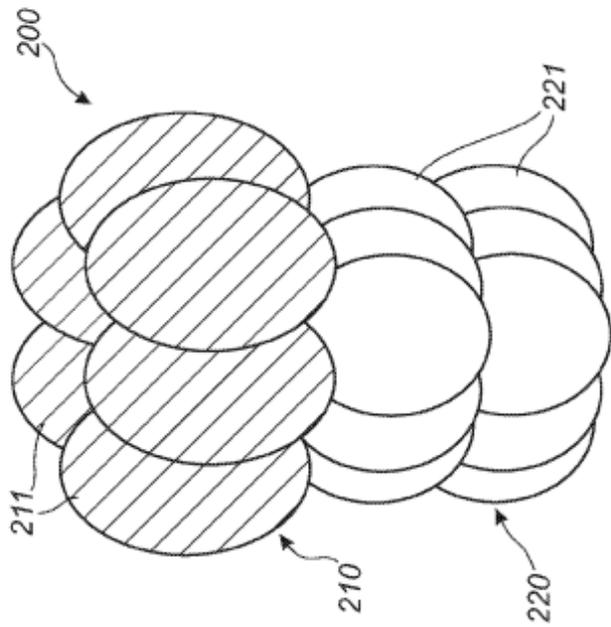


FIG. 2a

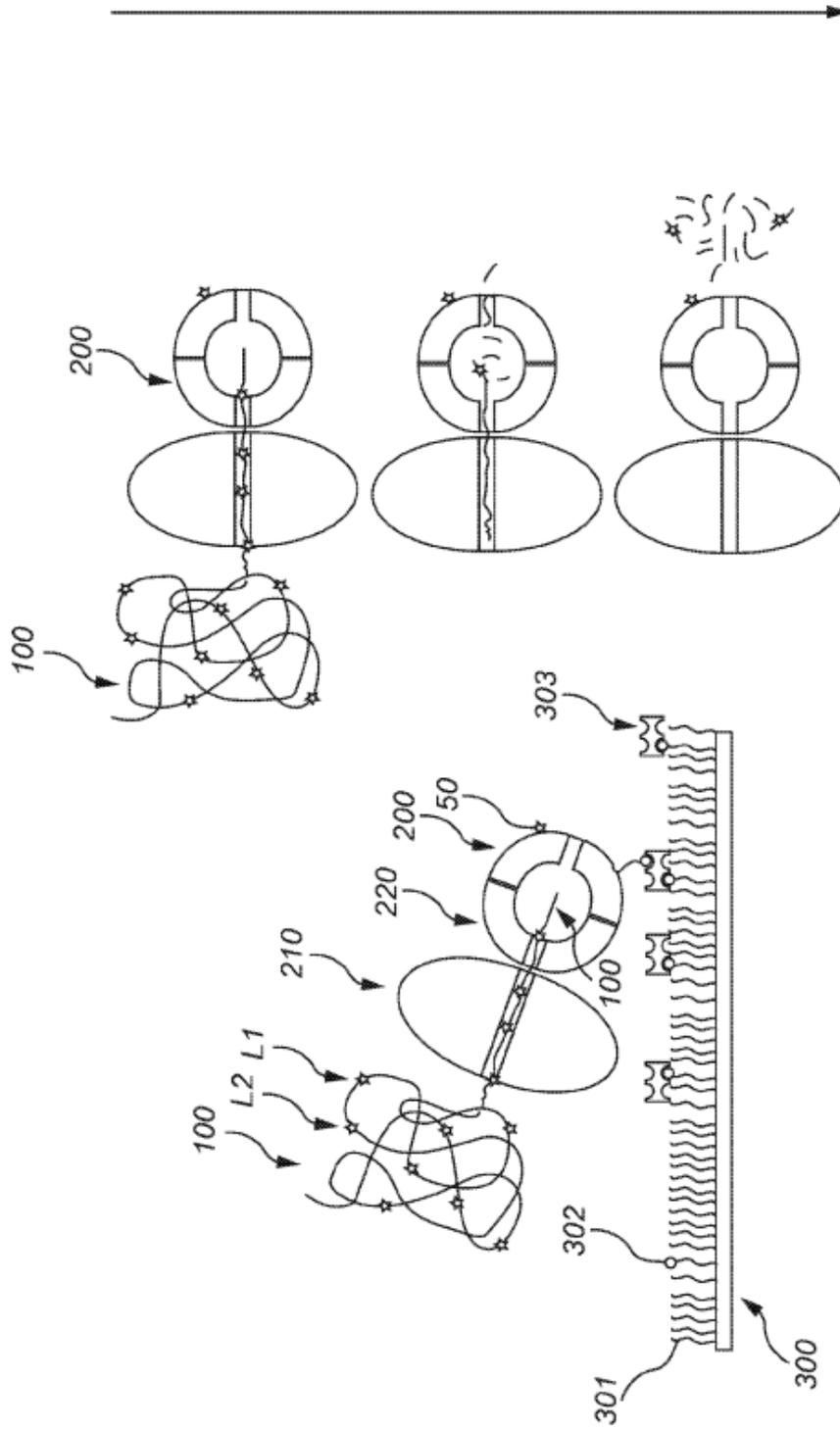


FIG. 3a

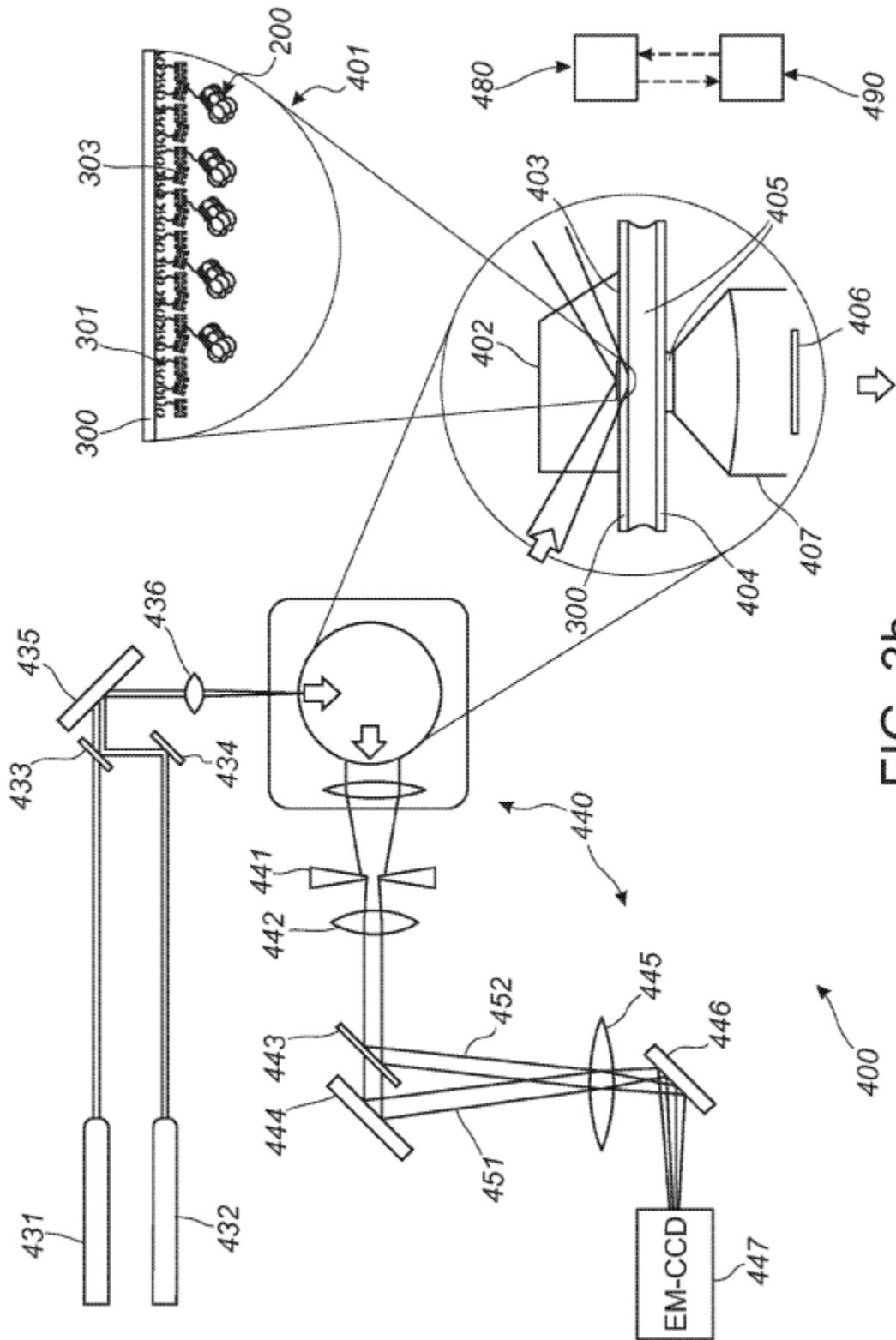


FIG. 3b

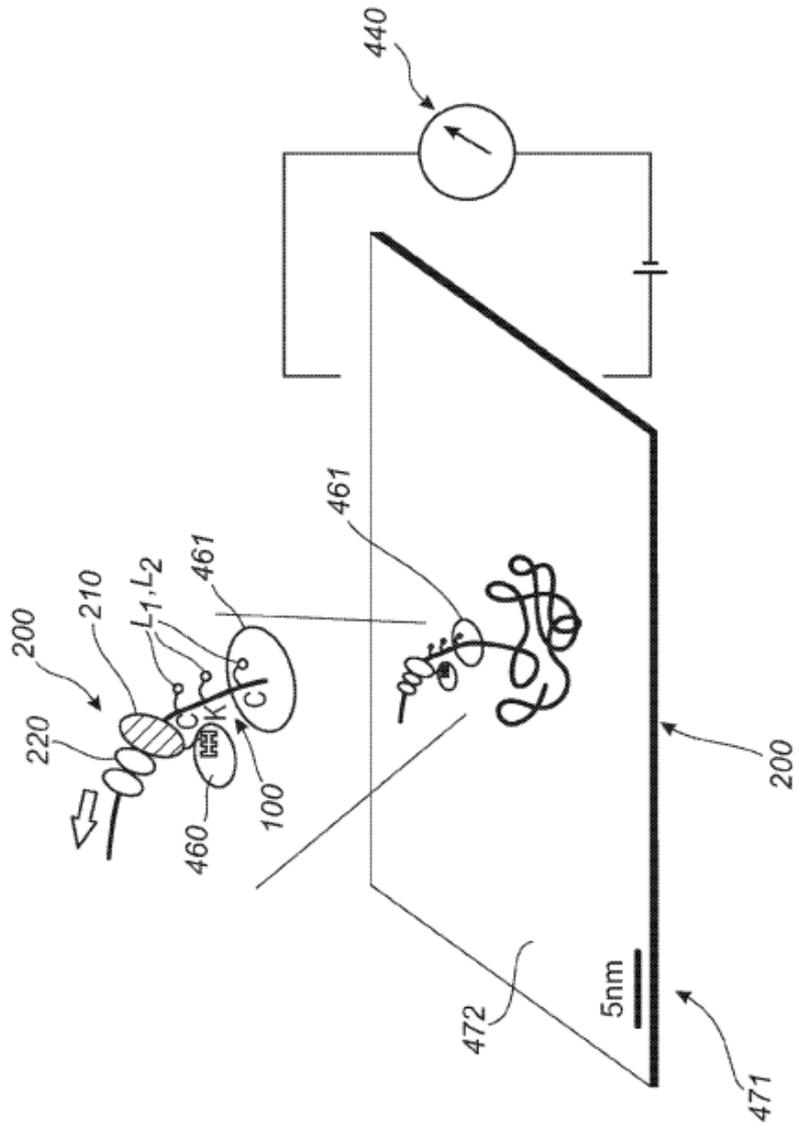


FIG. 3c

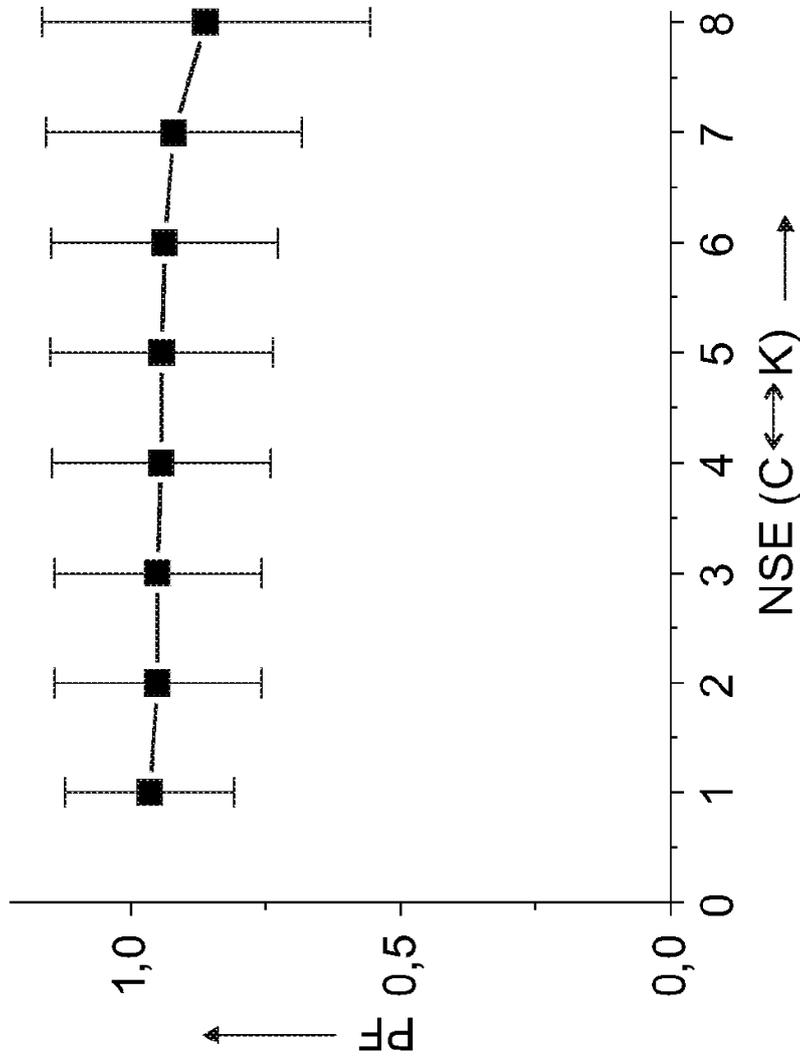


FIG. 4a

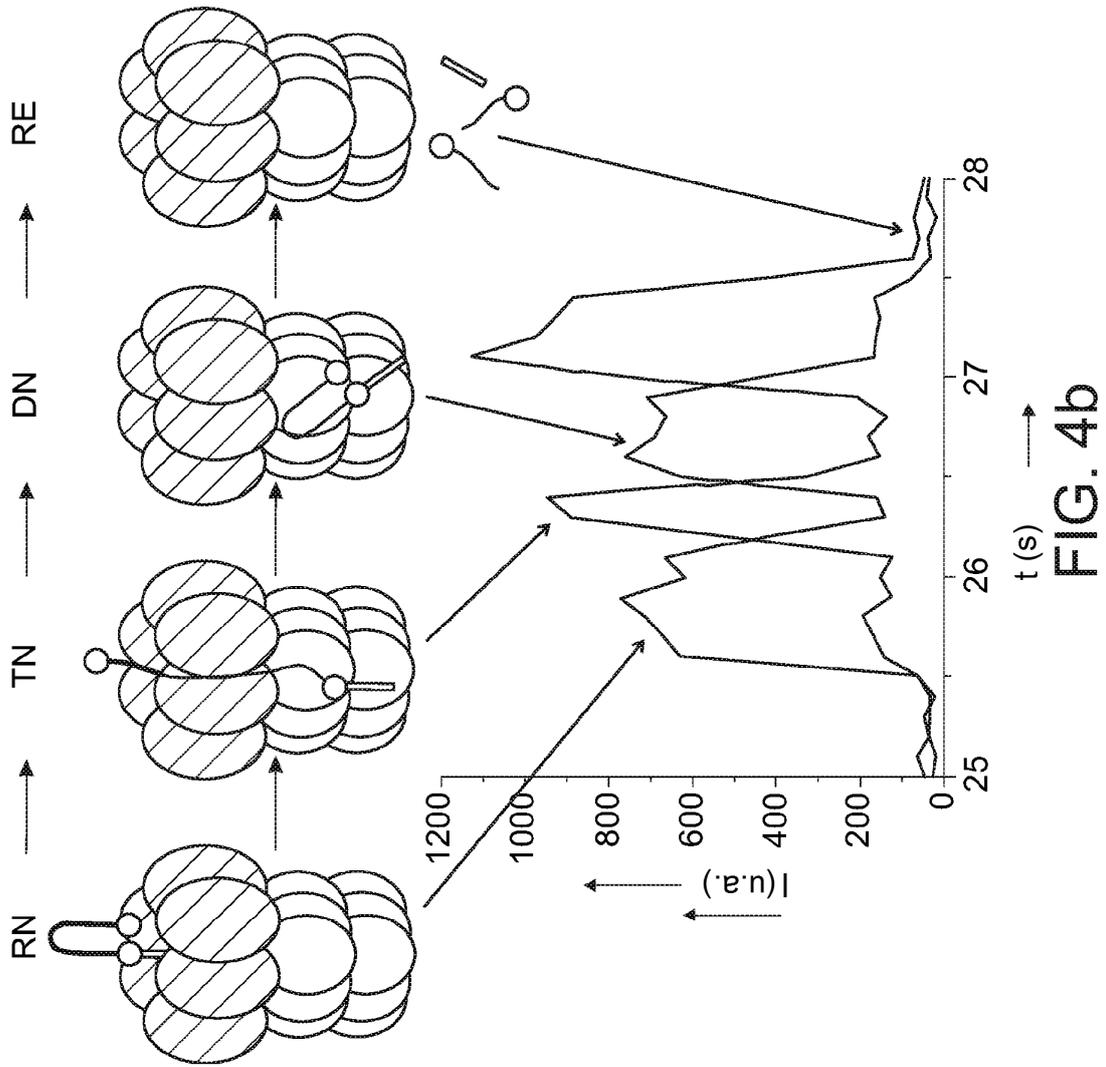


FIG. 4b