(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 999 861**

(51) Int. Cl.:

C12Q 1/6869 (2008.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **09.08.2021 PCT/EP2021/072112**

(87) Fecha y número de publicación internacional: **17.02.2022 WO22033998**

(96) Fecha de presentación y número de la solicitud europea: **09.08.2021 E 21763270 (2)**

(97) Fecha y número de publicación de la concesión europea: **23.10.2024 EP 4196608**

(54) Título: **Nucleósidos-5'-oligofosfato marcados con polímeros cargados positivamente, nanoporos que incorporan cargas negativas y procedimientos y sistemas que usan los mismos**

(30) Prioridad:

11.08.2020 US 202062706353 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
27.02.2025

(73) Titular/es:

F. HOFFMANN-LA ROCHE AG (100.00%)
Grenzacherstrasse 124
4070 Basel, CH

(72) Inventor/es:

AYER, ARUNA;
BERGMANN, FRANK;
CRISALLI, PETER;
HILLRINGHAUS, LARS;
KHAKSHOOR, OMID;
KUCHELMEISTER, HANNES;
NIE, RONGXIN;
PORTER, MARSHALL W.;
SEIDEL, CHRISTOPH;
SHIN, SEONG-HO;
TAING, MENG y
VARGAS, ADOLFO

(74) Agente/Representante:

LINAGE GONZÁLEZ, Rafael

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

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

Nucleósidos-5'-oligofosfato marcados con polímeros cargados positivamente, nanoporos que incorporan cargas negativas y procedimientos y sistemas que usan los mismos

Antecedentes

I. Campo

La presente divulgación se refiere a nucleósidos-5'-oligofosfato y usos de los mismos, y a nanoporos y usos de los mismos, incluyendo su uso en sistemas y procedimientos de secuenciación por síntesis basada en nanoporos.

II. Breve análisis de la técnica relacionada

En su forma más básica, los sistemas de secuenciación de nanoporos comprenden un electrodo sensor situado cerca de un nanoporo, de modo que el electrodo sensor puede detectar y registrar las características electroquímicas de los iones que fluyen a través del nanoporo. Cuando moléculas relativamente grandes ocupan el nanoporo, las características electroquímicas detectadas por el electrodo sensor cambian. La identidad de la molécula que ocupa el nanoporo se puede determinar, a continuación, basándose en el cambio de las características electroquímicas, tal como un cambio en la corriente que fluye a través del nanoporo o una caída en el voltaje medido. Se puede encontrar una visión general de los sistemas de secuenciación basada en nanoporos en Wang I y Feng.

Los nanoporos usados en estos sistemas de secuenciación típicamente son de tres tipos: nanoporos biológicos, nanoporos de estado sólido y nanoporos híbridos. Los nanoporos biológicos son moléculas formadoras de poros de naturales, especialmente proteínas tales como porinas, hemolisinas y similares. Las proteínas formadoras de poros comúnmente usadas incluyen la proteína α-hemolisina (α-HL) de *Staphylococcus aureus*, proteína de membrana externa G (ompG) de *Escherichia coli* y porina MspA (MspA) de *Mycobacterium smegmatis*. En algunos casos, como en el caso de ompG, el poro se forma a partir de una única subunidad de la proteína. En otros casos, como con α-HL y MspA, el poro es un conjunto de múltiples subunidades de la proteína formadora de poros. Por ejemplo, α-HL forma una estructura de poro heptamérica y MspA forma una estructura de poro octamérica. Se pueden encontrar nanoporos genomanipulados ejemplares basados en estas proteínas en, por ejemplo, los documentos WO 2016/069806 (α-HL), WO 2017/050728 (α-HL), WO 2017/184866 (α-HL), WO 2018/002125 (α-HL), WO 2012/178097 (α-HL), Gari (ompG), WO 2017/050722 (ompG), US 2015-0080242 (ompG), Manrao (MspA), Pavlenok (MspA), WO 2013/098562 (MspA), US 2014-0309402 (MspA), US 2013-0146457 (MspA) y Wang II (diversos). Los nanoporos de estado sólido son estructuras de poros fabricadas a partir de materiales sintéticos, por ejemplo, formando orificios de tamaño nanométrico en membranas sintéticas. Los materiales ejemplares a partir de los que se pueden formar nanoporos de estado sólido incluyen nitruros de silicio, sílice, alúmina, grafeno, nitruro de boro y disulfuro de molibdeno. Los nanoporos de estado sólido se revisan por Chen, Lee, Wasfi, Wang I y Feng. Los nanoporos híbridos incorporan tanto nanoporos biológicos como nanoporos de estado sólido. Por ejemplo, un nanoporo biológico (tal como un nanoporo de α-HL) se puede insertar en un nanoporo de estado sólido. Los nanoporos híbridos se revisan por Lee, Wasfi y Feng.

Un enfoque para la secuenciación de ácidos nucleicos basada en nanoporos implica hacer pasar ácidos nucleicos monocatenarios directamente a través del poro (denominado en el presente documento "secuenciación directa"). Cada nucleótido (o combinación única de nucleótidos) genera un cambio único en al menos una característica electroquímica del poro. Estos sistemas usan con frecuencia medios para controlar la tasa a la que el ácido nucleico se transloca a través del poro, tal como anclar enzimas al poro (incluyendo polimerasas y helicasas), retirar residuos cargados negativamente de y añadir residuos cargados positivamente al canal de poro, y añadir regiones bicatenarias al ácido nucleico monocatenario. Se analizan enfoques de secuenciación directa ejemplares por, por ejemplo, Feng, Manrao y Wang I.

Otro procedimiento implica un enfoque de secuenciación por síntesis (SBS) realizando una reacción de amplificación catalizada por polimerasa cerca de una abertura del nanoporo con moléculas de nucleótido polifosfato marcado. Cada nucleótido polifosfato marcado incluye un resto marcado distinto que genera una firma electroquímica única cuando reside en o cerca del nanoporo. A medida que los nucleótidos polifosfato marcados se incorporan en el amplicón, la marca pasa hacia o cerca del nanoporo y se registra la firma electroquímica de la marca. La secuencia del amplicón se deriva del orden en el que los restos de marca entran en el nanoporo. Se describen enfoques y materiales de SBS basada en marcas ejemplares para realizar dichos procedimientos en, por ejemplo, los documentos WO 2012-083249, WO 2013/154999, US 2014/0309144, US 9,017,937, WO 2015/148402, WO 2016/069806, WO 2016/144973, US 2016/0222363, US 2016/0333327, WO 2017/050728, WO 2017/184866, WO 2017/050722, US 2017/0267983, US 2018/0245147, US 2018/0094249, WO 2018/002125 y Kumar. Se han propuesto diversas marcas para su uso en dichos sistemas, incluyendo marcas basadas en polipéptidos (tales como marcas de polilisina) y polinucleótidos. Véanse, por ejemplo, los documentos US 8.652.779 y WO2017042038A1.

Sumario de la invención

En el presente documento se divulgan sistemas para la secuenciación por síntesis basada en nanoporos de polinucleótidos, comprendiendo el sistema un conjunto de nucleósidos-5'-oligofosfato que tienen una marca cargada positivamente, un nanoporo, una ácido nucleico polimerasa y un electrodo sensor en proximidad al nanoporo, teniendo el nanoporo un canal que tiene una pluralidad de restos cargados negativamente. Los sistemas presentados en el presente documento tienen un fondo significativamente reducido y una proporción señal/ruido mejorada en relación con sistemas similares que usan marcas cargadas negativamente y poros que carecen de cargas negativas adicionales.

También se divulgan en el presente documento construcciones de marcas para su uso en reacciones de secuenciación por síntesis basada en marcas, así como nucleósidos-5'-oligofosfato marcados que incorporan las mismas. En algunos casos, las construcciones de marcas incluyen un segmento cargado positivamente (SCP) de al menos 5 unidades de monómero de longitud que tiene una carga neta positiva de al menos +5 a pH 7,0. En algunos casos, el SCP comprende una pluralidad de unidades de repetición y cada unidad de repetición tiene una carga neta positiva y una densidad de carga de al menos +0,1.

También se divulgan en el presente documento polipéptidos para su uso en la generación de un nanoporo biológico o un componente biológico de un nanoporo híbrido. En algunos casos, el polipéptido contiene suficientes restos cargados negativamente no naturales de modo que un nanoporo formado a partir del mismo tiene suficiente carga negativa en el canal para mitigar la inserción de ácido nucleico en el canal durante el proceso de secuenciación por síntesis. En otros casos, el polipéptido contiene suficientes restos cargados negativamente naturales que típicamente están ausentes en los nanoporos usados para la secuenciación de ácidos nucleicos. En algunos modos de realización, el polipéptido es un componente de un nanoporo biológico, tal como los nanoporos de OmpG, MspA y alfa-hemolisina, entre otros. En algunos modos de realización, el polipéptido tiene al menos un 75 % de identidad de secuencia con cualquiera de las secuencias seleccionadas del grupo que consiste en SEQ ID NO: 1 y SEQ ID NO: 31-101, con la condición de que el polipéptido comprenda al menos un aminoácido cargado negativamente no natural en una posición correspondiente a la región de entrada y/o la región de barril beta de un poro homoheptamérico formado a partir de la SEQ ID NO: 1. En otros modos de realización, el polipéptido tiene al menos un 70 % de identidad con respecto a una secuencia seleccionada del grupo que consiste en SEQ ID NO: 23-28, con la condición de que dicho polipéptido comprenda un conjunto de hebras beta β1–β14 que tienen al menos un 90 % de identidad con respecto a las hebras beta β1–β14 de SEQ ID NO: 23 y al menos 1 de dichas hebras beta tenga un aminoácido cargado negativamente no natural. En otros modos de realización, el polipéptido tiene al menos un 70 % de identidad con respecto a SEQ ID NO: 29, con la condición de que el polipéptido tenga un ácido aspártico o ácido glutámico en cualquiera o cada una de las posiciones correspondientes a D90, D91, D93, D118, D134 y E139 de SEQ ID NO: 29.

También se divulgan en el presente documento procedimientos de secuenciación de un ácido nucleico, que comprenden: (A) proporcionar una pluralidad de complejos de secuenciación de nanoporos, comprendiendo cada complejo de secuenciación de nanoporos (a1) un electrodo sensor, (a2) un nanoporo insertado en una barrera electroquímicamente resistiva en proximidad al electrodo sensor, en el que el canal del nanoporo tiene una pluralidad de restos cargados negativamente, (a3) una ácido nucleico polimerasa asociada con el nanoporo, (a4) un ácido nucleico molde complejado con la ácido nucleico polimerasa, y (a) un conjunto de nucleósidos-5'-oligofosfato marcados que comprenden una marca de polímero cargada positivamente; (b) en los complejos de secuenciación de nanoporos, polimerizar un conjunto de nucleósidos-5'-oligofosfato en un ácido nucleico complementario del ácido nucleico molde por una reacción de amplificación de ácido nucleico dependiente de molde catalizada por la ácido nucleico polimerasa, en el que la marca de polímero del nucleósido-5'-oligofosfato marcado se mueve hacia o en proximidad al canal del nanoporo a medida que el nucleósido-5'-oligofosfato marcado se incorpora al ácido nucleico complementario, y en el que el movimiento de la marca de polímero hacia o en proximidad al canal cambia una característica de una corriente que fluye a través del nanoporo; (c) detectar el cambio en la característica de la corriente que fluye a través del nanoporo proporcionado por las marcas de polímero con el electrodo sensor y registrar el cambio en el sistema informático; y (d) correlacionar cada cambio registrado con uno de los nucleósidos-5'-oligofosfato marcados.

Otros detalles e invenciones se describen en detalle a continuación.

Breve descripción de los dibujos

El fichero de patente o solicitud contiene al menos un dibujo realizado en color. Las copias de esta publicación de patente o solicitud de patente con dibujo(s) en color se proporcionarán por la Oficina bajo petición y pago de la tasa necesaria.

La FIG. 1 es un experimento de secuenciación SBS-nano ejemplar que muestra el problema de la inserción de molde/cebador.

La FIG. 2A es una vista lateral de un poro de α-hemolisina ejemplar. Los residuos aminoacídicos correspondientes

al vestíbulo de canal están en turquesa. Los aminoácidos de los residuos correspondientes al sitio de constricción están en rojo. Los aminoácidos orientados hacia el canal del cuerpo de barril beta y la salida están en rosa.

La FIG. 2B es una sección transversal del poro de α-hemolisina de la FIG. 2A, que ilustra las diversas regiones del canal de nanoporo.

La FIG. 3 es un esquema de una celda de secuenciación de nanoporos que realiza un procedimiento de SBS basado en marcas.

La FIG. 4A es un diagrama de cinta de un nanoporo de OmpG monomérico. Las flechas **401** son hebras beta. También se ilustran los bucles **402** y los giros **403**.

La FIG. 4B es una sección transversal de un nanoporo de OmpG monomérico. Además de las regiones de hebra beta **401**, los bucles **402** y los giros **403**, se ilustra un sitio de restricción **404**.

La FIG. 5A muestra las diferentes regiones del polipéptido de OmpG monomérico de acuerdo con la SEQ ID NO: 23.

La FIG. 5B muestra las diferentes regiones del polipéptido de OmpG monomérico de acuerdo con la SEQ ID NO: 24.

La FIG. 5C muestra las diferentes regiones del polipéptido de OmpG monomérico de acuerdo con la SEQ ID NO: 25.

La FIG. 5D muestra las diferentes regiones del polipéptido de OmpG monomérico de acuerdo con la SEQ ID NO: 26.

La FIG. 5E muestra las diferentes regiones del polipéptido de OmpG monomérico de acuerdo con la SEQ ID NO: 27.

La FIG. 5F muestra las diferentes regiones del polipéptido de OmpG monomérico de acuerdo con la SEQ ID NO: 28.

La FIG. 6 es una vista superior de un modo de realización de un chip sensor de nanoporos.

La FIG. 7 ilustra un modo de realización de una celda de nanoporo de ejemplo en un chip sensor de nanoporos.

La FIG. 8 ilustra un modo de realización de una celda de nanoporo que realiza la secuenciación de nucleótidos usando la técnica de nano-SBS.

La FIG. 9 ilustra un modo de realización de un circuito eléctrico en una celda de nanoporo.

La FIG. 10 muestra puntos de datos de ejemplo capturados de una celda de nanoporo durante períodos brillantes y períodos oscuros de ciclos de CA.

La FIG. 11 es un gráfico de los niveles de expresión (eje x, unidades en A280) frente a la carga neta del monómero a pH 7,0 (eje y).

La FIG. 12 muestra geles de SDS-PAGE de autooligomerización de preparaciones de α-hemolisina monomérica. Los cuadros con rayas continuas muestran la tasa de migración esperada de heptámeros. Los cuadros con rayas discontinuas muestran la migración esperada de monómeros.

La FIG. 13A muestra geles de SDS-PAGE de autooligomerización de la preparación de α-hemolisina monomérica G1637 en presencia de una composición de tampón de ácido 2-(N-morfolino)etanosulfónico (MES) o bien una composición de tampón que contenía N-óxido de trimetilamina (TMAO) 3 M. Los carriles ilustrados con "M" son preparaciones antes del tratamiento de oligomerización. Los carriles ilustrados con "O" son preparaciones después del tratamiento de oligomerización. Los cuadros con rayas continuas muestran la tasa de migración esperada de heptámeros. Los cuadros con rayas discontinuas muestran la migración esperada de monómeros.

La FIG. 13B muestra geles de SDS-PAGE de autooligomerización de la preparación de α-hemolisina monomérica G1637 en presencia de: (carriles 1 y 2) acetato de sodio 50 mM, NaCl 50 mM y TMAO 3 M, pH 4,8; (carriles 3 y 4) MES 50 mM, NaCl 50 mM y TMAO 3 M, pH 6,0; (carriles 5 y 6) fosfato de potasio 50 mM, NaCl 50 mM y TMAO 3 M, pH 7,4; y (carriles 7 y 8) Tris 50 mM, NaCl 50 mM y TMAO 3 M, pH 8,0. Los carriles ilustrados con "M" son preparaciones antes del tratamiento de oligomerización. Los carriles ilustrados con "O" son preparaciones después del tratamiento de oligomerización. Los cuadros con rayas continuas muestran la tasa de migración esperada de heptámeros. Los cuadros con rayas discontinuas muestran la migración esperada de monómeros.

La FIG. 14A ilustra la oligomerización del monómero G1147 con un monómero de α-hemolisina natural (WT) marcado con SpyCatcher. Como se indica por las flechas, se formaron heteroheptámeros 1:6 (WT:G1147), así como heteroheptámeros 2:5.

La FIG. 14B ilustra la oligomerización con diversas proporciones de G1515 con respecto a WT, incluyendo ejemplos en los que G1515 está en exceso (proporciones de 3:1, 4:1, 5:1, 6:1, 7:1, 8:1 y 9:1), en los que G1515 y WT están presentes en cantidades equivalentes (proporción de 1:1) y en los que el monómero WT está en exceso (proporciones de 1:3, 1:5, 1:7 y 1:9). Las flechas indican la migración esperada de heteroheptámeros 1:6 (1 monómero WT-GFP-SC con respecto a 6 monómeros G1515) y heteroheptámeros 2:5 (2 monómeros WT-GFP-SC con respecto a 5 monómeros G1515).

La FIG. 14C muestra un desarrollo de cromatografía de intercambio catiónico de un heptámero generado a partir de una mezcla 9:1 de G1515 con respecto a WT.

La FIG. 14D es un gel de SDS-PAGE de fracciones de heptámero generada a partir de una mezcla 9:1 de G1515 con respecto a WT (+/-SC-GFP). P1-P6 corresponden a las fracciones marcadas de la fig. 14C. Los carriles marcados con un "-" son heptámeros no marcados. Los carriles marcados con un "+" son SC-GFP marcados. Las flechas indican la migración esperada de heptámeros 1:6 y 2:5 (es decir, 1 o 2 copias de los componentes marcados WT-SC-GFP).

La FIG. 14E es un desarrollo de cromatografía de intercambio catiónico (gráfico inferior) de un heptámero generado a partir de una mezcla 9:1 de G1713 a WT, así como un gel SDS-PAGE (imagen superior) de las fracciones resultantes P1-P4. Los carriles del gel de SDS-PAGE marcados con un "-" son heptámeros no marcados. Los carriles marcados con un "+" ilustran heptámeros marcados con SC-GFP. Las flechas muestran la migración esperada de heptámeros 1:6 y 2:5.

La FIG. 14F es un gel de SDS-PAGE de oligómeros purificados además a partir de la fracción P2 de la fig. 14E. Los carriles del gel de SDS-PAGE marcados con un "-" son heptámeros no marcados. Los carriles marcados con un "+" ilustran heptámeros marcados con SC-GFP. Las flechas indican la migración esperada de heptámeros 1:6 y 2:5.

La FIG. 14G es un gel de SDS-PAGE de oligómeros purificados formados a partir de G1752 + WT. Los carriles del gel de SDS-PAGE marcados con un "-" son heptámeros no marcados. Los carriles marcados con un "+" ilustran heptámeros marcados con SC-GFP. Las flechas indican la migración esperada de heptámeros 1:6 y 2:5.

La FIG. 14H es un gel de SDS-PAGE de oligómeros purificados formados a partir de G1753 + WT (las 2 columnas de la izquierda) y G1933 + WT (las 2 columnas de la derecha). Los carriles del gel de SDS-PAGE marcados con un "-" son heptámeros no marcados. Los carriles marcados con un "+" ilustran heptámeros marcados con SC-GFP. Las flechas indican la migración esperada de heptámeros 1:6 y 2:5.

La FIG. 15A es un gráfico que muestra las distribuciones de los niveles de marca frente a la densidad de probabilidad usando una membrana de DOPhPC. El eje X es la fracción del nivel de canal abierto registrado cuando se usa la marca indicada. El eje Y se informa como la densidad de probabilidad (es decir, el porcentaje de acontecimientos que tienen el nivel de marca registrado).

La FIG. 15B es un gráfico que muestra las distribuciones de los niveles de marca frente a la densidad de probabilidad usando una membrana de DPhPC. El eje X es la fracción del nivel de canal abierto registrado cuando se usa la marca indicada. El eje Y se informa como la densidad de probabilidad (es decir, el porcentaje de acontecimientos que tienen el nivel de marca registrado).

La FIG. 16A ilustra la estructura general de un desoxirribonucleósido hexafosfato marcado ejemplar (marcado como "dN6P"). Un resto de marca cargado positivamente (marcada como "marca") se fija al dN6P por medio de un enlace químico entre la marca y el grupo fosfato terminal del dN6P ("enlace").

La FIG. 16B ilustra un esquema de síntesis ejemplar para dN6P marcado.

La FIG. 16C ilustra un esquema de síntesis ejemplar para dT6P marcado que tiene una marca de acuerdo con la SEQ ID NO: 110 con un enlace de acuerdo con la fórmula 7.

La FIG. 17A muestra la intensidad de fluorescencia (eje y) frente al tiempo (eje x) para cada marca a una concentración de 15, 10, 5 y 1 µM con las variantes de Pol6 313, 431, 867, 1369 y 1743. La intensidad máxima de fluorescencia de las marcas positivas es aproximadamente un 10 % de la intensidad máxima de fluorescencia de la marca negativa. Se repitió el ensayo de desplazamiento de fluorescencia en el que se marcaron cada uno de dA6P, dT6P, dC6P y dG6P y el. La fig. 17B ilustra la K_{ext} usando la marca negativa, la fig. 17C ilustra la K_{ext} usando la marca 4Npa-K'$_{20}$-K$_4$-NH$_2$ y la fig. 17D ilustra la K_{ext} de Pol6 1743 en diversas concentraciones de dN6P marcado.

La FIG. 17B muestra la K_{ext} de cada una de las variantes de Pol6 313, 431, 867, 1369 y 1743 con 1, 5, 10 y 15 µM de una marca cargada negativamente (SP2).

La FIG. 17C muestra la K_{ext} de cada una de las variantes de Pol6 313, 431, 867, 1369 y 1743 con 1, 5, 10 y 15 µM de una marca cargada positivamente.

La FIG. 17D muestra la K_{ext} de variantes de Pol6 1743 con 1, 5, 10 y 15 µM de marca Sp2 y dos marcas cargadas positivamente diferentes.

La FIG. 18 muestra los resultados de un experimento de secuenciación usando dN6P de acuerdo con la tabla 13 que tienen la misma marca (P-C-83, P-A-82, P-G-80 y P-T-81). El eje Y es el nivel de señal registrado y el eje X es el tiempo en segundos. Se muestran dos réplicas. El gráfico A y el gráfico C muestran los experimentos de secuenciación completos de las dos réplicas. El gráfico B es un subconjunto del intervalo de tiempo del gráfico A (que va desde ~381,25 segundos hasta <382,75 segundos). El gráfico D es un subconjunto del intervalo de tiempo del gráfico D (que va desde ~360 segundos hasta ~410 segundos).

La FIG. 19 muestra los resultados de un experimento de secuenciación usando dN6P de acuerdo con la tabla 13 que tienen la marca de dos marcas diferentes (P-C-74 y P-A-78 (que tienen la misma marca entre sí), P-G-76 y P-T-70 (que tienen la misma marca entre sí, pero diferente de la marca de P-C-74 y P-A-78)). El eje Y es el nivel de señal registrado y el eje X es el tiempo en segundos. Se muestran dos réplicas. El gráfico A y el gráfico C muestran los experimentos de secuenciación completos de las dos réplicas. El gráfico B es un subconjunto del intervalo de tiempo del gráfico A (que va desde ~570 segundos hasta ~625 segundos). El gráfico D es un subconjunto del intervalo de tiempo del gráfico D (que va desde ~523,6 segundos hasta ~525,4 segundos). El cuadro azul y el cuadro amarillo son los diferentes niveles de paso de diferentes marcas respectivamente dentro del nanoporo durante un experimento de secuenciación.

Descripción detallada de la invención

La invención se describirá ahora en detalle a modo de referencia usando solo las siguientes definiciones y ejemplos.

A menos que se defina de otro modo en el presente documento, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que se entiende comúnmente por un experto en la técnica a la que pertenece la presente invención. Singleton *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2.ª ED., John Wiley and Sons, New York (1994), y Hale y Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) proporcionan a un experto un diccionario general de muchos de los términos usados en la presente invención. Los profesionales sanitarios se dirigen en particular a Sambrook *et al.*, 1989, y Ausubel FM *et al.*, 1993, para consultar definiciones y términos de la técnica. Se ha de entender que la presente invención no está limitada a la metodología, protocolos y reactivos particulares descritos, ya que estos pueden variar.

A menos que se indique de otro modo, los ácidos nucleicos se escriben de izquierda a derecha en una orientación de 5' a 3'; las secuencias de aminoácidos se escriben de izquierda a derecha en una orientación de amino a carboxi, respectivamente.

Los títulos proporcionados en el presente documento no son limitaciones de los diversos aspectos o modos de realización de la invención que se pueden tener por referencia a la memoria descriptiva como un todo. En consecuencia, los términos definidos de inmediato a continuación se definen más completamente por referencia a la memoria descriptiva como un todo.

I. Términos

Complejo de secuenciación de nanoporos activo: un complejo de secuenciación de nanoporos en el que la polimerasa está catalizando la polimerización basada en molde de los nucleósidos-5'-oligofosfato marcados y el electrodo sensor está detectando acontecimientos de captura en el nanoporo.

Alfa-hemolisina: como se usa en el presente documento, "alfa-hemolisina", "α-hemolisina" y "α-HL" se usan de manera intercambiable y se refieren a la proteína monomérica que se autoensambla en un canal transmembranario lleno de agua heptamérico (es decir, nanoporo). Dependiendo del contexto, el término también se puede referir al canal transmembranario formado por siete proteínas monoméricas.

Aminoácido amidado: un aminoácido en el que se ha amidado el extremo carboxílico.

Aminoácido: como se usa en el presente documento, el término "aminoácido" se refiere a cualquier compuesto que puede formar un enlace peptídico que tiene la estructura general H_aNX-COOH, en la que X es una cadena de alquilo (opcionalmente sustituida) de al menos un carbono de longitud, a es 1 o 2 con la condición de que cuando a

es 1, HN–X es una estructura cíclica. Esto incluye explícitamente, pero no se limita a, α-aminoácidos, β-aminoácidos, γ-aminoácidos y δ-aminoácidos y sus enantiómeros L y D. El término "aminoácido" se usa de manera intercambiable con "residuo aminoacídico" y se puede referir a un aminoácido libre y/o a un residuo aminoacídico de un péptido. Será evidente a partir del contexto en el que se usa el término si se refiere a un aminoácido libre o a un residuo de un péptido. A menos que se indique de otro modo, todas las secuencias de aminoácidos se representan en el presente documento por fórmulas con una orientación izquierda y derecha que es en la dirección convencional del extremo amínico al extremo carboxílico. A menos que se indique de otro modo, todos los residuos aminoacídicos de una secuencia de aminoácidos mencionados usando el código de aminoácido estándar de 1 o 3 letras se referirán al enantiómero L del α-aminoácido correspondiente.

Análogo de aminoácido: como se usa en el presente documento, el término "análogo de aminoácido" se referirá a cualquier estructura química que pueda formar enlaces similares a péptido y que tenga la misma estructura general de un aminoácido, excepto que uno o más de los carbonos de la cadena de alquilo X se reemplaza por otro resto (tal como un nitrógeno o un grupo fenilo) y/o la cadena lateral se localiza en una posición distinta de carbono a lo largo de la cadena principal (tal como en el extremo amínico). Los análogos de aminoácido ejemplares incluyen peptoides, azapéptidos, oligoureas, arilamidas, oligohidracidas y similares.

Derivado de aminoácido: en algunos casos, la memoria descriptiva se refiere a un "derivado de" un aminoácido específico o construcciones similares. Esto se interpretará para referirse al mismo aminoácido que tiene una clase específica de modificación química en la cadena lateral. Por ejemplo, el término "derivado cargado positivamente de un aminoácido alifático" se referirá a un aminoácido alifático en el que se ha modificado la cadena lateral alifática para tener un resto cargado positivamente. Asimismo, el término "derivado cargado positivamente de un aminoácido aromático" se referirá a un aminoácido aromático en el que se ha modificado la cadena lateral para tener un resto cargado positivamente.

Par de bases (pb): como se usa en el presente documento, par de bases se refiere a una asociación de adenina (A) con timina (T), adenina (A) con uracilo (U) o de citosina (C) con guanina (G) en un ácido nucleico bicatenario.

Acontecimiento de captura: una inserción de una marca de polímero en un nanoporo que es suficiente para generar un cambio en una característica de la corriente iónica que fluye a través del nanoporo de modo que el cambio sea detectable por un electrodo sensor.

Complementario: como se usa en el presente documento, el término "complementario" se refiere al concepto amplio de complementariedad de secuencia entre regiones de dos hebras polinucleotídicas o entre dos nucleótidos a través de emparejamiento de bases. Es conocido que un nucleótido adenina puede formar enlaces de hidrógeno específicos ("emparejamiento de bases") con un nucleótido que sea timina o uracilo. De forma similar, es conocido que un nucleótido citosina puede realizar emparejamiento de bases con un nucleótido guanina.

Casete de expresión: un "casete de expresión" o "vector de expresión" es una construcción de ácido nucleico generada de forma recombinante o sintética, con una serie de elementos de ácido nucleico especificados que permiten la transcripción de un ácido nucleico particular en una célula diana. El casete de expresión recombinante se puede incorporar en un plásmido, cromosoma, ADN mitocondrial, ADN de plástido, virus o fragmento de ácido nucleico. Típicamente, la porción de casete de expresión recombinante de un vector de expresión incluye, entre otras secuencias, una secuencia de ácido nucleico que se va a transcribir y un promotor.

Heterógeno: una construcción o secuencia de ácido nucleico "heterógeno" tiene una porción de la secuencia que no es natural para la célula en la que se expresa.

Heterógeno, con respecto a una secuencia de control se refiere a una secuencia de control (es decir, promotor o potenciador) que no funciona en la naturaleza para regular el mismo gen con una expresión que se está regulando actualmente. En general, las secuencias de ácido nucleico heterógeno no son endógenas para la célula o parte del genoma en el que están presentes, y se han añadido a la célula, por infección, transfección, transformación, microinyección, electroporación o similares. Una construcción de ácido nucleico "heterógeno" puede contener una combinación de secuencia de control/secuencia codificante de ADN que es igual a, o diferente de una combinación de secuencia de control/secuencia codificante de ADN encontrada en la célula natural.

Célula huésped: por el término "célula huésped" se quiere decir una célula que contiene un vector y apoya la replicación, y/o transcripción o transcripción y traducción (expresión) de la construcción de expresión. Las células huésped para su uso en la presente invención pueden ser células procariotas, tales como *E. coli* o *Bacillus subtilus*, o células eucariotas, tales como levadura, células vegetales, de insecto, de anfibio o de mamífero. En general, las células huésped son procariotas, por ejemplo, *E. coli*.

Aislado: una molécula "aislada" es una biomolécula que se separa de al menos otra molécula con la que normalmente se asocia, por ejemplo, en su entorno natural.

Mutación: como se usa en el presente documento, el término "mutación" se refiere a un cambio introducido en una

secuencia original, incluyendo, pero sin limitarse a, sustituciones, inserciones y/o deleciones (incluyendo truncamientos). Las consecuencias de una mutación incluyen, pero no se limitan a, la creación de un nuevo carácter, propiedad, función, fenotipo o rasgo no encontrado en la proteína codificada por la secuencia original.

Nanoporo: el término "nanoporo", como se usa en el presente documento, se refiere, en general, a un poro, canal o paso formado o proporcionado de otro modo en una barrera eléctricamente resistiva (tal como una membrana lipídica, una capa de silicio, una capa polimérica o una capa de grafeno) a través del que puede pasar una corriente iónica. A menos que se establezca de otro modo, el término genérico "nanoporo" incluirá nanoporos biológicos, nanoporos de estado sólido y nanoporos híbridos.

Complejo de secuenciación de nanoporos: un sitio en el que se puede realizar un procedimiento de secuenciación basada en nanoporos, que comprende, en general, al menos un nanoporo a través del que pueden fluir iones y un electrodo sensor configurado para detectar una característica de la corriente de iones que fluye a través del nanoporo (tal como la caída de voltaje).

En el contexto de un sistema o procedimiento de secuenciación de nucleótidos SBS basada en marca, el complejo de secuenciación de nanoporos comprende al menos (a) un nanoporo a través del que pueden fluir iones; (b) una ácido nucleico polimerasa fijada a o asociada de otro modo con el nanoporo en una configuración que posibilita que la polimerasa catalice la polimerización basada en molde de nucleósidos-5'-oligofosfato marcados de modo que una marca de polímero del nucleósido-5'-oligofosfato se pueda insertar en el canal del nanoporo mientras se polimeriza el nucleósido-5'-oligofosfato; y (c) un electrodo sensor configurado para detectar una característica de la corriente de iones que fluye a través del nanoporo.

Aminoácido natural: cualquier aminoácido de una secuencia de aminoácidos que, cuando se alinea con una secuencia de aminoácidos de referencia, es igual que el aminoácido que ocupa la posición correspondiente de la secuencia de referencia.

Resto cargado negativamente no natural: un componente de un nanoporo que tiene una carga neta negativa a pH 7,0 que no se encuentra en una estructura de referencia. Por ejemplo, cuando el nanoporo incluye un polipéptido, un "aminoácido cargado negativamente no natural" sería cualquier aminoácido que tenga una cadena lateral con una carga neta negativa a pH 7,0 que represente una sustitución o una inserción en una posición particular en relación con una secuencia de aminoácidos de referencia, o una que represente una modificación química de la cadena lateral de un aminoácido natural que dé como resultado una carga neta negativa a pH 7,0.

Molécula de ácido nucleico: el término "molécula de ácido nucleico" incluye moléculas de ARN, ADN y ADNc. Se entenderá que, como resultado de la degeneración del código genético, se puede producir una multitud de secuencias de nucleótidos que codifican una proteína dada, tal como α-hemolisina y/o variantes de la misma. La presente invención contempla todas las posibles secuencias de nucleótidos variantes.

Péptido: los términos "péptido" y "enlace peptídico" se referirán a cualquier enlace de cadena principal entre dos aminoácidos y/o análogos de aminoácido resultante de una reacción de condensación entre un resto de ácido carboxílico de un aminoácido o análogo de aminoácido y un grupo amino de un segundo aminoácido o análogo de aminoácido. A menos que esté claro de otro modo del contexto, estos términos se entenderán en todos los casos como que engloban (pero sin limitarse a) los enlaces entre α-aminoácidos, β-aminoácidos, γ-aminoácidos, δ-aminoácidos y combinaciones de los mismos, así como los enlaces entre restos de ácido carboxílico de cadena principal y restos de amino de cadena lateral (tales como con lisina enlazada a ε).

Cadena peptídica: el término "cadena peptídica" se referirá a cualquier secuencia de dos o más aminoácidos y/o análogos de aminoácido enlazados por enlaces peptídicos.

Peptidomimético: los términos "peptidomimético" y "enlace peptidomimético" se referirán a enlaces de cadena principal entre dos análogos de aminoácido o entre un aminoácido y un análogo de aminoácido, incluyendo, pero sin limitarse a, peptoides (aminoácidos en los que la cadena lateral se fija al grupo amino), azapéptidos (reemplazo del carbono α por un nitrógeno), oligourea (enlace peptídico reemplazado por un enlace de urea), arilamidas, oligohidracidas y similares.

Cadena peptidomimética: el término "cadena peptidomimética" se referirá a cualquier secuencia de dos o más aminoácidos y/o análogos de aminoácido enlazados por enlaces de cadena principal peptidomimética.

Polipéptido: a menos que se establezca de otro modo o a menos que esté claro de otro modo en base al contexto de la divulgación, la frase "polipéptido" se entenderá en su sentido más amplio y englobará cualquier secuencia de dos o más aminoácidos y/o análogos de aminoácido enlazados por enlaces peptídicos y/o enlaces peptidomiméticos.

Promotor: como se usa en el presente documento, el término "promotor" se refiere a una secuencia de ácido nucleico que funciona para dirigir la transcripción de un gen hacia 3'. El promotor será apropiado, en general, para

la célula huésped en la que se expresa el gen diana. El promotor conjuntamente con otras secuencias de ácido nucleico reguladoras transcripcionales y traduccionales (también denominadas "secuencias de control") son necesarios para expresar un gen dado. En general, las secuencias reguladoras transcripcionales y traduccionales incluyen, pero no se limitan a, secuencias promotoras, sitios de unión ribosómicos, secuencias de inicio y parada transcripcionales, secuencias de inicio y parada traduccionales y secuencias potenciadoras o activadoras.

Aminoácido proteinógeno: el enantiómero L de cualquier α-aminoácido codificado genéticamente.

Purificado: como se usa en el presente documento, "purificado" quiere decir que una molécula está presente en una muestra a una concentración de al menos un 95 % en peso, o al menos un 98 % en peso de la muestra en la que está contenida.

Marca: como se usa en el presente documento, el término "marca" se refiere a un resto detectable por nanoporo que puede ser átomos o moléculas, o un grupo de átomos o moléculas. Una marca puede proporcionar una firma óptica, electroquímica, magnética o electrostática (por ejemplo, inductiva, capacitiva), firma que se puede detectar con la ayuda de un nanoporo. Típicamente, cuando un nucleótido se fija a la marca, se le llama "nucleótido marcado".

Tiempo de paso: el término "tiempo de paso" o "TTT" quiere decir el tiempo que tarda una marca para pasar hacia el barril del nanoporo después de asociarse con una ácido nucleico polimerasa asociada con el nanoporo.

Variante: como se usa en el presente documento, el término "variante" de un polipéptido de referencia o de un ácido nucleico es cualquier molécula que contiene al menos un cambio molecular en relación con la molécula de referencia.

Vector: como se usa en el presente documento, el término "vector" se refiere a una construcción de ácido nucleico diseñada para la transferencia entre diferentes células huésped. Un "vector de expresión" se refiere a un vector que tiene la capacidad de incorporar y expresar fragmentos de ADN heterógeno en una célula exógena. Muchos vectores de expresión procariotas y eucariotas están disponibles comercialmente. La selección de vectores de expresión apropiados está dentro del conocimiento de los expertos en la técnica.

Porcentaje de homología: el término "% de homología" se usa de manera intercambiable en el presente documento con el término "% de identidad" en el presente documento y se refiere al nivel de identidad de secuencia de ácido nucleico o aminoácidos entre la secuencia de ácido nucleico que codifica uno cualquiera de los polipéptidos según la invención o la secuencia de aminoácidos del polipéptido según la invención, cuando se alinea usando un programa de alineación de secuencias. Por ejemplo, como se usa en el presente documento, un 80 % de homología quiere decir lo mismo que un 80 % de identidad de secuencia determinada por un algoritmo definido y, en consecuencia, un homólogo de una secuencia dada tiene más de un 80 % de identidad de secuencia en una longitud de la secuencia dada. Los niveles ejemplares de identidad de secuencia incluyen, pero no se limitan a, un 80, 85, 90, 95, 98 % o más de identidad de secuencia con respecto a una secuencia dada, por ejemplo, la secuencia codificante para uno cualquiera de los polipéptidos según la invención, como se describe en el presente documento. Los programas informáticos ejemplares que se pueden usar para determinar la identidad entre dos secuencias incluyen, pero no se limitan a, el paquete de programas BLAST, por ejemplo, BLASTN, BLASTX y TBLASTX, BLASTP y TBLASTN, disponibles públicamente en Internet. Véanse también, Altschul, *et al.*, 1990 y Altschul, *et al.*, 1997. Las búsquedas de secuencias se llevan a cabo típicamente usando el programa BLASTN cuando se evalúa una secuencia de ácido nucleico dada en relación con las secuencias de ácido nucleico en las secuencias de ADN de GenBank y otras bases de datos públicas. El programa BLASTX se puede usar para buscar secuencias de ácido nucleico que se han traducido en todos los marcos de lectura frente a secuencias de aminoácidos en las secuencias de proteínas de GenBank y otras bases de datos públicas. Tanto BLASTN como BLASTX se ejecutan usando parámetros predeterminados de una penalización por hueco abierto de 11,0 y una penalización por hueco extendido de 1,0, y utilizan la matriz BLOSUM-62. (Véase, por ejemplo, Altschul, S. F., *et al.*, Nucleic Acids Res. 25:3389-3402, 1997). A menos que se establezca de otro modo, la referencia a una alineación de dos secuencias de aminoácidos se referirá a una alineación que se puede obtener usando la herramienta de alineación de secuencias por pares EMBOSS Needle con la matriz BLOSUM62, configuración de ABRIR HUECO de 10, configuración de EXTENDER HUECO de 0,5, configuración de PENALIZACIÓN POR HUECO DE EXTREMO de "falso", configuración de ABRIR HUECO DE EXTREMO de 10 y configuración de EXTENDER HUECO DE EXTREMO de 0,5 (disponible en EMBL-EBI).

Plegómero: un oligómero con una tendencia característica a plegarse en una estructura específica en solución que se estabiliza por interacciones no covalentes entre subunidades no contiguas.

Subunidad de monómero: una subunidad estructural de un complejo proteico multimérico. Por ejemplo, un poro de α-hemolisina heptamérico comprende siete subunidades monoméricas de α-hemolisina. Una subunidad monomérica que no se ha oligomerizado en una subunidad multimérica se denomina en el presente documento "subunidad monomérica no oligomerizada".

II. Antecedentes del desarrollo

La FIG. 1 ilustra un experimento de secuenciación por síntesis (SBS) basada en marca estándar usando un nanoporo de α-hemolisina y marcas cargadas negativamente. La banda oscura en la parte superior es el nivel de canal abierto. **101** y una marca que ocupa el canal del nanoporo se registra como un cambio en la señal (en este caso, el nivel de conductancia) en relación con el canal abierto, dando como resultado diferentes marcas diferentes cambios en la señal **102a–102d**. Sin embargo, los autores de la presente invención han observado que con frecuencia se observa una banda de fondo persistente **103**. El fondo incrementado da como resultado señales de marca complicadas y un procesamiento de señales que se incrementa a medida que se incrementa la tasa de paso. Esto limita por naturaleza el rendimiento y la exactitud de la SBS basada en marca.

La presente divulgación demuestra que el emparejamiento de un conjunto de nucleósidos-5'-oligofosfato que tienen una marca cargada positivamente con nanoporos que tienen un canal con más cargas negativas que los nanoporos usados típicamente para la secuenciación basada en nanoporos mitiga este problema. Sin quedar vinculados por la teoría, el patrón anómalo puede resultar al menos en parte del paso del ácido nucleico molde y/o cebador a través del nanoporo, y esa adición de cargas negativas al canal puede proporcionar una fuerza de repulsión contra el molde y el cebador cargados negativamente.

III. Polipéptidos de alfa-hemolisina y ácidos nucleicos asociados y nanoporos

Los nanoporos de α-HL son estructuras heptaméricas formadas a partir de 7 subunidades monoméricas del polipéptido de α-HL de *Staphylococcus aureus*. Se pueden encontrar diversos enfoques para genomanipular nanoporos de α-HL para su uso en secuenciación basada en nanoporos, por ejemplo, en Ayub, Wang II, documentos WO 2014/100481, WO 2016/069806, WO 2017/050718, WO 2017/184866 y WO 2018/002125. Como se ilustra en la FIG. 2A, los nanoporos de α-HL tienen una región de caperuza 201 y una región de barril beta 202, con un canal 203 que se extiende axialmente a través de las regiones de caperuza y tronco. La FIG. 2B es una sección transversal del nanoporo de α-HL que muestra el canal 203, que se puede dividir en un vestíbulo 204, un sitio de constricción 205, un cuerpo de barril beta 206 y una salida de barril beta 207. Las referencias en el presente documento a la "región de barril beta" incluyen cada uno del sitio de constricción 205, el cuerpo de barril beta 206 y la salida de barril beta 207. Las referencias en el presente documento a "nanoporo de α-HL" se referirán a poros heptaméricos de 7 subunidades monoméricas de α-HL.

IIIA. Polipéptidos de α-HL que tienen al menos un aminoácido cargado negativamente no natural

En un aspecto, la presente divulgación se refiere a polipéptidos útiles para formar nanoporos de α-HL. Los polipéptidos divulgados en el presente documento comprenden una o más subunidades monoméricas de α-HL que tienen al menos un aminoácido cargado negativamente no natural (tal como ácido aspártico o ácido glutámico) en una posición correspondiente a la región de vestíbulo y/o la región de barril beta del nanoporo de α-HL. Una secuencia de aminoácidos correspondiente a una subunidad monomérica de α-HL natural se puede encontrar en la SEQ ID NO: 1. A menos que se indique de otro modo, toda la numeración de aminoácidos relacionada con las subunidades monoméricas de α-HL son con referencia a SEQ ID NO: 1. Cuando se hace referencia a una subunidad monomérica de α-HL "que comprende una sustitución en la posición X" o "que comprende una sustitución X#Y" se entenderá que quiere decir que la secuencia de aminoácidos de subunidad monomérica, cuando se alinea con la SEQ ID NO: 1, tiene una sustitución en la posición correspondiente a la posición mencionada de SEQ ID NO: 1. Como se usa en el presente documento, un "aminoácido no natural" es un aminoácido en una posición de la secuencia de aminoácidos de subunidad monomérica que representa una sustitución o inserción cuando se alinea con la SEQ ID NO: 1. En un modo de realización, los polipéptidos comprenden al menos una subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1.

En un modo de realización, la subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1 tiene al menos un aminoácido cargado negativamente no natural (tal como ácido aspártico o ácido glutámico) en una posición correspondiente a una posición de la SEQ ID NO: 1 enumerada en la tabla 1 o la tabla 2. La tabla 1 enumera los residuos aminoacídicos orientados al disolvente que se localizan en la entrada 304, la zona de constricción 305 o el barril beta 306 cuando una subunidad monomérica que consiste en SEQ ID NO: 1 se autoensambla en un nanoporo de α-HL homoheptamérico en presencia de DPhPC en solución acuosa de Tris-HCl 20 mM, pH 8,0, NaCl 200 mM a 37 °C. "N.º" indica la posición dentro de SEQ ID NO: 1, "AA" indica el aminoácido en la posición mencionada de SEQ ID NO: 1 y "Localización" indica la subregión del nanoporo de α-HL en la que se localiza el aminoácido.

N.º	AA	Localización	N.º	AA	Localización	N.º	AA	Localización

1	Ala	Entrada	17	Asn	Entrada	125	Thr	Barril beta
2	Asp	Entrada	18	Thr	Entrada	129	Thr	Salida
3	Ser	Entrada	19	Thr	Entrada	130	Gly	Salida
4	Asp	Entrada	20	Val	Entrada	131	Lys	Salida
5	Ile	Entrada	21	Lys	Entrada	133	Gly	Barril beta
6	Asn	Entrada	22	Thr	Entrada	135	Leu	Barril beta
7	Ile	Entrada	46	Lys	Entrada	137	Gly	Barril beta
8	Lys	Entrada	47	Asn	Entrada	139	Asn	Barril beta
9	Thr	Entrada	106	Ser	Entrada	141	Ser	Barril beta
10	Gly	Entrada	111	Glu	Constricción	143	Gly	Barril beta
11	Thr	Entrada	113	Met	Constricción	145	Thr	Barril beta
12	Thr	Entrada	115	Thr	Barril beta	147	Lys	Constricción
13	Asp	Entrada	117	Thr	Barril beta	149	Val	Entrada
14	Ile	Entrada	119	Gly	Barril beta	239	Ser	Entrada
15	Gly	Entrada	121	Asn	Barril beta			
16	Ser	Entrada	123	Asn	Barril beta			

Tabla 1

La tabla 2 enumera los residuos aminoacídicos orientados al disolvente (distintos de ácido aspártico o ácido glutámico) que se localizan en la entrada 204, la zona de constricción 205, el cuerpo de barril beta 206 y la salida de barril beta 207 cuando una subunidad monomérica de α-HL que consiste en SEQ ID NO: 1 se autoensambla en un nanoporo de α-HL homoheptamérico en presencia de DPhPC en solución acuosa de Tris-HCl 20 mM, pH 8,0, NaCl 200 mM a 37 °C:

Tabla 2

N.º	AA	Localización	N.º	AA	Localización	N.º	AA	Localización
1	Ala	Entrada	18	Thr	Entrada	125	Thr	Barril beta
3	Ser	Entrada	19	Thr	Entrada	129	Thr	Salida
5	Ile	Entrada	20	Val	Entrada	130	Gly	Salida
6	Asn	Entrada	21	Lys	Entrada	131	Lys	Salida
7	Ile	Entrada	22	Thr	Entrada	133	Gly	Barril beta
8	Lys	Entrada	46	Lys	Entrada	135	Leu	Barril beta
9	Thr	Entrada	47	Asn	Entrada	137	Gly	Barril beta
10	Gly	Entrada	106	Ser	Entrada	139	Asn	Barril beta
11	Thr	Entrada	113	Met	Constricción	141	Ser	Barril beta
12	Thr	Entrada	115	Thr	Barril beta	143	Gly	Barril beta
14	Ile	Entrada	117	Thr	Barril beta	145	Thr	Barril beta
15	Gly	Entrada	119	Gly	Barril beta	147	Lys	Constricción
16	Ser	Entrada	121	Asn	Barril beta	149	Val	Entrada
17	Asn	Entrada	123	Asn	Barril beta	239	Ser	Entrada

Adicionalmente o de forma alternativa, se puede colocar un aminoácido cargado negativamente no natural en el lado N terminal de la posición 1 (denominado a continuación en el presente documento "posición 0"). En algunos modos de realización, un aminoácido cargado negativamente no natural está en un número suficiente de posiciones de la tabla 1 o la tabla 2 (y/o en la posición 0) para obtener un canal que tiene una carga neta negativa cuando se forma en un nanoporo homoheptamérico. Como se usa en el contexto de los polipéptidos de α-HL, la "carga neta del canal" es la suma de las cargas de todas las cadenas laterales de aminoácidos orientadas al disolvente dentro del canal de un nanoporo de α-HL homoheptamérico formado a partir de la subunidad monomérica en presencia de DPhPC en solución acuosa de Tris-HCl 20 mM, pH 8,0, NaCl 200 mM a 37 °C. En un modo de realización, el al menos un aminoácido cargado negativamente no natural está en una posición correspondiente a una posición de SEQ ID NO: 1 seleccionada del grupo que consiste en 0, A1, K8, T9, G10, N17, K46, N47, M113, T117, N121, T129, G130, K131, K147 y V149. En algunos modos de realización, la subunidad monomérica comprende aminoácidos cargados negativamente no naturales en un número suficiente de posiciones de SEQ ID NO: 1 seleccionadas del grupo que consiste en 0, A1, K8, T9, G10, N17, K46, N47, M113, T117, N121, T129, G130, K131, K147 y V149 para obtener un canal que tiene una carga neta negativa.

En un modo de realización, la subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos

un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1 tiene al menos un aminoácido cargado negativamente no natural en una posición de la subunidad monomérica correspondiente a la entrada del canal. Además de ayudar a repeler el ácido nucleico molde y/o cebador, la(s) carga(s) negativa(s) en la entrada tienen el efecto añadido de incrementar la tasa de llegada y/o la tasa de paso de las marcas cargadas positivamente en relación con un nanoporo de α-HL homoheptamérico en el que las subunidades monoméricas consisten en SEQ ID NO: 1. Como se usa en el presente documento, la "tasa de llegada" del nanoporo de α-HL es una medida de la frecuencia con la que el nanoporo de α-HL captura la marca de una molécula de marca biotinilada. Por ejemplo, la tasa de llegada se puede determinar obteniendo un chip que tenga una pluralidad de poros de interés insertados en la bicapa, haciendo fluir una MARCA de estreptavidina-biotina a través del chip y midiendo el tiempo promedio entre acontecimientos de captura en cada uno de la pluralidad de poros (típicamente a una frecuencia de modulación de CA muy baja, tal como ~50 Hz). La tasa de llegada es el tiempo promedio entre acontecimientos en todos los poros. Como se usa en el presente documento, la "tasa de paso" del nanoporo de α-HL es la tasa a la que una marca unida a un complejo de secuenciación de nanoporos activo se captura por el poro. Por ejemplo, en experimentos realizados con complejos de secuenciación de nanoporos activos a >1 kHz, se determina la tasa de captura de marca para cada ciclo de modulación. Típicamente, la tasa de llegada de un poro se correlaciona razonablemente con la tasa de paso del poro en un complejo de secuenciación de nanoporos activo. Los poros ejemplares que tienen una captura de marca mejorada (es decir, más rápida) en los sistemas presentes incluyen poros que comprenden subunidades monoméricas que tienen un aminoácido cargado negativamente no natural en una posición correspondiente a uno o más de 0, A1, S3, I5, N6, I7, K8, T9, G10, T11, T12, I14, G15, S16, N17, T18, T19, V20, K21, T22, K46, N47, S106 y V149 de SEQ ID NO: 1. En un modo de realización, la(s) subunidad(es) monomérica(s) tiene(n) un aminoácido cargado negativamente no natural en las posiciones 1, 2, 3, 4, 5, 6 o 7 seleccionadas del grupo que consiste en 0, A1, S3, I5, N6, I7, K8, T9, G10, T11, T12, 114, G15, S16, N17, T18, T19, V20, K21, T22, K46, N47, S106 y V149 de SEQ ID NO: 1. En un modo de realización, el/los aminoácido(s) cargado(s) negativamente no natural(es) está(n) en las posiciones 1, 2, 3, 4, 5, 6 o 7 seleccionadas del grupo que consiste en A1, K8, T9, G10, N17, K46 y N47 de SEQ ID NO: 1. En otro modo de realización, el/los aminoácido(s) cargado(s) negativamente no natural(es) está(n) en las posiciones 1, 2 o 3 seleccionadas del grupo que consiste en 0, A1, N17 y N47 de SEQ ID NO: 1.

En un modo de realización, la subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1 tiene al menos un aminoácido cargado negativamente no natural en una posición correspondiente al barril beta. Las cargas negativas no naturales introducidas en este sitio tienen el efecto añadido de incrementar la tasa de paso de las marcas cargadas positivamente en relación con un nanoporo de α-HL que comprende 7 subunidades monoméricas que tienen un 100 % de identidad con SEQ ID NO: 1. Como se usa en el presente documento, "en el barril beta" incluye residuos en el sitio de constricción, residuos en el cuerpo del barril beta y residuos en la salida del barril beta. Los residuos ejemplares en el barril beta que se podrían modificar para introducir una carga negativa incluyen M113, T115, T117, N121, T129, G130, K131 y K147 de SEQ ID NO: 1. Como ejemplo, se proporciona una subunidad monomérica que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1, en el que la subunidad monomérica tiene un ácido aspártico o un ácido glutámico en una posición correspondiente a uno o más de M113, T115, T117, N121, T129, G130, K131 y K147 de la SEQ ID NO: 1.

En un modo de realización, la subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1 tiene al menos un aminoácido cargado negativamente no natural en una posición correspondiente a la salida de barril beta. Los residuos ejemplares en la salida de barril beta que se podrían modificar para introducir una carga neta negativa incluyen T129, G130 y K131 de SEQ ID NO: 1. Como ejemplo, se proporciona una subunidad monomérica que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1, en el que la subunidad monomérica tiene un ácido aspártico o un ácido glutámico en una posición correspondiente a T129, G130 y/o K131 de SEQ ID NO: 1.

En un modo de realización, la subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1 tiene al menos un aminoácido cargado negativamente no natural en una o más posiciones correspondientes al sitio de constricción. Las cargas negativas no naturales introducidas en este sitio tienen el efecto añadido de incrementar la tasa de paso de las marcas cargadas positivamente en relación con un nanoporo de α-HL que comprende 7 subunidades monoméricas que tienen un 100 % de identidad

con SEQ ID NO: 1. Los residuos ejemplares que se podrían modificar para introducir una carga negativa en el sitio de constricción incluyen M113 y K147 de SEQ ID NO: 1. Como ejemplo, la subunidad monomérica tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1, en el que la(s) subunidad(es) monomérica(s) tiene(n) un ácido aspártico o un ácido glutámico en una posición correspondiente a M113 y/o K147 de la SEQ ID NO: 1.

En un modo de realización, la subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1 tiene al menos un aminoácido cargado negativamente no natural en una o más posiciones correspondientes al vestíbulo y una o más posiciones correspondientes al barril beta. Como ejemplo, la(s) subunidad(es) monomérica(s) tiene(n) al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1, en el que la subunidad monomérica tiene un ácido aspártico o un ácido glutámico en: (a) una posición correspondiente a uno o más de 0, A1, K8, T9, G10, N17, K46 y N47 de SEQ ID NO: 1 y (b) una posición correspondiente a uno o más de M113, T115, T117, N121, T129, G130, K131 y K147 de SEQ ID NO: 1. Como otro ejemplo, la subunidad monomérica tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1, en el que la subunidad monomérica tiene un ácido aspártico o un ácido glutámico en una posición correspondiente a A1, N17 y N47 de la SEQ ID NO: 1 y en una posición correspondiente a uno o más de M113, T115, T117, N121, T129, G130, K131 y K147 de la SEQ ID NO: 1.

En un modo de realización, la subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1 tiene al menos un aminoácido cargado negativamente no natural en una o más posiciones correspondientes al vestíbulo y en una o más posiciones correspondientes a la salida al barril beta. Como ejemplo, se proporciona una subunidad monomérica que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1, teniendo la(s) subunidad(es) monomérica(s) un ácido aspártico o un ácido glutámico en: (a) una posición correspondiente a uno o más de A1 N17 y N47 de SEQ ID NO: 1 y (b) una posición correspondiente a K131 de SEQ ID NO: 1.

En un modo de realización, la subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con la SEQ ID NO: 1 tiene al menos un aminoácido cargado negativamente no natural en una o más posiciones correspondientes al sitio de constricción y una o más posiciones correspondientes a la salida al barril beta. Como ejemplo, se proporciona(n) la(s) subunidad(es) monomérica(s) que tiene(n) al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1, y tienen un ácido aspártico o un ácido glutámico en: (a) una posición correspondiente a cualquiera o ambos de M113 y K147 de SEQ ID NO: 1 y (b) una posición correspondiente a K131 de SEQ ID NO: 1.

En un modo de realización, la subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1 tiene al menos un aminoácido cargado negativamente no natural en una o más posiciones correspondientes al vestíbulo, una o más posiciones correspondientes al sitio de constricción y una o más posiciones correspondientes a la salida al barril beta. Como ejemplo, la(s) subunidad(es) monomérica(s) tiene(n) al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1, y tienen un ácido aspártico o un ácido glutámico en: (a) una posición correspondiente a uno o más de A1, N17 y N47 de SEQ ID NO: 1, (b) una posición correspondiente a cualquiera o ambos de M113 y K147 de SEQ ID NO: 1, y (c) una posición correspondiente a K131 de SEQ ID NO: 1.

En un modo de realización, la subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al

menos un 95 % de identidad con SEQ ID NO: 1 y que tiene al menos un aminoácido cargado negativamente no natural divulgado en el presente documento puede comprender además modificaciones adicionales. Un ejemplo incluye sustituciones que ensanchan el sitio de constricción. Estas sustituciones reemplazan la cadena lateral de los aminoácidos que forman el sitio de constricción con aminoácidos que tienen cadenas laterales más cortas y/o menos voluminosas. Los ejemplos incluyen las sustituciones E111A/S, M113 A/S y K147A/S/N. En un ejemplo, al menos 3 subunidades monoméricas del nanoporo de α-HL comprenden una o más sustituciones seleccionadas del grupo que consiste en E111A/S, M113A/S y K147A/S/N. En un ejemplo, al menos 4 subunidades monoméricas del nanoporo de α-HL comprenden una o más sustituciones seleccionadas del grupo que consiste en E111A/S, M113A/S y K147A/S/N. En un ejemplo, al menos 5 subunidades monoméricas del nanoporo de α-HL comprenden una o más sustituciones seleccionadas del grupo que consiste en E111A/S, M113A/S y K147A/S/N. En un ejemplo, al menos 6 subunidades monoméricas del nanoporo de α-HL comprenden una o más sustituciones seleccionadas del grupo que consiste en E111A/S, M113A/S y K147A/S/N (incluyendo subunidades monoméricas 6:1, en las que el componente "6" tiene sustituciones correspondientes a E111A/S, M113A/S y K147A/S/N). Otro ejemplo incluye sustituciones que controlan la capacidad de las subunidades monoméricas no oligomerizadas para autooligomerizarse. Por ejemplo, las subunidades monoméricas de α-HL que tienen sustituciones en H35 (por ejemplo, sustituciones H35G/L/D/E) sustancialmente no se oligomerizan siempre que se mantengan a temperatura ambiente o inferior (por ejemplo, 25 °C o menos), pero se oligomerizarán de forma estable cuando la temperatura se eleve a una temperatura mayor (por ejemplo, 35 °C). En un modo de realización ejemplar, las subunidades monoméricas de α-HL que tienen el uno o más aminoácidos cargados negativamente no naturales comprenden además una sustitución H35L. Otros ejemplos de estrategias de sustitución para controlar la autooligomerización y/o dirigir patrones específicos de oligomerización se divulgan, por ejemplo, en el documento WO/2017/050718. Otro ejemplo incluye sustituciones que mejoran el nivel de expresión de la(s) subunidad(es) monomérica(s) de α-HL en una célula recombinante usada para expresar la(s) subunidad(es) monomérica(s). Por ejemplo, se ha demostrado que las sustituciones H35L mejoran los niveles de expresión de subunidades monoméricas de α-HL que tienen una pluralidad de aminoácidos cargados negativamente no naturales en sistemas de expresión de *E. coli*. Otros ejemplos incluyen sustituciones que reducen el coeficiente de variación de la tasa de llegada del poro (CV), tales como D227N.

Los polipéptidos comprenden, en general, de 1 a 7 subunidades monoméricas de α-HL. En un modo de realización, los polipéptidos divulgados en el presente documento comprenden una única subunidad monomérica de α-HL. En otro modo de realización, el polipéptido comprende de 2 a 7 subunidades monoméricas de α-HL (denominado a continuación en el presente documento polipéptido de α-HL concatenado), incluyendo explícitamente polipéptidos que comprenden 2 subunidades monoméricas de α-HL, polipéptidos que comprenden 3 subunidades monoméricas de α-HL, polipéptidos que comprenden 4 subunidades monoméricas de α-HL, polipéptidos que comprenden 5 subunidades monoméricas de α-HL, polipéptidos que comprenden 6 subunidades monoméricas de α-HL y polipéptidos que comprenden 7 subunidades monoméricas de α-HL. Los procedimientos ejemplares de generación de polipéptidos de α-HL concatenados y las consideraciones para hacerlo se divulgan, por ejemplo, por Hammerstein y el documento US 2017-0088890 A1. En un modo de realización, cada subunidad monomérica del polipéptido de α-HL concatenado se separa de la(s) otra(s) subunidad(es) monomérica(s) por una secuencia conectora. En un modo de realización, la secuencia conectora es un conector flexible. Los conectores flexibles ejemplares se divulgan, por ejemplo, por Hammerstein y Chen III.

Los polipéptidos también pueden incluir componentes útiles para la purificación del polipéptido, tales como, por ejemplo, marcas de epítopo, sitios de escisión de proteasas, etc.

Los polipéptidos también pueden incluir entidades útiles para la fijación de otros agentes activos (tales como polimerasas) al polipéptido (denominados en el presente documento "componentes de fijación"). Los componentes de fijación ejemplares incluyen, por ejemplo, componentes del sistema de péptidos SpyTag/SpyCatcher (Zakeri *et al.* PNAS 109: E690-E697 2012), el sistema de ligadura química natural (Thapa *et al.*, Molecules 19:14461-14483 2014), el sistema de sortasa (Wu y Guo, J Carbohydr Chem 31:48-66 2012; Heck *et al.*, Appl Microbiol Biotechnol 97:461-475 2013)), los sistemas de transglutaminasa (Dennler *et al.*, Bioconjug Chem 25:569 578 2014), los sistemas de enlace de formilglicina (Rashidian *et al.*, Bio conjug Chem 24:1277-1294 2013), un sistema de fijación de química Clic u otras técnicas de ligadura química conocidas en la técnica.

IIIB. Ácidos nucleicos, casetes de expresión, vectores de expresión, células recombinantes y procedimientos de producción de polipéptidos

En otro aspecto de la presente divulgación, se proporcionan polinucleótidos aislados, comprendiendo dicho ácido nucleico una secuencia de nucleótidos que codifica un polipéptido que comprende una o más subunidades monoméricas de α-HL que tienen al menos un aminoácido cargado negativamente no natural (tal como ácido aspártico o ácido glutámico) en una posición correspondiente a la región de vestíbulo, la zona de constricción y/o el barril beta del nanoporo de α-HL, incluyendo expresamente los polipéptidos divulgados en la sección IIA. En un modo de realización, el ácido nucleico es un casete de expresión que comprende la secuencia de nucleótidos que codifica el polipéptido enlazado a un conjunto de elementos de transcripción de ácido nucleico (tales como promotores, potenciadores, codones de inicio y parada, sitios de unión ribosómica y similares) suficientes para la transcripción de la secuencia de nucleótidos que codifica el polipéptido en una célula procariota o eucariota o en un

sistema de expresión sin células.

En otro aspecto, se proporciona un vector que comprende el nucleótido que codifica el polipéptido. Los vectores pueden ser, por ejemplo, vectores de clonación o de expresión. Las cadenas principales de vector adecuadas incluyen, por ejemplo, las usadas de forma rutinaria en la técnica, tales como plásmidos, cromosomas artificiales, BAC o PAC. Numerosos vectores y sistemas de expresión están disponibles comercialmente en dichas empresas como Novagen (Madison, Wis.), Clonetech (Pal Alto, Calif.), Stratagene (La Jolla, Calif.) e Invitrogen/Life Technologies (Carlsbad, Calif.). Los vectores contienen típicamente una o más regiones reguladoras. Las regiones reguladoras incluyen, sin limitación, secuencias promotoras, secuencias potenciadoras, elementos de respuesta, sitios de reconocimiento de proteínas, elementos inducibles, secuencias de unión a proteína, regiones no traducidas (UTR) 5' y 3', sitios de inicio de la transcripción, secuencias de finalización, secuencias de poliadenilación, etc.

En otro modo de realización, se proporciona una célula huésped que comprende el vector de expresión. Por ejemplo, una célula huésped útil para la producción de polipéptidos se transforma o se transfecta de forma transitoria o estable con el vector de expresión. En otro aspecto de la presente divulgación, se proporciona un procedimiento de preparación de un polipéptido que comprende una o más subunidades monoméricas de α-HL que tienen al menos un aminoácido cargado negativamente no natural (tal como ácido aspártico o ácido glutámico) en una posición correspondiente a la región de vestíbulo, la zona de constricción y/o el barril beta del nanoporo de α-HL como se divulga en el presente documento (que incluye expresamente los polipéptidos divulgados en la sección IIA), comprendiendo el procedimiento (a) cultivar una célula huésped que comprende un vector de expresión como se divulga en el presente documento en condiciones suficientes para inducir la expresión del polipéptido, y (b) purificar el polipéptido a partir de la célula huésped. Dichos procedimientos son bien conocidos en la técnica y están disponibles comercialmente muchos sistemas para hacerlo.

IIIC. Nanoporos de α-HL

En un modo de realización, se proporciona un nanoporo de α-HL heptamérico o un nanoporo híbrido que comprende un nanoporo de α-HL heptamérico como componente biológico, que tiene una pluralidad de aminoácidos cargados negativamente no naturales (tales como ácido aspártico o ácido glutámico) en una o más de la región de vestíbulo, la zona de constricción y el barril beta del nanoporo de α-HL. Los nanoporos de α-HL divulgados en el presente documento tienen un canal que comprende al menos un aminoácido cargado negativamente no natural en una posición que se correlaciona con una o más subregiones del canal seleccionadas del grupo que consiste en la región de vestíbulo, la zona de constricción y el barril beta. Se proporcionan suficientes aminoácidos cargados negativamente no naturales en una o más de las localizaciones anteriores para que los ácidos nucleicos molde y/o cebadores sustancialmente se repelan. Adicionalmente o de forma alternativa, se proporcionan suficientes aminoácidos cargados negativamente no naturales en una o más de las localizaciones anteriores de modo que las marcas cargadas positivamente se puedan translocar a través del canal del nanoporo. Adicionalmente o de forma alternativa, se proporcionan suficientes aminoácidos cargados negativamente no naturales en una o más de las localizaciones anteriores de modo que se incrementa la tasa de llegada y/o la tasa de paso de una marca cargada positivamente en relación con la tasa de un nanoporo que tiene el aminoácido natural en el mismo sitio.

Cada subunidad monomérica del nanoporo de α-HL puede tener la misma secuencia de aminoácidos primaria (denominada "homoheptámero"), o al menos una subunidad monomérica del heptámero puede tener una secuencia de aminoácidos que sea diferente de la secuencia de aminoácidos de las otras subunidades monoméricas (denominada "heteroheptámero"). Los nanoporos de α-HL heteroheptaméricos se pueden denominar en el presente documento por una proporción de las especies de diferentes subunidades monoméricas usadas en el nanoporo. Por ejemplo, un "nanoporo de α-HL 6:1" tiene 6 subunidades monoméricas con la misma secuencia de aminoácidos y 1 subunidad monomérica con una secuencia de aminoácidos diferente. En un ejemplo de este tipo, la referencia al componente "6" querrá decir cada una de las 6 subunidades monoméricas idénticas, mientras que la referencia al componente "1" querrá decir la 1 subunidad monomérica con la secuencia de aminoácidos diferente. En algunos modos de realización, cada subunidad monomérica del nanoporo de α-HL se dispone en un polipéptido que no contiene subunidades monoméricas adicionales (denominado en el presente documento "subunidad monomérica no oligomerizada"). Los procedimientos ejemplares de preparación de homoheptámeros y heteroheptámeros a partir de subunidades monoméricas no oligomerizadas se divulgan en el documento US 2017-0088890 A1. Por ejemplo, se pueden generar heteroheptámeros 6:1 mezclando dos preparaciones de monómeros diferentes (por ejemplo, una en la que el monómero se modifica con una entidad que se puede usar para unirse a una polimerasa y otra entidad que no contiene una modificación de este tipo). La entidad que se pretende que esté en exceso en el heptámero resultante se proporciona en un exceso molar en relación con el otro heptámero en presencia de una membrana y se incuba la mezcla en una solución acuosa (tal como Tris-HCl 20 mM, pH 8,0, NaCl 200 mM o citrato de sodio 20 mM, pH 3, NaCl 400 mM, TWEEN20 al 0,1 % + TMAO 0,2 M) durante la noche a 37 °C. A continuación, se purifican los heptámeros resultantes por cromatografía de intercambio catiónico. En algunos modos de realización, se realiza la oligomerización en presencia de N-óxido de trimetilamina (TMAO), tal como TMAO de 0,1 a 5 M, TMAO de 1 a 4 M y similares. En un modo de realización, una subunidad monomérica de α-HL que tiene un conjunto de sustituciones en relación con SEQ ID NO: 1 que

comprende una sustitución H35G y al menos un aminoácido cargado negativamente no natural se oligomeriza en presencia de un tampón acuoso que comprende TMAO de 0,1 a 5 M a 37 °C. En otro modo de realización, una subunidad monomérica de α-HL que tiene un conjunto de sustituciones en relación con SEQ ID NO: 1 que comprende una sustitución H35G y al menos un aminoácido cargado negativamente no natural se oligomeriza en presencia de un tampón acuoso que comprende TMAO de 0,2 a 4 M a 37 °C. En otro modo de realización, una subunidad monomérica de α-HL que tiene un conjunto de sustituciones en relación con SEQ ID NO: 1 que comprende una sustitución H35G y al menos un aminoácido cargado negativamente no natural se oligomeriza en presencia de un tampón acuoso que comprende TMAO de aproximadamente 0,2 M a aproximadamente 3 M a 37 °C. En otros modos de realización, el nanoporo incluye al menos un conjunto de subunidades monoméricas concatenadas. Los procedimientos ejemplares de preparación de nanoporos de α-HL a partir de subunidades monoméricas concatenadas de subunidades monoméricas de α-HL se divulgan, por ejemplo, en Hammerstein y el documento US 2017-0088890 A1.

En un modo de realización, se proporciona un nanoporo de α-HL heptamérico que comprende 7 subunidades monoméricas que tienen al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1, en el que 1, 2, 3, 4, 5, 6 o 7 de las subunidades monoméricas tiene al menos un aminoácido cargado negativamente no natural (tal como ácido aspártico o ácido glutámico) en una posición correspondiente a una posición de SEQ ID NO: 1 enumerada en la tabla 1 o la tabla 2. En algunos modos de realización, el nanoporo de α-HL comprende un número suficiente de aminoácidos cargados negativamente no naturales en una posición de la tabla 1 o la tabla 2 para obtener un canal que tiene una carga neta negativa. Como se usa en el contexto de los nanoporos de α-HL, la "carga neta del canal" es la suma de las cargas de todas las cadenas laterales de aminoácidos orientadas al disolvente dentro del canal del nanoporo en presencia de DPhPC en solución acuosa de Tris-HCl 20 mM, pH 8,0, NaCl 200 mM a 37 °C. En un modo de realización, el al menos un aminoácido cargado negativamente no natural está en una posición correspondiente a una posición de SEQ ID NO: 1 seleccionada del grupo que consiste en A1, K8, T9, G10, N17, K46, N47, M113, T117, N121, T129, G130, K131, K147 y V149. En algunos modos de realización, el nanoporo de α-HL comprende un número suficiente de aminoácidos cargados negativamente no naturales en una posición de SEQ ID NO: 1 seleccionados del grupo que consiste en A1, K8, T9, G10, N17, K46, N47, M113, T117, N121, T129, G130, K131, K147 y V149 para obtener un canal que tiene una carga neta negativa.

Los nanoporos de α-HL descritos en el presente documento pueden incluir también una polimerasa fijada a los mismos. En un modo de realización, se fija una única polimerasa al nanoporo de α-HL. Las polimerasas ejemplares incluyen las derivadas de la ADN polimerasa del fago phiCPV4 de *Clostridium* (descrita por el n.º de acceso de GenBank YP_00648862, denominada en el presente documento "Pol6"), la ADN polimerasa phi29, la ADN pol T7, la ADN pol T4, la ADN pol 1 de *E. coli*, el fragmento Klenow, la ARN polimerasa T7 y la ARN polimerasa de *E. coli*, así como las subunidades y cofactores asociados. En un modo de realización, la polimerasa es una ADN polimerasa derivada de Pol6. Los derivados de Pol6 ejemplares útiles en la secuenciación basada en nanoporos se divulgan, por ejemplo, en los documentos US 2016/0222363, US 2016/0333327, US 2017/0267983, US 2018/0094249 y US 2018/0245147. Los procedimientos ejemplares de fijación de una polimerasa a un nanoporo de α-HL incluyen el sistema de péptidos SpyTag/SpyCatcher (Zakeri *et al*. PNAS 109: E690-E697 2012), el sistema de ligadura química natural (Thapa *et al*., Molecules 19:14461-14483 2014), el sistema de sortasa (Wu y Guo, J Carbohydr Chem 31:48-66 2012; Heck *et al*., Appl Microbiol Biotechnol 97:461-475 2013)), los sistemas de transglutaminasa (Dennler *et al*., Bioconjug Chem 25:569 578 2014), los sistemas de enlace de formilglicina (Rashidian *et al*., Bio conjug Chem 24:1277-1294 2013), los sistemas de fijación de química Click u otras técnicas de ligadura química conocidas en la técnica. En un modo de realización, la polimerasa se fija a una cadena lateral de aminoácido de una de las subunidades monoméricas. En un modo de realización, la polimerasa se fija a una cadena lateral de aminoácido de una de las subunidades monoméricas. En un modo de realización, el nanoporo de α-HL es un nanoporo 6:1, en el que la polimerasa se fija al componente "1". En un modo de realización, el nanoporo de α-HL es un nanoporo 6:1, en el que la polimerasa se fija al componente "1", y en el que la polimerasa es una ADN polimerasa. En otro modo de realización, el nanoporo de α-HL es un nanoporo 6:1, en el que la polimerasa se fija al componente "1", y en el que la polimerasa es una ADN polimerasa derivada de Pol6.

IV. Nucleósidos-5'-oligofosfato marcados

En un aspecto, se divulgan en el presente documento nucleótidos marcados, comprendiendo dichos nucleótidos marcados un resto de nucleósido-5'-oligofosfato enlazado covalentemente a un resto detectable por nanoporo (denominado a continuación en el presente documento "marca") que comprende un polímero cargado positivamente.

Las marcas de polímero comprenden un segmento de 5 unidades de monómero a 100 unidades de monómero de longitud que tiene una carga neta positiva de al menos +5 a pH 7,0 (denominado a continuación en el presente documento "segmento cargado positivamente" o "SCP"). Cuando se usan con nanoporos modificados como se describe en el presente documento, las marcas que tienen un SCP tienen tasas de captura y tasas de translocación mejoradas en relación con las marcas cargadas negativamente. Como se usa en el presente documento, una "unidad de monómero" es una subunidad monomérica de un polímero cuando se polimeriza. Las unidades de

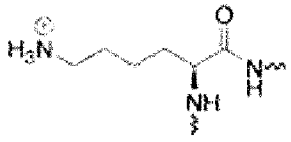
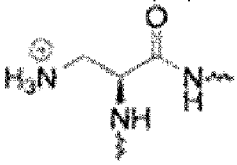
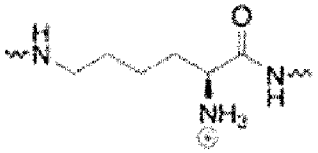
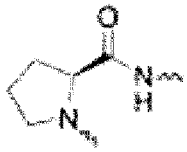
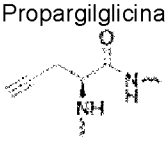
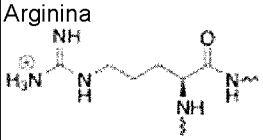
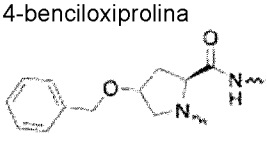
monómero ejemplares incluyen aminoácidos, análogos de aminoácido, poli(compuestos de amina) lineales o ramificados o cíclicos, compuestos de amina cuaternizada, fosfinas cuaternizadas, glicoles o centros metálicos que contienen ligandos coordinados. Como se usa en el presente documento, una "unidad de monómero cargada positivamente" es una unidad de monómero que tiene una carga neta positiva a pH 7,0 en una solución acuosa tamponada (por ejemplo, HEPES 20 mM o HEPES 50 mM). Como se usa en el presente documento, una "unidad de monómero sin carga" es una unidad de monómero que tiene una carga neta neutra en una solución acuosa tamponada de forma neutra (por ejemplo, HEPES 20 mM o HEPES 50 mM). Como se usa en el presente documento, una "unidad de monómero cargada negativamente" es una unidad de monómero que tiene una carga neta negativa en una solución acuosa tamponada de forma neutra (por ejemplo, HEPES 20 mM o HEPES 50 mM). En un modo de realización, el SCP tiene una proporción de al menos 1 unidad de monómero cargada positivamente con respecto a 5 unidades de monómero sin carga y una proporción de más de 1 unidad de monómero cargada positivamente con respecto a cada unidad de monómero cargada negativamente. En un modo de realización, el SCP comprende menos de 1 unidad de monómero cargada negativamente por cada 2 unidades de monómero. En un modo de realización, el SCP comprende menos de 1 unidad de monómero cargada negativamente por cada 3 unidades de monómero. En un modo de realización, el SCP comprende menos de 1 unidad de monómero cargada negativamente por cada 4 unidades de monómero. En un modo de realización, el SCP comprende menos de 1 unidad de monómero cargada negativamente por cada 5 unidades de monómero. En un modo de realización, el SCP comprende menos de 1 unidad de monómero cargada negativamente por cada 10 unidades de monómero. En un modo de realización, el SCP comprende menos de 1 unidad de monómero cargada negativamente por cada 20 unidades de monómero. En un modo de realización, el SCP no contiene ninguna unidad de monómero cargada negativamente. En un modo de realización, el SCP tiene al menos 5 monómeros de longitud, tiene una carga neta de al menos +5 a pH 7,0 y tiene una proporción de al menos 1 unidad de monómero cargada positivamente con respecto a 5 unidades de monómero sin carga y una proporción de más de 1 unidad de monómero cargada positivamente con respecto a cada unidad de monómero cargada negativamente. En un modo de realización, el SCP tiene al menos 8 monómeros de longitud, tiene una carga neta de al menos +5 a pH 7,0 y tiene una proporción de al menos 1 unidad de monómero cargada positivamente con respecto a 5 unidades de monómero sin carga y una proporción de más de 1 unidad de monómero cargada positivamente con respecto a cada unidad de monómero cargada negativamente. En un modo de realización, el SCP tiene al menos 5 monómeros de longitud y tiene una carga neta de al menos +5 a pH 7,0. En un modo de realización, el SCP tiene una densidad de carga de al menos 0,1. En otro modo de realización, el SCP tiene una densidad de carga de al menos 0,2. En otro modo de realización, el SCP tiene una densidad de carga de al menos 0,25. En otro modo de realización, el SCP tiene una densidad de carga de al menos 0,3. En otro modo de realización, el SCP tiene una densidad de carga de al menos 0,4. En otro modo de realización, el SCP tiene una densidad de carga de al menos 0,5. En otro modo de realización, el SCP tiene una densidad de carga de al menos 0,6. En otro modo de realización, el SCP tiene una densidad de carga de al menos 0,7. En otro modo de realización, el SCP tiene una densidad de carga de al menos 0,8. En otro modo de realización, el SCP tiene una densidad de carga de al menos 0,9. Como se usa en el contexto de un SCP, la "densidad de carga" se determina dividiendo la carga neta del SCP entre el número total de unidades de monómero del SCP.

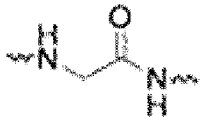
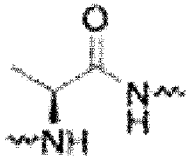
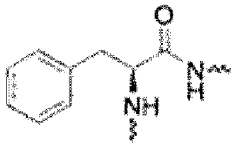
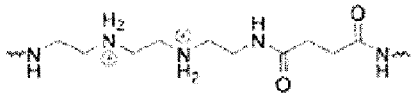
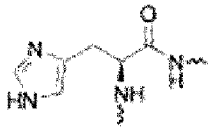
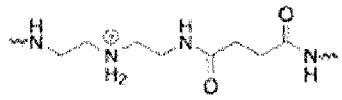
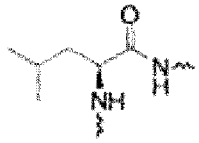
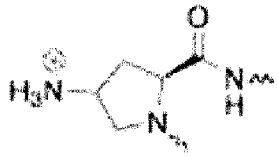
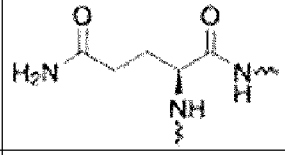
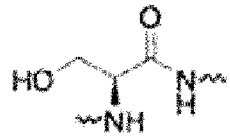
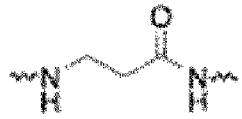
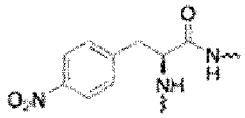
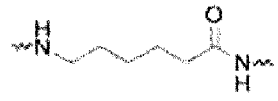
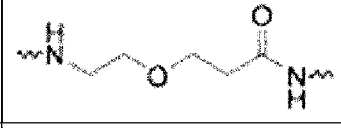
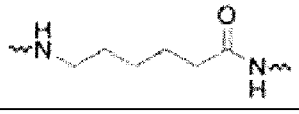
En un modo de realización, el SCP comprende una secuencia homopolimérica de unidades de monómero cargadas positivamente, con la condición de que dichas unidades de monómero cargadas positivamente no sean lisina enlazada a α. Las secuencias de SCP homopoliméricas ejemplares incluyen secuencias de aminoácidos o derivados de aminoácido, tales como lisina enlazada a épsilon, ácido aminoetil-piperacinacético, ácido trietilentriamino-succinámico, ácido dietilentriamino-succinámico, aminoetilglicina y 4-aminoprolina. En un modo de realización, el SCP comprende una secuencia heteropolimérica de unidades de monómero, en el que al menos una porción de las unidades de monómero son unidades de monómero cargadas positivamente.

En un modo de realización, el SCP es una secuencia heteropolimérica que comprende unidades de monómero sin carga y/o cargadas negativamente además de las unidades de monómero cargadas positivamente. Las unidades de monómero sin carga y las unidades de monómero cargadas negativamente pueden ser útiles en el SCP, por ejemplo, para ajustar la densidad de carga del SCP (por ejemplo, ajustando la distancia entre cada carga positiva en el SCP o neutralizando una porción de las unidades de monómero cargadas positivamente), para impartir estructuras secundarias específicas en el SCP (por ejemplo, induciendo giros en la cadena principal del SCP para formar una estructura helicoidal y/o estabilizando estructuras secundarias por medio de interacciones electrostáticas) y utilizando secciones lineales o ramificadas o cíclicas que contienen una única o múltiples cargas positivas. En algunos modos de realización, las unidades de monómero cargadas positivamente se distribuyen a lo largo de toda la longitud del SCP. La distribución de unidades de monómero cargadas positivamente se puede evaluar determinando la densidad de carga de longitudes predefinidas del SCP. Por ejemplo, si la longitud predefinida es de 10 unidades de monómero, se determina la densidad de carga de cada secuencia de 10 unidades de monómero dentro del SCP. En este contexto, la "densidad de carga" se determinaría dividiendo la carga neta de la secuencia entre el número de unidades de monómero dentro de la secuencia. En un modo de realización, al menos un 50 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,1. En un modo de realización, al menos un 60 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,1. En un modo de realización, al menos un 70 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,1. En un modo de realización, al menos un 80 % de las 5 secuencias de unidades de monómero del SCP tienen una

densidad de carga de al menos 0,1. En un modo de realización, al menos un 90 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,1. En un modo de realización, al menos un 50 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,2. En un modo de realización, al menos un 60 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,2. En un modo de realización, al menos un 70 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,2. En un modo de realización, al menos un 80 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,2. En un modo de realización, al menos un 90 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,2. En un modo de realización, al menos un 50 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,3. En un modo de realización, al menos un 60 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,3. En un modo de realización, al menos un 70 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,3. En un modo de realización, al menos un 80 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,3. En un modo de realización, al menos un 90 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,3. En un modo de realización, al menos un 50 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,4. En un modo de realización, al menos un 60 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,4. En un modo de realización, al menos un 70 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,4. En un modo de realización, al menos un 80 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,4. En un modo de realización, al menos un 90 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,4. En un modo de realización, al menos un 50 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,5. En un modo de realización, al menos un 60 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,5. En un modo de realización, al menos un 70 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,5. En un modo de realización, al menos un 80 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,5. En un modo de realización, al menos un 90 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,5. En un modo de realización, al menos un 50 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,6. En un modo de realización, al menos un 60 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,6. En un modo de realización, al menos un 70 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,6. En un modo de realización, al menos un 80 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,6. En un modo de realización, al menos un 90 % de las 5 secuencias de unidades de monómero del SCP tienen una densidad de carga de al menos 0,6.

En un modo de realización, el SCP comprende, consiste esencialmente en, o consiste en polímeros de unidades de monómero expuestas en la tabla 3:

K	Lisina 	Dap	Ácido diaminopropiónico 
K'	Lisina enlazada a ε 	Dapa	Ácido diaminopropiónico sustituido con aminoetilo
P	Prolina 	Pra	Propargilglicina 
R	Arginina 	BnoP	4-benciloxiprolina 

G	Glicina 	Apa	Ácido aminoetil-piperacinacético
A	Alanina 	PyrAla	Pirrolil-alanina
F	Fenilalanina 	TETA	Ácido trietilentriamino-succinámico 
H	Histidina 	DETA	Ácido dietilentriamino-succinámico 
L	Leucina 	4Ap	4-aminoprolina (ambos isómeros *cis* y *trans*) 
Q	Glutamina 	Aeg	Aminoetilglicina 
S	Serina 	bAla	β-alanina 
4Npa	4-Nitrofenilalanina 	Ahx	Ácido aminohexanoico 
2Aba	Ácido 2-aminobenzoico	3Aba	Ácido 3-aminobenzoico
4Aba	Ácido 4-aminobenzoico	OAhx	Ácido oxoaminohexanoico 
Ahx	Ácido aminohexanoico 	pnaT	Ácido peptidonucleico (timina)

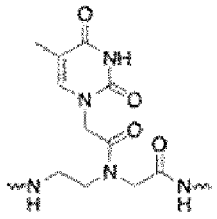
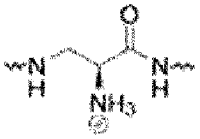
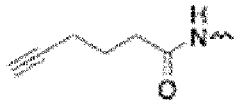
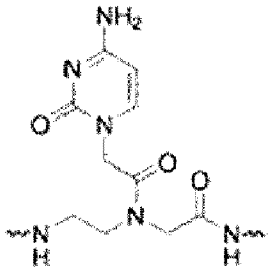
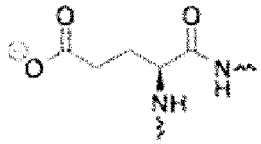
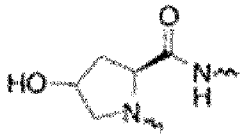
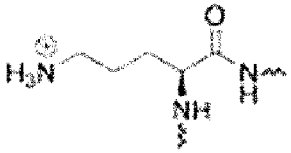
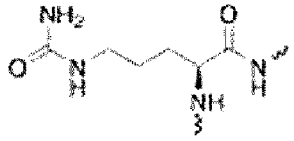
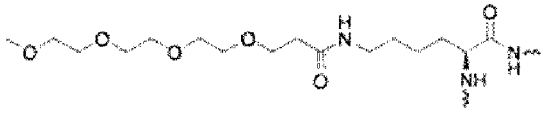
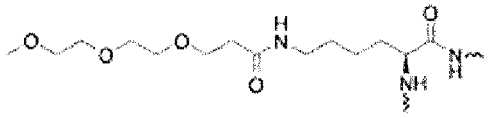
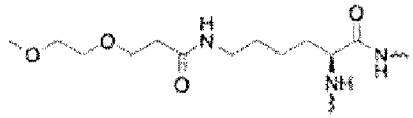
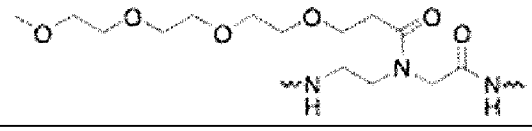
			
Dap'	Ácido gamma-diaminopropiónico 	Hex	Hexinilo 
PEG	Etilenglicol 	pnaC	Ácido peptidonucleico (citosina) 
Glu	Glutamato 	4-OHP	4-hidroxoprolina 
Orn	ornitina 	Cit	Citrulina 
KPeg4	Lisina-Peg4 		
KPeg3	Lisina-Peg3 		
Kpeg2	Lisina-Peg2 		
Aegpeg4	Aminoetilglicina-Peg4 		

Tabla 3

En un modo de realización, el SCP es un homopolímero de una unidad de monómero de la tabla 3, excepto polilisina. En otro modo de realización, el SCP es un heteropolímero que comprende, consiste esencialmente en, o consiste en unidades de monómero de la tabla 3.

En un modo de realización, el SCP comprende, consiste esencialmente en, o consiste en una cadena peptídica o una cadena peptidomimética. Cuando se usen en el contexto de estructuras de SCP, los términos "aminoácido" y "secuencias de aminoácidos" incluirán el enantiómero L, el enantiómero D y mezclas de los mismos, a menos que se indique de otro modo. A menos que se indique de otro modo, una referencia a un residuo aminoacídico específico se referirá al α-aminoácido. Las unidades de monómero de las cadenas peptídicas tradicionales son los denominados α-aminoácidos, en los que el extremo amínico y el extremo carboxílico están separados por un único carbono (denominado carbono α) al que se fija la cadena lateral. Sin embargo, se pueden insertar carbonos adicionales entre los extremos amínico y carboxílico, alargando de este modo la cadena principal del polipéptido. Los aminoácidos ejemplares modificados de esta manera incluyen β-aminoácidos (2 carbonos que separan los extremos amínico y carboxílico), γ-aminoácidos (3 carbonos que separan los extremos amínico y carboxílico) y δ-aminoácidos (4 carbonos que separan los extremos amínico y carboxílico). La cadena lateral de estos aminoácidos se puede producir en cualquiera de los carbonos entre los extremos carboxílico y amínico. Adicionalmente, la lisina tiene dos grupos amino que pueden formar un enlace peptídico: uno en el carbono α y otro en la cadena lateral en el carbono ε. Cuando se enlaza a un aminoácido contiguo por medio del grupo amino en el carbono α, se denominará aminoácido "lisina" o "K". Cuando se enlaza a un aminoácido contiguo por medio del grupo amino en el carbono ε, se denominará "lisina enlazada a ε" o "K"'. Los ejemplos adicionales de aminoácidos incluyen aminoácidos que tienen cadenas laterales no naturales, tales como 4-aminoprolina, 4-nitrofenilalanina, 4-benciloxiprolina, propargilglicina y pirrolil-alanina.

Una manera de garantizar la distribución de cargas positivas en toda la marca es diseñar el SCP para que contenga una pluralidad de unidades de repetición, teniendo cada una estructuras y/o densidades de carga iguales o similares. En un modo de realización, el SCP comprende, consiste esencialmente en, o consiste en, una o más estructuras de acuerdo con la fórmula 1:

$[\text{REPETICIÓN}]_a$ **Fórmula 1**,

en la que "REPETICIÓN" es una secuencia heteropolimérica, en la que al menos una unidad de monómero de la secuencia heteromonomérica tiene una carga neta positiva a pH 7,0, "a" es un número entero ≥ 1, y en la que "a" se selecciona de modo que la secuencia heteropolimérica tenga de 10 a 100 unidades de monómero de longitud y tenga una carga neta de al menos +5 a pH 7,0. En un modo de realización, la REPETICIÓN tiene una densidad de carga de al menos 0,1. En otro modo de realización, la REPETICIÓN tiene una densidad de carga de al menos 0,2. En otro modo de realización, la REPETICIÓN tiene una densidad de carga de al menos 0,25. En otro modo de realización, la REPETICIÓN tiene una densidad de carga de al menos 0,3. En otro modo de realización, la REPETICIÓN tiene una densidad de carga de al menos 0,4. En otro modo de realización, la REPETICIÓN tiene una densidad de carga de al menos 0,5. En otro modo de realización, la REPETICIÓN tiene una densidad de carga de al menos 0,6. En otro modo de realización, la REPETICIÓN tiene una densidad de carga de al menos 0,7. En otro modo de realización, la REPETICIÓN tiene una densidad de carga de al menos 0,8. En otro modo de realización, la REPETICIÓN tiene una densidad de carga de al menos 0,9. Como se usa en el presente documento, la "densidad de carga" se determina dividiendo la carga neta de la REPETICIÓN entre el número total de unidades de monómero de REPETICIÓN.

En otro modo de realización, el SCP comprende, consiste esencialmente en, o consiste en un único caso de fórmula 1, en la que "a" es mayor que 1. En un modo de realización de este tipo, se repite la misma unidad de repetición de fórmula 1.

En otro modo de realización, el SCP comprende, consiste esencialmente en, o consiste en múltiples casos diferentes de fórmula 1. En un modo de realización, el SCP comprende, consiste esencialmente en, o consiste en una estructura de fórmula la

REPETICIÓN 1-REPETICIÓN 2 **Fórmula 1a**,

o fórmula 1b:

REPETICIÓN 1-REPETICIÓN 2-REPETICIÓN 3 **Fórmula 1b**,

en la que cada una de la REPETICIÓN 1, la REPETICIÓN 2 y la REPETICIÓN 3 es una estructura de acuerdo con la fórmula 1 y al menos la REPETICIÓN 2 es diferente de la REPETICIÓN 1 y la REPETICIÓN 3.

En un modo de realización, se proporciona el SCP de acuerdo con la fórmula 1, fórmula la o fórmula 1b, en la que

la REPETICIÓN tiene una estructura de acuerdo con la fórmula 1c:

Xb-Yc **Fórmula 1c**,

o la fórmula 1d:

Yc-Xb **Fórmula 1d**,

en la que: X es una unidad de monómero que tiene una carga neta positiva; Y es una unidad de monómero diferente de X y polimerizable con X; y b y c son números enteros seleccionados de modo que la REPETICIÓN tenga una densidad de carga de al menos 0,1 a pH 7,0. En un modo de realización, b = 1, 2, 3, 4 o 5 y c = 1, 2, 3, 4 o 5. En un modo de realización, la REPETICIÓN es una estructura de fórmula 1c o fórmula 1d, en la que Y es una unidad de monómero sin carga, y b y c se seleccionan del grupo que consiste en: (1) b = 1 y c = 1; (2) b = 2, c = 1; (3) b = 2 y c = 2; (4) b = 2 y c = 3; y (5) b = 1 y c = 3. En un modo de realización, la REPETICIÓN es una unidad de repetición de fórmula 1c o 1d, en la que Y es una unidad de monómero sin carga, al menos uno de X e Y se selecciona para inducir una curvatura en la cadena principal de la unidad de repetición de modo que múltiples unidades de repetición entrelazadas entre sí tengan una estructura helicoidal, y en la que b y c se seleccionan del grupo que consiste en: (1) b=1 y c=1; (2) b=1 y c=2; (3) b=2 y c=1; (4) b=2 y c=2; y (5) b=2 y c=3.

En otro modo de realización, se proporciona el SCP de acuerdo con la fórmula 1, en la que la REPETICIÓN tiene una estructura de acuerdo con cualquiera de las fórmulas 1e - fórmula 1j:

X_b-Y_c-Z_d **Fórmula 1e,**
Z_d-Y_c-X_b **Fórmula 1f,**
X_b-Z_d-Y_c **Fórmula 1g,**
Y_b–Z_c–X_a **Fórmula 1h,**
Z_d-X_b-Y_c **Fórmula 1i,**
Y_c-X_b-Z_d **Fórmula 1j,**

en las que: X es una unidad de monómero que tiene una carga neta positiva; Y es una unidad de monómero polimerizable con X; Z es una unidad de monómero polimerizable con X e Y, en las que cada una de X, Y y Z es diferente de la unidad de monómero de inmediato contigua dentro de la estructura de REPETICIÓN; y b, c y d son números enteros seleccionados de modo que la REPETICIÓN tenga una densidad de carga de al menos 0,1 a pH 7,0. En un modo de realización, b = 1, 2, 3, 4 o 5; c = 1, 2, 3, 4 o 5; y d = 1, 2, 3, 4 o 5. En otro modo de realización, b = 1, 2, 3, 4 o 5; c = 1, 2, 3, 4 o 5; y d = 1, 2, 3, 4 o 5, y a + b + c ≤ 10. En este contexto, la frase "en las que cada una de X, Y y Z es diferente de la unidad de monómero de inmediato contigua dentro de la estructura de REPETICIÓN" quiere decir que cada una de X, Y y Z puede ser igual o diferente, con la condición de que al menos la variable "intermedia" en la estructura de repetición sea diferente de las otras variables. Por tanto, por ejemplo, en la fórmulas 1e y 1f, X y Z pueden ser iguales o diferentes, pero Y es diferente tanto de X como de Z. Como otro ejemplo, en la fórmulas 1g y 1h, X e Y pueden ser iguales o diferentes, pero Z es diferente tanto de X como de Y. Como otro ejemplo, en la fórmulas 1i y 1j, Y y Z pueden ser iguales o diferentes, pero X es diferente tanto de Y como de Z. En un modo de realización, la REPETICIÓN es una estructura de cualquiera de la fórmulas 1e a la fórmula 1j, en las que Y y Z son unidades de monómero sin carga, y b, c y d se seleccionan del grupo que consiste en: (1) b=1, c=1 y d=0; (2) b=1 y c+d=2; (3) b=2, c=1 y d=0; (4) b=2 y c+d=2; y (5) b=2 y c+d=3. En un modo de realización, X es una unidad de monómero cargada positivamente e Y y Z son unidades de monómero sin carga y al menos una de X, Y y Z se selecciona para inducir una curvatura en la cadena principal de la unidad de repetición, de modo que múltiples unidades de repetición entrelazadas entre sí tengan una estructura helicoidal, y en las que b, c y d se seleccionan del grupo que consiste en: (1) b=1, c=1 y d=0; (2) b=1 y c+d=2; (3) b=2, c=1 y d=0; (4) b=2 y c+d=2; y (5) b=2 y c+d=3.

En un modo de realización, las unidades de monómero de REPETICIÓN de fórmula 1 (así como X, Y y Z de las fórmulas 1c-1j) son péptidos y/o peptidomiméticos. En un modo de realización, el SCP comprende una pluralidad de unidades de repetición de acuerdo con la fórmula 1, en la que la REPETICIÓN es una cadena peptídica que comprende uno o más aminoácidos cargados positivamente seleccionados del grupo que consiste en lisina, lisina enlazada a ε, arginina, ornitina, derivados de aminoácidos alifáticos cargados positivamente (tales como aminoetilglicina), derivados de aminoácidos aromáticos cargados positivamente, derivados de prolina cargados positivamente (tales como 4-aminoprolina), Dap y Dapa. En un modo de realización, la REPETICIÓN es una estructura de acuerdo con cualquiera de la fórmula Ia-Ih, en las que X, Y y Z (si están presentes) se seleccionan cada uno de los restos de la tabla 3. En otros modos de realización, el SCP peptídico o peptidomimético también puede contener sustituyentes distintos de aminoácidos, tales como monómeros de etilenglicol (PEG), que pueden actuar como un conector hidrófobo flexible que puede coordinar el potasio para proporcionar una carga positiva, o hexinilo (Hex), que puede actuar como un resto reactivo al conector para fijar la marca a un nucleótido por medio de una reacción de Clic.

En otro modo de realización ejemplar, la REPETICIÓN es una estructura de acuerdo con cualquiera de la fórmula

la-lh, en las que al menos un Y y Z es un aminoácido o análogo de aminoácido sin carga, tal como un aminoácido alifático (tal como glicina, alanina, valina, leucina o isoleucina), un aminoácido aromático, prolina o un derivado de prolina sin carga. Por ejemplo, la inclusión de A, F, H y similares puede dar espaciamiento de las cargas a lo largo de la cadena principal de la molécula de marca para evitar la agregación de ADN o la inhibición de la ADN polimerasa o proporcionar volumen estérico adicional para impartir características de nivel deseadas o proporcionar una estructura secundaria más rígida a la molécula de marca. Como ejemplo, X es un aminoácido cargado positivamente seleccionado del grupo que consiste en lisina, β-lisina, γ-lisina, δ-lisina y lisina enlazada a ε; cada uno de Y y Z es un aminoácido sin carga seleccionado del grupo que consiste en un aminoácido alifático (tal como α-, β-, γ- o δ-glicina; α-, β-, γ- o δ-alanina; α-, β-, γ- o δ-valina; α-, β-, γ- o δ-leucina; o α-, β-, γ- o δ-isoleucina), un aminoácido aromático, prolina o un derivado de prolina; y b, c y d se seleccionan del grupo que consiste en: (1) b=1, c=1 y d=0; (2) b=1 y c+d=2; (3) b=2, c=1 y d=0; (4) b=2 y c+d=2; y (5) b=2 y c+d=3. Como otro ejemplo, Y y Z son aminoácidos sin carga; b, c y d se seleccionan del grupo que consiste en (1) b=1, c=1 y d=0; (2) b=1 y c+d=2; (3) b=2, c=1 y d=0; (4) b=2 y c+d=2; y (5) b=2 y c+d=3; y cada uno de X, Y y Z se selecciona de los restos de la tabla 3.

En un modo de realización, las unidades de monómero específicas del péptido o de la cadena peptidomemética se seleccionan para obtener un SCP que tiene una estructura secundaria específica. Por ejemplo, se pueden generar estructuras helicoidales por inclusión de prolina y derivados de prolina en las unidades de repetición. Las estructuras helicoidales proporcionan estructuras de marca más rígidas que probablemente tengan menos variabilidad de señal y brinden un nivel más discreto y también reduzcan la interacción potencial de la molécula de marca con la muestra de ADN o la polimerasa presente en la secuenciación. Adicionalmente, los aminoácidos que tienen cadenas laterales no naturales (por ejemplo, 4-aminoprolina, 4-nitrofenilalanina, 4-benciloxiprolina, propargilglicina y pirrolil-alanina) se pueden usar para generar rasgos característicos estructurales definidos para controlar el nivel de marca observado o permitir manipulaciones para la modificación posterior con los ligandos reactivos apropiados. Estos diferentes componentes básicos se pueden usar para diseñar SCP que tengan características específicamente definidas, por ejemplo, longitudes especificadas, rigidez de cadena principal, cadena lateral y/o densidad de carga a lo largo de la cadena principal, estructuras secundarias (tales como hélices y estructuras de láminas), potenciar la interacción de la polimerasa con la molécula de marca, evitar la agregación de ADN o potenciar las interacciones de paso a través del poro. Estos tipos de péptidos y peptidomiméticos con frecuencia se denominan plegómeros. Las referencias que analizan consideraciones de diseño para plegómeros peptídicos y peptidomiméticos incluyen a Licini, Martinek, Goodman, entre otros. Como ejemplo dentro de este modo de realización, el SCP es una cadena peptídica que tiene una estructura de plegómero, en la que una pluralidad de subunidades de monómero son prolina o derivados de una prolina. Por ejemplo, se puede proporcionar un SCP de fórmula 1, en el que la REPETICIÓN comprende una o más prolinas o derivados de prolina. En otro modo de realización, la REPETICIÓN de fórmula 1 comprende o consiste en una estructura de acuerdo con cualquiera de la fórmula 1c-1j, en las que al menos uno de Y y Z es una prolina o un derivado de prolina. Como otro ejemplo, el SCP comprende la estructura de fórmula 1, donde la REPETICIÓN es una estructura de acuerdo con cualquiera de la fórmula 1c-1j, y en las que: X se selecciona del grupo que consiste en lisina, β-lisina, γ-lisina, δ-lisina y lisina enlazada a ε; cada uno de Y y Z es un aminoácido sin carga seleccionado del grupo que consiste en un aminoácido alifático (tal como glicina, alanina, valina, leucina o isoleucina), un derivado sin carga de un aminoácido alifático (tal como β-, γ- o δ-glicina, β-, γ- o δ-alanina, β-, γ- o δ-valina, β-, γ- o δ-leucina, o β-, γ- o δ-isoleucina), un aminoácido aromático, un derivado sin carga de un aminoácido aromático, prolina o un derivado de prolina, con la condición de que Y o bien Z sea prolina o un derivado de prolina; y b, c y d se seleccionan del grupo que consiste en: (1) b=1, c=1 y d=0; (2) b=1 y c+d=2; (3) b=2, c=1 y d=0; (4) b=2 y c+d=2; y (5) b=2 y c+d=3. Como otro ejemplo, el SCP comprende la estructura de fórmula 1, en la que la REPETICIÓN es una estructura de acuerdo con cualquiera de la fórmula 1c-1j, en las que Y y Z son aminoácidos o análogos de aminoácido sin carga, con la condición de que al menos uno de Y y Z sea prolina o un derivado de prolina; b, c y d se seleccionan del grupo que consiste en (1) b=1, c=1 y d=0; (2) b=1 y c+d=2; (3) b=2, c=1 y d=0; (4) b=2 y c+d=2; y (5) b=2 y c+d=3; y cada uno de X, Y y Z se selecciona de los restos de la tabla 3.

Las combinaciones de variables ejemplares de la fórmula 1c-1j incluyen, pero no se limitan a, las combinaciones de la tabla 4a:

Variables de la fórmula 1c-1j						Densidad de carga	Variables de la fórmula 1c-1j						Densidad de carga
X	Y	Z	b	c	d		X	Y	Z	b	c	d	
K	P	-	1	1	-	0,5	K	S	P	1	1	2	0,25
K	P	-	2	1	-	0,66	K'	P	-	1	2	-	0,33
K	P	-	2	2	-	0,5	K'	P	-	2	1	-	0,66
K	P	-	2	3	-	0,4	K'	P	-	2	2	-	0,5
K	G	-	1	1	-	0,5	K'	BnoP	P	2	1	1	0,5
K	G	-	2	2	-	0,5	K	bAla	-	1	1	-	0,5
K	G	P	2	1	1	0,5	K	bAla	-	1	2	-	0,33
K	G	P	1	1	2	0,25	Ds	Dap	-	1	1	-	0,5

K	A	P	1	1	2	0,25	4Ap	P	-	1	1	-	0,5
K	A	-	1	3	-	0,25	4Ap	P	-	2	2	-	0,5
K	F	P	1	1	2	0,25	R	P	-	2	2	-	0,5
K	H	P	1	1	2	0,25	R	P	-	2	3	-	0,4
K	L	P	1	1	2	0,25	Aeg	P	-	1	2	-	0,33
K	Q	P	1	1	2	0,25	Aeg	P	-	2	2	-	0,5

Tabla 4a

Como se usa en la tabla 4a, la "densidad de carga" se determina dividiendo la carga neta de la unidad de repetición entre el número total de aminoácidos de la unidad de repetición. Las estructuras de REPETICIÓN específicas dentro del alcance de la fórmula 1 incluyen, pero no se limitan a, los péptidos de la tabla 4b y la secuencia retroinvertida de los mismos:

Repetición	SEQ ID NO	Repetición	SEQ ID NO	Repetición	SEQ ID NO	Repetición	SEQ ID NO
K–P	ND	K'$_2$-P$_2$	14	K'$_2$-Ahx	ND	Aeg$_2$-G	ND
K$_2$–P	ND	K'$_2$-BnoP-P	15	K'$_2$-G2	159	G-KPeg4$_2$-P	161
K$_2$–P$_2$	2	bAla–K	ND	K'$_2$-G	ND	Kpeg-G	ND
K$_2$-P$_3$	3	bAla$_2$–K	ND	K'$_2$-A	ND	Aeg-pnaT$_3$	162
K–G	ND	Ds–Dap	ND	K'$_2$-S	ND	G-Kpeg3-G-Kpeg3-P	163
K$_2$–G$_2$	4	4Ap–P	ND	K'$_2$-F	ND	Aegpeg4$_2$-P	ND
K$_2$–G–P	5	4Ap$_2$–P$_2$	16	K'$_2$-Bnop	ND	pnaC$_2$-P	ND
K–G–P$_2$	6	R$_2$–P$_2$	17	K'$_2$-Q	ND	K'-R-P$_3$	164
K–A–P$_2$	7	R$_2$–P$_3$	18	pnaT2-P	ND	K$_3$-P	165
K–A$_3$	8	Aeg–P$_2$	ND	K'-OAhx	ND	K-P-R-P	166
K–F–P$_2$	9	Aeg$_2$–P$_2$	19	K'$_2$-OAhx	ND	bAla-K	ND
K–H–P$_2$	10	Aeg$_2$-P	ND	Aeg-OAhx	ND	K$_2$-R-P	167
K–L–P$_2$	11	G-4Ap-P	ND	K'-OAhx$_2$	ND	K$_2$-pnaT	ND
K–Q–P$_2$	12	K'-K-P	ND	pnaT$_2$-G	ND	K'$_2$-pnaT	ND
K–S–P$_2$	13	K'$_2$-pnaT	ND	pnaT$_2$-K	ND	R-bAla	ND
K'-P$_2$	ND	Dap'$_2$-P$_2$	158	Kpeg4$_4$-P	ND	Acp$_2$-P	ND
K'$_2$-P	ND	K'-4Ap-P	ND	G-Kpeg-G-Kpeg-P	160	K'-K-P	ND

Tabla 4b

Como se usa en el presente documento, una "secuencia retroinvertida" de la REPETICIÓN divulgada es una unidad de repetición que tiene la misma secuencia de unidades de monómero, excepto en el orden inverso. Por tanto, por ejemplo, para un conjunto dado de variables dentro del alcance de la fórmula 1c:

X_b–Y_c **Fórmula 1c,**

la secuencia retroinvertida sería la secuencia de fórmula Id que contiene el mismo conjunto de variables:

Y_c–X_b **Fórmula 1d.**

En un modo de realización, la marca de polímero comprende, consiste en o consiste esencialmente en una estructura de acuerdo con la fórmula 2:

U–SCP–V **Fórmula 2**

en la que: U es opcional y, cuando está presente, comprende de 1 a 10 unidades de monómero; V es opcional y, cuando está presente, comprende de 1 a 10 unidades de monómero; y SCP es como se describe anteriormente. La estructura definida por la fórmula 2 tiene una carga neta positiva a pH 7,0. U se puede usar como un adaptador para fijar el SCP a un nucleósido oligofosfato, y las subunidades de monómero de U se pueden seleccionar para mantener o mejorar la interacción de la molécula de marca con la polimerasa utilizada en la secuenciación. V se puede variar para potenciar potencialmente las propiedades de atracción del SCP por el mutante de poro para mejorar las propiedades de tiempo de paso del sistema, por ejemplo, añadiendo cargas negativas adicionales. En algunos modos de realización, los constituyentes de U, SCP y V se seleccionan de modo que la marca de polímero tenga una longitud global suficiente para extenderse más allá del sitio de constricción del nanoporo cuando se hace pasar hacia el canal del nanoporo (por ejemplo, una longitud mínima igual a al menos la longitud desde el extremo

cis del vestíbulo hasta la salida del canal). En general son conocidas las dimensiones de diversos nanoporos y las distancias a los sitios de constricción. Véase, en general, Song y Wang II. Por ejemplo, un poro heptamérico de α-hemolisina tiene un canal que tiene ~10 nm de longitud con el sitio de constricción localizado a ~4,8 nm del vestíbulo. Véase Song. En un caso de este tipo, la estructura formada por U–SCP–V tendría al menos 4,8 nm de longitud para alcanzar el sitio de constricción y al menos 10 nm de longitud para extenderse a través de todo el canal del nanoporo. En un modo de realización, se proporciona una marca de polímero que tiene la estructura de acuerdo con la fórmula 2, en la que el SCP tiene una densidad de carga de al menos 0,25.

En un modo de realización, la marca de polímero de fórmula 2 comprende una cadena peptídica en la que el SCP tiene la estructura de fórmula 1. En un modo de realización, la longitud de cadena principal global de la cadena peptídica es al menos tan larga como un canal de nanoporo con el que se pretende usar la marca. Como se usa en el presente documento, el término "longitud de cadena principal global" se referirá a la longitud combinada de los enlaces que forman la cadena principal peptídica. En otro modo de realización, d es un número entero que da como resultado un SCP que tiene una longitud de cadena principal global que es al menos tan larga como un canal de un nanoporo con el que se pretende usar la marca. Por ejemplo, para una marca de péptido que se pretende usar con un poro de α-hemolisina, d puede ser un número entero que da como resultado un SCP que tiene una longitud de cadena principal global de al menos 10 nm. En otros modos de realización, la estructura de fórmula 1i es un plegómero, en la que d se selecciona de modo que la longitud global del plegómero sea al menos tan larga como un canal de un nanoporo con el que se pretende usar la marca. En un modo de realización, el SCP es una estructura de fórmula 1, en la que d es un número entero de modo que la cadena principal peptídica del SCP es de al menos 10 nm. En otro modo de realización, SCP es una estructura de fórmula 1, en la que la REPETICIÓN es una estructura de cualquiera de la fórmula 1c-1j, con la condición de que $a(c+d+d) \leq 100$, incluyendo, pero sin limitarse a ejemplos en los que d se selecciona de modo que: $7 \leq a(c+d+d) \leq 100$, o $7 \leq a(c+d+d) \leq 70$, o $7 \leq a(c+d+d) \leq 60$, u $8 \leq a(c+d+d) \leq 100$, u $8 \leq a(c+d+d) \leq 70$, u $8 \leq a(c+d+d) \leq 60$.

En un modo de realización, la marca de polímero de fórmula 2 en la que U y V tienen una carga neta neutra o una neta positiva. En un modo de realización, U comprende, consiste esencialmente en, o consiste en de 1 a 10 aminoácidos que tienen una carga neta neutra o una carga neta positiva a pH 7,0, tales como lisina, derivados de lisina (tales como lisina enlazada a ε), arginina, ornitina, aminoácidos alifáticos, derivados de aminoácidos alifáticos (tales como pirrolil-alanina, aminoetilglicina o propargilglicina), aminoácidos aromáticos, derivados de aminoácidos aromáticos (tales como 4-nitrofenilalanina), prolina, derivados de prolina (tales como 4-aminoprolina), Dap y Dapa. En otro modo de realización, U comprende, consiste esencialmente en, o consiste en uno o más aminoácidos seleccionados del grupo que consiste en lisina, derivados de lisina (tales como lisina enlazada a ε), arginina, ornitina, derivados de aminoácidos alifáticos (tales como aminoetilglicina o propargilglicina), derivados de aminoácidos aromáticos (tales como 4-nitrofenilalanina y pirrolil-alanina), derivados de prolina (tales como 4-aminoprolina), Dap y Dapa. Aún en otro modo de realización, U comprende, consiste esencialmente en, o consiste en uno o más aminoácidos seleccionados del grupo que consiste en lisina, lisina enlazada a ε, arginina, ornitina, aminoetilglicina, propargilglicina, 4-nitrofenilalanina, pirrolil-alanina, 4-aminoprolina, Dap y Dapa. En otro modo de realización, U tiene 1 o 2 aminoácidos de longitud, en el que los 1 o 2 aminoácidos se seleccionan del grupo que consiste en lisina, lisina enlazada a ε, arginina, ornitina, aminoetilglicina, propargilglicina, 4-nitrofenilalanina, pirrolil-alanina, 4-aminoprolina, Dap y Dapa. En un modo de realización, U tiene 1 o 2 aminoácidos de longitud, en el que los 1 o 2 aminoácidos se seleccionan del grupo que consiste en lisina, lisina enlazada a ε, arginina, propargilglicina, 4-nitrofenilalanina y pirrolil-alanina. En un modo de realización, U se selecciona del grupo que consiste en K, K2, K', K'K, K'K_2, K'R, Pra, PyrAla y 4Npa. En otro modo de realización, V es una cadena homopeptídica que comprende, consiste esencialmente en, o consiste en lisina, Dap o Dapa. En un modo de realización, el extremo C de V se modifica de modo que el resto de ácido carboxílico se reemplaza por una amida primaria.

En un modo de realización específico, el SCP de fórmula 2 tiene una estructura de acuerdo con la fórmula 2a:

$$U – [Xb – Y_c – Z_d]_a – V_e$$ **Fórmula 2a,**

o la fórmula 2b:

$$U – [Z_d – Y_c - X_b]_a – V_e$$ **Fórmula 2b,**

en la que: U está ausente o es una cadena de aminoácidos o análogos de aminoácido de 1 a 10 aminoácidos y/o análogos de aminoácido de longitud; a-c son números enteros y d es 0 o un número entero, con la condición de que $a(b+c+d) \leq 100$; V es un aminoácido o análogo de aminoácido cargado positivamente, y e es de 0 a 10. En un ejemplo dentro de este modo de realización, U es una cadena peptídica de 1 a 3 aminoácidos o análogos de aminoácido de longitud. En un modo de realización, V es un aminoácido o análogo de aminoácido cargado positivamente. En un modo de realización, U y V son aminoácidos o análogos de aminoácido cargados positivamente seleccionados de la tabla 3. Las estructuras de marca peptídica ejemplares dentro del alcance de la fórmula 2a o 2b incluyen, pero no se limitan a, las combinaciones de la tabla 5a:

U	X	Y	Z	V	b	c	d	a	e

K	K	P	–	K	2	2	0	1–25	1–10
K_2	K	P	—	K	1	1	0	1–50	1–10
K'K	K	P	–	K	2	3	0	1–25	1–10
K'K_2	K	P	–	K	1	1	0	1–50	1–10
K'R	R	P	–	R	2	2	0	1–25	1–10
K'R	R	P	–	R	2	3	0	1–20	1–10
K'	Aeg	P	–	K	2	2	0	1–25	1–10
K'	Aeg	P	–	K	1	2	0	1–33	1–10
K'K	K	P	–	K	2	1	0	1–33	1–10
K'K	K	P	–	K	2	3	0	1–20	1–10
K'	K'	P	–	K	2	1	0	1–33	1–10
–	K'	P	–	K	2	2	0	1–25	1–10
K'	K'	P	–	K	1	2	0	1–33	1–10
–	K'	BnoP	P	K	2	1	1	1–25	1–10
K'	K	G	P	K	1	1	2	1–25	1–10
K'	K'	P	–	K	1	2	0	1–33	1–10
–	K'	P	–	K	2	2	0	1–25	1–10
–	K'	P	–	K	2	2	0	1–25	1–10
K'	K	P	–	K	2	2	0	1–25	1–10
K'	K	P	–	K	2	2	0	1–25	1–10
K'	K	A	P	K	1	1	2	1–25	1–10
K'	K	F	P	K	1	1	2	1–25	1–10
K'	K	H	P	K	1	1	2	1–25	1–10
K'	K	L	P	K	1	1	2	1–25	1–10
K'	K	Q	P	K	1	1	2	1–25	1–10
K'	K	S	P	K	1	1	2	1–25	1–10
Pra	K	P	–	K	2	2	0	1–25	1–10
PyrAla	K	P	–	K	2	2	0	1–25	1–10
K'	K	P	–	K	2	2	0	1–25	1–10
K'	K	P	–	Dap	2	2	0	1–25	1–10
K'	K	P	–	R	2	2	0	1–25	1–10
K'	K	P	–	R	2	2	0	1–25	1–10
K'	K	P	–	Dap a	2	2	0	1–25	1–10
K'	bAla	K	–	K	1	1	0	1–50	1–10
K'	bAla	K	–	K	2	1	0	1–33	1–10
K'	Ds	Dap	–	K	1	1	0	1–50	1–10
K'	K	A	–	K	1	3	0	1–25	1–10
K'	K	G	–	K	1	1	0	1–50	1–10
K'	K	G	–	K	2	2	0	1–25	1–10
K'	K	G	P	K	2	1	1	1–25	1–10
K'	4Ap	P	–	K	2	2	0	1–25	1–10
K'	4Ap	P	–	K	2	1	0	1–33	1–10
4Npa	K	P	–	K	2	1	0	1–33	1–10
4Npa	Aeg	P	–	K	1	2	0	1–33	1–10
4Npa	K'	P	–	K	2	1	0	1–33	1–10
Hex-4NPA	Aeg	P	–	K	1	2	0	1–33	1–10
Hex-4NPA	K'	P	–	K	2	1	0	1–33	1–10
Hex-4NPA	Aeg	P	–	K	2	2	0	1–25	1–10
Hex-4NPA	Aeg	P	–	K	2	1	0	1–33	1–10
Pra	K'	P	–	K'	2	2	0	1-25	1–10
4NPA	G	4Ap	P	K	1	1	1	1–33	1–10
4NPA	K'	P	–	K'	2	1	0	1–33	1–10
Hex-4NPA	K	G	-	K	1	1	0	1–50	1–10
Hex-4NPA	K'	K	P	K	1	1	1	1–33	1–10

Hex-4NPA	K'	P	-	K	2	2	0	1–25	1–10
4NPA	K'	pnaT	-	K	2	1	0	1–33	1–10
Hex-4NPA	K	G	-	K	2	2	0	1–25	1–10
Hex-4NPA	Dap'	P	-	K	2	2	0	1–25	1–0
Hex-4NPA	K'	4Ap	P	K	1	1	1	1–33	1–10
Hex-4NPA	K'	Ahx	-	K'	2	1	0	1–33	1–10
Hex-4NPA	K'	G	-	K'	2	2	0	1–25	1–10
Hex-4NPA	K'	G	-	K'	2	1	0	1–33	1–10
Hex-4NPA	K'	A	-	K'	2	1	0	1–33	1–10
Hex-4NPA	K'	S	-	K'	2	1	0	1–33	1–10
Hex-4NPA	K'	F	-	K'	2	1	0	1–33	1–10
Hex-4NPA	K'	Bnop	-	K'	2	1	0	1–33	1–10
Hex-4NPA	K'	Q	-	K'	2	1	0	1–33	1–10
Hex-4NPA	K'	OAhx	-	K'	1	1	0	1-50	1-10
Hex-4NPA	K'	OAhx	-	K'	2	1	0	1–33	1–10
Hex-4NPA	Aeg	OAhx	-	K'	1	1	0	1–50	1–10
Hex-4NPA	K'	OAhx	-	K'	1	2	0	1–33	1–10
Hex-K	K	P	-	K	2	2	0	1–25	1–10
Hex-K	K	P	-	K	2	3	0	1–20	1–10
Hex-K'	K'	P	-	K	2	1	0	1–33	1–10
Hex-4NPA	K'	P	-	K	2	1	0	1–33	1–10
Hex-4NPA	Aeg	P	-	K	1	2	0	1–33	1–10
Hex-4NPA	Aeg	P	-	K	2	1	0	1–33	1–10
Hex-4NPA	Aeg	P	-	K	2	2	0	1–25	1–10
Hex-4NPA	K'	R	P	R	1	1	3	1–20	1–10
Hex-4NPA	K	P	-	K	3	1	0	1–25	1–10
Hex-4NPA	K	P	-	K	2	2	0	1–25	1–10
Hex-4NPA	K	G	-	K	1	1	0	1–50	1–10
Hex-4NPA	K	P	-	K	2	1	0	1–33	1–10
Hex-4NPA	U	K	-	K	1	1	0	1–50	1–10
Hex-4NPA	K	G	P	K	2	1	1	1–25	1–10
Hex-4NPA	K	G	-	K	2	2	0	1–25	1–10
Hex-4NPA	K	R	P	K	2	1	1	1–25	1–10
4NPA	K	pnaT	-	K	2	1	0	1–33	1–10
4NPA	K'	pnaT	-	K	2	1	0	1–33	1–10
4NPA	K	P	-	K	2	2	0	1–25	1–10
4NPA	R	U	-	K	1	1	0	1–50	1–10
Hex-4NPA	Acp	P	-	K	2	1	0	1–33	1–10
Hex-K'	Aeg	P	-	K	2	2	0	1–25	1–10
Pra	K'	P	-	K	2	2	0	1–25	1–10
4NPA	G	4Ap	P	K	1	1	1	1–33	1–10
4NPA	K'	P	-	K'	2	1	0	1–33	1–10
Hex-4NPA	K'	K	P	K	1	1	1	1–33	1–10
Hex-4NPA	Aeg	G	-	K	1	2	0	1–33	1–10

Tabla 5a

En otro modo de realización, se proporciona una marca dentro del alcance de la fórmula 2, en la que SCP es una secuencia homopolimérica de longitud f, y en la que las variables se seleccionan de un conjunto de variables de acuerdo con la tabla 5b.

Estructura de marca	U	SCP	V	f	e
K'–[Apa]$_f$–K$_e$	K'	Poli-Apa	K o K'	5-100	1–10
K'–[Ds]$_f$–K$_e$	K'	Poli-Ds	K o K'	5-100	1–10
K'–[Ts]$_f$–K$_e$	K'	Poli-Ts	K o K'	5-100	1–10

K′–[4Ap]$_f$–K$_e$	K′	Poli-4Ap	K o K′	5-100	1–10
4Npa–[K′]$_f$–K$_e$	4Npa	Poli-K′	K o K′	5-100	1–10

Tabla 5b

En otro modo de realización, se proporciona una marca dentro del alcance de la fórmula 2, en la que SCP es una secuencia heteropolimérica de una estructura de acuerdo con la fórmula 2c:

[REPETICIÓN 1]$_g$-[REPETICIÓN 2]$_h$-[REPETICIÓN 3]$_i$-[REPETICIÓN 4]$_j$ **Fórmula 2c**,

en la que REPETICIÓN 3 y REPETICIÓN 4 son opcionales; REPETICIÓN 1, REPETICIÓN 2, REPETICIÓN 3 (si está presente) y REPETICIÓN 4 (si está presente) se seleccionan independientemente del grupo que consiste en K′$_2$-P, PEG, Aeg$_2$-P, Aeg$_2$-G, Aeg-OAhx-P, pnaT$_2$-P, pnaT$_2$-G, pnaT$_2$-K, (Kpeg4)$_4$-P, (G-Kpeg)$_2$-P, G-(Kpeg4)$_2$-P, Kpeg-G, (G-(Kpeg3))$_2$-P, (Aegpeg4)$_2$-P y pnaC$_2$-P; g y h son números enteros de 1-30, i es un número entero de 1-30 si REPETICIÓN 3 está presente, j es un número entero de 1-30 si REPETICIÓN 4 está presente, g, h, i y j se seleccionan de modo que la longitud del SCP sea ≤ 100. En un modo de realización, REPETICIÓN 1 se selecciona del grupo que consiste en K′$_2$-P, PEG, Aeg$_2$-P, Aeg$_2$-G y Aeg-OAhx-P; REPETICIÓN 2 es diferente de REPETICIÓN 1 y se selecciona del grupo que consiste en pnaT$_2$-P, pnaT$_2$-G, pnaT$_2$-K, (Kpeg4)$_4$-P, (G-Kpeg)$_2$-P, G-(Kpeg4)$_2$-P, Kpeg-G, (G-(Kpeg3))$_2$-P, (Aegpeg4)$_2$-P, pnaC$_2$-P, PEG, Aeg$_2$-P, Aeg-OAhx-P; REPETICIÓN 2 es diferente de REPETICIÓN 1 y se selecciona del grupo que consiste en pnaT$_2$-P, pnaT$_2$-G, pnaT$_2$-K, (Kpeg4)$_4$-P, (G-Kpeg)$_2$-P, G-(Kpeg4)$_2$-P, Kpeg-G, (G-(Kpeg3))$_2$-P, (Aegpeg4)$_2$-P, pnaC$_2$-P, PEG, Aeg$_2$-P, Aeg-OAhx-P; REPETICIÓN 3, si está presente, es diferente de REPETICIÓN 2 y se selecciona del grupo que consiste en K′$_2$-P, Aeg$_2$-P, Aeg$_2$-G y pnaT$_2$-P; y REPETICIÓN 4, si está presente, es diferente de REPETICIÓN 3 y es Aeg$_2$-P. En un modo de realización específico, REPETICIÓN 1 es PEG, g es de 12-36 (incluyendo 12-24, 12 o 24), REPETICIÓN 2 se selecciona del grupo que consiste en pnaT$_2$-P, pnaT$_2$-G, pnaT$_2$-K y K′, h es de 1-5 (incluyendo 1, 2, 3, 4 o 5) y REPETICIÓN 3 y REPETICIÓN 4 están ausentes. En otro modo de realización específico, REPETICIÓN 1 es PEG, g es 24, REPETICIÓN 2 se selecciona del grupo que consiste en pnaT$_2$-P, pnaT$_2$-G, pnaT$_2$-K y K′, h es 3 y REPETICIÓN 3 y REPETICIÓN 4 están ausentes. En otro modo de realización, REPETICIÓN 1 se selecciona del grupo que consiste en K'$_2$-P, Aeg$_2$-P, Aeg$_2$-G y Aeg-OAhx-P; g es 3-10 (incluyendo 3, 4, 5, 6, 7, 8, 9 y 10); REPETICIÓN 2 es diferente de REPETICIÓN 1 y se selecciona del grupo que consiste en pnaT$_2$-P, pnaT$_2$-G, pnaT$_2$-K, (Kpeg4)$_4$-P, (G-Kpeg)$_2$-P, G-(Kpeg4)$_2$-P, Kpeg-G, (G-(Kpeg3))$_2$-P, (Aegpeg4)$_2$-P, pnaC$_2$-P, Aeg$_2$-P y Aeg-OAhx-P; h es de 1-5 (incluyendo h=1, h=2, h=3, h=4 y h=5); y REPETICIÓN 3 y REPETICIÓN 4 están ausentes. En otro modo de realización, REPETICIÓN 1 se selecciona del grupo que consiste en K′$_2$-P, Aeg$_2$-P, Aeg$_2$-G y Aeg-OAhx-P; g es 3-10 (incluyendo 3, 4, 5, 6, 7, 8, 9 y 10); REPETICIÓN 2 es diferente de REPETICIÓN 1 y se selecciona del grupo que consiste en pnaT$_2$-P, pnaT$_2$-G, pnaT$_2$-K, (Kpeg4)$_4$-P, (G-Kpeg)$_2$-P, G-(Kpeg4)$_2$-P, Kpeg-G, (G-(Kpeg3))$_2$-P, (Aegpeg4)$_2$-P, pnaC$_2$-P, Aeg$_2$-P y Aeg-OAhx-P; h es de 1-5 (incluyendo h=1, h=2, h=3, h=4 y h=5); y REPETICIÓN 3 es diferente de REPETICIÓN 2 y se selecciona del grupo que consiste en K′$_2$-P, Aeg$_2$-P, Aeg$_2$-G y pnaT$_2$-P; i es de 1-5 (incluyendo i=1, i=2, i=3, i=4 e i=5); REPETICIÓN 4, si está presente, es diferente de REPETICIÓN 3 y es Aeg$_2$-P; y si REPETICIÓN 4 está presente, j=1 o j=2. En otro modo de realización, REPETICIÓN 1 se selecciona del grupo que consiste en K′$_2$-P, Aeg$_2$-P, Aeg$_2$-G y Aeg-OAhx-P; g es 3-10 (incluyendo 3, 4, 5, 6, 7, 8, 9 y 10); REPETICIÓN 2 es PEG; h es 12; REPETICIÓN 3 se selecciona del grupo que consiste en K′$_2$-P, Aeg$_2$-P, Aeg$_2$-G y pnaT$_2$-P; i es de 1-5 (incluyendo i=1, i=2, i=3, i=4 e i=5); y REPETICIÓN 4 está ausente.

En un modo de realización, U, V, e, REPETICIÓN 1, REPETICIÓN 2, REPETICIÓN 3 y REPETICIÓN 4 se seleccionan de un conjunto de variables de acuerdo con la tabla 5c.

U	REPETICIÓN 1	REPETICIÓN 2	REPETICIÓN 3	REPETICIÓN 4	V	e
hex-4NPA	K′$_2$-P	pnaT$_2$-P	K′$_2$-P	-	K o K′	1–10
4NPA	PEG	pnaT$_2$-P	-	-	K o K′	1–10
4NPA	PEG	pnaT$_2$-G	-	-	K o K′	1–10
4NPA	PEG	pnaT$_2$-K	-	-	K o K′	1–10
4NPA	PEG	K′	-	-	K o K′	1–10
4NPA	Aeg$_2$-P	pnaT$_2$-P	Aeg$_2$-P	-	K o K′	1–10
4NPA	K′$_2$-P	(Kpeg4)$_4$-P	K′$_2$-P	-	K o K′	1–10
4NPA	K′$_2$-P	(G-Kpeg)$_2$-P	K′$_2$-P	-	K o K′	1–10
4NPA	K′$_2$-P	pnaT$_2$-P	K′$_2$-P	-	K o K′	1–10
Hex-4NPA	Aeg$_2$-G	pnaT$_2$-P	Aeg$_2$-G	-	K o K′	1–10
Hex-4NPA	Aeg$_2$-P	G-(Kpeg4)$_2$-P	Aeg$_2$-P	-	K o K′	1–10

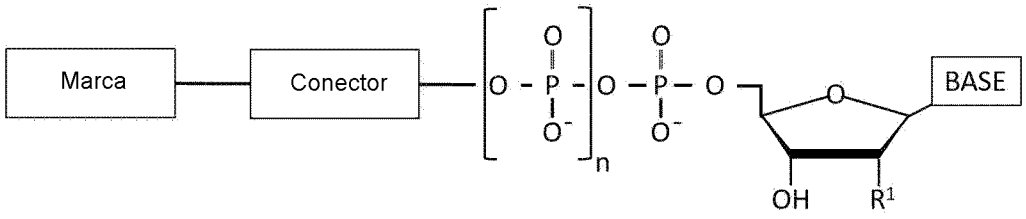
Hex-4NPA	Aeg_2-P	Kpeg-G	Aeg_2-P	-	K o K'	1–10
Hex-4NPA	Aeg_2-P	$(G-(Kpeg3))_2$-P	Aeg_2-P	-	K o K'	1–10
Hex-4NPA	Aeg_2-P	$(Aegpeg4)_2$-P	Aeg_2-P	-	K o K'	1–10
Hex-4NPA	Aeg_2-P	$pnaC_2$-P	Aeg_2-P	-	K o K'	1–10
Hex-4NPA	Aeg_2-P	PEG	Aeg_2-P	-	K o K'	1–10
Hex-4NPA	Aeg-OAhx-P	$pnaT_2$-P	Aeg_2-P	-	K o K'	1–10
Hex-4NPA	Aeg-OAhx-P	Aeg_2-P	$pnaT_2$-P	Aeg_2-P	K o K'	1–10
Hex-4NPA	Aeg_2-P	Aeg-OAhx-P	$pnaT_2$-P	Aeg_2-P	K o K'	1–10
Hex-4NPA	Aeg_2-P	Aeg-OAhx-P	-	-	K o K'	1–10

Tabla 5c

En un modo de realización, se proporciona la marca como se describe en la tabla 5a, 5b o 5c, en la que el ácido carboxílico C terminal de V se reemplaza por una amida.

Cualquiera de las marcas de polímero como se describe en el presente documento se puede enlazar covalentemente a un nucleósido oligofosfato. En algunos modos de realización, la marca de polímero se fija covalentemente a la base del nucleósido. En otros modos de realización, la marca de polímero se enlaza covalentemente al resto glucídico del nucleósido. Aún en otros modos de realización, la marca de polímero se enlaza covalentemente a un grupo fosfato. En un modo de realización, la marca se fija al nucleósido oligofosfato de modo que el resto "U" (si está presente) está próximo al nucleósido oligofosfato y el resto "V" está distal al nucleósido oligofosfato. En otro modo de realización, la marca se fija al nucleósido oligofosfato por medio del aminoácido N terminal de la marca, directamente o bien por medio de un resto conector.

En un modo de realización, se proporciona un nucleósido oligofosfato marcado, que tiene una estructura de acuerdo con la fórmula 3:



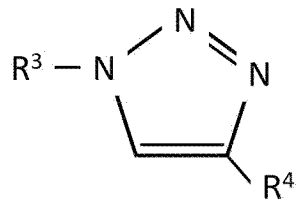
Fórmula 3

en la que la base se selecciona de adenina, citosina, guanina, timina y uracilo; R^1 es H, OH, $O-CH_3$, o F; n es de 2 a 12; marca es cualquiera de las marcas de polímero que tienen la carga neta positiva como se describe anteriormente; y conector es un resto químico que resulta de enlazar covalentemente un oligofosfato nucleotídico no marcado a la marca. Las estructuras de conector ejemplares incluyen éster, éter, tioéter, amina, amida, imida, carbonato, carbamato, escuarato, tiazol, tiazolidina, hidrazona, oxima, triazol, dihidropiridacina, fosfodiéster, polietilenglicol (PEG), aducto de Pictet-Spengler y cualquier combinación de los mismos. En un modo de realización, el conector tiene una estructura de fórmula 4:

$R^4 – R^2 – R^3$ **Fórmula 4**

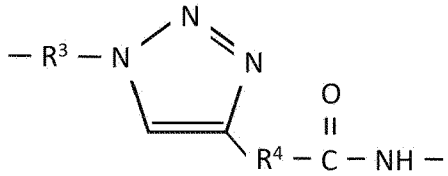
en la que R^2 se selecciona del grupo que consiste en enlace éster, enlace éter, enlace tioéter, enlace amina, enlace amida, enlace imida, enlace carbonato, enlace carbamato, enlace escuarato, enlace tiazol, enlace tiazolidina, enlace hidrazona, enlace oxima, enlace triazol, enlace dihidropiridacina, enlace fosfodiéster, enlace polietilenglicol (PEG), aducto de Pictet-Spengler y cualquier combinación de los mismos, R^3 comprende una cadena de carbonos saturada o insaturada, ramificada o no ramificada, sustituida o no sustituida de al menos 2 carbonos de longitud enlazada covalentemente en un extremo a uno de los restos fosfato del oligofosfato nucleotídico y en el otro extremo a R^2, y R^4 comprende una cadena de carbonos saturada o insaturada, ramificada o no ramificada, sustituida o no sustituida de al menos 2 carbonos de longitud enlazada covalentemente en un extremo a R^2 y en el otro extremo a la marca. En un modo de realización, las cadenas de carbono R^3 y R^4 tienen cada una menos de 20 carbonos de longitud. En un modo de realización, R^3 tiene de 5 a 20 carbonos de longitud y R^4 tiene de 2 a 12 carbonos de longitud. En un modo de realización, R^3 tiene de 5 a 20 carbonos de longitud y R^4 tiene de 2 a 12 carbonos de longitud, con la condición de que R^3 sea más larga que R^4. En un modo de realización, el enlace entre el CONECTOR y el oligofosfato implica el grupo fosfato terminal. En un modo de realización, el enlace entre la MARCA y el CONECTOR implica un grupo amino del residuo N terminal de la MARCA. En un modo de realización, el grupo carboxi del residuo C terminal de la MARCA se convierte en una amida, y el enlace entre la MARCA y el CONECTOR implica el grupo amino de la amida C terminal de la MARCA. En un modo de realización, el enlace

entre la MARCA y el CONECTOR implica un grupo carboxi del residuo C terminal de la MARCA. Los esquemas para generar estructuras de conector de fórmula 4 incluyen los expuestos en el documento US 2018-0057870. En un modo de realización, el conector tiene una estructura de fórmula 4, en la que R^2 es un triazol de fórmula 4a:



Fórmula 4a

en la que la suma de los carbonos en la cadena de carbonos R_3 y la cadena de carbonos R_4 es ≤ 96. En un modo de realización, la marca es una cadena peptídica y el conector tiene una estructura de fórmula 4, en la que R^2 es un triazol de fórmula 4b:



Fórmula 4b

y en la que el –NH– del enlace peptídico conectado a R^4 se aporta por el aminoácido o análogo de aminoácido N terminal de la cadena peptídica.

V. Sistemas y procedimientos de secuenciación SBS

En un modo de realización, se proporciona un sistema para realizar secuenciación por síntesis (SBS) de ácidos nucleicos, comprendiendo el sistema: (a) un nanoporo que tiene un canal que tiene suficientes restos cargados negativamente para repeler sustancialmente el ácido nucleico molde y/o cebador, (b) una ácido nucleico polimerasa asociada con el nanoporo, (c) un conjunto de oligofosfatos nucleotídicos dispuestos en una solución electrolítica, comprendiendo dichos oligofosfatos nucleotídicos una marca cargada positivamente que puede pasar a través del nanoporo de (a), y (d) al menos una posición de electrodo para registrar una característica de una corriente que fluye a través del canal.

La FIG. 3 ilustra un modo de realización ejemplar de un complejo de secuenciación de nanoporos 300 para realizar una secuenciación de nucleótidos por SBS basada en marca. Una barrera eléctricamente resistiva 301 separa una solución electrolítica a granel 302 de una segunda solución electrolítica 303. Un nanoporo 304 se dispone en la barrera eléctricamente resistiva 301, y el canal del nanoporo 305 proporciona una ruta a través de la que pueden fluir los iones entre el electrólito a granel 302 y el segundo electrólito 303. Un electrodo de trabajo 306 se dispone en el lado de la barrera eléctricamente resistiva 301 que contiene el segundo electrólito 303 (denominado el "lado *trans*" de la barrera eléctricamente resistiva) y se sitúa cerca del nanoporo 304. Un contraelectrodo 307 se sitúa en el lado de la barrera eléctricamente resistiva 301 que contiene el electrólito a granel 302 (denominado el "lado *cis*" de la barrera eléctricamente resistiva). Una fuente de señales 308 se adapta para aplicar una señal de voltaje entre el electrodo de trabajo 306 y el contraelectrodo 307. Una polimerasa 309 se asocia con el nanoporo 304, y un ácido nucleico molde cebado 310 se asocia con la polimerasa. El electrólito a granel 302 incluye cuatro diferentes nucleósidos oligofosfato marcados con polímero 311 (marca ilustrada como 311a). La polimerasa 309 cataliza la incorporación de los nucleótidos marcados con polímero 311 en un amplicón del molde. Cuando un nucleósido oligofosfato marcado con polímero 311 forma un complejo correctamente con la polimerasa 309, la marca 311a se puede atraer (por ejemplo, cargar) hacia el nanoporo por una fuerza eléctrica, tal como una fuerza generada en presencia de un campo eléctrico generado por un voltaje aplicado a través de la barrera eléctricamente resistiva 301 y/o el nanoporo 304. Mientras la marca 311a ocupa el canal del nanoporo 304, afecta el flujo iónico a través del nanoporo 304, generando de este modo una señal de bloqueo iónico 312. Cada nucleótido 311 tiene una marca de polímero única 311a que genera una señal de bloqueo iónico única debido a la estructura química distintiva y/o al tamaño de la marca 311a. Al identificar la señal de bloqueo iónico única 312, se puede identificar la identidad de las marcas únicas 311a (y, por lo tanto, el nucleótido 310 con el que se asocia). Este proceso se repite iterativamente con cada nucleótido 310 incorporado al amplicón.

VA. Nanoporos para su uso en el sistema

Los nanoporos de los presentes sistemas comprenden una concentración de carga negativa dispuesta en o cerca del canal del nanoporo. Sin quedar vinculado por la teoría, se cree que las cargas negativas repelen los ácidos nucleicos cargados negativamente y, por tanto, disuaden la inserción de moldes, amplicones, cebadores, etc., en el nanoporo.

En un modo de realización, el nanoporo es un nanoporo biológico. Las proteínas comúnmente usadas para generar nanoporos biológicos incluyen α-hemolisina (α-HL) de *Staphylococcus aureus* (secuencia sin procesar de longitud completa canónica divulgada en el n.º de acceso de Uniprot P09616-1 (SEQ ID NO: 20)), nanoporo de porina G de membrana externa (OmpG) de *Escherichia coli* (secuencia sin procesar de longitud completa canónica divulgada en el n.º de acceso de Uniprot P76045-1 (SEQ ID NO: 21)) y porina A de *Mycobacterium smegmatis* (MspA) (secuencia sin procesar de longitud completa canónica divulgada en el n.º de acceso de Uniprot A0QR29-1 (SEQ ID NO: 22)). Otros nanoporos biológicos ejemplares incluyen nanoporo de leucocidina, nanoporo de porina F de membrana externa (OmpF), nanoporo de citolisina A (ClyA), nanoporo de fosfolipasa A de membrana externa, nanoporo de lipoproteína autotransportadora (NalP) de *Neisseria*, nanoporo de WZA, nanoporo de canal selectivo catiónico NfpA/NfpB de *Nocardia farcinica*, nanoporo de lisenina, aerolisina y nanoporo de la subunidad G de crecimiento dependiente de sigma S de curlina (CsgG). Aún en otros modos de realización, el nanoporo es un nanoporo híbrido que comprende un nanoporo biológico. En otros modos de realización, el nanoporo es un nanoporo de estado sólido que comprende una carga negativa localizada en la abertura del canal.

VA1. Nanoporos de α-HL

Los nanoporos de α-HL son estructuras heptaméricas formadas a partir de 7 subunidades monoméricas del polipéptido de α-HL de *Staphylococcus aureus*. Se pueden encontrar diversos enfoques para genomanipular nanoporos de α-HL para su uso en secuenciación basada en nanoporos, por ejemplo, en Ayub, Wang II, documentos WO 2014/100481, WO 2016/069806, WO 2017/050718, WO 2017/184866 y WO 2018/002125. Como se ilustra en la FIG. 2A, los nanoporos de α-HL tienen una región de caperuza 201 y una región de barril beta 202, con un canal 203 que se extiende axialmente a través de las regiones de caperuza y tronco. La FIG. 2B es una sección transversal del nanoporo de α-HL que muestra el canal 203, que se puede dividir en un vestíbulo 204, un sitio de constricción 205, un cuerpo de barril beta 206 y una salida de barril beta 207. Las referencias en el presente documento a la "región de barril beta" incluyen cada uno del sitio de constricción 205, el cuerpo de barril beta 206 y la salida de barril beta 207. Las referencias en el presente documento a "nanoporo de α-HL" se referirán a poros heptaméricos de 7 subunidades monoméricas de α-HL.

Los nanoporos de α-HL útiles en los presentes sistemas tienen un canal que comprende un aminoácido cargado negativamente no natural en una pluralidad de posiciones orientadas al disolvente dentro del canal. En un modo de realización, una o más de las subunidades monoméricas de α-HL (incluyendo 1, 2, 3, 4, 5, 6 o 7 de las subunidades) del nanoporo tienen al menos un aminoácido cargado negativamente no natural (tal como ácido aspártico o ácido glutámico) en una posición correspondiente a la región de vestíbulo y/o la región de barril beta del nanoporo de α-HL. Una secuencia de aminoácidos correspondiente a una subunidad monomérica de α-HL natural se puede encontrar en la SEQ ID NO: 1. A menos que se indique de otro modo, toda la numeración de aminoácidos relacionada con las subunidades monoméricas de α-HL son con referencia a SEQ ID NO: 1. Cuando se hace referencia a una subunidad monomérica de α-HL "que comprende una sustitución en la posición n.º" o "que comprende una sustitución X#Y" se entenderá que quiere decir que la secuencia de aminoácidos de subunidad monomérica, cuando se alinea con la SEQ ID NO: 1, tiene una sustitución en la posición correspondiente a la posición mencionada de SEQ ID NO: 1. Como se usa en el presente documento, un "aminoácido no natural" es un aminoácido en una posición de la secuencia de aminoácidos de subunidad monomérica que representa una sustitución o inserción cuando se alinea con la SEQ ID NO: 1. En un modo de realización, los polipéptidos comprenden al menos una subunidad monomérica de α-HL que tiene al menos un 75 % de identidad, al menos un 80 % de identidad, al menos un 85 % de identidad, al menos un 90 % de identidad, al menos un 91 % de identidad, al menos un 92 % de identidad, al menos un 93 % de identidad, al menos un 94 % de identidad o al menos un 95 % de identidad con SEQ ID NO: 1.

Se proporcionan suficientes aminoácidos cargados negativamente no naturales para que (a) los ácidos nucleicos molde y/o cebadores tengan menos probabilidades de insertarse en el poro en relación con los nanoporos de una α-HL que tienen el residuo aminoacídico natural, y (b) las marcas cargadas positivamente se puedan translocar a través del canal del nanoporo. Los nanoporos de α-HL ejemplares se divulgan anteriormente en la sección II.

VA2. Nanoporos de ompG

En un modo de realización, el nanoporo es un nanoporo de ompG o un nanoporo híbrido que comprende un nanoporo de ompG como componente biológico. Como se ilustra en las FIGS. 4A y 4B, los nanoporos de ompG **400** se forman a partir de un polipéptido monocatenario dispuesto con los siguientes rasgos característicos estructurales: un canal que comprende 14 hebras beta β1–β14 **401**; 7 regiones de bucle: L1–L7 **402**; y 6 vueltas cortas T1–T6 **403**. Un sitio de constricción **404** se localiza en el interior del canal. El canal tiene un lado *cis* **405**, que

corresponde a la porción extracelular en la proteína natural, y un lado *trans* **406**, que corresponde a la porción intracelular en la proteína natural. Las referencias que describen nanoporos de ompG genomanipulados para su uso en aparatos de secuenciación basados en nanoporo incluyen Gari, Grosse, Chen II y el documento WO 2017/050722.

Un polipéptido ejemplar usado para formar nanoporos de ompG se divulga en SEQ ID NO: 23, que es una forma madura de SEQ ID NO: 21 que carece del péptido señal N terminal de aminoácido 21. Los aminoácidos correspondientes a las hebras beta, bucles y vueltas de SEQ ID NO: 23 se ilustran en la fig. 5A. Como se ilustra en la fig. 5A, las hebras beta corresponden a los siguientes residuos aminoacídicos de SEQ ID NO: 23: β1 = 7–15, β2 = 33–40, β3 = 43–50, β4 = 69–78, β5 = 84–94, β6 = 109–121, β7 = 126–138, β8 = 150–160, β9 = 166–175, β10 = 190–201, β11 = 204–210, β12 = 237–243, β13 = 248–254 y β14 = 273–280.

En un modo de realización, el nanoporo es un nanoporo de ompG que comprende un canal formado por un conjunto de hebras beta β1–β14 que tienen al menos un 90 % de identidad con respecto a las hebras beta β1–β14 de SEQ ID NO: 23, en el que el conjunto de hebras beta β1–β14 tiene al menos un aminoácido cargado negativamente no natural en relación con las hebras beta β1–β14 de SEQ ID NO: 23. En un modo de realización, el conjunto de hebras beta β1–β14 tiene al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o al menos 14 aminoácidos cargados negativamente no naturales en relación con las hebras beta β1–β14 de SEQ ID NO: 23. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o cada una de las hebras beta β1–β14 comprende al menos 1 de los aminoácidos cargados negativamente no naturales.

En un modo de realización, el nanoporo de ompG comprende una secuencia de aminoácidos que tiene al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 % o al menos un 95 % de identidad de secuencia con respecto a SEQ ID NO: 23, teniendo el nanoporo de ompG un canal que comprende un conjunto de hebras beta β1–β14 que tienen al menos un aminoácido cargado negativamente no natural en relación con las hebras beta β1–β14 de SEQ ID NO: 23. En un modo de realización, el conjunto de hebras beta β1–β14 tiene al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o al menos 14 aminoácidos cargados negativamente no naturales en relación con las hebras beta β1–β14 de SEQ ID NO: 23. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o cada una de las hebras beta β1–β14 comprende al menos 1 de los aminoácidos cargados negativamente no naturales.

Los nanoporos de ompG divulgados en el presente documento pueden contener modificaciones adicionales que mejoran la capacidad del nanoporo que se va a usar en biodetección. Por ejemplo, es conocido que la región de bucle L6 provoca la apertura de poro; el bloqueo espontáneo de la corriente a través del poro durante un potencial aplicado. Se han usado diferentes estrategias para mitigar la apertura de poro, incluyendo el truncamiento del bucle (véase, por ejemplo, Gari, Grosse y el documento WO 2017/050722), reduciendo la movilidad del bucle, por ejemplo, al introducir un enlace disulfuro al añadir residuos de cisteína a los extremos extracelulares de las hebras β12 y β13 (tales como G230 y D262 de SEQ ID NO: 23) (Chen II) y/o introduciendo un ancla lipídica en L6, por ejemplo, por alquilación de una cisteína genomanipulada (tal como, por ejemplo, por introducción de una sustitución I226C). La movilidad del bucle L6 se puede minimizar además optimizando los enlaces de hidrógeno entre las hebras β11 y β12 por deleción del residuo D215 (Chen II y documento WO 2017/050722). En el presente documento se divulga un polipéptido de ompG ejemplar que tiene una región de bucle L6 truncada en SEQ ID NO: 24, correspondiente a SEQ ID NO: 23 con modificaciones ΔR216–E227 y E229A. Los cambios en relación con SEQ ID NO: 23 se ilustran en la fig. 5B, con los residuos delecionados ilustrados por guiones (-) y las sustituciones

ilustradas por cuadros sombreados. En el presente documento se divulga un polipéptido de ompG ejemplar que tiene una región de bucle estabilizada y estabilización de enlaces de hidrógeno entre las hebras β11 y β12 en SEQ ID NO: 25, correspondiente a SEQ ID NO: 23 con modificaciones G230C, D262C y ΔD215. Los cambios en relación con SEQ ID NO: 23 se ilustran en la fig. 5C, con los residuos delecionados ilustrados por guiones (-) y las sustituciones ilustradas por cuadros sombreados. En el presente documento se divulga un polipéptido de ompG ejemplar que tiene una cisteína genomanipulada en el bucle L6 al que se puede fijar un ancla lipídica en SEQ ID NO: 26, correspondiente a SEQ ID NO: 23 con una sustitución I226C. Los cambios en relación con SEQ ID NO: 23 se ilustran en la fig. 5D, con los residuos delecionados ilustrados por guiones (-) y las sustituciones ilustradas por cuadros sombreados. En el presente documento se divulga un polipéptido de ompG ejemplar que tiene una región de bucle L6 truncada y estabilización de enlaces de hidrógeno entre las hebras β11 y β12 en SEQ ID NO: 27, correspondiente a SEQ ID NO: 23 con modificaciones ΔD215–E227 + E229A. Los cambios en relación con SEQ ID NO: 23 se ilustran en la fig. 5E, con los residuos delecionados ilustrados por guiones (-) y las sustituciones ilustradas por cuadros sombreados. En el presente documento se divulga un polipéptido de ompG ejemplar que tiene una cisteína genomanipulada en el bucle L6 al que se puede fijar un ancla lipídica y estabilización de enlaces de hidrógeno entre las hebras β11 y β12 en SEQ ID NO: 28, correspondiente a SEQ ID NO: 23 con una modificación ΔD215 + I226C. Los cambios en relación con SEQ ID NO: 23 se ilustran en la fig. 5F, con los residuos delecionados ilustrados por guiones (-) y las sustituciones ilustradas por cuadros sombreados.

En un modo de realización, el nanoporo es un nanoporo de ompG que comprende: (a) un canal formado por un conjunto de hebras beta β1–β14 que tienen al menos un 90 % de identidad con respecto a las hebras beta β1–β14 de SEQ ID NO: 23, en el que el conjunto de hebras beta β1–β14 tiene al menos un aminoácido cargado negativamente no natural en relación con las hebras beta β1–β14 de SEQ ID NO: 23, (b) una modificación seleccionada del grupo que consiste en una región de bucle L6 truncada o delecionada o una región de bucle L6 estabilizada, y opcionalmente (c) una modificación que estabiliza enlaces de hidrógeno entre las hebras β11 y β12. En un modo de realización, el conjunto de hebras beta β1–β14 tiene al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o al menos 14 aminoácidos cargados negativamente no naturales en relación con las hebras beta β1–β14 de SEQ ID NO: 23. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o cada una de las hebras beta β1–β14 comprende al menos 1 de los aminoácidos cargados negativamente no naturales.

En un modo de realización, el nanoporo es un nanoporo de ompG que comprende: (a) un canal formado por un conjunto de hebras beta β1–β14 que tienen al menos un 90 % de identidad con respecto a las hebras beta β1–β14 de SEQ ID NO: 23, en el que el conjunto de hebras beta β1–β14 tiene al menos un aminoácido cargado negativamente no natural en relación con las hebras beta β1–β14 de SEQ ID NO: 23, y (b) una región de bucle L6 que tiene modificaciones ΔR216–E227 y E229A en relación con SEQ ID NO: 23. En un modo de realización, el conjunto de hebras beta β1–β14 tiene al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o al menos 14 aminoácidos cargados negativamente no naturales en relación con las hebras beta β1–β14 de SEQ ID NO: 23. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o cada una de las hebras beta β1–β14 comprende al menos 1 de los aminoácidos cargados negativamente no naturales. En un modo de realización, el nanoporo es un nanoporo de ompG que tiene al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 % o al menos un 95 % de identidad de secuencia con respecto a SEQ ID NO: 24.

En un modo de realización, el nanoporo es un nanoporo de ompG que comprende: (a) un canal formado por un conjunto de hebras beta β1–β14 que tienen al menos un 90 % de identidad con respecto a las hebras beta β1–β14 de SEQ ID NO: 23, en el que el conjunto de hebras beta β1–β14 tiene al menos un aminoácido cargado

negativamente no natural en relación con las hebras beta β1–β14 de SEQ ID NO: 23, y (b) modificaciones G230C, D262C y ΔD215 en relación con SEQ ID NO: 23. En un modo de realización, el conjunto de hebras beta β1–β14 tiene al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o al menos 14 aminoácidos cargados negativamente no naturales en relación con las hebras beta β1–β14 de SEQ ID NO: 23. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o cada una de las hebras beta β1–β14 comprende al menos 1 de los aminoácidos cargados negativamente no naturales. En un modo de realización, el nanoporo es un nanoporo de ompG que tiene al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 % o al menos un 95 % de identidad de secuencia con respecto a SEQ ID NO: 25.

En un modo de realización, el nanoporo es un nanoporo de ompG que comprende: (a) un canal formado por un conjunto de hebras beta β1–β14 que tienen al menos un 90 % de identidad con respecto a las hebras beta β1–β14 de SEQ ID NO: 23, en el que el conjunto de hebras beta β1–β14 tiene al menos un aminoácido cargado negativamente no natural en relación con las hebras beta β1–β14 de SEQ ID NO: 23, y (b) una modificación I226C en relación con SEQ ID NO: 23, en la que dicha cisteína está opcionalmente alquilada. En un modo de realización, el conjunto de hebras beta β1–β14 tiene al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o al menos 14 aminoácidos cargados negativamente no naturales en relación con las hebras beta β1–β14 de SEQ ID NO: 23. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o cada una de las hebras beta β1–β14 comprende al menos 1 de los aminoácidos cargados negativamente no naturales. En un modo de realización, el nanoporo es un nanoporo de ompG que tiene al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 % o al menos un 95 % de identidad de secuencia con respecto a SEQ ID NO: 26.

En un modo de realización, el nanoporo es un nanoporo de ompG que comprende: (a) un canal formado por un conjunto de hebras beta β1–β14 que tienen al menos un 90 % de identidad con respecto a las hebras beta β1–β14 de SEQ ID NO: 23, en el que el conjunto de hebras beta β1–β14 tiene al menos un aminoácido cargado negativamente no natural en relación con las hebras beta β1–β14 de SEQ ID NO: 23, y (b) modificaciones ΔD215–E227 + E229A en relación con SEQ ID NO: 23. En un modo de realización, el conjunto de hebras beta β1–β14 tiene al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o al menos 14 aminoácidos cargados negativamente no naturales en relación con las hebras beta β1–β14 de SEQ ID NO: 23. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o cada una de las hebras beta β1–β14 comprende al menos 1 de los aminoácidos cargados negativamente no naturales. En un modo de realización, el nanoporo es un nanoporo de ompG que tiene al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 % o al menos un 95 % de identidad de secuencia con respecto a SEQ ID NO: 27.

En un modo de realización, el nanoporo es un nanoporo de ompG que comprende: (a) un canal formado por un conjunto de hebras beta β1–β14 que tienen al menos un 90 % de identidad con respecto a las hebras beta β1–β14 de SEQ ID NO: 23, en el que el conjunto de hebras beta β1–β14 tiene al menos un aminoácido cargado negativamente no natural en relación con las hebras beta β1–β14 de SEQ ID NO: 23, (b) una modificación I226C en relación con SEQ ID NO: 23, en la que dicha cisteína está opcionalmente alquilada, y (c) una modificación ΔD215 en relación con SEQ ID NO: 23. En un modo de realización, el conjunto de hebras beta β–β14 tiene al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o al menos 14 aminoácidos cargados negativamente no naturales en relación con las hebras beta β1–β14 de SEQ ID NO: 23. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la entrada *cis* al canal y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor del sitio de constricción y al menos una porción de los aminoácidos cargados negativamente no naturales se localiza alrededor de la salida *trans* al canal. En un modo de realización, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13 o cada una de las hebras beta β1–β14 comprende al menos 1 de los aminoácidos cargados negativamente no naturales. En un modo de realización, el nanoporo es un nanoporo de ompG que tiene al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 % o al menos un 95 % de identidad de secuencia con respecto a SEQ ID NO: 28.

VA3. Nanoporos de MspA

En un modo de realización, el nanoporo es un nanoporo de MspA. Los nanoporos de MspA comprenden típicamente 8 subunidades monoméricas de MspA. Las referencias que describen nanoporos de MspA genomanipulados para su uso en aparatos de secuenciación basada en nanoporos incluyen Butler, Manrao, Pavlenok, los documentos WO 2013/098562, US 2014-0309402, US 2013-0146457 y Wang II. En el presente documento se divulga una secuencia de aminoácidos ejemplar usada para genomanipular subunidades monoméricas de MspA para su uso en secuenciación basada en nanoporos en SEQ ID NO: 29. Los nanoporos de MspA para su uso en la secuenciación de ácidos nucleicos basada en nanoporos tienen típicamente un sitio de constricción neutro (por ejemplo, realizando sustituciones D90N, D91N y D93N en relación con SEQ ID NO: 29) y aminoácidos cargados positivamente cerca de la entrada de canal *cis* (por ejemplo, sustituyendo uno o más de D118, D134, E139 de SEQ ID NO: 29 con un aminoácido cargado positivamente, tales como las sustituciones D118R, D134R y/o E139K). Estos cambios se realizan típicamente para potenciar la interacción del ADN monocatenario con el poro. Por el contrario, los nanoporos de MspA del presente sistema comprenden un aminoácido cargado negativamente en cada una de estas posiciones.

En un modo de realización, se proporciona un nanoporo de MspA octamérico que comprende 8 subunidades monoméricas que tienen al menos un 70, al menos un 80, al menos un 90 o al menos un 95 % de identidad con SEQ ID NO: 29, en el que al menos 4, al menos 5, al menos 6, al menos 7 o las 8 subunidades monoméricas tienen un ácido aspártico o ácido glutámico en posiciones correspondientes a D90, D91, D93, D118, D134 y E139 de SEQ ID NO: 29.

VB. Conjuntos de nucleósidos oligofosfato marcados

El sistema comprende además un conjunto de al menos 4 nucleósidos oligofosfato marcados, comprendiendo el conjunto un nucleósido oligofosfato de adenosina que tiene una primera marca de polímero, un nucleósido oligofosfato de citosina que tiene una segunda marca de polímero, un nucleósido oligofosfato de guanina que tiene una tercera marca de polímero, y un nucleósido oligofosfato de timina que tiene una cuarta marca de polímero o un nucleósido oligofosfato de uracilo que tiene una cuarta marca de polímero, y cada una de las primera a cuarta marcas de polímero tiene las siguientes características: (a) puede ocupar el canal del nanoporo de una manera que genera una señal de bloqueo iónico que es detectable en condiciones de funcionamiento normales del complejo de secuenciación de nanoporos; (b) se puede liberar del nucleósido oligofosfato después de que se genera la señal de bloqueo iónico; y (c) puede fluir a través del nanoporo. En algunos modos de realización, cada señal de bloqueo iónico generada por cada marca de polímero se puede distinguir de la señal de bloqueo iónico generada por cada una de las otras marcas de polímero en condiciones de funcionamiento normales del complejo de secuenciación de nanoporos. En otros modos de realización (por ejemplo, cuando se pretende que los nucleósidos oligofosfato fluyan de a uno en uno en el sistema), al menos algunas de las marcas de polímero pueden generar señales de bloqueo iónico que son eficazmente indistinguibles entre sí en condiciones de funcionamiento normales del complejo de secuenciación de nanoporos.

En un modo de realización, las marcas de polímero tienen una carga neta neutra a neta positiva. En algunos modos de realización, al menos uno de los al menos 4 nucleósidos oligofosfato tiene una marca de polímero con carga

neta positiva. En algunos modos de realización, al menos 2 de los al menos 4 nucleósidos oligofosfato marcados tienen una marca de polímero con carga neta positiva. En algunos modos de realización, al menos 3 de los al menos 4 nucleósidos oligofosfato marcados tienen una marca de polímero con carga neta positiva. En algunos modos de realización, al menos 4 de los nucleósidos oligofosfato marcados tienen una marca de polímero con carga neta positiva. En algunos modos de realización, cada nucleósido oligofosfato marcado tiene una marca de polímero con carga neta positiva. En un modo de realización, la marca de polímero con carga neta positiva es una marca de péptido de acuerdo con la fórmula 2, la fórmula 2a o la fórmula 2b. En un modo de realización, el conjunto de nucleósidos oligofosfato marcados comprende un desoxiadenosina oligofosfato (dAP), un desoxicitosina oligofosfato (dCP), un desoxiguanina oligofosfato (dGP) y un desoxitimina oligofosfato (dTP), en el que cada uno de dAP, dCP, dGP y dTP se marca con una marca de péptido de acuerdo con la fórmula 2, que incluye, pero sin limitarse a, las marcas de péptido expuestas en la tabla 5. En un modo de realización, cada uno de dAP, dCP, dGP y dTP tiene una estructura de acuerdo con la fórmula 3, en la que MARCA es la marca peptídica de acuerdo con la fórmula 2, 2a o 2b, incluyendo, pero sin limitarse a las marcas peptídicas expuestas en la tabla 5.

VC. Ácido nucleico polimerasas

Los presentes sistemas y procedimientos pueden incorporar cualquier ácido nucleico polimerasa que sea útil en la secuenciación SBS y pueda realizar una polimerización específica de secuencia de nucleósido polifosfatos que tengan 4, 5, 6, 7, 8 o más fosfatos.

En algunos modos de realización, la ADN polimerasa dependiente de ADN es una variante de un polipéptido natural que tiene actividad ADN polimerasa dependiente de ADN, tal como Pol6, ADN polimerasa phi29, ADN pol T7, ADN pol T4, ADN pol 1 de *E. coli*, fragmento Klenow, así como subunidades y cofactores asociados. En un modo de realización, la ADN polimerasa dependiente de ADN es una ADN polimerasa derivada de Pol6. Una secuencia natural marcada con His de Pol6 está disponible en SEQ ID NO: 30. Los derivados de Pol6 ejemplares útiles en la secuenciación basada en nanoporos se divulgan, por ejemplo, en los documentos US 2016/0222363, US 2016/0333327, US 2017/0267983, US 2018/0094249 y US 2018/0245147.

VD. Chips sensores de nanoporos

En algunos modos de realización, se proporciona un chip sensor de nanoporos, que comprende una matriz de celdas de nanoporos que comprenden un complejo de secuenciación de nanoporos como se describe anteriormente.

La FIG. 6 es una vista superior de un modo de realización de un chip sensor de nanoporos 600 que tiene una matriz 640 de celdas de nanoporos 650. Cada celda de nanoporo 650 incluye un circuito de control integrado en un sustrato de silicio del chip sensor de nanoporos 600. En algunos modos de realización, se incluyen paredes laterales 636 en la matriz 740 para separar grupos de celdas de nanoporos 650 de modo que cada grupo pueda recibir una muestra diferente para su caracterización. Cada celda de nanoporo se puede usar para secuenciar un ácido nucleico. En algunos modos de realización, el chip sensor de nanoporos 600 incluye una placa de cubierta 630. En algunos modos de realización, el chip sensor de nanoporos 600 también incluye una pluralidad de clavijas 610 para interactuar con otros circuitos, tales como un procesador de ordenador.

En algunos modos de realización, el chip sensor de nanoporos 600 incluye múltiples chips en un mismo paquete, tales como, por ejemplo, un módulo multichip (MCM) o un sistema en paquete (SiP). Los chips pueden incluir, por ejemplo, una memoria, un procesador, una matriz de puertas programables en campo (FPGA), un circuito integrado específico de la aplicación (ASIC), convertidores de datos, una interfaz de E/S de alta velocidad, etc.

En algunos modos de realización, el chip sensor de nanoporos 600 se acopla a (por ejemplo, conecta a) una estación de trabajo de nanochip 620, que puede incluir diversos componentes para llevar a cabo (por ejemplo, llevar a cabo automáticamente) diversos modos de realización de los procesos divulgados en el presente documento. Estos procesos pueden incluir, por ejemplo, mecanismos de suministro de analito, tales como pipetas para suministrar suspensión de lípidos u otra suspensión de estructura de membrana, solución de analito y/u otros líquidos, suspensiones o sólidos. Los componentes de la estación de trabajo de nanochip pueden incluir además brazos robóticos, uno o más procesadores de ordenador y/o memoria. Se puede detectar una pluralidad de polinucleótidos en la matriz 640 de celdas de nanoporos 650. En algunos modos de realización, cada celda de nanoporo 650 se puede abordar individualmente.

A. Estructura de celda de secuenciación de nanoporos

La FIG. 7 ilustra un modo de realización de una celda de nanoporo de ejemplo 800 en un chip sensor de nanoporos, tal como la celda de nanoporo 750 en el chip sensor de nanoporos 700 de la FIG. 7, que se puede usar para caracterizar un polinucleótido. La celda de nanoporo 700 puede incluir un pocillo 705 formado por capas dieléctricas 701 y 704; una membrana, tal como una bicapa lipídica 714 formada sobre el pocillo 705; y una cámara de muestra 715 sobre la bicapa lipídica 714 y separada del pocillo 705 por la bicapa lipídica 714. El pocillo 705 puede contener un volumen de electrólito 706, y la cámara de muestra 715 puede contener electrólito a granel 708

que contiene un nanoporo, por ejemplo, un complejo molecular transmembranario de nanoporo de proteína soluble (PNTMC), y el analito de interés (por ejemplo, una molécula de ácido nucleico que se va a secuenciar).

La celda de nanoporo 700 puede incluir un electrodo de trabajo 702 en el fondo del pocillo 705 y un contraelectrodo 710 dispuesto en la cámara de muestra 715. Una fuente de señales 728 puede aplicar una señal de voltaje entre el electrodo de trabajo 702 y el contraelectrodo 710. Se puede insertar un único nanoporo (por ejemplo, un PNTMC) en la bicapa lipídica 714 por un proceso de electroporación provocado por la señal de voltaje, formando de este modo un nanoporo 716 en la bicapa lipídica 714. Las membranas individuales (por ejemplo, bicapas lipídicas 714 u otras estructuras de membrana) en la matriz no pueden estar conectadas entre sí ni química ni eléctricamente. De este modo, cada celda de nanoporo de la matriz puede ser un aparato de secuenciación independiente, que produce datos únicos para la molécula de polímero individual asociada con el nanoporo que actúa sobre el analito de interés y modula la corriente iónica a través de la bicapa lipídica de otro modo impermeable.

Como se muestra en la FIG. 2, se puede formar la celda de nanoporo 700 sobre un sustrato 730, tal como un sustrato de silicio. Se puede formar la capa dieléctrica 701 sobre el sustrato 730. El material dieléctrico usado para formar la capa dieléctrica 701 puede incluir, por ejemplo, vidrio, óxidos, nitruros y similares. Se puede formar un circuito eléctrico 722 para controlar la estimulación eléctrica y para procesar la señal detectada desde la celda de nanoporo 700 sobre el sustrato 730 y/o dentro de la capa dieléctrica 701. Por ejemplo, se puede formar una pluralidad de capas metálicas estampadas (por ejemplo, de metal 1 a metal 6) en la capa dieléctrica 701, y se puede fabricar una pluralidad de dispositivos activos (por ejemplo, transistores) sobre el sustrato 730. En algunos modos de realización, la fuente de señales 728 se incluye como parte del circuito eléctrico 722. El circuito eléctrico 722 puede incluir, por ejemplo, amplificadores, integradores, convertidores de analógico a digital, filtros de ruido, lógica de control de retroalimentación y/u otros componentes diversos. El circuito eléctrico 722 se puede acoplar además a un procesador 724 que se acopla a una memoria 726, donde el procesador 724 puede analizar los datos de secuenciación para determinar secuencias de las moléculas de polímero que se han secuenciado en la matriz.

El electrodo de trabajo 702 se puede formar sobre la capa dieléctrica 701 y puede formar al menos una parte del fondo del pocillo 705. En algunos modos de realización, el electrodo de trabajo 702 es un electrodo de metal. Para la conducción no faradaica, el electrodo de trabajo 702 se puede hacer de metales u otros materiales que son resistentes a la corrosión y la oxidación, tales como, por ejemplo, platino, oro, nitruro de titanio y grafito. Por ejemplo, el electrodo de trabajo 702 puede ser un electrodo de platino con platino electrodepositado. En otro ejemplo, el electrodo de trabajo 702 puede ser un electrodo de trabajo de nitruro de titanio (TiN). El electrodo de trabajo 702 puede ser poroso, incrementando de este modo su área de superficie y una capacidad resultante asociada con el electrodo de trabajo 702. Debido a que el electrodo de trabajo de una celda de nanoporo puede ser independiente del electrodo de trabajo de otra celda de nanoporo, en la presente divulgación el electrodo de trabajo se puede denominar electrodo de celda.

Se puede formar la capa dieléctrica 704 encima de la capa dieléctrica 701. La capa dieléctrica 704 forma las paredes que rodean el pocillo 705. El material dieléctrico usado para formar la capa dieléctrica 704 puede incluir, por ejemplo, vidrio, óxido, mononitruro de silicio (SiN), poliimida u otro material aislante hidrófobo adecuado. La superficie superior de la capa dieléctrica 704 se puede silanizar. La silanización puede formar una capa hidrófoba 720 encima de la superficie superior de la capa dieléctrica 704. En algunos modos de realización, la capa hidrófoba 720 tiene un espesor de aproximadamente 1,5 nanómetros (nm).

El pocillo 705 formado por las paredes de capa dieléctrica 704 incluye un volumen de electrólito 706 por encima del electrodo de trabajo 702. El volumen de electrólito 706 se puede tamponar y puede incluir uno o más de los siguientes: cloruro de litio (LiCl), cloruro de sodio (NaCl), cloruro de potasio (KCl), glutamato de litio, glutamato de sodio, glutamato de potasio, acetato de litio, acetato de sodio, acetato de potasio, cloruro de calcio ($CaCl_2$), cloruro de estroncio ($SrCl_2$), cloruro de manganeso ($MnCl_2$) y cloruro de magnesio ($MgCl_2$). En algunos modos de realización, el volumen de electrólito 706 tiene un espesor de aproximadamente tres micrómetros (µm).

Como también se muestra en la FIG. 7, se puede formar una membrana sobre la capa dieléctrica 704 y que abarca todo el pocillo 705. En algunos modos de realización, la membrana incluye una monocapa lipídica 718 formada sobre la capa hidrófoba 720. A medida que la membrana alcanza la abertura del pocillo 705, la monocapa lipídica 708 se puede convertir en la bicapa lipídica 714 que abarca toda la abertura del pocillo 705. La bicapa lipídica puede comprender o consistir en fosfolípido, por ejemplo, seleccionado de difitanoil-fosfatidilcolina (DPhPC), 1,2-difitanoil-sn-glicero-3-fosfocolina, 1,2-di-O-fitanil-sn-glicero-3-fosfocolina (DOPhPC), palmitoil-oleoil-fosfatidilcolina (POPC), éster dioleoil-fosfatidil-metílico (DOPME), dipalmitoilfosfatidilcolina (DPPC), fosfatidilcolina, fosfatidiletanolamina, fosfatidilserina, ácido fosfatídico, fosfatidilinositol, fosfatidilglicerol, esfingomielina, 1,2-di-O-fitanil-sn-glicerol, 1,2-dipalmitoil-sn-glicero-3-fosfoetanolamina-N-[metoxi(polietilenglicol)-350], 1,2-dipalmitoil-sn-glicero-3-fosfoetanolamina-N-[metoxi(polietilenglicol)-550], 1,2-dipalmitoil-sn-glicero-3-fosfoetanolamina-N-[metoxi(polietilenglicol)-750], 1,2-dipalmitoil-sn-glicero-3-fosfoetanolamina-N-[metoxi(polietilenglicol)-1000], 1,2-dipalmitoil-sn-glicero-3-fosfoetanolamina-N-[metoxi(polietilenglicol)-7000], 1,2-dioleoil-sn-glicero-3-fosfoetanolamina-N-lactosil, ganglósido GM1, lisofosfatidilcolina (LPC) o cualquier

combinación de los mismos.

Como se muestra, la bicapa lipídica 714 está incluida en un único nanoporo 716, por ejemplo, formado por un único PNTMC. Como se describe anteriormente, se puede formar el nanoporo 716 insertando un único PNTMC en la bicapa lipídica 714 por electroporación. El nanoporo 716 puede ser lo suficientemente grande como para permitir el paso de al menos una porción del analito de interés y/o iones pequeños (por ejemplo, Na^+, K^+, Ca^{2+}, Cl^-) entre los dos lados de la bicapa lipídica 714.

La cámara de muestra 715 está sobre la bicapa lipídica 714 y puede contener una solución del analito de interés para su caracterización. La solución puede ser una solución acuosa que contiene electrólito a granel 708 y tamponada a una concentración de iones óptima y mantenida a un pH óptimo para mantener abierto el nanoporo 716. El nanoporo 716 cruza la bicapa lipídica 714 y proporciona la única ruta para el flujo iónico desde el electrólito a granel 708 al electrodo de trabajo 702. Además de los nanoporos (por ejemplo, los PNTMC) y el analito de interés, el electrólito a granel 708 puede incluir además uno o más de los siguientes: cloruro de litio (LiCl), cloruro de sodio (NaCl), cloruro de potasio (KCl), glutamato de litio, glutamato de sodio, glutamato de potasio, acetato de litio, acetato de sodio, acetato de potasio, cloruro de calcio ($CaCl_2$), cloruro de estroncio ($SrCl_2$), cloruro de manganeso ($MnCl_2$) y cloruro de magnesio ($MgCl_2$).

El contraelectrodo (CE) 710 puede ser un sensor de potencial electroquímico. En algunos modos de realización, el contraelectrodo 710 se comparte entre una pluralidad de celdas de nanoporos y, por lo tanto, se puede denominar electrodo común. En algunos casos, el potencial común y el electrodo común pueden ser comunes a todas las celdas de nanoporos, o al menos a todas las celdas de nanoporos dentro de una agrupación particular. El electrodo común se puede configurar para aplicar un potencial común al electrólito a granel 708 en contacto con el nanoporo 716. El contraelectrodo 710 y el electrodo de trabajo 702 se pueden acoplar a la fuente de señales 728 para proporcionar un estímulo eléctrico (por ejemplo, polarización de voltaje) a través de la bicapa lipídica 714, y se pueden usar para detectar características eléctricas de la bicapa lipídica 714 (por ejemplo, resistencia, capacidad, caída de voltaje y flujo de corriente iónica). En algunos modos de realización, la celda de nanoporo 700 también puede incluir un electrodo de referencia 712.

Aunque estos modos de realización se refieren a bicapas lipídicas, también se aprecia que los presentes sistemas pueden usar cualquier membrana semipermeable que permita el flujo transmembranario de agua pero que tenga una permeabilidad limitada o nula al flujo de iones u otros osmolitos. Por ejemplo, los procedimientos y sistemas divulgados se pueden usar con membranas que sean poliméricas. En algunos modos de realización, la membrana es un copolímero. En algunos modos de realización, la membrana es un copolímero tribloque. En un modo de realización ejemplar, la membrana es un copolímero tribloque A-B-A en el que "A" es poli-b-(metiloxazolina) y "B" es poli(dimetilsiloxano)-poli-b-(metiloxazolina) (membrana Pmoxa-PDMS-Pmoxa).

En algunos modos de realización, se realizan diversas comprobaciones durante la creación de la celda de nanoporo como parte de la calibración. Una vez que se crea una celda de nanoporo, se pueden realizar etapas de calibración adicionales, por ejemplo, para identificar celdas de nanoporos que están funcionando como se desea (por ejemplo, un nanoporo en la celda). Dichas comprobaciones de calibración pueden incluir comprobaciones físicas, calibración de voltaje, calibración de canal abierto e identificación de celdas con un único nanoporo.

Señales de detección de celda de secuenciación de nanoporos

Las celdas de nanoporos en el chip sensor de nanoporos, tales como las celdas de nanoporos 750 en el chip sensor de nanoporos 700, pueden posibilitar la secuenciación en paralelo usando una técnica de secuenciación por síntesis basada en nanoporos (nano-SBS) de una única molécula.

La FIG. 8 ilustra un modo de realización de una celda de nanoporo 800 que realiza la secuenciación de nucleótidos usando la técnica nano-SBS. En la técnica nano-SBS, una molde 832 que se va a secuenciar (por ejemplo, una molécula de ácido nucleico u otro analito de interés) y un cebador se pueden introducir en el electrólito a granel 808 en la cámara de muestra de la celda de nanoporo 800. Como ejemplo, el molde 832 puede ser circular o lineal. Un cebador de ácido nucleico se puede hibridar con una porción del molde 832 a la que se pueden añadir cuatro nucleótidos 838 marcados con polímero de forma diferente.

En algunos modos de realización, una enzima (por ejemplo, una polimerasa 834, tal como una ADN polimerasa) se asocia con el nanoporo 816 para su uso en la síntesis de una hebra complementaria al molde 832. Por ejemplo, la polimerasa 834 se puede fijar covalentemente al nanoporo 816. La polimerasa 834 puede catalizar la incorporación de los nucleótidos 838 al cebador usando una molécula de ácido nucleico monocatenario como molde. Los nucleótidos 838 pueden comprender especies de marca ("marcas"), siendo el nucleótido uno de cuatro tipos diferentes: A, T, G o C para desoxirribonucleótidos, o A, U, G o C para ribonucleótidos. Cuando un nucleótido marcado forma un complejo correctamente con la polimerasa 834, la marca se puede atraer (por ejemplo, cargar) hacia el nanoporo por una fuerza eléctrica, tal como una fuerza generada en presencia de un campo eléctrico generado por un voltaje aplicado a través de la bicapa lipídica 814 y/o el nanoporo 816. La cola de la marca se puede situar en el barril del nanoporo 816. La marca contenida en el barril del nanoporo 816 puede generar una

señal de bloqueo iónico única 840 debido a la estructura química y/o tamaño distintivos de la marca, identificando electrónicamente de este modo la base añadida a la que se fija la marca.

Como se usa en el presente documento, una marca "cargada" o "de paso" es una que se sitúa en y/o permanece en o cerca del nanoporo durante una cantidad de tiempo apreciable, por ejemplo, de 0,1 milisegundos (ms) a 10000 ms. En algunos casos, una marca se carga en el nanoporo antes de liberarse del nucleótido. En algunos casos, la probabilidad de que una marca cargada pase a través del (y/o se detecte por el) nanoporo después de liberarse tras un acontecimiento de incorporación de nucleótido es adecuadamente alta, por ejemplo, de un 80 % a un 99 %.

En algunos modos de realización, antes de que la polimerasa 834 se conecte al nanoporo 816, la conductancia del nanoporo 816 es alta, tal como, por ejemplo, aproximadamente 300 *picosiemens* (300 pS). A medida que la marca se carga en el nanoporo, se genera una señal de conductancia única (por ejemplo, la señal 840) debido a la estructura química y/o el tamaño distintivos de la marca. Por ejemplo, la conductancia del nanoporo puede ser de aproximadamente 60 pS, 80 pS, 100 pS o 120 pS, correspondiente cada una a uno de los cuatro tipos de nucleótidos marcados. A continuación, la polimerasa puede experimentar una reacción de isomerización y transfosforilación para incorporar el nucleótido a la molécula de ácido nucleico en crecimiento y liberar la molécula de marca.

En algunos casos, algunos de los nucleótidos marcados pueden no coincidir (bases complementarias) con una posición actual de la molécula de ácido nucleico (molde). Los nucleótidos marcados que no están emparejados por base con la molécula de ácido nucleico también pueden pasar a través del nanoporo. Estos nucleótidos no emparejados se pueden rechazar por la polimerasa dentro de una escala de tiempo que es más corta que la escala de tiempo durante la que los nucleótidos emparejados correctamente permanecen asociados con la polimerasa. Las marcas unidas a nucleótidos no emparejados pueden pasar a través del nanoporo rápidamente y detectarse durante un corto período de tiempo (por ejemplo, menos de 10 ms), mientras que las marcas unidas a nucleótidos emparejados se pueden cargar en el nanoporo y detectarse durante un período de tiempo largo (por ejemplo, al menos 10 ms). Por lo tanto, los nucleótidos no emparejados se pueden identificar por un procesador hacia 3' en base al menos en parte en el tiempo durante el que se detecta el nucleótido en el nanoporo.

Se puede medir una conductancia (o equivalentemente la resistencia) del nanoporo que incluye la marca cargada (de paso) por medio de un valor de señal (por ejemplo, voltaje o una corriente que pasa a través del nanoporo), proporcionando de este modo una identificación de la especie de marca y, por tanto, del nucleótido en la posición actual. En algunos modos de realización, se aplica una señal de corriente continua (CC) a la celda de nanoporo (por ejemplo, de modo que la dirección en la que la marca se mueve a través del nanoporo no se invierta). Sin embargo, hacer funcionar un sensor de nanoporo durante períodos de tiempo largos usando una corriente continua puede cambiar la composición del electrodo, desequilibrar las concentraciones de iones a través del nanoporo y tener otros efectos indeseables que pueden afectar la vida útil de la celda de nanoporo. La aplicación de una forma de onda de corriente alterna (CA) puede reducir la electromigración para evitar estos efectos indeseables y tiene determinadas ventajas como se describe a continuación. Los procedimientos de secuenciación de ácidos nucleicos descritos en el presente documento que utilizan nucleótidos marcados son completamente compatibles con los voltajes de CA aplicados y, por lo tanto, se puede usar una forma de onda de CA para lograr estas ventajas.

La capacidad de recargar el electrodo durante el ciclo de detección de CA puede ser ventajosa cuando se usan electrodos de sacrificio, electrodos que cambian el carácter molecular en las reacciones que transportan corriente (por ejemplo, electrodos que comprenden plata) o electrodos que cambian el carácter molecular en las reacciones que transportan corriente. Un electrodo se puede agotar durante un ciclo de detección cuando se usa una señal de corriente continua. La recarga puede evitar que el electrodo alcance un límite de agotamiento, tal como que se agote completamente, lo que puede ser un problema cuando los electrodos son pequeños (por ejemplo, cuando los electrodos son lo suficientemente pequeños como para proporcionar una matriz de electrodos que tiene al menos 500 electrodos por milímetro cuadrado). La vida útil del electrodo en algunos casos varía con, y al menos depende en parte de, la anchura del electrodo.

Son conocidas en la técnica condiciones adecuadas para medir corrientes iónicas que pasan a través de los nanoporos y en el presente documento se proporcionan ejemplos. La medición se puede llevan a cabo con un voltaje aplicado a través de la membrana y el poro. En algunos modos de realización, el voltaje usado varía de -400 mV a +400 mV. El voltaje usado está preferentemente en un intervalo que tiene un límite inferior seleccionado de -400 mV, -300 mV, -200 mV, -150 mV, -100 mV, -50 mV, -20 mV y 0 mV y un límite superior seleccionado independientemente de +10 mV, +20 mV, +50 mV, +100 mV, +150 mV, +200 mV, +300 mV y +400 mV. El voltaje usado puede estar más preferentemente en el intervalo de 100 mV a 240 mV y lo más preferentemente en el intervalo de 160 mV a 240 mV. Es posible incrementar la discriminación entre diferentes nucleótidos por un nanoporo usando un potencial aplicado incrementado. La secuenciación de ácidos nucleicos usando formas de onda de CA y nucleótidos marcados se describe en la publicación de patente de EE. UU. n.º US 2014/0134616 titulada "Nucleic Acid Sequencing Using Tags", presentada el 6 de noviembre de 2013. Además de los nucleótidos marcados descritos en el documento US 2014/0134616, la secuenciación se puede realizar usando análogos de nucleótido que carecen de un resto glucídico o acíclico, por ejemplo, nucleósidos trifosfato de (S)-glicerol (gNTP)

de las cinco nucleobases comunes: adenina, citosina, guanina, uracilo y timina (Horhota *et al.* Organic Letters, 8:5345-5347 [2006]).

Circuito eléctrico de la celda de secuenciación de nanoporos

La FIG. 9 ilustra un modo de realización de un circuito eléctrico 900 (que puede incluir porciones del circuito eléctrico 722 en la FIG. 7) en una celda de nanoporo, tal como la celda de nanoporo 800. Como se describe anteriormente, en algunos modos de realización, el circuito eléctrico 900 incluye un contraelectrodo 910 que se puede compartir entre una pluralidad de celdas de nanoporos o todas las celdas de nanoporos en un chip sensor de nanoporos y, por lo tanto, también se puede denominar electrodo común. El electrodo común se puede configurar para aplicar un potencial común al electrólito a granel (por ejemplo, electrólito a granel 808) en contacto con la bicapa lipídica (por ejemplo, bicapa lipídica 814) en las celdas de nanoporos conectándose a una fuente de voltaje $V_{LÍQ}$ 920. En algunos modos de realización, se utiliza un modo no faradaico de CA para modular el voltaje $V_{LÍQ}$ con una señal de CA (por ejemplo, una onda cuadrada) y aplicarla al electrólito a granel en contacto con la bicapa lipídica en la celda de nanoporo. En algunos modos de realización, $V_{LÍQ}$ es una onda cuadrada con una magnitud de ±200-250 mV y una frecuencia entre, por ejemplo, 25 y 400 Hz. El electrólito a granel entre el contraelectrodo 910 y la bicapa lipídica (por ejemplo, la bicapa lipídica 814) se puede modelar por un condensador grande (no mostrado), tal como, por ejemplo, de 100 µF o más grande.

La FIG. 9 también muestra un modelo eléctrico 922 que representa las propiedades eléctricas de un electrodo de trabajo 902 (por ejemplo, el electrodo de trabajo 802) y la bicapa lipídica (por ejemplo, la bicapa lipídica 814). El modelo eléctrico 922 incluye un condensador 926 (C_{Bicapa}) que modela una capacidad asociada a la bicapa lipídica y una resistencia 928 (R_{PORO}) que modela una resistencia variable asociada con el nanoporo, que puede cambiar en base a la presencia de una marca particular en el nanoporo. El modelo eléctrico 922 también incluye un condensador 924 que tiene una capacidad de doble capa ($C_{Doble\ capa}$) y que representa las propiedades eléctricas del electrodo de trabajo 902 y del pocillo 705. El electrodo de trabajo 902 se puede configurar para aplicar un potencial distintivo independiente de los electrodos de trabajo en otras celdas de nanoporos.

El dispositivo de paso 906 es un interruptor que se puede usar para conectar o desconectar la bicapa lipídica y el electrodo de trabajo del circuito eléctrico 900. El dispositivo de paso 906 se puede controlar por la línea de control 907 para posibilitar o imposibilitar que se aplique un estímulo de voltaje a través de la bicapa lipídica en la celda de nanoporo. Antes de que se depositen los lípidos para formar la bicapa lipídica, la impedancia entre los dos electrodos puede ser muy baja porque el pocillo de la celda de nanoporo no está sellado y, por lo tanto, el dispositivo de paso 906 se puede mantener abierto para evitar una condición de cortocircuito. El dispositivo de paso 906 se puede cerrar después de que se haya depositado el disolvente lipídico en la celda de nanoporo para sellar el pocillo de la celda de nanoporo.

El circuito 900 puede incluir además un condensador integrador en chip 908 (n_{Con}). El condensador integrador 908 se puede precargar usando una señal de reinicio 903 para cerrar el interruptor 901, de modo que el condensador integrador 908 se conecte a una fuente de voltaje V_{PRE} 905. En algunos modos de realización, la fuente de voltaje V_{PRE} 905 proporciona un voltaje de referencia constante con una magnitud de, por ejemplo, 900 mV. Cuando el interruptor 901 se cierra, el condensador integrador 908 se puede precargar al nivel de voltaje de referencia de la fuente de voltaje V_{PRE} 905.

Después de que el condensador integrador 908 se precarga, la señal de reinicio 903 se puede usar para abrir el interruptor 901 de modo que el condensador integrador 908 se desconecte de la fuente de voltaje V_{PRE} 905. En este punto, dependiendo del nivel de la fuente de voltaje $V_{LÍQ}$, el potencial del contraelectrodo 910 puede estar en un nivel mayor que el del potencial del electrodo de trabajo 902 (y el condensador integrador 908) o viceversa. Por ejemplo, durante una fase positiva de una onda cuadrada desde una fuente de voltaje $V_{LÍQ}$ (por ejemplo, el período brillante u oscuro del ciclo de señales de fuente de voltaje de CA), el potencial del contraelectrodo 910 está en un nivel mayor que el potencial del electrodo de trabajo 902. Durante una fase negativa de la onda cuadrada de la fuente de voltaje $V_{LÍQ}$ (por ejemplo, el período oscuro o brillante del ciclo de señales de fuente de voltaje de CA), el potencial del contraelectrodo 910 está en un nivel menor que el del potencial del electrodo de trabajo 902. Por tanto, en algunos modos de realización, el condensador integrador 908 se puede cargar además durante el período brillante a partir del nivel de voltaje precargado de la fuente de voltaje V_{PRE} 905 a un nivel mayor, y descargarse durante el período oscuro a un nivel menor, debido a la diferencia de potencial entre el contraelectrodo 910 y el electrodo de trabajo 902. En otros modos de realización, la carga y la descarga se producen en períodos oscuros y períodos brillantes, respectivamente.

El condensador integrador 908 se puede cargar o descargar durante un período de tiempo fijo, dependiendo de la tasa de muestreo de un convertidor de analógico a digital (ADC) 935, que puede ser mayor que 1 kHz, 5 kHz, 10 kHz, 100 kHz o más. Por ejemplo, con una tasa de muestreo de 1 kHz, el condensador integrador 908 se puede cargar/descargar durante un período de aproximadamente 1 ms y, a continuación, el nivel de voltaje se puede muestrear y convertir por el ADC 935 al final del período de integración. Un nivel de voltaje particular correspondería a una especie de marca particular en el nanoporo y, por tanto, correspondería al nucleótido en una posición actual en el molde.

Después de muestrearse por el ADC 935, el condensador integrador 908 se puede precargar de nuevo usando la señal de reinicio 903 para cerrar el interruptor 901, de modo que el condensador integrador 908 se conecte a la fuente de voltaje V_{PRE} 905 de nuevo. Las etapas de precarga del condensador integrador 908, espera de un período de tiempo fijo para que el condensador integrador 908 se cargue o descargue, y muestreo y conversión del nivel de voltaje del condensador integrador por el ADC 935 se pueden repetir en ciclos en todo el proceso de secuenciación.

Un procesador digital 930 puede procesar los datos de salida de ADC, por ejemplo, para normalización, almacenamiento intermedio de datos, filtrado de datos, compresión de datos, reducción de datos, extracción de acontecimientos o ensamblaje de datos de salida de ADC de la matriz de celdas de nanoporos en diversas tramas de datos. En algunos modos de realización, el procesador digital 930 realiza un procesamiento posterior adicional, tal como la determinación de base. El procesador digital 930 se puede implementar como equipo informático (por ejemplo, en una unidad de procesamiento de gráficos (GPU), FPGA, ASIC, etc.) o como una combinación de equipo informático y programa informático.

En consecuencia, la señal de voltaje aplicada a través del nanoporo se puede usar para detectar estados particulares del nanoporo. Uno de los posibles estados del nanoporo es un estado de canal abierto cuando un polifosfato fijado a marca está ausente del barril del nanoporo, también denominado en el presente documento estado de no paso a través del nanoporo. Otros cuatro estados posibles del nanoporo corresponden cada uno a un estado cuando uno de los cuatro tipos diferentes de nucleótidos de polifosfato fijados a marca (A, T, G o C para desoxirribonucleótidos, o A, U, G o C para ribonucleótidos) está contenido en el barril del nanoporo. Aún otro posible estado del nanoporo es cuando la bicapa lipídica se rompe.

Cuando se mide el nivel de voltaje en el condensador integrador 908 después de un período de tiempo fijo, los diferentes estados de un nanoporo pueden dar como resultado mediciones de diferentes niveles de voltaje. Esto se debe a que la tasa de caída de voltaje (disminución por descarga o incremento por carga) en el condensador integrador 908 (es decir, la inclinación de la pendiente de un gráfico de voltaje en el condensador integrador 908 frente al tiempo) depende de la resistencia de nanoporo (por ejemplo, la resistencia de la resistencia R_{PORO} 928). Más en particular, como la resistencia asociada con el nanoporo en diferentes estados es diferente debido a las estructuras químicas distintivas de las moléculas (marcas), se pueden observar diferentes tasas correspondientes de caída de voltaje y se pueden usar para identificar los diferentes estados del nanoporo. La curva de caída de voltaje puede ser una curva exponencial con una constante de tiempo de RC $\tau = RC$, donde R es la resistencia asociada con el nanoporo (es decir, resistencia R_{PORO} 928) y C es la capacidad asociada con la membrana (es decir, condensador C_{Bicapa} 926) en paralelo con R. Una constante de tiempo de la celda de nanoporo puede ser, por ejemplo, aproximadamente 200-500 ms. La curva de caída puede no ajustarse exactamente a una curva exponencial debido a la implementación detallada de la bicapa, pero la curva de caída puede ser similar a una curva exponencial y ser repetitiva, permitiendo, por tanto, la detección de marcas.

En algunos modos de realización, la resistencia asociada con el nanoporo en un estado de canal abierto está en el intervalo de 100 MOhm a 20 GOhm. En algunos modos de realización, la resistencia asociada con el nanoporo en un estado donde una marca está en el interior del barril del nanoporo puede estar dentro del intervalo de 200 MOhm a 100 GOhm. En otros modos de realización, se omite el condensador integrador 908, ya que el voltaje que se conduce al ADC 935 variará todavía debido a la caída de voltaje en el modelo eléctrico 922.

La tasa de la caída del voltaje en el condensador integrador 908 se puede determinar de diferentes maneras. Como se expone anteriormente, la tasa de la caída de voltaje se puede determinar midiendo una caída de voltaje durante un intervalo de tiempo fijo. Por ejemplo, se puede medir el voltaje en el condensador integrador 908, en primer lugar, por el ADC 935 en el tiempo t1 y, a continuación, se mide el voltaje de nuevo por el ADC 935 en el tiempo t2. La diferencia de voltaje es mayor cuando la pendiente del voltaje en el condensador integrador 908 frente al tiempo es más pronunciada, y la diferencia de voltaje es más pequeña cuando la pendiente de la curva de voltaje es menos pronunciada. Por tanto, se puede usar la diferencia de voltaje como una métrica para determinar la tasa de la caída del voltaje en el condensador integrador 908 y, por tanto, el estado de la celda de nanoporo.

En otros modos de realización, la tasa de la caída de voltaje se determina midiendo una duración de tiempo que se requiere para una cantidad seleccionada de caída de voltaje. Por ejemplo, se puede medir el tiempo requerido para que el voltaje descienda o se incremente desde un primer nivel de voltaje V1 hasta un segundo nivel de voltaje V2. El tiempo requerido es menor cuando la pendiente de la curva de voltaje frente a tiempo es más pronunciada, y el tiempo requerido es mayor cuando la pendiente de la curva de voltaje frente a tiempo es menos pronunciada. Por tanto, se puede usar el tiempo medido requerido como una métrica para determinar la tasa de la caída del voltaje en el condensador integrador n_{con} 908 y, por tanto, el estado de la celda de nanoporo. Un experto en la técnica apreciará los diversos circuitos que se pueden usar para medir la resistencia del nanoporo, por ejemplo, incluyendo las técnicas de medición de valor de señal, tales como mediciones de voltaje o corriente.

En algunos modos de realización, el circuito eléctrico 900 no incluye un dispositivo de paso (por ejemplo, el dispositivo de paso 906) y un condensador adicional (por ejemplo, el condensador integrador 908 (n_{con})) que se

fabrican en el chip, facilitando de este modo la reducción en el tamaño del chip de secuenciación basada en nanoporos. Debido a la naturaleza delgada de la membrana (bicapa lipídica), la capacidad asociada con la membrana (por ejemplo, el condensador 926 (C_{Bicapa})) por sí sola puede ser suficiente para crear la constante de tiempo de RC requerida sin necesidad de la capacidad en el chip adicional. Por lo tanto, se puede usar el condensador 926 como condensador integrador y se puede precargar por la señal de voltaje V_{PRE} y posteriormente descargarse o cargarse por la señal de voltaje $V_{LÍQ}$. La eliminación del condensador adicional y del dispositivo de paso que de otro modo se fabrican en el chip del circuito eléctrico puede reducir significativamente la superficie ocupada de una única celda de nanoporo en el chip de secuenciación de nanoporos, facilitando de este modo el escalamiento del chip de secuenciación de nanoporos para incluir cada vez más celdas (por ejemplo, tener millones de celdas en un chip de secuenciación de nanoporos).

B. Muestreo de datos en una celda de nanoporo

Para realizar la secuenciación de un ácido nucleico, se puede muestrear el nivel de voltaje del condensador integrador (por ejemplo, el condensador integrador 908 (n_{con}) o el condensador 926 (C_{Bicapa})) y convertir por el ADC (por ejemplo, el ADC 935) mientras se añade un nucleótido marcado al ácido nucleico. La marca del nucleótido se puede empujar hacia el barril del nanoporo por el campo eléctrico a través del nanoporo que se aplica a través del contraelectrodo y el electrodo de trabajo, por ejemplo, cuando el voltaje aplicado es de modo que $V_{LÍQ}$ es menor que V_{PRE}.

1. Paso

Un acontecimiento de paso es cuando un nucleótido marcado se fija al molde (por ejemplo, un fragmento de ácido nucleico) y la marca se mueve dentro y fuera del barril del nanoporo. Este movimiento puede ocurrir múltiples veces durante un acontecimiento de paso. Cuando la marca está en el barril del nanoporo, la resistencia del nanoporo puede ser mayor y una corriente menor puede fluir a través del nanoporo.

Durante la secuenciación, es posible que una marca no esté en el nanoporo en algunos ciclos de CA (denominado estado de canal abierto), donde la corriente es la más alta debido a la menor resistencia del nanoporo. Cuando una marca se atrae hacia el barril del nanoporo, el nanoporo está en modo brillante. Cuando la marca se expulsa fuera del barril del nanoporo, el nanoporo está en modo oscuro.

2. Período brillante y oscuro

Durante un ciclo de CA, se puede muestrear el voltaje en el condensador integrador múltiples veces por el ADC. Por ejemplo, en un modo de realización, se aplica una señal de voltaje de CA a través del sistema a, por ejemplo, aproximadamente 100 Hz, y una tasa de adquisición del ADC puede ser de aproximadamente 2000 Hz por celda o mayor (incluyendo, por ejemplo, aproximadamente 4000 Hz por celda). Por tanto, puede haber al menos aproximadamente 20 puntos de datos (mediciones de voltaje) capturados por ciclo de CA (ciclo de una forma de onda de CA). Los puntos de datos correspondientes a un ciclo de la forma de onda de CA se pueden denominar conjunto. En un conjunto de puntos de datos para un ciclo de CA, puede haber un subconjunto capturado cuando, por ejemplo, $V_{LÍQ}$ es menor que V_{PRE}, que puede corresponder a un modo (período) brillante cuando la marca se fuerza hacia el barril del nanoporo. Otro subconjunto puede corresponder a un modo (período) oscuro cuando la marca se empuja fuera del barril del nanoporo por el campo eléctrico aplicado cuando, por ejemplo, $V_{LÍQ}$ es mayor que V_{PRE}.

3. Voltajes medidos

Para cada punto de datos, cuando se abre el interruptor 901, el voltaje en el condensador integrador (por ejemplo, el condensador integrador 908 (n_{con}) o el condensador 926 (C_{Bicapa})) cambiará de manera decreciente como resultado de la carga/descarga por $V_{LÍQ}$, por ejemplo, como un incremento de V_{PRE} a $V_{LÍQ}$ cuando $V_{LÍQ}$ es mayor que V_{PRE} o una disminución de V_{PRE} a $V_{LÍQ}$ cuando $V_{LÍQ}$ es menor que V_{PRE}. Los valores de voltaje final pueden diferir de $V_{LÍQ}$ a medida que se carga el electrodo de trabajo. La tasa de cambio del nivel de voltaje en el condensador integrador se puede regir por el valor de la resistencia de la bicapa, que puede incluir el nanoporo, que a su vez puede incluir una molécula (por ejemplo, una marca de nucleótidos marcados) en el nanoporo. El nivel de voltaje se puede medir en un tiempo predeterminado después de que se abre el interruptor 901.

El interruptor 901 puede funcionar a la tasa de adquisición de datos. El interruptor 901 se puede cerrar durante un período de tiempo relativamente corto entre dos adquisiciones de datos, típicamente justo después de una medición realizada por el ADC. El interruptor permite recopilar múltiples puntos de datos durante cada subperíodo (brillante u oscuro) de cada ciclo de CA de $V_{LÍQ}$. Si el interruptor 901 permanece abierto, el nivel de voltaje en el condensador integrador y, por tanto, el valor de salida del ADC, cae completamente y permanece allí. Si en cambio se cierra el interruptor 901, el condensador integrador se precarga de nuevo (a V_{PRE}) y queda listo para otra medición. Por tanto, el interruptor 901 permite recopilar múltiples puntos de datos para cada subperíodo (brillante u oscuro) de cada ciclo de CA. Dichas mediciones múltiples pueden permitir una mayor resolución con un ADC fijo (por ejemplo, de 8 bits a 14 bits debido al mayor número de mediciones, que se pueden promediar). Las

mediciones múltiples también pueden proporcionar información cinética sobre la molécula que pasa a través del nanoporo. La información de tiempo puede permitir la determinación de cuánto tiempo tarda en pasar. Esto también se puede usar para ayudar a determinar si se secuencian múltiples nucleótidos que se añaden a la hebra de ácido nucleico.

La FIG. 10 muestra puntos de datos de ejemplo capturados de una celda de nanoporo durante períodos brillantes y períodos oscuros de ciclos de CA. En la FIG. 10, el cambio en los puntos de datos se exagera con propósitos ilustrativos. El voltaje (V_{PRE}) aplicado al electrodo de trabajo o al condensador integrador está en un nivel constante, tal como, por ejemplo, 900 mV. Una señal de voltaje 1010 ($V_{LÍQ}$) aplicada al contraelectrodo de las celdas de nanoporos es una señal de CA mostrada como una onda rectangular, donde el ciclo de trabajo puede ser cualquier valor adecuado, tal como menor que o igual a un 50 %, por ejemplo, aproximadamente un 40 %.

Durante un período brillante 1020, la señal de voltaje 1010 ($V_{LÍQ}$) aplicada al contraelectrodo es menor que el voltaje V_{PRE} aplicado al electrodo de trabajo, de modo que una marca se puede forzar hacia el barril del nanoporo por el campo eléctrico provocado por los diferentes niveles de voltaje aplicados en el electrodo de trabajo y el contraelectrodo (por ejemplo, debido a la carga en la marca y/o flujo de los iones). Cuando se abre el interruptor 1001, el voltaje en un nodo antes del ADC (por ejemplo, en un condensador integrador) disminuirá. Después de capturar un punto de datos de voltaje (por ejemplo, después de un período de tiempo especificado), se puede cerrar el interruptor 1001 y el voltaje en el nodo de medición se incrementará de vuelta a V_{PRE} de nuevo. El proceso se puede repetir para medir múltiples puntos de datos de voltaje. De esta manera, se pueden capturar múltiples puntos de datos durante el período brillante.

Como se muestra en la FIG. 10, un primer punto de datos 1022 (también denominado primer punto delta (FPD)) en el período brillante después de un cambio en el signo de la señal de $V_{LÍQ}$ puede ser menor que los puntos de datos posteriores 1024. Esto se puede deber a que no hay marca en el nanoporo (canal abierto) y, por tanto, tiene una baja resistencia y una alta tasa de descarga. En algunos casos, el primer punto de datos 1022 puede superar el nivel de $V_{LÍQ}$ como se muestra en la FIG. 10. Esto puede estar provocado por la capacidad de la bicapa que acopla la señal al condensador en el chip. Los puntos de datos 1024 se pueden capturar después de que se haya producido un acontecimiento de paso, es decir, se fuerza una marca hacia el barril del nanoporo, donde la resistencia del nanoporo y, por tanto, la tasa de descarga del condensador integrador dependen del tipo particular de marca que se fuerza hacia el barril del nanoporo. Los puntos de datos 1024 pueden disminuir ligeramente para cada medición debido a la carga acumulada en $C_{Doble\ capa}$ 1024, como se menciona a continuación.

Durante un período oscuro 1030, la señal de voltaje 1010 ($V_{LÍQ}$) aplicada al contraelectrodo es mayor que el voltaje (V_{PRE}) aplicado al electrodo de trabajo, de modo que cualquier marca se expulsaría fuera del barril del nanoporo. Cuando se abre el interruptor 1001, el voltaje en el nodo de medición se incrementa porque el nivel de voltaje de la señal de voltaje 1010 ($V_{LÍQ}$) es mayor que V_{PRE}. Después de capturar un punto de datos de voltaje (por ejemplo, después de un período de tiempo especificado), el interruptor 1001 se puede cerrar y el voltaje en el nodo de medición disminuirá de vuelta a V_{PRE} de nuevo. El proceso se puede repetir para medir múltiples puntos de datos de voltaje. Por tanto, se pueden capturar múltiples puntos de datos durante el período oscuro, incluyendo un primer punto delta 1032 y puntos de datos posteriores 1034. Como se describe anteriormente, durante el período oscuro, cualquier marca de nucleótido se expulsa fuera del nanoporo y, por tanto, se obtiene una información mínima sobre cualquier marca de nucleótido, además de para su uso en la normalización.

La FIG. 10 también muestra que durante el período brillante 1040, aunque la señal de voltaje 1010 ($V_{LÍQ}$) aplicada al contraelectrodo es menor que el voltaje (V_{PRE}) aplicado al electrodo de trabajo, no se produce ningún acontecimiento de paso (canal abierto). Por tanto, la resistencia del nanoporo es baja y la tasa de descarga del condensador integrador es alta. Como resultado, los puntos de datos capturados, incluyendo un primer punto de datos 1042 y puntos de datos posteriores 1044, muestran niveles de voltaje bajo.

Se podría esperar que el voltaje medido durante un período brillante u oscuro sea aproximadamente el mismo para cada medición de una resistencia constante del nanoporo (por ejemplo, realizada durante un modo brillante de un ciclo de CA dado mientras una marca está en el nanoporo), pero este puede no ser el caso cuando la carga se acumula en el condensador de doble capa 924 ($C_{Doble\ capa}$). Esta acumulación de carga puede provocar que la constante de tiempo de la celda de nanoporo se alargue. Como resultado, el nivel de voltaje se puede mover, provocando de este modo que el valor medido disminuya para cada punto de datos en un ciclo. Por tanto, dentro de un ciclo, los puntos de datos pueden cambiar un poco de un punto de datos a otro, como se muestra en la FIG. 10.

Se pueden encontrar detalles adicionales con respecto a las mediciones en, por ejemplo, la publicación de patente de EE. UU. n.º 2016/0178577 titulada "Nanopore-Based Sequencing With Varying Voltage Stimulus", la publicación de patente de EE. UU. n.º 2016/0178554 titulada "Nanopore-Based Sequencing With Varying Voltage Stimulus", la solicitud de patente de EE. UU. n.º 15/085.700 titulada "Non-Destructive Bilayer Monitoring Using Measurement Of Bilayer Response To Electrical Stimulus" y la solicitud de patente de EE. UU. n.º 15/085.713 titulada "Electrical Enhancement Of Bilayer Formation".

4. Normalización y llamada de bases

Para cada celda de nanoporo utilizable del chip sensor de nanoporos, se puede realizar un modo de producción para secuenciar ácidos nucleicos. Los datos de salida de ADC capturados durante la secuenciación se pueden normalizar para proporcionar una mayor exactitud. La normalización puede tener en cuenta efectos de desviación, tales como conformación de ciclo, variaciones de ganancia, desviación de inyección de carga y desviación de valor de referencia. En algunas implementaciones, los valores de señal de un ciclo de período brillante correspondiente a un acontecimiento de paso se pueden aplanar de modo que se obtenga un único valor de señal para el ciclo (por ejemplo, un promedio) o se pueden realizar ajustes a la señal medida para reducir la caída intraciclo (un tipo de efecto de conformación de ciclo). Las variaciones de ganancia, en general, escalan toda la señal y cambian en el orden de cientos a miles de segundos. Como ejemplo, las variaciones de ganancia se pueden desencadenar por cambios en la solución (resistencia de poro) o cambios en la capacidad de bicapa. La desviación de valor de referencia se produce en una escala de tiempo de ~100 ms y se refiere a una desviación de voltaje en el electrodo de trabajo. La desviación de valor de referencia puede estar impulsada por cambios en una proporción de rectificación eficaz a partir del paso como resultado de la necesidad de mantener el equilibrio de carga en la celda de secuenciación desde el período brillante hasta el período oscuro.

Después de la normalización, los modos de realización pueden determinar agrupamientos de voltajes para los canales de paso, donde cada agrupamiento corresponde a una especie de marca diferente y, por tanto, a un nucleótido diferente. Los agrupamientos se pueden usar para determinar las probabilidades de que un voltaje dado corresponda a un nucleótido dado. Como otro ejemplo, los agrupamientos se pueden usar para determinar voltajes de corte para discriminar entre diferentes nucleótidos (bases).

VI. Ejemplos

La presente invención se describe con mayor detalle en los siguientes ejemplos, que pretenden ilustrar, pero no limitar, la invención reivindicada.

VIA. Ejemplo 1: generación de α-HL con sustituciones de aminoácido cargado negativamente

VIA1. Generación de monómeros variantes

Se adquirió el ADN que codifica un α-HL natural que tiene la secuencia de aminoácidos de SEQ ID NO: 1 de una fuente comercial. Se realizaron las modificaciones de secuencia por mutagénesis dirigida al sitio usando un kit de mutagénesis dirigida a múltiples sitios QuikChange (Stratagene, La Jolla, CA). Todos los polinucleótidos modificados codifican una variante de α-HL que tiene al menos un aminoácido cargado negativamente no natural y una marca de epítopo escindible (tal como un conector C terminal/TEV/marca His). Otras modificaciones incluyen modificaciones que controlan la oligomerización (por ejemplo, sustituciones H35G/L y, en algunos casos, H144A), ensanchan el sitio de constricción (por ejemplo, E111A, K147A y/o M113D), mejoran el CV de poro (D227N). La tabla 6 enumera los monómeros de α-HL ejemplares codificados por los polinucleótidos modificados:

Tabla 6

ID	Modificaciones en relación con SEQ ID NO: 1	AA negativo no natural	Otras sustituciones / Propósito	SEQ ID NO
G617	A1D / H35G / H144A	Ent.: A1D	CO: H35G, H144A	31
G619	K8D/H35G/H144A	Ent.: K8D	CO: H35G, H144A	32
G623	H35L/K46D	Ent. K46D	CO: H35L	33
G697	H35G / N47D / H144A / V149D	Ent.: N47D, V149D	CO: H35G, H144A	34
G724	A1D / K8D / H35G / N47D / H144A / V149D	Ent. A1D,K8D,N47D,V149D	CO: H35G, H144A	35
G1143	K8D / D13N / H35G / H144A	Ent. K8D	CO: H35G, H144A EX: D13N	36
G1147	K8D / D13N/ H35G / H144A / V149D	Ent. K8D, V149D	EX: D13N CO: H35G, H144A	37
G1313	0D / H35G / H144A	Ent. 0D	CO: H35G, H144A	38
G1314	0D / A1D / H35G / H144A	Ent. 0D, A1D	CO: H35G, H144A	39
G1315	0D / H35G / N47D / H144A	Ent. 0D, N47D	CO: H35G, H144A	40
G1316	0D / K8D / H35G / H144A	Ent. 0D, K8D	CO: H35G, H144A	41
G1317	0D / A1D / K8D / H35G / H144A	Ent. 0D, A1D, K8D	CO: H35G, H144A	42
G1429	A1D / H35G / M113D / T115D / T117D/N121D / H144A	Ent. A1D Con. M113D BB: T115D, T117D, N121D	CO: H35G, H144A	43

G1440	A1D / S3D / G10D / H35G /K46D/ S106D / H144A	Ent. A1D,S3D,G10D,K46D,S 106D	CO: H35G, H144A	44
G1441	G10D / H35G / M113D / T115D / T117D/N121D / H144A	Ent. G10D Con.: M113D BB: 115D, T117D, N121D	CO: H35G, H144A	45
G1143	K8D / D13N / H35G / H144A	Ent.: K8D	CO: H35L, H144A EX: D13N	46
G1147	K8D / D13N / H35G / V149D	Ent.: K8D, V149D	CO: H35L EX: D13N	47
G1313	0D / H35G / H144A	Ent.: 0D	CO: H35G, H144A	48
G1314	0D / A1D / H35G / H144A	Ent.: 0D, A1D	CO: H35G, H144A	49
G1315	0D / H35G / N47D / H144A	Ent.: 0D, N47D	CO: H35G, H144A	50
G1316	0D / K8D / H35G / H144A	Ent.: 0D, K8D	CO: H35G, H144A	51
G1317	0D / A1D / K8D / H35G / H144A	Ent.: 0D, A1D, K8D	CO: H35G, H144A	52
G1515	H35G / E111A / M113D / T115D / T117D / N121D / H144A/K147A	Con.: M113D BB: T115D, T117D, N121D	WC: E111A, K147A, M113D CO: H144A, H35G	53
G1608	A1D / T9D / H35G / N47K / T129D / G130D / H144A	Ent.: A1D, T9D BB: T129D, G130D	CO: H144A, H35G EX: N47K	54
G1609	A1D / H35G / N47K / T129D / G130D / H144A	Ent.: A1D BB: T129D, G130D	CO: H144A, H35G EX: N47K	55
G1633	H35G / M113D / T115D / T117D/N121D	Con.: M113D BB: T115D, T117D, N121D	CO: H35G	56
G1637	A1D / H35G / N47K / T129D / G130D / K131D / H144A / K147D	Ent.: A1D Const.: K147D BB: T129D, G130D Salida: K131D	CO: H144A, H35G EX: N47K	57
G1646	A1D / N17D / H35G / K131D/H144A	Ent.: A1D, N17D Salida: K131D	CO: H144A, H35G	58
G1687	A1D / K131D / K147D	Ent.: A1D Salida: K131D K147D	Ninguna	59
G1698	N17D / H35G / N47K / G77K / E111A / D113A / T115D / T117D / N121D / D128K / K147A / N173K	Ent.: N17D BB: T115D, T117D, N121D	WC: E111A, K147A CO: H35G EX: N47K pI: G77K, D128K, N173K	60
G1701	N17D / N47K / G77K / E111A/D113A/T115D/	Ent.: N17D BB: T115D, T117D, N121D	WC: E111A, D113A, K147A CO: H35G	61
	T117D / N121D / D128K / K147A/N173K		EX: N47K pI: G77K, D128K, N173K	
G1713	A1D / H35L / K131D / K147D	Ent.: A1D Con.: K147D Salida: K131D	CO: H35L	62
G1748	A1D / K8D / H35L / K131D / H144A / K147D	Ent.: A1D, K8D Con.: K147D Salida: K131D	CO: H35L, H144A	63
G1749	A1D / N17D / H35L / K131D / H144A / K147D	Ent.: A1D, N17D Con.: K147D Salida: K131D	CO: H35L, H144A	64
G1751	A1D / K8D / H35L / T115D / T117D / K131D / H144A / K147D	Ent.: A1D, K8D Con.: K147D BB: T115D, T117D Salida: K131D	CO: H35L, H144A	65
G1752	N17D / H35L / N47D / E111A / M113D / T115D / T117D / N121D / K147A / D227N	Ent.: N17D, N47D Con.: M113D BB: T115D, T117D, N121D	WC: E111A, K147A CO: H35L CV: D227N	66
G1753	K8D / H35L / N47D / E111A / M113D / T115D / T117D / N121D / K147A / D227N	Ent.: K8D, N47D Con.: M113D BB: T115D, T117D,	WC: E111A, K147A CO: H35L CV: D227N	67

		N121D		
G1754	A1D / T9D / H35L / K131D / H144A / K147D	Ent.: A1D, T9D Con.: K147D Salida: K131D	CO: H35L, H144A	68
G1755	A1D / K8D / H35L / N121D / K131D / H144A / K147D	Ent.: A1D, K8D Con.: K147D BB: N121D Salida: K131D	CO: H35L, H144A	69
G1756	A1D / K8D / H35L / N123D / K131D / H144A / K147D	Ent.: A1D, K8D Con.: K147D BB: N123D Salida: K131D	CO: H35L, H144A	70
G1757	A1D / N17D / H35L / N121D / K131D / H144A / K147D	Ent.: A1D, N17D Con.: K147D BB: N121D Salida: K131D	CO: H35L, H144A	71
G1758	A1D / N17D / H35L / N123D / K131D / H144A / K147D	Ent.: A1D, N17D Con.: K147D BB: N123D Salida: K131D	CO: H35L, H144A	72
G1814	A1D / N17D / H35L / N47D / E111A / M113D / T115D / T117D / N121D / K147A/D227N	Ent.: A1D, N17D, N47D Con: M113D BB: T115D, T117D, N121D	CO: H35L WC: E111A, K147A, M113D CV: D227N	73
G1815	N17D / H35L / N47D / M113D / T115D / T117D / N121D / K147A / D227N	Ent.: N17D, N47D Con: M113D BB: T115D, T117D, N121D	CO: H35L WC: M113D, K147A CV: D227N	74
G1821	K8D / T9D / D13N / H35G	Ent.: K8D, T9D	CO: H35G EX: D13N	75
G1823	K8D / D13N / H35G / S106D	Ent.: K8D, S106D	CO: H35G EX: D13N	76
G1824	S3D / K8D / D13N / H35G	Ent.: S3D, K8D	CO: H35G EX: D13N	77
G1825	K8D / D13N / N17D / H35G	Ent.: K8D, N17D	CO: H35G EX: D13N	78
G1827	K8D / D13N / H35G / S239D	Ent.: K8D S239D	CO: H35G EX: D13N	79
G1828	K8D / D13N / H35G / N47D	Ent.: K8D, N47D	CO: H35G EX: D13N	80
G1829	K8D / D13N / H35G / K46D	Ent.: K8D, K46D	CO: H35G EX: D13N	81
G1831	K8D / D13N / H35G / M113D	Ent.: K8D Con.: M113D	CO: H35G WC: M113D EX: D13N	82
G1832	K8D / D13N / H35G / T115D	Ent.: K8D BB: T115D	CO: H35G EX: D13N	83
G1833	K8D / D13N / H35G / T117D	Ent.: K8D BB: T117D	CO: H35G EX: D13N	84
G1834	K8D / D13N / H35G / N121D	Ent.: K8D BB: N121D	CO: H35G EX: D13N	85
G1835	K8D / D13N / H35G / T129D	Ent.: K8D BB: T129D	CO: H35G EX: D13N	86
G1836	K8D / D13N / H35G / G130D	Ent.: K8D BB: G130D	CO: H35G EX: D13N	87
G1837	K8D / D13N / H35G / K131D	Ent.: K8D Salida: K131D	CO: H35G EX: D13N	88
G1838	K8D / D13N / H35G / K288D	Ent.: K8D, K288D	CO: H35G EX: D13N	89
G1843	K8D / D13N / H35G / E111A	Ent.: K8D	CO: H35G, E111A EX: D13N	90
G1868	K8D / D13N / H35G / K46D / M113D / E111A / N121D	Ent.: K8D, K46D Con.: M113D BB: N121D	CO: H35G, E111A	91

G1910	A1D / N17D / H35L / K131D / G133D / K147D	Ent.: A1D, N17D Con.: K147D BB: G133D Salida: K131D	CO: H35L, H144A	92
G1911	A1D / N17D / H35L / K131D / L135D / K147D	Ent.: A1D, N17D Con.: K147D BB: L135D Salida: K131D	CO: H35L, H144A	93
G1914	K8D / N17D / H35L / N47D / M1 13D / T115D / T117D / N121D / K147A / D227N	Ent.: K8D, N17D, N47D Con.: M113D BB: T115D, T117D, N121D	CO: H35G, E111A WC: M113D, K147A CV: D227N	94
G1918	A1D / H35L / D127G / D128K / K131D / K147D	Ent.: A1D Con.: K147D Salida: K131D	CO: H35L, H144A LI: D127G, D128K	95
G1933	A1D/K8D / N17D / H35L / N121D / K131D / K147D	Ent.: A1D, K8D, N17D Con.: K147D BB: N121D Salida: K131D	CO: H35L	96
G1953	K8D / T9D / D13N / H35G / K46D	Ent.: K8D, T9D K46D	CO: H35G EX: D13N	97
G1954	K8D / T9D / D13N / H35G / N47D	Ent.: K8D, T9D N47D	CO: H35G EX: D13N	98
G1956	K8D / T9D / D13N / H35G /K46D/N121D	Ent.: K8D, T9D K46D BB: N121D	CO: H35G EX: D13N	99
G2027	T18D / H35L / K131D / K147D	Ent.: T18D Con.: K147D Salida: K131D	CO: H35L	100
G2080	K8N / H35L / E111A / M113D / T115D / T117D / K147A	Ent.: K8D Con.: M113D BB: T115D, T117D	CO: H35L, E111A WC: M113D, K147A	101
Clave: Ent. = entrada del poro heptamérico; Con. = sitio de constricción; BB = barril beta; Salida = salida del barril beta; CO = modificaciones que controlan la oligomerización; WC = modificaciones que ensanchan el sitio de constricción; CV = coeficiente de variación de la tasa de llegada; EX = Mejorar la expresión; pI = incrementar el punto isoeléctrico; LI = vida útil mejorada				

Se transformó un plásmido pPR-IBA2 (IBA Life Sciences, Alemania) que comprendía un polinucleótido(s) de α-HL modificado(s) como se expone anteriormente en la línea celular BL21 DE3 de *E. coli* (Thermo Fisher, Waltham, MA, EE. UU.) y se cultivaron las células transformadas para detectar la expresión de proteína de acuerdo con las instrucciones del fabricante. Se recogieron las células cultivadas por centrifugación y, a continuación, se lisaron por medio de sonificación. Se purificaron los polipéptidos que tenían la marca de epítopo escindible a partir del lisado por cromatografía en columna de afinidad (columnas PhyTip, PhyNexus, Inc., San José, CA). Se escindieron las marcas de epítopo de los polipéptidos y se separaron los monómeros de α-HL de las marcas escindidas y los polipéptidos no escindidos por medio de cromatografía en columna de afinidad (columnas PhyTip, PhyNexus, Inc., San José, CA).

VIA2. Expresión de monómeros variantes

Se cribaron algunas de las variantes de la tabla 6 para determinar el nivel de expresión por espectrofotometría. En resumen, se recogieron células de 1 ml de cultivo de expresión y se eluyó la variante expresada que tenía la marca de epítopo a un volumen de 100 µl. Se registró la absorbancia A280 del eluato y se promedió para todas las colonias de cada variante. Se usó como control una variante que tenía las sustituciones H144A, N47K y H35G en relación con la SEQ ID NO: 1 (variante G432). Los resultados se muestran en la tabla 7:

ID de variante	Prom. A280	ID de variante	Prom. A280	ID de variante	Prom. A280
G367	0,040	G1389	0,032	G1318	0,024
G616	0,166	G1390	0,034	G1319	0,039
G617	0,074	G1391*	0,403	G1589	0,030
G618	0,099	G1513	0,042	G1590	0,031
G619	0,111	G1516	0,029	G363	0,000
G620	0,064	G1514	0,037	G365	0,012
G621	0,067	G1589	0,017	G557	0,102

G622	0,120	G1590	0,012	G403	0,195
G623*	0,577	G1591	0,014	G559	0,136
G653	0,094	G432 (n.º 1)	2,273	G596	0,016
G679	0,108	G432 (n.º 2)	1,752	G597	0,015
G680	0,123	G1598	0,056	G681	0,147
G725	0,024	G1609*	1,490	G727	0,020
G1599	0,073	G1610*	0,908	G1143*	1,120
G1600	0,066	G1637*	0,883	G1147*	1,084
G1608*	0,907	G1642	0,064	G1145*	0,219
G1309	0,079	G1643	0,075	G1148	0,060
G1310	0,053	G1644*	0,557	G1149	0,107
G1314	0,026	G1645	0,080	G1150	0,072
G1313	0,049	G1647	0,022	G1308	0,089
G1315	0,038	G1651	0,176		
G1316	0,033	G728	0,014		
* Variantes en las que al menos una colonia tiene una A280 de ≥0,3 (de alta expresión)					

Tabla 7

Adicionalmente, se calculó la carga neta de un homoheptámero formado a partir de un subconjunto de estos monómeros. Se calculó la carga neta sumando la carga a pH 7,5 de todas las cadenas laterales de aminoácido localizadas en una posición orientada hacia el disolvente de un canal lleno de disolvente. Las cargas netas calculadas se representaron gráficamente frente a la A280 del monómero. Los resultados se muestran en la fig. 11.

VIA3. Formación de homoheptámeros

Se sometieron a prueba las variantes consideradas "de alta expresión" en la sección VA2 para determinar su capacidad de formar oligómeros heptaméricos. Se solubilizó el lípido difitanoilfosfatidilcolina (DPhPC) en cualquiera de Tris 50 mM, NaCl 200 mM, pH 8 hasta una concentración final de 50 mg/ml y se añadió a los monómeros de α-HL hasta una concentración lipídica final de 5 mg/ml. Se incubó la mezcla durante la noche a 37 °C. Después de esto, se añadió n-octil-β-D-glucopiranósido (βOG) a una concentración final del 5 % (peso/volumen) para solubilizar la mezcla de lípidos-proteínas resultante. Se centrifugó la muestra para depurar los agregados de proteína y los complejos lipídicos restantes y se recogió el sobrenadante. Se realizaron las preparaciones de monómero diluyendo el monómero a una concentración de proteína final de 5 mg/ml en Tris 50 mM, NaCl 50 mM, pH 8. Se fraccionaron las muestras por electroforesis en gel de poliacrilamida con dodecilsulfato de sodio (SDS-PAGE) junto con las muestras de la preparación de monómero. En la fig. 12 se ilustran imágenes ejemplares de geles. Los cuadros negros muestran la tasa de migración esperada de los heptámeros. Como se ilustra en la fig. 12, los monómeros G623 (gel inferior, columnas 1 y 2), G1143 (gel superior, columnas 1 y 2) y G1147 (gel superior, columnas 3 y 4) pudieron formar heptámeros.

Se intentó la oligomerización del monómero G1637 en condiciones alternativas.

En primer lugar, se repitió la oligomerización usando una composición de tampón de ácido 2-(N-morfolino)etanosulfónico (MES) [MES 50 mM , pH 6,0, NaCl 50 mM] o bien una composición de tampón que contenía N-óxido de trimetilamina (TMAO) 3 M [TMAO 3 M, Tris 50 mM, NaCl 50 mM, pH 8]. Se solubilizó DPhPC en cualquiera de los tampones hasta una concentración final de 50 mg/ml y se añadió al monómero de α-HL hasta una concentración lipídica final de 5 mg/ml. Se incubó la mezcla durante la noche a 37 °C. Después de esto, se añadió n-octil-β-D-glucopiranósido (βOG) a una concentración final del 5 % (peso/volumen) para solubilizar la mezcla de lípidos-proteínas resultante. Se centrifugó la muestra para depurar los agregados de proteína y los complejos lipídicos restantes y se recogió el sobrenadante. Se realizaron las preparaciones de monómero diluyendo el monómero hasta una concentración final de proteína de 5 mg/ml en cualquiera de los tampones. En la fig. 13 A se ilustran imágenes ejemplares de geles. Los cuadros negros continuos muestran la tasa de migración esperada de heptámeros, mientras que los cuadros negros discontinuos muestran la tasa de migración esperada de monómero. Como se ilustra en la fig. 13A, la oligomerización se produjo en presencia de TMAO, pero no del tampón MES.

En segundo lugar, se usaron los siguientes tampones para la oligomerización de G1637: (1) acetato de sodio 50 mM, NaCl 50 mM y TMAO 3 M, pH 4,8; (2) MES 50 mM, NaCl 50 mM y TMAO 3 M, pH 6,0; (3) fosfato de potasio 50 mM, NaCl 50 mM y TMAO 3 M, pH 7,4; y (4)Tris 50 mM, NaCl 50 mM y TMAO 3 M, pH 8,0. Se solubilizó DPhPC en los tampones hasta una concentración final de 50 mg/ml y se añadió al monómero de α-HL hasta una concentración lipídica final de 5 mg/ml. Se incubó la mezcla durante la noche a 37 °C. Después de esto, se añadió n-octil-β-D-glucopiranósido (βOG) a una concentración final del 5 % (peso/volumen) para solubilizar la mezcla de

lípidos-proteínas resultante. Se centrifugó la muestra para depurar los agregados de proteína y los complejos lipídicos restantes y se recogió el sobrenadante. Se realizaron las preparaciones de monómero diluyendo el monómero hasta una concentración final de proteína de 5 mg/ml en cualquiera de los tampones. En la fig. 13B se ilustran imágenes ejemplares de geles. Los cuadros negros continuos muestran la tasa de migración esperada de heptámeros, mientras que los cuadros negros discontinuos muestran la tasa de migración esperada de monómero. Como se ilustra en la fig. 13B, la oligomerización se produjo en las 4 composiciones de tampón, produciéndose el rendimiento de heptámeros más alto en los tampones de fosfato de potasio y Tris.

VIA4. Formación y purificación de heteroheptámeros 6:1

La capacidad de formar poros 6:1 es útil para aplicaciones de secuenciación porque permite controlar de cerca el número de polimerasas por pocillo activo. Por lo tanto, se sometieron a prueba variantes para determinar su capacidad de formar heteroheptámeros 6:1. Se mezclaron variantes con un ligando en las siguientes combinaciones:

Variante	Ligando	Variante: Ligando (p/p)
G1147	WT	4:1, 5:1, 6:1, 7:1, 8:1, 9:1
G1515	WT	1:6, 1:7, 1:5, 1:3, 1:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1
G1713	WT	9:1
G1752	WT	9:1
WT = SEQ ID NO: 1 con modificación de SpyTag en el extremo C		

Tabla 8

Se solubilizó lípido DPhPC en Tris 50 mM, NaCl 200 mM, pH 8 hasta una concentración final de 50 mg/ml y se añadió a la mezcla de monómeros de α-HL hasta una concentración lipídica final de 5 mg/ml. Se incubó la mezcla durante la noche a 37 °C. Después de esto, se añadió n-octil-β-D-glucopiranósido (βOG) a una concentración final del 5 % (peso/volumen) para solubilizar la mezcla de lípidos-proteínas resultante. Se centrifugó la muestra para depurar los agregados de proteína y los complejos lipídicos restantes y se recogió el sobrenadante. Se purificaron las muestras por cromatografía de intercambio catiónico para enriquecer la fracción oligomérica y se fraccionaron por SDS-PAGE. Se mezcló un conjugado de proteína fluorescente verde SpyCatcher (SC-GFP) con algunas alícuotas de los heptámeros resultantes para fijar la SC-GFP al SpyTag C terminal del monómero natural. La masa añadida de la SC-GFP alteró la migración de SDS-PAGE esperada, de modo que los heteroheptámeros 1:6 (WT:G1147) migraron a una tasa diferente a los heteroheptámeros 2:5 o bien los homoheptámeros. En las FIG. 14A-14G se ilustran imágenes ejemplares de geles (y en algunos casos los gráficos cromatográficos adjuntos).

La FIG. 14A ilustra la oligomerización del monómero G1147 con un monómero de α-hemolisina natural (WT). Se mezclaron los monómeros G1147 y WT-SC-GFP con G1147 en exceso en proporciones (p/p) de 4:1, 5:1, 6:1, 7:1, 8:1 y 9:1. El conjugado de SC-GFP añadió masa adicional al monómero natural, de modo que los heteroheptámeros 1:6 (WT:G1147) migraron a una tasa diferente a los heteroheptámeros 2:5 o bien los homoheptámeros. Como se indica por las flechas, se formaron heteroheptámeros 1:6 (1 monómero WT con respecto a 6 monómeros G1147), así como heteroheptámeros 2:5.

Las FIG. 14B - FIG. 14D ilustran la oligomerización del monómero G1515 con el monómero WT. La FIG. 14B ilustra la oligomerización con diversas proporciones de G1515 con respecto a WT, incluyendo ejemplos en los que G1515 está en exceso (proporciones de 3:1, 4:1, 5:1, 6:1, 7:1, 8:1 y 9:1), en los que G1515 y WT están presentes en cantidades equivalentes (proporción de 1:1) y en los que el monómero WT está en exceso (proporciones de 1:3, 1:5, 1:7 y 1:9). Como se indica por las flechas, se formaron heteroheptámeros 1:6 (1 monómero WT con respecto a 6 monómeros G1515), así como heteroheptámeros 2:5 (2 monómeros WT con respecto a 5 monómeros G1515). La FIG. 14C y la FIG. 14D muestran un desarrollo de cromatografía de intercambio catiónico (FIG. 14C) y un gel de SDS-PAGE de las fracciones resultantes (FIG. 14D) de un heptámero generado a partir de una mezcla 9:1 de G1515 con respecto a WT. Se recogieron 6 fracciones (marcadas P1-P6) (FIG. 14C). Se analizaron las fracciones por SDS-PAGE, con resultados que se muestran en la FIG. 14D.

La FIG. 14E ilustra la purificación por desarrollo de cromatografía de intercambio catiónico de un heptámero generado a partir de una mezcla 9:1 de G1713 con respecto a WT. Se recogieron 4 fracciones (marcadas P1-P4) y se analizaron las fracciones por SDS-PAGE, con resultados que se muestran en la FIG. 14E. La fracción P2 pareció tener el más alto rendimiento de oligómero 1:6 en comparación con el oligómero 2:5. La fracción P2 se purificó además posteriormente usando cromatografía de afinidad de metal inmovilizado (IMAC) y se analizó por SDS-PAGE. Los resultados se muestran en la FIG. 14F. Como se puede observar, el fraccionamiento por cromatografía de intercambio catiónico seguido de IMAC da como resultado una pureza muy alta de heptámero 1:6. Se repitió este proceso con heptámeros formados a partir de G1752 + WT (FIG. 14G), G1753 + WT (FIG. 14H, izquierda) y G1933 + WT (FIG. 14H, derecha), mostrando de forma similar cada uno de ellos una pureza muy alta de heptámero 1:6.

Adicionalmente, se han generado los siguientes poros:

Número P	Componente 1	Copias de comp. 1	Componente 2	Copias de comp. 2
P-0022	G617	7	ND	ND
P-0023	G617	6	G1030	1
P-0051	G724	6	pA018	1
P-0186	G1441	6	pA018	1
P-0192	G1429	7	ND	ND
P-0194	G1440	6	pA018	1
P-0234	G1515	7	ND	ND
P-0236	G697	7	ND	ND
P-0237	G1316	7	ND	ND
P-0238	G1317	7	ND	ND
P-0241	G1313	7	ND	ND
P-0242	G1314	7	ND	ND
P-0243	G1315	7	ND	ND
P-0262	G1646	7	ND	ND
P-0263	G1608	7	ND	ND
P-0264	G1609	7	ND	ND
P-0271	G1687	7	ND	ND
P-0272	G1701	7	ND	ND
P-0273	G1698	7	ND	ND
P-0284	G1637	7	ND	ND
P-0285	G1147	7	ND	ND
P-0290	G1713	6	pA018	1
P-303	G1752	6	pA018	1
P-0326	G1748	6	pA018	1
P-0327	G1749	6	pA018	1
P-0331	G1754	6	pA018	1
P-0332	G1756	6	pA018	1
P-0338	G1755	6	pA018	1
P-0340	G1757	6	pA018	1
P-0341	G1758	6	pA018	1
P-0360	G1918	6	pA018	1
P-0361	G1910	6	pA018	1
P-0362	G1911	6	pA018	1
P-0365	G1751	6	pA018	1
P-0372	G1933	6	pA018	1
P-0339	G1814	6	pA018	1
P-0359	G1815	6	pA018	1
P-0366	G1914	6	pA018	1
P-0371	G1753	6	pA018	1
P-0408	G1814	6	pA018b	1
P-0430	G2084	6	G2132	1

Tabla 9

Como se usa en la tabla 9, "pA018" se refiere a una α-HL natural que tiene un extremo C modificado con una construcción de SpyTag. "pA018b" se refiere a la construcción pA018 excepto que el extremo C se ha modificado para contener un residuo de asparagina no natural al que se fija la construcción de SpyTag. "G2132" se refiere a SEQ ID NO: 1 con las siguientes modificaciones: sustituciones H35L, E111A, M113D, T115D, T117D, K147A y la modificación C terminal de pA018b.

VIB. Ejemplo 2: nucleósidos oligofosfato marcados positivamente

VIB1. Marcas biotiniladas

Para identificar estructuras de marcas candidatas para su uso en el desarrollo de nucleósidos polifosfato, se sintetizaron marcas de polímero biotiniladas para su uso en experimentos de captura libre con poros de α-HL modificados. Se sintetizaron los polímeros por síntesis de péptidos en fase sólida con Fmoc usando un sintetizador prelude X (Gyros Protein Technologies AG) en resina de amida ChemMatrix® Rink (Gyros Protein Technologies AG) usando N-óxido de hexafluorofosfato de N-[(7-aza-1H-benzotriazol-1-il)(dimetilamino)-metileno]-N-metilmetanaminio (HATU) como reactivo de acoplamiento con N-metilmorfolina (NMM) como base, NMM al 20 % en N,N-dimetilformamida (DMF) como solución de desprotección.

Se generaron las siguientes marcas de polímero:

ID	Secuencia de marca	SEQ ID	ID	Secuencia de marca	SEQ ID
14	K-$(K_2P_2)_{10}$-K_4-NH_2	102	77	hex-4NPA-$(Dap'_2$-$P_2)_{13}$-K_4-NH_2	181
15	K_2-$(KP)_{20}$-K_4-NH_2	103	78	hex-4NPA-(K′-4Ap-P)$_{15}$-K_4-NH_2	182
16	K′K-$(K_2P_2)_{10}$-K_4-NH_2	104	79	hex-4NPA-$(K'_2$-Ahx$)_5$-K′$_4$-NH_2	183
17	K′K_2-$(KP)_{20}$-K_4-NH_2	105	80	hex-4NPA-$(K'_2$-$G_2)_5$-K′$_4$-NH_2	184
18	K′R-$(R_2P_2)_{10}$-R_4-NH_2	106	81	Hex-4NPA-$(K'_2$-G$)_{10}$-K′$_4$-NH_2	185
19	K′R-$(R_2P_3)_{10}$-R_4-NH_2	107	82	Hex-4NPA-$(K'_2$-A$)_{10}$-K′$_4$-NH_2	186
20	K′$(Aeg_2$-$P_2)_{10}$-K_4-NH_2	108	83	Hex-4NPA-$(K'_2$-S$)_{10}$-K′$_4$-NH_2	187
21	K′(Aeg-$P_2)_{10}$-K_4-NH_2	109	84	Hex-4NPA-$(K'_2$-F$)_{10}$-K′$_4$-NH_2	188
22	K′K-$(K_2P)_{10}$-K_4-NH_2	110	85	Hex-4NPA-$(K'_2$-Bnop$)_{10}$-K′$_4$-NH_2	189
23	K′K-$(K_2P_3)_{10}$-K_4-NH_2	111	86	Hex-4NPA-$(K'_2$-Q$)_{10}$-K′$_4$-NH_2	190
24	K′-$(K'_2P)_{10}$-K_4-NH_2	112	87	hex-4NPA-$(K'_2P)_8$-$(pnaT_2$-P$)_3$-$(K'_2$-P$)_1$-K′$_4$-NH_2	191
25	$(K'_2P_2)_{10}$-K_4-NH_2	113	88	hex-4NPA-(K′$_2$P$)_6$-$(pnaT_2$-P$)_3$-$(K'_2$-P$)_3$-K′$_4$-NH_2	192
26	K′-(K′$P_2)_{10}$-K_4-NH_2	114	89	hex-4NPA-$(K'_2P)_4$-$(pnaT_2$-P$)_3$-$(K'_2$P)5-K′$_4$-NH_2	193
27	$(K'_2$-BnoP-P$)_{10}$-K_4-NH_2	115	90	Hex-4NPA-(K′-OAhx$)_{10}$-K′$_4$-NH_2	194
28	K′-$(KGP_2)_{10}$-K_4-NH_2	116	91	Hex-4NPA-$(K'_2$-OAhx$)_{10}$-K′$_4$-NH_2	195
29	K′-$(K'P_2)_5$-K_4-NH_2	117	92	prop-PEG_{24}-Hex-K'_2-K_4NH_2	196
30	$(K'_2p_2)_4$-K_4-NH_2	118	93	prop-PEG_{24}-NHCO-PEG_{12}-Hex-K'_2-K_4-NH_2	197
31	(K'_2p_2)7-K_4-NH_2	119	94	prop-PEG_{24}-NHCO-PEG_{24}-Hex-K'_2-K_4NH_2	198
32	K′-$(K_2P_2)_4$-K_4-NH_2	120	95	Hex-4NPA-(Aeg-OAhx$)_{10}$-K′$_4$-NH_2	199
33	K′-$(K_2P_2)_7$-K_4-NH_2	121	96	Hex-4NPA-(K′-$OAhx_2)_{10}$-K′$_4$-NH_2	200
34	K′-$(KAP_2)_7$-K_4-NH_2	122	97	4NPA-PEG_{24}-$(pnaT_2$-P$)_3$-K_4-NH_2	201
35	K′-$(KFP_2)_7$-K_4-NH_2	123	98	4NPA-PEG_{24}-$(pnaT_2$-G$)_3$-K_4-NH_2	202

36	K′-(KHP$_2$)$_7$-K$_4$-NH$_2$	124	99	4NPA-PEG$_{24}$-(pnaT$_2$-K)$_3$-K$_4$-NH$_2$	203
37	K′-(KLP$_2$)$_7$-K$_4$-NH$_2$	125	100	4NPA-PEG$_{24}$-K′$_2$-K$_4$-NH	204
38	K′-(KQP$_2$)$_7$-K$_4$-NH$_2$	126	101	4NPA-(Aeg$_2$-P)$_8$-(tPNA$_2$-P)$_3$-Aeg$_2$-P-K$_4$-NH$_2$	205
39	K′-(KSP$_2$)$_7$-K$_4$-NH$_2$	127	102	4NPA-(K′$_2$-P)$_8$-((Kpeg4)$_4$-P)$_3$-K′$_2$-P-K$_4$-NH$_2$	206
40	Pra-(K$_2$P$_2$)$_7$-K$_4$-NH$_2$	128	103	4NPA-(K′$_2$P)$_8$-(G-Kpeg-G-Kpeg-P)$_3$-K′$_2$-P-K$_4$-NH$_2$	207
41	Pra-(K′$_2$P$_2$)$_7$-K$_4$-NH$_2$	129	104	4NPA-(K′$_2$-P)$_7$-(tPNA$_2$-P)$_4$-K′$_2$-P-K$_4$-NH$_2$	208
42	PyrAla-(K$_2$P$_2$)$_7$-K$_4$-NH$_2$	130	105	4NPA-(K′$_2$-P)$_8$-(tPNA$_2$-K)$_3$-K′$_2$-P-K$_4$-NH$_2$	209
43	K′-(K$_2$P$_2$)$_{10}$-K$_4$-NH$_2$	131	106	4NPA-(K′$_2$-P)$_8$-(tPNA$_2$-P)$_3$-(K′$_2$-P)$_3$-K$_4$-NH$_2$	210
44	K′-(K$_2$P$_2$)$_{13}$-K$_4$-NH$_2$	132	107	Hex-4NPA-(Aeg$_2$-G)$_6$-(tPNA$_2$-P)$_3$-(Aeg$_2$-G)$_3$-K$_4$-NH$_2$	211
45	K′-(K$_2$P$_2$)$_{16}$-K$_4$-NH$_2$	133	108	Hex-4NPA-(Aeg$_2$-G)$_7$-(tPNA$_2$-P)$_3$-(Aeg$_2$-G)$_2$-K$_4$-NH$_2$	212
46	K′-(K$_2$P$_2$)$_{10}$-K$_8$-NH$_2$	134	109	Hex-4NPA-(Aeg$_2$P)$_8$-((G-KPeg$_4$)$_2$-P)$_3$-(Aeg$_2$P)$_1$-K$_4$-NH$_2$	213
47	K′-(K$_2$P$_2$)$_{10}$-(Dap)$_4$-NH$_2$	135	110	hex-4NPA-(Aeg$_2$P)$_8$-(Kpeg-G)$_8$-(Aeg$_2$-P)$_1$-K$_4$-NH$_2$	214
48	K′-(K$_2$P$_2$)$_{10}$-R$_4$-NH$_2$	136	111	hex-4NPA-(Aeg$_2$-P)$_7$-(pnaT$_2$-P)$_4$-(Aeg$_2$-P)$_1$-K$_4$-NH$_2$	215
49	K′-(K$_2$P$_2$)$_{10}$-R$_8$-NH$_2$	137	112	hex-4NPA-(Aeg$_2$P)$_8$-(Aeg-pnaT$_3$)$_2$-Aeg-K$_4$-NH$_2$	216
50	K′-(K$_2$P$_2$)$_{10}$-(Dapa)$_4$-NH$_2$	138	113	Hex-4NPA-PEG$_{24}$-K′$_2$-K$_4$-NH$_2$	217
51	K′-(bAlaK)$_{15}$-K$_4$-NH$_2$	139	114	Hex-4NPA-(Aeg2-P)$_8$-(G-Kpeg3-G-Kpeg3-P)$_3$-Aeg$_2$-P-K$_4$-NH$_2$	218
52	K′-(bAla$_2$K)$_{10}$-K$_4$-NH$_2$	140	115	Hex-4NPA-(Aeg2-P)$_7$-(G-Kpeg3-G-Kpeg3-P)$_4$-(Aeg2-P)$_1$-K$_4$-NH$_2$	219
53	K′-(Apa)$_{15}$-K$_4$-NH$_2$	141	116	Hex-4NPA-(Aeg$_2$-P)$_8$-(Aegpeg4$_2$-P)$_3$-(Aeg$_2$-P)$_1$-K$_4$-NH$_2$	220
54	K′-(Ds)$_{10}$-K$_4$-NH$_2$	142	117	Hex-4NPA-(Aeg$_2$P)$_8$-(pnaC$_2$-P)$_3$-(Aeg$_2$-P)$_1$-K$_4$-NH$_2$	221
55	K′-(Ds-Dap)$_8$-K$_4$-NH$_2$	143	118	Hex-K-(K$_2$-P$_2$)$_{10}$-K$_4$-NH$_2$	222
56	K′-(Ts)$_8$-K$_4$-NH$_2$	144	119	Hex-K-(K$_2$-P$_3$)$_{10}$-K$_4$-NH$_2$	223
57	K′-(KA$_3$)$_{10}$-K$_4$-NH$_2$	145	120	Hex-K′-(K′$_2$-P)$_{10}$-K$_4$-NH$_2$	224
58	K′-(KG)$_{20}$-K$_4$-NH$_2$	146	121	Hex-4NPA-(K′$_2$-P)$_{10}$-K$_4$-NH$_2$	225
59	K′-(K$_2$G$_2$)$_{10}$-K$_4$-NH$_2$	147	122	Hex-4NPA-(bAla-K)$_{20}$-K$_4$-NH$_2$	226
60	K′-(K$_2$GP)$_{10}$-K$_4$-NH2	148	123	hex-K′-(AEG$_2$-P$_2$)$_{10}$-K$_4$-NH$_2$	227
61	K′-(4Ap$_2$-P$_2$)$_{10}$-K$_4$-NH$_2$	149	124	Hex-4NPA-PEG24-(Aeg2P)-K4-NH2	228
62	K′-(4Ap-P)$_{20}$-K$_4$-NH$_2$	150	125	Hex-4NPA-(Aeg2P)6-PEG12-(Aeg2P)-K4-NH2	229
63	K′-(4Ap)$_{40}$-K$_4$-NH$_2$	151	126	Hex-4NPA-(Aeg-OAhx-P)7-(pnaT2-P)4-(Aeg2-P)1-K4-NH2	230
64	Hex-4NPA-K′$_{20}$-K$_4$-NH$_2$	168	127	Hex-4NPA-(Aeg-OAhx-P)4-(Aeg2-P)3-(pnaT2-P)4-(Aeg2-P)1-K4-NH2	231

65	Hex-4NPA-(Aeg-P_2)$_{10}$-K_4-NH_2	169	128	Hex-4NPA-(Aeg2-P)3-(Aeg-OAhx-P)4-(pnaT2-P)4-(Aeg2-P)1-K4-NH2	232
66	Hex-4NPA-(K'$_2$-P)$_{10}$-K_4-NH_2	170	129	Hex-4NPA-(Aeg2-P)5-(Aeg-OAhx-P)2-(pnaT2-P)4-(Aeg2-P)1-K4-NH2	233
67	Hex-4NPA-(Aeg$_2$-P_2)$_{10}$-K4-NH2	171	130	hex-4NPA-(Aeg-G2)11-K4-NH2	234
68	4NPA-(Aeg$_2$-P)$_{10}$-K_4-NH_2	172	131	hex-4NPA-(Aeg-K-P)11-K4-NH2	235
69	pra-(K'$_2$-P_2)$_7$-K_4-NH_2	173	132	hex-4NPA-(Aeg-K-BnOP)11-K4-NH2	236
70	4NPA-(G-4Ap-P)$_{15}$-K_4-NH_2	174	133	hex-4NPA-(Aeg2P)7-(Aeg-K-P)5-K4-NH2	237
71	4NPA-(K'$_2$-P)$_{10}$-K'$_4$-NH_2	175	134	Hex-4NPA-(Aeg2P)7-(Aeg(Peg4)-Aeg-P)4-(Aeg2P)1-K4-NH2	238
72	Hex-4NPA-(K-G)$_{20}$-K_4-NH_2	176	135	Hex-4NPA-(Aeg2P)7-(Aeg-pnaT-P)4-(Aeg2P)-K4-NH2	239
73	Hex-4NPA-(K'-K-P)$_{10}$-K_4-NH_2	177	136	Hex-4NPA-(K'2P2)13-R4	240
74	Hex-4NPA-(K'$_2$P$_2$)$_{10}$-K_4-NH_2	178	137	Hex-4NPA-(AEG2-P)12-K4-NH2	241
75	4NPA-(K'$_2$-pnaT)$_{10}$-K_4-NH_2	179	138	Hex-4NPA-(Acp2-P)12-K4-NH2	242
76	hex-4NPA-(K_2-G_2)$_{13}$-K_4-NH_2	180	Vacío		

Tabla 10

Las marcas anteriores se exponen con el aminoácido aminoterminal a la izquierda y el aminoácido carboxiterminal a la derecha. El resto –NH_2 en el extremo carboxílico en cada una de las marcas anteriores indica que el aminoácido carboxiterminal es un aminoácido amidado. Se N-biotinilaron las marcas de péptido en resina usando éster de *N*-hidroxisuccinimida de biotina (biotina-NHS) / *N,N*-diisopropiletilamina (DIEA) en DMF. Se desprotegieron los péptidos completamente y se escindieron de la resina usando una solución de escisión de ácido trifluoroacético (TFA) / triisopropilsilano (TIS) / agua. Se precipitaron los péptidos con éter y se purificaron además usando cromatografía de líquidos de alto rendimiento en fase inversa (RP-HPLC) con un sistema de disolvente de TFA al 0,1 %/agua/acetonitrilo. Se determinó la cantidad de estos péptidos usando una prueba de ácido 4'-hidroxiazobenceno-2-carboxílico (HABA).

Se sometieron a prueba diversas de las estructuras de marca biotiniladas en un chip basado en α-hemolisina para determinar (a) las señales de nanoporo generadas por la marca y (b) la tasa de llegada. Se usó el n.º de ID de poro P-0234 (véase la tabla 9) como poro en dos composiciones de membrana diferentes (DOPhPC y DPhPC). La mediana de la fracción de canal abierto (fracción OC) y la media de la tasa de llegada (AR) se informan en la tabla 11 (DOPhPC) y la tabla 12 (DPhPC). Las distribuciones a nivel de marca se ilustran en la fig. 15A (DOPhPC) y la fig. 15B (DPhPC). Como se puede observar, se pueden lograr diferentes niveles de marca.

ID de marca	Fracción OC	AR	ID de marca	Fracción OC	AR
Biotina-15x(lisina)-amidada	ND	ND	22	0,106	0,62
Biotina-30x(lisina)-amidada	0,130	0,13	23	0,183	0,2
Biotina-40x(lisina)-amidada	0,028	0,87	24	0,208	5,79
14	0,116	1,42	25	0,261	1,71
15	0,123	1,23	26 + DOPhPC	0,351	0,64
16	0,136	1,04	26 + DPhPC	0,460	0,40
17	0,125	1,33	27	0,124	11,06
20	0,354	1,43	28	0,274	0,99

21	0,453	0,95			

Tabla 11

ID de marca	Fracción OC	AR	ID de marca	Fracción OC	AR
Biotina-15x(lisina)-amidada	0,70	1,83	24	0,55	0,3
Biotina-30x(lisina)-amidada	0,32	0,14	25	0,58	0,45
Biotina-40x(lisina)-amidada	0,16	0,15	26	0,71	0,3
14	0,46	0,61	27	0,41	0,47
15	0,46	0,55	28	0,76	0,3
16	0,46	0,42	29	0,71	1,96
17	0,47	0,48	30	0,69	1,70
20	0,76	0,275	31	0,58	0,41
21	0,62	2	32	0,72	2,66
22	0,38	0,24	33	0,64	0,35
23	0,49	0,18			

Tabla 12

VIB2. Nucleótidos marcados

Se generaron nucleótidos marcados que tienen la estructura general de la fig. 16A, que comprende un desoxirribonucleósido hexafosfato (marcado como "dN6P") fijado a un resto de marca cargada positivamente (marcada como "marca") por medio de un enlace químico entre la marca y el grupo fosfato terminal del dN6P. El esquema general de reacción se expone en la fig. 16B. En resumen, se proporcionó una molécula de dN6P no marcado **1601**, teniendo el dN6P no marcado la estructura de fórmula 5:

$$\mathrm{dN6P - (CH_2)_{11} - N_3}$$

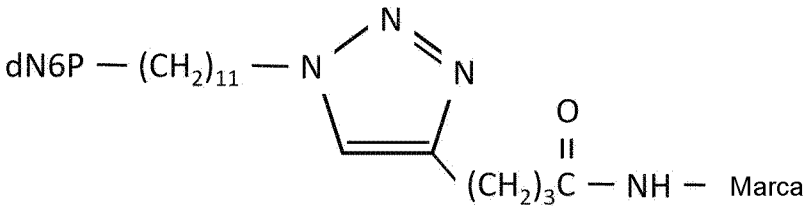
Fórmula 5

en la que "dN6P" se selecciona del grupo que consiste en desoxiadenosina hexafosfato (dA6P), desoxicitidina hexafosfato (dC6P), desoxitimidina hexafosfato (dT6P) y desoxiguanosina hexafosfato (dG6P), y en la que $(CH_2)_{11}$ es una cadena de 11 carbonos no ramificada en la que uno de los carbonos terminales se une al fosfato terminal de dN6P y el otro carbono terminal se fija al grupo acida. También se proporcionó una molécula de marca peptídica que tiene un resto 5-hexinamida fijado al aminoácido N terminal **1602**, teniendo la molécula de marca peptídica una estructura de acuerdo con la fórmula 6:

$$\mathrm{HC \equiv C - (CH_2)_3 - \overset{\overset{O}{||}}{C} - NH - marca}$$

Fórmula 6

en la que: "marca" es una estructura de marca de fórmula 2 en la que el grupo ácido carboxílico del aminoácido C terminal se convierte en un grupo amida, y en la que el resto 5-hexinamida (Hex) se localiza en la amina α del aminoácido N terminal a menos que el aminoácido N terminal sea lisina enlazada a ε (K'), caso en el que Hex se localiza en la amina ε de K'. Se hizo reaccionar la molécula de dN6P no marcado **1601** con la molécula de marca peptídica **1602** en presencia de una solución de CuBr/tris-hidroxipropiltriazolilmetilamina (THPTA) **1603** para obtener un dN6P marcado que tenía un enlace 1,2,3-triazol **1604** de acuerdo con la estructura de la fórmula 7:



Fórmula 7

en la que "dN6P" se selecciona del grupo que consiste en desoxiadenosina hexafosfato (dA6P), desoxicitidina hexafosfato (dC6P), desoxitimidina hexafosfato (dT6P) y desoxiguanosina hexafosfato (dG6P); en la que $(CH_2)_{11}$ es una cadena de hidrocarburo de 11 carbonos insaturada y no ramificada en la que uno de los carbonos terminales se une al fosfato terminal de dN6P y el otro carbono terminal se une al anillo de triazol, y marca es la estructura de marca de la fórmula 2, el "NH" que forma el enlace peptídico ilustrado se aporta por la amina α del aminoácido N terminal a menos que el aminoácido N terminal sea lisina enlazada a ε (K'), caso en el que se aporta el "NH" que forma el enlace peptídico ilustrado por la amina ε de K'. En la tabla 13 se enumeran nucleósidos hexafosfato marcados ejemplares sintetizados de acuerdo con este procedimiento.

ID	dN6P	Secuencia de marca (SEQ ID NO)	ID	dN6P	Secuencia de marca (SEQ ID NO)
P-A-00002	dA6P	K_{40}-NH_2(SEQ ID NO: 152)	P-T-00079	dT6P	4Npa-(K'$_2$-P)$_{10}$-K_4-NH_2 (SEQ ID NO: 155)
P-T-00069	dT6P	K-(K_2P_2)$_{10}$-K_4-NH_2 (SEQ ID NO: 102)	P-G-00080	dG6P	4Npa-K'$_{20}$-K_4-NH_2 (SEQ ID NO: 156)
P-T-00070	dT6P	K'-(Aeg-P_2)$_{10}$-K_4-NH_2 (SEQ ID NO: 109)	P-T-00081	dT6P	4Npa-K'$_{20}$-K_4-NH_2 (SEQ ID NO: 156)
P-T-00071	dT6P	(K'$_2P_2$)$_{10}$-K_4-NH_2 (SEQ ID NO: 113)	P-A-00082	dA6P	4Npa-K'$_{20}$-K_4-NH_2 (SEQ ID NO: 156)
P-T-00072	dT6P	K'-(K_2P)$_{10}$-K_4-NH_2 (SEQ ID NO: 153)	P-C-00083	dC6P	4Npa-K'$_{20}$-K_4-NH_2 (SEQ ID NO: 156)
P-T-00073	dT6P	K'-(K_2P_3)$_{10}$-K_4-NH_2 (SEQ ID NO: 154)	P-T-00084	dT6P	4Npa-(Aeg-P_2)$_{10}$-K_4-NH_2 (SEQ ID NO: 157)
P-C-00074	dC6P	K'-(Aeg$_2$-P_2)$_{10}$-K_4-NH_2 (SEQ ID NO: 108)	P-C-00085	dC6P	4Npa-(Aeg-P_2)$_{10}$-K_4-NH_2 (SEQ ID NO: 157)
P-T-00075	dT6P	K'-(K'$_2$P)$_{10}$-K_4-NH_2 (SEQ ID NO: 112)	P-G-00086	dG6P	4Npa-(K'$_2$-P)$_{10}$-K_4-NH_2 (SEQ ID NO: 155)
P-G-00076	dG6P	K'-(Aeg-P_2)$_{10}$-K_4-NH_2 (SEQ ID NO: 109)	P-C-00087	dC6P	4Npa-(K'$_2$-P)$_{10}$-K_4-NH_2 (SEQ ID NO: 155)
P-G-00077	dG6P	K_{40}-NH_2 (SEQ ID NO: 152)	P-G-00088	dG6P	4Npa-(K'$_2$-P)$_{10}$-K_4-NH_2 (SEQ ID NO: 155)
P-A-00078	dA6P	K'-(Aeg$_2$-P_2)$_{10}$-K_4-NH_2 (SEQ ID NO: 108)			

Tabla 13

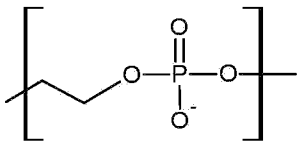
Las marcas anteriores se exponen con el aminoácido N terminal a la izquierda y el aminoácido C terminal a la derecha. El resto $–NH_2$ en el extremo carboxílico de las marcas indica que el aminoácido carboxiterminal es un aminoácido amidado.

El protocolo de síntesis para la marca P-T-00079 se expone a continuación y en la fig. 16C para ilustrar este procedimiento. Se disolvió la marca de péptido sólido Hex-4Npa-(K'$_2$-P)$_{10}$-K_4-NH_2 (SEQ ID NO: 155 con una 5-hexinamida N terminal) **1605** en agua desionizada para alcanzar una concentración de 10 mM y se ajustó el pH de la solución a 7. Se añadió una solución de NaCl, seguida de dT6P-$(CH_2)_{11}$–N_2 **1606**. Una solución recién preparada de CuBr/tris-hidroxipropiltriazolilmetilamina (THPTA) se mezcló con la solución de péptido/nucleótido **1607**. Para algunas secuencias peptídicas, se observó una solución turbia o precipitación. Se sometió a ultrasonido la solución de reacción brevemente durante aproximadamente 30 segundos y se agitó en vórtex vigorosamente

durante aproximadamente un minuto. Se colocó la solución de mezcla en un termomezclador a 25 °C mezclándose durante la noche. Se disolvió el precipitado por adición gradual (5 µl a la vez) de solución de EDTA (1 M) y se agitó en vórtex vigorosamente. Se purificó la solución por cromatografía de líquidos de alto rendimiento preparativa (prep-HPLC) usando una columna de polímero (Hamilton PS-DVB) con el siguiente gradiente: Acetonitrilo al 5-50 %, ácido trifluoroacético (TFA) al 0,085 % con agua desionizada, TFA al 0,1 % en 30 minutos. Se agruparon las fracciones recogidas y se volvieron de neutras a básicas con acetato de trietilamina (TEAA) 0,1 M usando una solución 2 M. Se ajustó el pH a 7 con trietilamina. Se concentró la solución para retirar el acetonitrilo y las sales volátiles y se liofilizó la solución acuosa restante durante la noche para dar un residuo aceitoso. Se recogió el aceite en 500 µl de TEAA (2 M) y se pasó a través de una columna de desalinización (cartucho Glen-Pak), eluyendo con TEAA 0,1 M (pH 7,5). Se recogió cada volumen de fracción (500 µl) y se analizó por un espectrómetro de masas. Se liofilizaron las fracciones para dar un rendimiento de un 20-40 % del producto sólido blanco ilustrado en **1608**.

VIB3. Extensión de molde usando marcas positivas

Se realizó una serie de experimentos para demostrar la capacidad de polimerizar las marcas cargadas positivamente que tienen dN6P. Se usaron dos estructuras de dN6P marcadas positivamente de acuerdo con la fórmula 7: (a) la fórmula 7 en la que la marca es 4Npa-K′$_{20}$–K$_4$–NH$_2$ (SEQ ID NO: 156); y (b) la fórmula 7 en la que la marca es K′–(Aeg-P$_2$)$_{10}$–K$_4$–NH$_2$ (SEQ ID NO: 109). Adicionalmente, se usó un dN6P marcado negativamente como control. El dN6P marcado negativamente tenía una estructura general de fórmula 7, excepto que en lugar de usar una marca de polímero cargado positivamente, la marca tenía la estructura (sp2)$_8$–T$_6$–(sp2)$_{16}$–C3 ("marca sp2", SEQ ID NO: 243), en la que T es desoxitimidina, C3 es un grupo propanol 3′ resultante de la síntesis de oligonucleótidos usando un soporte sólido funcionalizado con 3-(4,4'-dimetoxitritil)-1,3-propandiol o un espaciador inicial de fosforamidita C3 (3-(4,4'-dimetoxitritiloxi)propil-1-[(2-cianoetil)-(N,N-diisopropil)]-fosforamidita) cuando se usa un soporte de síntesis universal,y "sp2" es una unidad de monómero de fórmula 8:



Fórmula 8.

En algunos experimentos, solo se usó dG6P marcado y los nucleótidos restantes fueron dA6P, dC6P y dT6P no marcados. En otros experimentos, se marcó cada uno de dA6P, dT6P, dG6P y dC6P con la marca indicada.

Se realizó un ensayo de desplazamiento fluorescente usando 4 ADN polimerasas Pol6 variantes diferentes en combinación con dG6P marcado con dA6P, dC6P y dT6P no marcados. En resumen, se hibridó un molde de horquilla con un fluoróforo a un cebador con una molécula extintora. Tras la extensión del molde de horquilla por la polimerasa, se desplazó el cebador extintor y se midió la señal fluorescente. Se midió el cambio en la fluorescencia a lo largo del tiempo en tiempo real y se usó para determinar las tasas de extensión. Los gráficos de la fig. 17A muestran la intensidad de fluorescencia (eje y) frente al tiempo (eje x) para cada marca en concentraciones de 15, 10, 5 y 1 µM con las variantes de Pol6 313, 431, 867, 1369 y 1743. La intensidad máxima de fluorescencia de las marcas positivas es aproximadamente un 10 % de la intensidad máxima de fluorescencia de la marca negativa. Se repitió el ensayo de desplazamiento de fluorescencia en el que se marcaron cada uno de dA6P, dT6P, dC6P y dG6P, y se calculó la Kext de cada polimerasa con cada concentración de marca. La figura 17B ilustra la K_{ext} usando la marca negativa, la fig. 17C ilustra la K_{ext} usando la marca 4Npa-K′$_{20}$–K$_4$–NH$_2$ y la fig. 17D ilustra la K_{ext} de Pol6 1743 en diversas concentraciones de dN6P marcado. Aunque se observó una reducción en la tasa de catálisis cuando se usaron marcas cargadas positivamente, los datos, no obstante, demuestran que el dN6P marcado positivamente se puede polimerizar por una ADN polimerasa.

Se midió la tasa de extensión de polimerasa (K_{ext}) de una variante de Pol6 1743 mientras estaba fijada a cada uno de los siguientes heptámeros de α-HL:

ID de poro	Constituyentes de poro	
	Copia	Sustituciones en relación con SEQ ID NO: 1
P269	6	H35G, E111N, M113A, D127G, D128K, T129G, K131G, H144A, K147N, V149K
	1	H35G, E111N, M113A, D127G, D128G, T129G, K131G, K147N
P290	6	A1D,H35L,K131D,H144A,K147D

	1	WT
P303	6	N17D, G35L, K47D, E111A, M113D, T115D, T117D, N121D, H144A, K147A, D227N
	1	WT

Tabla 14

En cada caso, una Pol6 variante se fija al sustituyente "1" del heptámero por medio del sistema SpyCatcher/SpyTag. En cada caso, se usaron dA6P, dG6P y dC6P no marcados. En un caso, se usó dT6P no marcado (sin marca). En otro caso, se usó dT6P marcado con la marca sp2 (marca sp2). En el tercer caso, se usó dT6P marcado con K′-(Aeg-P_2)$_{10}$–K_4–NH_2 (marca positiva). La K_{ext} observada se informa en la tabla 15 (unidad = bases/s):

	P269	**P290**	**P303**
Sin marca	1,99	2,15	2,00
Marca positiva	1,19	1,37	1,24
Marca sp2	1,91	2,01	1,88

Tabla 15

Como se puede observar, la polimerización avanzó con el dN6P marcado positivamente cuando la ADN polimerasa se fijó a un poro de α-hemolisina.

VIB4. Detección por nanoporos de polimerización de dN6P marcado positivamente

Se realizó una serie de experimentos para demostrar que se pueden detectar marcas cargadas positivamente en una plataforma de secuenciación de α-hemolisina. En cada experimento, se usó el heptámero de α-HL P290 como se describe en la tabla 9 como nanoporo.

En un experimento se usaron los siguientes dN6P descritos en la tabla 13: P-C-83, P-A-82, P-G-80 y P-T-81. Cada uno de estos dN6P marcados tiene la misma marca positiva: 4Npa–K′$_{20}$–K_4–NH_2. Como se ilustra en la fig. 18, se observó un único nivel de marca como se esperaba.

En otro experimento, se usaron los siguientes dN6P descritos en la tabla 13: P-C-74, P-A-78, P-G-76 y P-T-70. En este caso, dC6P y dA6P comparten la misma marca [K′–(Aeg$_2$-P_2)$_{10}$–K_4-NH_2] (SEQ ID NO: 108), mientras que dG6P y dT6P comparten una marca diferente [K′–(Aeg-P_2)$_{10}$–K_4–NH_2] (SEQ ID NO: 109). Como se ilustra en la fig. 19, se observaron dos niveles de marca como se esperaba. Adicionalmente, la tasa de captura observada (es decir, el número de células que presentan un acontecimiento de captura) con estas marcas fue menor que la de las marcas que contenían 4Npa como aminoácido N terminal.

Estos datos demuestran que se pueden generar y detectar múltiples niveles de marca con dN6P marcado positivamente en una reacción de secuenciación por síntesis usando un poro de α-hemolisina modificado para contener múltiples aminoácidos cargados negativamente no naturales.

VIB5. Nucleótidos marcados adicionales

Se generaron nucleótidos marcados adicionales como se describe anteriormente. Las marcas con señales claras se informan a continuación en la tabla 16, con el nivel de marca observado y la carga neta.

	dN6P	**Marca**	**Nivel de marca**	**Carga**
P-A-109	dA6P	SEQ ID NO: 176	~0,2	24
P-A-112	dA6P	SEQ ID NO: 226	~0,2	24
P-A-114	dA6P	SEQ ID NO: 180	~0,2	30
P-A-122	dA6P	SEQ ID NO: 179	~0,5	24
P-A-212	dA6P	SEQ ID NO: 241	~0,6	28

P-A-213	dA6P	SEQ ID NO: 242	~0,45	28
P-A-78	dA6P	SEQ ID NO: 227	~0,65	25
P-C-102	dC6P	SEQ ID NO: 129	~0,6	18
P-C-120	dC6P	SEQ ID NO: 174	~0,55	19
P-C-131	dC6P	SEQ ID NO: 175	~0,5	24
P-C-134	dC6P	SEQ ID NO: 177	~0,45	24
P-C-135	dC6P	SEQ ID NO: 178	~0,5	24
P-C-136	dC6P	SEQ ID NO: 179	~0,55	24
P-C-138	dC6P	SEQ ID NO: 181	~0,4	30
P-C-140	dC6P	SEQ ID NO: 182	~0,5	34
P-C-144	dC6P	SEQ ID NO: 183	~0,62	14
P-C-145	dC6P	SEQ ID NO: 184	~0,65	14
P-C-146	dC6P	SEQ ID NO: 185	~0,7	24
P-C-147	dC6P	SEQ ID NO: 186	~0,62	24
P-C-148	dC6P	SEQ ID NO: 187	~0,55	24
P-C-151	dC6P	SEQ ID NO: 189	~0,43	24
P-C-152	dC6P	SEQ ID NO: 188	~0,45	24
P-C-157	dC6P	SEQ ID NO: 191	~0,05	22
P-C-158	dC6P	SEQ ID NO: 192	~0,05	22
P-C-159	dC6P	SEQ ID NO: 193	~0,42	22
P-C-160	dC6P	SEQ ID NO: 194	~0,55	14
P-C-161	dC6P	SEQ ID NO: 195	~0,62	24
P-C-162	dC6P	SEQ ID NO: 196	~0,76	6
P-C-163	dC6P	SEQ ID NO: 197	~0,72	6
P-C-164	dC6P	SEQ ID NO: 198	~0,72	6
P-C-165	dC6P	SEQ ID NO: 199	~0,8	14
P-C-166	dC6P	SEQ ID NO: 200	~0,58	14
P-C-169	dC6P	SEQ ID NO: 201	~0,75	4
P-C-170	dC6P	SEQ ID NO: 204	~0,75	6
P-C-172	dC6P	SEQ ID NO: 202	~0,72	4
P-C-193	dC6P	SEQ ID NO: 228	~0,75	6
P-C-194	dC6P	SEQ ID NO: 229	~0,55	18
P-C-196	dC6P	SEQ ID NO: 230	~0,6	13
P-C-197	dC6P	SEQ ID NO: 231	~0,6	16
P-C-198	dC6P	SEQ ID NO: 232	~0,6	16
P-C-199	dC6P	SEQ ID NO: 233	~0,55	18
P-C-220	dC6P	SEQ ID NO: 234	~0,7	26
P-C-83	dC6P	SEQ ID NO: 168	~0,55	24
P-C-85	dC6P	SEQ ID NO: 169	~0,6	14

P-C-87	dC6P	SEQ ID NO: 170	~0,5	24
P-C-94	dC6P	SEQ ID NO: 172	~0,55	24
P-C-94	dC6P	SEQ ID NO: 171	~0,55	24
P-G-115	dG6P	SEQ ID NO: 177	~0,5	24
P-G-116	dG6P	SEQ ID NO: 178	~0,4	24
P-G-185	dG6P	SEQ ID NO: 235	~0,5	26
P-G-186	dG6P	SEQ ID NO: 236	~0,4	26
P-G-195	dG6P	SEQ ID NO: 237	~0,55	28
P-G-207	dG6P	SEQ ID NO: 238	~0,6	24
P-G-208	dG6P	SEQ ID NO: 239	~0,5	24
P-T-103	dT6P	SEQ ID NO: 173	~0,45	19
P-T-123	dT6P	SEQ ID NO: 181	~0,5	30
P-T-124	dT6P	SEQ ID NO: 182	~0,4	34
P-T-130	dT6P	SEQ ID NO: 240	~0,15	30
P-T-175	dT6P	SEQ ID NO: 205	~0,15	22
P-T-178	dT6P	SEQ ID NO: 206	~0,2	22
P-T-179	dT6P	SEQ ID NO: 207	~0,15	22
P-T-180	dT6P	SEQ ID NO: 208	~0,05	20
P-T-181	dT6P	SEQ ID NO: 210	~0,08	26
P-T-183	dT6P	SEQ ID NO: 211	~0,35	22
P-T-184	dT6P	SEQ ID NO: 212	~0,3	22
P-T-189	dT6P	SEQ ID NO: 213	~0,15	22
P-T-190	dT6P	SEQ ID NO: 214	~0,55	22
P-T-192	dT6P	SEQ ID NO: 216	~0,25	23
P-T-200	dT6P	SEQ ID NO: 217	~0,7	6
P-T-209	dT6P	SEQ ID NO: 220	~0,35	22
P-T-210	dT6P	SEQ ID NO: 221	~0,15	22
P-T-69	dT6P	SEQ ID NO: 222	~0,35	25
P-T-73	dT6P	SEQ ID NO: 223	~0,45	25
P-T-75	dT6P	SEQ ID NO: 224	~0,45	25
P-T-79	dT6P	SEQ ID NO: 225	~0,45	24
P-T-84	dT6P	SEQ ID NO: 169	~0,6	14
P-T-91	dT6P	SEQ ID NO: 171	~0,6	24

Tabla 16

X. Referencias

1. Ayub *et al., Nucleobase Recognition by Truncated α-Hemolysin Pores*, ACS Nano, 2015, vol. 9, número 8, 7895–7903.

2. Bhattacharya *et al., Rectification of the Current in α-Hemolysin Pore Depends on the Cation Type: The Alkali Series Probed by Molecular Dynamics Simulations and Experiments*, The Journal of Physical Chemistry, 2011, vol.

115, número 10, pp. 4255–4264.

3. Chen y Liu, *Fabrication and Applications of Solid-State Nanopores*, 2019, Sensors, vol. 19, número 8, E1886 ("Chen I").

4. Chen *et al., Outer membrane protein G: Engineering a quiet pore for biosensing*, Proceedings of the National Academy of Sciences of the United States of America, 2008, vol. 105, número 17, pp. 6272–77 ("Chen II").

5. Chen *et al., Fusion Protein Linkers: Property, Design and Functionality*, Advanced Drug Delivery Reviews, 15 de octubre de 2013, vol. 65, número 10, pp. 1357–1369 ("Chen III")

6. Chen, *DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present*, Frontiers in Microbiology, 2014, vol. 5, artículo 305 ("Chen IV").

7. Feng *et al., Nanopore-based Fourth-generation DNA Sequencing Technology*, 2015, Genomics, Proteomics & Bioinformatics, vol. 13, número 1, pp. 4–16.

8. Gari *et al., Quiet Outer Membrane Protein G (OmpG) Nanopore for Biosensing*, ACS Sensors, 14 de abril de 2019, vol. 4, pp. 1230–35.

9. Goodman *et al., Foldamers as versatile frameworks for the design and evolution of function*, 2007, vol. 3, pp. 252–62.

10. Hammerstein *et al., Subunit dimers of α-hemolysin expand the engineering toolbox for protein Nanopores*, Journal of Biological Chemistry, vol. 286, número 16, pp.14324–34.

11. Kumar *et al., PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis*, 2012, Scientific Reports, vol. 2, art. 684.

12. Lee *et al., Recent progress in solid-state nanopores*, 2018, Advanced Materials, vol. 30, número 42.

13. Licini *et al., Oligopeptide Foldamers: From Structure to Function*, European Journal of Organic Chemistry, 2005, vol. 2005, número 6, pp. 969–77.

14. Manrao *et al., Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase*, 2012, Nature Biotechnology, vol. 30, pp. 349–53.

15. Martinek y Fueloep, *Peptidic foldamers: ramping up diversity*, 2012, Chemical Society Reviews, vol. 41, número 2, pp. 687–702.

16. Noskov *et al.*, *Ion Permeation through the α-Hemolysin Channel: Theoretical Studies Based on Brownian Dynamics and Poisson-Nernst-Plank Electrodiffusion Theory*, Biophysical Journal, 2004, vol. 87, número 4, pp. 2299–2309.

17. Pavlenok y Niederweis, *Hetero-oligomeric MspA pores in* Mycobacterium smegmatis, 2016, FEMS Microbiology Letters, Vol. 363, número 7, fnw046.

18. Song *et al., Structure of Staphylococcal α-Hemolysin, a Heptameric Transmembrane Pore*, Science, 1996, vol. 274, n.º 5294, pp. 1859–65.

19. Stoddart *et al., Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore*, Proceedings of the National Academy of Sciences of the United States of America, 2009, vol. 106, número 19, pp. 7702–7707.

20. Szuromi, *Synthesizing graphene nanopores*, 2018, Science, vol. 360, número 6385, pp. 166–68.

21. Wang *et al., The evolution of nanopore sequencing*, 2015, Frontiers in Genetics, vol. 5, art. 449 ("Wang I").

22. Wang *et al., Engineering of Protein Nanopores for Sequencing, Chemical or Protein Sensing and Disease Diagnosis*, 2018, Current Opinions in Biotechnology, vol. 51, pp. 80-89 ("Wang II").

23. *Wasfi et al., Graphene-based nanopore approaches for DNA sequencing: A literature review*, 2018, Biosensors and Bioelectronics, vol. 119, pp. 191-203.

REIVINDICACIONES

1. Un nucleósido-5'-oligofosfato enlazado covalentemente a una marca de polímero, en el que dicha marca de polímero comprende un segmento cargado positivamente (SCP), en el que dicho SCP es un homopolímero o un heteropolímero que comprende de 5 unidades de monómero a 100 unidades de monómero, que tiene una carga neta positiva de al menos +5 a pH 7,0, y que tiene una densidad de carga de al menos 0,1, con la condición de que dicho SCP no sea un homopolímero de arginina o lisina enlazada a α, en el que el SCP comprende una estructura de acuerdo con la fórmula 1:

$(REPETICIÓN)_a$ Fórmula 1,

en la que:

"REPETICIÓN" es una secuencia heteropolimérica de al menos 2 unidades de monómero, en la que REPETICIÓN tiene una densidad de carga de al menos 0,1; y

"a" es un número entero ≥ 2, con la condición de que dicho número entero sea suficiente para dar como resultado un SCP con una carga neta de al menos +5 a pH 7,0 y una longitud global de 5 a 100 unidades de monómero.

2. El nucleósido-5'-oligofosfato de la reivindicación 1, en el que los aminoácidos o análogos de aminoácido se seleccionan del grupo que consiste en lisina (K), ácido diaminopropiónico (Dap), lisina enlazada a ε (K'), ácido diaminopropiónico sustituido con aminoetilo (Dapa), prolina (P), propargilglicina (Pra), arginina (R), 4-benciloxiprolina (BnoP), glicina (G), ácido aminoetil-piperacinacético (Apa), alanina (A), pirrolil-alanina (PyrAla), fenilalanina (F), ácido trietilentriamino-succinámico (TETA), histidina (H), ácido dietilentriamino-succinámico (DETA), leucina (L), 4-aminoprolina (4Ap), glutamina (Q), aminoetilglicina (Aeg), serina (S), β-alanina (bAla), 4-nitrofenilalanina (4Npa), ácido aminohexanoico (Ahx), ácido 2-aminobenzoico (2Aba), ácido 3-aminobenzoico (3Aba), ácido 4-aminobenzoico (4Aba), ácido oxoaminohexanoico (OAhx), ácido peptidonucleico (timina) (pnaT), ácido gamma-diaminopropiónico (Dap'), ácido peptidonucleico (citosina) (pnaC), glutamato (Glu), 4-hidroxoprolina (4-OHP), ornitina (Orn), citrulina (Cit), lisina-Peg4 (KPeg4), lisina-Peg3 (KPeg3), lisina-Peg2 (KPeg2) y aminoetilglicina-Peg4 (Aegpeg4).

3. El nucleósido-5'-oligofosfato de la reivindicación 2, en el que el SCP comprende además uno o más restos hexinilo (Hex) o etilenglicol (PEG).

4. El nucleósido-5'-oligofosfato de cualquiera de las reivindicaciones 1-3, en el que la marca de polímero tiene una estructura de fórmula 2:

U – SCP –V Fórmula 2,

en la que:

U es opcional y, cuando está presente, comprende de 1 a 10 unidades de monómero;

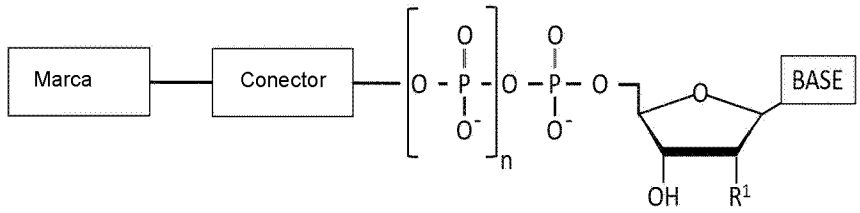
V es opcional y, cuando está presente, comprende de 1 a 10 unidades de monómero; y en el que dicha estructura de fórmula 2 tiene una carga neta positiva de al menos +5 a pH 7,0 y una densidad de carga de al menos 0,1.

5. El nucleósido-5'-oligofosfato de cualquiera de la reivindicación 4, en el que U y V tienen una carga neta neutra o una neta positiva.

6. El nucleósido-5'-oligofosfato de la reivindicación 5, en el que U comprende uno o más aminoácidos o análogos de aminoácido seleccionados del grupo que consiste en lisina, un derivado de lisina, arginina, ornitina, un aminoácido alifático, un derivado de un aminoácido alifático, un aminoácido aromático, un derivado de un aminoácido aromático, prolina, un derivado de prolina, Dap y Dapa.

7. El nucleósido-5'-oligofosfato de la reivindicación 6, en el que U tiene de 1 a 3 aminoácidos o análogos de aminoácido de longitud, en el que dichos aminoácidos o análogos de aminoácido se seleccionan del grupo que consiste en K, K', 4Npa, Pra y PyrAla.

8. El nucleósido-5'-oligofosfato de cualquiera de las reivindicaciones 1-7, que tiene una estructura de acuerdo con la fórmula 3:



Fórmula 3

en la que la base se selecciona de adenina, citosina, guanina, timina y uracilo; R^1 se selecciona de H y OH; n es de 2 a 12; el conector es un conector que comprende una cadena enlazada covalentemente de 2 a 100 átomos; y la marca es la marca de polímero.

9. El nucleósido-5'-oligofosfato de la reivindicación 8, en el que el conector comprende un grupo químico seleccionado del grupo que consiste en: éster, éter, tioéter, amina, amida, imida, carbonato, carbamato, escuarato, tiazol, tiazolidina, hidrazona, oxima, triazol, dihidropiridacina, fosfodiéster, polietilenglicol (PEG), aducto de Pictet-Spengler y cualquier combinación de los mismos.

10. El nucleósido-5'-oligofosfato de la reivindicación 9, en el que el conector tiene una estructura de fórmula 4:

$R^4 – R^2 – R^3$ Fórmula 4

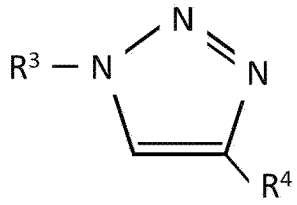
en la que:

R^2 se selecciona del grupo que consiste en enlace éster, enlace éter, enlace tioéter, enlace amina, enlace amida, enlace imida, enlace carbonato, enlace carbamato, enlace escuarato, enlace tiazol, enlace tiazolidina, enlace hidrazona, enlace oxima, enlace triazol, enlace dihidropiridacina, enlace fosfodiéster, enlace polietilenglicol (PEG), aducto de Pictet-Spengler y cualquier combinación de los mismos;

R^3 comprende una cadena de carbonos saturada o insaturada, ramificada o no ramificada, sustituida o no sustituida de al menos 2 carbonos de longitud enlazada covalentemente en un extremo a uno de los restos fosfato del oligofosfato nucleotídico y en el otro extremo a R^2; y

R^4 comprende una cadena de carbonos saturada o insaturada, ramificada o no ramificada, sustituida o no sustituida de al menos 2 carbonos de longitud enlazada covalentemente en un extremo a R^2 y en el otro extremo a la marca de polímero.

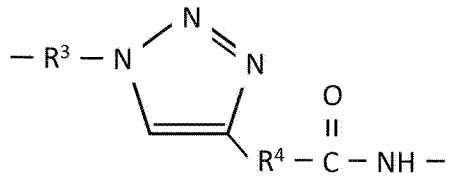
11. El nucleósido-5'-oligofosfato de la reivindicación 10, en el que R^2 es un triazol de fórmula 4a:



Fórmula 4a

en la que la suma de los carbonos en la cadena de carbonos R_3 y la cadena de carbonos R_4 es ≤ 96.

12. El nucleósido-5'-oligofosfato de la reivindicación 10, en el que R^2 es un triazol de fórmula 4b:



Fórmula 4b

y en la que el –NH– del enlace peptídico conectado a R^4 se aporta por el aminoácido o el análogo de aminoácido N

terminal de la marca de polímero.

13. Un kit que comprende:

un primer nucleósido-5'-oligofosfato de acuerdo con cualquiera de las reivindicaciones 1-12, en el que la base es adenosina,

un segundo nucleósido-5'-oligofosfato de acuerdo con cualquiera de las reivindicaciones 1-12, en el que la base es citosina,

un tercer nucleósido-5'-oligofosfato de acuerdo con cualquiera de las reivindicaciones 1-12, en el que la base es guanina, y

un cuarto nucleósido-5'-oligofosfato de acuerdo con cualquiera de las reivindicaciones 1-12, en el que la base es timina o uracilo;

en el que la marca de polímero de cada uno de los primer a cuarto nucleósidos-5'-oligofosfato es diferente.

14. Un sistema para realizar secuenciación por síntesis (SBS) de ácidos nucleicos, comprendiendo el sistema:

(a) un chip que comprende una pluralidad de electrodos sensores;

(b) una barrera electroquímicamente resistiva dispuesta sobre una superficie del chip, en el que la barrera tiene un lado *cis* y un lado *trans*;

(c) una primera solución electrolítica en el lado *cis* de la barrera;

(d) una segunda solución electrolítica en el lado *trans* de la barrera;

(e) una pluralidad de nanoporos que comprenden un canal que tiene suficientes restos cargados negativamente para repeler sustancialmente un ácido nucleico molde y/o cebador, en el que los nanoporos se disponen en la barrera de modo que el canal permite el intercambio de iones entre la primera solución electrolítica y la segunda solución electrolítica, y en el que al menos una porción de los nanoporos están lo suficientemente cerca de uno de los electrodos sensores para que el electrodo sensor pueda detectar al menos una característica de una corriente eléctrica que fluye a través del canal del nanoporo;

(f) un sistema informático en comunicación electrónica con los electrodos sensores, en el que el sistema informático se adapta para registrar la característica de la corriente eléctrica que fluye a través del nanoporo que se detecta por el electrodo sensor;

(g) una ácido nucleico polimerasa asociada con el nanoporo en el lado *cis* de la barrera, en el que la ácido nucleico polimerasa puede catalizar una reacción de amplificación de ácido nucleico dependiente de molde en la primera solución electrolítica; y

(h) un conjunto de nucleósidos-5'-oligofosfato dispuestos en la primera solución electrolítica, incluyendo el conjunto al menos un nucleósido-5'-oligofosfato de adenosina marcado con polímero, un nucleósido-5'-oligofosfato de guanina marcado con polímero, un nucleósido-5'-oligofosfato de citosina marcado con polímero y un nucleósido-5'-oligofosfato de timidina marcado con polímero o un nucleósido-5'-oligofosfato de uracilo marcado con polímero, en el que cada uno de los nucleósidos-5'-oligofosfato marcados con polímero es el nucleósido-5'-oligofosfato de acuerdo con cualquiera de las reivindicaciones 1-12.

15. Un procedimiento de secuenciación por síntesis (SBS) de secuenciación de un ácido nucleico molde, comprendiendo el procedimiento:

proporcionar un sistema de acuerdo con la reivindicación 14 que tiene una pluralidad de complejos de secuenciación de nanoporos activos, comprendiendo cada complejo de secuenciación de nanoporos activo:

- al menos uno de los electrodos sensores;
- uno de los nanoporos insertados en la barrera en proximidad al electrodo sensor, en el que una corriente fluye a través del nanoporo y una característica de la corriente se detecta por el electrodo sensor;
- la ácido nucleico polimerasa asociada con el nanoporo; y
- el ácido nucleico molde complejado con la ácido nucleico polimerasa;

en los complejos de secuenciación de nanoporos activos, incorporar los nucleósidos-5'-oligofosfato marcados en un ácido nucleico complementario del ácido nucleico molde por una reacción de amplificación de ácido nucleico dependiente de molde catalizada por la ácido nucleico polimerasa, en el que la marca de polímero del nucleósido-5'-oligofosfato marcado se mueve hacia o en proximidad al canal del nanoporo a medida que el nucleósido-5'-oligofosfato marcado se incorpora al ácido nucleico complementario, y en el que el movimiento de la marca de polímero hacia o en proximidad al canal cambia la característica de la corriente que fluye a través del nanoporo;

detectar el cambio en la característica de la corriente que fluye a través del nanoporo provocado por las marcas de polímero con el electrodo sensor y registrar el cambio en el sistema informático; y

correlacionar cada cambio registrado con uno de los nucleósidos-5'-oligofosfato marcados, generando de este modo una secuencia del ácido nucleico complementario generado en ese electrodo.

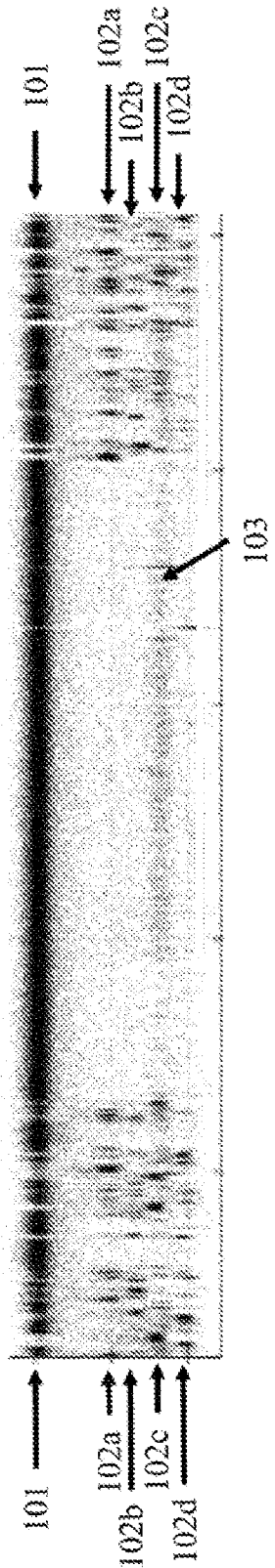


Fig. 1

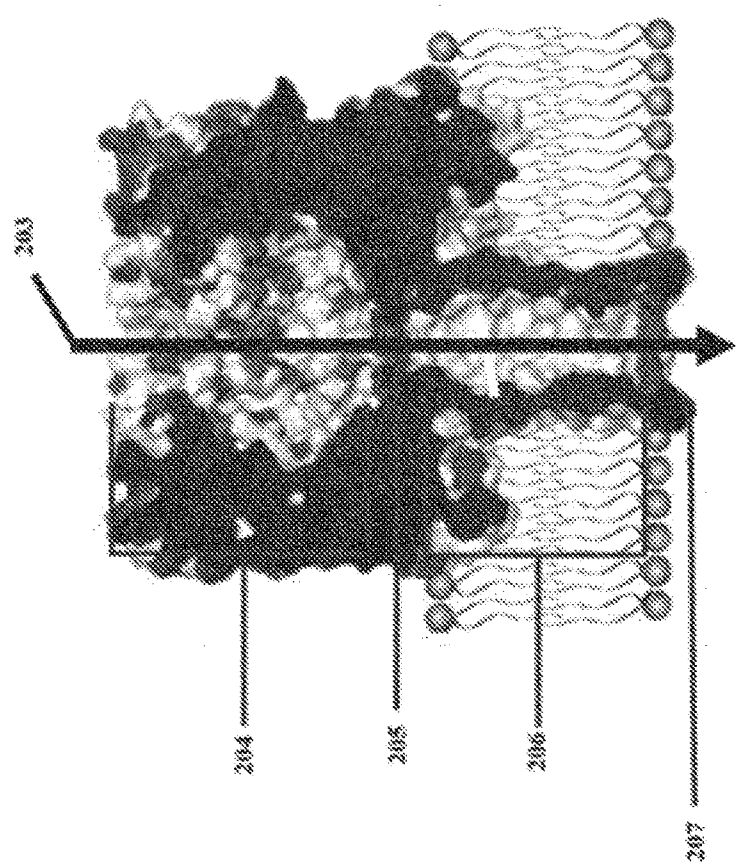


Fig. 2B

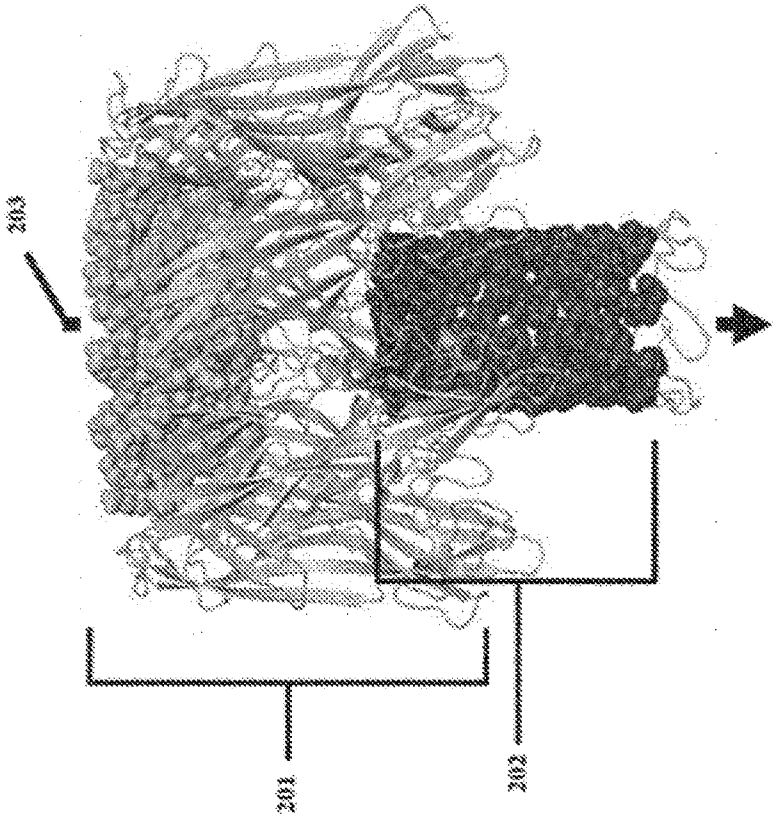


Fig. 2A

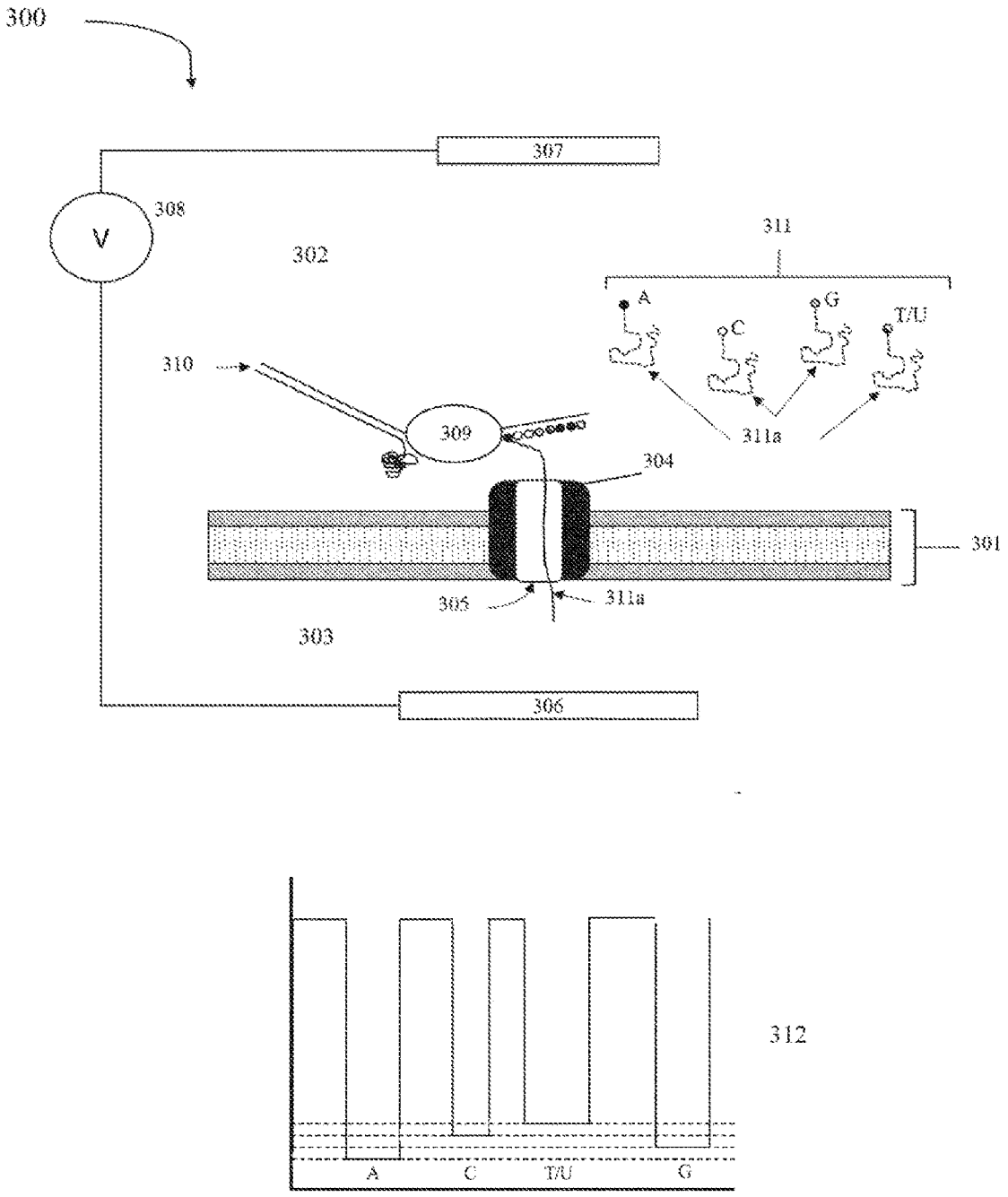
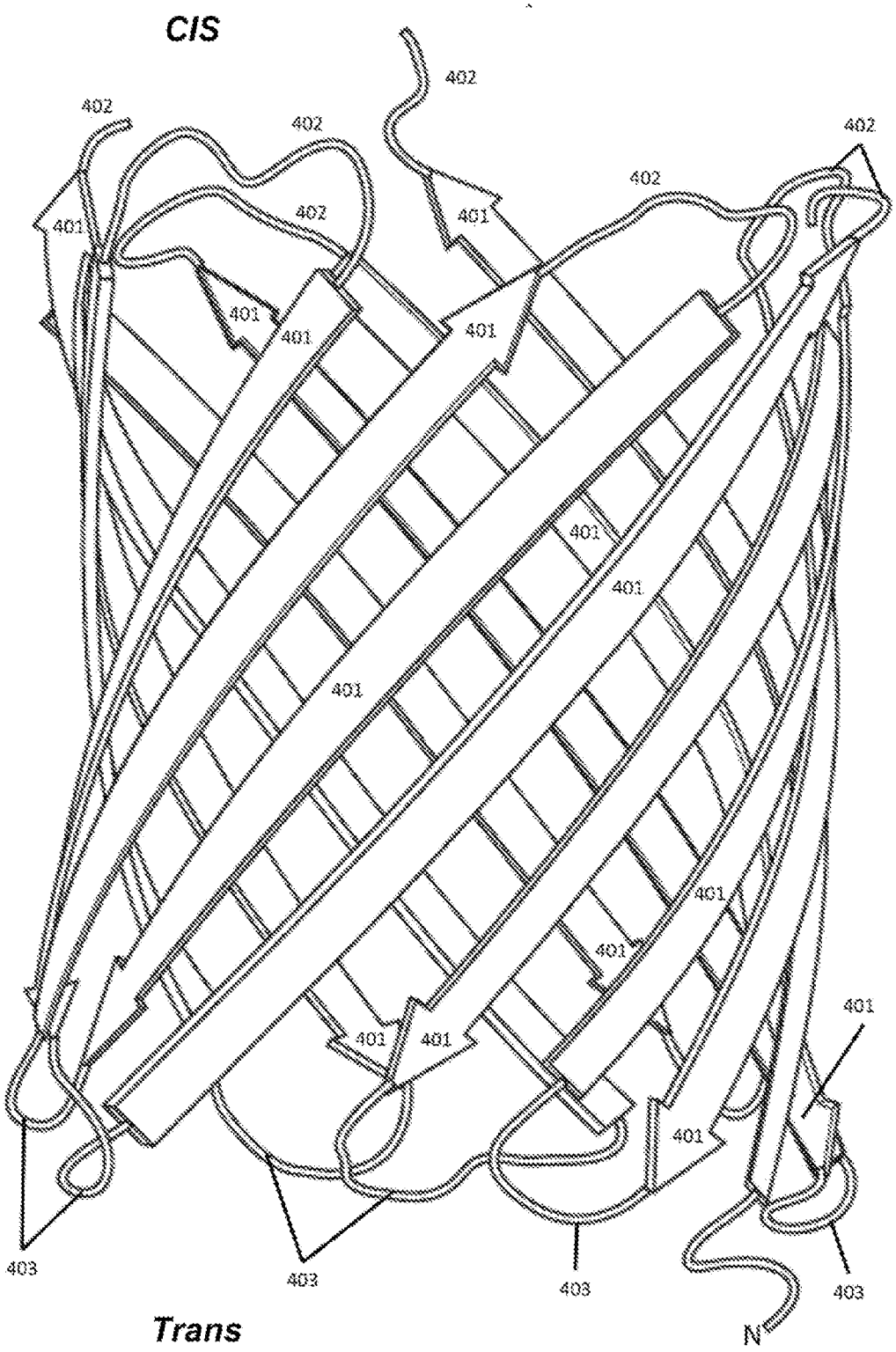
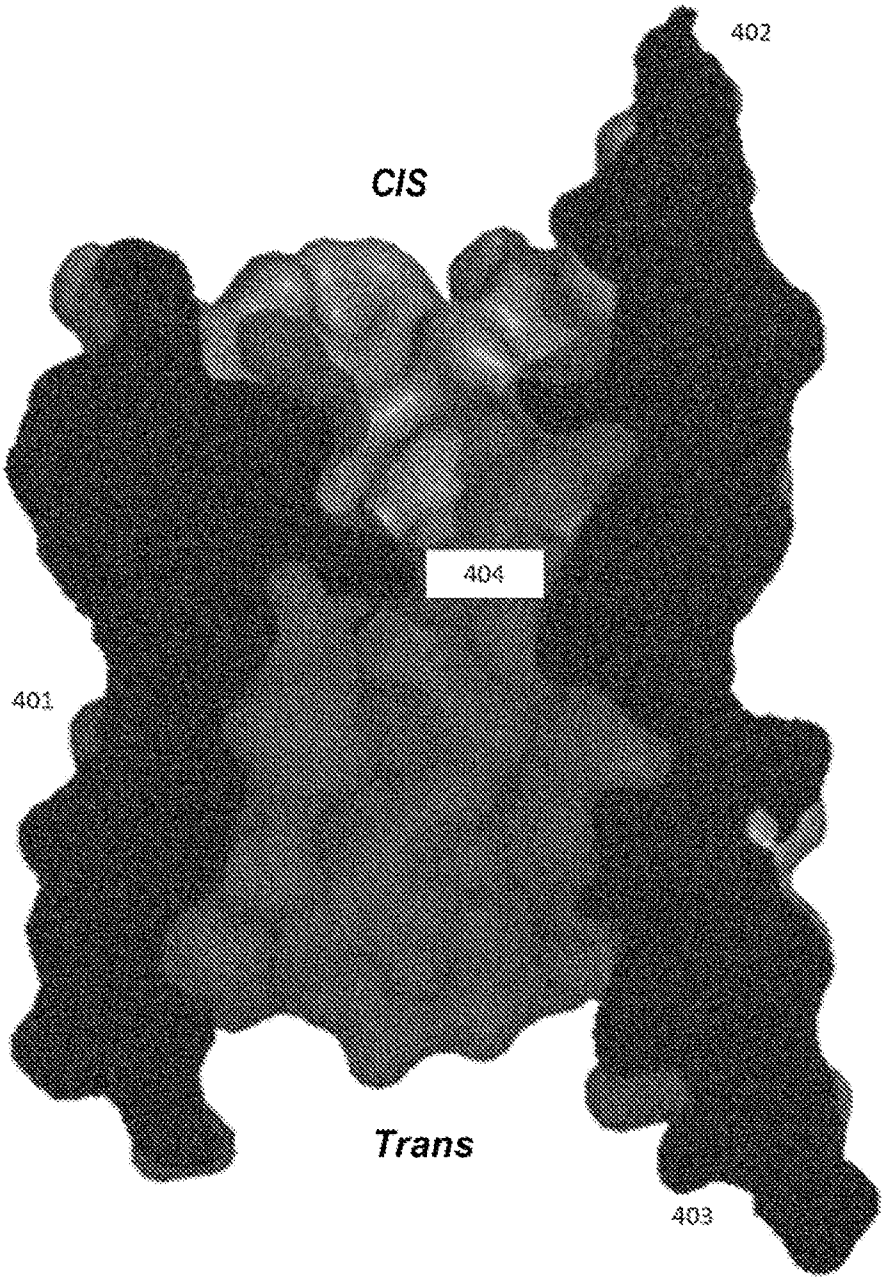

300
307
308
V
302
311
A
C
G
T/U
310
309
311a
304
301
305
311a
303
306
312
A
C
T/U
G

Fig. 3



400

Fig. 4A



400

Fig. 4B

SEQ ID NO: 23

Fig. 5A

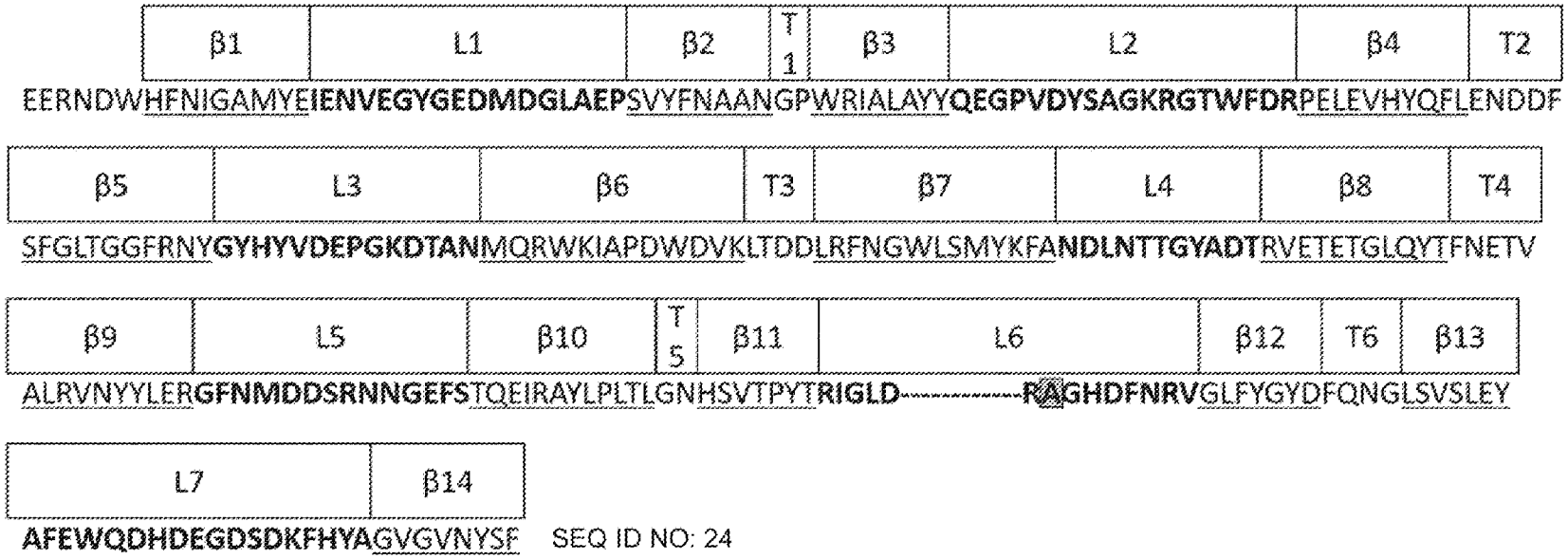

β1
L1
β2
T1
β3
L2
β4
T2
EERNDWHFNIGAMYEIENVEGYGEDMDGLAEPSVYFNAANGPWRIALAYYQEGPVDYSAGKRGTWFDRPELEVHYQFLENDDF
β5
L3
β6
T3
β7
L4
β8
T4
SFGLTGGFRNYGYHYVDEPGKDTANMQRWKIAPDWDVKLTDDLRFNGWLSMYKFANDLNTTGYADTRVETETGLQYTFNETV
β9
L5
β10
T5
β11
L6
β12
T6
β13
ALRVNYYLERGFNMDDSRNNGEFSTQEIRAYLPLTLGNHSVTPYTRIGLD----------RAGHDFNRVGLFYGYDFQNGLSVSLEY
L7
β14
AFEWQDHDEGDSDKFHYAGVGVNYSF
SEQ ID NO: 24

Fig. 5B

β1	L1	β2	T1	β3	L2	β4	T2

EERNDWHFNIGAMYEIENVEGYGEDMDGLAEPSVYFNAANGPWRIALAYYQEGPVDYSAGKRGTWFDRPELEVHYQFLENDDF

β5	L3	β6	T3	β7	L4	β8	T4

SFGLTGGFRNYGYHYVDEPGKDTANMQRWKIAPDWDVKLTDDLRFNGWLSMYKFANDLNTTGVADTRVETETGLQYTFNETV

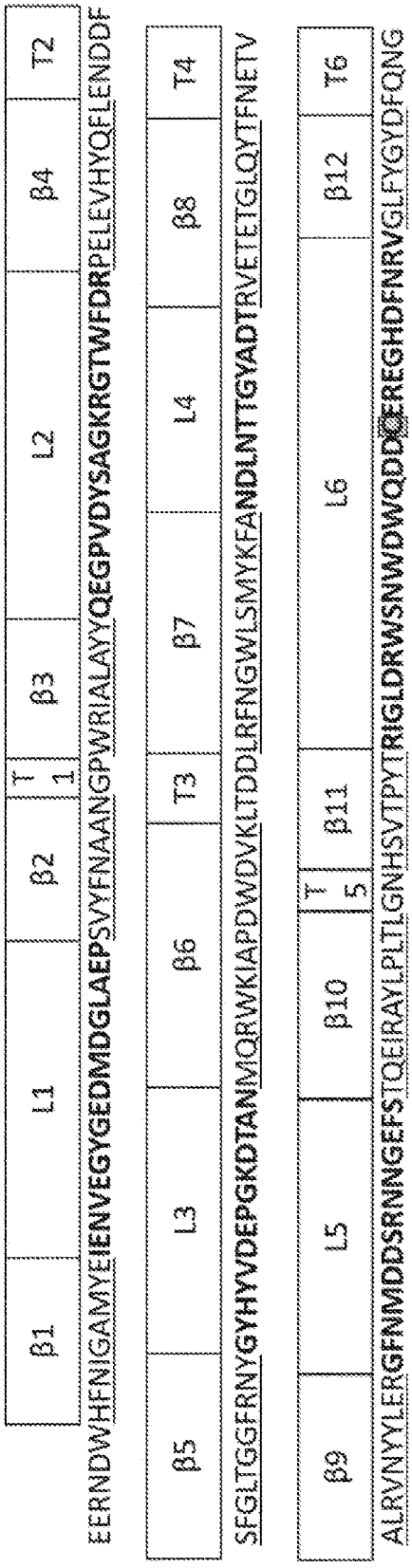
β9	L5	β10	T5	β11	L6	β12	T6

ALRVNYYLERGFNMDDSRNNGEFSTQEIRAYLPLTLGNHSVTPYTRIGL-RWSNWDWQDDIERE□HDFNRVGLFYGYDFQNG

β13	L7	β14

LSVSLEYAFEWQDH□EGDSDKFHYAGVGVNYSF SEQ ID NO: 25

Fig. 5C



SEQ ID NO: 26

Fig. 5D

 SEQ ID NO: 27

Fig. 5E

β1
L1
β2
T1
β3
L2
β4
T2
EERNDWHFNIGAMYEIENVEGYGEDMDGLAEPSVYFNAANGPWRIALAYYQEGPVDYSAGKRGTWFDRPELEVHYQFLENDDF
β5
L3
β6
T3
β7
L4
β8
T4
SFGLTGGFRNYGYHYVDEPGKDTANMQRWKIAPDWDVKLTDDLRFNGWLSMYKFANDLNTTGYADTRVETETGLQYTFNETV
β9
L5
β10
T5
β11
L6
β12
T6
ALRVNYYLERGFNMDDSRNWGEFSTQEIRAYLPLTLGNHSVTPYTRIGL-RWSNWDWQDDCEREGHDFNRVGLFYGYDFQNG
β13
L7
β14
LSVSLEYAFEWQDHDEGDSDKFHYAGVGVNYSF
SEQ ID NO: 28

Fig. 5F

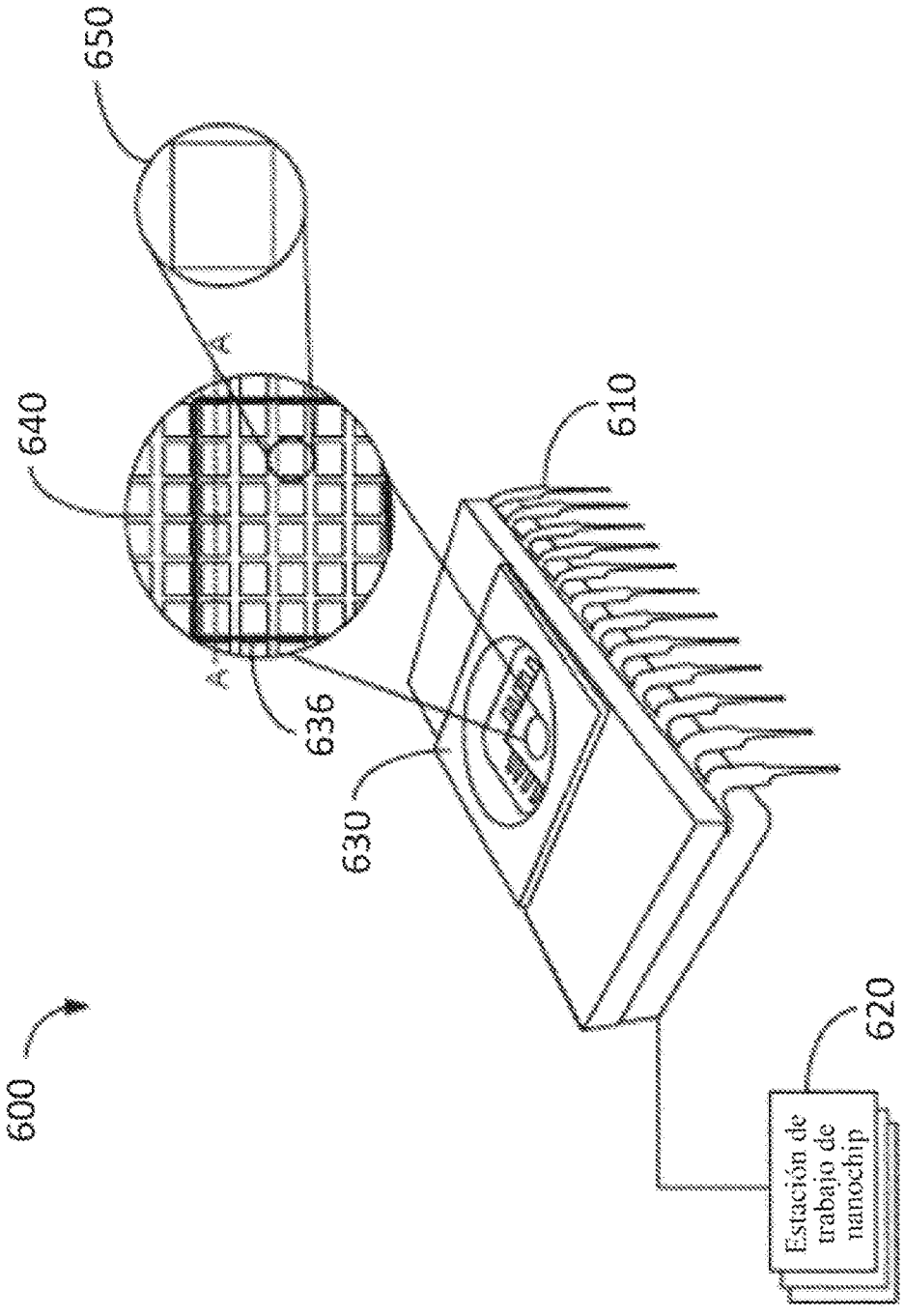


Fig. 6

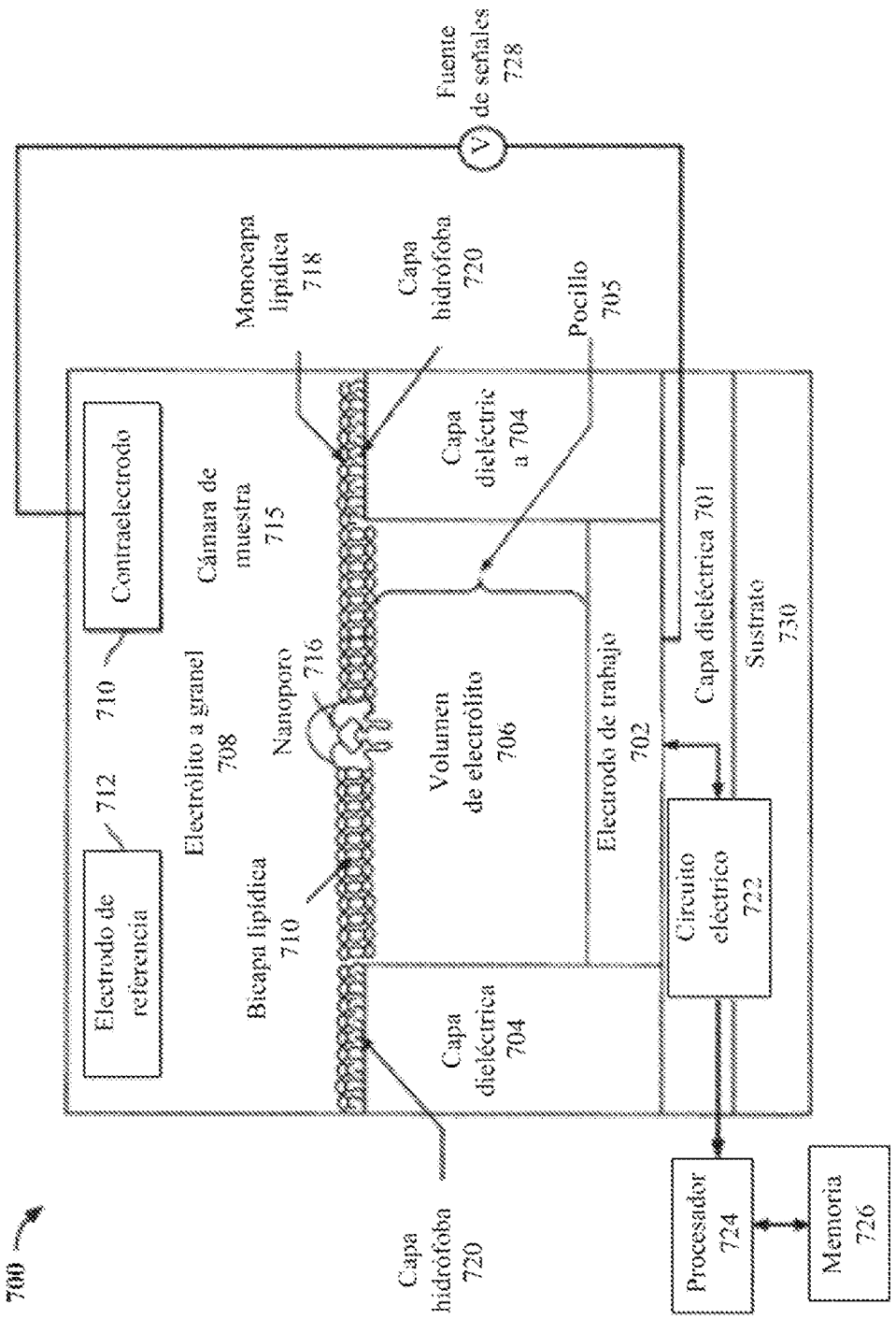

700
Fuente de señales 728
V
Monocapa lipídica 718
Capa hidrófoba 720
Pocillo 705
Contraelectrodo
Cámara de muestra 715
Capa dieléctrica 704
Capa dieléctrica 701
Sustrato 730
Electrodo de referencia
712
710
Electrolito a granel 708
Nanoporo 716
Volumen de electrolito 706
Electrodo de trabajo 702
Bicapa lipídica 710
Circuito eléctrico 722
Capa dieléctrica 704
Capa hidrófoba 720
Procesador 724
Memoria 726

Fig. 7

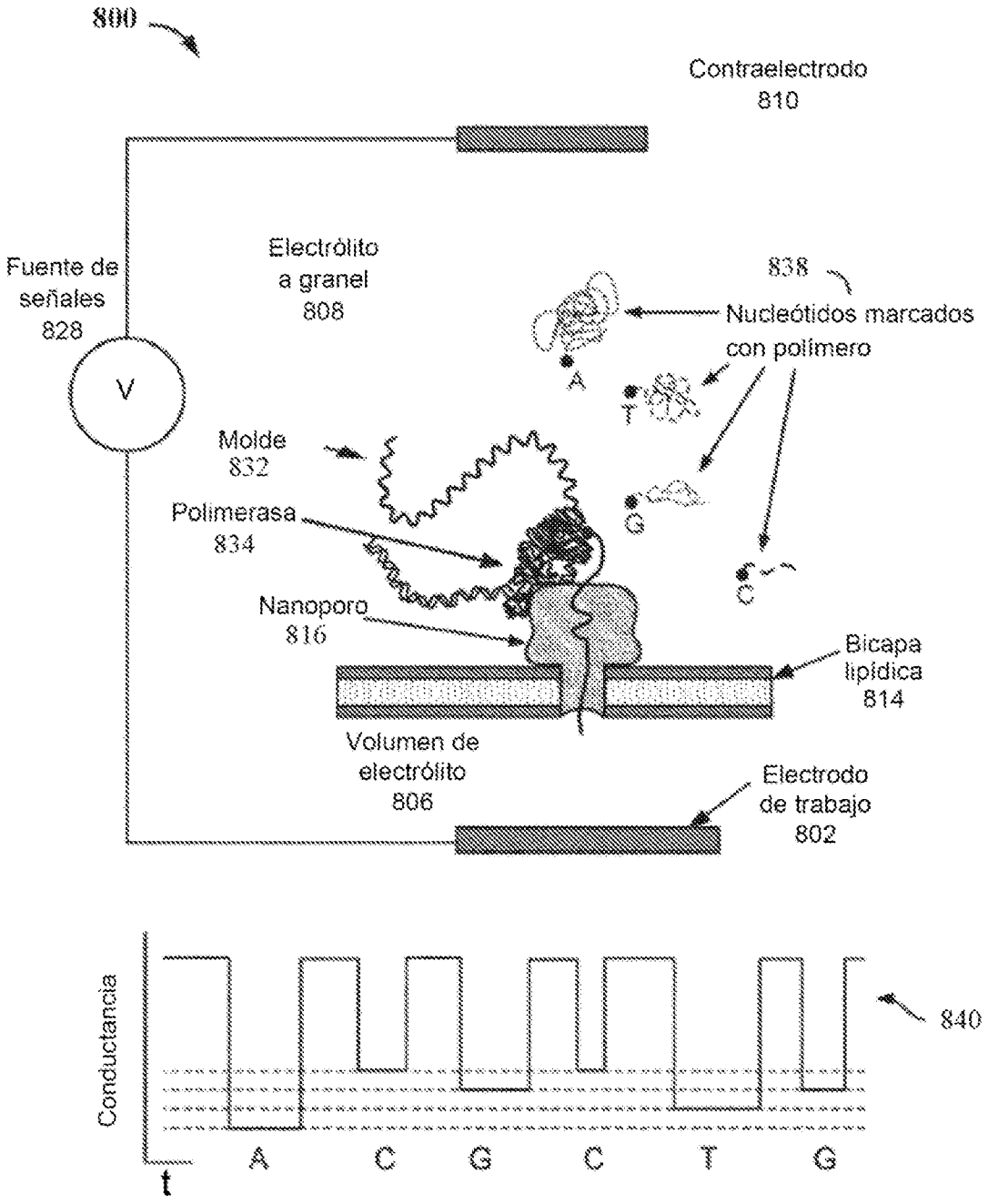

800
Contraelectrodo
810
Fuente de
señales
828
V
Electrólito
a granel
808
838
Nucleótidos marcados
con polímero
A
T
G
C
Molde
832
Polimerasa
834
Nanoporo
816
Bicapa
lipídica
814
Volumen de
electrólito
806
Electrodo
de trabajo
802
Conductancia
840
t
A
C
G
C
T
G

Fig. 8

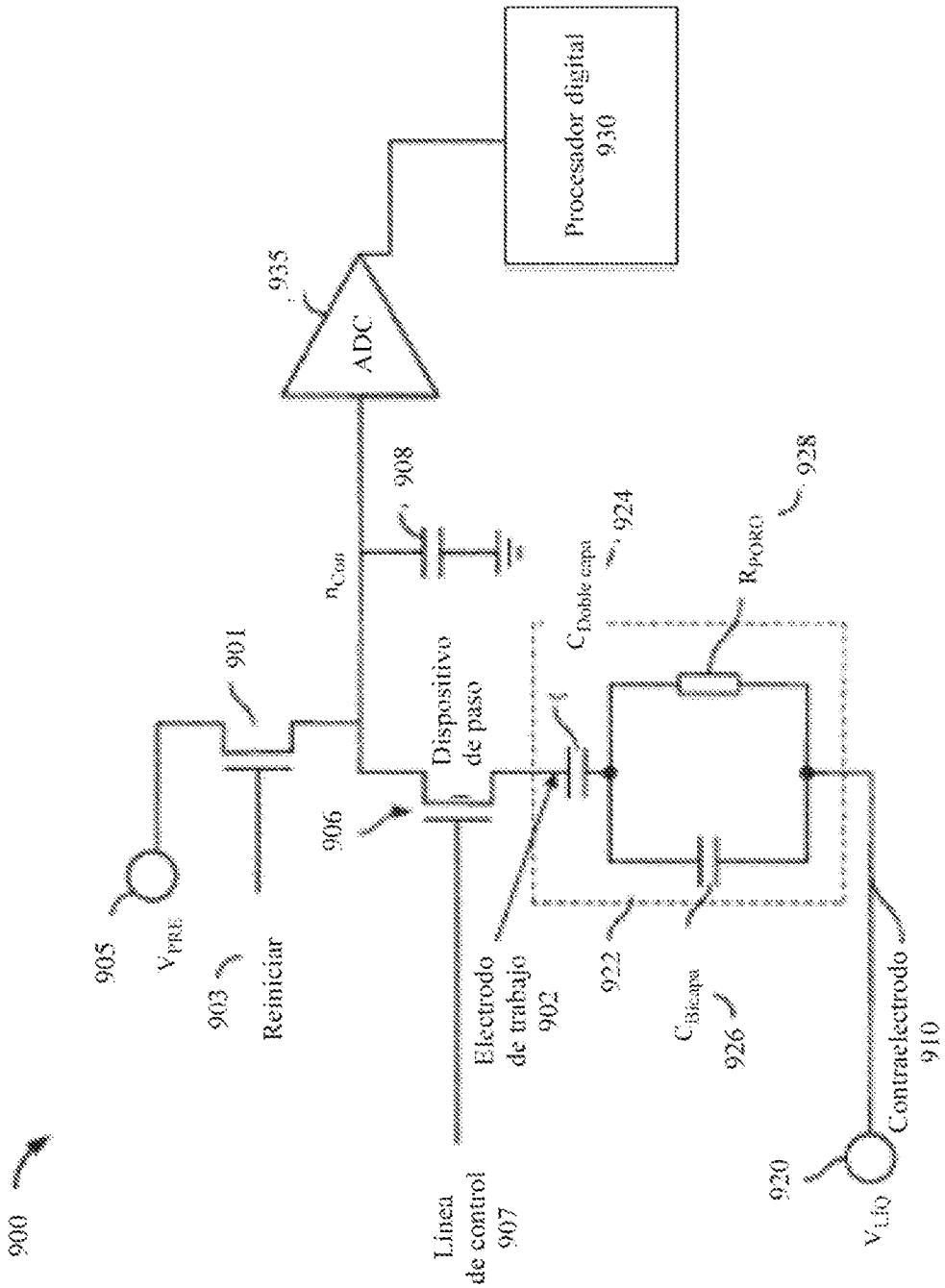

Procesador digital
930
935
ADC
908
n_{Con}
924
$C_{Doble\ capa}$
928
R_{PORO}
901
Dispositivo
de paso
906
905
V_{PRE}
903
Reiniciar
Electrodo
de trabajo
902
922
C_{Bicapa}
926
Línea
de control
907
920
V_{LIQ}
Contraelectrodo
910
900

Fig. 9

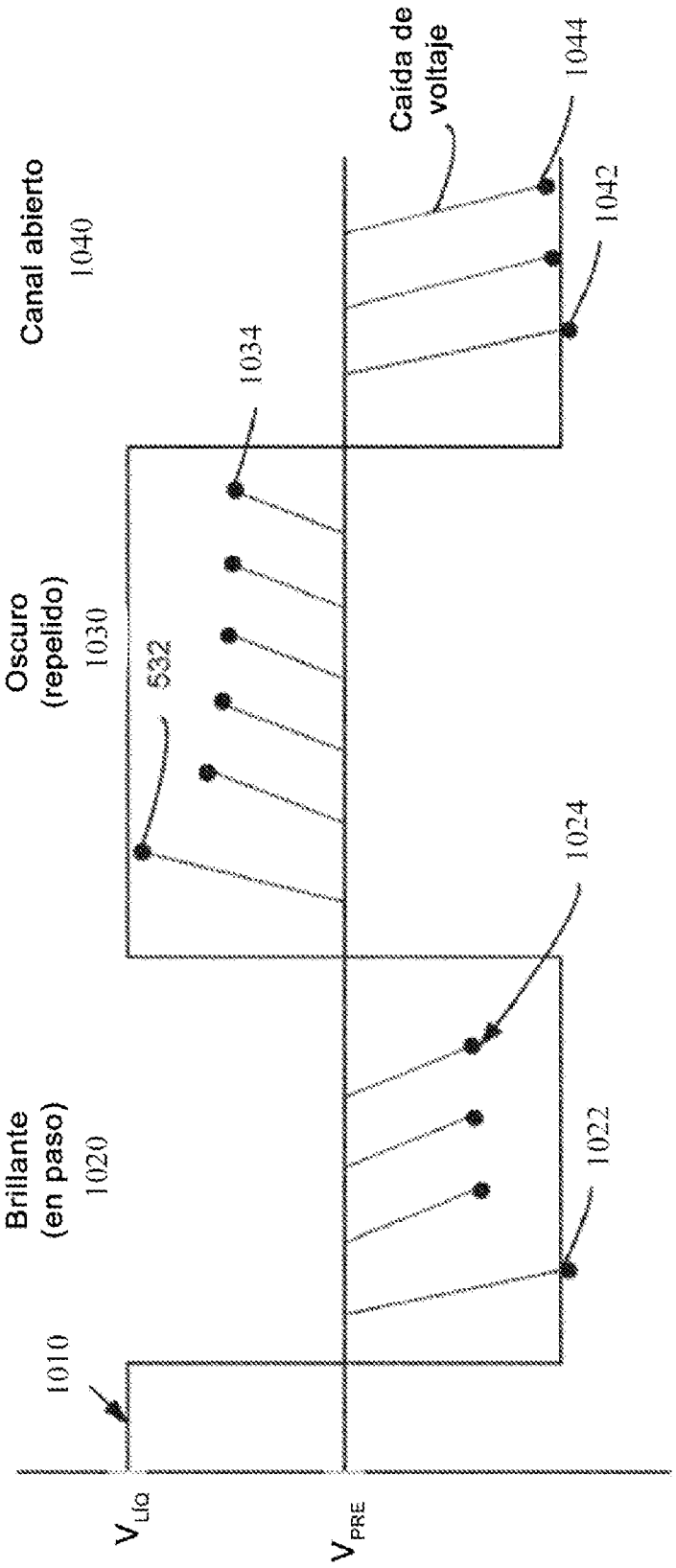

Caída de voltaje
1044
1042
Canal abierto
1040
1034
Oscuro (repelido)
1030
532
1024
Brillante (en paso)
1020
1022
1010
V_{LIQ}
V_{PRE}

Fig. 10

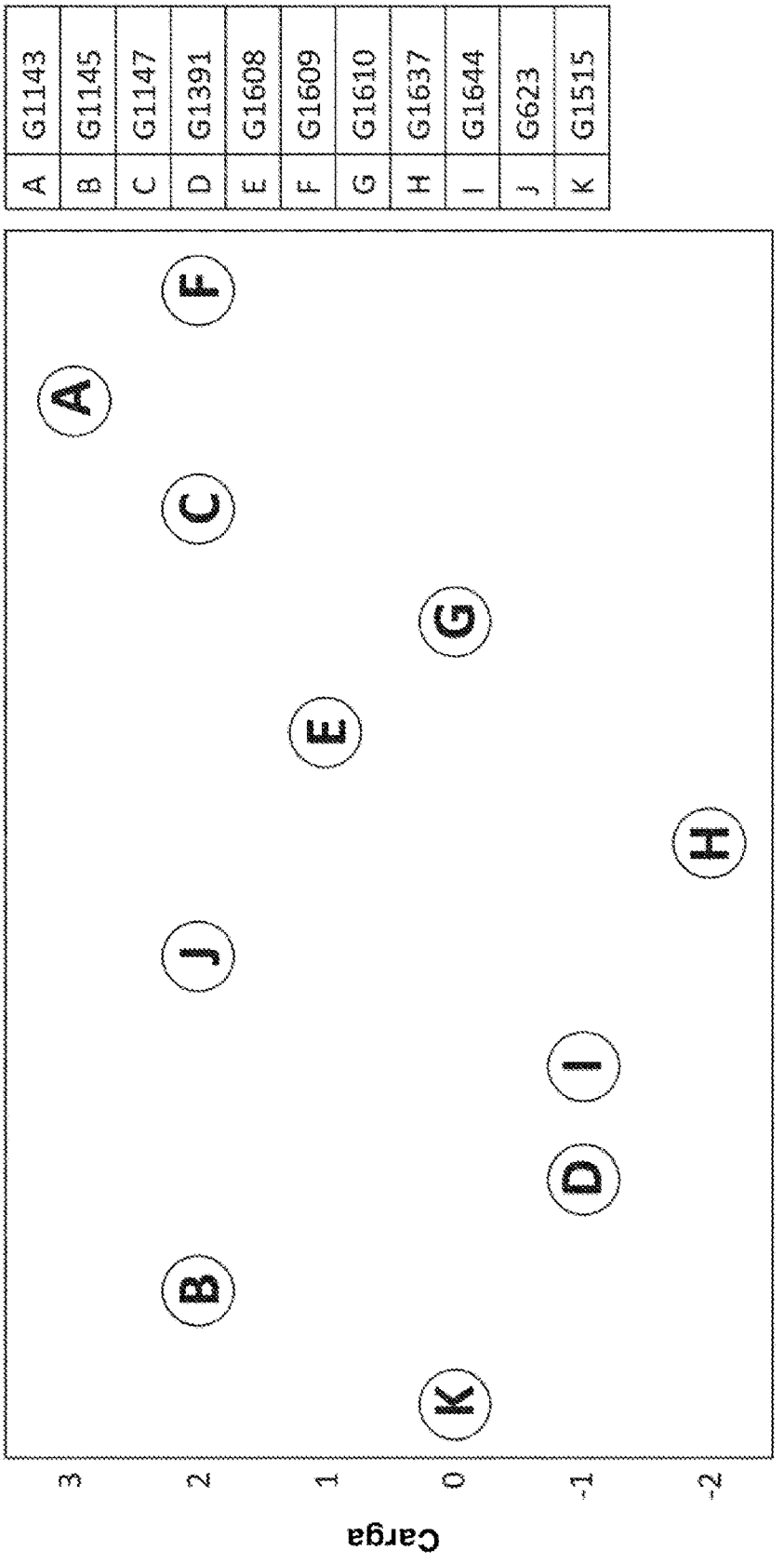
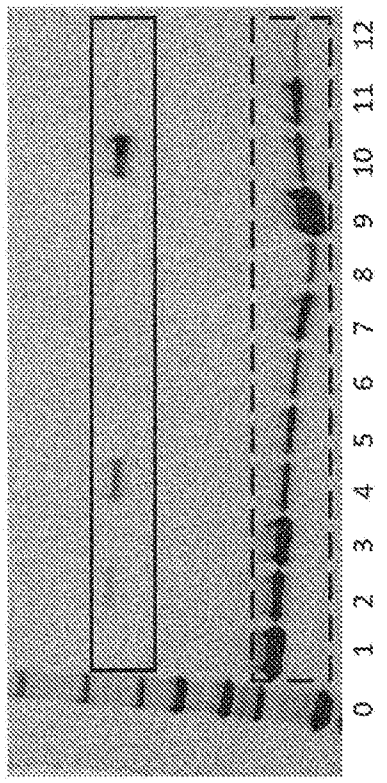
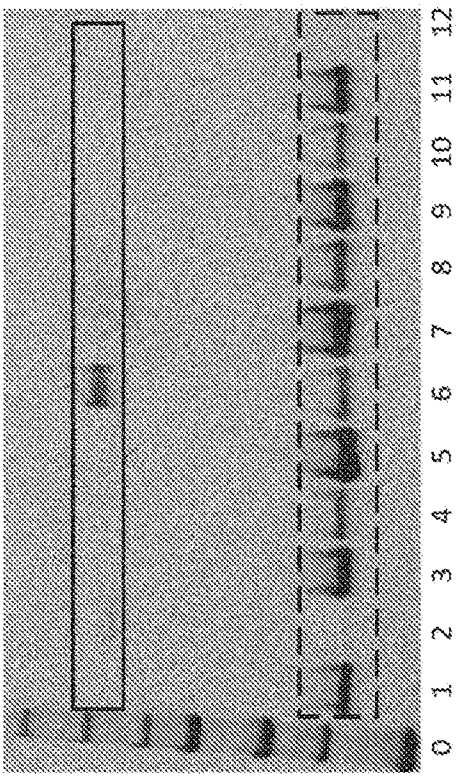

A G1143
B G1145
C G1147
D G1391
E G1608
F G1609
G G1610
H G1637
I G1644
J G623
K G1515
Carga
3
2
1
0
-1
-2
>0
0,219
0,403
0,557
0,577
0,883
0,907
0,908
1,08
1,12
1,49
A280

Fig. 11

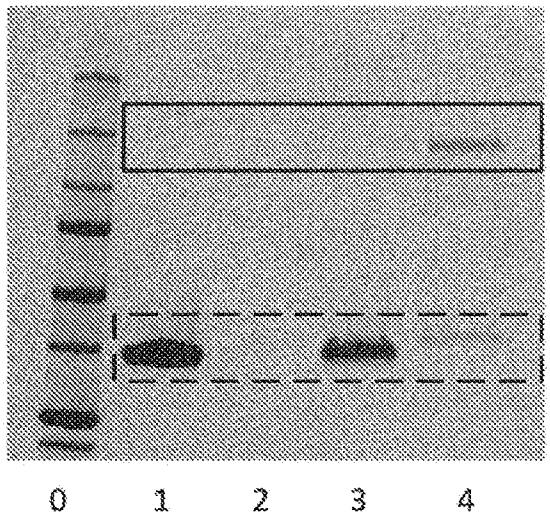


0	MW	7	G1145(O)
1	G1143(M)	8	G1391(O)
2	G1143(O)	9	G432(M)
3	G1147(M)	10	G432(O)
4	G1147(O)	11	G1644(M)
5	G1145(M)	12	G1644(O)
6	G1391(M)	MW = marcadores de peso molecular (M) = monómeros (O) = oligomerizado	



0	MW	7	G1609(O)
1	G623(M)	8	G1609(O)
2	G623(O)	9	G1610(M)
3	G1608(M)	10	G1610(O)
4	G1608(O)	11	G1637(M)
5	G432(M)	12	G1637(O)
6	G432(O)	MW = marcadores de peso molecular (M) = monómeros (O) = oligomerizado	

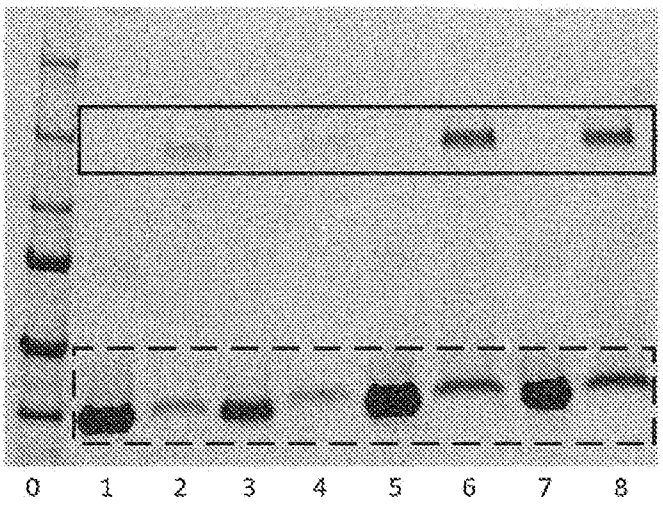
Fig. 12



0	MW
1	MES (M)
2	MES (O)
3	TMAO (M)
4	TMAO (O)

MW = marcadores de peso molecular
(M) = formulación monomérica
(O) = formulación oligomérica

Fig. 13A



0	MW
1	Acetato de sodio (M)
2	Acetato de sodio (O)
3	MES (M)
4	MES (O)
5	Fosfato de potasio (M)
6	Fosfato de potasio (O)
7	Tris (M)
8	Tris (O)

MW = marcadores de peso molecular
(M) = formulación monomérica
(O) = formulación oligomérica

Fig. 13B

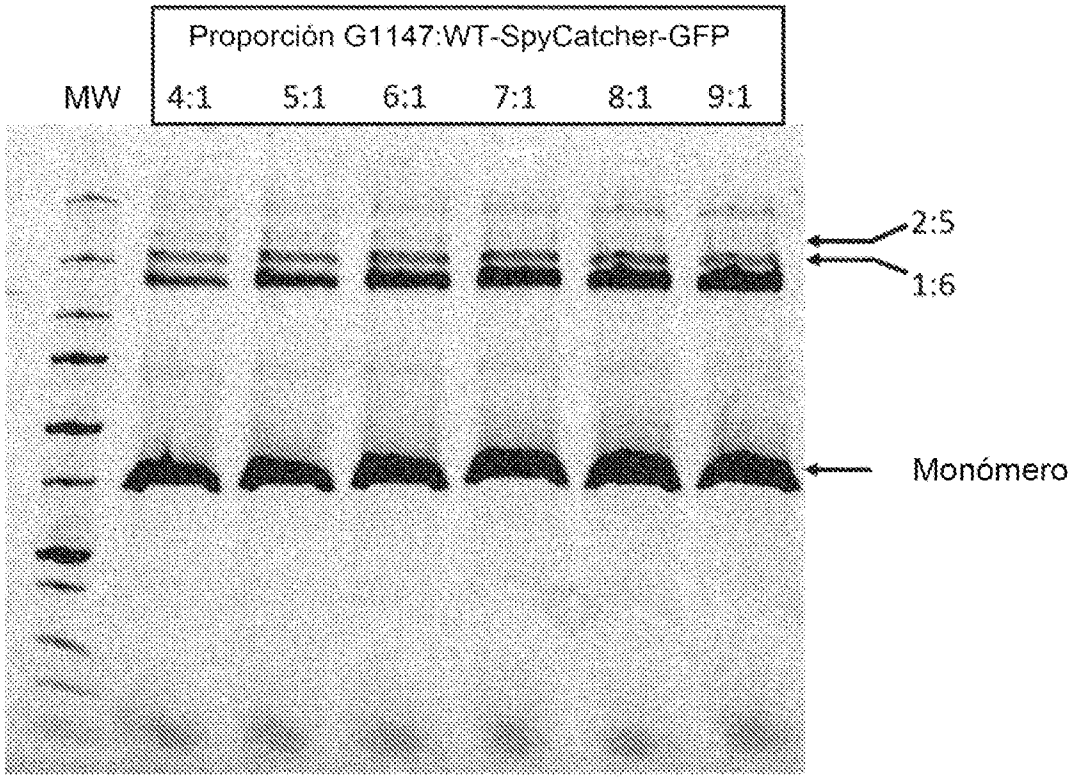

Proporción G1147:WT-SpyCatcher-GFP
MW
4:1
5:1
6:1
7:1
8:1
9:1
2:5
1:6
Monómero

Fig. 14A

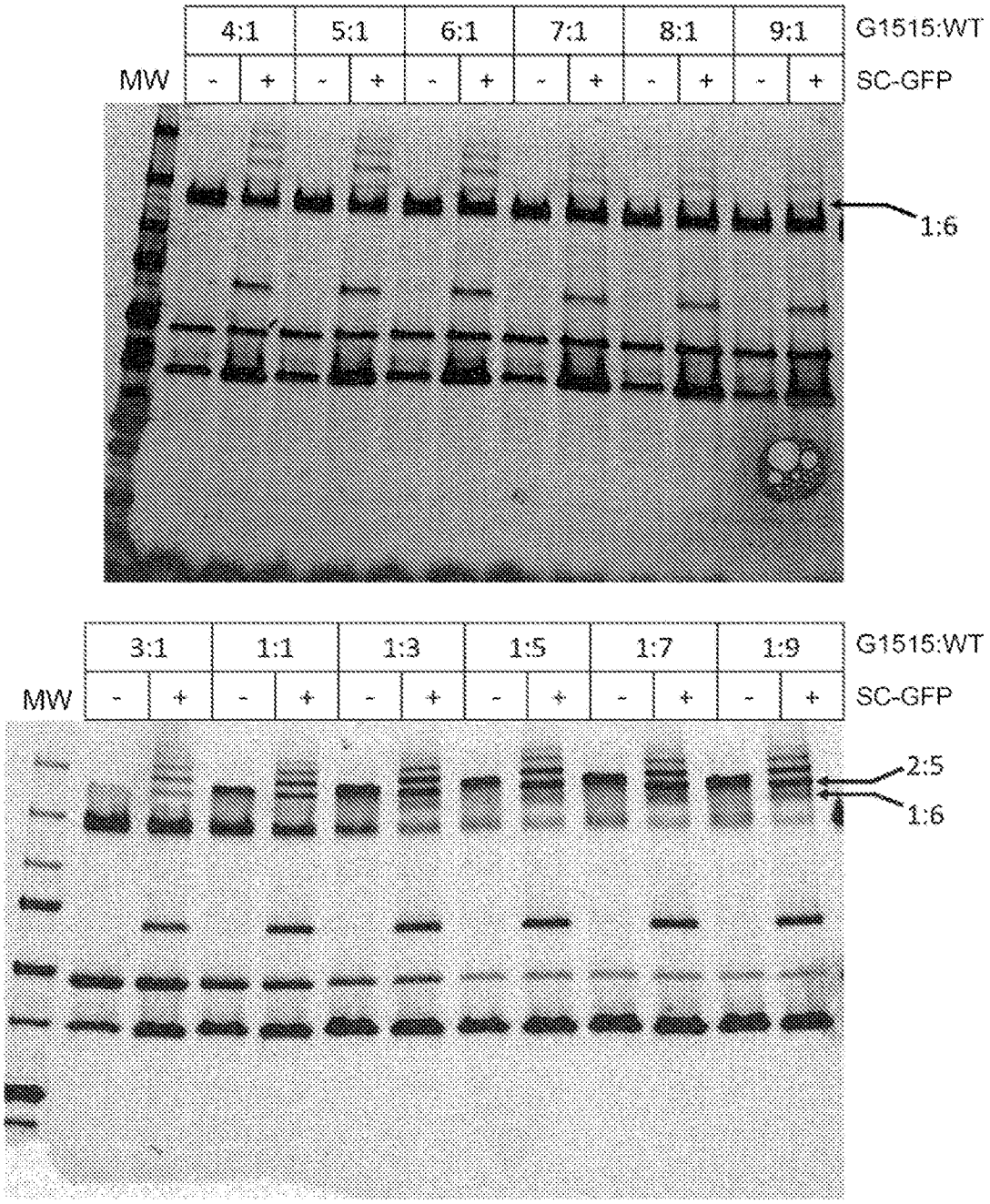

4:1
5:1
6:1
7:1
8:1
9:1
G1515:WT
MW
- + - + - + - + - + - +
SC-GFP
1:6
3:1
1:1
1:3
1:5
1:7
1:9
G1515:WT
MW
- + - + - + - + - + - +
SC-GFP
2:5
1:6

Fig. 14B

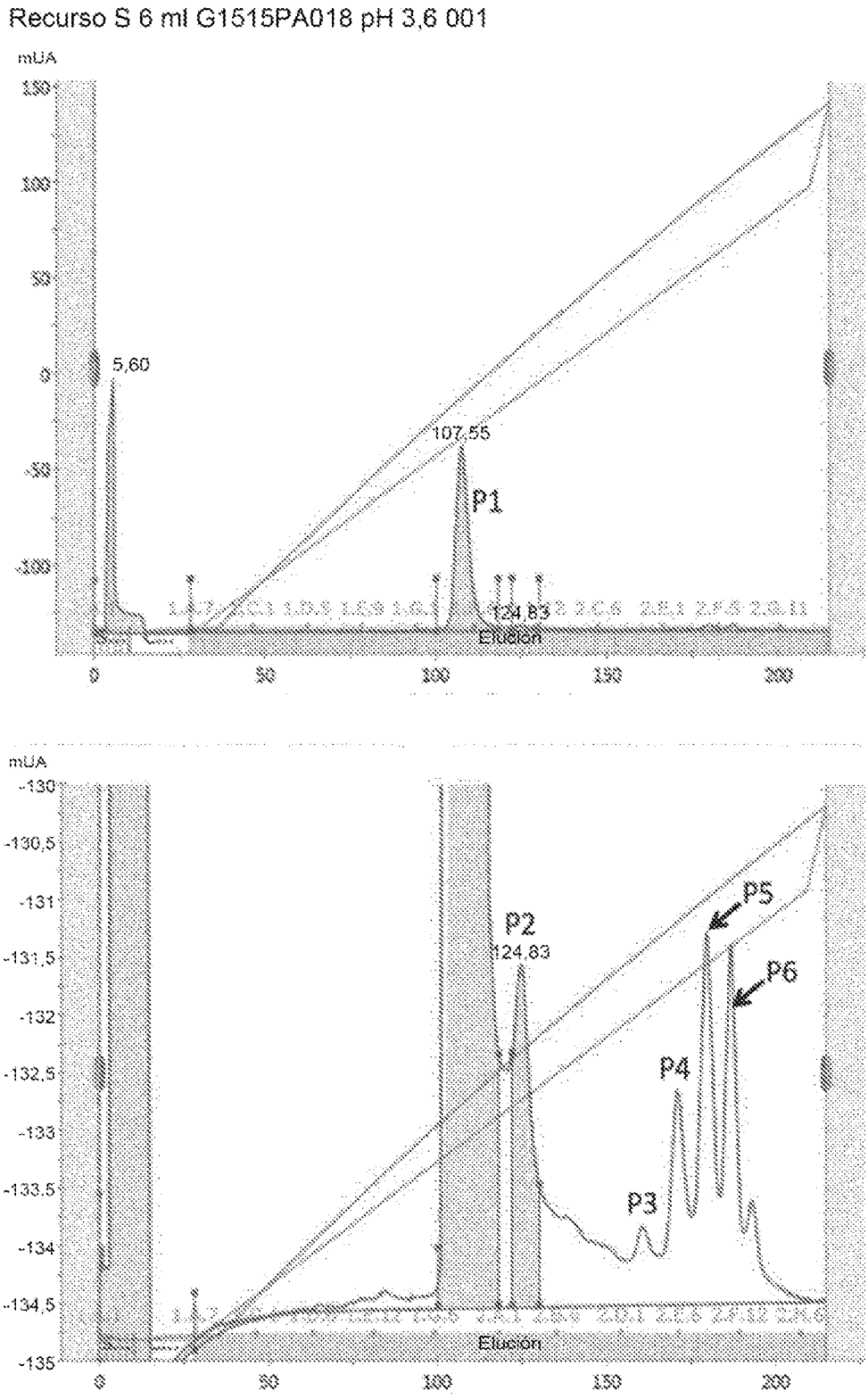

Recurso S 6 ml G1515PA018 pH 3,6 001
mUA
5,60
107,55
P1
124,83
Elución
mUA
P2
124,83
P5
P6
P4
P3
Elución

Fig. 14C

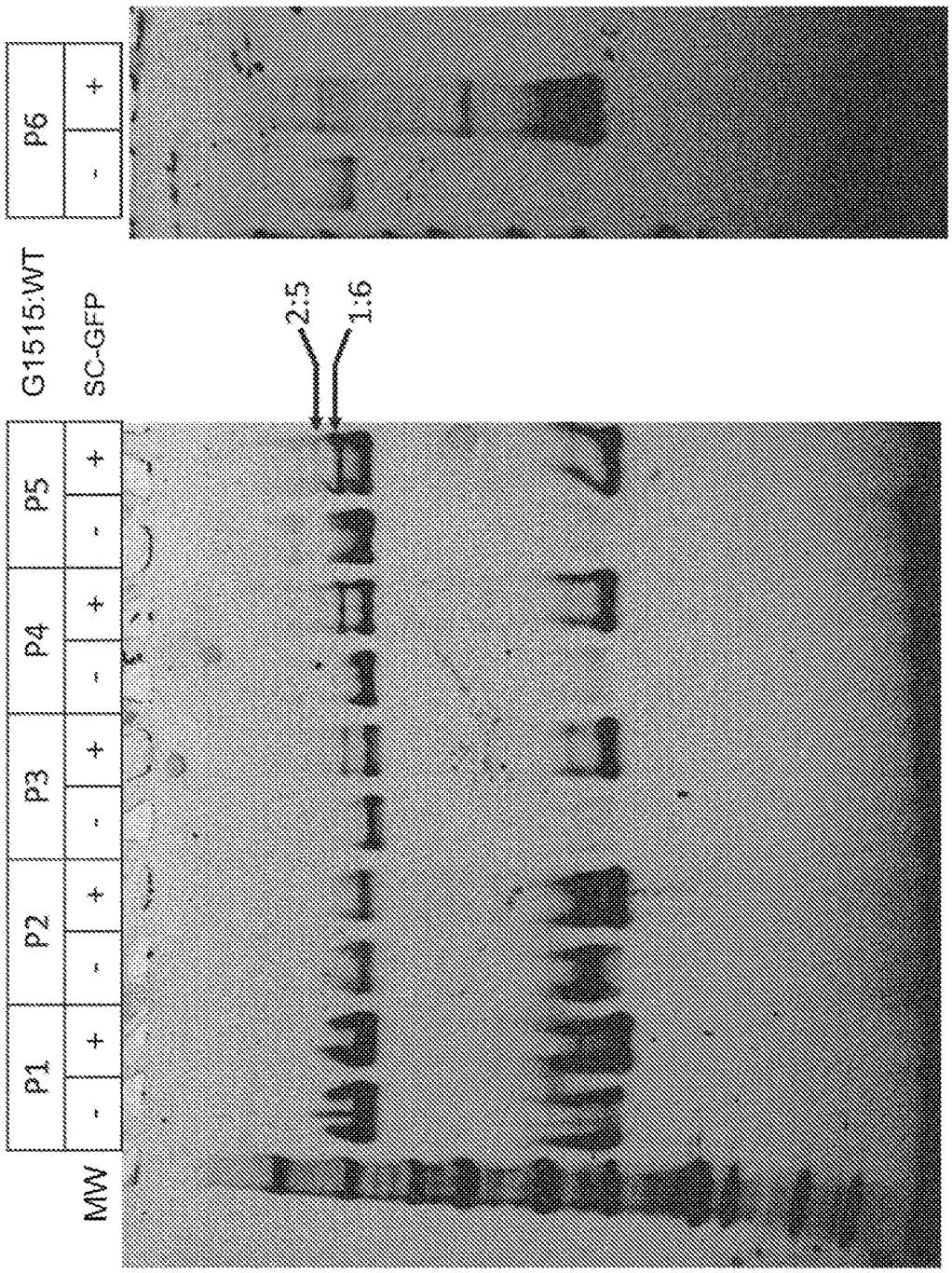
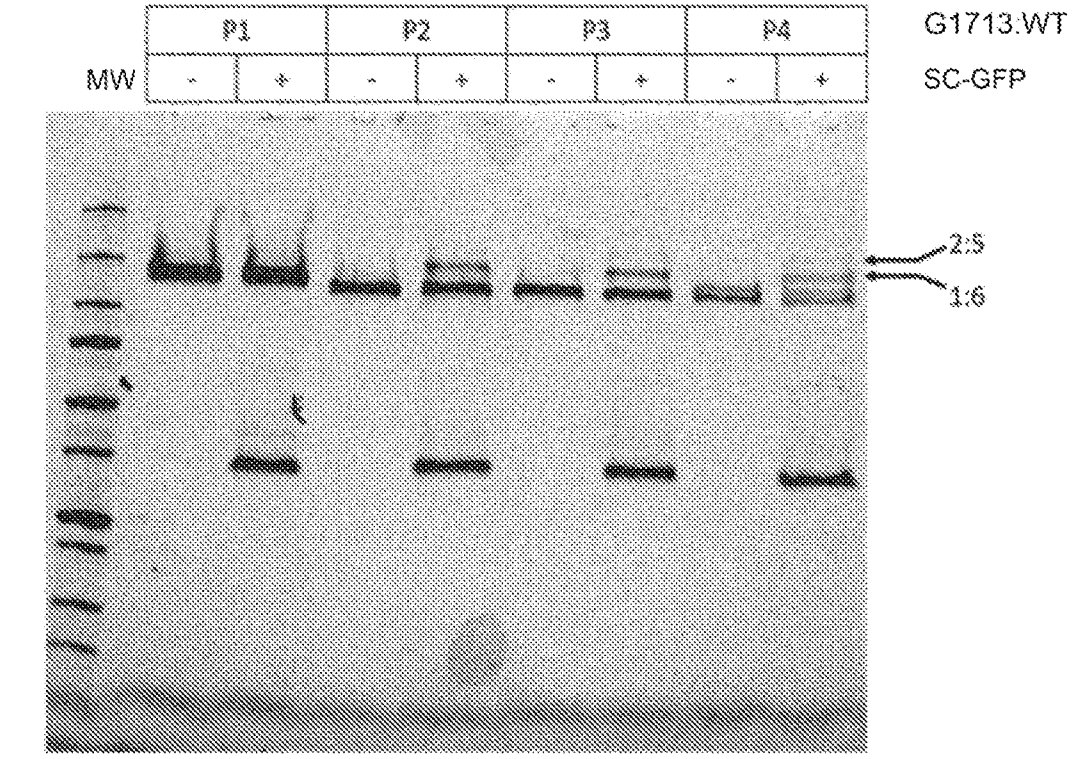


Fig. 14D


P1
P2
P3
P4
G1713:WT
MW
-
+
-
+
-
+
-
+
SC-GFP
2:5
1:6

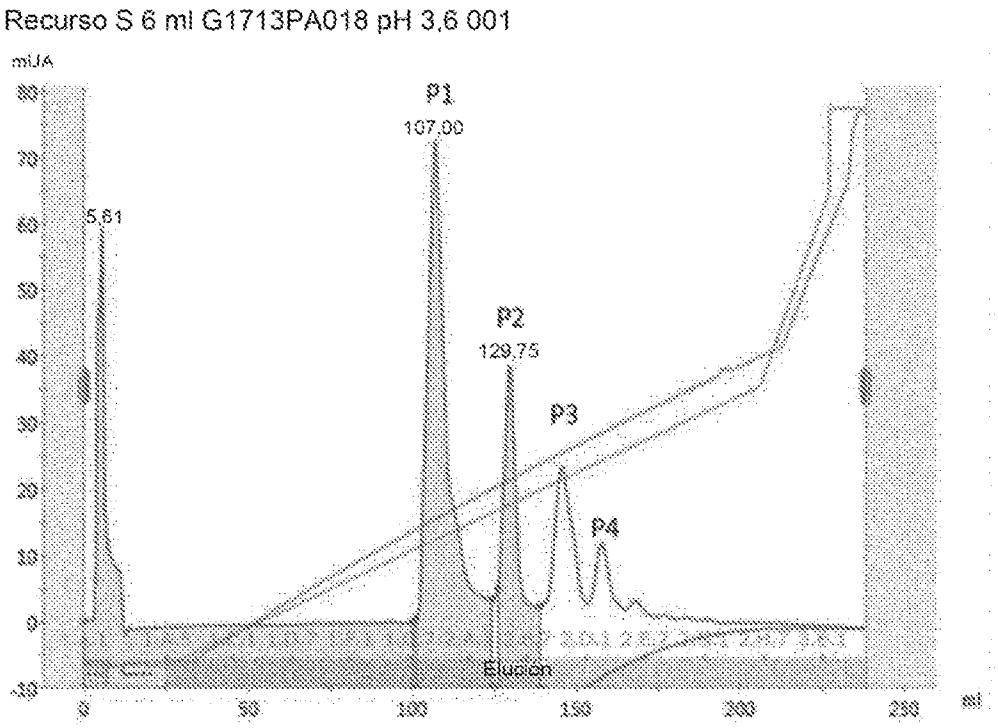

Recurso S 6 ml G1713PA018 pH 3,6 001
mUA
P1
107,00
5,61
P2
129,75
P3
P4
Elución
ml

Fig. 14E

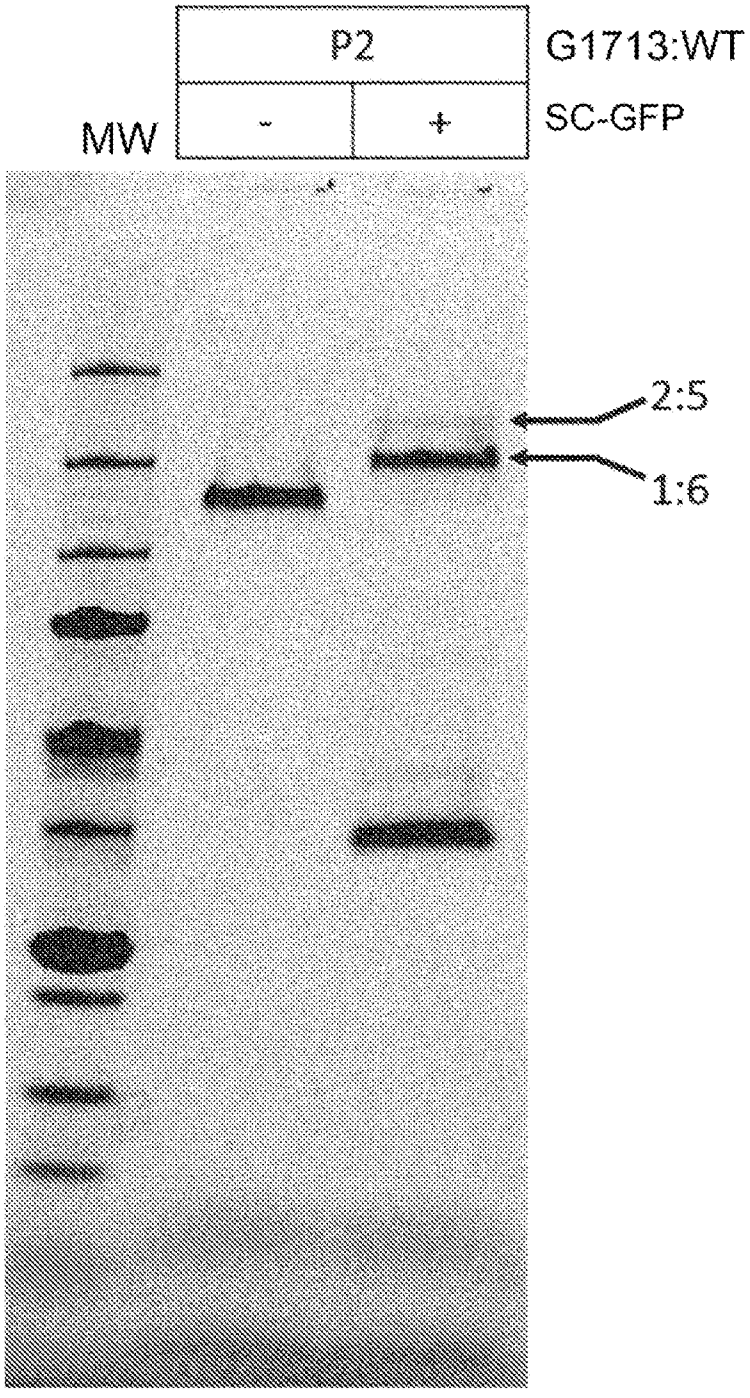


Fig. 14F

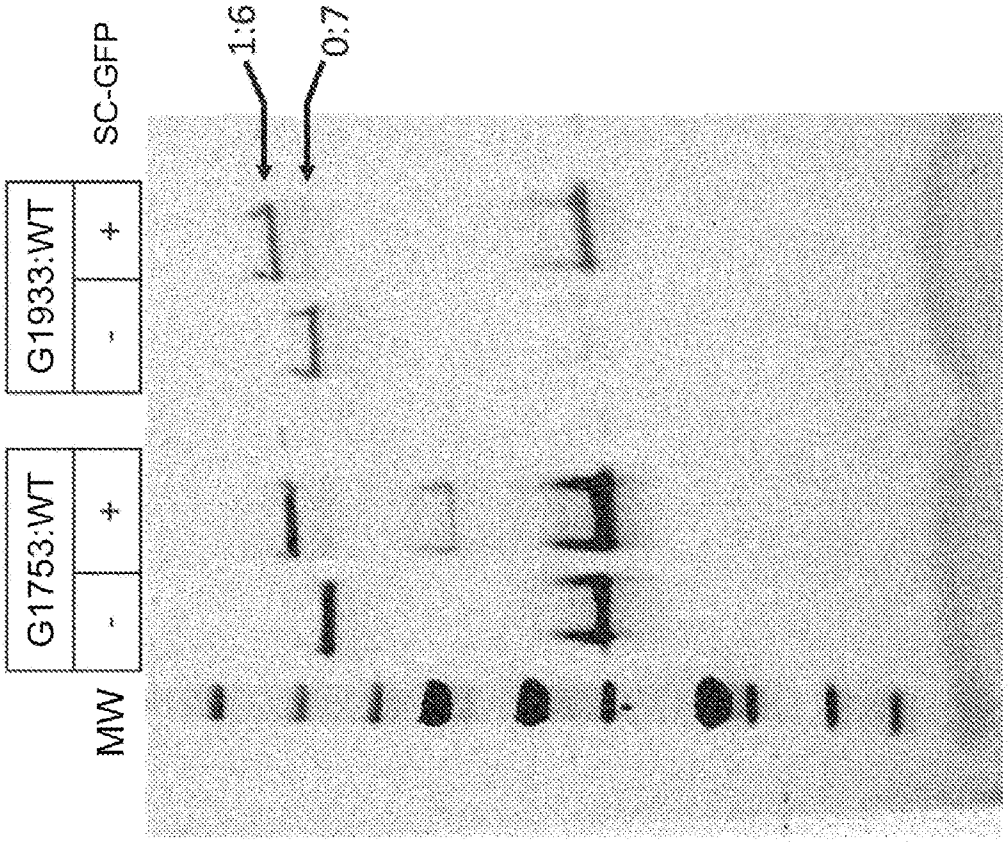


Fig. 14H

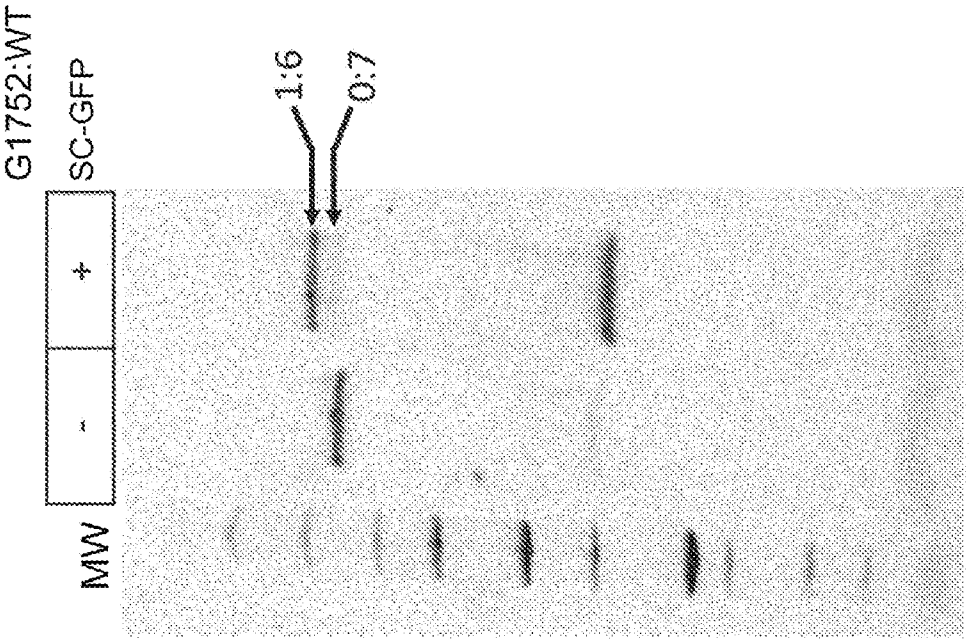


Fig. 14G

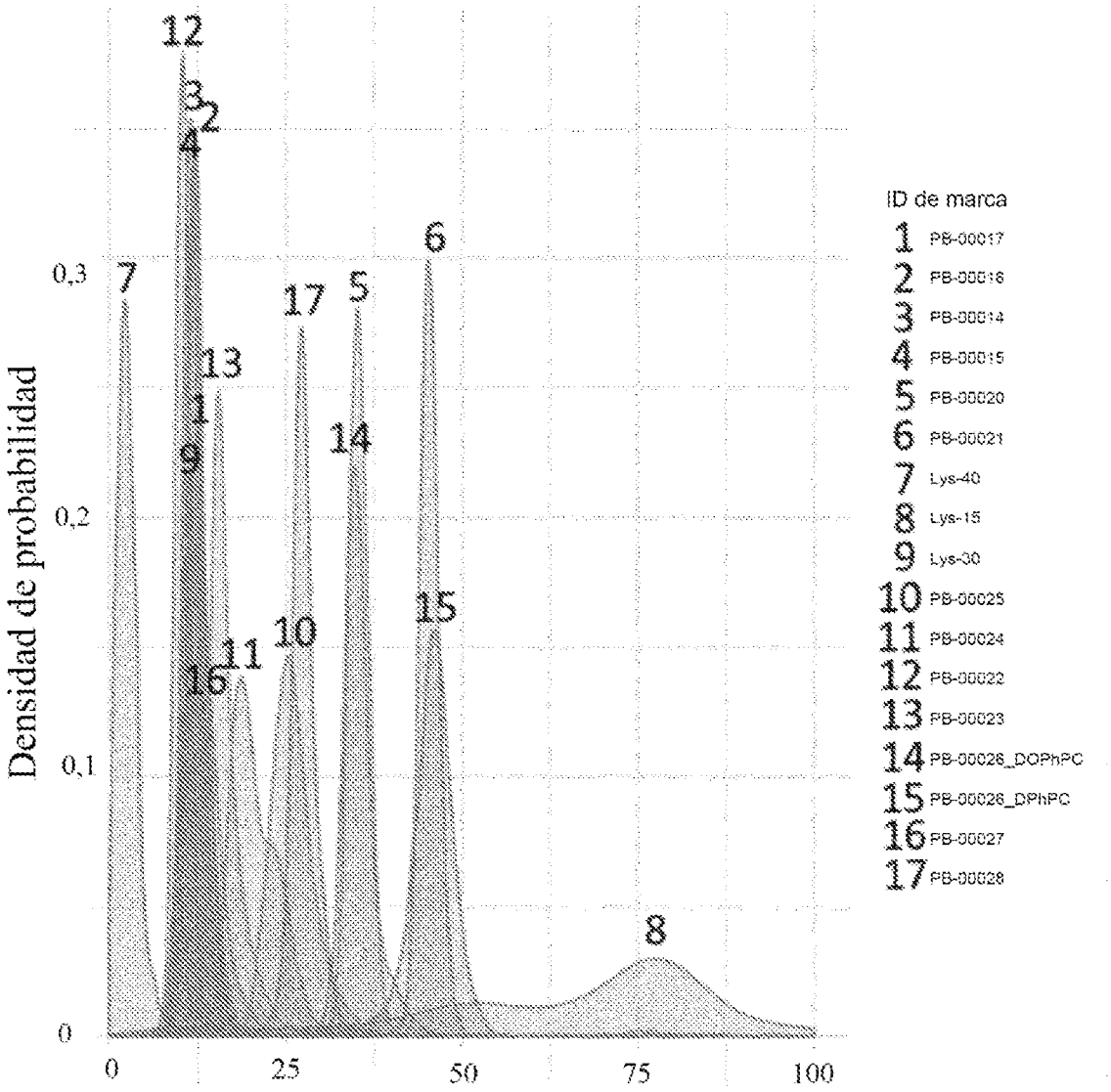

Densidad de probabilidad
0,3
0,2
0,1
0
0
25
50
75
100
Mediana de la fracción OC durante el paso de marca (%)
ID de marca
1 PB-00017
2 PB-00016
3 PB-00014
4 PB-00015
5 PB-00020
6 PB-00021
7 Lys-40
8 Lys-15
9 Lys-30
10 PB-00025
11 PB-00024
12 PB-00022
13 PB-00023
14 PB-00026_DOPhPC
15 PB-00026_DPhPC
16 PB-00027
17 PB-00028

Fig. 15A

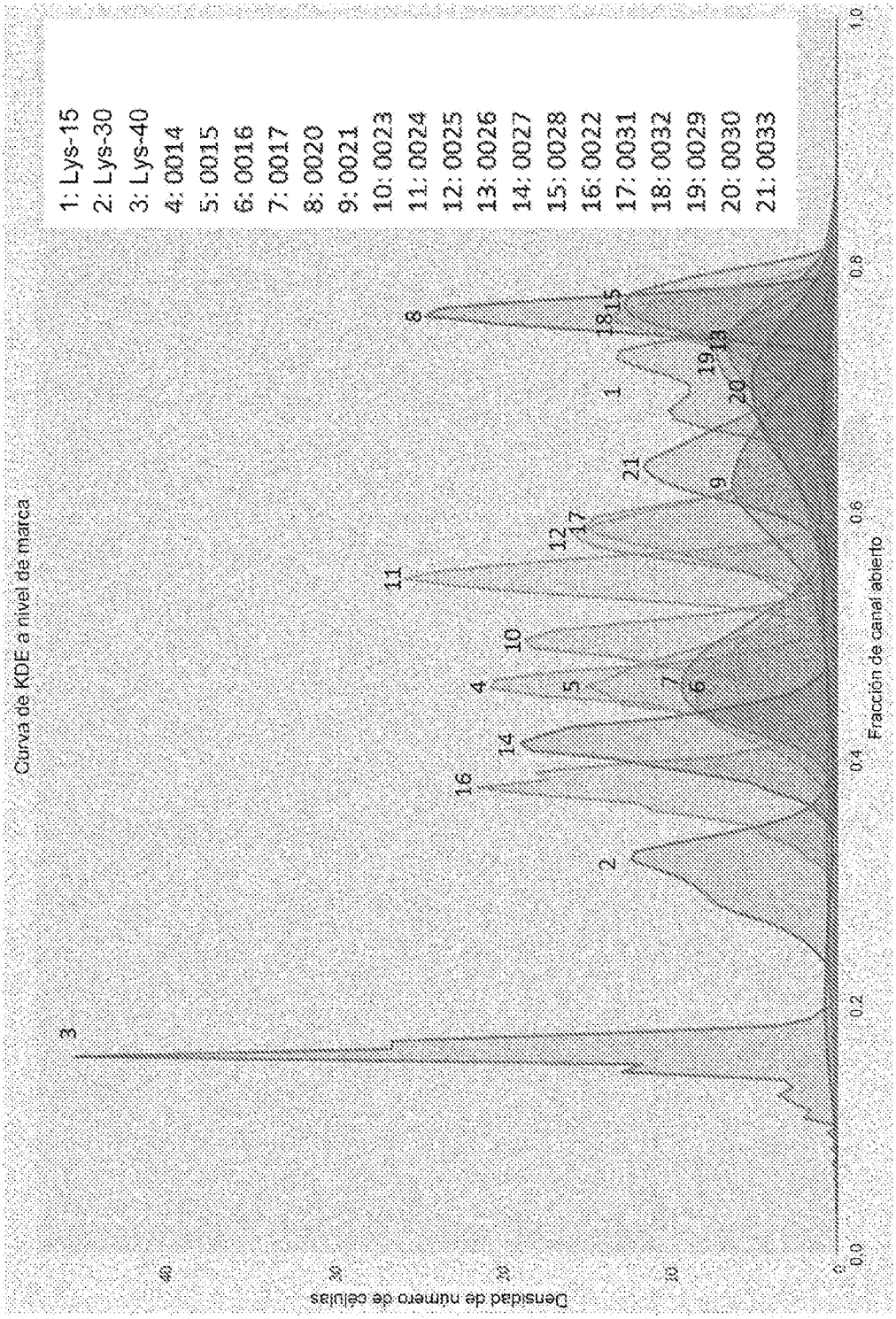

Curva de KDE a nivel de marca
Densidad de número de células
Fracción de canal abierto
0.0
0.2
0.4
0.6
0.8
1.0
1: Lys-15
2: Lys-30
3: Lys-40
4: 0014
5: 0015
6: 0016
7: 0017
8: 0020
9: 0021
10: 0023
11: 0024
12: 0025
13: 0026
14: 0027
15: 0028
16: 0022
17: 0031
18: 0032
19: 0029
20: 0030
21: 0033

Fig. 15B

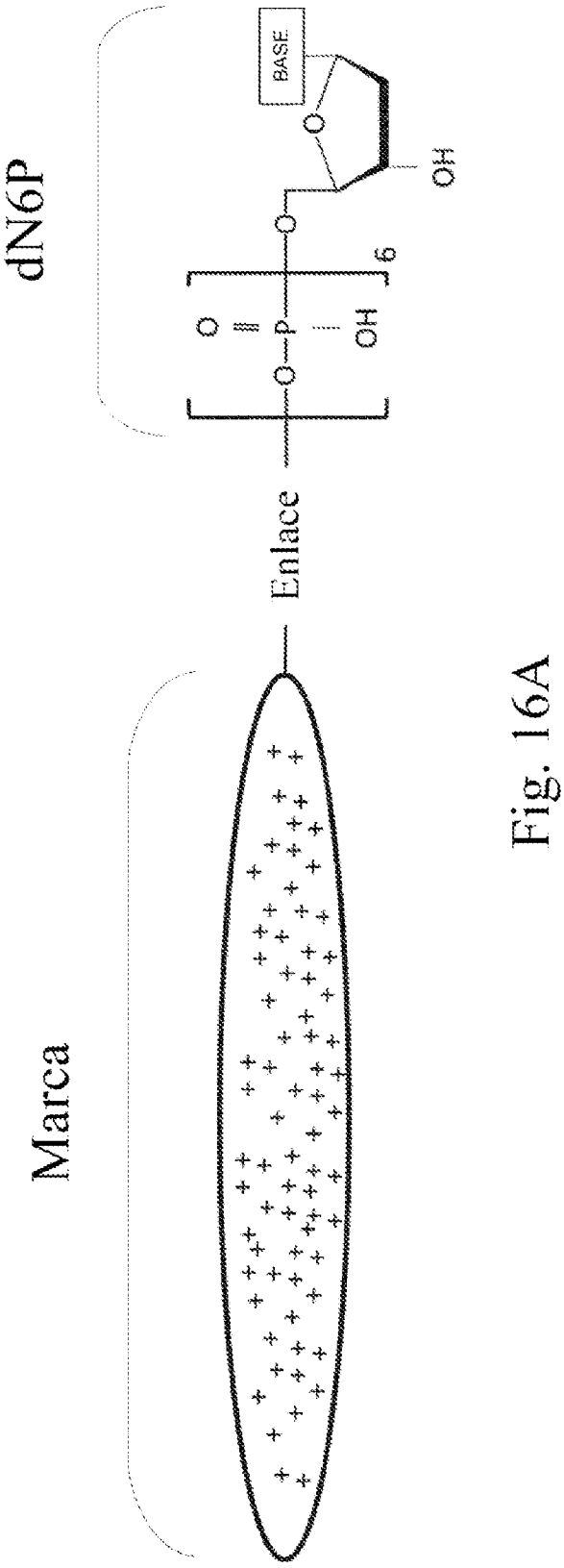


Fig. 16A

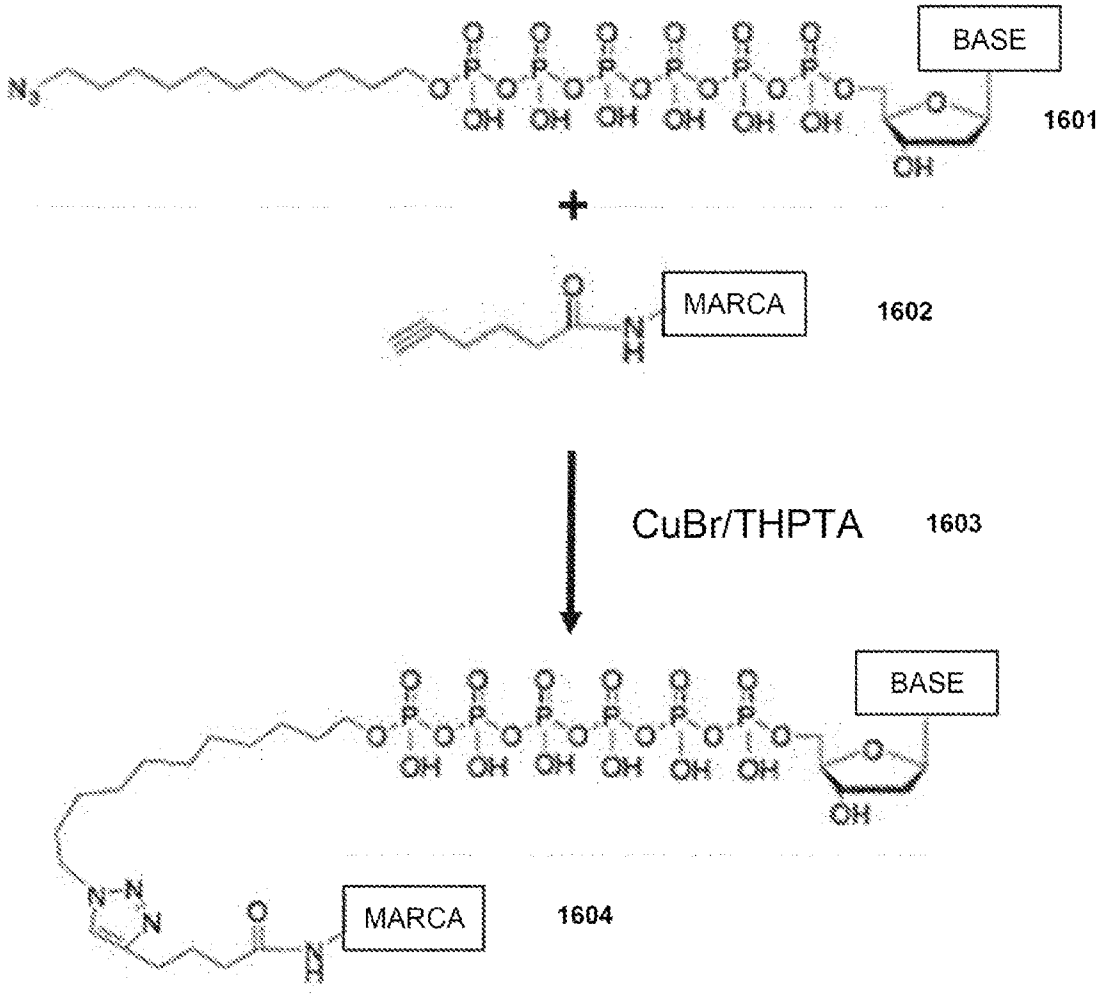


Fig. 16B

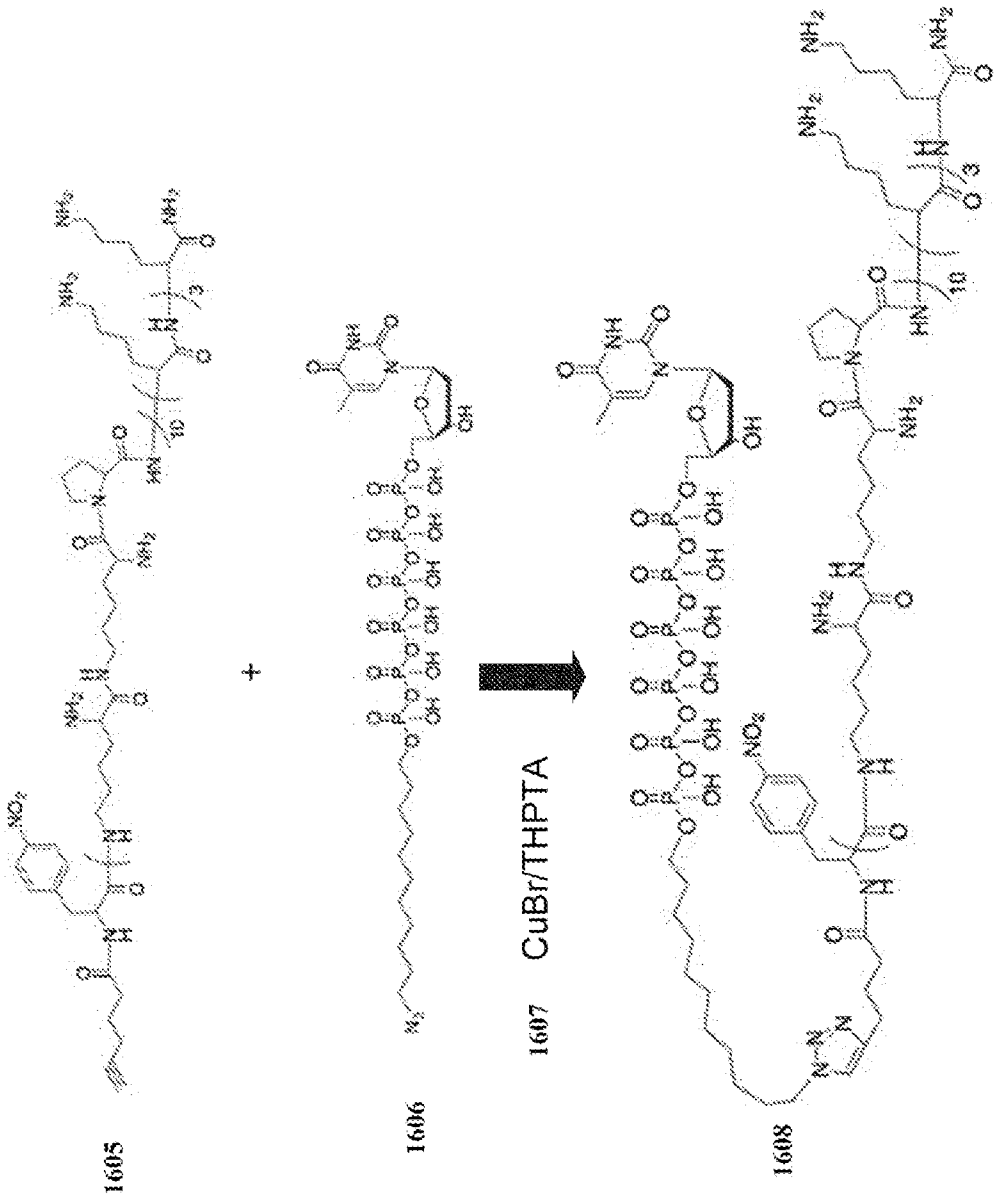


Fig. 16C

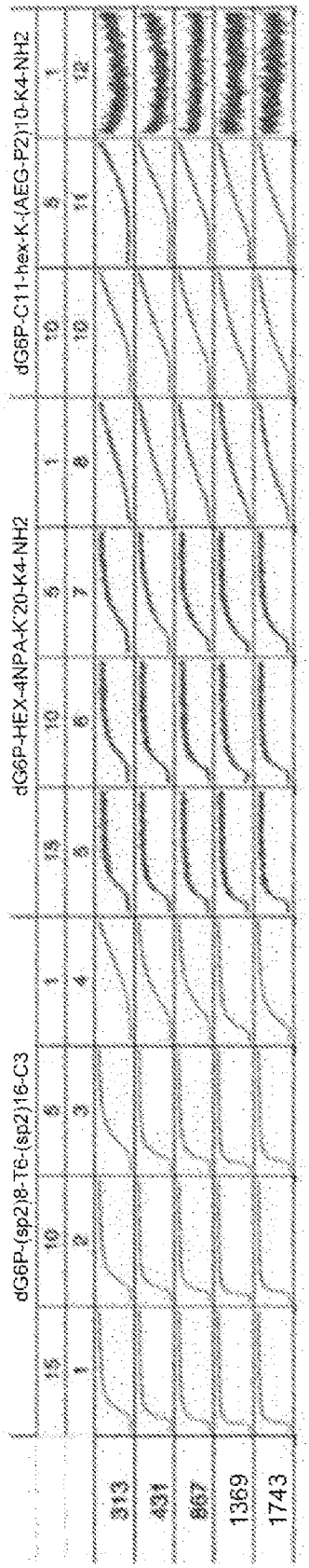

dG6P-(sp2)8-T6-(sp2)16-C3
dG6P-HEX-4NPA-K'20-K4-NH2
dG6P-C11-hex-K-(AEG-P2)10-K4-NH2
313
431
867
1369
1743

Fig. 17A

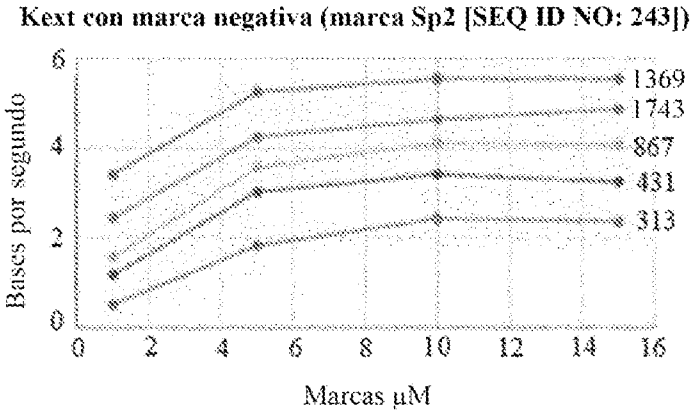


Fig. 17B

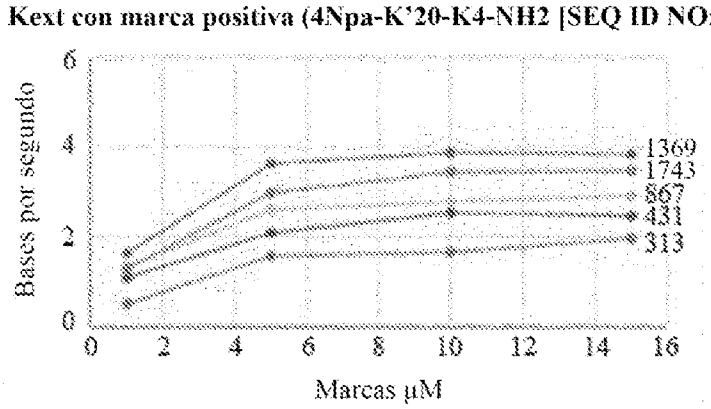


Fig. 17C

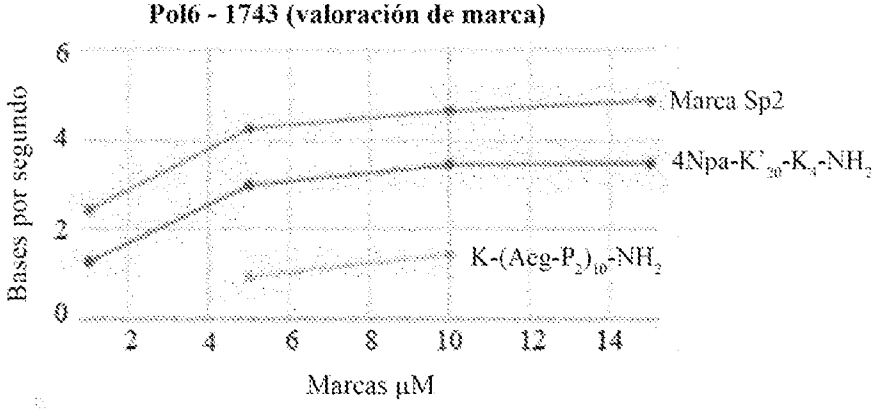


Fig. 17D

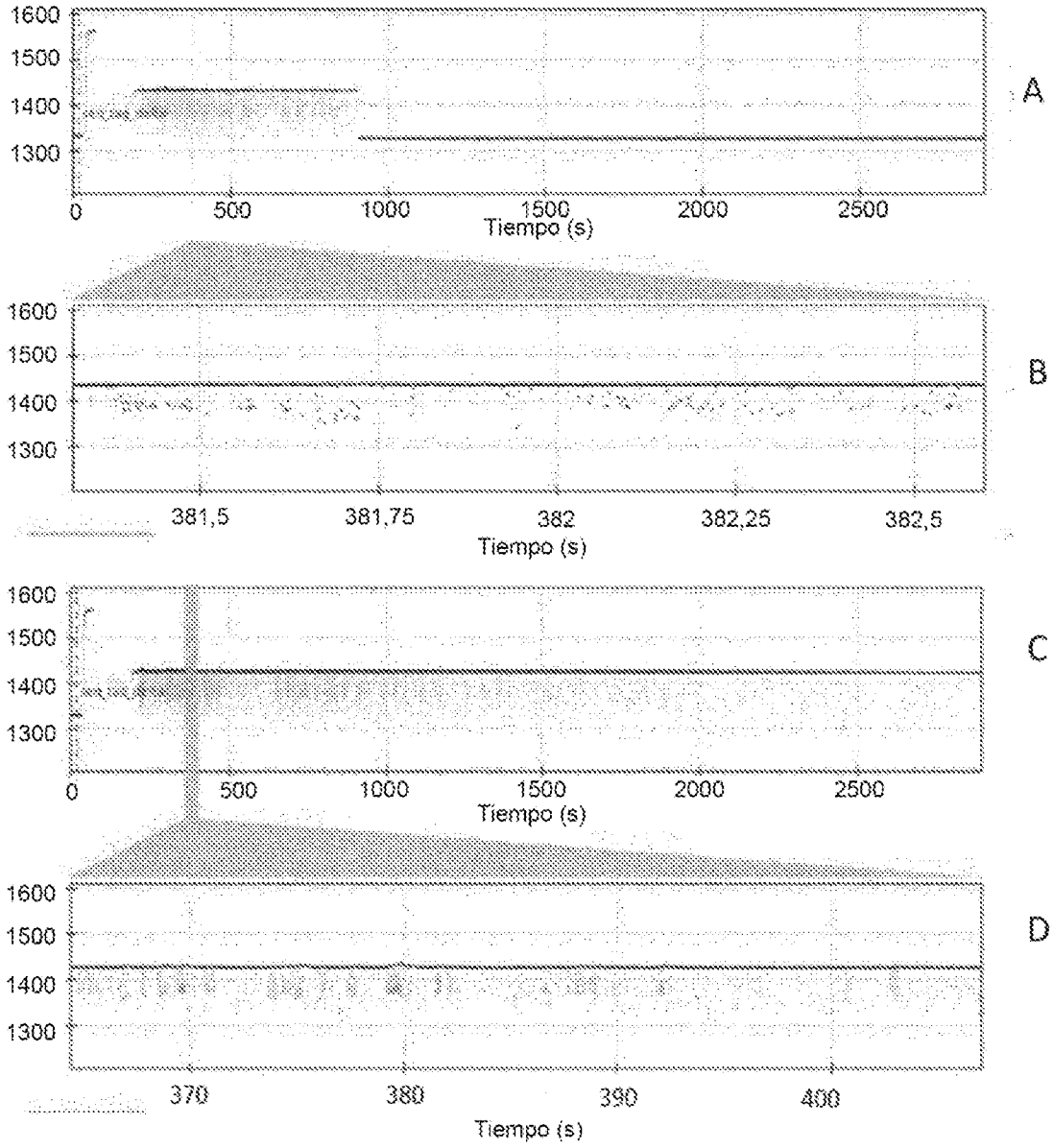

1600
1500
1400
1300
0
500
1000
1500
2000
2500
Tiempo (s)
A
1600
1500
1400
1300
381,5
381,75
382
382,25
382,5
Tiempo (s)
B
1600
1500
1400
1300
0
500
1000
1500
2000
2500
Tiempo (s)
C
1600
1500
1400
1300
370
380
390
400
Tiempo (s)
D

Fig. 18

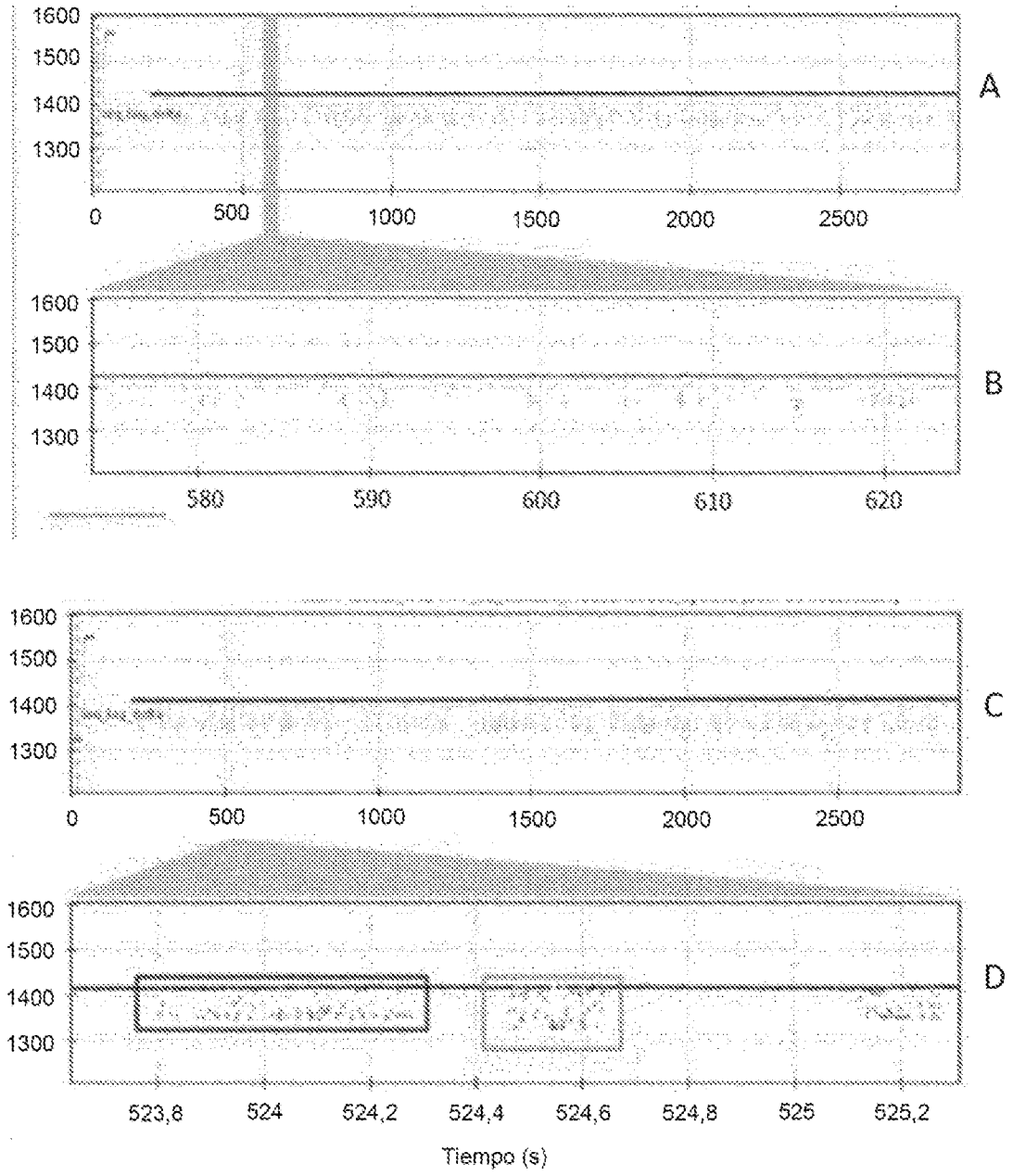

1600
1500
1400
1300
0
500
1000
1500
2000
2500
A
1600
1500
1400
1300
580
590
600
610
620
B
1600
1500
1400
1300
0
500
1000
1500
2000
2500
C
1600
1500
1400
1300
523,8
524
524,2
524,4
524,6
524,8
525
525,2
Tiempo (s)
D

Fig. 19