

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 988 083**

51 Int. Cl.:

C07H 19/02 (2006.01)
C07H 21/04 (2006.01)
C12Q 1/6869 (2008.01)
C07H 19/10 (2006.01)
C07H 19/20 (2006.01)
C07H 23/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **16.12.2019** **PCT/US2019/066670**
87 Fecha y número de publicación internacional: **25.06.2020** **WO20131759**
96 Fecha de presentación y número de la solicitud europea: **16.12.2019** **E 19839220 (1)**
97 Fecha y número de publicación de la concesión europea: **17.04.2024** **EP 3898643**

54 Título: **Nucleótidos protegidos reversiblemente en 3' para secuenciación por síntesis usando nanoporos**

30 Prioridad:

19.12.2018 US 201862781638 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
19.11.2024

73 Titular/es:

F. HOFFMANN-LA ROCHE AG (100.0%)
Grenzacherstrasse 124
4070 Basel, CH

72 Inventor/es:

BERGMANN, FRANK;
CRISALLI, PETER;
HEINDL, DIETER;
KHAKSHOOR, OMID y
TAING, MENG

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 988 083 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Nucleótidos protegidos reversiblemente en 3' para secuenciación por síntesis usando nanoporos

5 **Campo de la divulgación**

La presente divulgación se refiere en general a nucleótidos protegidos en 3' adecuados para su uso en secuenciación.

10 **Antecedentes de la divulgación**

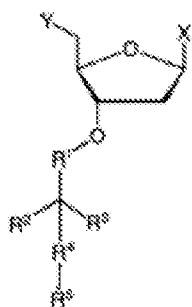
La importancia de la secuenciación de ADN se ha incrementado notablemente desde sus inicios hace cuatro décadas. Se reconoce como una tecnología fundamental para la mayoría de las áreas de la biología y la medicina y como la base del nuevo paradigma de la medicina personalizada y de precisión. La información sobre los genomas y epigenomas de los individuos puede ayudar a revelar su propensión a padecer enfermedades, su pronóstico clínico y su respuesta a los tratamientos, pero la aplicación rutinaria de la secuenciación del genoma en medicina requerirá datos completos entregados de manera oportuna y rentable.

La secuenciación de ácidos nucleicos basada en nanoporos es un enfoque que se ha estudiado ampliamente. En las últimas dos décadas, ha habido un gran interés en aprovechar los nanoporos para la caracterización de polímeros y para distinguir nucleótidos de manera rápida, con bajo coste y de molécula individual. Por ejemplo, Kasianowicz *et al.*, caracterizaron polinucleótidos monocatenarios a medida que se translocaban eléctricamente a través de un nanoporo de alfa hemolisina incrustado en una bicapa lipídica (véase, por ejemplo, Kasianowicz, J. (1996), Characterization of Individual Polynucleotide Molecules using a Membrane Channel. Proc. Natl. Acad. Sci., 93, 13770-3). Se demostró que durante la translocación de polinucleótidos el bloqueo parcial de la abertura del nanoporo se podía medir como una disminución de la corriente iónica. De forma similar, Gundlach *et al.*, demostraron un procedimiento de secuenciación de ADN que usaba un nanoporo de bajo ruido derivado de *Mycobacterium smegmatis* ("MspA") junto con un proceso llamado secuenciación interrumpida doble (véase, por ejemplo, Derrington, I. *et al.* (2010), Nanopore DNA Sequencing with MspA. Proc. Natl. Acad. Sci., 107(37), 16060-16065). Aquí, se usó un dúplex bicatenario para mantener temporalmente la porción monocatenaria del ácido nucleico en la constricción de MspA. Akesson *et al.*, (véase, por ejemplo, la publicación PCT n.º WO/20150344945) divulgan procedimientos para caracterizar polinucleótidos en un nanoporo que utilizan un motor molecular situado de forma contigua para controlar la tasa de translocación del polinucleótido a través o contigua a la abertura del nanoporo.

En general, se han seguido tres enfoques de secuenciación por nanoporos: secuenciación de hebras en la que se identifican las bases de ADN a medida que pasan secuencialmente a través de un nanoporo, secuenciación por nanoporos basada en exonucleasas en la que los nucleótidos se escinden enzimáticamente uno por uno de una molécula de ADN y se supervisan a medida que se capturan y pasan a través del nanoporo, y un enfoque de secuenciación por síntesis (SBS) por nanoporos en el que marcas de polímeros identificables se unen a nucleótidos y se registran en nanoporos durante la síntesis de ADN catalizada por enzimas. Todos estos procedimientos tienen en común la necesidad de un control preciso de las velocidades de reacción de modo que cada base se determine en orden. La secuenciación de hebras requiere un procedimiento para ralentizar el paso del ADN a través del nanoporo y decodificar una pluralidad de bases dentro del canal; para este propósito se han desarrollado enfoques de trinquete que aprovechan los motores moleculares. La secuenciación basada en exonucleasas requiere la liberación de cada nucleótido lo suficientemente cerca del poro para garantizar su captura y su tránsito a través del poro a una velocidad lo suficientemente lenta como para obtener una señal de corriente iónica válida. Además, ambos procedimientos se basan en distinciones entre las cuatro bases naturales, dos purinas relativamente similares y dos pirimidinas similares. El enfoque de SBS por nanoporos utiliza marcas de polímeros sintéticos unidas a los nucleótidos que se diseñan específicamente para producir firmas de bloqueo de corriente iónica exclusivas y fácilmente distinguibles para la determinación de secuencias. El documento WO2013/191793 describe nucleótidos que tienen un resto de bloqueo reversible en 3' usados para la secuenciación a través de secuenciación por síntesis de nanoporos. El documento divulga un compuesto que tiene un grupo protector en 3' que comprende un grupo ácido y un conector que no es un resto CH₂.

55 **Breve resumen de la divulgación**

En un aspecto de la presente divulgación están nucleótidos o sales de los mismos que tienen la fórmula (IIIA):



(III A)

en la que

5 X es una nucleobase o una nucleobase marcada;

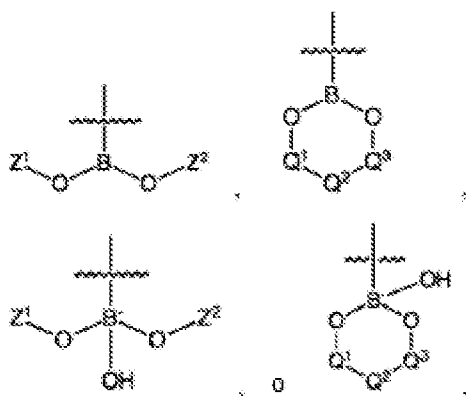
Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

10 R^1 es un enlace, $-CH_2-$, $-C(O)-O-$, $-C(O)-NR^a-$ o $-C(O)-R^x-$;

R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-[(C(R^a)(R^b))_p-O]_q-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$, $-CN$ o $-NO_2$;

15 R^4 es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-CH=CH-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u $-O-C(O)-$ aril-;

20 R^5 es $-(C(R^a)(R^b))_n-N_3$, $-(C(R^a)(R^b))_n-N^+C^-$, $-(C(R^a)(R^b))_n-CN$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:



25 Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

30 Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

35 R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

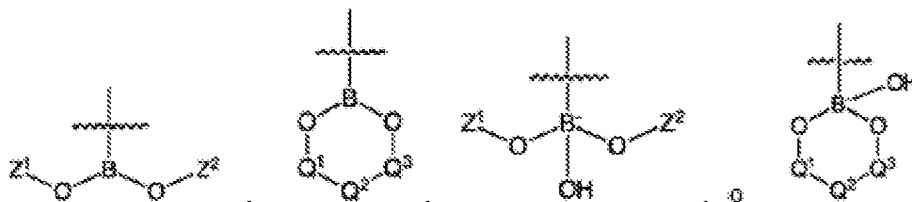
n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3;

siempre que cuando R^1 y R^4 sean ambos enlaces y cuando R^2 y R^3 sean ambos H, entonces R^5 no es una acida (por ejemplo, R^5 no es N^3).

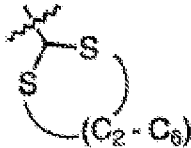
5

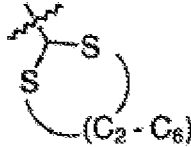
En algunos modos de realización, R^5 es:



- 10 En algunos modos de realización, al menos uno de Z^1 o Z^2 es H. En algunos modos de realización, R^1 es un enlace y al menos uno de Z^1 o Z^2 es H. En algunos modos de realización, R^1 es $-CH_2-$ y al menos uno de Z^1 o Z^2 es H. En algunos modos de realización, Z^1 y Z^2 son H. En algunos modos de realización, R^5 es $-B(OH)_2$.

- 15 En algunos modos de realización, R^1 es un enlace. En algunos modos de realización, R^1 es un enlace y R^4 es un enlace. En algunos modos de realización, R^1 es un enlace y R^4 es un enlace, y al menos uno de R^2 o R^3 es H. En algunos modos de realización, R^1 es un enlace y R^4 es un grupo arilo de 6 miembros. En algunos modos de realización, R^1 es un enlace y R^4 es un grupo arilo de 6 miembros, y al menos uno de R^2 o R^3 es H. En algunos modos de realización, R^1 es $-C(O)-O-$. En algunos modos de realización, R^1 es $-C(O)-O-$ y R^4 es un grupo arilo de 6 miembros. En algunos modos de realización, el grupo arilo de 6 miembros incluye al menos un sustituyente, en el que el al menos un sustituyente se selecciona del grupo que consiste en metilo y etilo. En algunos modos de realización, R^1 es $-C(O)-O-$ y R^4 es un grupo arilo de 6 miembros al menos uno de R^2 o R^3 es H. En algunos modos de realización, R^1 es un enlace y R^4 es $-CH=CH-$.

- 25 En algunos modos de realización, R^5 es , donde C_2-C_6 representa una cadena de alquilo de 2 a 6 carbonos saturada que puede estar sustituida o no sustituida. En algunos modos de realización, R^1 es un enlace y

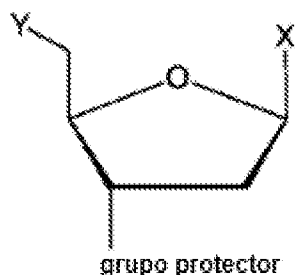
R^5 es , donde C_2-C_6 representa una cadena de alquilo de 2 a 6 carbonos saturada que puede estar sustituida o no sustituida.

- 30 En algunos modos de realización, R^5 es $-(C(R^a)(R^b))_n-CN$. En algunos modos de realización, R^1 es un enlace y R^5 es $-(C(R^a)(R^b))_n-CN$. En otros modos de realización, R^5 es $-(C(R^a)(R^b))_n-N^+C^-$. En otros modos de realización, R^1 es un enlace y R^5 es $-(C(R^a)(R^b))_n-N^+C^-$.

- 35 En algunos modos de realización, R^1 es un enlace, R^2 es $-[(C(R^a)(R^b))_p-O]_q-(R^a)$ o $-C(O)-OR^a-$, y en los que al menos un R^a es un grupo alquilo C_1-C_6 . En algunos modos de realización, R^1 es $-C(O)-R^x-$. En algunos modos de realización, R^1 es $-C(O)-R^x-$ y R^5 es $-(C(R^a)(R^b))_n-N_3$. En algunos modos de realización, R^5 se deriva de un 1,4-epoxi-1,4-dihidronaftaleno sustituido o no sustituido.

- 40 En algunos modos de realización, R^1 es $-CH_2-$ y R^5 es $-B(OZ^1)(OZ^2)$. En algunos modos de realización, Z^1 y Z^2 se seleccionan independientemente del grupo que consiste en H, metilo y etilo. En algunos modos de realización, R^1 es $-CH_2-$ y R^5 es $-B(OH)_2$. En algunos modos de realización, R^1 es $-CH_2-$ y R^5 es $-B(OH)_2$.

En otro aspecto de la presente divulgación están nucleótidos o sales de los mismos que tienen la fórmula (I):



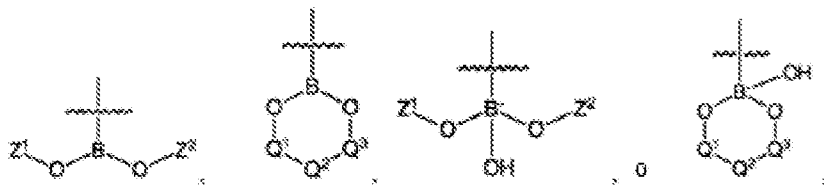
(I)

en la que

5 X es una nucleobase o una nucleobase marcada;

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

10 'grupo protector' tiene la estructura:



donde

15 Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

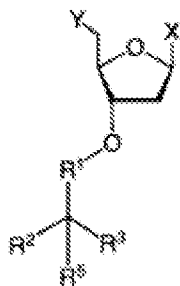
20 Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2; y

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido.

25 En algunos modos de realización, el 'grupo protector' es $-B(OZ^1)(OZ^2)$. En algunos modos de realización, el 'grupo protector' es $-B(OZ^1)(OZ^2)$, donde Z^1 y Z^2 se seleccionan independientemente de un grupo alquilo C_1-C_4 . En algunos modos de realización, el 'grupo protector' es $-B(OMe)(OMe)$. En algunos modos de realización, el 'grupo protector' es $-B(OEt)(OEt)$. En algunos modos de realización, el 'grupo protector' es $-B(OZ^1)(OH)$. En algunos modos de realización, el 'grupo protector' es $-B(OH)_2$. En algunos modos de realización, el 'grupo protector' es $-B(OH)_3$.

En otro aspecto de la presente divulgación están nucleótidos o sales de los mismos que tienen la fórmula (IV):



(IV)

35

en la que

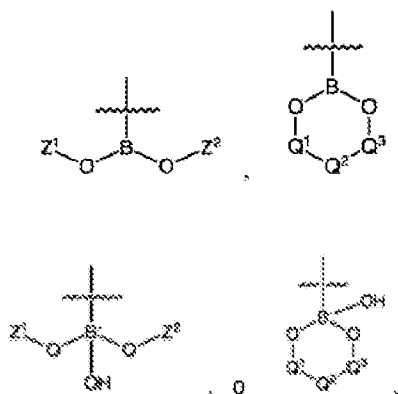
X es una nucleobase o una nucleobase marcada;

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R^1 es un enlace, $-CH_2-$, $-C(O)-O-$, $-C(O)-NR^a-$, $-C(O)-R^x-$;

R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-[(C(R^a)(R^b))_p-O]_q-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$, $-CN$ o $-NO_2$;

R^5 es $-(C(R^a)(R^b))_n-N_3$, $-(C(R^a)(R^b))_n-N^+C^-$, $-(C(R^a)(R^b))_n-CN$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:



Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

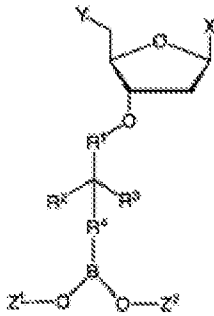
n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3;

con la condición de que cuando R^1 es un enlace y R^2 y R^3 son ambos H, R^5 no es una acida (por ejemplo, R^5 no es N_3).

En algunos modos de realización, R^5 es $-B(OZ^1)(OZ^2)$. En algunos modos de realización, R^5 es $-B(OZ^1)(OZ^2)$, donde Z^1 y Z^2 se seleccionan independientemente de un grupo alquilo C_1-C_4 . En algunos modos de realización, R^5 es $-B(OMe)(OMe)$. En algunos modos de realización, R^5 es $-B(OEt)(OEt)$. En algunos modos de realización, R^5 es $-B(OZ^1)(OH)$. En algunos modos de realización, R^5 es $-B(OH)_2$. En algunos modos de realización, el 'grupo protector' es $-B(OH)_3$.

En otro aspecto de la presente divulgación están nucleótidos o sales de los mismos que tienen la fórmula (VA):



(VA).

en la que

5 X es una nucleobase o una nucleobase marcada;

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

10 R^1 es un enlace, $-CH_2-$, $-C(O)-O-$, $-C(O)-NR^a-$, $-C(O)-R^x-$;

R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-[(C(R^a)(R^b))_p-O]_q-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$, $-CN$ o $-NO_2$;

15 R^4 es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-CH=CH-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u $-O-C(O)-aril-$;

20 Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

25 R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3.

30 En algunos modos de realización, Z^1 y Z^2 se seleccionan independientemente de metilo o etilo. En algunos modos de realización, Z^1 y Z^2 se seleccionan independientemente de metilo o etilo, y R^1 es un enlace. En algunos modos de realización, Z^1 y Z^2 se seleccionan independientemente de metilo o etilo, y R^1 es un enlace, y R^2 y R^3 son H.

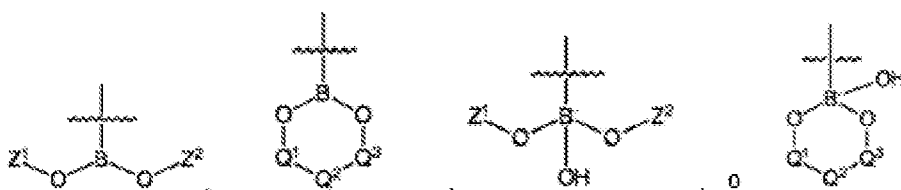
35 En algunos modos de realización, Z^1 es H y Z^2 se selecciona de metilo o etilo. En algunos modos de realización, Z^1 es H y Z^2 se selecciona de metilo o etilo, y R^1 es un enlace. En algunos modos de realización, Z^1 es H y Z^2 se selecciona independientemente de metilo o etilo, y R^1 es un enlace, y R^2 y R^3 son H.

40 La presente divulgación también proporciona procedimientos de secuenciación de una secuencia de ácido nucleico, utilizando el procedimiento cualquiera de los nucleótidos descritos en el presente documento (por ejemplo, los nucleótidos de cualquiera de las fórmulas (I) o (II)). En algunos modos de realización, la secuencia de nucleótidos de una porción de un ácido nucleico diana o fragmento del mismo se puede determinar usando una variedad de procedimientos y dispositivos. Los ejemplos no limitantes de procedimientos de secuenciación incluyen secuenciación electroforética, secuenciación por síntesis, secuenciación por ligación, secuenciación por hibridación, secuenciación de molécula individual y procedimientos de secuenciación ultrarrápida.

45 En otro aspecto de la presente divulgación está un procedimiento para determinar la secuencia de un polinucleótido monocatenario diana, que comprende supervisar la incorporación secuencial de nucleótidos complementarios, en el que cada uno de los nucleótidos complementarios tiene la estructura de cualquiera de las fórmulas (I), (II), (IIIA) - (IIID), (IV), (VA) - (VD) o (VIIIA) - (VIIIF) como se describe en el presente documento; y en el que la identidad de cada nucleótido complementario incorporado se determina a través de la detección de una marca (por ejemplo, un resto detectable) liberada de o unida a cada uno de los nucleótidos complementarios. En algunos modos de realización, la marca se libera de cada uno de los nucleótidos complementarios en las mismas condiciones usadas para desproteger el grupo de protección en 3' de nucleótidos complementarios de cualquiera de las fórmulas (I), (II), (IIIA) - (IIID), (IV), (VA) - (VD) o (VIIIA) - (VIIIF). En algunos modos de realización, la marca

se libera por escisión después de su detección. En algunos modos de realización, el procedimiento utiliza al menos cuatro nucleótidos complementarios diferentes de cualquiera de las fórmulas (I), (II), (IIIA) - (IIID), (IV), (VA) - (VD) o (VIIIA) - (VIIIF), teniendo cada uno de los al menos cuatro nucleótidos complementarios diferentes una nucleobase diferente. En algunos modos de realización, la marca se acopla a una nucleobase del nucleótido complementario. En algunos modos de realización, cada uno de los al menos cuatro nucleótidos complementarios diferentes tiene una marca diferente. En algunos modos de realización, cada uno de los al menos cuatro nucleótidos complementarios diferentes tiene una marca diferente, pero donde cada uno de los al menos cuatro nucleótidos complementarios diferentes comprende el mismo grupo protector, grupo bloqueador o resto R^5 (tal como esos grupos que se representan en las fórmulas (I), (II), (IIIA) - (IIID), (IV), (VA) - (VD) o (VIIIA) - (VIIIF) en el presente documento).

En otro aspecto de la presente divulgación está un procedimiento de secuenciación de una secuencia de ácido nucleico que comprende (a) realizar una reacción de polimerización con la ayuda de una única polimerasa (por ejemplo, ADN polimerasa o una variante o mutante de la misma) acoplada a un nanoporo, incorporando la reacción de polimerización uno de al menos cuatro nucleótidos protegidos en 3' diferentes en una hebra de polinucleótido creciente complementaria a una molécula de ácido nucleico monocatenario derivada de la muestra de ácido nucleico, comprendiendo cada nucleótido protegido en 3' individual de los al menos cuatro nucleótidos protegidos en 3' diferentes un grupo bloqueador que incluye un resto que tiene una estructura:



donde

Z^1 y Z^2 son independientemente H, un grupo alquilo C_1 - C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

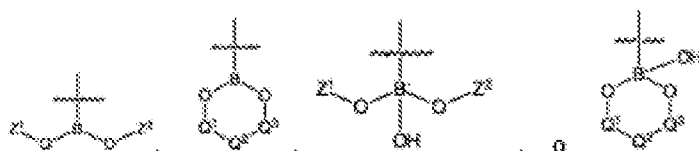
R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido; y

(b) detectar, con la ayuda del nanoporo, la incorporación del nucleótido protegido en 3' o un subproducto del mismo en la hebra de polinucleótido creciente; y (c) simultáneamente con o bien posterior a la etapa de detección de la incorporación del nucleótido protegido en 3' o un subproducto del mismo, desproteger el nucleótido protegido en 3' para proporcionar un nucleótido desprotegido en 3'. En algunos modos de realización, los al menos cuatro nucleótidos protegidos en 3' diferentes comprenden cada uno un resto detectable diferente. En algunos modos de realización, la etapa de detección comprende detectar los restos detectables diferentes asociados con cada uno de los al menos cuatro nucleótidos protegidos en 3' diferentes. En algunos modos de realización, el procedimiento comprende además correlacionar los restos detectables detectados asociados con cada uno de los al menos cuatro nucleótidos protegidos en 3' diferentes con un tipo de nucleótido. En algunos modos de realización, el procedimiento comprende además generar una secuencia de ácido nucleico de la molécula de ácido nucleico basándose en una evaluación de los restos detectables detectados durante la polimerización. En algunos modos de realización, la desprotección comprende poner en contacto el nucleótido protegido en 3' con un oxidante (por ejemplo, peróxido de hidrógeno). En algunos modos de realización, el procedimiento comprende además detectar un subproducto de la desprotección de los nucleótidos protegidos en 3'.

En otro aspecto de la presente divulgación está un procedimiento de secuenciación de una secuencia de ácido nucleico con la ayuda de un nanoporo, que comprende: (a) realizar una reacción de polimerización con la ayuda de una única polimerasa acoplada a un nanoporo, incorporando la reacción de polimerización uno de al menos cuatro nucleótidos protegidos en 3' diferentes en una hebra de polinucleótido creciente complementaria a una molécula de ácido nucleico monocatenaria derivada de la muestra de ácido nucleico, cada nucleótido protegido en 3' individual de los al menos cuatro nucleótidos protegidos en 3' diferentes se seleccionan de los nucleótidos de cualquiera de las fórmulas (I), (II), (IIIA) - (IIID), (IV), (VA) - (VD) o (VIIIA) - (VIIIF) como se describe en el presente documento; (b) detectar, con la ayuda del nanoporo, la incorporación del nucleótido protegido en 3' o un subproducto del mismo en la hebra de polinucleótido creciente; y (c) simultáneamente con o bien posterior a la etapa de detección de la incorporación del nucleótido protegido en 3' o el subproducto derivado del mismo, desproteger el nucleótido

protegido en 3' para proporcionar un nucleótido desprotegido en 3'.

En otro aspecto de la presente divulgación está un procedimiento de secuenciación de una muestra de ácido nucleico con la ayuda de un nanoporo, que comprende: (a) realizar una reacción de polimerización con la ayuda de una única enzima (por ejemplo, una polimerasa) acoplada al nanoporo, incorporando la reacción de polimerización uno de al menos cuatro nucleótidos protegidos en 3' diferentes en una hebra de polinucleótido creciente complementaria a una molécula de ácido nucleico monocatenario de la muestra de ácido nucleico, comprendiendo cada nucleótido protegido en 3' individual de los al menos cuatro nucleótidos protegidos en 3' diferentes un grupo protector que incluye un resto que tiene una estructura:



en la que

Z^1 y Z^2 son independientemente H, un grupo alquilo C_1 - C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

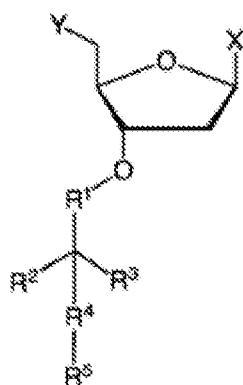
Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido; y

(b) detectar, con la ayuda del nanoporo, la incorporación del nucleótido protegido en 3' en la hebra de polinucleótido creciente; y (c) introducir un oxidante para desproteger un grupo protector en 3' y escindir una marca del nucleótido protegido en 3' para proporcionar un grupo OH en 3' (es decir, para proporcionar un nucleótido desprotegido). En algunos modos de realización, los al menos cuatro nucleótidos protegidos en 3' diferentes comprenden cada uno una marca diferente. En algunos modos de realización, la etapa de detección comprende detectar las marcas diferentes asociadas con cada uno de los al menos cuatro nucleótidos protegidos en 3' diferentes. En algunos modos de realización, el procedimiento comprende además correlacionar la marca detectada asociada con cada uno de los al menos cuatro nucleótidos protegidos en 3' diferentes con un tipo de nucleótido. En algunos modos de realización, el procedimiento comprende además generar una secuencia de ácido nucleico de la molécula de ácido nucleico basándose en una evaluación de las marcas detectadas durante la polimerización. En algunos modos de realización, la desprotección comprende poner en contacto el nucleótido protegido en 3' con un oxidante. En algunos modos de realización, el oxidante se selecciona del grupo que consiste en peróxido de hidrógeno, peryodato de sodio, perclorato de sodio, peroxinitrato u otro agente oxidante apropiado.

En otro aspecto de la presente divulgación está un procedimiento para secuenciar una molécula de ácido nucleico, comprendiendo el procedimiento: (a) obtener un chip que comprende una pluralidad de nanoporos individualmente direccionables, en el que cada nanoporo individualmente direccionable de la pluralidad de nanoporos individualmente direccionables comprende un nanoporo en una membrana que se dispone contigua a un electrodo, en el que el nanoporo se enlaza a una enzima y en el que cada nanoporo individualmente direccionable se adapta para detectar una marca que se libera de un nucleótido protegido en 3', en el que el nucleótido protegido en 3' (o cualquier sal del mismo) se representa por la fórmula (IIIA):



(BIA)_n

en la que

5 X es una nucleobase o una nucleobase marcada;

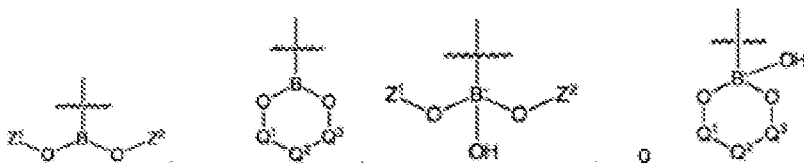
Y es $-\text{O}-\text{P}(\text{O})(\text{OH})-[\text{O}-\text{P}(\text{O})(\text{OH})]_z-\text{OH}$ u $-\text{O}-\text{P}(\text{O})(\text{OH})-[\text{O}-\text{P}(\text{O})(\text{OH})]_z-\text{O}$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

10 R¹ es un enlace, $-\text{CH}_2-$, $-\text{C}(\text{O})-\text{O}-$, $-\text{C}(\text{O})-\text{NR}^a-$ o $-\text{C}(\text{O})-\text{R}^x-$;

R² y R³ son cada uno independientemente H, un grupo alquilo C₁-C₆ saturado o insaturado, un grupo arilo o heteroarilo C₅-C₆, un halógeno, $-\text{C}(\text{R}^a)(\text{R}^b)_p-\text{O}]_q-(\text{R}^a)$, $-\text{C}(\text{O})-\text{OR}^a$, $-\text{C}(\text{O})-\text{N}(\text{R}^a)(\text{R}^b)$, $-(\text{C}(\text{R}^a)(\text{R}^b))_n-\text{N}^+\text{C}^-$, $-\text{CN}$ o $-\text{NO}_2$;

15 R⁴ es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-\text{CH}=\text{CH}-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u $-\text{O}-\text{C}(\text{O})-\text{aril}-$;

20 R⁵ es $-(\text{C}(\text{R}^a)(\text{R}^b))_n-\text{N}_3$, $-(\text{C}(\text{R}^a)(\text{R}^b))_n-\text{N}^+\text{C}^-$, $-(\text{C}(\text{R}^a)(\text{R}^b))_n-\text{CN}$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:



25 Z¹ y Z² son independientemente H, un grupo alquilo C₁-C₄ o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q¹ y Q³ son cada uno independientemente un enlace, $-\text{C}(\text{R}^e)(\text{R}^f)-$ o $-\text{C}(\text{O})-$;

30 Q² es un enlace o $-\text{C}(\text{R}^e)(\text{R}^f)_w-$, donde w es 1 o 2, u o-fenileno;

35 R^a y R^b son cada uno independientemente H o un grupo alquilo C₁-C₆ saturado;

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

n es 0 o un número entero que varía de 1 a 3; y

40 p y q son cada uno independientemente cero o un número entero que varía de 1 a 3;

siempre que cuando R¹ y R⁴ sean ambos enlaces y cuando R² y R³ sean ambos H, entonces R⁵ no es una acida (por ejemplo, R⁵ no es N₃);

45 (b) dirigir la molécula de ácido nucleico contigua a o en proximidad al nanoporo;

(c) con la ayuda de la enzima, polimerizar el nucleótido protegido en 3' a lo largo de la molécula de ácido nucleico para generar una hebra de polinucleótido que es complementaria a al menos una porción de la molécula de ácido nucleico; (d) desproteger el grupo protector en 3' del nucleótido protegido en 3' después de la incorporación del nucleótido protegido en 3' en la hebra de polinucleótido; y (e) detectar la marca liberada del nucleótido protegido en 3' con la ayuda del electrodo, en el que la marca liberada fluye a través del o en proximidad al nanoporo. En algunos modos de realización, la marca se libera antes de la desprotección del grupo protector en 3' del nucleótido protegido en 3'. En algunos modos de realización, la marca se libera simultáneamente con la desprotección del nucleótido protegido en 3'. En algunos modos de realización, la marca se desprotege antes de la liberación del grupo protector. En algunos modos de realización, la etapa de desprotección del nucleótido protegido en 3' comprende poner en contacto el nucleótido protegido en 3' con un oxidante, por ejemplo, peróxido de hidrógeno. En algunos modos de realización, se usan las mismas condiciones de reacción y/o reactivos para la desprotección del grupo protector en 3' y para liberar la marca.

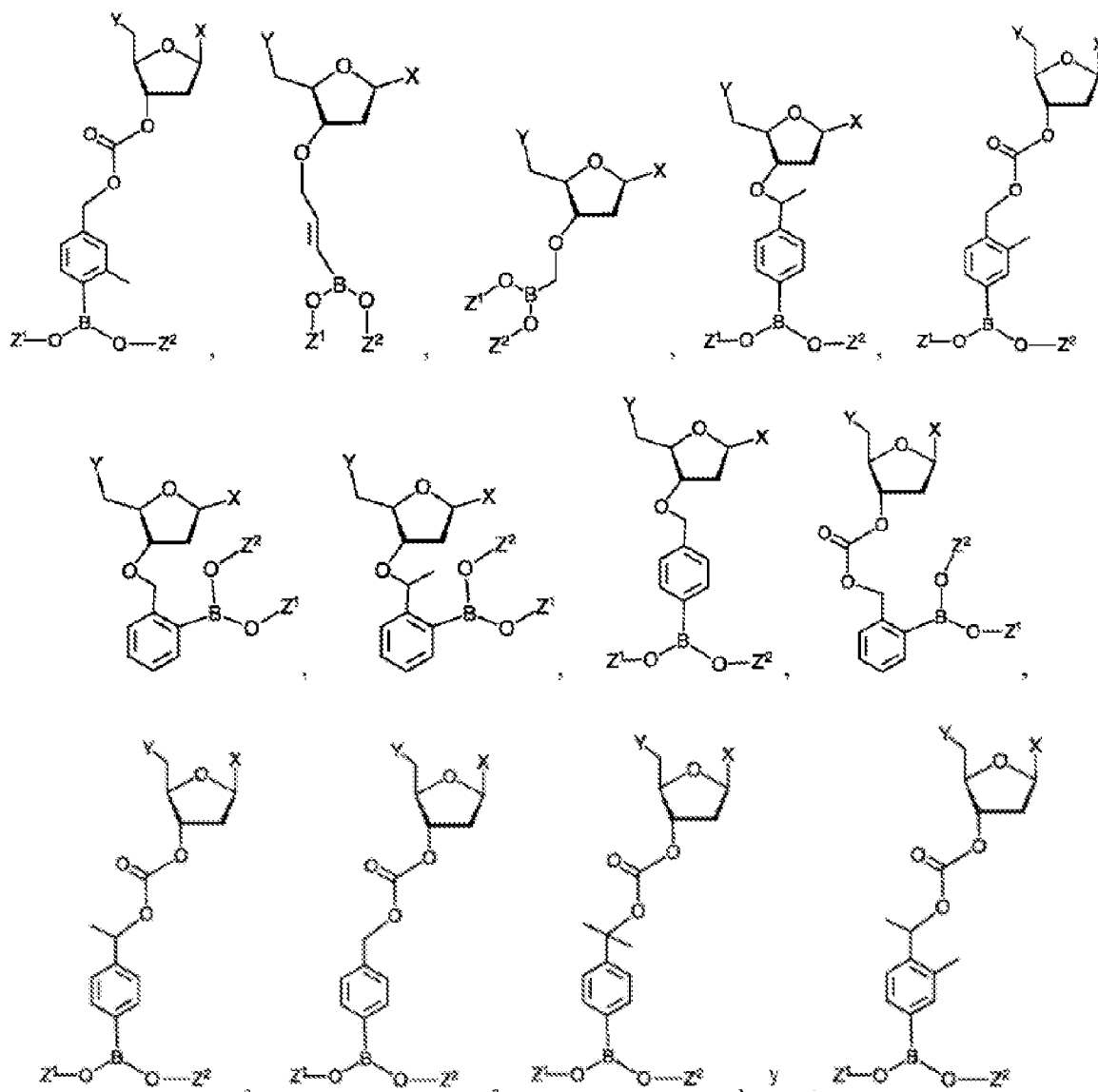
En otro aspecto de la presente divulgación está un kit que comprende: un biochip que comprende un sustrato semiconductor que tiene una pluralidad de pocillos con una densidad de al menos aproximadamente 250 pocillos/mm²; y un electrodo dispuesto en cada uno de la pluralidad de pocillos; y al menos uno de los nucleótidos de cualquiera de las fórmulas (I), (II), (IIIA) - (IIID), (IV), (VA) - (VD) o (VIIIA) - (VIIF) en el presente documento. En algunos modos de realización, el al menos uno de los nucleótidos comprende un resto derivado de ácido bórico o un derivado o análogo del mismo. En algunos modos de realización, la densidad de los pocillos es de al menos aproximadamente 500 pocillos/mm². En algunos modos de realización, al menos uno de los nucleótidos comprende un resto detectable, en el que el resto detectable se acopla a una nucleobase por medio de un conector escindible. En algunos modos de realización, el conector escindible se puede escindir usando un oxidante. En algunos modos de realización, el resto detectable incluye oligonucleótidos, oligopéptidos, polipéptidos, oligofosfatos, grupos PEG y otros restos.

En otro aspecto de la presente divulgación está un kit que comprende: (a) al menos uno de los nucleótidos de una cualquiera de las fórmulas (I), (II), (IIIA) - (IIID), (IV), (VA) - (VD) o (VIIIA) - (VIIF) en el presente documento; y (b) un oxidante. En algunos modos de realización, el kit comprende además una base. En algunos modos de realización, el al menos uno de los nucleótidos de una cualquiera de las fórmulas (I), (II), (IIIA) - (IIID), (IV), (VA) - (VD) o (VIIIA) - (VIIF) se proporciona en un primer recipiente; y en el que el oxidante se proporciona en un segundo recipiente.

En otro aspecto de la presente divulgación está un ensamblaje que comprende: un depósito, comprendiendo el depósito al menos uno de los nucleótidos de una cualquiera de las fórmulas (I), (II), (IIIA) - (IIID), (IV), (VA) - (VD) o (VIIIA) - (VIIF); un biochip que comprende un sustrato semiconductor que tiene una pluralidad de pocillos con una densidad de al menos aproximadamente 250 pocillos/mm²; un electrodo dispuesto en cada uno de la pluralidad de pocillos; y un contraelectrodo dispuesto en una superficie del depósito orientada hacia el biochip. En algunos modos de realización, al menos uno de los nucleótidos comprende un resto derivado de ácido bórico o un derivado o análogo del mismo. En algunos modos de realización, el depósito comprende además un subproducto derivado de la desprotección del nucleótido. En algunos modos de realización, el subproducto es ácido bórico o un derivado o análogo del mismo.

En otro aspecto de la presente divulgación está una composición que comprende: un polinucleótido y al menos uno de ácido bórico o un derivado o análogo del mismo, una metiluro de quinona, un acetal, un aldehído, un acrilaldehído y 4-metilen-2,5-ciclohexadien-1-ona. En algunos modos de realización, la composición comprende además un nucleótido de fórmula (VIA) (o una sal del mismo), o una cantidad mínima de un nucleótido de fórmula (VIA) (o una sal del mismo).

En otro aspecto de la presente divulgación está un nucleótido o sal del mismo, en el que el nucleótido se selecciona del grupo que consiste en:



donde

5

X es una nucleobase o una nucleobase marcada;

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5; y

10

Z¹ y Z² son independientemente H o grupo alquilo C₁-C₄.

En algunos modos de realización, Z¹ y Z² son independientemente H, metilo o etilo. En algunos modos de realización, al menos uno de Z¹ o Z² es H.

15

En algunos modos de realización, el nucleótido o sal del mismo comprende una marca seleccionada del grupo que consiste en uno o más de etilenglicol o un polímero derivado de etilenglicol, un aminoácido, un carbohidrato, un péptido, un tinte (incluyendo fluoróforos), un compuesto quimioluminiscente, una marca de masa, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un oligonucleótido, un oligonucleótido modificado, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alquénilo, un grupo alquinilo, un grupo ácido y cualquier combinación de los mismos. En algunos modos de realización, la nucleobase o sal de la misma comprende una marca basada en PEG. En algunos modos de realización, la nucleobase marcada comprende un conector escindible. En algunos modos de realización, el conector escindible se escinde con un oxidante (por ejemplo, peróxido de hidrógeno).

25

Breve descripción de las figuras

Para un entendimiento general de los rasgos característicos de la divulgación, se hace referencia a los dibujos. En los dibujos, se han usado números de referencia similares a lo largo de todos para identificar elementos idénticos.

La FIG. 1 ilustra la secuenciación de ADN de única molécula por un nanoporo con nucleótidos marcados en el polímero (140). Cada uno de los cuatro nucleótidos lleva una marca diferente. Durante la secuenciación por nanoporos, estas marcas, unidas por medio del fosfato terminal en 5' del nucleótido, se liberan en el nanoporo (130) una a la vez, donde producen firmas de bloqueo de corriente exclusivas (150).

La FIG. 2 ilustra un modo de realización de una célula 100 en un chip de secuenciación basada en nanoporos.

La FIG. 3A proporciona un diagrama de flujo que ilustra las etapas de la incorporación enzimática de un nucleótido protegido en un polinucleótido creciente, por ejemplo, un polinucleótido creciente que es complementario a una hebra de ácido nucleico que se secuencia de acuerdo con algunos modos de realización de la presente divulgación.

La FIG. 3B proporciona un diagrama de flujo que ilustra las etapas de la incorporación enzimática de un nucleótido marcado protegido en un polinucleótido creciente, por ejemplo, un polinucleótido creciente que es complementario a una hebra de ácido nucleico que se secuencia de acuerdo con algunos modos de realización de la presente divulgación.

La FIG. 3C ilustra un procedimiento para la secuenciación de ácidos nucleicos.

La FIG. 4 muestra un ejemplo de una señal generada por el paso de marcas a través de un nanoporo.

La FIG. 5 ilustra un esquema para desproteger un nucleótido protegido para proporcionar un nucleótido desprotegido que tiene un grupo 3'-OH libre.

La FIG. 6 ilustra una reacción de extensión de ADN usando análogos de nucleótidos con bases marcadas.

Descripción detallada

También se debe entender que, a menos que se indique claramente lo contrario, en cualquier procedimiento reivindicado en el presente documento que incluya más de una etapa o acto, el orden de las etapas o actos del procedimiento no se limita necesariamente al orden en que las etapas o actos del procedimiento se enumeran.

Como se usa en el presente documento, los términos en singular "un", "una" y "el/la" incluyen referentes al plural a menos que el contexto lo indique claramente de otro modo. De forma similar, se pretende que la palabra "o" incluya "y" a menos que el contexto lo indique claramente de otro modo. El término "incluye" se define de forma inclusiva, de modo que "incluye A o B" quiere decir incluyendo A, B o A y B.

Como se usa en el presente documento en la memoria descriptiva y en las reivindicaciones, se debe entender que "o" tiene el mismo significado que "y/o" como se define anteriormente. Por ejemplo, cuando se separan elementos en una lista, se interpretará que "o" o "y/o" son inclusivos, es decir, la inclusión de al menos uno, pero también incluyendo más de uno, de una serie o lista de elementos y, opcionalmente, elementos no enumerados adicionales. Solo los términos que claramente indiquen lo contrario, tales como "solo uno de" o "exactamente uno de" o, cuando se usa en las reivindicaciones, "que consiste en", se referirán a la inclusión de exactamente un elemento de una serie o lista de elementos. En general, solo se interpretará que el término "o" como se usa en el presente documento indica alternativas exclusivas (es decir, "uno o el otro, pero no ambos") cuando esté precedido por términos de exclusividad, tales como "uno... o bien", "uno de", "solo uno de" o "exactamente uno de". "Que consiste esencialmente en", cuando se usa en las reivindicaciones, tendrá su significado habitual como se usa en el campo del derecho de patentes.

Como se usan en el presente documento, los términos "que comprende", "que incluye", "que tiene" y similares se usan de manera intercambiable y tienen el mismo significado. De forma similar, se usan de manera intercambiable "comprende", "incluye", "tiene" y similares y tienen el mismo significado. Específicamente, cada uno de los términos se define de forma consecuente con la definición común del derecho de patentes de Estados Unidos de "que comprende" y, por lo tanto, se interpreta que es un término abierto que quiere decir "al menos lo siguiente" y también se interpreta que no excluye rasgos característicos, limitaciones, aspectos, etc., adicionales. Por tanto, por ejemplo, "un dispositivo que tiene los componentes a, b y c" quiere decir que el dispositivo incluye al menos los componentes a, b y c. De forma similar, la frase: "un procedimiento que implica las etapas a, b y c" quiere decir que el procedimiento incluye al menos las etapas a, b y c. Además, aunque las etapas y procesos se puedan explicar en el presente documento en un orden particular, el experto en la técnica reconocerá que el orden de las etapas y procesos puede variar.

Como se usa en el presente documento en la memoria descriptiva y en las reivindicaciones, la frase "al menos uno", en referencia a una lista de uno o más elementos, se debe entender que quiere decir al menos un elemento seleccionado de uno cualquiera o más de los elementos en la lista de elementos, pero que no incluye necesariamente al menos uno de cada uno de los elementos específicamente enumerados dentro de la lista de elementos y que no excluye ninguna combinación de elementos en la lista de elementos. Esta definición también permite que puedan estar presentes opcionalmente elementos distintos de los elementos identificados específicamente en la lista de elementos a los que se refiere la frase "al menos uno", bien relacionados o no relacionados con los elementos identificados específicamente. Por tanto, como un ejemplo no limitante, "al menos uno de A y B" (o, de forma equivalente, "al menos uno de A o B," o de forma equivalente "al menos uno de A y/o B") se puede referir, en un modo de realización, a al menos uno, incluyendo opcionalmente más de uno, A sin B presente (e incluyendo opcionalmente elementos distintos de B); en otro modo de realización, a al menos uno, incluyendo opcionalmente más de uno, B sin A presente (e incluyendo opcionalmente elementos distintos de A); aún en otro modo de realización, a al menos uno, incluyendo opcionalmente más de uno, A y al menos uno, incluyendo opcionalmente más de uno, B (e incluyendo opcionalmente otros elementos); etc.

La referencia a lo largo de toda esta memoria descriptiva a "un modo de realización" quiere decir que un rasgo característico, estructura o característica particular descrito en relación con el modo de realización se incluye en al menos un modo de realización. Por tanto, las apariciones de las frases "en un modo de realización" en diversos lugares a lo largo de toda esta memoria descriptiva no necesariamente se refieren todas al mismo modo de realización. Además, los rasgos característicos, estructuras o características particulares se pueden combinar de cualquier manera adecuada en uno o más modos de realización.

Como se usan en el presente documento, los términos "análogo" o "derivado" se usan de acuerdo con su significado habitual simple en Química y Biología y se refieren a un compuesto químico que es estructuralmente similar a otro compuesto (es decir, un compuesto denominado "de referencia") pero difiere en composición, por ejemplo, en el reemplazo de un átomo por un átomo de un elemento diferente, o en la presencia de un grupo funcional particular, o el reemplazo de un grupo funcional por otro grupo funcional, o la estereoquímica absoluta de uno o más centros quirales del compuesto de referencia. En consecuencia, un análogo es un compuesto que es similar o comparable en función y apariencia pero no en estructura u origen a un compuesto de referencia.

Como se usa en el presente documento, el término "alifático" quiere decir una cadena de hidrocarburo lineal o ramificada, que puede ser saturada o mono o poliinsaturada. Un grupo alifático insaturado contiene uno o más enlaces dobles y/o triples. Las ramificaciones de la cadena de hidrocarburo pueden incluir cadenas lineales así como elementos cíclicos no aromáticos. La cadena de hidrocarburo, que puede, a menos que se establezca de otro modo, tener cualquier longitud y contener cualquier número de ramificaciones. Tanto la cadena principal como las ramificaciones pueden contener además heteroátomos como por ejemplo B, N, O, P, S, Se o Si.

Como se usa en el presente documento, el término "alquilo", por sí mismo o como parte de otro sustituyente, quiere decir, a menos que se establezca de otro modo, una cadena lineal (es decir, no ramificada) o ramificada, o una combinación de las mismas, que puede ser totalmente saturada, mono- o poliinsaturada y puede incluir radicales di- y multivalentes, que tienen el número de átomos de carbono designado (es decir, C₁-C₁₀ quiere decir de uno a diez carbonos). Un "alquilo" no es ciclado. Los ejemplos de radicales hidrocarburos saturados incluyen, pero no se limitan a, grupos tales como metilo, etilo, n-propilo, isopropilo, n-butilo, t-butilo, isobutilo, sec-butilo, (ciclohexil)metilo, homólogos e isómeros de, por ejemplo, n-pentilo, n-hexilo, n-heptilo, n-octilo y similares. Un grupo alquilo insaturado es aquel que tiene uno o más enlaces dobles o enlaces triples. Los ejemplos de grupos alquilo insaturados incluyen, pero no se limitan a, vinilo, 2-propenilo, crotilo, 2-isopentenilo, 2-butenilo, 2,4-pentadienilo, 3-(1,4-pentadienilo), etinilo, 1- y 3-propinilo, 3-butenilo y los homólogos e isómeros superiores. Un alcoxi es un alquilo unido al resto de la molécula por medio de un átomo de oxígeno (-O-).

Como se usa en el presente documento, el término "alquilenos", por sí mismo o como parte de otro sustituyente, quiere decir, a menos que se establezca de otro modo, un radical divalente derivado de un alquilo, como se ejemplifica, pero no limitado por, -CH₂CH₂CH₂CH₂-. Típicamente, un grupo alquilo (o alquilenos) tendrá de 1 a 24 átomos de carbono, prefiriéndose en la presente divulgación aquellos grupos que tengan 10 o menos átomos de carbono. Un "alquilo inferior" o "alquilenos inferior" es un grupo alquilo o alquilenos de cadena más corta, que en general tiene ocho o menos átomos de carbono.

Como se usa en el presente documento, el término "alquilenos" quiere decir, en algunos modos de realización, un grupo alquilenos lineal, ramificado o cíclico que tiene de uno a tres átomos de carbono y puede ser, por ejemplo, un grupo metileno, un grupo etileno, un grupo propileno, un grupo isopropileno o un grupo c-propileno.

Como se usa en el presente documento, el término "aromático" quiere decir, a menos que se establezca de otro modo, un resto hidrocarburo cíclico plano de enlaces dobles conjugados, que puede ser un anillo único o incluir múltiples anillos condensados o enlazados de forma covalente. La cadena principal del resto hidrocarburo cíclico puede, a menos que se establezca de otro modo, tener cualquier longitud y contener cualquier número de heteroátomos, como por ejemplo N, O y S.

Como se usa en el presente documento, el término "heteroalquilo", por sí mismo o en combinación con otro término, quiere decir, a menos que se establezca de otro modo, una cadena lineal o ramificada estable, o combinaciones de las mismas, que consiste en al menos un átomo de carbono y al menos un heteroátomo seleccionado del grupo que consiste en O, N, P, Si y S, y en el que los átomos de nitrógeno, fósforo y azufre pueden estar opcionalmente oxidados, y el heteroátomo de nitrógeno puede estar opcionalmente cuaternizado. El/los heteroátomo(s) O, N, P, S y Si se pueden colocar en cualquier posición interior del grupo heteroalquilo o en la posición en la que el grupo alquilo se une al resto de la molécula. Un heteroalquilo no es ciclado. Los ejemplos incluyen, pero no se limitan a: $-\text{CH}_2-\text{CH}_2-\text{O}-\text{CH}_3$, $-\text{CH}_2-\text{CH}_2-\text{NH}-\text{CH}_3$, $-\text{CH}_2-\text{CH}_2-\text{N}(\text{CH}_3)-\text{CH}_3$, $-\text{CH}_2-\text{S}-\text{CH}_2-\text{CH}_3$, $-\text{CH}_2-\text{O}-\text{CH}_3$, $-\text{S}(\text{O})-\text{CH}_3$, $-\text{CH}_2-\text{CH}_2-\text{S}(\text{O})_2-\text{CH}_3$, $-\text{CH}=\text{CH}-\text{O}-\text{CH}_3$, $-\text{Si}(\text{CH}_3)_3$, $-\text{CH}_2-\text{CH}=\text{N}-\text{OCH}_3$, $-\text{CH}=\text{CH}-\text{N}(\text{CH}_3)-\text{CH}_3$, $-\text{O}-\text{CH}_3$, $-\text{O}-\text{CH}_2-\text{CH}_3$ y $-\text{CN}$. Hasta dos heteroátomos pueden ser consecutivos, tal como, por ejemplo, $-\text{CH}_2-\text{NH}-\text{OCH}_3$.

Como se usa en el presente documento, los términos "cicloalquilo" y "heterocicloalquilo", por sí mismos o en combinación con otros términos, quieren decir, a menos que se establezca de otro modo, versiones cíclicas de "alquilo" y "heteroalquilo", respectivamente. El cicloalquilo y el heterocicloalquilo no son aromáticos. Los cicloalquilos y heterocicloalquilos pueden estar además sustituidos, por ejemplo, con cualquiera de los sustituyentes descritos en el presente documento.

Cada uno de los términos anteriores (por ejemplo, "alquilo", "aromático", "heteroalquilo", "cicloalquilo", etc.) incluye formas tanto sustituidas como no sustituidas del radical indicado. En ese sentido, siempre que un grupo o resto se describe como "sustituido" u "opcionalmente sustituido" (o "que tiene opcionalmente" o "que comprende opcionalmente") ese grupo puede estar no sustituido o sustituido con uno o más de los sustituyentes indicados. Asimismo, cuando un grupo se describe como "sustituido o no sustituido" si está sustituido, el/los sustituyente(s) se puede(n) seleccionar de uno o más de los sustituyentes indicados. Si no se indican sustituyentes, quiere decir que el grupo "opcionalmente sustituido" o "sustituido" indicado puede estar sustituido con uno o más grupos seleccionados individual e independientemente de alquilo, alquenilo, alquinilo, cicloalquilo, cicloalquenilo, cicloalquinilo, arilo, heteroarilo, heteroalícicilo, aralquilo, heteroaralquilo, (heteroalícicil)alquilo, hidroxilo, hidroxilo protegido, alcoxi, ariloxi, acilo, mercapto, alquiltio, ariltio, ciano, cianato, halógeno, tiocarbonilo, O-carbamilo, N-carbamilo, O-tiocarbamilo, N-tiocarbamilo, C-amido, N-amido, S-sulfonamido, N-sulfonamido, C-carboxi, C-carboxi protegido, O-carboxi, isocianato, tiocianato, isotiocianato, nitro, sililo, sulfenilo, sulfenilo, sulfonilo, haloalquilo, haloalcoxi, trihalometanosulfonilo, trihalometanosulfonamido, un éter, amino (por ejemplo, un grupo amino monosustituido o un grupo amino disustituido) y derivados protegidos de los mismos. Cualquiera de los grupos anteriores puede incluir uno o más heteroátomos, incluyendo O, N o S. Por ejemplo, cuando un resto está sustituido con un grupo alquilo, ese grupo alquilo puede comprender un heteroátomo seleccionado de O, N o S (por ejemplo, $-(\text{CH}_2-\text{CH}_2-\text{O}-\text{CH}_2-\text{CH}_3)$).

Como se usa en el presente documento, los términos "acoplar" o "acoplamiento" se refieren a la unión, enlace (por ejemplo, enlace covalente) o enlace de una molécula o átomo a otra molécula o átomo.

Como se usa en el presente documento, los términos "halo" o "halógeno", por sí mismos o como parte de otro sustituyente, quieren decir, a menos que se establezca de otro modo, un átomo de flúor, cloro, bromo o yodo. Adicionalmente, quiere decir que términos tales como "haloalquilo" incluyen monohaloalquilo y polihaloalquilo.

Por ejemplo, el término "haloalquilo (C_1-C_4)" incluye, pero no se limita a, fluorometilo, difluorometilo, trifluorometilo, 2,2,2-trifluoroetilo, 4-clorobutilo, 3-bromopropilo y similares.

Como se usa en el presente documento, quiere decir que los términos "heteroátomo" o "heteroátomo de anillo" incluyen boro (B), oxígeno (O), nitrógeno (N), azufre (S), fósforo (P) y silicio (Si). En algunos modos de realización, un "anillo heterocíclico" puede comprender uno o más heteroátomos.

Como se usa en el presente documento, el "sustituyente inferior" o "grupo sustituyente inferior" quiere decir un grupo seleccionado de todos los sustituyentes descritos anteriormente para un "grupo sustituyente", en el que cada alquilo sustituido o no sustituido es un alquilo C_1-C_8 sustituido o no sustituido, cada heteroalquilo sustituido o no sustituido es un heteroalquilo de 2 a 8 miembros sustituido o no sustituido, cada cicloalquilo sustituido o no sustituido es un cicloalquilo C_5-C_7 sustituido o no sustituido, y cada heterocicloalquilo sustituido o no sustituido es un heterocicloalquilo de 5 a 7 miembros sustituido o no sustituido.

Como se usa en el presente documento, el término "nanoporo" se refiere a un poro, canal o paso formado o de otro modo proporcionado en una membrana. Un nanoporo se puede definir por una molécula (por ejemplo, proteína) en una membrana. Una membrana puede ser una membrana orgánica, tal como una bicapa lipídica, o una membrana sintética, tal como una membrana formada de un material polimérico. Como se usa en el presente documento, el término "polímero" se define como que incluye homopolímeros y copolímeros. El término "homopolímero" se define como un polímero derivado de una única especie de monómero. El término "copolímero" se define como un polímero derivado de más de una especie de monómero, incluyendo copolímeros que se obtienen por copolimerización de dos especies de monómeros, los obtenidos de tres especies de monómeros ("terpolímeros"),

los obtenidos de cuatro especies de monómeros ("cuaterpolímeros"), etc. El nanoporo se puede disponer contiguo o en proximidad a un circuito sensor, tal como, por ejemplo, un semiconductor complementario de óxido metálico (CMOS) o un circuito de transistor de efecto de campo (FET). Un nanoporo puede tener una anchura o diámetro característico del orden de 0,1 nanómetros (nm) a aproximadamente 1000 nm. Algunos nanoporos son proteínas.

5 La alfa hemolisina es un ejemplo de un nanoporo proteico. En algunos modos de realización, el nanoporo es un nanoporo en estado sólido (por ejemplo, un nanoporo en estado sólido es típicamente un orificio de tamaño nanométrico formado en una membrana sintética (normalmente SiNx o SiO₂)).

10 Como se usa en el presente documento, el término "complejo de secuenciación por nanoporos" se refiere a un nanoporo unido o acoplado a una enzima, por ejemplo, una polimerasa, que a su vez se asocia con un polímero, por ejemplo, un molde de polinucleótido. El complejo de secuenciación por nanoporos se sitúa en una membrana, por ejemplo, una bicapa lipídica, donde funciona para identificar componentes poliméricos, por ejemplo, nucleótidos o aminoácidos.

15 Como se usa en el presente documento, los términos "secuenciación por nanoporos" o "secuenciación basada en nanoporos" se refieren a un procedimiento que determina la secuencia de un polinucleótido con la ayuda de un nanoporo. En algunos modos de realización, la secuencia del polinucleótido se determina de manera dependiente del molde. Los procedimientos divulgados en el presente documento no se limitan a ningún procedimiento, sistema o dispositivo de secuenciación por nanoporos.

20 Como se usa en el presente documento, el término "ácido nucleico" se refiere a una molécula que comprende una o más subunidades de ácido nucleico. Un ácido nucleico puede incluir una o más subunidades (nucleobases naturales, sintéticas o modificadas) que incluyen, pero sin limitarse a, adenosina (A), citosina (C), guanina (G), timina (T) y uracilo (U). Los derivados de estas bases se ejemplifican en PCR Systems, Reagents and Consumables (Perkin Elmer Catalog 1996-1997, Roche Molecular Systems, Inc., Branchburg, N.J., EE. UU.). En algunos ejemplos, un ácido nucleico es ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN), o derivados de los mismos. Un ácido nucleico puede ser monocatenario o bicatenario. Un ácido nucleico puede incluir cualquier molécula de ácido nucleico, incluyendo, sin limitación, ADN, ARN e híbridos o variantes de los mismos.

30 Como se usa en el presente documento, el término "nucleobase" se refiere a un resto heterocíclico que se puede emparejar de forma no covalente con otra nucleobase. Una "nucleobase natural" o una "nucleobase no modificada" (usadas de manera intercambiable) se refieren a una nucleobase que no está modificada con respecto a su forma natural. Asimismo, una "nucleobase modificada" quiere decir cualquier sustitución y/o cambio de una nucleobase natural. Las modificaciones o sustituciones de nucleobases (o bases) son estructuralmente distinguibles de, pero funcionalmente intercambiables con, nucleobases naturales o sintéticas no modificadas. Las nucleobases tanto naturales como modificadas pueden participar en el enlace de hidrógeno. Dichas modificaciones de nucleobases pueden impartir estabilidad frente a nucleasa, afinidad de unión o alguna otra propiedad biológica beneficiosa a los compuestos antisentido. Las nucleobases modificadas incluyen nucleobases sintéticas y naturales tales como, por ejemplo, 5-metilcitosina (5-me-C). Determinadas sustituciones de nucleobases, incluyendo sustituciones de 5-metilcitosina, son en particular útiles para incrementar la afinidad de unión de un compuesto antisentido por un ácido nucleico diana. Por ejemplo, se ha demostrado que las sustituciones de 5-metilcitosina incrementan la estabilidad del dúplex de ácido nucleico en 0,6-1,2 °C (Sanghvi, Y. S., Crooke, S. T. y Lebleu, B., eds., *Antisense Research and Applications*, CRC Press, Boca Raton, 1993, pp. 276-278). Las nucleobases modificadas adicionales incluyen, pero no se limitan a, 5-hidroximetilcitosina, xantina, hipoxantina, 7-metilguanina, 2-aminoadenina, 2-aminopurina, iso-C, iso-G, tioT, tioG, 5,6-dihidouracilo, 6-metiladenina, 2-propilguanina y otros derivados alquílicos de adenina y guanina, 2-tiouracilo, 2-tiotimina y 2-tiocitosina, 5-halouracilo y citosina tales como 5-bromo, 5-trifluorometilo y otros uracilos y citosinas 5-sustituídos, 5-propinil (-C≡C-CH₃) uracilo y citosina y otros derivados alquílicos de bases pirimidínicas, 6-aza uracilo, citosina y timina, uracil-5-ilo (pseudouracilo), 4-tiouracilo, 8-halo, 8-amino, 8-tiol, 8-tioalquilo, 8-hidroxi y otras adeninas y guaninas 8-sustituídas, 7-metilguanina y 7-metiladenina, 2-F-adenina, 2-amino-adenina, 8-azaguanina y 8-azaadenina, 7-desazaguanina y 7-desazaadenina y 3-desazaguanina y 3-desazaadenina, 8-aza-7-desazaguanina y 8-aza-7-desazaadenina. Se describen nucleobases adicionales en Greco *et al.*, *Synthesis and site-specific incorporation of a simple fluorescent pyrimidine*, *Nature Protocols*, vol. 2, n.º 2, 2007; Dien *et al.*, *Progress Toward a Semi-Synthetic Organism with an Unrestricted Expanded Genetic Alphabet*, *J. Am. Chem. Soc.* 2018, 140, 16115-16123; Zhang *et al.*, *Evolution of Functional Six-Nucleotide DNA*, *J. Am. Chem. Soc.* 2015, 137, 6734-6737; Biondi *et al.*, *Artificially Expanded Genetic Information Systems for New Aptamer Technologies*, *Biomedicines* 2018, 6, 53; Liu *et al.*, *Helix-Forming Properties of Size-Expanded DNA, an Alternative Four-Base Genetic Form*, *J. Am. Chem. Soc.* 9 vol. 127, n.º 5, 2005, 1396-1402; Tor *et al.*, *Designing new isomorphous fluorescent nucleobase analogues: the thieno[3,2-d]pyrimidine core*, *Tetrahedron* 63 (2007) 3608-3614; Laos *et al.*, *Directed Evolution of Polymerases to Accept Nucleotides with Nonstandard Hydrogen Bond Patterns*, *Biochemistry* 2013, 52, 5288-5294; Krueger *et al.*, *Synthesis and Properties of Size-expanded DNAs: Toward Designed, Functional Genetic Systems*, *Acc Chem Res.* Febrero 2007, 40(2): 141-150; Srivatsan *et al.*, *A highly fluorescent nucleoside analog based on thieno[3,4-d]pyrimidine senses mismatched pairing*, *Org. Biomol. Chem.*, 2008, 6, 1334-1338; Kim *et al.*, *Synthesis and Properties of 5-Cyano-Substituted Nucleoside Analog with a Donor-Donor-Acceptor Hydrogen-Bonding Pattern*, *J. Org. Chem.* 2012, 77, 3664-3669; y Noe *et al.*, *Oligodeoxynucleotides Containing Multiple Thiophene-Modified Isomorphous Fluorescent Nucleosides*, *J. Org. Chem.* 2013, 78, 8123-8128.

Como se usa en el presente documento, el término "nucleósido" se refiere a una nucleobase unida de forma covalente a un glúcido, tal como ribosa o 2'-desoxirribosa.

Como se usa en el presente documento, el término "nucleótido" se refiere a un nucleósido unido de forma covalente a un fosfato o polifosfato, tal como adenosina 5'-monofosfato (AMP), adenosina 5'-difosfato (ADP), adenosina 5'-trifosfato (ATP), adenosina 5'-tetrafosfato o sus derivados 2'-desoxi. Como se usa en el presente documento, el término "oligonucleótido" se refiere a un oligómero de unidades de monómeros nucleotídicos o nucleosídicos en el que el oligómero incluye opcionalmente unidades de monómeros no nucleotídicos y/u otros grupos químicos unidos en posiciones internas y/o externas del oligómero. El oligómero puede ser natural o sintético y puede incluir oligonucleótidos naturales u oligómeros que incluyen nucleósidos con bases no naturales (o modificadas), restos de glúcido, enlaces fosfodiéster-análogo y/o quiralidades de unidades de monómeros alternativas y estructuras isoméricas (por ejemplo, enlace en de 5' a 2', L-nucleósidos, nucleósidos de anómero α).

Los oligonucleótidos ejemplares útiles como marcas detectables por nanoporos en la composición y los procedimientos de la presente divulgación incluyen las estructuras de marcas de oligonucleótidos mostradas en la tabla 4.

Como se usa en el presente documento, el término "polimerasa" se refiere a cualquier enzima que puede catalizar una reacción de polimerización. Los ejemplos de polimerasas incluyen, sin limitación, una polimerasa de ácido nucleico, una transcriptasa o una ligasa. Una polimerasa puede ser una enzima de polimerización. Una "ADN polimerasa" cataliza la polimerización de desoxinucleótidos. Una "ARN polimerasa" cataliza la polimerización de ribonucleótidos. Un polímero puede incluir una retrotranscriptasa, una enzima usada para generar ADN complementario (ADNc) a partir de un molde de ARN.

Como se usa en el presente documento, un "polinucleótido" es un polímero u oligómero que comprende al menos dos nucleótidos. Un polinucleótido u oligonucleótido puede comprender un polinucleótido u oligonucleótido de ADN, un polinucleótido u oligonucleótido de ARN o una o más secciones de polinucleótido u oligonucleótido de ADN y/o polinucleótido u oligonucleótido de ARN.

Como se usa en el presente documento, los términos "grupo reactivo" o "grupo funcional reactivo" se refieren a un grupo funcional que se puede asociar químicamente con, interactuar con, hibridar con, formar enlaces de hidrógeno con o acoplar con un grupo funcional de un resto diferente. En algunos modos de realización, una "reacción" entre dos grupos reactivos o dos grupos funcionales reactivos puede querer decir que se forma un enlace covalente entre dos grupos reactivos o dos grupos funcionales reactivos; o puede querer decir que los dos grupos reactivos o dos grupos funcionales reactivos se asocian entre sí, interactúan entre sí, se hibridan entre sí, forman enlaces de hidrógeno entre sí, etc.

En algunos modos de realización, la "reacción" incluye por tanto acontecimientos de unión, tales como la unión de un hapteno con un anticuerpo anti-hapteno, o una molécula huésped que se asocia con una molécula huésped supramolecular.

Como se usa en el presente documento, el término "secuenciación" se refiere a la determinación del orden y la posición de las bases en un ácido nucleico.

Como se usa en el presente documento, el término "marca" se refiere a un resto detectable 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), que se puede detectar con la ayuda de un nanoporo.

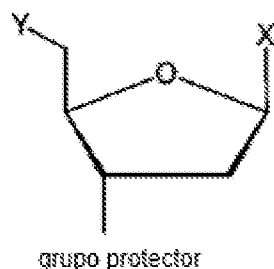
Los títulos proporcionados en el presente documento son solo por conveniencia y no interpretan el alcance o significado de los modos de realización divulgados.

Nucleótidos

La presente divulgación se dirige a nucleótidos y/o nucleósidos (o cualquier sal de los mismos) que incluyen un glúcido, por ejemplo, ribosa o desoxirribosa, que tienen un grupo protector extraíble en 3'. En algunos modos de realización, los nucleótidos y/o nucleósidos de la presente divulgación incluyen una nucleobase marcada, por ejemplo, una base que incluye un resto detectable. En algunos modos de realización, el grupo protector extraíble en 3' cumple al menos un criterio tal como (i) la capacidad de una enzima (por ejemplo, una polimerasa) para incorporar con exactitud y eficazmente los nucleótidos o nucleósidos que llevan los grupos protectores extraíbles en 3' en un oligonucleótido creciente; (ii) la disponibilidad de condiciones suaves para una desprotección rápida y cuantitativa, y (iii) la capacidad de la enzima para reiniciar la síntesis de oligómeros posterior a la etapa de desprotección. En algunos modos de realización, el grupo protector 3' se retira en condiciones no reductoras (por ejemplo, usando peróxido de hidrógeno u otro oxidante).

En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de acuerdo con la

presente divulgación tienen una estructura representada por la fórmula (I):



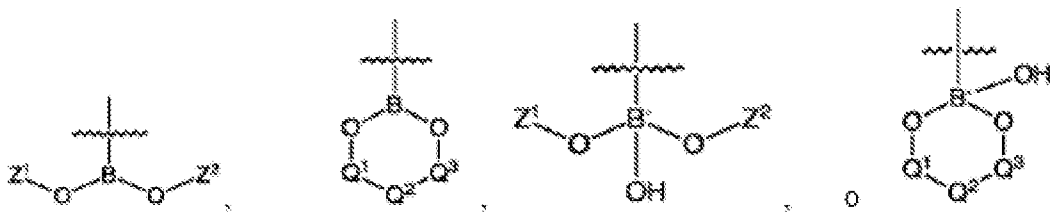
(I)

5 en la que

X es una nucleobase o una nucleobase marcada;

10 Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

'grupo protector' tiene la estructura:



15 donde

Z¹ y Z² son independientemente H, un grupo alquilo C₁-C₄ o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

20 Q¹ y Q³ son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

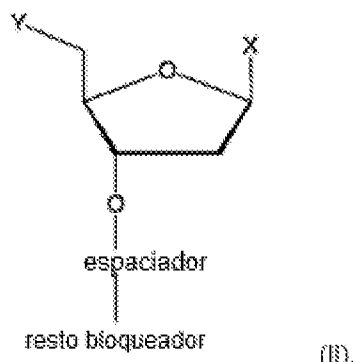
Q² es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2; y

25 R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido.

En algunos modos de realización, el oligonucleótido de Y se acopla por medio de su extremo 3'.

30 En algunos modos de realización, el 'grupo protector' es $-B(OZ^1)(OZ^2)$. En algunos modos de realización, uno de Z¹ o Z² es H, metilo o etilo. En algunos modos de realización, ambos de Z¹ y Z² se seleccionan independientemente de H, metilo o etilo. En algunos modos de realización, Z¹ y Z² se seleccionan independientemente de metilo o etilo. En algunos modos de realización, el 'grupo protector' es $-B(OH)_2$. En algunos modos de realización, el 'grupo protector' es $-B(OH)_3$.

35 En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de acuerdo con la presente divulgación tienen una estructura representada por la fórmula (II):

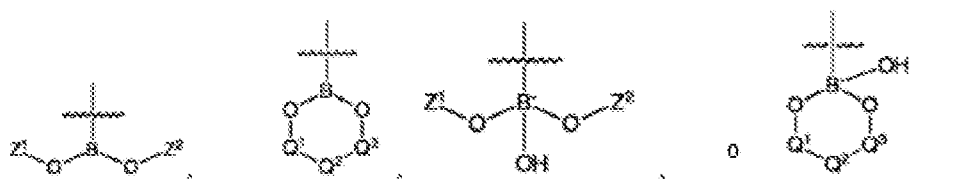


X es una nucleobase o una nucleobase marcada;

- 5 Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

'espaciador' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 16 átomos de carbono y opcionalmente sustituido con uno o más heteroátomos;

- 10 'resto bloqueador' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 20 átomos de carbono y opcionalmente sustituido con uno o más heteroátomos, y siempre que el 'resto bloqueador' incluya un grupo ácido, un grupo isonitrilo, un grupo ciano, un grupo heterocicloalquilo de 5 a 8 miembros que tiene al menos un heteroátomo seleccionado de O, N, S o Se, un
- 15 resto derivado de un 1,4-epoxi-1,4-dihidronaftaleno sustituido o no sustituido, o un grupo que tiene la estructura:



donde

- 20 Z¹ y Z² son independientemente H, un grupo alquilo C₁-C₄ o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q¹ y Q³ son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

- 25 Q² es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2; y

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

- 30 siempre que cuando el 'resto bloqueador' sea $-N_3$, el 'espaciador' no es $-CH_2-$.

En algunos modos de realización, el 'espaciador' es un grupo alifático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado que tiene entre 1 y 12 átomos de carbono y que incluye opcionalmente un grupo carbonilo. En otros modos de realización, el 'espaciador' es un grupo alifático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado que tiene entre 1 y 10 átomos de carbono e incluye

35 opcionalmente una cetona, un éster o una amida. En otros modos de realización más, el 'espaciador' es un grupo alifático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado que tiene entre 1 y 6 átomos de carbono e incluye un grupo carbonilo, por ejemplo, una cetona, un éster, un carbonato, un carbamato, un uretano o una amida.

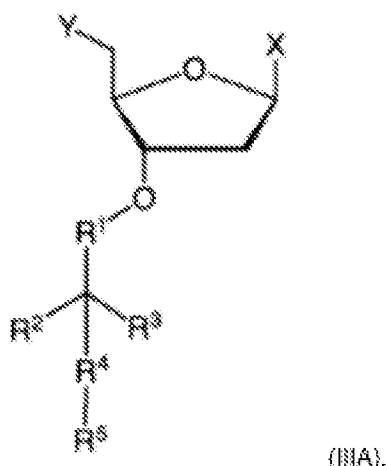
- 40 En algunos modos de realización, el 'resto bloqueador' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 20 átomos de carbono, y opcionalmente sustituido con uno o más heteroátomos, y siempre que el 'resto bloqueador' incluya un grupo isonitrilo, un grupo ciano, un grupo heterocicloalquilo de 5 a 8 miembros que tiene al menos un heteroátomo seleccionado de O, N, S o Se, un resto derivado de un 1,4-epoxi-1,4-dihidronaftaleno sustituido o no sustituido, o un resto derivado de ácido
- 45

borónico o un derivado o análogo del mismo. En algunos modos de realización, el 'resto bloqueador' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 16 átomos de carbono. En otros modos de realización, el 'resto bloqueador' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 12 átomos de carbono. Aún en otros modos de realización, el 'resto bloqueador' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 6 átomos de carbono.

En otros modos de realización, el 'resto bloqueador' incluye un grupo $-B(OZ^1)(OZ^2)$, donde Z^1 y Z^2 se seleccionan independientemente de H, metilo, etilo, isopropilo o *terc*-butilo. En otros modos de realización, el 'resto bloqueador' incluye un grupo $-B(OZ^1)(OZ^2)$, donde Z^1 y Z^2 se seleccionan independientemente de H, metilo o etilo. En otros modos de realización, el 'resto bloqueador' incluye un grupo $-B(OZ^1)(OZ^2)$, donde solo uno de Z^1 o Z^2 es H. Aún en otros modos de realización, el 'resto bloqueador' incluye un grupo $-B(OZ^1)(OZ^2)$, donde tanto Z^1 como Z^2 son H. En algunos modos de realización, el 'grupo protector' es $-B(OH)_3$.

En otros modos de realización, el 'resto bloqueador' incluye un grupo heterocicloalquilo de 5 a 8 miembros que incluye al menos un átomo de azufre o un átomo de selenio. En otros modos de realización, el 'resto bloqueador' incluye un grupo heterocicloalquilo de 5 a 8 miembros que incluye dos átomos de azufre o dos átomos de selenio, donde los dos átomos de azufre o los dos átomos de selenio están separados entre sí por un átomo de carbono (por ejemplo, $-S-C-S-$ o $-Se-C-Se-$). En algunos modos de realización, el grupo heterocicloalquilo de 5 a 8 miembros es un grupo 1,3-ditiano.

En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de acuerdo con la presente divulgación tienen una estructura como se representa por la fórmula (IIIA):



en la que

X es una nucleobase o una nucleobase marcada;

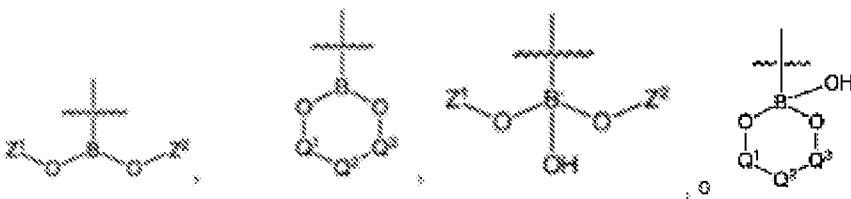
Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R^1 es un enlace, $-CH_2-$, $-C(O)-O-$, $-C(O)-NR^a-$ o $-C(O)-R^x-$;

R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-[(C(R^a)(R^b))_p-O]_q-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$, $-CN$ o $-NO_2$;

R^4 es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-CH=CH-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u $-O-C(O)-aril-$;

R^5 es $-(C(R^a)(R^b))_n-N_3$, $-(C(R^a)(R^b))_n-N^+C^-$, $-(C(R^a)(R^b))_n-CN$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:



donde

Z^1 y Z^2 son independientemente H, un grupo alquilo C_1 - C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1 - C_6 saturado;

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3;

siempre que cuando R^1 y R^4 sean ambos enlaces y cuando R^2 y R^3 sean ambos H, entonces R^5 no es una acida.

En algunos modos de realización, R^1 es un $-CH_2-$ y R^5 es $-B(OZ^1)(OZ^2)$. En algunos modos de realización, R^1 es un $-CH_2-$, R^4 es un enlace y R^5 es $-B(OZ^1)(OZ^2)$. En algunos modos de realización, R^1 es un $-CH_2-$, R^4 es un enlace y R^5 es $-B(OZ^1)(OZ^2)$, y al menos uno de R^2 o R^3 es H.

En algunos modos de realización, al menos uno de R^2 o R^3 es un H. En otros modos de realización, uno de R^2 o R^3 es $-[(C(R^a)(R^b))_p-O]_q-(R^a)$ o $-C(O)-OR^a-$. Aún en otros modos de realización, uno de R^2 o R^3 es $-[(C(R^a)(R^b))_p-O]_q-(R^a)$ o $-C(O)-OR^a-$ y R^a es metilo o etilo. En otros modos de realización, uno de R^2 o R^3 se selecciona de metilo, etilo, isopropilo o *tert*-butilo. Aún en otros modos de realización, al menos uno de R^2 o R^3 es H, R^5 es $-B(OZ^1)(OZ^2)$, y uno de Z^1 o Z^2 es H. En incluso otros modos de realización, tanto R^2 como R^3 son H, R^5 es $-B(OZ^1)(OZ^2)$, y ambos de Z^1 y Z^2 son H. En algunos modos de realización, R^1 es un enlace, R^2 y R^3 son ambos H, R^4 es un enlace y R^5 es $-B(OZ^1)(OZ^2)$. En algunos modos de realización, el 'grupo protector' es $-B(OH)_3$.

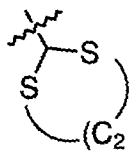
En algunos modos de realización, R^4 es un grupo fenilo o un grupo heterocicloalquilo de 5 o 6 miembros que está sustituido con al menos un resto seleccionado del grupo que consiste en $-O-R^6$, halógeno, un grupo alquilo C_1 - C_6 , un grupo alqueno C_2 - C_6 , $-C(O)-R^c$, $-C(O)-N(R^c)(R^d)$ o $-NO_2$, donde R^c y R^d son cada uno independientemente H o un grupo alquilo C_1 - C_6 saturado; y en los que R^6 es un grupo alquilo de C_1 a C_6 , $-N-(R^a)(R^b)$ o $-O(R^a)$. En otros modos de realización, R^6 es un grupo alquilo de C_1 a C_4 . Aún en otros modos de realización, R^6 se selecciona de metilo, etilo, isopropilo o *tert*-butilo.

En algunos modos de realización, R^4 es un grupo fenilo sustituido o no sustituido. En otros modos de realización, R^4 es un grupo fenilo sustituido con uno o más grupos aceptores de electrones (por ejemplo, nitro ($-NO_2$), ciano ($-CN$), carboxamida ($-C(O)NH_2$), trifluorometilo ($-CF_3$)). En algunos modos de realización, el grupo fenilo está sustituido con 1, 2 o 3 grupos aceptores de electrones. Aún en otros modos de realización, R^4 es un grupo fenilo sustituido con un grupo aceptor de electrones. En otros modos de realización, R^4 es un grupo fenilo no sustituido.

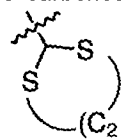
En algunos modos de realización, R^4 es un grupo heterocicloalquilo sustituido. En otros modos de realización, R^4 es un grupo heterocicloalquilo sustituido con uno o más grupos aceptores de electrones (por ejemplo, nitro ($-NO_2$), ciano ($-CN$), carboxamida ($-C(O)NH_2$), trifluorometilo ($-CF_3$)). Aún en otros modos de realización, R^4 es un grupo heterocicloalquilo sustituido con un grupo aceptor de electrones. En otros modos de realización, R^4 es un grupo heterocicloalquilo no sustituido.

En otros modos de realización, R^4 es un grupo fenilo o un grupo heterocicloalquilo de 5 o 6 miembros que está sustituido con al menos un sustituyente alquilo, donde cada sustituyente alquilo se selecciona de metilo, etilo, isopropilo o *tert*-butilo.

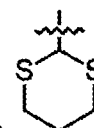
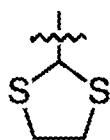
En algunos modos de realización, R^5 es $-B(OZ^1)(OZ^2)$, y Z^1 y Z^2 se seleccionan independientemente de H, metilo, etilo, isopropilo o *terc*-butilo. En otros modos de realización, R^5 es $-B(OZ^1)(OZ^2)$, y uno de Z^1 o Z^2 es H. Aún en otros modos de realización, R^5 es $-B(OZ^1)(OZ^2)$, y tanto Z^1 como Z^2 son H. En algunos modos de realización, R^4 es un enlace, R^5 es $-B(OZ^1)(OZ^2)$, y Z^1 y Z^2 se seleccionan independientemente de H, metilo, etilo, isopropilo o *terc*-butilo. En otros modos de realización, R^4 es un enlace, R^5 es $-B(OZ^1)(OZ^2)$, y tanto Z^1 como Z^2 son H.



En algunos modos de realización, R^5 es $(C_2 - C_6)$, donde $C_2 - C_6$ representa una cadena de alquilo de 2 a 6 carbonos saturada que puede estar sustituida o no sustituida. En algunos modos de realización, el grupo

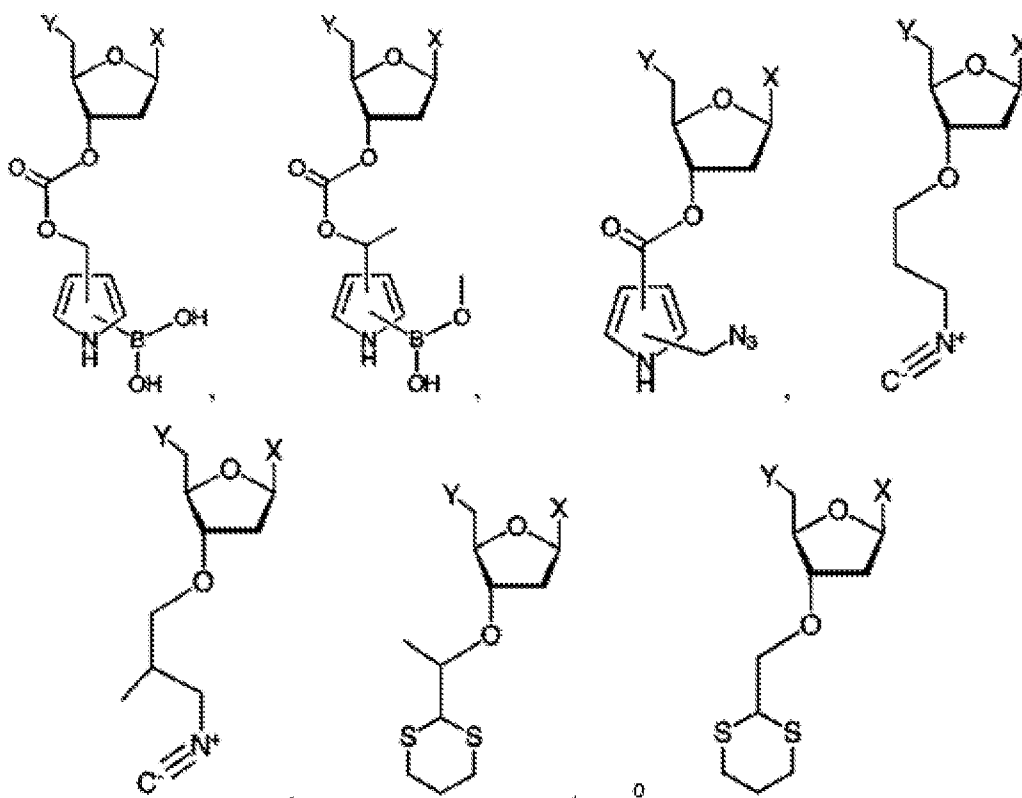


contiene al menos un sustituyente seleccionado de metilo, etilo, isopropilo o *terc*-butilo. En



10 algunos modos de realización, R^5 es . En otros modos de realización, R^5 es .

Los ejemplos no limitantes de los nucleótidos (incluyendo cualquier sal de los mismos) de fórmula (IIIA) incluyen:



15 en la que

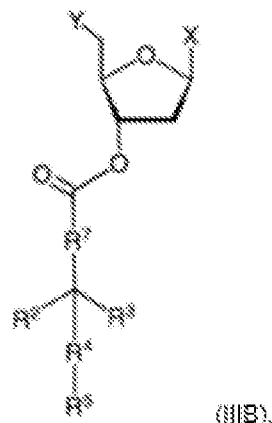
X es una nucleobase o una nucleobase marcada; e

20 Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5.

En algunos modos de realización, el grupo heterocíclico en los ejemplos identificados anteriormente está sustituido

en las posiciones 2,3 o las posiciones 2,5.

En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura como se representa por la fórmula (IIIB):



en la que

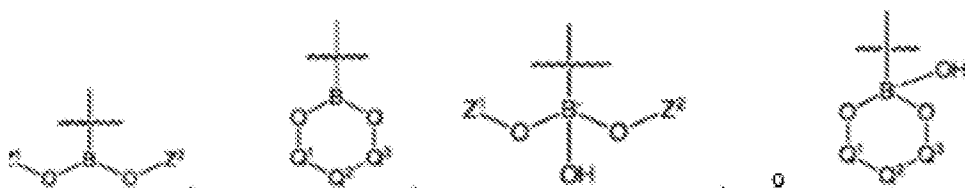
X es una nucleobase o una nucleobase marcada;

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-(C(R^a)(R^b))_p-O)_q-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$, $-CN$ o $-NO_2$;

R^4 es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-CH=CH-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u $-O-C(O)-aril-$;

R^5 es $-(C(R^a)(R^b))_n-N_3$, $-(C(R^a)(R^b))_n-N^+C^-$, $-(C(R^a)(R^b))_n-CN$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:



Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

R^7 es O o NR^a ;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

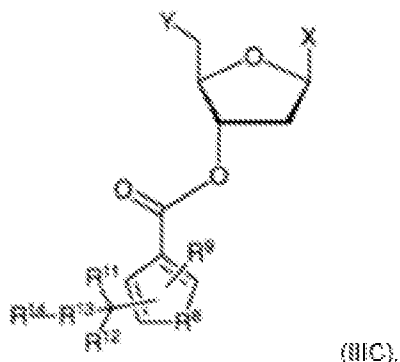
n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3.

En algunos modos de realización, al menos uno de R^2 o R^3 es H. En otros modos de realización, R^2 y R^3 son ambos H. Aún en otros modos de realización, uno de R^2 o R^3 es $-\text{[C(R}^a\text{)(R}^b\text{)]}_p\text{-O]}_q\text{-(R}^a\text{)}$ o $-\text{C(O)-OR}^a-$, y R^a es metilo o etilo. En algunos modos de realización, R^2 seleccionado de metilo, etilo, isopropilo o *terc*-butilo, y R^3 es H. Aún en otros modos de realización, al menos uno de R^2 o R^3 es H, R^5 es $-\text{B(OZ}^1\text{)(OZ}^2\text{)}$, y uno de Z^1 o Z^2 es H. Incluso en otros modos de realización, tanto R^2 como R^3 son H, R^5 es $-\text{B(OZ}^1\text{)(OZ}^2\text{)}$, y uno de Z^1 o Z^2 es H. Aún incluso en otros modos de realización, tanto R^2 como R^3 son H, R^5 es $-\text{B(OZ}^1\text{)(OZ}^2\text{)}$, y ambos de Z^1 y Z^2 son H.

En algunos modos de realización, R^7 es O. En otros modos de realización, R^7 es O, R^5 es $-\text{B(OZ}^1\text{)(OZ}^2\text{)}$, y Z^1 y Z^2 se seleccionan independientemente de H, metilo, etilo, isopropilo o *terc*-butilo. Aún en otros modos de realización, R^7 es O y R^5 es $-\text{B(OH)}_2$. Aún en otros modos de realización, R^7 es O, R^4 es fenilo o un fenilo sustituido (por ejemplo, uno sustituido con un grupo metilo, un grupo etilo, un grupo isopropilo o un grupo *terc*-butilo), y R^5 es $-\text{B(OH)}_2$. En otros modos de realización, R^7 es O, R^4 es un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido (por ejemplo, uno sustituido con un grupo metilo, un grupo etilo, un grupo isopropilo o un grupo *terc*-butilo), y R^5 es $-\text{B(OH)}_2$.

En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura representada por la fórmula (III C):



en la que

X es una nucleobase o una nucleobase marcada;

Y es $-\text{O-P(O)(OH)-[O-P(O)(OH)]}_z\text{-OH}$ u $-\text{O-P(O)(OH)-[O-P(O)(OH)]}_z\text{-O-oligonucleótido}$, donde z es 0 o un número entero que varía de 1 a 5;

R^8 es O, NR^e o S;

R^9 es H, halógeno, un grupo alquilo $\text{C}_1\text{-C}_4$, $-\text{C(O)-R}^a$, $-\text{N(R}^a\text{)(R}^b\text{)}$ o $-\text{NO}_2$;

R^{11} y R^{12} son cada uno independientemente H o un grupo alquilo $\text{C}_1\text{-C}_4$ lineal o ramificado;

R^{13} es un enlace, $-\text{CH}_2-$ o $-\text{CH}_2\text{-CH}_2-$;

R^{14} es $-\text{N}_3$, $-\text{C(R}^a\text{)(R}^b\text{)]}_n\text{-N}^+\text{C}^-$, $-\text{C(R}^a\text{)(R}^b\text{)]}_n\text{-CN}$ o $-\text{B(OZ}^1\text{)(OZ}^2\text{)}$, donde Z^1 y Z^2 se seleccionan independientemente de H, metilo, etilo, isopropilo o *terc*-butilo;

R^a y R^b son cada uno independientemente H o un grupo alquilo $\text{C}_1\text{-C}_6$ saturado;

R^e es H o un grupo alquilo $\text{C}_1\text{-C}_4$; y

n es 0 o un número entero que varía de 1 a 3.

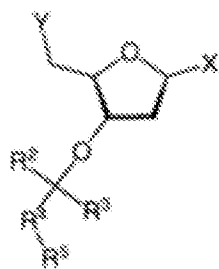
En algunos modos de realización, el anillo heterocíclico está sustituido en la posición 2,3 o en la posición 2,5, por ejemplo, $-\text{C(R}^{11}\text{)(R}^{12}\text{)-R}^{13}\text{-R}^{14}$ está presente en la posición 3 o la posición 5 del anillo heterocíclico.

En algunos modos de realización, al menos uno de R^{11} o R^{12} es H. En otros modos de realización, R^{11} se selecciona de metilo, etilo, isopropilo o *terc*-butilo, y R^{12} es H. Aún en otros modos de realización, al menos uno de R^{11} o R^{12} es H, R^{14} es $-\text{B(OZ}^1\text{)(OZ}^2\text{)}$, y uno de Z^1 o Z^2 es H. En otros modos de realización, tanto R^{11} como R^{12} son H, R^{14} es $-\text{B(OZ}^1\text{)(OZ}^2\text{)}$, Z^1 es H y Z^2 es un grupo alquilo $\text{C}_1\text{-C}_4$. En otros modos de realización, R^9 es H, uno de R^{11}

o R^{12} es H, R^{14} es $-B(OZ^1)(OZ^2)$, Z^1 es H y Z^2 es un grupo alquilo C_1-C_4 . Incluso en otros modos de realización, tanto R^{11} como R^{12} son H y R^{14} es $-B(OH)_2$.

5 En algunos modos de realización, al menos uno de R^{11} o R^{12} es H y R^{14} es N_3 . En otros modos de realización, R^{11} y R^{12} son ambos H y R^{14} es N_3 . En algunos modos de realización, R^{14} es N_3 . En algunos modos de realización, R^{13} es un enlace y R^{14} es N_3 . En algunos modos de realización, R^{13} es un enlace, R^{14} es N_3 y R^9 es H. En algunos modos de realización, R^{13} es un enlace, R^{14} es N_3 y R^8 es O. En otros modos de realización, R^{13} es un enlace, R^{14} es N_3 y R^8 es NR^e . Aún en otros modos de realización, R^{13} es un enlace, R^{14} es N_3 y R^8 es S.

10 En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura representada por la fórmula (IIID):



(IIID).

15 en la que

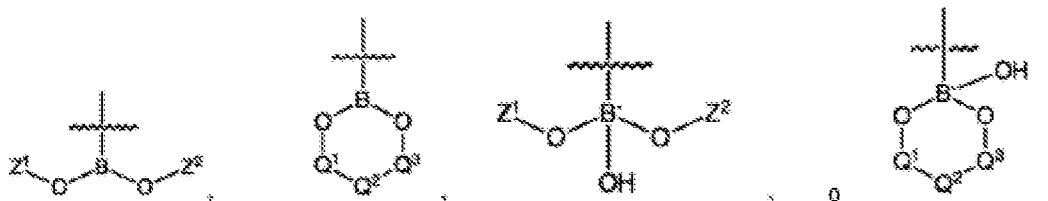
X es una nucleobase o una nucleobase marcada;

20 Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-[(C(R^a)(R^b))_p-O]_q-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$, $-CN$ o $-NO_2$;

25 R^4 es un enlace, un grupo fenilo sustituido o no sustituido, $-CH=CH-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido/no sustituido u $-O-C(O)-aril-$; y

30 R^5 es $-(C(R^a)(R^b))_n-N_3$, $-(C(R^a)(R^b))_n-N^+C^-$, $-(C(R^a)(R^b))_n-CN$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:



35 donde

Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

40 Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

45 R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3;

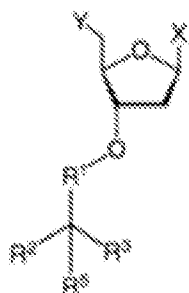
con la condición de que cuando R^2 y R^3 son H y R^4 es un enlace, entonces R^5 no es una acida.

En algunos modos de realización, R^4 es un enlace. En otros modos de realización, R^4 es un enlace y R^5 es $-B(OH)_2$. Aún en otros modos de realización, R^4 es un enlace, R^5 es $-(C(R^a)(R^b))_n-N_3$, y al menos uno de R^2 o R^3 es metilo, etilo, isopropilo o *terc*-butilo.

En algunos modos de realización, R^4 es un grupo fenilo, R^5 es $-B(OZ^1)(OZ^2)$, y uno de Z^1 o Z^2 es H. En otros modos de realización, R^4 es un grupo fenilo y R^5 es un $-B(OH)_2$. En otros modos de realización, R^4 es un grupo fenilo sustituido y R^5 es un $-B(OH)_2$, en el que el grupo fenilo está sustituido con al menos un grupo alquilo de C_1 a C_4 . Aún en otros modos de realización, R^4 es un grupo fenilo sustituido y R^5 es un $-B(OH)_2$, en el que el grupo fenilo está sustituido con un único sustituyente seleccionado de un grupo metilo, un grupo etilo, un grupo isopropilo o un grupo *terc*-butilo.

En algunos modos de realización, R^4 es $-CH=CH-$, R^5 es $-B(OZ^1)(OZ^2)$, y uno de Z^1 o Z^2 es H, metilo o etilo. En otros modos de realización, R^4 es $-CH=CH-$, R^5 es $-B(OZ^1)(OZ^2)$, y uno de Z^1 o Z^2 es H, metilo o etilo. En otros modos de realización, R^4 es $-CH=CH-$, R^5 es $-B(OZ^1)(OZ^2)$, y uno de Z^1 o Z^2 es H. Aún en otros modos de realización, al menos un R^2 o R^3 es H, R^4 es $-CH=CH-$, R^5 es $-B(OZ^1)(OZ^2)$ y uno de Z^1 o Z^2 es H. Aún en otros modos de realización, al menos un R^2 o R^3 es H, R^4 es $-CH=CH-$, R^5 es $-B(OZ^1)(OZ^2)$ y uno de Z^1 o Z^2 es H, metilo o etilo. En otros modos de realización, al menos un R^2 o R^3 es H, R^4 es $-CH=CH-$ y R^5 es un $-B(OH)_2$.

En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura como se representa por la fórmula (IV):



(IV),

en la que

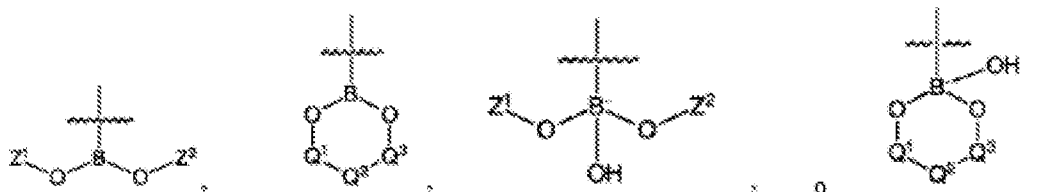
X es una nucleobase o una nucleobase marcada;

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R^1 es un enlace, $-CH_2-$, $-C(O)-O-$, $-C(O)-NR^a-$ o $-C(O)-R^x-$;

R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-(C(R^a)(R^b))_p-O-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$, $-CN$ o $-NO_2$;

R^5 es $-(C(R^a)(R^b))_n-N_3$, $-(C(R^a)(R^b))_n-N^+C^-$, $-(C(R^a)(R^b))_n-CN$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:



donde

Z^1 y Z^2 son independientemente H, un grupo alquilo C_1 - C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

5 Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1 - C_6 saturado;

10 R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

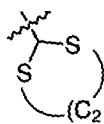
15 n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3;

20 con la condición de que cuando R^1 es un enlace y R^2 y R^3 son ambos H, R^5 no es una acida.

En algunos modos de realización, R^1 es un enlace. En otros modos de realización, al menos uno de R^2 o R^3 es H. Aún en otros modos de realización, R^1 es un enlace y al menos uno de R^2 o R^3 es H. En otros modos de realización, R^1 es un enlace, solo uno de R^2 o R^3 es H y R^5 es $-(C(R^a)(R^b))_n-N_3$.

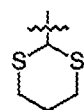
25 En algunos modos de realización, uno de R^2 o R^3 se selecciona de metilo, etilo, isopropilo o *terc*-butilo. En otros modos de realización, uno de R^2 o R^3 se selecciona de metilo, etilo, isopropilo o *terc*-butilo, R^5 es $-B(OZ^1)(OZ^2)$, y Z^1 y Z^2 se seleccionan independientemente de H, metilo, etilo, isopropilo o *terc*-butilo. En algunos modos de realización, R^5 es $-B(OZ^1)(OZ^2)$, y Z^1 y Z^2 se seleccionan independientemente de H, metilo, etilo, isopropilo o *terc*-butilo. En otros modos de realización, R^5 es $-B(OZ^1)(OZ^2)$ y uno de Z^1 o Z^2 es H. Aún en otros modos de realización, R^5 es $-B(OZ^1)(OZ^2)$ y tanto Z^1 como Z^2 son H. Aún incluso en otros modos de realización, R^5 es $-B(OZ^1)(OZ^2)$, tanto Z^1 como Z^2 son H y al menos uno de R^2 o R^3 es H.



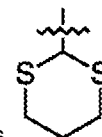
En algunos modos de realización, R^5 es es $(C_2 - C_6)$, donde C_2 - C_6 representa una cadena de alquilo de 2 a 6

35 carbonos saturada que puede estar sustituida o no sustituida. En algunos modos de realización, R^5 es . En

otros modos de realización, R^5 es . Aún en otros modos de realización, al menos un R^2 o R^3 es H y R^5 es



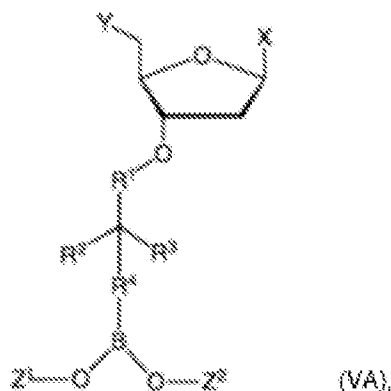
. Aún en otros modos de realización, R^1 es un enlace, al menos un R^2 o R^3 es H y R^5 es



40 En algunos modos de realización, R^1 es un $-CH_2-$ y R^5 es $-B(OZ^1)(OZ^2)$. En algunos modos de realización, R^1 es un $-CH_2-$, R^4 es un enlace y R^5 es $-B(OZ^1)(OZ^2)$. En algunos modos de realización, R^1 es un $-CH_2-$, R^4 es un enlace y R^5 es $-B(OZ^1)(OZ^2)$, y al menos uno de R^2 o R^3 es H.

En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura como se representa por la fórmula (VA):

45



en la que

5 X es una nucleobase o una nucleobase marcada;

Y es $-\text{O}-\text{P}(\text{O})(\text{OH})-\text{[O}-\text{P}(\text{O})(\text{OH})]_z-\text{OH}$ u $-\text{O}-\text{P}(\text{O})(\text{OH})-\text{[O}-\text{P}(\text{O})(\text{OH})]_z-\text{O-oligonucleótido}$, donde z es 0 o un número entero que varía de 1 a 5;

10 R^1 es un enlace, $-\text{CH}_2-$, $-\text{C}(\text{O})-\text{O}-$, $-\text{C}(\text{O})-\text{NR}^a-$ o $-\text{C}(\text{O})-\text{R}^x-$;

R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-\text{[(C(R}^a)(\text{R}^b))_p-\text{O}]_q-(\text{R}^a)$, $-\text{C}(\text{O})-\text{OR}^a$, $-\text{C}(\text{O})-\text{N}(\text{R}^a)(\text{R}^b)$, $-\text{(C(R}^a)(\text{R}^b))_n-\text{N}^+\text{C}^-$, $-\text{CN}$ o $-\text{NO}_2$;

15 R^4 es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-\text{CH}=\text{CH}-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u $-\text{O}-\text{C}(\text{O})-\text{aril}-$;

20 Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

25 R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

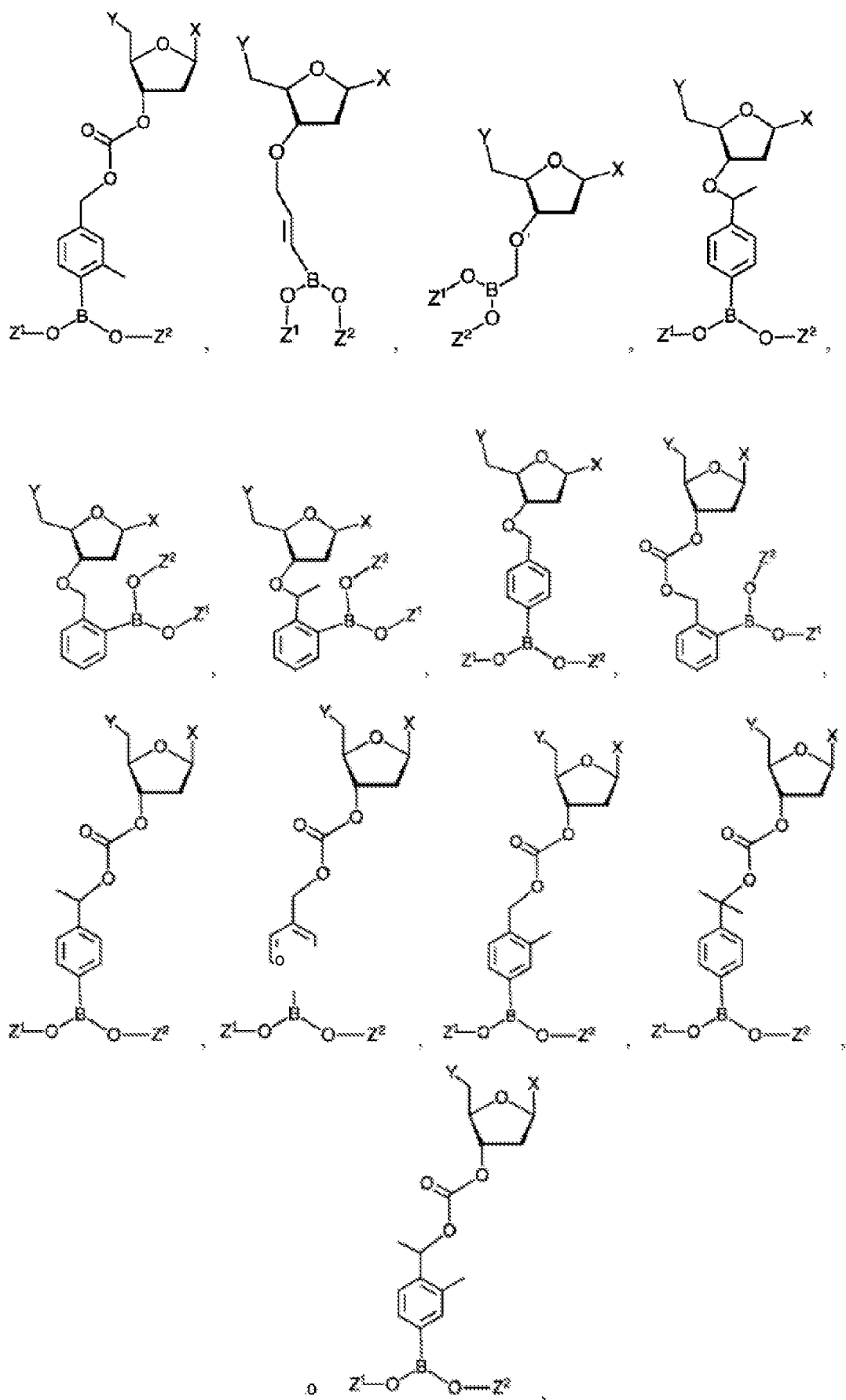
n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3.

30 En algunos modos de realización, R^2 y R^3 se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo. En otros modos de realización, R^1 es un enlace, y R^2 y R^3 se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo. Aún en otros modos de realización, R^1 es un enlace, R^2 y R^3 se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo, y R^4 es un enlace.

35 En algunos modos de realización, R^1 es $-\text{CH}_2-$, y Z^1 y Z^2 se seleccionan independientemente de H, metilo o etilo. En algunos modos de realización, R^1 es $-\text{CH}_2-$, Z^1 y Z^2 se seleccionan independientemente de H, metilo o etilo y R^4 es un enlace.

40 Los ejemplos no limitantes de los nucleótidos (incluyendo cualquier sal de los mismos) de fórmula (VA) incluyen los siguientes:



donde

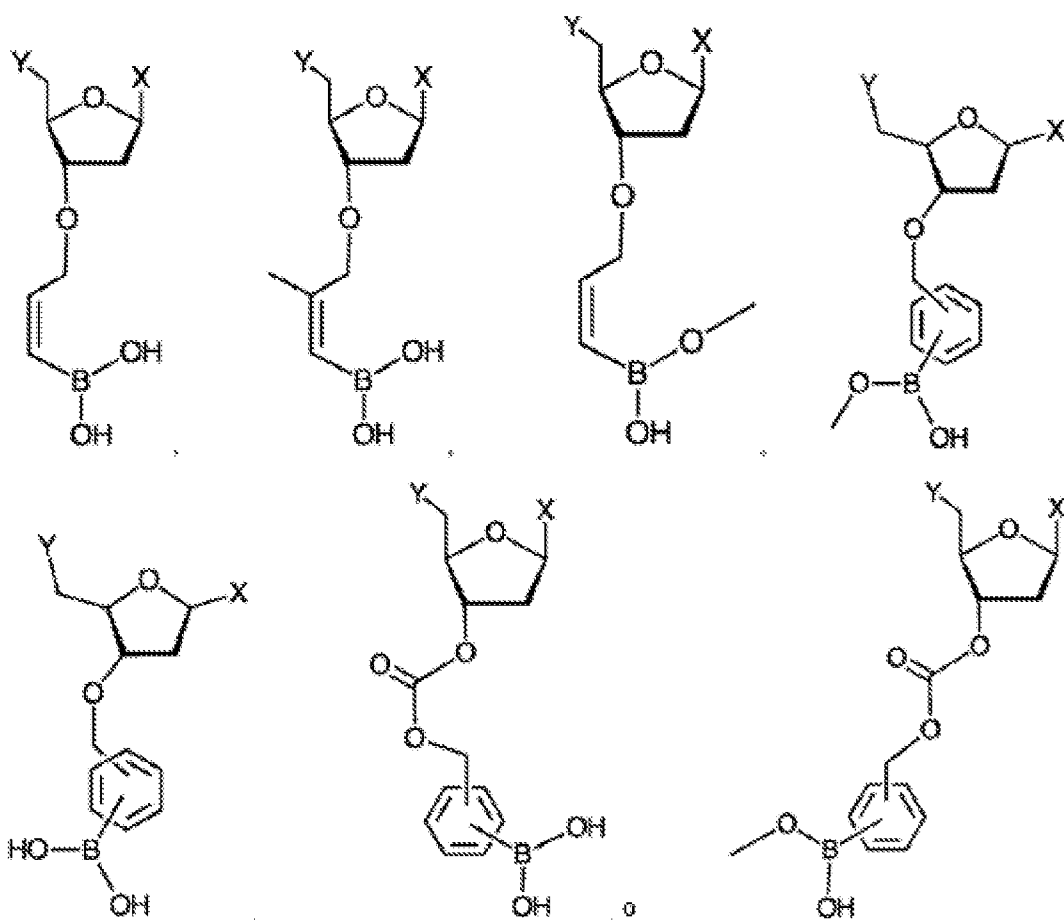
X es una nucleobase o una nucleobase marcada;

5 Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5; y

10 Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo.

En algunos modos de realización, Z^1 y Z^2 se seleccionan independientemente de H, metilo o etilo. En algunos modos de realización, Z^1 y Z^2 son ambos H.

15 Los ejemplos adicionales no limitantes de los nucleótidos (incluyendo cualquier sal de los mismos) de fórmula (VA) incluyen los siguientes:



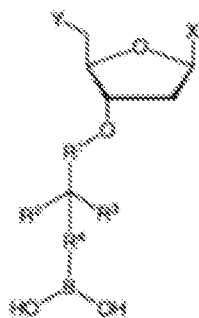
20 donde

X es una nucleobase o una nucleobase marcada; e

25 Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5.

En algunos modos de realización, el anillo heterocíclico está sustituido en las posiciones *orto* y *para* con un grupo $-B(OH)_2$.

30 En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura representada por la fórmula (VB):



(VB),

en la que

5 X es una nucleobase o una nucleobase marcada;

Y es $-\text{O}-\text{P}(\text{O})(\text{OH})-[\text{O}-\text{P}(\text{O})(\text{OH})]_z-\text{OH}$ u $-\text{O}-\text{P}(\text{O})(\text{OH})-[\text{O}-\text{P}(\text{O})(\text{OH})]_z-\text{O}$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

10 R^1 es un enlace, $-\text{CH}_2-$, $-\text{C}(\text{O})-\text{O}-$, $-\text{C}(\text{O})-\text{NR}^a-$ o $-\text{C}(\text{O})-\text{R}^x-$;

R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-\text{C}(\text{R}^a)(\text{R}^b)_p-\text{O}]_q-(\text{R}^a)$, $-\text{C}(\text{O})-\text{OR}^a$, $-\text{C}(\text{O})-\text{N}(\text{R}^a)(\text{R}^b)$, $-\text{C}(\text{R}^a)(\text{R}^b)_n-\text{N}^+\text{C}^-$, $-\text{CN}$ o $-\text{NO}_2$;

15 R^4 es un enlace, un grupo arilo de 5 o 6 miembros sustituido o no sustituido, $-\text{CH}=\text{CH}-$ o un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

20 R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

n es 0 o un número entero que varía de 1 a 3; y

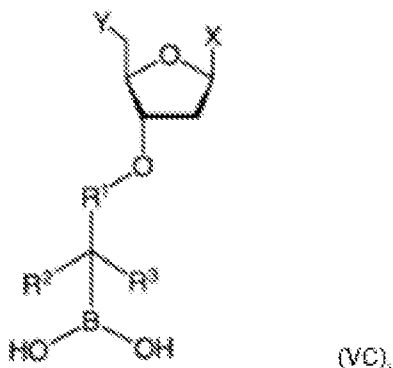
25 p y q son cada uno independientemente cero o un número entero que varía de 1 a 3.

En algunos modos de realización R^1 es $-\text{CH}_2-$. En algunos modos de realización R^1 es $-\text{CH}_2-$ y R^4 es un enlace.

30 En algunos modos de realización, R^2 y R^3 se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo. En otros modos de realización, R^1 es un enlace, y R^2 y R^3 se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo. Aún en otros modos de realización, R^1 es un enlace, R^2 y R^3 se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo, y R^4 es un enlace.

35 En algunos modos de realización, R^1 es un enlace, R^2 y R^3 se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo, y R^4 es un grupo fenilo que tiene al menos un sustituyente. En otros modos de realización, R^1 es un enlace, R^2 y R^3 se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo, y R^4 es un grupo fenilo que tiene al menos un sustituyente, en el que el al menos un sustituyente se selecciona de metilo, etilo, isopropilo o *terc*-butilo. Aún en otros modos de realización, R^1 es un enlace, R^2 y R^3 se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo, y R^4 es un grupo fenilo no sustituido.

40 En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura representada por la fórmula (VC):



en la que

- 5 X es una nucleobase o una nucleobase marcada;

Y es $-\text{O}-\text{P}(\text{O})(\text{OH})-[\text{O}-\text{P}(\text{O})(\text{OH})]_z-\text{OH}$ u $-\text{O}-\text{P}(\text{O})(\text{OH})-[\text{O}-\text{P}(\text{O})(\text{OH})]_z-\text{O}$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

- 10 R^1 es un enlace, $-\text{CH}_2-$, $-\text{C}(\text{O})-\text{O}-$, $-\text{C}(\text{O})-\text{NR}^a-$ o $-\text{C}(\text{O})-\text{R}^x-$;

R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-\text{C}(\text{R}^a)(\text{R}^b)_p-\text{O}]_q-(\text{R}^a)$, $-\text{C}(\text{O})-\text{OR}^a$, $-\text{C}(\text{O})-\text{N}(\text{R}^a)(\text{R}^b)$, $-\text{C}(\text{R}^a)(\text{R}^b)_n-\text{N}^+\text{C}^-$, $-\text{CN}$ o $-\text{NO}_2$;

- 15 R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

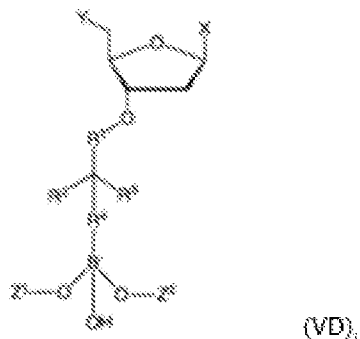
- 20 n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3.

En algunos modos de realización, R^1 es un enlace y al menos uno de R^2 o R^3 es H.

- 25 En algunos modos de realización R^1 es $-\text{CH}_2-$. En algunos modos de realización R^1 es $-\text{CH}_2-$ y al menos uno de R^2 o R^3 es H.

- 30 En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura como se representa por la fórmula (VD):



en la que

- 35 X es una nucleobase o una nucleobase marcada;

Y es $-\text{O}-\text{P}(\text{O})(\text{OH})-[\text{O}-\text{P}(\text{O})(\text{OH})]_z-\text{OH}$ u $-\text{O}-\text{P}(\text{O})(\text{OH})-[\text{O}-\text{P}(\text{O})(\text{OH})]_z-\text{O}$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R¹ es un enlace, -CH₂-, -C(O)-O-, -C(O)-NR^a- o -C(O)-R^x-;

R² y R³ son cada uno independientemente H, un grupo alquilo C₁-C₆ saturado o insaturado, un grupo arilo o heteroarilo C₅-C₆, un halógeno, -[(C(R^a)(R^b))_pO]_q-(R^a), -C(O)-OR^a, -C(O)-N(R^a)(R^b), -(C(R^a)(R^b))_n-N⁺C⁻, -CN o -NO₂;

R⁴ es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, -CH=CH-, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u -O-C(O)-aril-;

Z¹ y Z² son independientemente H, un grupo alquilo C₁-C₄ o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

R^a y R^b son cada uno independientemente H o un grupo alquilo C₁-C₆ saturado;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

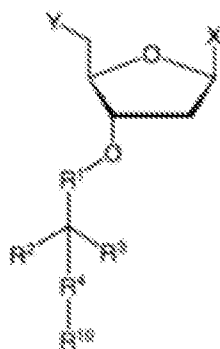
n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3.

En algunos modos de realización, R² y R³ se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo. En otros modos de realización, R¹ es un enlace, y R² y R³ se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo. Aún en otros modos de realización, R¹ es un enlace, R² y R³ se seleccionan independientemente de H, metilo, etilo, isopropilo y *terc*-butilo, y R⁴ es un enlace.

En algunos modos de realización, R¹ es -CH₂-, y Z¹ y Z² se seleccionan independientemente de H, metilo o etilo. En algunos modos de realización, R¹ es -CH₂-, Z¹ y Z² se seleccionan independientemente de H, metilo o etilo y R⁴ es un enlace.

En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura representada por la fórmula (VIA):



(VIA).

en la que

X es una nucleobase o una nucleobase marcada;

Y es -O-P(O)(OH)-[O-P(O)(OH)]_z-OH u -O-P(O)(OH)-[O-P(O)(OH)]_z-O-oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R¹ es un enlace, -CH₂-, -C(O)-O-, -C(O)-NR^a-, -C(O)-R^x-;

R² y R³ son cada uno independientemente H, un grupo alquilo C₁-C₆ saturado o insaturado, un grupo arilo o heteroarilo C₅-C₆, un halógeno, -[(C(R^a)(R^b))_pO]_q-(R^a), -C(O)-OR^a, -C(O)-N(R^a)(R^b), -(C(R^a)(R^b))_n-N⁺C⁻, -CN o -NO₂;

R⁴ es un enlace, un grupo fenilo sustituido o no sustituido, -CH=CH- o un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido, u -O-C(O)-aril-;

R¹⁰ es -OH, -(C(R^a)(R^b))_u-OH, -(C(R^a)(R^b))_u-C(O)H o un grupo cicloalquilo de 5 a 8 miembros que comprende dos grupos -S(O)- situados 1,3 entre sí;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1 - C_6 saturado;

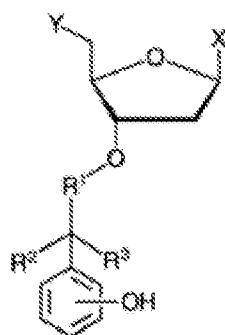
R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

n es 0 o un número entero que varía de 1 a 3;

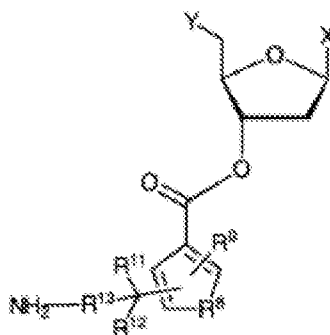
p y q son cada uno independientemente cero o un número entero que varía de 1 a 3; y

u es un número entero que varía de 1 a 3.

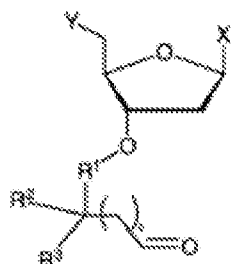
En algunos modos de realización, los nucleótidos (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura representada por una cualquiera de las fórmulas (VIB), (VIC), (VID), (VIE) y (VIF):



(VIB),



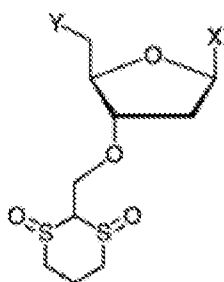
(VIC),



(VID),



(VIE) y



(VIF),

en las que

X es una nucleobase o una nucleobase marcada;

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R^1 es un enlace, $-CH_2-$, $-C(O)-O-$, $-C(O)-NR^a-$, $-C(O)-R^x-$;

R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1 - C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5 - C_6 , un halógeno, $-(C(R^a)(R^b))_p-O]_q-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$ $-CN$ o $-NO_2$;

R^8 es O, NR^e o S;

R^9 es H, halógeno, un grupo alquilo C_1-C_4 , $-C(O)-R^a$, $-N(R^a)(R^b)$ o $-NO_2$;

R^{11} y R^{12} son cada uno independientemente H o un grupo alquilo C_1-C_4 lineal o ramificado;

R^{13} es un enlace, $-CH_2-$ o $-CH_2-CH_2-$;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

R^e es H o un grupo alquilo C_1-C_4 ;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

n es 0 o un número entero que varía de 1 a 3;

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3; y

u es un número entero que varía de 1 a 3.

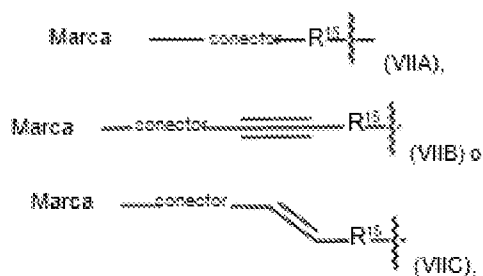
En algunos modos de realización, el anillo heterocíclico está sustituido en las posiciones *orto* y *para* con un grupo $-B(OH)_2$.

En algunos modos de realización, los compuestos de una cualquiera de las fórmulas (VIA) a (VIF) pueden estar presentes como una mezcla, tal como una mezcla que incluye ácido bórico u otro subproducto. En algunos modos de realización, los compuestos de una cualquiera de las fórmulas (VIA) a (VIF) pueden estar presentes como una mezcla con un nucleótido de fórmula (I) o fórmula (II).

Nucleobases y nucleobases marcadas

Como se describe en el presente documento, los nucleótidos o nucleósidos de la presente divulgación pueden incluir una nucleobase o una nucleobase marcada. Por "nucleobase" se quiere decir cualquier base nitrogenada adecuada para su inclusión dentro de un oligonucleótido (por ejemplo, una molécula de ARN o ADN), incluyendo bases naturales y bases sintéticas. En algunos modos de realización, la nucleobase se selecciona de adenina, citosina, guanina, timina y uracilo o un derivado o análogo de las mismas. En otros modos de realización, la nucleobase es 7-desazaguanina, 7-desazaadenina o 5-metilcitosina.

En algunos modos de realización, una nucleobase marcada tiene una estructura como se representa por cualquiera de las fórmulas (VIIA), (VIIB), (VIIC):



en las que

R^{15} es una nucleobase;

'conector' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 50 átomos de carbono y opcionalmente sustituido con uno o más heteroátomos; y

'marca' incluye una especie detectable.

Los procedimientos de unión de marcas a los nucleótidos del presente se divulgan en la publicación de solicitud de patente de Estados Unidos n.º 2014/0134616.

En algunos modos de realización, el 'conector' se puede unir en cualquier posición en la nucleobase siempre que

5 todavía se pueda llevar a cabo el emparejamiento de bases de Watson-Crick. En algunos modos de realización, y en el contexto de bases de purina, el conector se une por medio de la posición 7 de una 7-desazapurina, por medio de una purina modificada en 8, por medio de una adenina modificada en N-6 o una guanina modificada en N-2. En algunos modos de realización, y en el contexto de las pirimidinas, la unión es por medio de la posición 5 en citosina, timina o uracilo y la posición N-4 en citosina.

10 En algunos modos de realización, el 'conector' incluye desde entre 1 y 50 átomos de carbono. En otros modos de realización, el 'conector' incluye desde entre 2 y 25 átomos de carbono. Aún en otros modos de realización, el 'conector' incluye desde entre 5 y 20 átomos de carbono. En otros modos de realización, el 'conector' incluye desde entre 10 y 20 átomos de carbono. En algunos modos de realización, el 'conector' tiene un peso molecular que varía de aproximadamente 20 g/mol a aproximadamente 600 g/mol. En otros modos de realización, el 'conector' tiene un peso molecular que varía de aproximadamente 40 g/mol a aproximadamente 500 g/mol. En otros modos de realización, el 'conector' tiene un peso molecular que varía de aproximadamente 50 g/mol a aproximadamente 500 g/mol. En algunos modos de realización, el 'conector' tiene una longitud que varía desde entre aproximadamente 0,5 nm a aproximadamente 80 nm. En algunos modos de realización, el 'conector' tiene una longitud que varía desde entre aproximadamente 0,5 nm a aproximadamente 50 nm.

20 En algunos modos de realización, el 'conector' comprende un grupo que se puede escindir, por ejemplo, un grupo fotoescindible, un grupo enzimáticamente escindible, un grupo químicamente escindible, un grupo escindible a determinados pH. El uso del término "conector escindible" no quiere decir que implica que se requiera que todo el conector se retire de la nucleobase. Más bien, se puede localizar un sitio de escisión dentro del conector en una posición en el conector que asegure que parte del conector permanezca unida a la nucleobase después de la escisión. El uso de un conector escindible asegura que la marca se pueda retirar, si se requiere, después de la detección, evitando cualquier señal de interferencia con cualquier nucleótido marcado incorporado posteriormente. Los conectores escindibles son conocidos en la técnica y se puede aplicar química convencional para unir un conector a una nucleobase y una marca. El conector se puede escindir por cualquier procedimiento adecuado, incluyendo exposición a ácidos, bases, nucleófilos, electrófilos, radicales, metales, agentes reductores u oxidantes, luz, temperatura, enzimas, etc. Los conectores adecuados se pueden adaptar a partir de grupos bloqueadores químicos estándar, como se divulga en Greene y Wuts, *Protective Groups in Organic Synthesis*, John Wiley & Sons. Otros conectores escindibles adecuados usados en síntesis en fase sólida se divulgan en Guillier *et al.* (Chem. Rev. 100:2092-2157, 2000).

35 En algunos modos de realización, se cree que el conector puede comprender opcionalmente una o más unidades espaciadoras. El espaciador distancia la nucleobase del sitio de escisión, marca o conector. La longitud del conector no es importante siempre que la marca se mantenga a una distancia suficiente del nucleótido para no interferir con ninguna interacción entre el nucleótido y una enzima, por ejemplo, polimerasa. En algunos modos de realización, una nucleobase marcada comprende la estructura -[nucleobase]-[conector]-[extensor]_y-[marca], donde el 'extensor' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 60 átomos de carbono y opcionalmente sustituido con uno o más heteroátomos; e y es 0, 1 o 2.

45 Los conectores adecuados incluyen, pero no se limitan a, conectores de disulfuro, conectores lábiles a ácidos, incluyendo conectores de dialcoxibencilo, conectores de Sieber, conectores de indol, conectores de Sieber de t-butilo, conectores escindibles de forma electrófila, conectores escindibles de forma nucleófila, conectores fotoescindibles, escisión en condiciones reductoras, condiciones oxidantes, escisión por medio del uso de conectores de anclaje de liberación controlada y escisión por mecanismos de eliminación.

50 Los conectores fotoescindibles se han usado ampliamente en la química de carbohidratos. En algunos modos de realización, la luz requerida para activar la escisión no afecta a los otros componentes de los nucleótidos modificados. Por ejemplo, si se usa un fluoróforo como marca, en algunos modos de realización el fluoróforo se elige de modo que absorba luz de una longitud de onda diferente a la requerida para escindir la molécula conectora. Los conectores adecuados incluyen aquellos basados en compuestos de o-nitrobencilo y compuestos de nitroveratrilo (a modo de ejemplo, se puede usar un alcohol 5-hidroxi-2-nitrobencilo como material de partida). En un modo de realización, el conector fotoescindible es un resto 2-nitrobencilo. También se pueden usar conectores basados en la química de benzoína (Lee *et al.*, J. Org. Chem. 64:3454-3460, 1999).

60 Por ejemplo, se puede introducir un grupo en el conector que se puede escindir tras la exposición a una fuente de radiación electromagnética que tiene una longitud de onda de entre aproximadamente 200 nm a aproximadamente 400 nm (UV) o entre aproximadamente 400 nm a aproximadamente 800 nm (visible). En algunos modos de realización, el grupo fotoescindible con luz UV o visible se selecciona del grupo que consiste en grupos arilcarbonilmetilo (incluyendo 4-acetil-2-nitrobencilo, dimetilfenacilo (DMP), 2-(alcoximetil)-5-metil- α -cloroacetofenonas, 2,5-dimetilbenzoil-oxiranos y grupos benzoína: 3',5'-dimetoxibenzoína (DMB)), grupos o-nitrobencilo (incluyendo 1-(2-nitrofenil)etilo (NPE), 1-(metoximetil)-2-nitrobenceno, 4,5-dimetoxi-2-nitrobencilo (DMNB); grupos α -carboxinitrobencilo (α -CNB), o-nitro-2-fenetiloxicarbonilo, incluyendo derivados de 1-(2-nitrofenil)etiloxicarbonilo y 2-nitro-2-fenetilo, y o-nitroanilidas tales como 5-bromo-7-nitroindolinas aciladas); grupos cumarin-4-il-metilo (incluyendo derivados de 7-metoxicumarina);

xantenos 9-sustituídos y grupos arilmetilo (incluyendo grupos o-hidroxiarilmetilo).

En algunos modos de realización, se puede introducir un grupo en el conector que se puede escindir tras la exposición a una fuente de radiación electromagnética que tiene una longitud de onda de entre aproximadamente 700 nm a aproximadamente 1000 nm. Los grupos fotoescindibles de infrarrojo cercano adecuados incluyen grupos cianina, incluyendo heptametinocianinas sustituidas con dialquilamina C4.

Aún en otros modos de realización, el 'conector' incluye grupos químicamente escindibles que se pueden escindir por diferentes reactivos químicos, incluyendo agentes reductores o por cambios inducidos en el pH (por ejemplo, escisión del grupo a un pH de menos de aproximadamente 7). Los grupos químicamente escindibles adecuados incluyen grupos basados en disulfuro; grupos diazobenceno (incluyendo estructuras de ácido 2-(2-alcoxi-4-hidroxi-fenilazo)benzoico, sensibles a ditionito de sodio); grupos basados en enlaces éster (pH alto); y conectores sensibles a ácidos (tales como conector de dialcoxidifenilsilano o acilhidrazona). Se puede escindir un conector escindible de diol vecinal por NaIO_4 , tal como se describe en "A simple and effective cleavable linker for chemical proteomics applications", *Mol Cell Proteomics*, enero de 2013; 12(1): 237-44. doi: 10.1074/mcp.M112.021014. Publicación electrónica del 1 de octubre de 2012. Los grupos escindibles enzimáticamente adecuados incluyen grupos escindibles con tripsina y grupos escindibles con proteasa V8. Otros conectores que se pueden utilizar en cualquiera de los nucleótidos de la presente divulgación se divulgan en la publicación de solicitud de patente de Estados Unidos n.º 2018/0057870.

Se cree que los conectores escindidos de forma electrófila se escinden por protones e incluyen escisiones sensibles a los ácidos. Los conectores adecuados incluyen los sistemas bencílicos modificados tales como tritilo, ésteres p-alcoxibencílicos y de p-alcoxibencilamidas. Otros conectores adecuados incluyen grupos *tert*-butiloxycarbonilo (Boc) y el sistema acetal. Para la preparación de moléculas conectoras adecuadas también se puede considerar el uso de metales tífilos, tales como níquel, plata o mercurio, en la escisión de tioacetal u otros grupos protectores que contienen azufre.

La escisión nucleófila también es un procedimiento bien reconocido en la preparación de moléculas conectoras. Se pueden usar grupos tales como ésteres que son lábiles en agua (es decir, que se pueden escindir simplemente a pH básico) y grupos que son lábiles para nucleófilos no acuosos. Los iones fluoruro se pueden usar para escindir enlaces silicio-oxígeno en grupos tales como triisopropilsilano (TIPS) o t-butildimetilsilano (TBDMS).

Existen muchos conectores conocidos que son susceptibles a la escisión reductora. La hidrogenación catalítica usando catalizadores basados en paladio se ha usado para escindir grupos bencilo y benciloxycarbonilo. La reducción de enlaces disulfuro también es conocida en la técnica.

Los enfoques basados en oxidación son bien conocidos en la técnica. Estos incluyen la oxidación de grupos p-alcoxibencilo y la oxidación de conectores de azufre y selenio. El uso de yodo acuoso para escindir disulfuros y otros conectores basados en azufre o selenio también está dentro del alcance de la divulgación. En algunos modos de realización, el conector comprende un grupo $-\text{CH}_2-\text{O}-\text{CH}[\text{B}(\text{OH})_2]-\text{CH}_2-$ conector. En algunos modos de realización, los conectores comprenden conectores basados en 1,2-diol ($-\text{CH}(\text{OH})-\text{CH}(\text{OH})-$).

Los conectores de anclaje de liberación controlada son aquellos que se escinden en dos etapas. En algunos modos de realización, la primera etapa es la generación de un centro nucleófilo reactivo seguido de una segunda etapa que implica una ciclación intramolecular que da como resultado la escisión. Por ejemplo, los enlaces de éster levulínico se pueden tratar con hidracina o fotoquímica para liberar una amina activa, que a continuación se puede ciclar para escindir un éster en cualquier otra parte de la molécula (Burgess *et al.*, *J. Org. Chem.* 62:5165-5168, 1997).

También se pueden usar reacciones de eliminación. Por ejemplo, se puede utilizar la eliminación catalizada por bases de grupos tales como Fmoc y cianoetilo, y la eliminación reductora de sistemas alílicos catalizada por paladio.

Una marca puede ser cualquier grupo químico o molécula que se puede detectar. En algunos modos de realización, la marca puede ser cualquier grupo químico o molécula que se puede detectar en un nanoporo, por ejemplo, por su carga, conformación, tamaño o cualquier combinación de estos. En algunos modos de realización, una marca comprende uno o más de etilenglicol o un polímero derivado de etilenglicol, un aminoácido, un carbohidrato, un péptido, un tinte (incluyendo fluoróforos), un compuesto quimioluminiscente, una marca de masa, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un oligonucleótido, un oligonucleótido modificado, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alqueno, un grupo alquino, un grupo ácido o una combinación de los mismos. Otras marcas se divulgan por Fuller *et al.*, "Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array", *PNAS*, 10 de mayo de 2016, 113 (19) 5233-5238. En algunos modos de realización, la marca comprende además una cantidad adecuada de lisinas o argininas para equilibrar la cantidad de grupos fosfato en el compuesto. Los ejemplos de otras marcas adecuadas incluyen las marcas descritas en las publicaciones PCT n.ºs WO/1991/006678, WO/2018/191389 y

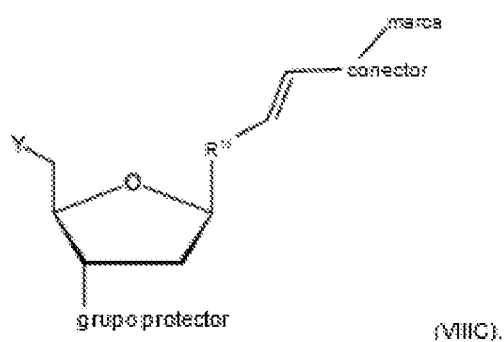
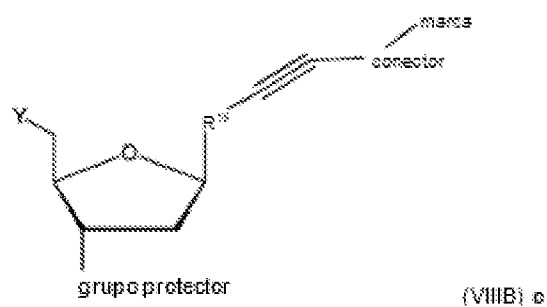
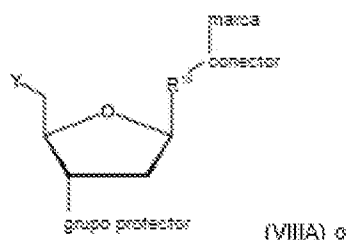
WO/2004/018497.

En algunos modos de realización, la marca es una marca de masa que incluye uno o más grupos indicadores distinguibles por la masa y, por tanto, que se pueden analizar por espectrometría de masas. Los grupos indicadores pueden ser químicamente diferentes y, por tanto, distinguirse entre sí por su peso molecular. De forma alternativa, los grupos indicadores pueden ser químicamente idénticos, pero distinguirse entre sí por contener diferentes isótopos (por ejemplo, $^{12}\text{C}/^{13}\text{C}$ e $^1\text{H}/^2\text{H}$). El resto de marca y/o los grupos indicadores son adecuados o se adaptan para el análisis por espectrometría de masas, por ejemplo, después de la escisión por medios fotoquímicos u otros adecuados. Los ejemplos de marcas de masa adecuadas incluyen las enumeradas en las patentes de EE. UU. n.ºs 7.132.519, 9.291.597 y 10.078.083. Los ejemplos adicionales de marcas de masa adecuadas incluyen marcas de electróforo, tales como las divulgadas por Xu L. *et al.*, "Electrophore Mass Tag Dideoxy DNA Sequence", *Anal. Chem.* 1 de septiembre de 1997; 69(17):3595-602.

En algunos modos de realización, la marca es un polímero. En algunos modos de realización, la marca es un polímero de polietilenglicol (PEG). En modos de realización donde el polímero es PEG, se puede seleccionar un polímero de PEG que tenga cualquier número de unidades de etilenglicol. Por ejemplo, el número de unidades de etilenglicol en el polímero de PEG puede variar desde entre 1 y 100. En algunos casos, el número de unidades de etilenglicol en el polímero de PEG es diferente para cada tipo de nucleótido. Por ejemplo, para cuatro tipos de nucleótidos diferentes, cada uno puede comprender una marca diferente que tiene 16, 20, 24 o bien 36 unidades de etilenglicol en el polímero de PEG. En algunos modos de realización, el polímero basado en PEG es lineal. En otros modos de realización, el polímero basado en PEG está ramificado, es decir, el polímero basado en PEG comprende múltiples cadenas de PEG. En algunos casos, la marca comprende además un resto identificable adicional, tal como un tinte basado en cumarina, o un derivado o análogo de un tinte basado en cumarina. En algunos casos, el polímero tiene carga. En algunos casos, el polímero no tiene carga y la marca se detecta en una alta concentración de sal (por ejemplo, 3-4 M). Se describen ejemplos adicionales de marcas en las publicaciones de EE. UU. n.ºs 2015/0368710, 2018/0073071, 2015/0111759, 2013/0264207, 2013/0244340, 2014/0134616 y 2018/0112257. En particular, las publicaciones de EE. UU. n.ºs 2015/0368710 y 2015/0111759 indicadas anteriormente, describen el uso de nucleótidos marcados para SBS por nanoporos y divulgan el posible uso de un único nucleótido unido a una única marca que comprende cadenas de PEG ramificadas. Otros nucleótidos marcados con PEG se divulgan por Shiv Kumar *et al.*, "PEG-Labeled Nucleotides and Nanopore Detection for Single Molecule DNA Sequencing by Synthesis", *Scientific Reports* 2, artículo número 684 (2012); y por Carl Fuller *et al.*, "Real-time single molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array", *Proc. Natl. Acad. Sci. USA* 2016; 10 de mayo; 113(19):5233-5238.

En algunos modos de realización, la marca es un fluoróforo. Los fluoróforos pertenecen a varias clases químicas comunes, incluyendo cumarinas, fluoresceínas (o derivados y análogos de la fluoresceína), rodaminas, oxacinas (incluyendo resorufinas), BODIPY, luminóforos y cianinas. Se pueden encontrar ejemplos adicionales de moléculas fluorescentes en *Molecular Probes Handbook - A Guide to Fluorescent Probes and Labeling Technologies*, Molecular Probes, Eugene, OR, ThermoFisher Scientific, 11.ª edición. En otros modos de realización, el fluoróforo se selecciona de derivados de xanteno, derivados de cianina, derivados de escuaraina, derivados de naftaleno, derivados de cumarina, derivados de oxadiazol, derivados de antraceno, derivados de pireno, derivados de oxacina, derivados de acridina, derivados de arilmetina y derivados de tetrapirrol. En otros modos de realización, el resto fluorescente se selecciona de un tinte CF (disponible en Biotium), sondas DRAQ y CyTRAK (disponibles en BioStatus), BODIPY (disponible en Invitrogen), Alexa Fluor (disponibles en Invitrogen), DyLight Fluor (por ejemplo, DyLight 649) (disponible en Thermo Scientific, Pierce), Atto y Tracy (disponibles en Sigma Aldrich), FluoProbes (disponibles en Interchim), tintes Abberior (disponibles en Abberior), tintes DY y MegaStokes (disponibles en Dyomics), tintes Sulfo Cy (disponibles en Cyandye), HiLyte Fluor (disponible en AnaSpec), tintes Seta, SeTau y Square (disponibles en SETA BioMedicals), tintes Quasar y Cal Fluor (disponibles en Biosearch Technologies), tintes SureLight (disponibles en APC, RPEPerCP, ficobilisomas) (Columbia Biosciences) y APC, APCXL, RPE, BPE (disponibles en Phyco-Biotech, Greensea, Prozyme, Flogen).

De acuerdo con lo anterior, los ejemplos no limitantes de nucleótidos marcados (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura representada por cualquiera de las fórmulas (VIII A), (VIII B) o (VIII C):



en la que

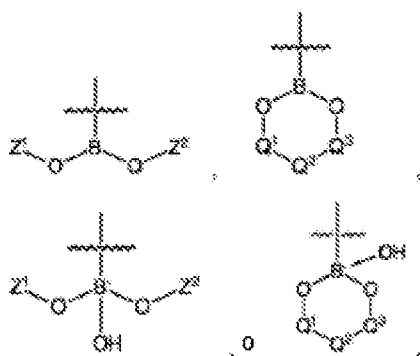
5

R^{15} es una nucleobase;

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

10

'grupo protector' tiene la estructura:



15

Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

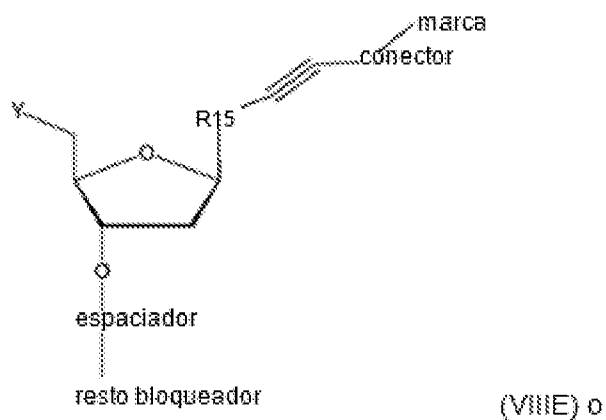
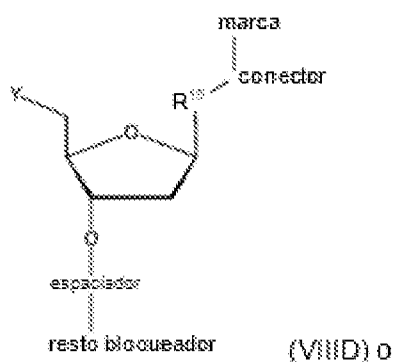
Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

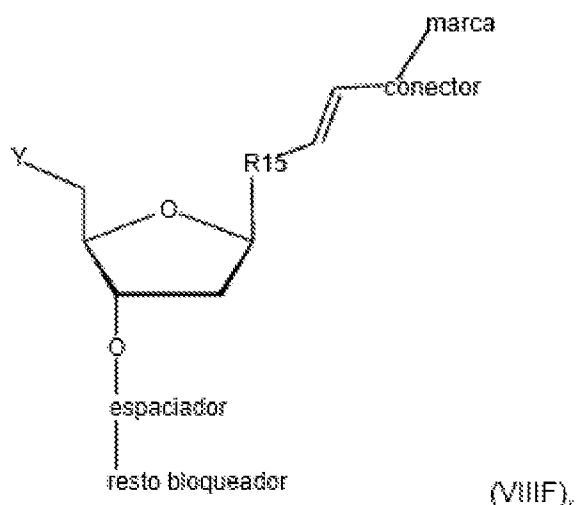
5 R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

10 'conector' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 50 átomos de carbono y opcionalmente sustituido con uno o más heteroátomos; y

'marca' es una especie detectable, incluyendo cualquiera de las especies enumeradas en el presente documento.

15 Otros ejemplos no limitantes de nucleótidos marcados (incluyendo cualquier sal de los mismos) de la presente divulgación tienen una estructura representada por una cualquiera de las fórmulas (VIID), (VIIE) o (VIIF):





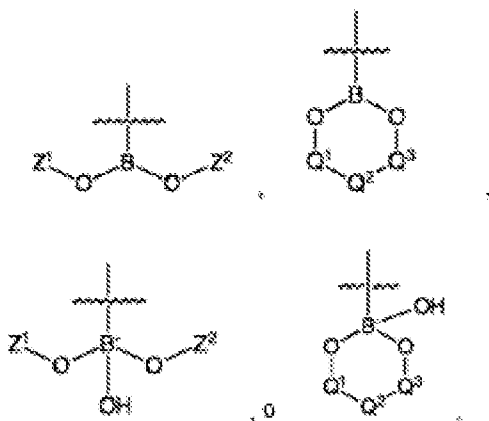
en la que

5 R^{15} es una nucleobase;

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

10 'espaciador' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 16 átomos de carbono y opcionalmente sustituido con uno o más heteroátomos;

15 'resto bloqueador' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 20 átomos de carbono, y opcionalmente sustituido con uno o más heteroátomos, y siempre que el 'resto bloqueador' incluya un grupo ácido, un isonitrilo, un grupo ciano, un grupo heterocicloalquilo de 5 a 8 miembros que tiene dos heteroátomos seleccionados de O, N, S o Se, un resto derivado de un 1,4-epoxi-1,4-dihidronaftaleno sustituido o no sustituido o un grupo que tiene la estructura:



20

Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

25 Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

30 R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

'conector' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 50 átomos de carbono y opcionalmente sustituido con uno o más heteroátomos; y

'marca' es una especie detectable, incluyendo cualquiera de las especies enumeradas en el presente documento;

siempre que cuando el 'resto bloqueador' sea $-N_3$, el 'espaciador' no es $-CH_2-$.

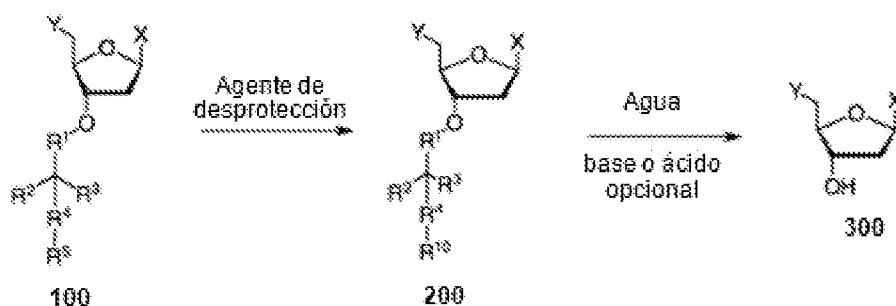
Desprotección de nucleótidos protegidos

La presente divulgación también proporciona procedimientos de desprotección de cualquiera de los nucleótidos protegidos expuestos en el presente documento. Tras la incorporación de un nucleótido protegido en un oligonucleótido creciente como se describe en el presente documento, el nucleótido protegido se puede desproteger para proporcionar un nucleótido desprotegido. En modos de realización donde el nucleótido incluye una marca o un marcador, la desprotección se puede producir antes o después de la retirada de la marca o el marcador. En otros modos de realización, la desprotección se puede producir simultáneamente con la retirada de la marca o el marcador, si está presente.

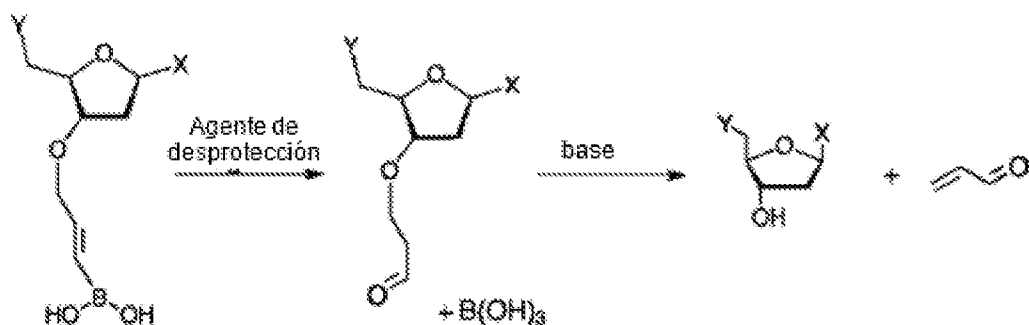
Con referencia a la FIG. 5, los nucleótidos protegidos **500** de acuerdo con la presente divulgación se pueden tratar con un agente de desprotección para proporcionar un intermedio **501**. En algunos modos de realización, la desprotección se lleva a cabo usando un reactivo no reductor. En otros modos de realización, la desprotección se lleva a cabo usando un oxidante. Aún en otros modos de realización, la desprotección se lleva a cabo usando un reactivo neutro redox. Los ejemplos de reactivos adecuados incluyen, sin limitación, peróxido de hidrógeno, peryodato de sodio, perclorato de sodio o peroxinitrato. Otros ejemplos de reactivos adecuados incluyen, sin limitación, una tetracina o un derivado de la misma, o esterasa. Aún otros ejemplos de reactivos adecuados incluyen (tris(2-carboxietil)fosfina) y ditiotretol.

En algunos modos de realización, el intermedio **501** experimenta una eliminación o hidrólisis espontánea para proporcionar el nucleótido desprotegido **502**. Por ejemplo, el intermedio puede experimentar una eliminación 1,6 o una eliminación beta dependiendo de la configuración del intermedio **501** y las condiciones a las que está presente el intermedio **501**. En algunos modos de realización, a continuación, el intermedio **501** se puede tratar opcionalmente con una base para proporcionar el nucleótido desprotegido **502**. En algunos modos de realización, las bases adecuadas incluyen, pero no se limitan a, piperidina, 1,8-diazabicyclo[5.4.0]undec-7-eno (DBU), amoníaco, diisopropiltilamina e hidróxido de sodio.

El esquema 1 a continuación ilustra además los procesos de desprotección de un nucleótido protegido **100** (tal como un nucleótido protegido que tiene la fórmula (IIIA)) en presencia de un agente de desprotección, por ejemplo, peróxido de hidrógeno, peryodato de sodio, perclorato de sodio o $NaONO_2$, para proporcionar un intermedio **200** (tal como un intermedio que tiene la fórmula (VIA)), que a continuación puede experimentar eliminación o hidrólisis espontánea para proporcionar el nucleótido desprotegido **300**. Como se describe en el presente documento, cada uno de los nucleótidos **100**, **200** o **300** se puede incorporar en un oligonucleótido y puede incluir una marca o marcador. Aún otras estrategias de desprotección ilustrativas se ilustran en los esquemas 2 - 9 expuestos a continuación.

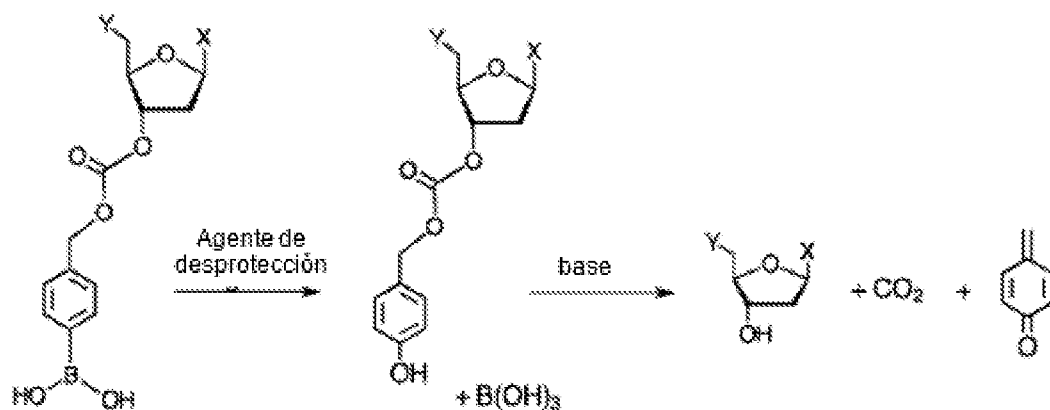


Esquema 1: ilustra la desprotección de un nucleótido protegido, donde R^1 , R^2 , R^3 , R^4 , R^5 , R^{10} , Y y X son como se describe en el presente documento.



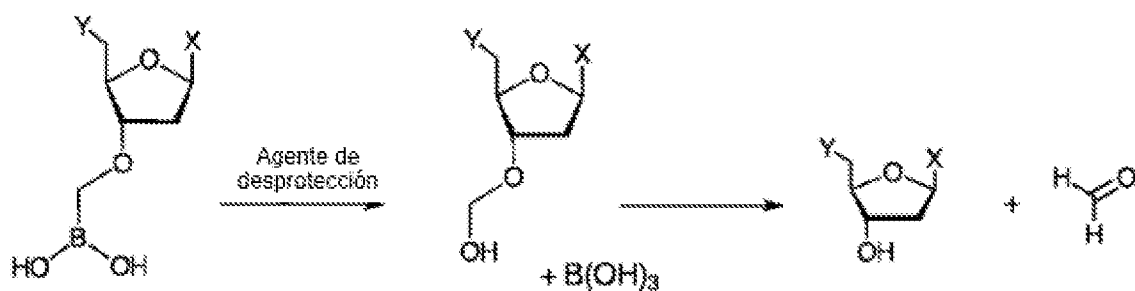
Esquema 2: ilustra la desprotección de un nucleótido protegido que incluye un resto boronato de acuerdo con algunos modos de realización de la presente divulgación, donde Y y X son como se describe en el presente documento.

5



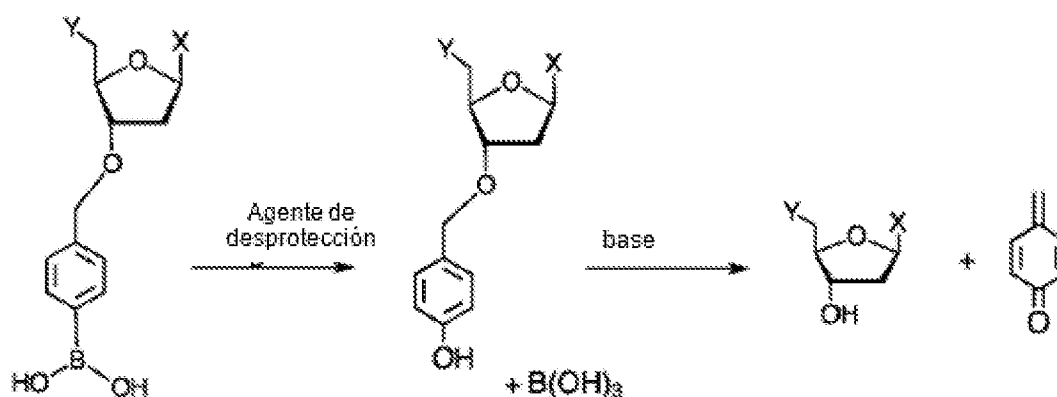
Esquema 3: ilustra la desprotección de un nucleótido protegido que incluye un resto boronato de acuerdo con algunos modos de realización de la presente divulgación donde Y y X son como se describe en el presente documento.

10

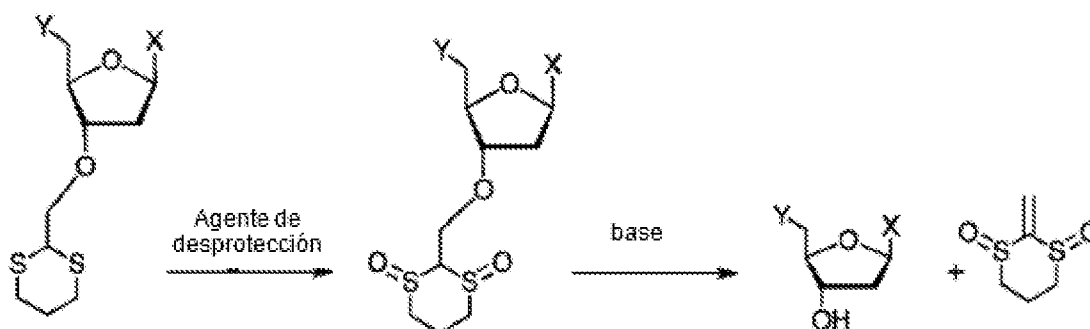


Esquema 4: ilustra la desprotección de un nucleótido protegido que incluye un resto boronato de acuerdo con algunos modos de realización de la presente divulgación donde Y y X son como se describe en el presente documento.

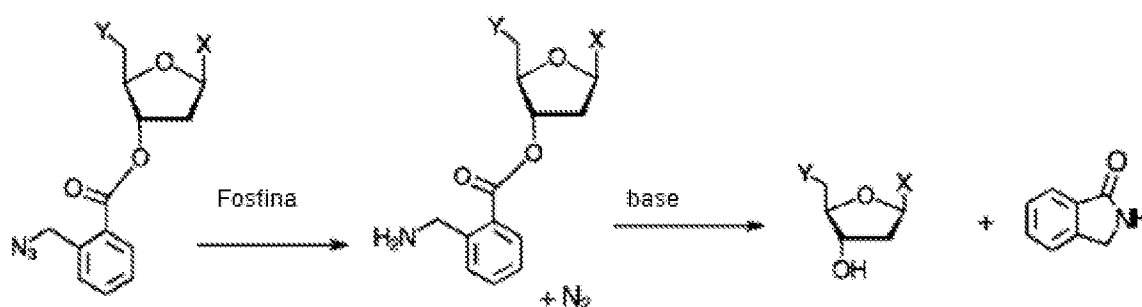
15



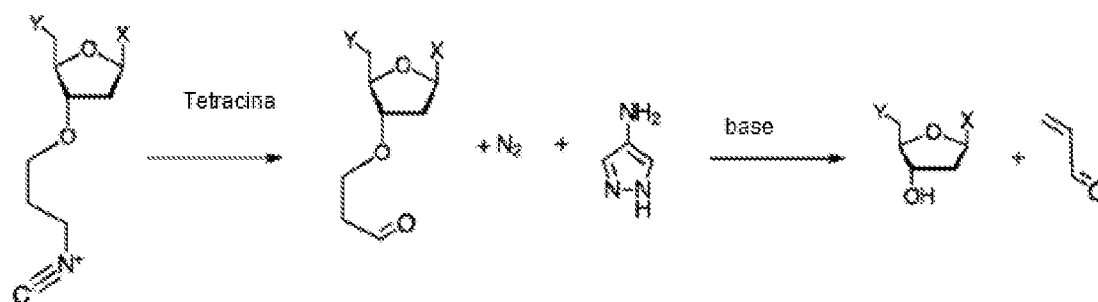
Esquema 5: ilustra la desprotección de un nucleótido protegido que incluye un resto boronato de acuerdo con algunos modos de realización de la presente divulgación donde Y y X son como se describe en el presente documento.



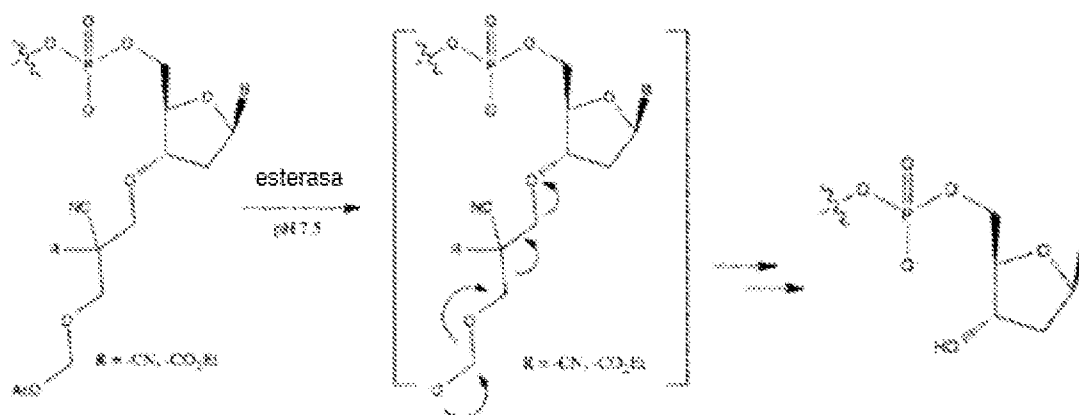
Esquema 6: ilustra la desprotección de un nucleótido protegido que incluye un resto del grupo 1,3-ditiano de acuerdo con algunos modos de realización de la presente divulgación donde Y y X son como se describe en el presente documento.



Esquema 7: ilustra la desprotección de un nucleótido protegido que incluye un resto acida de acuerdo con algunos modos de realización de la presente divulgación donde Y y X son como se describe en el presente documento.



Esquema 8: ilustra la desprotección de un nucleótido protegido que incluye un resto isonitrilo de acuerdo con algunos modos de realización de la presente divulgación donde Y y X son como se describe en el presente documento.



Esquema 9: ilustra la desprotección de un nucleótido protegido que incluye un resto malonato de acuerdo con algunos modos de realización de la presente divulgación.

Secuenciación por nanoporos

La presente divulgación también proporciona procedimientos de secuenciación por nanoporos, con lo que el procedimiento utiliza cualquiera de los nucleótidos o nucleótidos marcados descritos en el presente documento, incluyendo cualquiera de los nucleótidos (o sales de los mismos) de fórmulas (I), (II), (IIIA), (IIIB), (IIIC), (IIID), (IV), (VA), (VB), (VC), (VD), (VIIIA), (VIIIB), (IIIC), (VIID), (VIIIE) y (VIIIF).

Visión general

La secuenciación por nanoporos de un polinucleótido, por ejemplo, ADN o ARN, se puede lograr por secuenciación de hebras y/o exosecuenciación de la secuencia de polinucleótidos. En algunos modos de realización, la secuenciación de hebras comprende procedimientos con los que las bases nucleotídicas de una hebra de polinucleótido de muestra se determinan directamente a medida que los nucleótidos del molde de polinucleótido se hacen pasar a través del nanoporo. En algunos modos de realización, la secuenciación de ácido nucleótido basada en nanoporos usa una mezcla de cuatro análogos de nucleótidos que se pueden incorporar por una enzima en una hebra creciente. En algunos modos de realización, se puede secuenciar un polinucleótido haciéndolo pasar a través de un poro microscópico en una membrana. En algunos modos de realización, las bases se pueden identificar por la forma en que afectan a los iones que fluyen a través del poro de un lado de la membrana al otro. En algunos modos de realización, una molécula de proteína puede "descomprimir" una hélice de ADN en dos hebras. Una segunda proteína puede crear un poro en la membrana y contener una molécula "adaptadora". Un flujo de iones a través del poro puede crear una corriente, con lo que cada base puede bloquear el flujo de iones en diferente grado, alterando la corriente. La molécula adaptadora puede mantener las bases en su lugar el tiempo suficiente para que se identifiquen electrónicamente (véase la publicación PCT n.º WO/2018/034745 y las publicaciones de solicitudes de patente de Estados Unidos n.ºs 2018/0044725 y 2018/0201992).

En algunos modos de realización, se pueden usar nanoporos para secuenciar moléculas de ácido nucleico indirectamente, es decir, la secuenciación indirecta puede incluir cualquier procedimiento donde una molécula de ácido nucleico polimerizada no pase a través del nanoporo durante la secuenciación. En estos modos de realización, la molécula de ácido nucleico puede estar localizada al menos parcialmente en el vestíbulo del nanoporo, pero no en el poro (es decir, la porción más estrecha) del nanoporo. La molécula de ácido nucleico

puede pasar dentro de cualquier distancia adecuada del y/o en proximidad al nanoporo, y opcionalmente dentro de una distancia tal que los subproductos liberados de los acontecimientos de incorporación de nucleótidos, por ejemplo, marcas escindidas de nucleótidos marcados, incluyendo las expuestas en al menos las fórmulas (I) y (II), se detectan en el nanoporo.

En algunos modos de realización, cada análogo de nucleótido tiene un resto de marca unido de forma covalente que proporciona una firma identificable y distinguible cuando se detecta con un nanoporo. La enzima que extiende la hebra (por ejemplo, ADN polimerasa) se une específicamente al compuesto de nucleótido marcado que es complementario a una hebra de ácido nucleico molde que se hibrida a la hebra de ácido nucleico creciente en su sitio activo. A continuación, la enzima que extiende la hebra acopla catalíticamente (es decir, incorpora) el resto de nucleótido complementario del compuesto de nucleótido marcado al extremo de la hebra de ácido nucleico creciente. La finalización del acontecimiento de incorporación catalítica da como resultado la liberación del resto de marca y el resto de oligofosfato (menos el un fosfato incorporado en la hebra creciente) que, a continuación, pasa a través del nanoporo contiguo. Sin embargo, incluso antes de que experimente el proceso catalítico que lo libera del nucleótido incorporado, el resto de marca de un compuesto de nucleótido marcado entra en el poro del nanoporo bajo un potencial aplicado y, de este modo, altera el flujo de iones positivos de fondo a través del nanoporo. En general, la presencia de un resto de marca en un nanoporo da como resultado una disminución (o bloqueo) del flujo de iones positivos a través del nanoporo. Esta "corriente de bloqueo" se detecta como una señal que es un porcentaje de (o está por debajo de) la corriente de "canal abierto" (u "OC") resultante del flujo de iones positivos a través del nanoporo sin presencia de resto de marca.

En algunos modos de realización, la secuenciación basada en nanoporos utiliza una enzima, tal como una localizada en proximidad a un nanoporo, que incorpora nucleótidos protegidos, por ejemplo, aquellos de al menos las fórmulas (I) o (II) en el presente documento, en una cadena polinucleotídica creciente (naciente), en la que la cadena polinucleotídica creciente es complementaria a una hebra de ácido nucleico molde correspondiente. Los acontecimientos de incorporación de nucleótidos se catalizan por la enzima, tal como la ADN polimerasa o cualquier mutante o variante de la misma, y usan interacciones de pares de bases con una molécula de molde para elegir entre los nucleótidos disponibles para su incorporación en cada localización. Los "acontecimientos de incorporación de nucleótidos", como se usa ese término en el presente documento, quiere decir la incorporación de un nucleótido protegido (incluyendo cualquiera de los de las fórmulas (I) o (II)) en una cadena polinucleotídica creciente. En algunos modos de realización, los subproductos de acontecimientos de incorporación de nucleótidos se pueden detectar por el nanoporo. En algunos modos de realización, un subproducto se puede correlacionar con la incorporación de un tipo de nucleótido dado. En algunos modos de realización, el subproducto pasa a través del nanoporo y/o genera una señal detectable en el nanoporo (véase, por ejemplo, la FIG. 4). Las moléculas de marca liberadas, tales como cualquiera de las marcas identificadas anteriormente, son ejemplos de subproductos de acontecimientos de incorporación de nucleótidos. A modo de ejemplo, la FIG. 1 representa una ADN polimerasa (120) unida en estrecha proximidad a un nanoporo (130). Un molde de polinucleótido (170) que se va a secuenciar se añade junto con un cebador (el molde se asocia con la enzima). A este complejo de secuenciación por nanoporos (incluyendo el cebador), se le añaden cuatro nucleótidos marcados de forma diferente (140) al grueso de la fase acuosa. Después de la incorporación catalizada por polimerasa del nucleótido correcto, la marca se liberará y pasará a través del nanoporo (130) para generar una señal de bloqueo de corriente iónica exclusiva (150), identificando de este modo electrónicamente la base añadida porque cada una de las marcas tiene estructuras químicas distintas. Detalles adicionales relacionados con dichos sistemas y procedimientos de secuenciación basados en nanoporos se describen en las patentes de Estados Unidos n.ºs 9.605.309 y 9.557.294.

En algunos modos de realización, un procedimiento para secuenciar una molécula de ácido nucleico comprende (a) polimerizar nucleótidos marcados protegidos (por ejemplo, usando una enzima que incorpora un nucleótido marcado a la vez usando una primera molécula de ácido nucleico como molde) en el que una marca asociada con un nucleótido individual se libera tras la polimerización y donde los nucleótidos marcados protegidos tienen una de las fórmulas (I) o (II); y (b) detectar la marca liberada con la ayuda de un nanoporo. En algunos modos de realización, el grupo bloqueador de cualquier nucleótido protegido en 3' se retira y a continuación los procesos se repiten de forma iterativa. En algunos modos de realización, la enzima extrae de un agrupamiento de nucleótidos marcados protegidos. Como se indica en el presente documento, cada tipo de nucleótido protegido se acopla a una molécula de marca diferente de modo que cuando las marcas se liberan y pasan cerca o a través del nanoporo, se pueden diferenciar entre sí en base a una señal que se genera (véase, por ejemplo, la FIG. 1). En algunos modos de realización, cada marca puede tener una señal detectable diferente, por ejemplo, intensidades de señal diferentes, amplitudes de señal diferentes, etc., que se pueden interpretar tal como por algoritmos de llamada de base.

En algunos modos de realización, una marca liberada fluye a través del nanoporo o en estrecha proximidad al nanoporo de modo que un circuito sensor detecta una señal eléctrica asociada con la marca a medida que pasa a través o cerca del nanoporo (véase la FIG. 1). Una señal detectada (es decir, datos de secuenciación) se puede recoger y almacenar en una localización de memoria y usarse más tarde para construir una secuencia del ácido nucleico. La señal recogida se puede procesar para justificar cualquier anomalía en la señal detectada, tal como errores. Los detectores de nanoporos adecuados se describen en las publicaciones de solicitud de patente de Estados Unidos n.ºs 2011/0193570 y 2018/0073071. Asimismo, las patentes de Estados Unidos n.ºs 9.377.437 y

8.324.914 describen la recogida y análisis de señales eléctricas de sistemas de secuenciación basados en nanoporos.

En algunos modos de realización, las enzimas acopladas o de otro modo conjugadas con nanoporos incluyen enzimas procesadoras de polinucleótidos, por ejemplo, ADN y ARN polimerasas, retrotranscriptasas, exonucleasas y desplegasas. En algunos modos de realización, la enzima es una helicasa. En algunos modos de realización, la enzima puede ser una enzima natural o puede ser una forma variante de la enzima natural. En algunos modos de realización, la enzima es una variante de polimerasa. Por ejemplo, las variantes de polimerasa pueden incluir al menos una alteración en una posición correspondiente a H223, N224, Y225, H227, I295, Y342, T343, I357, S360, L361, I363, S365Q, S366, Y367, P368, D417, E475, Y476, F478, K518, H527, T529, M531, N535, G539, P542, N545, Q546, A547, L549, I550, N552, G553, F558, A596, G603, A610, V615, Y622, C623, D624, I628, Y629, R632, N635, M641, A643, I644, T647, I648, T651, I652, K655, W656, D657, V658, H660, F662 y L690. Otras variantes de polimerasa adecuadas se divulgan en la publicación de solicitud de patente de Estados Unidos n.º 2016/0222363. Aún otras enzimas adecuadas se divulgan en la patente de Estados Unidos n.º 9.797.009. Incluso otras enzimas adecuadas se divulgan en la publicación de solicitud de patente de Estados Unidos n.º 2016/0257942.

En algunos modos de realización, los nanoporos del complejo de secuenciación por nanoporos incluyen, sin limitación, nanoporos biológicos, nanoporos de estado sólido y nanoporos híbridos de estado biológico-sólido. Los nanoporos biológicos de los complejos de secuenciación por nanoporos incluyen OmpG de *E. coli*, *sp.*, *Salmonella sp.*, *Shigella sp.* y *Pseudomonas sp.*, citolisina A (ClyA) y alfa hemolisina de *S. aureus sp.*, MspA de *M. esmegmatis sp.* Los nanoporos pueden ser nanoporos naturales, nanoporos variantes o nanoporos variantes modificados. Véase, por ejemplo, la publicación de solicitud de patente de Estados Unidos n.º 2017/0088588. En algunos modos de realización, el nanoporo variante del complejo de secuenciación por nanoporos se genomanipula para reducir el ruido de corriente iónica del nanoporo original del que se deriva. Aún otros nanoporos se describen en las publicaciones de solicitud de patente de Estados Unidos n.ºs 2017/0268052, 2017/0356037 y 2018/0201993. Cualquier variante de nanoporo ahora conocida o descubierta posteriormente se puede cribar de acuerdo con los procedimientos descritos en el presente documento, tal como simultáneamente con el cribado de una o más variantes de enzimas (por ejemplo, para identificar un par de variante de nanoporo y variante de enzima que proporciona propiedades deseables).

El nanoporo se puede formar o incluir de otro modo en una membrana dispuesta contigua a un electrodo sensor de un circuito sensor, tal como un circuito integrado. El circuito integrado puede ser un circuito integrado específico de la aplicación (ASIC). En algunos ejemplos, el circuito integrado es un transistor de efecto de campo o un semiconductor complementario de óxido metálico (CMOS). El circuito sensor se puede situar en un chip u otro dispositivo que tenga el nanoporo, o fuera del chip o dispositivo, tal como en una configuración fuera de chip. El semiconductor puede ser cualquier semiconductor, incluyendo, sin limitación, semiconductores del grupo IV (por ejemplo, silicio) y del grupo III-V (por ejemplo, arseniuro de galio, disulfuro de molibdeno). Los procedimientos para ensamblar complejos de secuenciación por nanoporos se describen en la publicación de solicitud de patente de EE. UU. n.º 2017/0268052. Otros procedimientos adecuados para complejar cada uno de los diferentes moldes con conjugados de nanoporo-enzima incluyen los descritos en las publicaciones PCT n.ºs WO2014/074727, WO2006/028508 y WO2012/083249.

La FIG. 2 ilustra un modo de realización de una celda **160** en un chip de secuenciación basado en nanoporos. En algunos modos de realización, se forma una membrana **102** sobre la superficie de la célula. En algunos modos de realización, la membrana **102** es una bicapa lipídica. El grueso del electrolito **114** que contiene complejos moleculares transmembranarios de nanoporos de proteínas (PNTMC) y el analito de interés se coloca directamente sobre la superficie de la celda. En algunos modos de realización, un único PNTMC **104** se inserta en la membrana **102** por electroporación. En algunos modos de realización, las membranas individuales de la matriz no están conectadas ni química ni eléctricamente entre sí. Por tanto, cada celda en la matriz es una máquina de secuenciación independiente, que produce datos exclusivos para la única molécula de polímero asociada con el PNTMC. En algunos modos de realización, el PNTMC **104** funciona sobre los analitos y modula la corriente iónica a través de la bicapa que de otro modo sería impermeable.

Con referencia continua a la FIG. 2, los circuitos de medición analógicos **112** se conectan a un electrodo de metal **170** (por ejemplo, un electrodo compuesto de rutenio, oxígeno, titanio o nitrógeno) cubierto por una película delgada de electrolito **108**. En algunos modos de realización, la película delgada de electrolito **108** se aísla del grueso del electrolito **114** por la membrana impermeable a iones **102**. El PNTMC **104** cruza la membrana **102** y proporciona la única vía para que la corriente iónica fluya del grueso del líquido al electrodo de trabajo **170**. En algunos modos de realización, la celda también incluye un contraelectrodo (CE) **116**, que es un sensor de potencial electroquímico. En algunos modos de realización, la celda también incluye un electrodo de referencia **117**.

Un chip para secuenciar una muestra de ácido nucleico puede comprender una pluralidad de nanoporos individualmente direccionables. Un nanoporo individualmente direccionable de la pluralidad puede contener al menos un nanoporo formado en una membrana dispuesta contigua a un circuito integrado. En algunos modos de realización, cada nanoporo individualmente direccionable puede detectar una marca asociada con un nucleótido

individual.

Se pueden proporcionar múltiples sensores de nanoporos como matrices, tales como matrices presentes en un chip o biochip. La matriz de nanoporos puede tener cualquier número adecuado de nanoporos. En algunos casos, la matriz comprende aproximadamente 200, aproximadamente 400, aproximadamente 600, aproximadamente 800, aproximadamente 1000, aproximadamente 1500, aproximadamente 2000, aproximadamente 3000, aproximadamente 4000, aproximadamente 5000, aproximadamente 10000, aproximadamente 15000, aproximadamente 20000, aproximadamente 40000, aproximadamente 60000, aproximadamente 80000, aproximadamente 100000, aproximadamente 200000, aproximadamente 400000, aproximadamente 600000, aproximadamente 800000, aproximadamente 1000000 y similares nanoporos. Los biochips y los procedimientos para fabricar biochips se describen en la publicación PCT n.º WO2015/061511. En la publicación de solicitud de patente de Estados Unidos n.º 2017/0268052 se describen otros biochips adecuados que comprenden una pluralidad de nanoporos. Aún otras matrices de nanoporos adecuadas se describen en la patente de Estados Unidos n.º 8.986.928.

Incorporación de nucleótidos protegidos o nucleótidos marcados protegidos en un oligómero

La presente divulgación proporciona procedimientos de síntesis de un oligómero, tal como un oligonucleótido usado para la secuenciación por nanoporos, derivándose el oligómero de los nucleótidos protegidos o de los nucleótidos marcados protegidos divulgados en el presente documento.

La FIG. 3A expone un diagrama de flujo que ilustra el procedimiento general de sintetizar un oligonucleótido usando los nucleótidos de cualquiera de las fórmulas (I), (II), (IIIA), (IIIB), (IIIC), (IIID), (IV), (VA), (VB), (VC), (VD), (VIIIA), (VIIIB), (IIIC), (IIID), (VIIIE) y (VIIIF). En la etapa **320**, se introduce un nucleótido protegido o un nucleótido marcado protegido, tal como en un complejo de secuenciación por nanoporos. En algunos modos de realización, el nucleótido protegido o nucleótido marcado protegido se introduce como un "agrupamiento" de nucleótidos protegidos o nucleótidos marcados protegidos, en los que el agrupamiento puede comprender diferentes nucleótidos protegidos o nucleótidos marcados protegidos, difiriendo cada uno al menos en la nucleobase o nucleobase marcada acoplada al mismo. En algunos modos de realización, el agrupamiento comprende nucleótidos protegidos o nucleótidos marcados protegidos que se pueden hibridar con un nucleótido A, T, G o C en el ADN que se secuencia (o, en el caso de la secuenciación de ARN, A, G, C y U). En algunos modos de realización, se introducen al menos cuatro nucleótidos protegidos diferentes en un agrupamiento.

En la etapa **321**, el nucleótido protegido o el nucleótido marcado protegido se incorpora enzimáticamente a una hebra de oligonucleótido creciente. A modo de ejemplo, una enzima polimerasa puede extraer del agrupamiento de nucleótidos protegidos o nucleótidos marcados protegidos e incorporar enzimáticamente el nucleótido protegido o el nucleótido marcado protegido en la hebra de oligonucleótido creciente. En la etapa **322**, el nucleótido protegido o el nucleótido marcado protegido se desprotege, tal como se describe en el presente documento. En modos de realización donde se incorpora un nucleótido marcado protegido, el procedimiento puede comprender además liberar la marca y detectar la marca liberada. En algunos modos de realización, la marca se libera antes de la desprotección (etapa **322**) pero después de la incorporación enzimática (etapa **321**). De forma alternativa, y en otros modos de realización, la marca se libera después de la desprotección (etapa **322**) pero antes de la incorporación enzimática de un segundo nucleótido protegido o nucleótido marcado protegido (etapa **323**). Aún en otros modos de realización, la liberación de la marca y la desprotección se pueden producir en la misma etapa (por ejemplo, se pueden elegir reactivos que actúan sobre el grupo protector para facilitar la desprotección mientras actúan también sobre un conector escindible para posibilitar la liberación de la marca acoplada al mismo). La etapa **323** indica que el proceso se repite de forma iterativa hasta que se secuencia una secuencia de ácido nucleico de secuencia completa, por ejemplo, una secuencia de ADN o ARN, con el nanoporo. El experto apreciará que si se introduce un agrupamiento de nucleótidos en la etapa **320**, a continuación se pueden repetir (etapa **323**) las etapas **321** y **322** según sea necesario de modo que se pueda secuenciar un polinucleótido de acuerdo con la presente divulgación.

La FIG. 3B expone un diagrama de flujo que ilustra las etapas de la incorporación enzimática de un nucleótido marcado protegido (etapa **331**), tal como uno que incluye un resto escindible como se describe en el presente documento, seguido de la retirada de la marca del nucleótido marcado protegido (**332**), y posterior desprotección del nucleótido incorporado (ahora un nucleótido protegido desmarcado) (etapa **333**). En algunos modos de realización, cada uno de los diferentes nucleótidos marcados protegidos se distingue por la señal detectable distintiva que produce la marca cuando se incorpora a una nueva hebra complementaria por una enzima que extiende la hebra. En algunos modos de realización, las marcas liberadas pueden fluir a través de un nanoporo después de que se liberan del nucleótido. En algunos modos de realización, se aplica un voltaje para tirar de las marcas a través del nanoporo. El experto apreciará que las etapas de incorporación enzimática, retirada de marca y desprotección se pueden repetir (etapa **334**) según sea necesario de modo que se pueda secuenciar un polinucleótido de acuerdo con la presente divulgación.

El procedimiento se ilustra además en la FIG. 3C. Como se muestra, la hebra de ácido nucleico **300** pasa a través o en proximidad (pero no a través como se indica por la flecha en **301**) al nanoporo **302**. Una enzima **303** (por

ejemplo, ADN polimerasa) extiende una cadena de ácido nucleico creciente **304** incorporando un nucleótido protegido o un nucleótido marcado protegido a la vez usando una primera molécula de ácido nucleico como molde **300** (es decir, la enzima cataliza acontecimientos de incorporación de nucleótidos).

En algunos modos de realización, y con referencia continua a la FIG. 3C, la enzima extrae de un agrupamiento de nucleótidos marcados protegidos (círculos con relleno en la indicación **305**) unidos a moléculas de marca (círculos sin relleno en la indicación **305**). Cada tipo de nucleótido marcado protegido se une a una molécula de marca diferente de modo que cuando las marcas se liberan y pasan a través del nanoporo **306**, se pueden diferenciar entre sí en base a la señal que se genera en el nanoporo. En algunos modos de realización, a medida que la marca pasa hacia y/o a través del nanoporo, puede generar un cambio electrónico. En algunos modos de realización, el cambio electrónico es un cambio en la amplitud de la corriente, un cambio en la conductancia del nanoporo o cualquier combinación de los mismos. Entre las características de señal detectables, solas o en combinación, que se pueden usar para distinguir los nucleótidos marcados protegidos en un procedimiento de detección con nanoporos está el cambio en el flujo iónico provocado por la presencia de la marca en el nanoporo, que a su vez da como resultado un cambio en la medida del nivel de corriente a través de los electrodos del sistema de detección con nanoporos (bajo potencial de CC o bien de CA). "Flujo iónico", como se usa en el presente documento, se refiere al movimiento de iones, típicamente en una solución, debido a una fuerza electromotriz, tal como el potencial entre un ánodo y un cátodo. El flujo iónico típicamente se puede medir como corriente o la caída de un potencial electrostático. En consecuencia, en algunos modos de realización, la presente divulgación proporciona un conjunto de nucleótidos marcados protegidos, cada uno con una marca diferente, en el que cada marca diferente provoca un flujo de iones diferente a través del poro dando como resultado un nivel de corriente de marca detectable diferente a través de los electrodos cuando se sitúa en el nanoporo.

La FIG. 4 proporciona un ejemplo de señales diferentes que se generan por marcas diferentes a medida que se detectan por el nanoporo. Se detectan cuatro intensidades de señal diferentes (**401**, **402**, **403** y **404**). Estas pueden corresponder a cuatro marcas diferentes, tales como marcas incluidas dentro de cualquiera de los nucleótidos marcados protegidos divulgados en el presente documento. Por ejemplo, la marca presentada en el nanoporo y/o liberada por la incorporación de adenosina (A) puede generar una señal con una amplitud **401**. Una marca presentada en el nanoporo y/o liberada por la incorporación de citosina (C) puede generar una señal con una amplitud mayor **403**; una marca presentada en el nanoporo y/o liberada por la incorporación de guanina (G) puede generar una señal con una amplitud incluso mayor **404**; y una marca presentada en el nanoporo y/o liberada por la incorporación de timina (T) puede generar una señal con una amplitud aún mayor **402**. La señal puede volver a un nivel de referencia **405** entre detecciones en algunos casos.

La FIG. 6 ilustra además la incorporación de nucleótidos marcados protegidos en una hebra de polinucleótido creciente. Como se ilustra, cada nucleótido marcado protegido incluye un grupo R que representa un grupo protector de acuerdo con la presente divulgación (por ejemplo, R puede representar el 'grupo protector' o el resto bloqueador-espaciador- de las fórmulas (I) y (II), respectivamente. En este modo de realización particular, la escisión de la marca (es decir, marcas 1, 2, 3 y 4) también da como resultado la retirada del grupo R, es decir, el resto protector, de modo que se pueden introducir de forma iterativa nucleótidos marcados protegidos marcados adicionales.

Secuenciación por síntesis

La presente divulgación también proporciona procedimientos de secuenciación por síntesis (SBS), con lo que el procedimiento utiliza cualquiera de los nucleótidos o nucleótidos marcados descritos en el presente documento, incluyendo cualquiera de los nucleótidos (o sales de los mismos) de fórmulas (I), (II), (IIIA), (IIIB), (IIIC), (IIID), (IV), (VA), (VB), (VC), (VD), (VIIIA), (VIIIB), (VIIC), (VIID), (VIIIE) y (VIIIF). Las técnicas de SBS en general implican la extensión enzimática de una hebra de ácido nucleico naciente a través de la adición iterativa de nucleótidos contra una hebra de molde. Cada adición de nucleótido consulta una o unas pocas bases de la hebra de molde. En un tipo ejemplar de SBS, la secuenciación del ciclo se logra por la adición gradual de nucleótidos finalizadores reversibles que contienen, por ejemplo, un marcador de tinte escindible o fotoblanqueable (véase, por ejemplo, la publicación de solicitud PCT n.º WO 91/06678. En algunos modos de realización, se añaden secuencialmente uno o más nucleótidos marcados protegidos a una cadena de polinucleótidos que se extiende en la dirección 5' a 3' para formar un polinucleótido extendido complementario a un ácido nucleico molde que se va a secuenciar. La identidad de la base presente en uno o más de los nucleótidos marcados protegidos añadidos se puede determinar en una etapa de detección o de formación de imágenes, tal como después de cada incorporación de nucleótido.

Por ejemplo, un procedimiento para determinar la secuencia de un polinucleótido monocatenario diana usando secuenciación por síntesis comprende: (a) proporcionar nucleótidos marcados protegidos (tales como cualquiera de los nucleótidos marcados protegidos de las fórmulas (I), (II), (IIIA), (IIIB), (IIIC), (IIID), (IV), (VA), (VB), (VC), (VD), (VIIIA), (VIIIB), (VIIC), (VIID), (VIIIE) y (VIIIF); (b) incorporar un nucleótido marcado protegido en un complemento de un polinucleótido monocatenario diana; (c) detectar la marca del nucleótido marcado protegido de la etapa (b), determinando de este modo el tipo de nucleótido incorporado; (d) retirar el grupo protector en 3' y la marca del nucleótido marcado protegido de la etapa (b); y (e) repetir opcionalmente las etapas (b) - (d) una o más veces; determinando de este modo la secuencia del polinucleótido monocatenario diana. En algunos modos de

realización, el procedimiento puede comprender además liberar la marca y detectar la marca liberada. En algunos modos de realización, el nucleótido marcado protegido se desprotege tal como se describe en el presente documento. En algunos modos de realización, la marca se libera antes de la desprotección del grupo protector en 3' pero después de la incorporación enzimática. En otros modos de realización, la marca se libera después de la desprotección del grupo protector en 3' pero antes de la incorporación enzimática de un segundo nucleótido protegido o nucleótido marcado protegido. En algunos modos de realización, la escisión de la marca y la desprotección del grupo protector en 3' tienen lugar sustancialmente de forma simultánea. En algunos modos de realización, los nucleótidos marcados protegidos proporcionados incluyen al menos cuatro nucleótidos marcados protegidos diferentes (tal como un agrupamiento de nucleótidos marcados protegidos), cada nucleótido marcado protegido diferente incluye una marca diferente de modo que cada uno de los nucleótidos marcados protegidos diferentes se pueden distinguir entre sí.

En algunos modos de realización, el procedimiento para determinar la secuencia de un polinucleótido diana usando secuenciación por síntesis se puede llevar a cabo poniendo en contacto el polinucleótido diana por separado con los nucleótidos marcados protegidos diferentes para formar el complemento al del polinucleótido diana y detectando la incorporación de los nucleótidos marcados protegidos. Como se indica en el presente documento, un procedimiento de este tipo hace uso de la polimerización, con lo que una enzima polimerasa extiende la hebra complementaria incorporando el nucleótido marcado protegido correcto complementario al de la diana. La reacción de polimerización también requiere un cebador específico para iniciar la polimerización. Para cada ciclo, la incorporación del nucleótido marcado se lleva a cabo por la enzima polimerasa (tal como las indicadas en el presente documento) y a continuación se determina el acontecimiento de incorporación.

En algunos modos de realización, los procedimientos de secuenciación se llevan a cabo con el polinucleótido diana dispuesto sobre un soporte sólido. Se pueden inmovilizar múltiples polinucleótidos diana sobre el soporte sólido a través de moléculas conectoras, o se pueden unir a partículas, por ejemplo, microesferas, que también se pueden unir a un material de soporte sólido. En algunos modos de realización, los polinucleótidos se pueden unir al soporte sólido por una serie de medios, incluyendo el uso de interacciones biotina-avidina. Los soportes sólidos adecuados incluyen, pero no se limitan a, portaobjetos y perlas de vidrio, superficies de cerámica y silicio y materiales plásticos. En algunos modos de realización, el soporte es una superficie plana. En otros modos de realización también se pueden usar perlas microscópicas (microesferas) y que a su vez se pueden unir a otro soporte sólido por medios conocidos. Las microesferas pueden ser de cualquier tamaño adecuado, típicamente en el intervalo de desde aproximadamente 10 nm a aproximadamente 100 nm de diámetro. En algunos modos de realización, los polinucleótidos se unen directamente a una superficie plana, por ejemplo, una superficie de vidrio plana. En algunos modos de realización, la unión es a través de un enlace covalente. Los ejemplos no limitantes de matrices adecuadas se describen en la publicación de solicitud PCT n.º WO 00/06770. El procedimiento de secuenciación se puede llevar a cabo tanto en una única molécula de polinucleótido como en matrices de moléculas de múltiples polinucleótidos, es decir, matrices de distintas moléculas de polinucleótidos individuales y matrices de distintas regiones que comprenden múltiples copias de una molécula de polinucleótido individual. En algunos modos de realización, las matrices de única molécula permiten que cada polinucleótido individual se resuelva por separado. En algunos modos de realización, se cree que la secuenciación de matrices de molécula única de forma no destructiva permite que se forme una matriz direccionable espacialmente.

Para llevar a cabo la reacción de la polimerasa, en algunos modos de realización se hibrida una secuencia de cebador con el polinucleótido diana, reconociéndose la secuencia de cebador por la enzima polimerasa y actuando como un sitio de iniciación para la extensión posterior de la hebra complementaria. La secuencia de cebador se puede añadir como un componente separado con respecto al polinucleótido diana. De forma alternativa, el cebador y el polinucleótido diana pueden ser cada uno parte de una molécula monocatenaria, formando la porción de cebador un dúplex intramolecular con una parte de la diana, es decir, una estructura de bucle en horquilla. En algunos modos de realización, esta estructura se puede inmovilizar en el soporte sólido en cualquier punto de la molécula. En algunos modos de realización, los nucleótidos marcados protegidos de la presente divulgación a continuación se ponen en contacto con el polinucleótido diana, para permitir que se produzca la polimerización. En algunos modos de realización, los nucleótidos marcados protegidos se pueden añadir secuencialmente, es decir, adición separada de cada tipo de nucleótido (por ejemplo, un nucleótido marcado protegido que incorpora una nucleobase tal como A, T, G o C), o añadirse juntos. En algunos modos de realización, si los nucleótidos marcados protegidos se añaden juntos, cada tipo diferente de nucleótido marcado protegido que tiene una nucleobase diferente se marca con una marca diferente. En algunos modos de realización, se deja que la etapa de polimerización avance durante un tiempo suficiente para permitir la incorporación de un nucleótido marcado protegido. En algunos modos de realización, los nucleótidos marcados protegidos no incorporados se retiran a continuación, por ejemplo, sometiendo la matriz a una etapa de lavado y, a continuación, se puede llevar a cabo la detección de las marcas incorporadas. En algunos modos de realización, el procedimiento comprende además una etapa de desprotección, como se indica en el presente documento. En algunos modos de realización, después de la detección, la marca se puede retirar usando condiciones adecuadas que escindan el conector.

En algunos modos de realización, cada uno de los nucleótidos marcados protegidos se puede poner en contacto con la diana secuencialmente, con la retirada de los nucleótidos marcados protegidos no incorporados antes de la adición del siguiente nucleótido marcado protegido, donde la detección y retirada de la marca se lleva a cabo

después de la adición de cada nucleótido marcado protegido, o bien después de la adición de los cuatro nucleótidos marcados protegidos. En otros modos de realización, todos los tipos diferentes de nucleótidos marcados protegidos se ponen en contacto con la diana simultáneamente, es decir, una composición que comprende todos los nucleótidos marcados protegidos diferentes se ponen en contacto con la diana, y los nucleótidos no incorporados se retiran antes de la detección y posterior a la retirada de la(s) marca(s).

En algunos modos de realización, los procedimientos pueden comprender una primera etapa y una segunda etapa, donde en la primera etapa, una primera composición que comprende dos de los cuatro nucleótidos se pone en contacto con la diana, y los nucleótidos marcados protegidos no incorporados se retiran antes de detección y posterior a la retirada de la marca, y donde en la segunda etapa, una segunda composición que comprende los dos nucleótidos marcados protegidos no incluidos en la primera composición se pone en contacto con la diana, y los nucleótidos marcados protegidos no incorporados se retiran antes de detección y posterior a la retirada de la marca, y donde la primera etapa y la segunda etapa se pueden repetir opcionalmente una o más veces después de la desprotección del grupo protector en 3' del nucleótido incorporado.

En algunos modos de realización, los procedimientos descritos en el presente documento también pueden comprender una primera etapa y una segunda etapa, donde en la primera etapa, una composición que comprende uno de cuatro nucleótidos marcados protegidos diferentes se pone en contacto con la diana, y los nucleótidos marcados protegidos no incorporados se retiran antes de la detección y posterior a la retirada de la marca, y donde en la segunda etapa, una segunda composición que comprende los tres nucleótidos marcados protegidos no incluidos en la primera composición se pone en contacto con la diana, y los nucleótidos marcados protegidos no incorporados se retiran antes de la detección y posterior a la retirada de la marca, y donde la primera etapa y la segunda etapa se pueden repetir opcionalmente una o más veces después de la desprotección del grupo protector en 3' del nucleótido incorporado.

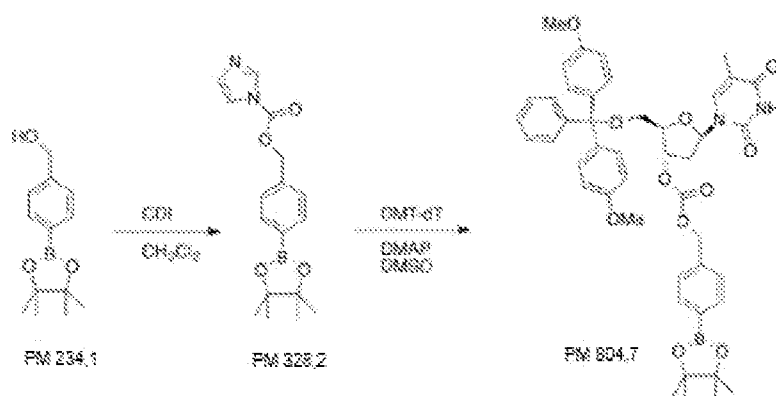
En algunos modos de realización, los procedimientos descritos en el presente documento también pueden comprender una primera etapa y una segunda etapa, donde en la primera etapa, una primera composición que comprende tres de los cuatro nucleótidos marcados protegidos diferentes se pone en contacto con la diana, y los nucleótidos no incorporados se retiran antes de la detección y posterior a la retirada de la marca, y donde en la segunda etapa, una composición que comprende el nucleótido marcado protegido no incluido en la primera composición se pone en contacto con la diana, y los nucleótidos marcados protegidos no incorporados se retiran antes de la detección y posterior a la retirada de la marca, y donde la primera etapa y la segunda etapa se pueden repetir opcionalmente una o más veces después de la desprotección del grupo protector en 3' del nucleótido incorporado.

En algunos modos de realización, el procedimiento para determinar la secuencia de un polinucleótido diana comprende supervisar la incorporación secuencial de nucleótidos marcados protegidos complementarios, en el que los nucleótidos marcados protegidos comprenden una marca detectable enlazada al nucleótido marcado protegido por medio de un conector escindible, y con lo que se detecta la incorporación supervisando la marca, y en el que el procedimiento comprende además una etapa de desprotección para permitir que se produzca otra incorporación de nucleótido marcado protegido.

Componentes y procedimientos adicionales para la secuenciación por síntesis se describen en las patentes de Estados Unidos n.ºs 9.605.310 y 9.441.272.

Ejemplos

Preparación de DMT-dT protegido:

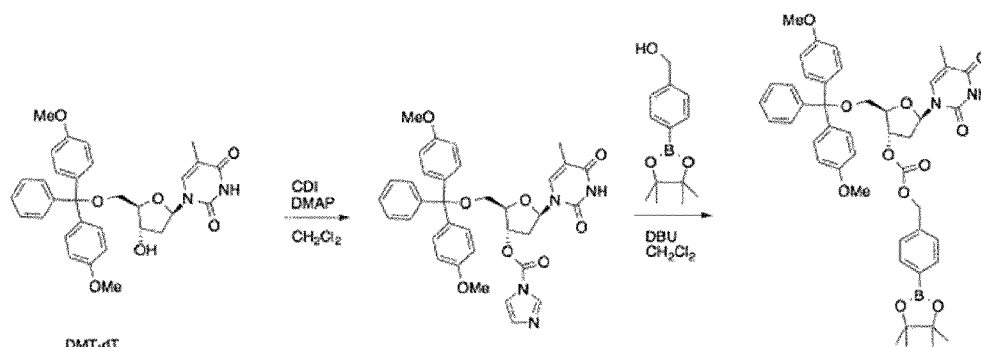


Se colocaron 255 mg (1,1 mmol) de éster pinacólico del ácido 4-(hidroximetil)bencenoborónico y 340 mg de CDI

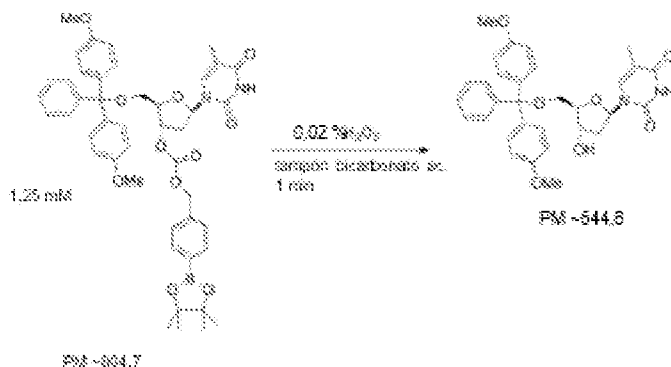
(2,1 mmol) en un tubo de 15 ml. Se añadieron 1,5 ml de diclorometano seco y se agitó la solución durante 30 minutos. A continuación, se diluyó hasta 10 ml con acetato de etilo y se lavó con 5 ml de agua (x2). A continuación, se secó sobre sulfato de sodio anhidro y se concentró a vacío para dar 300 mg de sólido blanco (rendimiento de un 80 %).

Se disolvieron 440 mg de DMT-dT (0,81 mmol) en 1 ml de DMSO anhidro. Se añadieron 122 mg de DMAP (1,0 mmol) y se agitó para disolver. A continuación, se añadió a los 300 mg del sólido blanco anterior (0,91 mmol) y se agitó bajo atmósfera de nitrógeno durante la noche. A continuación, se disolvió en 100 ml de acetato de etilo y se lavó con 100 ml de agua (x3) y salmuera. Se secó sobre sulfato de sodio anhidro y se evaporó a vacío para dar 800 mg de un aceite viscoso. Se purificó sobre una combi flash en 40 g de gel de sílice con disolvente hexanos/acetato de etilo (triethylamina al 1 %) de 50:50 a 0:100 durante 15 min. La fracción pura con una masa de 803 (modo negativo) se concentró a vacío para dar 30 mg de boronato-carbonato-DMT-dT como una película viscosa.

Preparación de boronato-carbonato-DMT-dT



Desprotección de DMT-dT protegido:



Reacción desenmascaradora:

Se disolvieron 30 mg del producto boronato-carbonato-DMT-dT anterior en 3 ml de acetonitrilo para preparar una solución de 10 mg/ml (aproximadamente 12,5 mM). A continuación se preparó la siguiente reacción:

Se añadieron 10 µl de 10 mg/ml de boronato-carbonato-DMT-dT a 90 µl de tampón de bicarbonato de sodio 20 mM, pH ~8,5. La solución se volvió turbia debido a la baja solubilidad del boronato-carbonato-DMT-dT en agua. Se añadieron 2 µl de H₂O₂ ac. al 1 % y se agitó en vórtex la reacción durante 1 minuto. La solución se volvió transparente de nuevo. Se diluyeron 10 µl en 1 ml de agua y se analizó por CL-EM (en modo negativo). El pico de masa principal de 543 mostró el desenmascaramiento del compuesto hasta llegar a DMT-dT. No se observó material de partida.

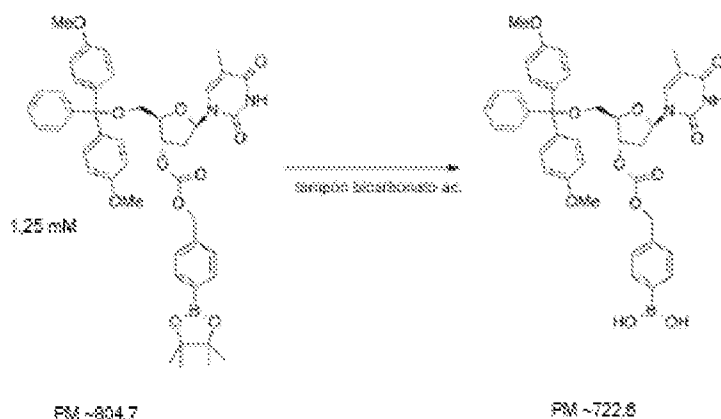
Reacción de control:

Se añadieron 10 µl de 10 mg/ml de boronato-carbonato-DMT-dT a 90 µl de tampón de bicarbonato de sodio 20 mM, pH ~8,5. La solución se volvió turbia. Se añadieron 2 µl de agua y se agitó en vórtex la reacción durante 10 minutos. La solución todavía estaba turbia pero en menor medida. Se diluyeron 10 µl en 1 ml de agua y se analizó por CL-EM en modo negativo). Se observaron dos picos de masa principales: 803 del material de partida y 721 de

la hidrólisis del éster pinacólico.

A continuación, se agitó la reacción de control a t.a. durante 40 horas. Se diluyeron 10 µl en 1 ml de agua y se analizó por CL-EM en modo negativo). Se observó el único pico de masa principal 721 procedente de la hidrólisis del éster pinacólico.

5



Preparación de un nucleótido protegido que tiene un grupo malonilo

10

Se disuelven 10 mmol del compuesto bis(hidroximetilo) 1 en 8 ml de piridina seca. Se añaden lentamente 9 mmol de cloruro de *tert*-butildimetilsililo disueltos en 2 ml de piridina seca y se agita la reacción durante 3 días. Se evapora la mezcla hasta sequedad a vacío y se disuelve en 50 ml de CH₂Cl₂. Se lava con 50 ml de agua y se extrae la solución acuosa con 50 ml de CH₂Cl₂. Se combinan todas las fracciones orgánicas, se secan sobre sulfato de sodio anhidro y se concentra a vacío. Se purifica además el producto 2 por cromatografía en gel de sílice usando hexanos/acetato de etilo como disolvente. En este ejemplo, R puede ser -CN o -C(O)O-Et.

15



20

Se disuelven 10 mmol del compuesto 2 en una solución premezclada de 20 ml de anhídrido acético, 6 ml de ácido acético y 30 ml de DMSO. Se agita la reacción durante la noche y a continuación se diluye cuidadosamente con 150 ml de Na₂CO₃ acuoso al 10 % frío. Se extrae el producto con éter dietílico (5x50 ml). Se combinan las fases orgánicas, se seca sobre sulfato de sodio anhidro y se concentran a vacío. Se purifica además el producto 3 por cromatografía en gel de sílice con CH₂Cl₂ como disolvente. En este ejemplo, R puede ser -CN o -C(O)O-Et.

25



30

Se disuelven 10 mmol de compuesto en 50 ml de CH₂Cl₂ seco y se agita la solución bajo argón. Se añadieron 12 ml (12 mmol) de solución de cloruro de sulfurilo 1 M en CH₂Cl₂ en 3 alícuotas y se agitó la reacción durante 1 hora bajo argón. Se retira el disolvente a vacío y se disuelve el residuo en 30 ml de CH₂Cl₂ seco. Se añade una solución de 17 mmol de acetato de potasio y 7,5 mmol de dibenzo-18-corona-6 en 30 ml de CH₂Cl₂ seco y se agita la reacción durante 1,5 h. Se añaden 80 ml de acetato de etilo y se lava la fase orgánica con 100 ml de agua. Se seca sobre sulfato de sodio y se concentra a vacío. Se purifica además el producto 4 por cromatografía en gel de sílice con CH₂Cl₂ como disolvente. En este ejemplo, R puede ser -CN o -C(O)O-Et.

35



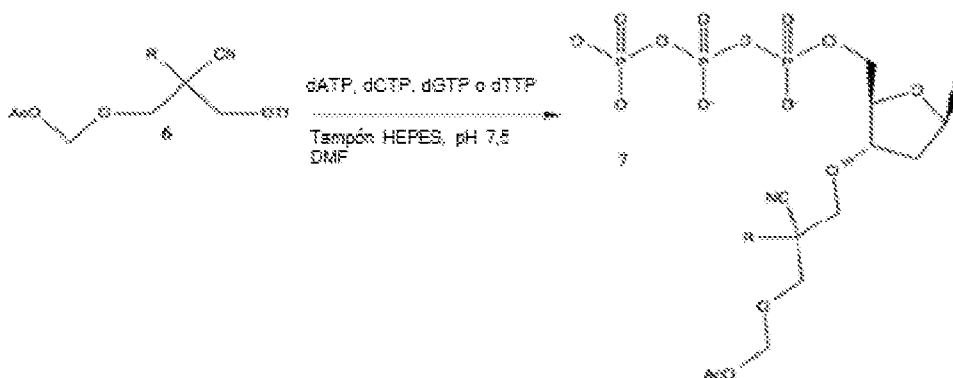
Se disuelven 10 mmol del compuesto 4 en 30 ml de THF seco y se añaden 2 ml (12 mmol) de trihidrofluoruro de trietilamina. Se agita la reacción durante 1 semana. Se añaden 18 ml de acetato de trietilamonio 2 M acuoso y se evapora la mezcla hasta sequedad a vacío. Se purifica el residuo que incluye el compuesto 5 por cromatografía en gel de sílice con CH_2Cl_2 y metanol como disolventes. En este ejemplo, R puede ser $-\text{CN}$ o $-\text{C}(\text{O})\text{O}-\text{Et}$.



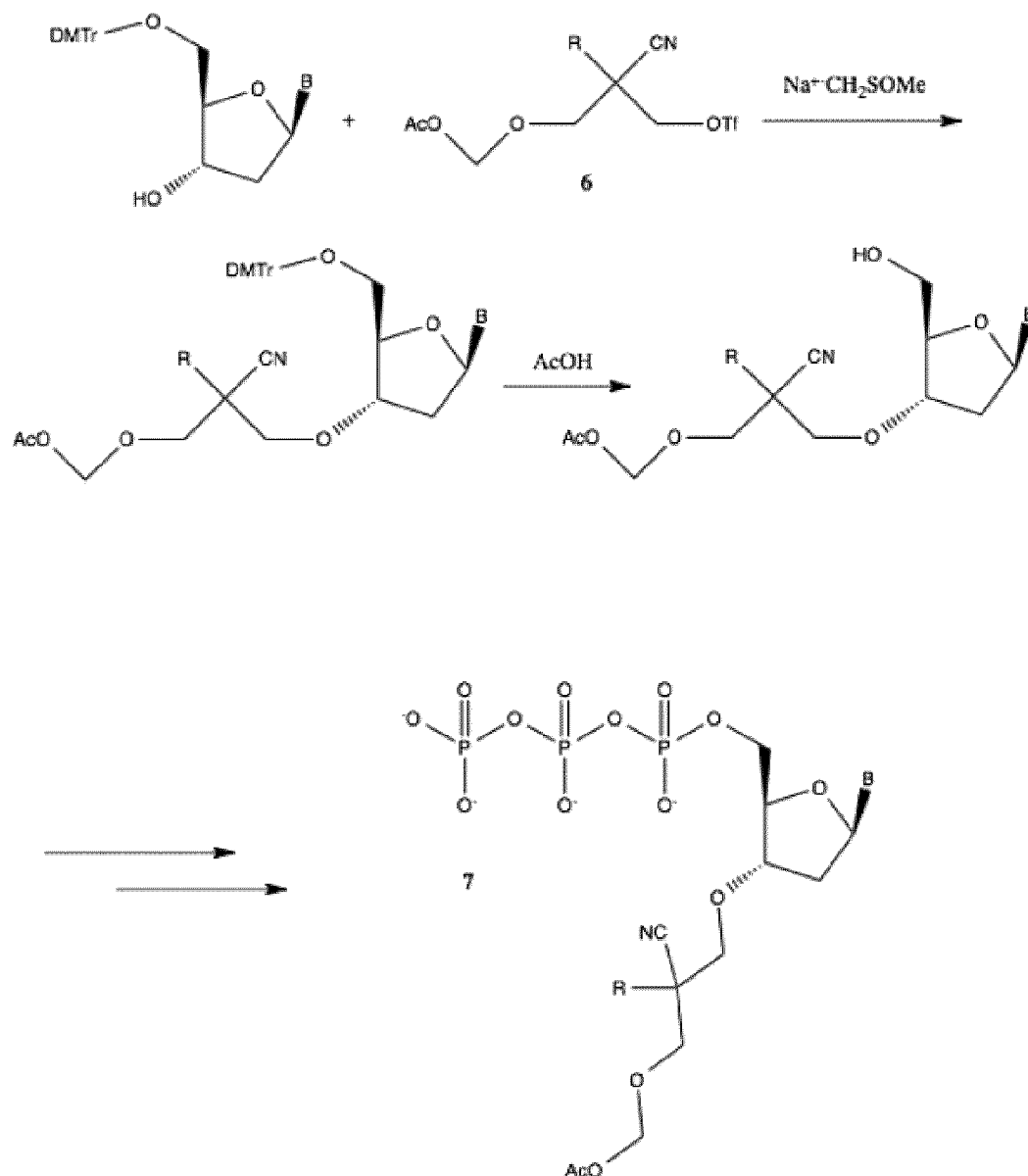
Se disuelven 10 mmol del compuesto 5 en 20 ml de 1,2-dicloroetano y se añaden 20 mmol de piridina y se agita la solución a -10°C . Se añaden 12 mmol de anhídrido trifluorometanosulfónico preenfriado en 10 ml de 1,2-dicloroetano y se agita la mezcla durante 30 min a -10°C . A continuación se desactiva con 200 ml de NaHCO_3 acuoso al 5 % y se agita la mezcla durante 30 min a temperatura ambiente. Se separa la capa orgánica, se seca sobre sulfato de sodio y se concentra a vacío. El residuo que incluye el compuesto 6 se usa directamente en la siguiente etapa sin purificación adicional. En este ejemplo, R puede ser $-\text{CN}$ o $-\text{C}(\text{O})\text{O}-\text{Et}$.



Se disuelven 2 mmol del compuesto 6 en 1 ml de DMF y se añade a una solución de 1 ml de dATP, dCTP, dGTP o dTTP 1 M en tampón HEPES 50 mM, pH 7,5. Se agita la reacción a 37°C durante 24 horas. A continuación, se retira el disolvente a vacío y se purifica el residuo por HPLC de fase inversa usando bicarbonato de trietilamonio 50 mM, pH 7,5 y acetonitrilo como disolventes. Se retira el disolvente y se precipita el producto como sal de sodio usando perclorato de sodio en acetona. Se analiza la identidad y pureza del producto 7 por procedimientos analíticos de CL-EM. En este ejemplo, R puede ser $-\text{CN}$ o $-\text{C}(\text{O})\text{O}-\text{Et}$.

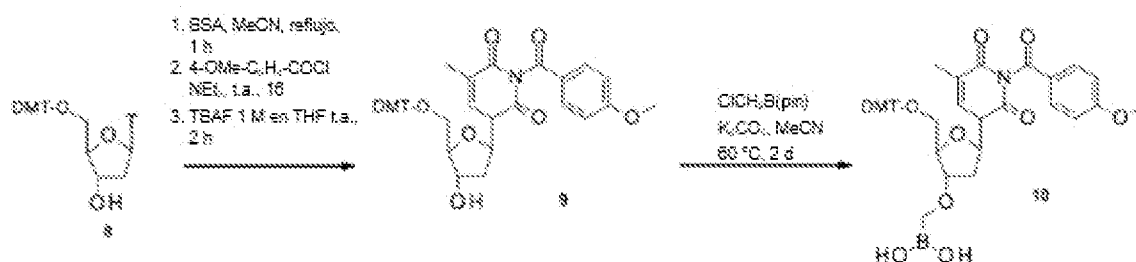


A continuación se expone un procedimiento alternativo para la preparación:



Se disuelven los desoxinucleótidos protegidos con 5'-dimetoxitritilo (DMTr-) en THF seco y se tratan lentamente con 1,2 equivalentes de dimislo sódico. Después de agitar la reacción durante 30 minutos, se añaden lentamente a la mezcla 1,2 equivalentes de triflatos (compuesto 6). Después de que se completa la reacción, se evapora el disolvente y se purifica el producto por cromatografía en columna en gel de sílice. Se retira el grupo protector DMTr por tratamiento con ácido acético y se obtienen los trifosfatos (compuesto 7) por medio de procedimientos de preparación de trifosfato clásicos y se purifican por procedimientos de purificación de intercambio iónico y HPLC de fase inversa. En este ejemplo, R puede ser $-\text{CN}$ o $-\text{C}(\text{O})\text{O}-\text{Et}$.

5'-O-DMT-N3-anisoil-timidina 9



A una solución de 5'-O-DMT-timidina **8** (5,08 g, 9,33 mmol) en MeCN seco (100 ml) se le añadió *N,O*-bis-trimetilsilil-acetamida (4,60 ml, 3,83 g, 18,8 mmol). Se mantuvo a reflujo la mezcla de reacción durante 1 h y posteriormente se enfrió hasta t.a. Seguidamente, se añadieron cloruro de 4-metoxibenzoilo (1,65 ml, 2,08 g, 12,2 mmol) y NEt₃ (2,60 ml, 1,89 g, 18,7 mmol) y se agitó la mezcla de reacción durante la noche a t.a. Se añadió TBAF (1 M en THF, 28,5 ml, 28,5 mmol) y se agitó la mezcla resultante a t.a. Después de 2 h, se concentró la mezcla a vacío y se disolvió el producto bruto en EtOAc (200 ml). Se lavó la fase orgánica (2 x 100 ml de NaHCO₃ sat.; 100 ml de salmuera) y se secó sobre Na₂SO₄. Se aisló 5'-O-DMT-*N*3-anisoil-timidina **9** (1,95 g, 2,78 mmol, 31 % en tres etapas) como un sólido amarillo después de la purificación por cromatografía ultrarrápida en gel de sílice usando *n*-hexano/acetato de etilo (2/1→1/2) como fase móvil.

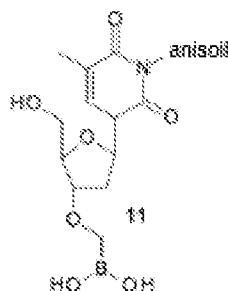
ESI-EM: m/z = 677,9 (C₃₉H₃₇N₂O₉ (M-H⁻); calc. 677,2 (M-H⁻)).

Ácido 3'-O-metilborónico-5'-O-DMT-*N*3-anisoil-timidina **10**

Se disolvió 5'-O-DMT-*N*3-anisoil-timidina **9** (648 mg, 1,00 mmol) en MeCN seco (20 ml) bajo atmósfera de argón y en presencia de tamices moleculares activados (4 Å). Se añadió K₂CO₃ (415 mg, 3,00 mmol) y se agitó la suspensión durante 30 min a t.a. Se añadió 2-(clorometil)-4,4,5,5-tetrametil-1,3,2-dioxaborolano (351 µl, 442 mg, 2,00 mmol) y se agitó la mezcla de reacción a 60 °C. Después de 24 h, se añadió K₂CO₃ adicional (138 mg, 1,00 mmol) y del cloruro de alquilo (176 µl, 222 mg, 1,00 mmol). Se continuó agitando a 60 °C durante otras 16 h. Se enfrió la mezcla hasta t.a., se filtró y se concentró a vacío. Se purificó el producto bruto por cromatografía ultrarrápida en gel de sílice usando *n*-hexano/acetato de etilo (1/2→1/4) y diclorometano/metanol (2/1) como fase móvil. La purificación final por RP-HPLC dio el ácido 3'-O-metilborónico-5'-O-DMT-*N*3-anisoil-timidina **10** deseado (167 mg, 0,23 mmol, 23 %) como un sólido blanco después de la liofilización.

ESI-EM: m/z = 735,4 (C₄₀H₄₀BN₂O₁₁ (M-H⁻); calc. 735,3 (M-H⁻)).

Ácido 3'-O-metilborónico-*N*3-anisoil-timidina **11**

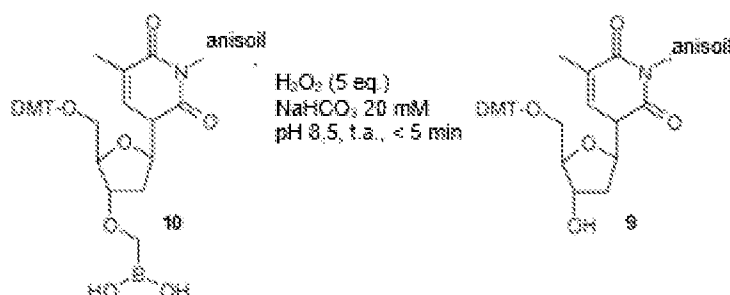


Se formó ácido 3'-O-metilborónico-*N*3-anisoil-timidina **11** como un producto secundario durante la síntesis de ácido 3'-O-metilborónico-5'-O-DMT-*N*3-anisoil-timidina **10** por desprotección parcial del grupo protector 5'-DMT y se aisló en el transcurso de la purificación por RP-HPLC. Se aisló ácido 3'-O-metilborónico-*N*3-anisoil-timidina **11** (5 mg, 11,5 µmol) como un sólido blanco después de la liofilización.

ESI-EM: m/z = 433,2 (C₁₉H₂₂BN₂O₉ (M-H⁻); calc. 433,1 (M-H⁻)).

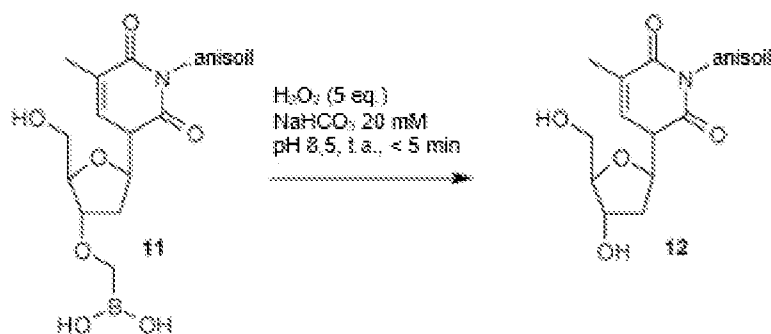
Desprotección

Ácido 3'-O-metilborónico-5'-O-DMT-*N*3-anisoil-timidina **10** → 5'-O-DMT-*N*3-anisoil-timidina **9**



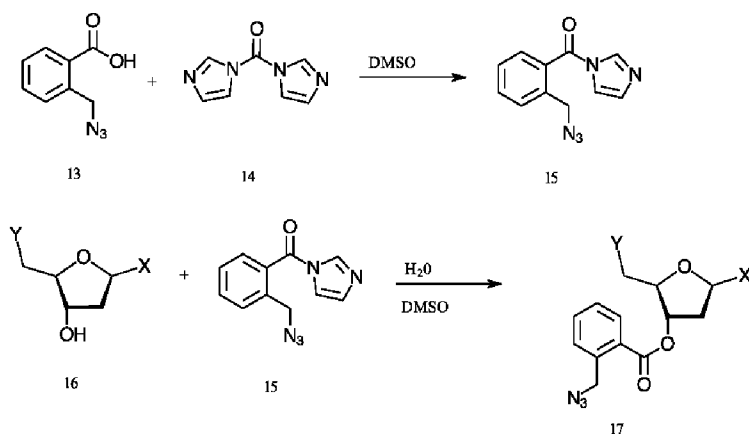
Se disolvió el nucleósido protegido por ácido metilborónico **10** (1,23 mg, 1,67 μ mol) en 134 μ l de MeCN ($c(\text{nucleósido}) = 12,5 \text{ mM}$). Se diluyeron 20 μ l de esta solución con 80 μ l de MeCN y se complementaron con 100 μ l de NaHCO_3 40 mM ac. ($c(\text{nucleósido}) = 1,25 \text{ mM}$). Se añadieron 50 μ l de la solución a 1 μ l de H_2O_2 al 1 % ac. y se agitó la mezcla de reacción a t.a. Como se indica por el análisis de HPLC-EM, la conversión cuantitativa del material de partida **10** (ESI-EM: $m/z = 735,4$ ($\text{C}_{40}\text{H}_{40}\text{BN}_2\text{O}_{11}$ (M-H^-); calc. 735,3 (M-H^-)) en el alcohol deseado **9** (ESI-EM: $m/z = 677,5$ ($\text{C}_{39}\text{H}_{37}\text{N}_2\text{O}_9$ (M-H^-); calc. 677,2 (M-H^-)) avanzó en menos de 5 minutos.

Ácido 3'-O-metilborónico-N3-anisoil-timidina **11** $\rightarrow$ N3-anisoil-timidina **12**



Se disolvió el nucleósido protegido por ácido metilborónico **11** (0,75 mg, 1,73 μ mol) en 138 μ l de MeCN ($c(\text{nucleósido}) = 12,5 \text{ mM}$). Se diluyeron 20 μ l de esta solución con 160 μ l de NaHCO_3 20 mM ac. ($c(\text{nucleósido}) = 1,25 \text{ mM}$). Se añadieron 50 μ l de la solución a 1 μ l de H_2O_2 al 1 %. Como se indica por el análisis de HPLC-EM, la conversión cuantitativa del material de partida **11** (ESI-EM: $m/z = 433,3$ ($\text{C}_{19}\text{H}_{22}\text{BN}_2\text{O}_9$ (M-H^-); calc. 433,1 (M-H^-)) en el alcohol deseado **12** (ESI-EM: $m/z = 375,3$ ($\text{C}_{18}\text{H}_{19}\text{N}_2\text{O}_7$ (M-H^-); calc. 375,1 (M-H^-)) avanzó en menos de 5 minutos.

Síntesis de nucleótido trifosfato bloqueado con aril(acidometilo)



Las estructuras identificadas anteriormente, Y y X, son como se describe en el presente documento.

Se disolvieron 1,22 gramos de ácido 2-(acidometil)benzoico **13** en 8,1 ml de DMSO anhidro y se agitó la solución transparente a temperatura ambiente. Se añadieron 1,11 gramos de N,N'-carbonyldiimidazol **14** en porciones con agitación (1 equivalente). Tras la disolución del CDI, se agitó la reacción a temperatura ambiente durante 2 horas y se usó la mezcla resultante (producto:imidazol, 1:1) como una solución madre 0,85 M de la acilimidazolida activada **15**.

Se colocaron 55 mg de dATP **16** (sal disódica) en un tubo de reacción. Se añadieron 182 μ l de solución de acilimidazol activado **15** (0,85 M en DMSO), seguida de 500 μ l de DMSO anhidro y 500 μ l de agua. Se incubó la solución transparente resultante a 55 $^{\circ}\text{C}$ durante la noche y a continuación se enfrió hasta temperatura ambiente. La CLEM analítica confirmó la formación del producto. **17** (PM obs. 649,04, PM calc. 649), así como un producto secundario difosfato (PM obs. 569,07). Se purificó el material por HPLC de fase inversa en una columna C18 usando un gradiente de acetonitrilo en acetato de trietilamonio 0,1 M, pH 7,5. Se recogieron las fracciones que contenían el producto y se liofilizó para dar un sólido blanco.

Reducción con TCEP de acidas bloqueadas

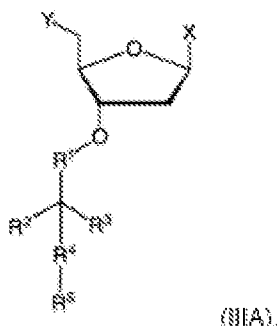
Se trató una solución 1 mM del dATP bloqueado con éster de (acidometil)benzoato **17** en HEPES 30 mM, pH 7,5, con TCEP (Tris(2-carboxietil)fosfina) 15 mM y se incubó a temperatura ambiente. Para el análisis por CLEM, se analiza una dilución 10x de la mezcla de reacción en agua usando un gradiente de acetonitrilo en acetato de trietilamonio 100 mM y detección a 260 nM y por ESI-EM en modo negativo. Antes de la escisión, el dATP bloqueado eluye a los 4 min en el gradiente de CL con un PM de 649,03 (modo negativo). Tras la reducción de la acida y la ciclación intramolecular y desprotección, el dATP liberado eluye a 1,8 min en el gradiente de CLEM con un PM de 489,99 (modo negativo).

Incorporación de un único nucleótido:

La eficacia de la incorporación de trifosfato 7 por una polimerasa se examina en un experimento de inserción de un único nucleótido usando un cebador de 15-mero y moldes de 25-mero que contienen las bases complementarias en los sitios de inserción. El éxito de la inserción se analiza usando procedimientos analíticos de CL-EM.

REIVINDICACIONES

1. Un nucleótido o una sal del mismo de fórmula (IIIA):



5

en la que

X es una nucleobase o una nucleobase marcada;

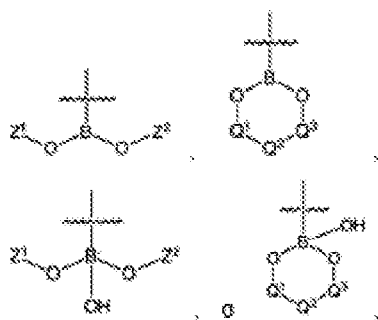
- 10 Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R^1 es un enlace, $-CH_2-$, $-C(O)-O-$, $-C(O)-NR^a-$ o $-C(O)-R^x-$;

- 15 R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-[(C(R^a)(R^b))_p-O]_q-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$, $-CN$ o $-NO_2$;

- 20 R^4 es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-CH=CH-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u $-O-C(O)-aril-$;

R^5 es $-(C(R^a)(R^b))_n-N_3$, $-(C(R^a)(R^b))_n-N^+C^-$, $-(C(R^a)(R^b))_n-CN$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:



25

Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

- 30 Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

35

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

40

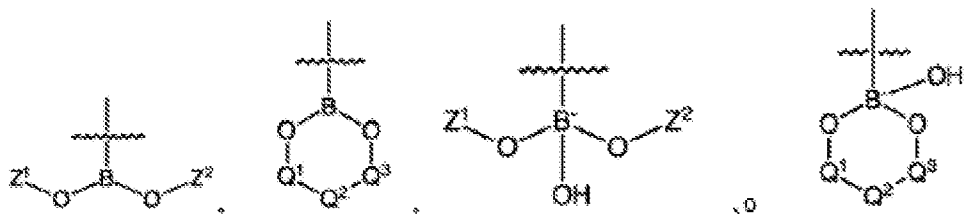
n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3;

siempre que cuando R^1 y R^4 sean ambos enlaces y cuando R^2 y R^3 sean ambos H, entonces R^5 no es una acida.

5

2. El nucleótido de la reivindicación 1, en el que R^5 es:



10

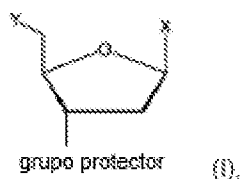
3. El nucleótido de una cualquiera de las reivindicaciones 1 - 2, en el que R^1 es $-C(O)-O-$.

4. El nucleótido de una cualquiera de las reivindicaciones 1 - 2, en el que R^1 es un enlace.

15

5. El nucleótido de una cualquiera de las reivindicaciones 1 - 2, en el que R^1 es $-CH_2-$ y R^5 es $-B(OZ^1)(OZ^2)$.

6. Un nucleótido o una sal del mismo de fórmula (I):



20

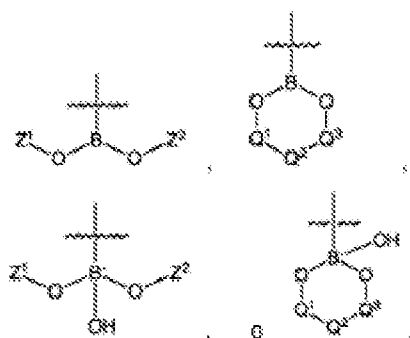
en la que

X es una nucleobase o una nucleobase marcada;

25

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O-$ oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

'grupo protector' tiene la estructura:



30

donde

35

Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2; y

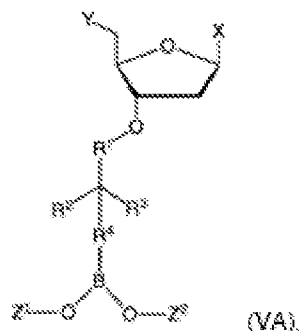
Re y Rf son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido.

7. El nucleótido de la reivindicación 6, en el que el 'grupo protector' es $-B(OZ^1)(OZ^2)$.

8. El nucleótido de la reivindicación 6, en el que Z^1 y Z^2 se seleccionan independientemente del grupo que consiste en H, metilo y etilo.

9. El nucleótido de la reivindicación 6, en el que el 'grupo protector' es $-B(OH)_2$.

10. Un nucleótido o una sal del mismo de fórmula (VA):



en la que

X es una nucleobase o una nucleobase marcada;

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R¹ es un enlace, un -CH₂-, -C(O)-O-, -C(O)-NR^a- o -C(O)-R^x-;

R² y R³ son cada uno independientemente H, un grupo alquilo C₁-C₆ saturado o insaturado, un grupo arilo o heteroarilo C₅-C₆, un halógeno, -[(C(R^a)(R^b))P-O]_q-(R^a), -C(O)-OR^a, -C(O)-N(R^a)(R^b), -(C(R^a)(R^b))_n-N⁺C⁻, -CN o -NO₂.

R⁴ es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, -CH=CH-, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u -O-C(O)-aril-;

donde Z^1 y Z^2 son independientemente H o un grupo alquilo C_1 - C_4 ;

R^a y R^b son cada uno independientemente H o un grupo alquilo C₁-C₆ saturado;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3.

11. El nucleótido de la reivindicación 10, en el que al menos uno de Z^1 o Z^2 es H.

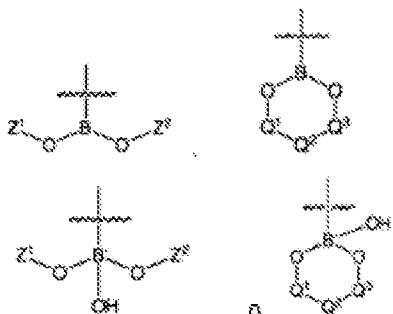
12. El nucleótido de la reivindicación 10, en el que tanto Z^1 como Z^2 son H.

13. El nucleótido de una cualquiera de las reivindicaciones 10-11, en el que R^1 es un enlace.

14. El nucleótido de la reivindicación 10, en el que R¹ es -CH₂- y tanto Z¹ como Z² son H.

15. Un procedimiento de secuenciación de una muestra de ácido nucleico con la ayuda de un nanoporo, que comprende:

(a) realizar una reacción de polimerización con la ayuda de una única polimerasa acoplada al nanoporo, incorporando la reacción de polimerización uno de al menos cuatro nucleótidos protegidos en 3' diferentes en una hebra de polinucleótido creciente complementaria a una molécula de ácido nucleico monocatenaria derivada de la muestra de ácido nucleico, incluyendo cada nucleótido protegido en 3' individual de los al menos cuatro nucleótidos protegidos en 3' diferentes que comprende un grupo bloqueador un resto que tiene una estructura:



donde

Z¹ y Z² son independientemente H, a grupo alquilo C₁-C₄, un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q¹ y Q³ son cada uno independientemente un enlace, -C(R^e)(R^f)- o -C(O)-;

Q² es un enlace, o-fenileno o -[C(R^e)(R^f)]_w-, donde w es 1 o 2; y

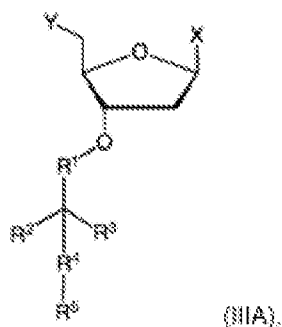
R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

(b) detectar, con la ayuda del nanoporo, la incorporación del nucleótido protegido en 3' o un subproducto del mismo en la hebra de polinucleótido creciente; y

(c) simultáneamente con o bien posterior a la etapa de detectar la incorporación del nucleótido protegido en 3' o el subproducto del mismo, desproteger el nucleótido protegido en 3' para proporcionar un nucleótido desprotegido en 3'.

16. Un procedimiento para secuenciar una molécula de ácido nucleico, comprendiendo el procedimiento:

(a) obtener un chip que comprende una pluralidad de nanoporos individualmente direccionables, en el que cada nanoporo individualmente direccionable de la pluralidad de nanoporos individualmente direccionables comprende un nanoporo en una membrana que se dispone contigua a un electrodo, en el que el nanoporo se enlaza a una enzima y en el que cada nanoporo individualmente direccionable se adapta para detectar una marca que se libera de un nucleótido protegido en 3', en el que el nucleótido protegido en 3' se representa por la fórmula (III A):



en la que

X es una nucleobase marcada;

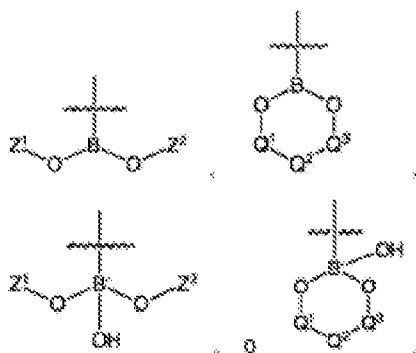
Y es -O-P(O)(OH)-[O-P(O)(OH)]_z-OH u -O-P(O)(OH)-[O-P(O)(OH)]_z-O-oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R^1 es un enlace, $-\text{CH}_2-$, $-\text{C}(\text{O})-\text{O}-$, $-\text{C}(\text{O})-\text{NR}^a-$ o $-\text{C}(\text{O})-\text{R}^x-$;

5 R^2 y R^3 son cada uno independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-\text{[(C(R}^a\text{)(R}^b\text{))}_p\text{O]}_q-(\text{R}^a)$, $-\text{C}(\text{O})-\text{OR}^a$, $-\text{C}(\text{O})-\text{N}(\text{R}^a)(\text{R}^b)$, $-(\text{C(R}^a\text{)(R}^b\text{))}_n\text{N}^+\text{C}^-$, $-\text{CN}$ o $-\text{NO}_2$;

10 R^4 es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-\text{CH}=\text{CH}-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u $-\text{O}-\text{C}(\text{O})-\text{aril}-$;

R^5 es $-(\text{C(R}^a\text{)(R}^b\text{))}_n\text{N}_3$, $-(\text{C(R}^a\text{)(R}^b\text{))}_n\text{N}^+\text{C}^-$, $-(\text{C(R}^a\text{)(R}^b\text{))}_n\text{CN}$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:



15 Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

20 Q^1 y Q^3 son cada uno independientemente un enlace, $-\text{C(R}^e\text{)(R}^f\text{)}-$ o $-\text{C}(\text{O})-$;

Q^2 es un enlace, o-fenileno o $-\text{[C(R}^e\text{)(R}^f\text{))}_w-$, donde w es 1 o 2;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

25 R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

30 n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3; siempre que cuando R^1 y R^4 son ambos enlaces y cuando R^2 y R^3 son ambos H, entonces R^5 no es un grupo ácido;

35 (b) dirigir la molécula de ácido nucleico contigua a o en proximidad al nanoporo;

(c) con la ayuda de la enzima, polimerizar el nucleótido protegido en 3' a lo largo de la molécula de ácido nucleico para generar una hebra de polinucleótido que es complementaria a al menos una porción de la molécula de ácido nucleico;

40 (d) desproteger el nucleótido protegido en 3' después de la incorporación del nucleótido protegido en 3' en la hebra de polinucleótido; y

45 (e) detectar la marca liberada del nucleótido protegido en 3' con la ayuda del electrodo, en el que la marca liberada fluye a través del o en proximidad al nanoporo,

donde las etapas (d) y (e) se pueden realizar en cualquier orden.

50 17. Un procedimiento de secuenciación de una muestra de ácido nucleico con la ayuda de un nanoporo, que comprende: (a) realizar una reacción de polimerización con la ayuda de una única polimerasa acoplada al nanoporo, incorporando la reacción de polimerización al menos uno de los nucleótidos protegidos de cualquiera de las reivindicaciones 1 a 14 en una hebra de polinucleótido creciente complementaria a una molécula de ácido

nucleico monocatenaria derivada de la muestra de ácido nucleico, detectando, con la ayuda del nanoporo, la incorporación del al menos un nucleótido protegido o un subproducto del mismo en la hebra de polinucleótido creciente; y simultáneamente con o bien posterior a la etapa de detectar la incorporación del nucleótido protegido o el subproducto del mismo, desproteger el nucleótido protegido para proporcionar un nucleótido desprotegido en 3'.

5

18. Un kit que comprende:

(a) un biochip que comprende:

10 (i) un sustrato semiconductor que tiene una pluralidad de pocillos con una densidad de al menos aproximadamente 250 pocillos/mm²; y

(ii) un electrodo dispuesto en cada uno de la pluralidad de pocillos; y

15 (b) al menos uno de los nucleótidos de una cualquiera de las reivindicaciones 1 a 14.

19. Un ensamblaje que comprende:

20 (a) un depósito, comprendiendo el depósito al menos uno de los nucleótidos de una cualquiera de las reivindicaciones 1 a 14;

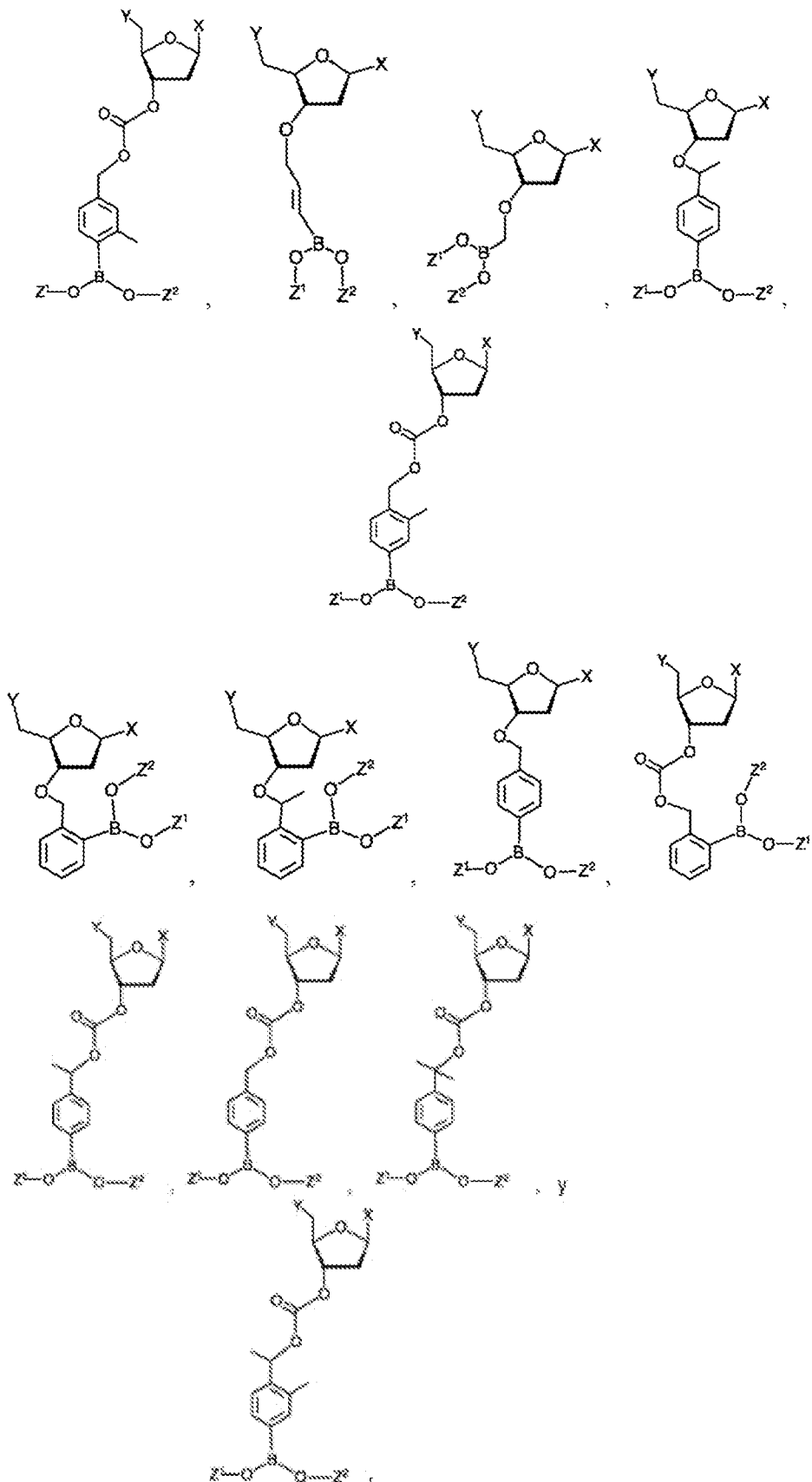
(b) un biochip que comprende:

25 (iii) un sustrato semiconductor que tiene una pluralidad de pocillos con una densidad de al menos aproximadamente 250 pocillos/mm²; y

(iv) un electrodo dispuesto en cada uno de la pluralidad de pocillos; y

30 (c) un contraelectrodo dispuesto en una superficie del depósito orientada hacia el biochip.

20. Un nucleótido o sal del mismo seleccionado del grupo que consiste en:



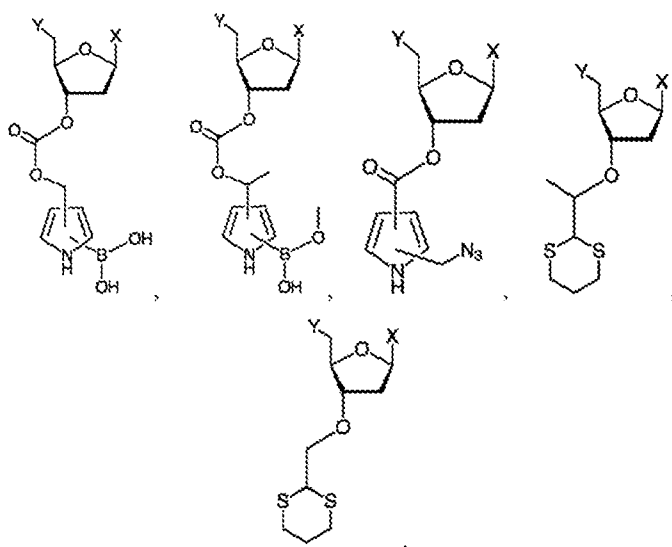
donde

5 X es una nucleobase o una nucleobase marcada;

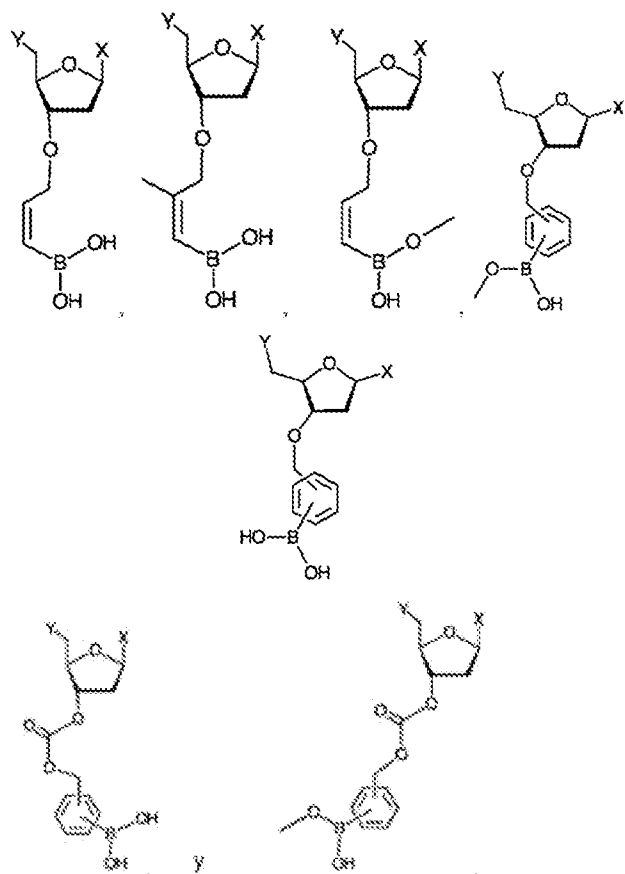
Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5; y

- 5 Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo.

21. Un nucleótido o sal del mismo seleccionado del grupo que consiste en:



10

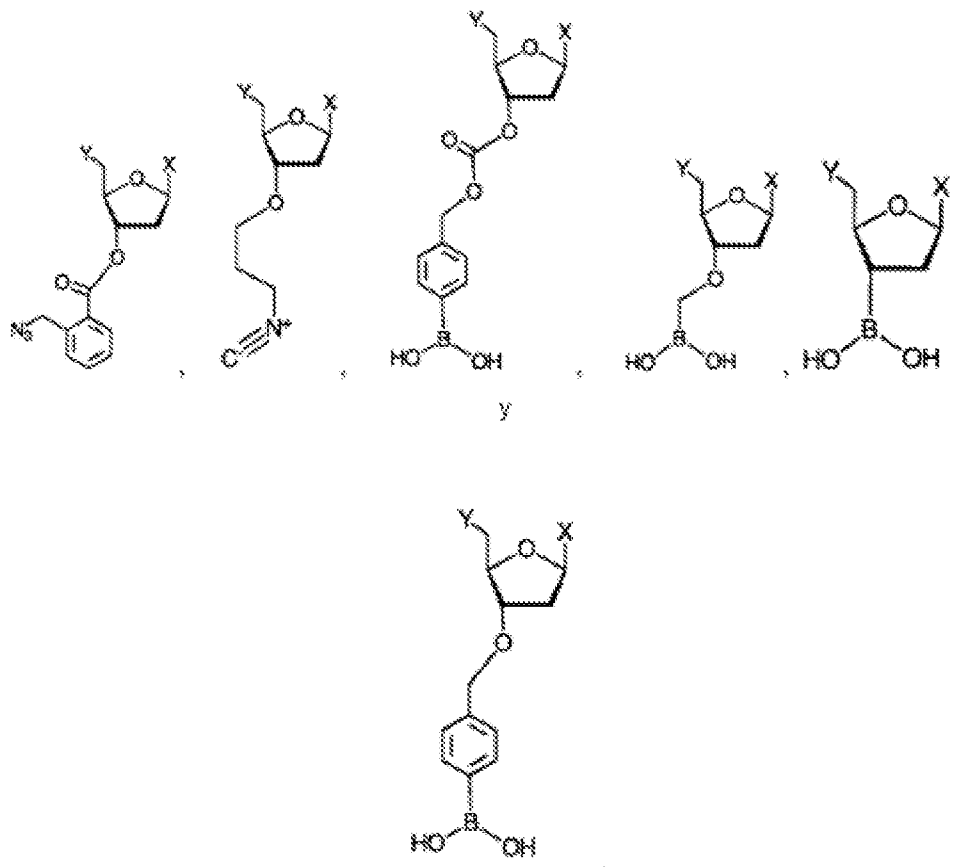


donde

X es una nucleobase o una nucleobase marcada; e

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5.

22. Un nucleótido o sal del mismo seleccionado del grupo que consiste en:



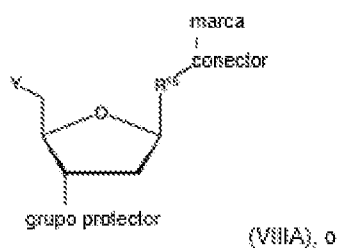
donde

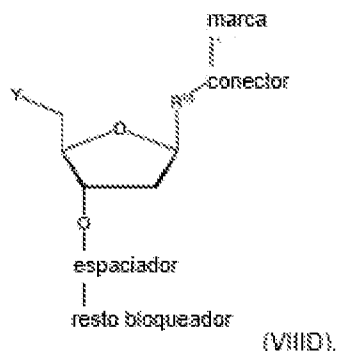
X es una nucleobase o una nucleobase marcada; e

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5.

23. Un kit que comprende al menos dos nucleótidos diferentes de una cualquiera de las reivindicaciones 20 a 22, en el que cada uno de los al menos dos nucleótidos diferentes comprende una nucleobase marcada diferente.

24. Un nucleótido o sal del mismo que tiene una cualquiera de las fórmulas (VIII A) o (VIII D):



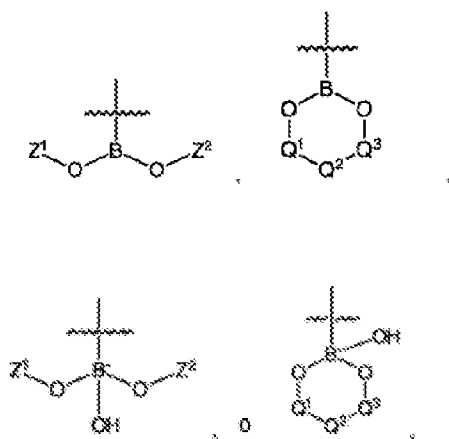


en la que

R^{15} es una nucleobase;

5 Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

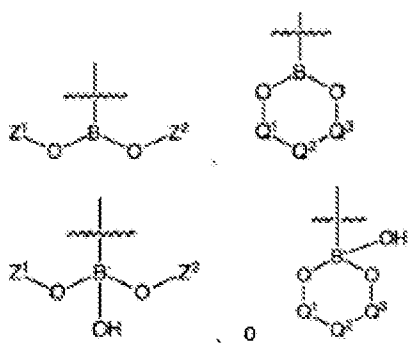
10 'grupo protector' tiene la estructura:



15 'conector' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 50 átomos de carbono y opcionalmente sustituido con uno o más heteroátomos; y

20 'marca' es una especie detectable, incluyendo cualquiera de las especies enumeradas en el presente documento, 'espaciador' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 16 átomos de carbono y opcionalmente sustituido con uno o más heteroátomos; 'resto bloqueador' es un grupo alifático o aromático, saturado o insaturado, sustituido o no sustituido, de cadena lineal o ramificado, que tiene entre 1 y 20 átomos de carbono, y opcionalmente sustituido con uno o más heteroátomos, y siempre que el 'resto bloqueador' incluya un grupo ácido, un grupo isonitrilo, un grupo heterocicloalquilo de 5 a 8 miembros que tiene al menos un heteroátomo seleccionado de O, N, S o Se, un resto derivado de un 1,4-epoxi-1,4-dihidronaftaleno sustituido o no sustituido o un grupo que tiene la estructura:

25



Z^1 y Z^2 son independientemente H, un grupo alquilo C_1 - C_4 , un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

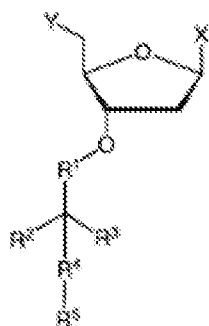
Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2; y

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

siempre que cuando el 'resto bloqueador' sea $-N_3$, el 'espaciador' no es $-CH_2-$.

25. Un procedimiento para determinar la secuencia de un polinucleótido monocatenario diana usando secuenciación por síntesis comprende: (a) proporcionar los nucleótidos marcados protegidos de una cualquiera de las reivindicaciones 1 - 14, 20 - 22 y 24; (b) incorporar el nucleótido marcado protegido en un complemento de un polinucleótido monocatenario diana; (c) detectar la marca del nucleótido marcado protegido de la etapa (b), determinando de este modo el tipo de nucleótido incorporado; (d) desproteger el nucleótido marcado protegido incorporado y retirar la marca del nucleótido marcado protegido de la etapa (b); y (e) repetir opcionalmente las etapas (b) - (d) una o más veces; determinando de este modo la secuencia del polinucleótido monocatenario diana.

26. Un nucleótido o una sal del mismo de fórmula (IIIA):



(IIIA).

en la que

X es una nucleobase o una nucleobase marcada;

Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R^1 es un enlace, $-CH_2-$, $-C(O)-O-$, $-C(O)-NR^a-$ o $-C(O)-R^x-$;

R^2 es H;

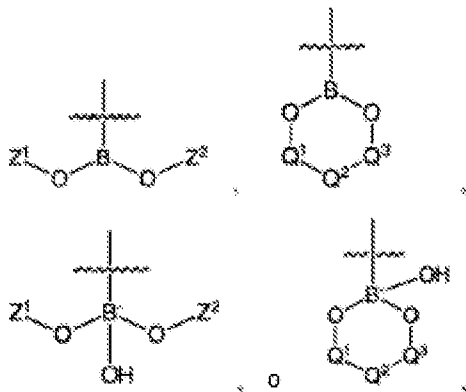
R^3 es un grupo alquilo C_1 - C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5 - C_6 , un halógeno, $-[(C(R^a)(R^b))_p-O]_q-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$, $-CN$ o $-NO_2$;

R^4 es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-CH=CH-$, un grupo heterocicloalquilo

de 5 o 6 miembros sustituido o no sustituido u $-\text{O}-\text{C}(\text{O})-\text{aril}-$;

R^5 es $-(\text{C}(\text{R}^a)(\text{R}^b))_n-\text{N}_3$, $-(\text{C}(\text{R}^a)(\text{R}^b))_n-\text{N}^+\text{C}^-$, $-(\text{C}(\text{R}^a)(\text{R}^b))_n-\text{CN}$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:

5



Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

10

Q^1 y Q^3 son cada uno independientemente un enlace, $-\text{C}(\text{R}^e)(\text{R}^f)-$ o $-\text{C}(\text{O})-$;

Q^2 es un enlace, o-fenileno o $-\text{C}(\text{R}^e)(\text{R}^f)_w-$, donde w es 1 o 2;

15

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

20

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

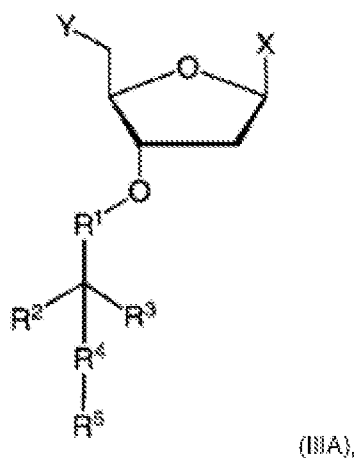
n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3;

25

siempre que cuando R^1 y R^4 sean ambos enlaces y cuando R^2 y R^3 sean ambos H, entonces R^5 no es una acida.

27. Un nucleótido o una sal del mismo de fórmula (IIIA):



30

en la que

X es una nucleobase o una nucleobase marcada;

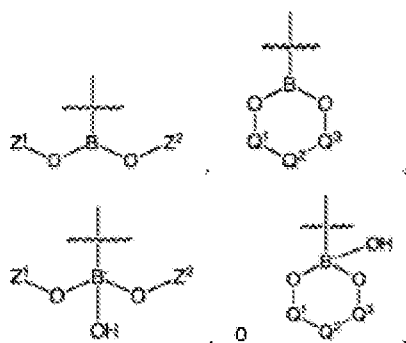
Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R¹ es un enlace, $-CH_2-$, $-C(O)-O-$, $-C(O)-NR^a-$ o $-C(O)-R^x-$;

R² y R³ son independientemente un grupo alquilo C₁-C₆ saturado o insaturado, un grupo arilo o heteroarilo C₅-C₆, un halógeno, $-[(C(R^a)(R^b))_p-O]_q-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$, $-CN$ o $-NO_2$;

R⁴ es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-CH=CH-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u $-O-C(O)-aril-$;

R⁵ es $-(C(R^a)(R^b))_n-N_3$, $-(C(R^a)(R^b))_n-N^+C^-$, $-(C(R^a)(R^b))_n-CN$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:



Z¹ y Z² son independientemente H, un grupo alquilo C₁-C₄ o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q¹ y Q³ son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

Q² es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

R^a y R^b son cada uno independientemente H o un grupo alquilo C₁-C₆ saturado;

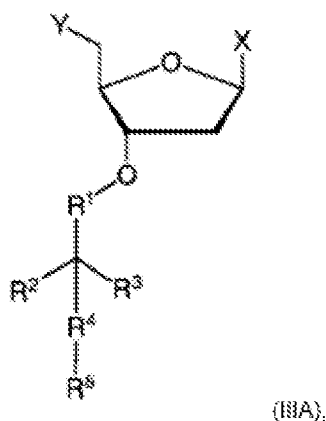
R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3.

28. Un nucleótido o una sal del mismo de fórmula (IIIA):



en la que

X es una nucleobase o una nucleobase marcada;

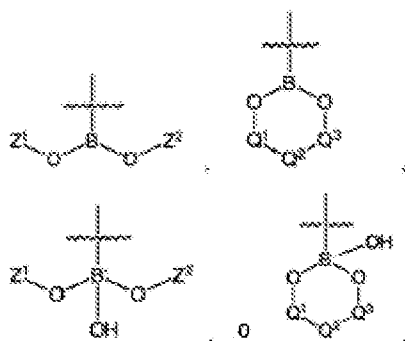
- 5 Y es $-O-P(O)(OH)-[O-P(O)(OH)]_z-OH$ u $-O-P(O)(OH)-[O-P(O)(OH)]_z-O$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R^1 es un enlace, $-CH_2-$, $-C(O)-O-$, $-C(O)-NR^a-$ o $-C(O)-R^x-$;

- 10 R^2 y R^3 son independientemente H, un grupo alquilo C_1-C_6 saturado o insaturado, un grupo arilo o heteroarilo C_5-C_6 , un halógeno, $-[(C(R^a)(R^b))_p-O]_q-(R^a)$, $-C(O)-OR^a$, $-C(O)-N(R^a)(R^b)$, $-(C(R^a)(R^b))_n-N^+C^-$, $-CN$ o $-NO_2$;

R^4 es un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-CH=CH-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u $-O-C(O)-aril-$;

- 15 R^5 es $-(C(R^a)(R^b))_n-N_3$, $-(C(R^a)(R^b))_n-N^+C^-$, $-(C(R^a)(R^b))_n-CN$, un grupo cicloalquilo de 5 a 8 miembros que comprende dos átomos de azufre situados 1, 3 entre sí o un grupo que tiene la estructura:



- 20 Z^1 y Z^2 son independientemente H, un grupo alquilo C_1-C_4 o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q^1 y Q^3 son cada uno independientemente un enlace, $-C(R^e)(R^f)-$ o $-C(O)-$;

- 25 Q^2 es un enlace, o-fenileno o $-[C(R^e)(R^f)]_w-$, donde w es 1 o 2;

R^a y R^b son cada uno independientemente H o un grupo alquilo C_1-C_6 saturado;

- 30 R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

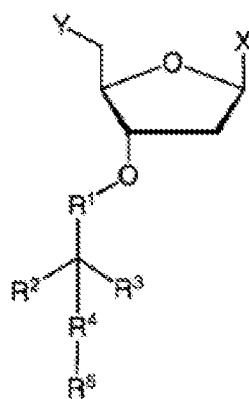
R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

- 35 n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3;

siempre que cuando R^1 y R^4 sean ambos enlaces y cuando R^2 y R^3 sean ambos H, entonces R^5 no es una acida.

- 40 29. Un nucleótido o una sal del mismo de fórmula (IIIA):



(IIA).

en la que

X es una nucleobase o una nucleobase marcada;

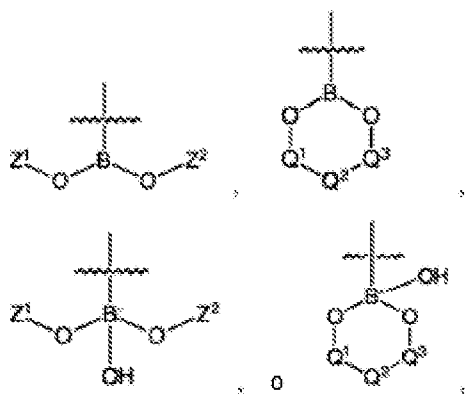
Y es $-\text{O}-\text{P}(\text{O})(\text{OH})-[\text{O}-\text{P}(\text{O})(\text{OH})]_z-\text{OH}$ u $-\text{O}-\text{P}(\text{O})(\text{OH})-[\text{O}-\text{P}(\text{O})(\text{OH})]_z-\text{O}$ -oligonucleótido, donde z es 0 o un número entero que varía de 1 a 5;

R¹ es un enlace, $-\text{CH}_2-$, $-\text{C}(\text{O})-\text{O}-$, $-\text{C}(\text{O})-\text{NR}^a-$ o $-\text{C}(\text{O})-\text{R}^x-$;

R² y R³ son cada uno independientemente H, un grupo alquilo C₁-C₆ saturado o insaturado, un grupo arilo o heteroarilo C₅-C₆, un halógeno, $-\text{C}(\text{R}^a)(\text{R}^b)_p-\text{O}-(\text{R}^a)$, $-\text{C}(\text{O})-\text{OR}^a$, $-\text{C}(\text{O})-\text{N}(\text{R}^a)(\text{R}^b)$, $-(\text{C}(\text{R}^a)(\text{R}^b))_n-\text{N}^+\text{C}^-$, $-\text{CN}$ o $-\text{NO}_2$;

R⁴ es un enlace, un grupo arilo de 5 a 7 miembros sustituido o no sustituido, $-\text{CH}=\text{CH}-$, un grupo heterocicloalquilo de 5 o 6 miembros sustituido o no sustituido u $-\text{O}-\text{C}(\text{O})-\text{aril}-$;

R⁵ es



Z¹ y Z² son independientemente H, un grupo alquilo C₁-C₄ o un grupo arilo de 5 a 6 miembros opcionalmente sustituido con uno o más grupos hidroxilo;

Q¹ y Q³ son cada uno independientemente un enlace, $-\text{C}(\text{R}^e)(\text{R}^f)-$ o $-\text{C}(\text{O})-$;

Q² es un enlace, o-fenileno o $-\text{C}(\text{R}^e)(\text{R}^f)_w-$, donde w es 1 o 2;

R^a y R^b son cada uno independientemente H o un grupo alquilo C₁-C₆ saturado;

R^e y R^f son independientemente H, metilo, etilo, isopropilo o un grupo arilo de 5 o 6 miembros sustituido o no sustituido;

R^x es un grupo aromático o grupo heteroaromático de 5 o 6 miembros sustituido o no sustituido;

n es 0 o un número entero que varía de 1 a 3; y

p y q son cada uno independientemente cero o un número entero que varía de 1 a 3;

siempre que cuando R^1 y R^4 sean ambos enlaces y cuando R^2 y R^3 sean ambos H, entonces R^5 no es una acida.

5

30. Un procedimiento para determinar una secuencia de un polinucleótido monocatenario diana usando secuenciación por síntesis comprende: (a) proporcionar los nucleótidos marcados protegidos de una cualquiera de las reivindicaciones 26-29; (b) incorporar el nucleótido marcado protegido en un complemento de un polinucleótido monocatenario diana; (c) detectar la marca del nucleótido marcado protegido de la etapa (b), determinando de este modo el tipo de nucleótido incorporado; (d) desproteger el nucleótido marcado protegido incorporado y retirar la marca del nucleótido marcado protegido de la etapa (b); y (e) repetir opcionalmente las etapas (b) - (d) una o más veces; determinando de este modo la secuencia del polinucleótido monocatenario diana.

10

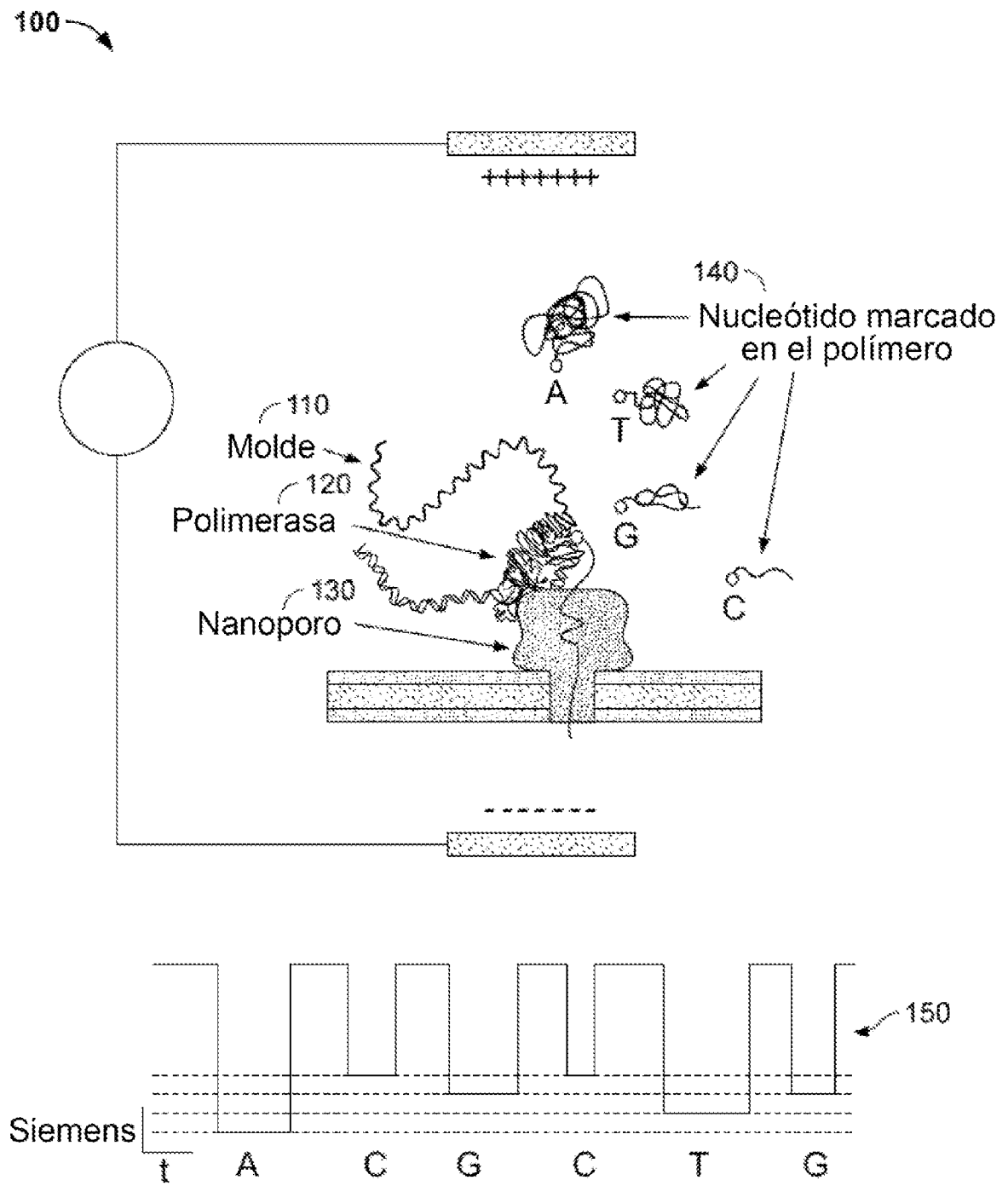


FIG. 1

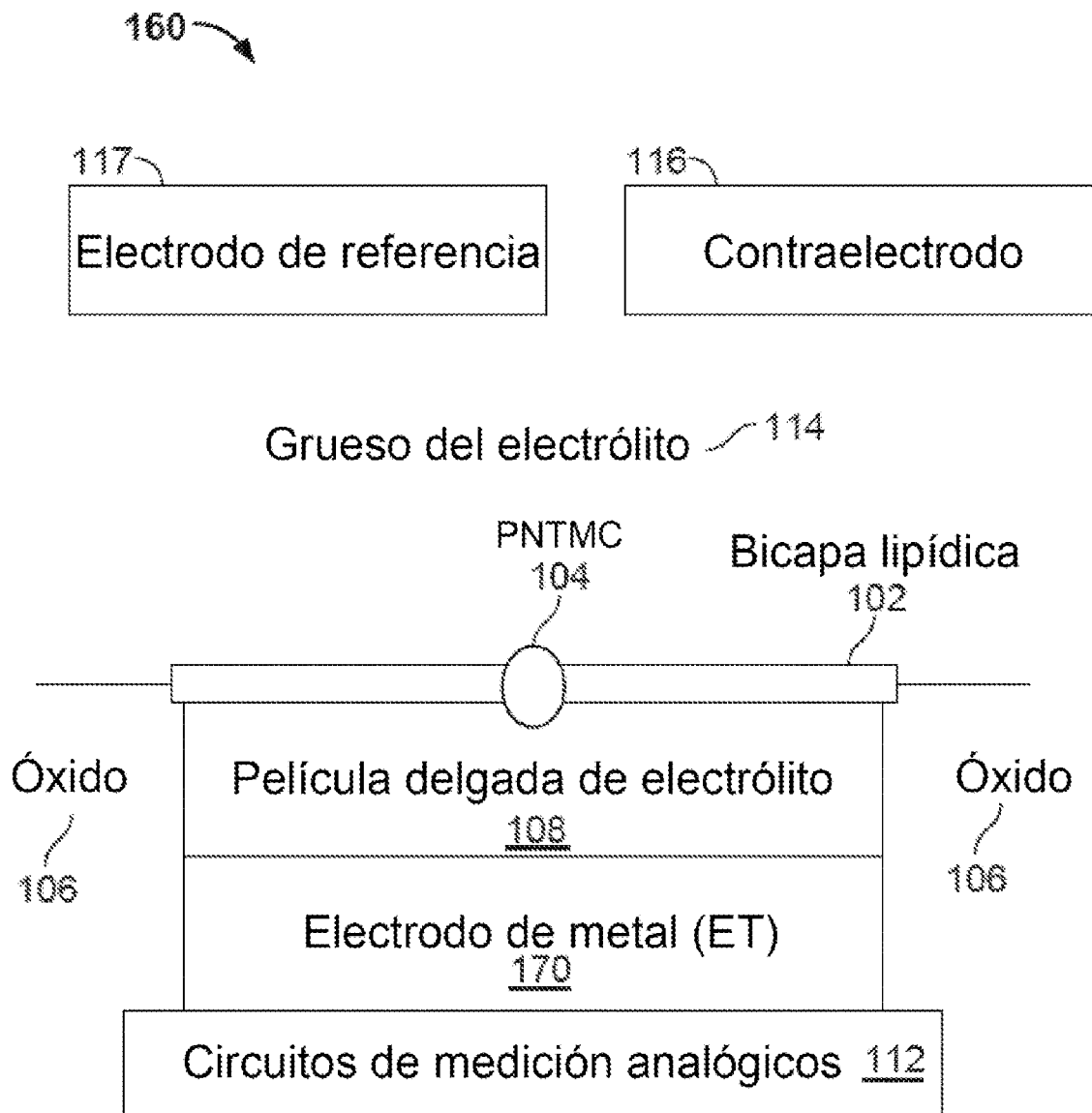


FIG. 2

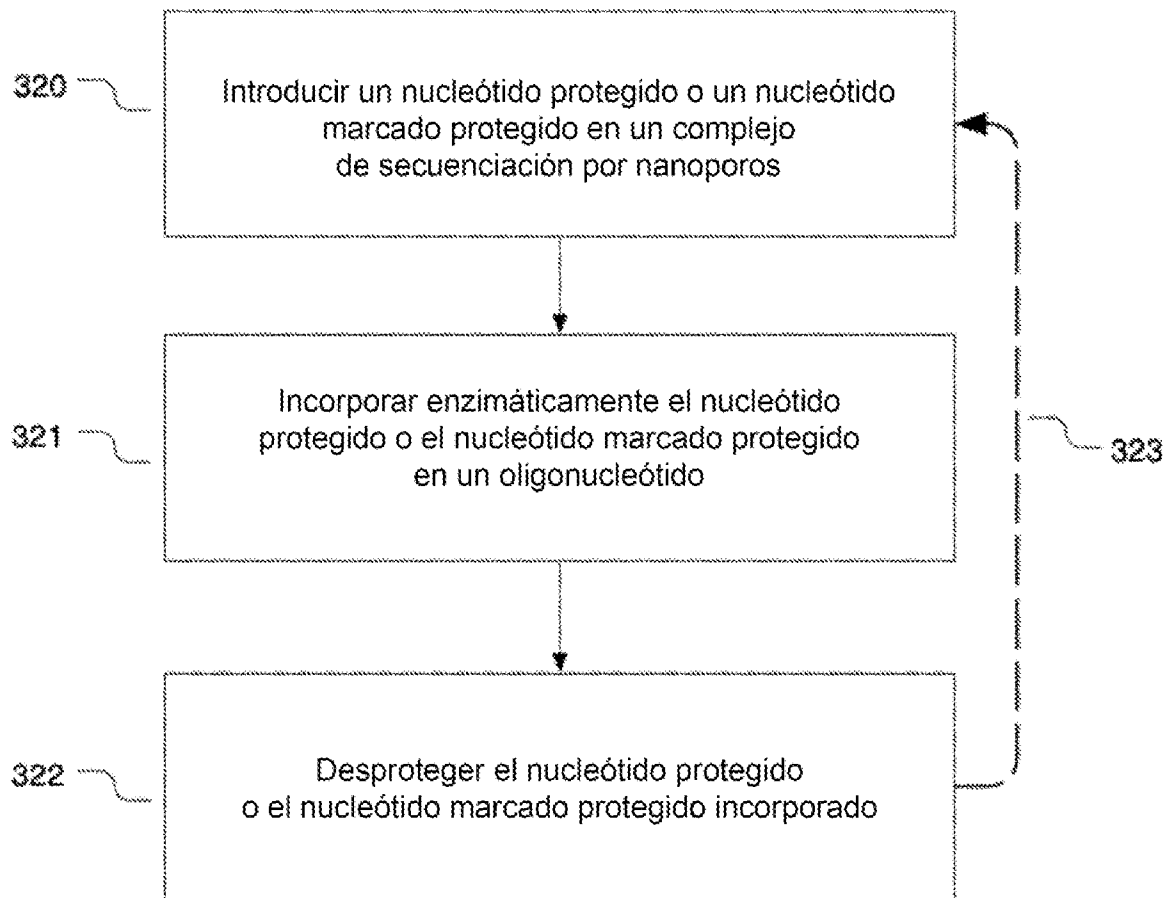


FIG. 3A

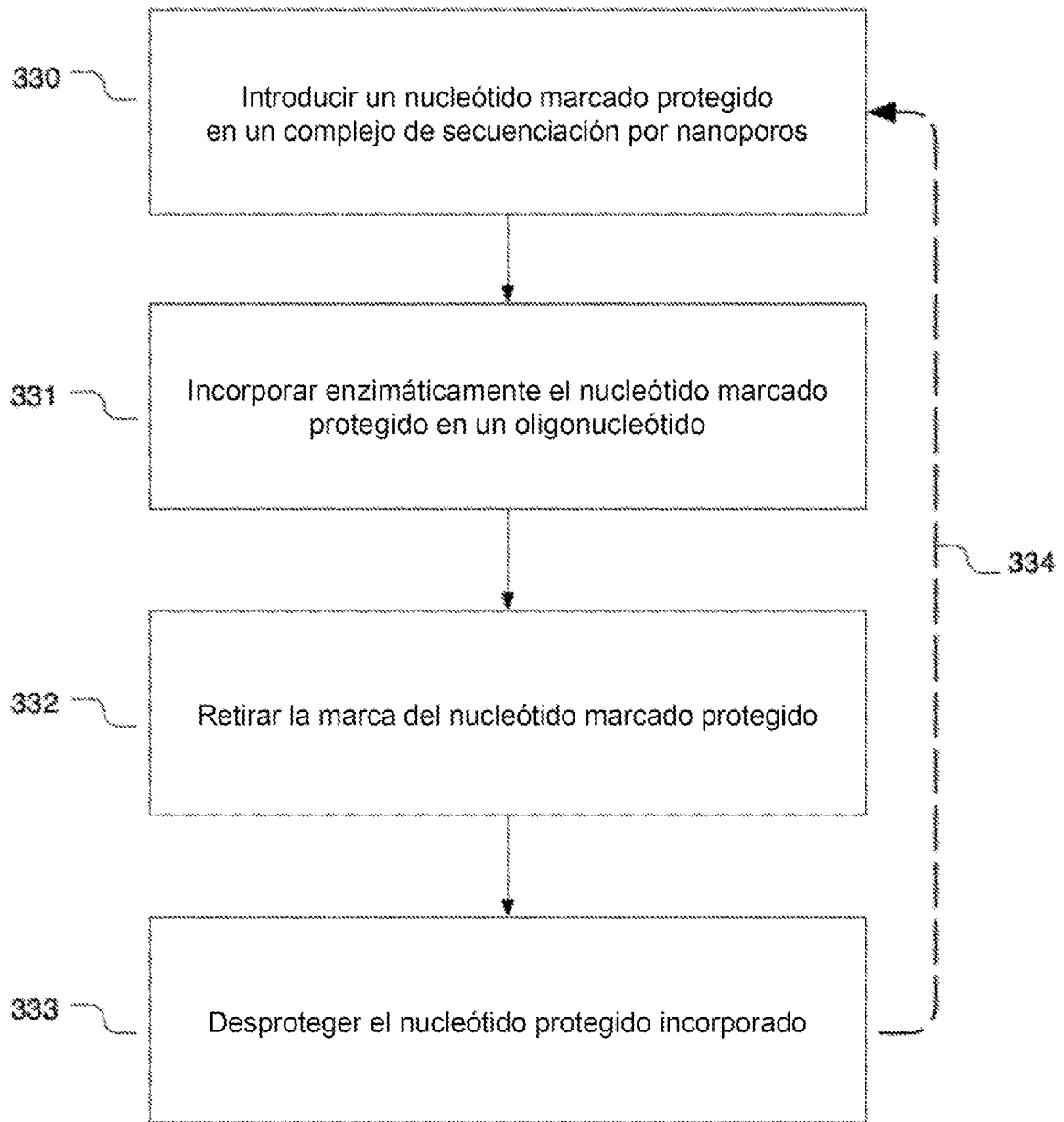


FIG. 3B

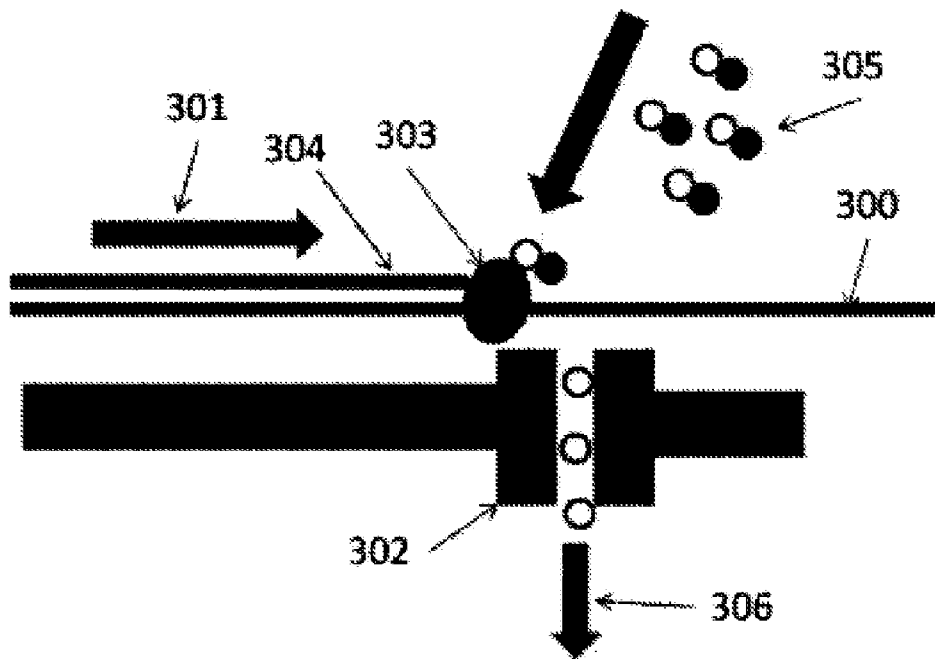


FIG. 3C

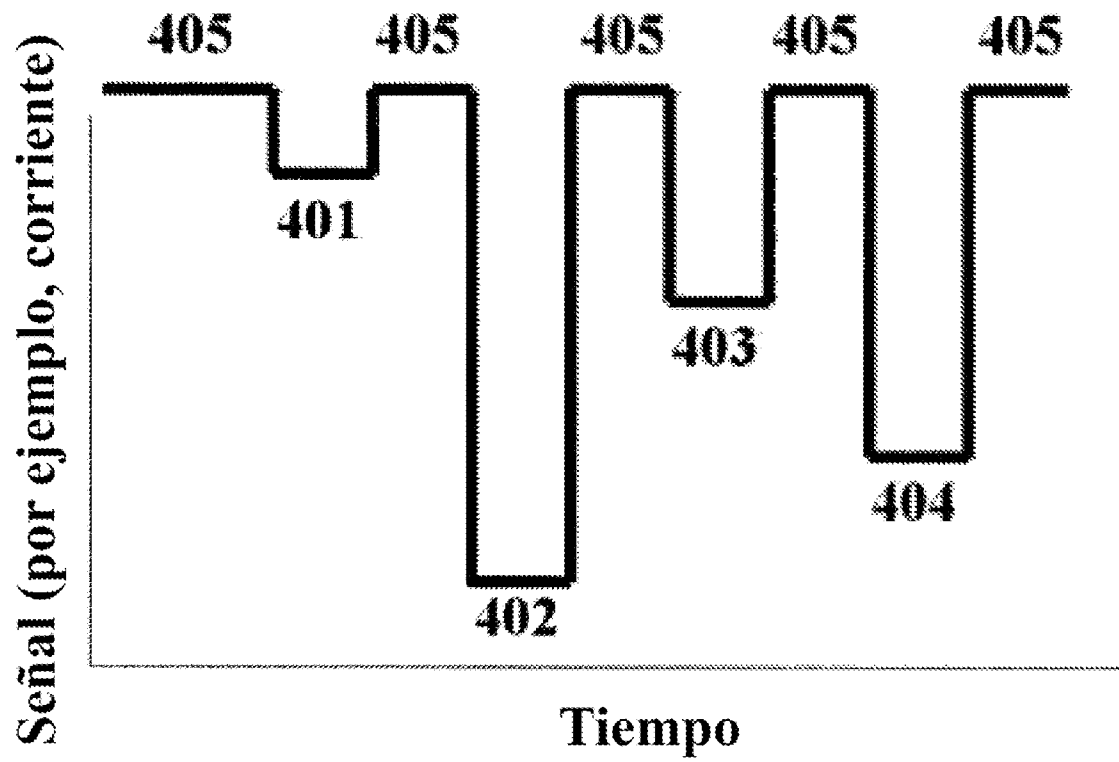


FIG. 4

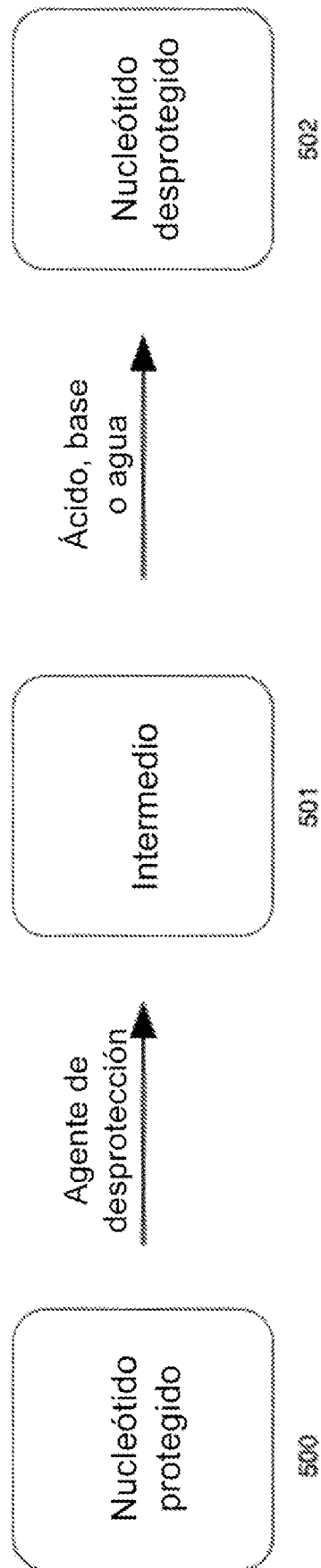


FIG. 5

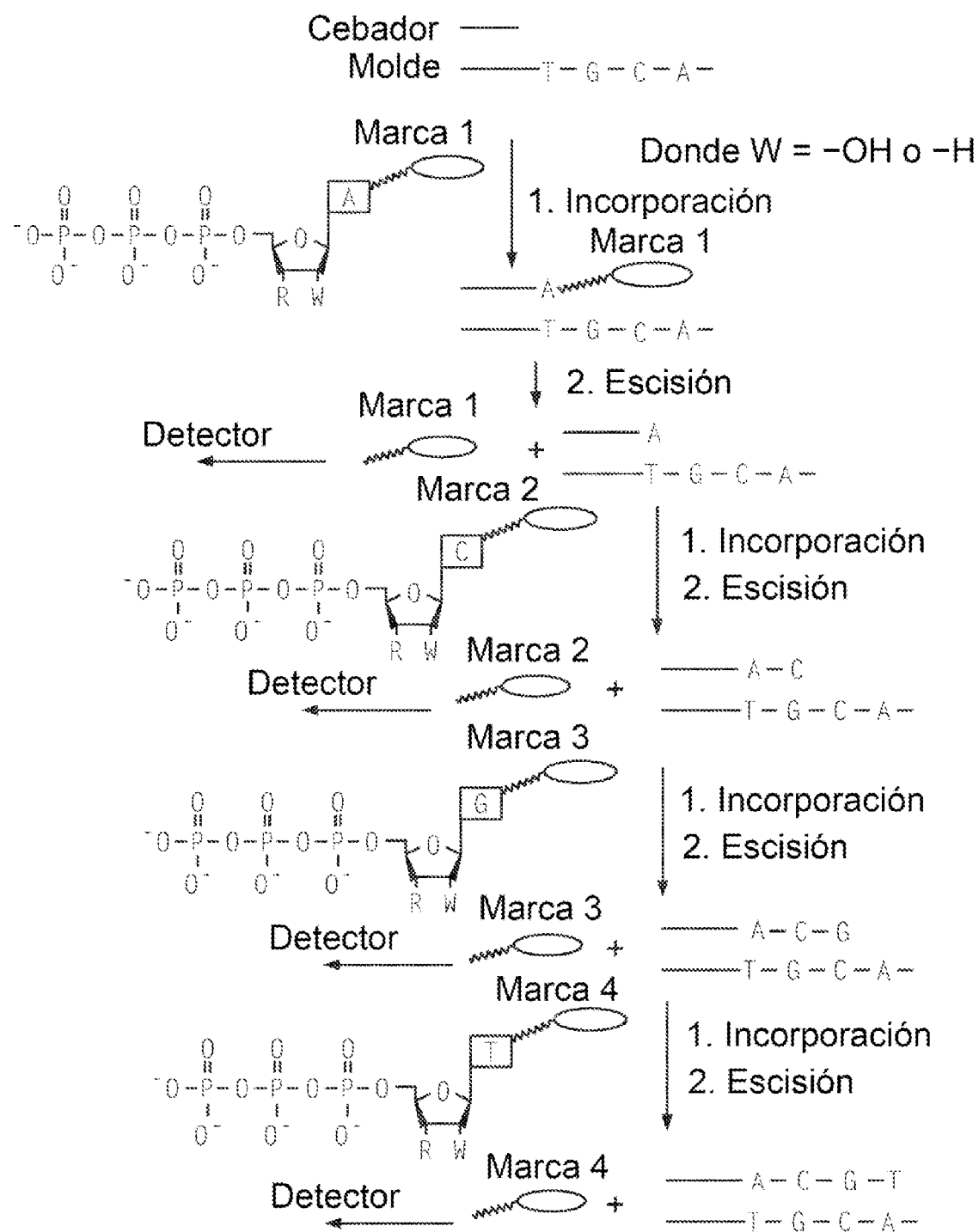


FIG. 6