

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 779 699**

51 Int. Cl.:

C12Q 1/68 (2008.01)

C12Q 1/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **08.04.2013 PCT/US2013/035630**

87 Fecha y número de publicación internacional: **27.12.2013 WO13191793**

96 Fecha de presentación y número de la solicitud europea: **08.04.2013 E 13807639 (3)**

97 Fecha y número de publicación de la concesión europea: **23.10.2019 EP 2864502**

54 Título: **Secuenciación de ácidos nucleicos mediante detección en nanoporos de moléculas de etiqueta**

30 Prioridad:

20.06.2012 US 201261662329 P

20.06.2012 US 201261662334 P

14.03.2013 US 201361781353 P

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

18.08.2020

73 Titular/es:

**THE TRUSTEES OF COLUMBIA UNIVERSITY IN
THE CITY OF NEW YORK (50.0%)**

West 116th Street and Broadway

New York, NY 10027, US y

ROCHE SEQUENCING SOLUTIONS, INC. (50.0%)

72 Inventor/es:

JU, JINGYUE;

DAVIS, RANDALL y

CHEN, ROGER

74 Agente/Representante:

ARIAS SANZ, Juan

ES 2 779 699 T3

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

DESCRIPCIÓN

Secuenciación de ácidos nucleicos mediante detección en nanoporos de moléculas de etiqueta

Antecedentes

5 La secuenciación de ácidos nucleicos es el procedimiento para determinar las bases de ácido nucleico de un ácido nucleico. Tal información de secuencia puede ser útil en el diagnóstico y/o tratamiento de un sujeto. Por ejemplo, la secuencia de ácido nucleico de un sujeto puede usarse para identificar, diagnosticar y posiblemente desarrollar tratamientos para enfermedades genéticas. Como otro ejemplo, la investigación de patógenos puede conducir al tratamiento para enfermedades contagiosas.

10 Hay métodos disponibles que pueden usarse para secuenciar un ácido nucleico. Sin embargo, tales métodos son caros y pueden no proporcionar información de secuencia dentro de un periodo de tiempo y con una precisión que pueden ser necesarios para diagnosticar y/o tratar a un sujeto.

Sumario

15 Los métodos de secuenciación de ácidos nucleicos que implican una molécula de ácido nucleico monocatenario que pasa a través de un nanoporo pueden tener una sensibilidad insuficiente. Las bases de ácido nucleico que comprende la molécula de ácido nucleico (por ejemplo, adenina (A), citosina (C), guanina (G), timina (T) y/o uracilo (U)) pueden no proporcionar una señal suficientemente diferenciada unas de otras. En particular, las purinas (es decir, A y G) tienen un tamaño, forma y carga similares entre sí y proporcionan una señal insuficientemente diferenciada en algunos casos. Además, las pirimidinas (es decir, C, T y U) tienen un tamaño, forma y carga similares entre sí y proporcionan una señal insuficientemente diferenciada en algunos casos. En el presente documento se reconoce la necesidad de métodos mejorados para la identificación de moléculas de ácido nucleico y la secuenciación de ácidos nucleicos.

20 Un aspecto de la presente divulgación proporciona un método para secuenciar una molécula de ácido nucleico, comprendiendo el método: (a) proporcionar un chip que comprende una pluralidad de nanoporos individualmente direccionables, en el que un nanoporo individualmente direccionable de dicha pluralidad de nanoporos individualmente direccionables comprende un nanoporo en una membrana que está dispuesta adyacente a un electrodo, en el que dicho nanoporo está unido a una ácido nucleico polimerasa, y en el que cada nanoporo individualmente direccionable está adaptado para detectar una etiqueta que se libera a partir de un nucleótido marcado con etiqueta tras la polimerización de dicho nucleótido marcado con etiqueta; (b) dirigir dicha molécula de ácido nucleico adyacente a o en proximidad de dicho nanoporo; (c) con la ayuda de dicha polimerasa, polimerizar nucleótidos a lo largo de dicha molécula de ácido nucleico para generar una cadena que es complementaria a al menos una porción de dicha molécula de ácido nucleico, en el que durante la polimerización se libera una etiqueta a partir de un nucleótido individual de dichos nucleótidos, y en el que dicha etiqueta liberada fluye a través o en proximidad de dicho nanoporo; y (d) detectar la etiqueta con la ayuda de dicho electrodo, en el que la etiqueta se detecta después de liberarse a partir de dicho nucleótido individual. En algunas realizaciones, dicha detección de (d) comprende además identificar dicha etiqueta. En algunos casos, el método comprende además correlacionar dicha etiqueta identificada con un tipo de dicho nucleótido individual. En algunos casos, el método comprende además generar, con la ayuda de un procesador informático, una secuencia de ácido nucleico de la molécula de ácido nucleico basándose en una evaluación de las etiquetas detectadas durante la polimerización.

35 Otro aspecto de la presente divulgación proporciona un método para la secuenciación de ácidos nucleicos, comprendiendo el método: (a) ligar una horquilla de ácido nucleico sobre un extremo de una molécula de ácido nucleico bicatenario; (b) disociar la molécula de ácido nucleico bicatenario y horquilla para formar un molde de ácido nucleico monocatenario; (c) extender un cebador hibridado al molde de ácido nucleico monocatenario usando nucleótidos marcados con etiqueta, en el que una etiqueta asociada con un nucleótido individual se libera tras la extensión; y (d) detectar la etiqueta liberada con la ayuda de un nanoporo, determinando así la secuencia de ácido nucleico de la molécula de ácido nucleico bicatenario. En algunos casos, el método comprende además dirigir la etiqueta liberada a partir del nucleótido individual a través del nanoporo. En algunos casos, el método comprende además dirigir la etiqueta liberada a partir del nucleótido individual hasta una ubicación adyacente al nanoporo.

40 Otro aspecto de la presente divulgación proporciona un método para la secuenciación de ácidos nucleicos, comprendiendo el método: (a) polimerizar nucleótidos marcados con etiqueta a una primera tasa, en el que una etiqueta asociada con un nucleótido individual se libera tras la polimerización; y (b) detectar la etiqueta liberada haciendo pasar la etiqueta a través de un nanoporo a una segunda tasa, en el que la segunda tasa es mayor que o igual a la primera tasa.

45 Otro aspecto de la presente divulgación proporciona un método para la secuenciación de ácidos nucleicos, comprendiendo el método: (a) polimerizar nucleótidos marcados con etiqueta, en el que una etiqueta asociada con un nucleótido individual se libera tras la polimerización; y (b) detectar la etiqueta liberada con la ayuda de un nanoporo. En algunos casos, el método comprende además dirigir la etiqueta liberada a partir del nucleótido individual a través del nanoporo. En algunos casos, el método comprende además dirigir la etiqueta liberada a partir del nucleótido individual hasta una ubicación adyacente al nanoporo.

Otro aspecto de la presente divulgación proporciona un método para la secuenciación de ácidos nucleicos, que comprende detectar, con la ayuda de un nanoporo, la incorporación de un nucleótido en una molécula de ácido nucleico, en el que la molécula de ácido nucleico no pasa a través del nanoporo.

5 Otro aspecto de la presente divulgación proporciona un método para la secuenciación de ácidos nucleicos, que comprende detectar un subproducto de un acontecimiento de incorporación de nucleótido individual con la ayuda de un nanoporo.

Un método para secuenciar una molécula de ácido nucleico, que comprende distinguir entre acontecimientos de incorporación de nucleótido individuales con una precisión de más de 4σ .

10 Otro aspecto de la presente divulgación proporciona un método para la secuenciación de ácidos nucleicos, comprendiendo el método: (a) proporcionar una matriz de nanoporos, en el que un nanoporo individual en dicha matriz está acoplado a una ácido nucleico polimerasa; y (b) polimerizar nucleótidos marcados con etiqueta con la polimerasa, en el que un nucleótido individual marcado con etiqueta comprende una etiqueta, y en el que la etiqueta se libera y se detecta con la ayuda del nanoporo.

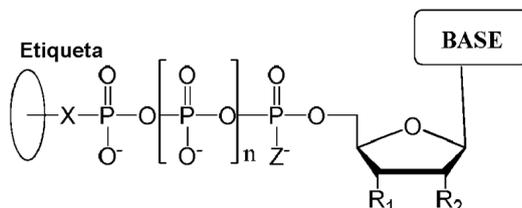
15 Otro aspecto de la presente divulgación proporciona un nucleótido marcado con etiqueta, en el que el nucleótido comprende una etiqueta que puede escindirarse en un acontecimiento de polimerización de nucleótido y detectarse con la ayuda de un nanoporo en un chip que comprende una matriz de nanoporos.

20 Otro aspecto de la presente divulgación proporciona un sistema para secuenciar una molécula de ácido nucleico, que comprende: (a) un chip que comprende una pluralidad de nanoporos individualmente direccionables, en el que un nanoporo individualmente direccionable de dicha pluralidad de nanoporos individualmente direccionables comprende al menos un nanoporo en una membrana dispuesta adyacente a un electrodo, en el que cada nanoporo individualmente direccionable está adaptado para ayudar en la detección de una etiqueta liberada a partir de un nucleótido marcado con etiqueta tras la incorporación de dicho nucleótido marcado con etiqueta en una cadena de ácido nucleico que es complementaria a dicha molécula de ácido nucleico; y (b) un procesador informático acoplado a dichos nanoporos individualmente direccionables, en el que dicho procesador informático está programado para ayudar en la caracterización de una secuencia de ácido nucleico de dicha molécula de ácido nucleico basándose en señales eléctricas recibidas a partir de dicha pluralidad de nanoporos individualmente direccionables, en el que una señal eléctrica individual está asociada con una etiqueta que se libera a partir de un nucleótido marcado con etiqueta después de la incorporación de dicho nucleótido marcado con etiqueta en una cadena de ácido nucleico que es complementaria a dicha molécula de ácido nucleico.

30 Otro aspecto de la presente divulgación proporciona un método para secuenciar una molécula de ácido nucleico, comprendiendo el método proporcionar una matriz de sitios individualmente direccionables a una densidad de al menos aproximadamente 500 sitios por mm^2 , teniendo cada sitio un nanoporo unido a una ácido nucleico polimerasa, y, en un sitio dado de la matriz, polimerizar nucleótidos marcados con etiqueta con una polimerasa, en el que tras la polimerización se libera una etiqueta y se detecta mediante un nanoporo en el sitio dado. En algunos casos, el método comprende además dirigir generar, con la ayuda de un procesador, una secuencia de ácido nucleico de la molécula de ácido nucleico basándose en las etiquetas detectadas.

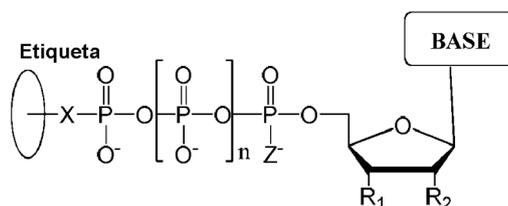
40 Otro aspecto de la presente divulgación proporciona un sistema de medición de conductancia que comprende: (a) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (b) unos medios para aplicar un campo eléctrico a través de la barrera; (c) unos medios para medir un cambio en el campo eléctrico; (d) al menos una polimerasa unida al poro; y (e) más de una enzima fosfatasa unida al poro.

Otro aspecto de la presente divulgación proporciona un compuesto que tiene la estructura:

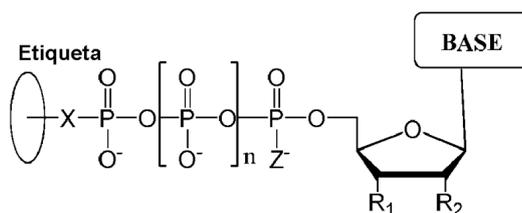


45 en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alquenoilo, un grupo alquínilo, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que la base es adenina, guanina, citosina, timina, uracilo o un derivado de una de estas bases, en la que n es 1, 2, 3 ó 4, y en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto.

Otro aspecto de la presente divulgación proporciona una composición que comprende cuatro tipos diferentes de un compuesto que tiene la estructura:

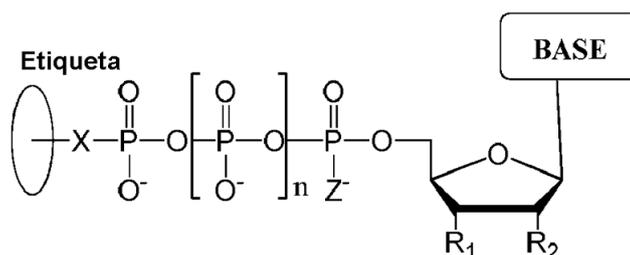


- 5 en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alqueno, un grupo alquilo, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que n es 1, 2, 3 ó 4, en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto, en la que la base de un primer tipo de compuesto es adenina o un derivado de la misma, la base de un segundo tipo de compuesto es guanina o un derivado de la misma, la base de un tercer tipo de compuesto es citosina o un derivado de la misma, y la base de un cuarto tipo de compuesto es timina o un derivado de la misma o uracilo o un derivado del mismo, y en la que la etiqueta en cada tipo de compuesto es diferente de la etiqueta en cada uno de los otros tres tipos de compuesto.
- 10
- 15 En algunos casos, la composición comprende además un quinto tipo de compuesto que difiere de cada uno de los cuatro tipos de compuesto en la base y en la etiqueta, en la que la base del quinto tipo de compuesto es uracilo o un derivado del mismo si la base del cuarto tipo de compuesto es timina o un derivado de la misma, o en la que la base del quinto tipo de compuesto es timina o un derivado de la misma si la base del cuarto tipo de compuesto es uracilo o un derivado del mismo.
- 20 Otro aspecto de la presente divulgación proporciona un método para determinar la identidad de un compuesto que comprende: (a) poner en contacto el compuesto con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; y (b)
- 25 registrar el cambio en el campo eléctrico cuando el compuesto se transloca a través del poro en el que el cambio en el campo eléctrico es el resultado de la interacción entre el compuesto, el electrolito y el poro, y es indicativo del tamaño, la carga y la composición del compuesto, permitiendo así la correlación entre el cambio y los valores predeterminados para determinar la identidad del compuesto. En algunos casos, el método comprende además una etapa de tratar el compuesto con una enzima fosfatasa antes de la etapa (a).
- 30 Otro aspecto de la presente divulgación proporciona un método para determinar si un compuesto es una etiqueta o un precursor de la etiqueta que comprende: (a) poner en contacto el compuesto con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; (b) registrar el cambio en el campo eléctrico cuando el compuesto se transloca a través del poro; y (c) comparar el cambio en el campo eléctrico con valores predeterminados correspondientes a la etiqueta y al precursor de la etiqueta, determinando así si el compuesto es la etiqueta o el precursor de la misma. En algunos casos, el método comprende además una etapa de ajustar la polarización de corriente del campo eléctrico en la etapa (a).
- 35
- 40 Otro aspecto de la presente divulgación proporciona un método para determinar la secuencia de nucleótidos de un ADN monocatenario, comprendiendo el método: (a) poner en contacto el ADN monocatenario con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; (iv) al menos una polimerasa unida al poro; y (v) más de una enzima fosfatasa unida al poro, y una composición que comprende cuatro tipos diferentes de un compuesto que tiene la estructura:
- 45



en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alqueniilo, un grupo alquinilo, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que n es 1, 2, 3 ó 4, en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto, en la que la base de un primer tipo de compuesto es adenina o un derivado de la misma, la base de un segundo tipo de compuesto es guanina o un derivado de la misma, la base de un tercer tipo de compuesto es citosina o un derivado de la misma, y la base de un cuarto tipo de compuesto es timina o un derivado de la misma, y en la que la etiqueta en cada tipo de compuesto es diferente de la etiqueta en cada uno de los otros tres tipos de compuesto, en el que el ADN monocatenario está en una disolución de electrolito en contacto con la polimerasa unida al poro y en el que el ADN monocatenario tiene un cebador hibridado a una porción del mismo, en condiciones que permiten que la polimerasa catalice la incorporación de uno de los compuestos en el cebador si el compuesto es complementario al residuo de nucleótido del ADN monocatenario inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ADN, en el que la incorporación del compuesto da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo, en el que la enzima fosfatasa unida al poro escinde la etiqueta a partir del polifosfato para liberar la etiqueta; (b) determinar qué compuesto se ha incorporado en el cebador para formar el producto de extensión de ADN en la etapa (a) aplicando un campo eléctrico a través de la barrera y midiendo un cambio electrónico a través del poro resultante de que la etiqueta generada en la etapa (a) se transloca a través del poro, en el que el cambio electrónico es diferente para cada tipo de etiqueta, identificando así el residuo de nucleótido en el ADN monocatenario complementario al compuesto incorporado; y (c) realizar de manera repetida la etapa (b) para cada residuo de nucleótido del ADN monocatenario que está secuenciándose, determinando así la secuencia de nucleótidos del ADN monocatenario.

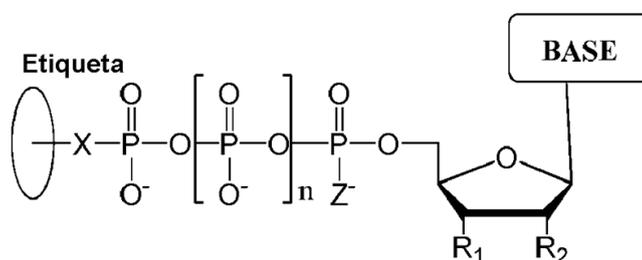
Otro aspecto de la presente divulgación proporciona un método para determinar la secuencia de nucleótidos de un ADN monocatenario, comprendiendo el método: (a) poner en contacto el ADN monocatenario con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; (iv) al menos una polimerasa unida al poro; y (v) más de una enzima fosfatasa unida al poro, y un compuesto que tiene la estructura:



en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alqueniilo, un grupo alquinilo, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que la base es adenina, guanina, citosina, timina o un derivado de una de esas bases, en la que n es 1, 2, 3 ó 4, y en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto, en el que el ADN monocatenario está en una disolución de electrolito en contacto con la polimerasa unida al poro y en el que el ADN monocatenario tiene un cebador hibridado a una porción del mismo, en condiciones que permiten que la polimerasa catalice la incorporación del compuesto en el cebador si es complementario al residuo de nucleótido del ADN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ADN, en el que si no se incorpora el compuesto, repetir de manera iterativa la puesta en contacto con diferentes compuestos hasta que se incorpora un compuesto, con la condición de que (1) el tipo de base en el compuesto es diferente del tipo de base en cada uno de los compuestos

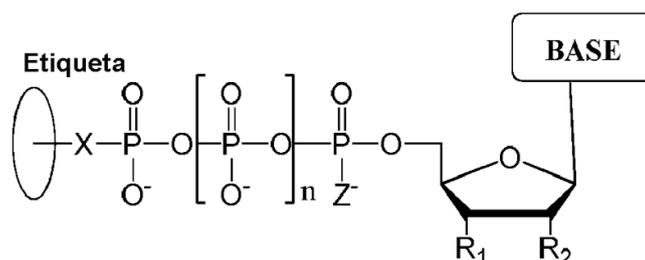
anteriores, y (2) el tipo de etiqueta en el compuesto es diferente del tipo de etiqueta en cada uno de los compuestos anteriores, en el que la incorporación del compuesto da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo, en el que la enzima fosfatasa unida al poro escinde la etiqueta a partir del polifosfato para liberar la etiqueta; (b) determinar qué compuesto se ha incorporado en el cebador para formar el producto de extensión de ADN en la etapa (a) aplicando un campo eléctrico a través de la barrera y midiendo un cambio electrónico a través del poro resultante de que la etiqueta generada en la etapa (a) se transloca a través del poro, en el que el cambio electrónico es diferente para cada tipo de etiqueta, identificando así el residuo de nucleótido en el ADN monocatenario complementario al compuesto incorporado; y (c) realizar de manera iterativa las etapas (a) y (b) para cada residuo de nucleótido del ADN monocatenario que está secuenciándose, determinando así la secuencia de nucleótidos del ADN monocatenario.

Otro aspecto de la presente divulgación proporciona un método para determinar la secuencia de nucleótidos de un ARN monocatenario, comprendiendo el método: (a) poner en contacto el ARN monocatenario con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; (iv) al menos una polimerasa unida al poro; y (v) más de una enzima fosfatasa unida al poro, y una composición que comprende cuatro tipos diferentes de un compuesto que tiene la estructura:



en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, 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 alquínico, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que n es 1, 2, 3 ó 4, en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto, en la que la base de un primer tipo de compuesto es adenina o un derivado de la misma, la base de un segundo tipo de compuesto es guanina o un derivado de la misma, la base de un tercer tipo de compuesto es citosina o un derivado de la misma, y la base de un cuarto tipo del compuesto es uracilo o un derivado del mismo, y en la que la etiqueta en cada tipo de compuesto es diferente de la etiqueta en cada uno de los otros tres tipos de compuesto, en el que el ARN monocatenario está en una disolución de electrolito en contacto con la polimerasa unida al poro y en el que el ARN monocatenario tiene un cebador hibridado a una porción del mismo, en condiciones que permiten que la polimerasa catalice la incorporación de uno de los compuestos en el cebador si el compuesto es complementario al residuo de nucleótido del ARN monocatenario inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ARN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ARN, en el que la incorporación del compuesto da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo, en el que la enzima fosfatasa unida al poro escinde la etiqueta a partir del polifosfato para liberar la etiqueta; (b) determinar qué compuesto se ha incorporado en el cebador para formar el producto de extensión de ARN en la etapa (a) aplicando un campo eléctrico a través de la barrera y midiendo un cambio electrónico a través del poro resultante de que la etiqueta generada en la etapa (a) se transloca a través del poro, en el que el cambio electrónico es diferente para cada tipo de etiqueta, identificando así el residuo de nucleótido en el ARN monocatenario complementario al compuesto incorporado; y (c) realizar de manera repetida la etapa (b) para cada residuo de nucleótido del ARN monocatenario que está secuenciándose, determinando así la secuencia de nucleótidos del ARN monocatenario.

Otro aspecto de la presente divulgación proporciona un método para determinar la secuencia de nucleótidos de un ARN monocatenario, comprendiendo el método: (a) poner en contacto el ARN monocatenario con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; (iv) al menos una polimerasa unida al poro; y (v) más de una enzima fosfatasa unida al poro, y un compuesto que tiene la estructura:



en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alquenoilo, un grupo alquínilo, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que la base es adenina, guanina, citosina, uracilo o un derivado de una de esas bases, en la que n es 1, 2, 3 ó 4, y en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto, en el que el ARN monocatenario está en una disolución de electrolito en contacto con la polimerasa unida al poro y en el que el ARN monocatenario tiene un cebador hibridado a una porción del mismo, en condiciones que permiten que la polimerasa catalice la incorporación del compuesto en el cebador si es complementario al residuo de nucleótido del ARN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ARN, en el que si no se incorpora el compuesto, repetir de manera iterativa la puesta en contacto con diferentes compuestos hasta que se incorpora un compuesto, con la condición de que (1) el tipo de base en el compuesto es diferente del tipo de base en cada uno de los compuestos anteriores, y (2) el tipo de etiqueta en el compuesto es diferente del tipo de etiqueta en cada uno de los compuestos anteriores, en el que la incorporación del compuesto da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo, en el que la enzima fosfatasa unida al poro escinde la etiqueta a partir del polifosfato para liberar la etiqueta; (b) determinar qué compuesto se ha incorporado en el cebador para formar el producto de extensión de ARN en la etapa (a) aplicando un campo eléctrico a través de la barrera y midiendo un cambio electrónico a través del poro resultante de que la etiqueta generada en la etapa (a) se transloca a través del poro, en el que el cambio electrónico es diferente para cada tipo de etiqueta, identificando así el residuo de nucleótido en el ARN monocatenario complementario al compuesto incorporado; y (c) realizar de manera iterativa las etapas (a) y (b) para cada residuo de nucleótido del ARN monocatenario que está secuenciándose, determinando así la secuencia de nucleótidos del ARN monocatenario.

Otro aspecto de la presente divulgación proporciona un sistema de medición de conductancia que comprende: (a) una barrera eléctricamente resistiva que separa al menos unas disoluciones de electrolito primera y segunda; (b) dicha barrera eléctricamente resistiva comprende al menos un poro con un diámetro a escala nanométrica; (c) al menos un compuesto con una etiqueta en al menos una de dichas disoluciones de electrolito primera y segunda; (d) estando dicho al menos un poro configurado para permitir que se impulse una corriente iónica a través de dichas disoluciones de electrolito primera y segunda mediante un potencial aplicado; (e) comprendiendo dicho al menos un poro una característica configurada para escindir la etiqueta a partir del compuesto para liberar la etiqueta; y (f) unos medios de medición de la corriente iónica y unos medios de registro de su transcurso en el tiempo como serie temporal, incluyendo periodos de tiempo en los que el al menos un poro no está obstruido por la etiqueta y también periodos de tiempo en los que la etiqueta provoca pulsos de conductancia reducida. En algunos casos, la etiqueta tiene un tiempo de residencia en el poro que es mayor que las limitaciones de ancho de banda de corriente iónica y el ruido de disparo de corriente de dichos medios de medición de la corriente iónica.

Otro aspecto de la presente divulgación proporciona un método para delinear segmentos de una serie temporal de conductancia para dar regiones estadísticamente compatibles con el nivel de conductancia de poro no obstruido, y pulsos de conductancia reducida, y también segmentos estadísticamente estacionarios dentro de pulsos de conductancia reducida individuales, generándose dicha serie temporal de conductancia con un sistema de medición de conductancia que comprende: una barrera eléctricamente resistiva que separa al menos unas disoluciones de electrolito primera y segunda; dicha barrera eléctricamente resistiva comprende al menos un poro con un diámetro a escala nanométrica; al menos un compuesto con una etiqueta en al menos una de dichas disoluciones de electrolito primera y segunda; estando dicho al menos un poro configurado para permitir que se impulse una corriente iónica a través de dichas disoluciones de electrolito primera y segunda mediante un potencial aplicado; comprendiendo dicho al menos un poro una característica configurada para escindir la etiqueta a partir del compuesto para liberar la etiqueta; y unos medios de medición de la corriente iónica y unos medios de registro de dicha serie temporal de conductancia, incluyendo periodos de tiempo en los que el al menos un poro no está obstruido por dicha etiqueta y también periodos de tiempo en los que dicha etiqueta provoca pulsos de conductancia reducida; seleccionándose dicho método para delinear segmentos de una serie temporal de conductancia del grupo que consiste en: (a) una decodificación de Viterbi de la secuencia de estados de probabilidad máxima de una densidad continua de un modelo de Markov oculto estimada a partir de la serie temporal de conductancia sin procesar; (b) una delineación de las regiones de pulsos de conductancia reducida mediante comparación con un umbral para la desviación con

respecto al nivel de conductancia de poro abierto; y (c) unos medios para caracterizar pulsos de conductancia reducida estimando las tendencias centrales de los niveles de corriente iónica para cada segmento, o mediante medición de tendencias centrales y duración de segmento en conjunto, seleccionándose la medición de tendencia central de segmento del grupo que consiste en: (i) un parámetro medio de una componente gaussiana de un primer GMM estimado a partir de la serie temporal de conductancia como parte de un modelo de Markov oculto de densidad continua; (ii) una media aritmética; (iii) una media recortada; (iv) una mediana; y (v) un estimador de máximo a posteriori de ubicación de muestra, o un estimador de probabilidad máxima de ubicación de muestra.

En algunos casos, el método comprende además al menos uno de: (a) una estimación de probabilidad máxima de un segundo modelo mixto gaussiano basándose en las medidas de tendencia central de segmentos de conductancia; (b) un hallazgo de picos por medio de interpolación y suavizado de la densidad de probabilidad empírica de las estimaciones de tendencias centrales de segmentos de la serie temporal de conductancia y hallazgo de raíces de las derivadas de las funciones de interpolación; y (c) otros medios de localización de los modos de estimador de distribución multimodal.

Otro aspecto de la presente divulgación proporciona un método para determinar al menos un parámetro de un compuesto en una disolución que comprende las etapas de: colocar un primer fluido en un primer depósito; colocar un segundo fluido en un segundo depósito; comprendiendo al menos uno de dichos fluidos primero y segundo al menos un compuesto, en el que el compuesto es un nucleótido marcado con etiqueta o una etiqueta escindida a partir de un nucleótido marcado con etiqueta; estando dicho primer fluido en dicho primer depósito separado de dicho segundo fluido en dicho segundo depósito con una barrera eléctricamente resistiva; comprendiendo dicha barrera eléctricamente resistiva al menos un poro; hacer pasar una corriente iónica a través de dicho primer fluido, dicho al menos un poro y dicho segundo fluido con un potencial eléctrico entre dichos fluidos primero y segundo; medir la corriente iónica que pasa a través de dicho al menos un poro y la duración de cambios en la corriente iónica; llevándose a cabo la medición de la corriente iónica durante un periodo de tiempo suficiente para medir una reducción en la corriente iónica provocada por el compuesto que interacciona con dicho al menos un poro; y determinar al menos un parámetro del compuesto analizando matemáticamente los cambios en la corriente iónica y la duración de los cambios en la corriente iónica a lo largo del periodo de tiempo; comprendiendo dicho análisis matemático al menos una etapa seleccionada del grupo que consiste en: (a) un parámetro medio de una componente gaussiana de un primer GMM estimado a partir de la serie temporal de conductancia como parte de un modelo de Markov oculto de densidad continua; (b) una extracción de media de acontecimientos; (c) asignación de estado de acontecimiento de probabilidad máxima; (d) detección y cálculo de promedio de umbral; (e) análisis de ventana deslizante; (f) una media aritmética; (g) una media recortada; (h) una mediana; e (i) un estimador de máximo a posteriori de ubicación de muestra, o un estimador de probabilidad máxima de ubicación de muestra.

Otro aspecto de la presente divulgación proporciona un nucleótido marcado con etiqueta, en el que el nucleótido comprende una etiqueta que puede escindirse en un acontecimiento de polimerización de nucleótido y detectarse con la ayuda de un nanoporo.

Un aspecto adicional de la presente divulgación proporciona un método para la secuenciación de ácidos nucleicos, comprendiendo el método proporcionar una matriz de sitios individualmente direccionables, teniendo cada sitio un nanoporo unido a una ácido nucleico polimerasa, y, en un sitio dado de dicha matriz, polimerizar nucleótidos marcados con etiqueta con una polimerasa, en el que se libera una etiqueta y se detecta mediante un nanoporo en dicho sitio dado.

Aspectos y ventajas adicionales de la presente divulgación resultarán fácilmente evidentes para los expertos en esta técnica a partir de la siguiente descripción detallada, en la que sólo se muestran y se describen realizaciones ilustrativas de la presente divulgación. Tal como se constatará, la presente divulgación puede presentar realizaciones adicionales y diferentes, y sus diversos detalles pueden presentar modificaciones en diversos aspectos evidentes, todo ello sin apartarse la divulgación. Por consiguiente, los dibujos y la descripción deben considerarse como de naturaleza ilustrativa y no limitativa.

Incorporación mediante referencia

Breve descripción de los dibujos

Las características novedosas de la invención se exponen con particularidad en las reivindicaciones adjuntas. Se obtendrá una mejor comprensión de las características y ventajas de la presente invención haciendo referencia a la siguiente descripción detallada que expone realizaciones ilustrativas, en las que se usan los principios de la invención, y los dibujos adjuntos de los que:

La figura 1 muestra esquemáticamente las etapas del método.

Las figuras 2A, 2B y 2C muestran ejemplos de detectores de nanoporos, en las que la figura 2A tiene el nanoporo dispuesto sobre el electrodo, la figura 2B tiene el nanoporo insertado en una membrana sobre un pocillo y la figura 2C tiene el nanoporo sobre un electrodo sobresaliente.

La figura 3 ilustra un método para la secuenciación de ácidos nucleicos.

- La figura 4 muestra un ejemplo de una señal generada mediante el paso de etiquetas a través de un nanoporo.
- La figura 5 muestra una matriz de detectores de nanoporos.
- La figura 6 muestra un ejemplo de una molécula de etiqueta unida al fosfato de un nucleótido.
- La figura 7 muestra ejemplos de ubicaciones de etiqueta alternativas.
- 5 La figura 8 muestra un ejemplo de nucleótidos marcados con etiqueta.
- La figura 9 muestra etiqueta detectable-polifosfato y etiqueta detectable.
- La figura 10 muestra un método para preparar una cadena de molde.
- La figura 11 muestra una configuración de chip a modo de ejemplo que comprende un nanoporo.
- La figura 12 muestra una configuración de matriz de célula de chip de prueba a modo de ejemplo.
- 10 La figura 13 muestra un conjunto de circuitos analógico de célula a modo de ejemplo.
- Las figuras 14A y 14B muestran ejemplos de circuitos ultracompactos.
- La figura 15 muestra un ejemplo de síntesis de análogos de nucleótido de cumarina-PEG-dG4P.
- La figura 16 muestra un ejemplo de caracterización de las etiquetas liberadas mediante MALDI-TOF-EM; etiquetas de cumarina-PEG-NH₂ generadas mediante hidrólisis ácida de cumarina-PEG16-dG4P produciendo cumarina-PEG16-NH₂ (azul), cumarina-PEG20-dG4P produciendo cumarina-PEG20-NH₂ (verde), cumarina-PEG24-dG4P produciendo cumarina-PEG24-NH₂ (naranja) y cumarina-PEG36-dG4P produciendo cumarina-PEG36-NH₂ (rojo), son idénticas a las etiquetas liberadas correspondientes generadas en reacciones de extensión de polimerasa tras tratamiento con fosfatasa alcalina, tal como se muestra mediante análisis de MALDI-TOF-EM; se muestra una imagen compuesta de cuatro espectros de EM obtenidos por separado; se muestran las estructuras de las etiquetas de cumarina-PEG-NH₂.
- 15 20
- La figura 17 muestra un ejemplo de distinción de etiquetas liberadas en nanoporos de proteína a nivel de detección de moléculas individuales; se combinaron cuatro compuestos de cumarina-PEG_n-NH₂ (n = 16, 20, 24 y 36), derivados a partir de los cuatro nucleótidos comparables mediante hidrólisis ácida, y se diluyeron en KCl 4 M, Tris 10 mM, pH 7,2 para medición de nanoporos; (izquierda) los datos de serie temporal indican que cuando estas etiquetas de PEG entran en un único canal iónico de α -hemolisina, provocan bloqueos de corriente que son característicos de su tamaño; a la derecha un histograma del bloqueo de corriente medio provocado por moléculas individuales muestra una resolución de línea base con un ancho de banda de medición de 10 kHz; las barras coloreadas en la parte superior representan la distribución de 6 σ de los datos (suponiendo distribuciones gaussianas para cada una de cuatro etiquetas de PEG que pueden representar cada uno de los cuatro nucleótidos de ADN), lo que sugiere que puede distinguirse una única base con una precisión mejor que 1 de cada 300.000 acontecimientos, representada en esta figura usando las designaciones A, C, G y T, que pueden producirse cuando se usan cuatro nucleótidos diferentes con cuatro PEG de longitud diferente para secuenciación de ADN.
- 25 30
- La figura 18 muestra un sistema informático configurado para controlar un secuenciador.
- La figura 19 muestra un histograma de lecturas de corriente de célula.
- 35 La figura 20 muestra una representación gráfica de corriente medida en picoamperios frente al tiempo medido en segundos para 4 etiquetas diferentes.
- La figura 21 muestra el autoensamblaje de proteína α -hemolisina en una bicapa lipídica para formar un canal iónico y un tramo de ácido nucleico que pasa a través de la misma (parte superior), con las firmas electrónicas correspondientes generadas (parte inferior) (Vercoutere *et al.* 2001 y Deamer *et al.* 2002).
- 40 La figura 22 muestra estructuras de nucleótidos desoxirribonucleótido de trifosfato de adenosina, desoxirribonucleótido de trifosfato de guanosina, desoxirribonucleótido de trifosfato de citosina y desoxirribonucleótido de trifosfato de timidina.
- La figura 23 muestra estructuras de cuatro 5'-polifosfatos de nucleósido marcados con etiqueta en fosfato.
- La figura 24 muestra la síntesis de 5'-trifosfatos de nucleósido marcados con etiqueta en fosfato.
- 45 La figura 25 muestra la síntesis de 5'-tetrafosfatos de nucleósido marcados con etiqueta en fosfato.
- La figura 26 muestra la síntesis de 5'-pentafosfatos de nucleósido marcados con etiqueta en el fosfato terminal.
- La figura 27 muestra a) unión de oligo-3' a 5'-fosfato, b) unión de oligo-5' a 5'-fosfato, c) resto detectable después de

reacción de polimerasa.

La figura 28(A) muestra la síntesis de 5'-trifosfatos de nucleósido de base modificada.

La figura 28(B) muestra la escisión de 5'-trifosfato de nucleósido de base modificada con TCEP.

5 La figura 29 muestra la síntesis de 5'-trifosfatos de nucleósido modificados en 3'-O; (A) dNTP unidos a 3'-O-2-nitrobencilo; (B) dNTP unidos a 3'-O-azidometilo; (C) resto detectable después de extensión de polimerasa y escisión por TCEP; y (D) resto detectable después de extensión de polimerasa y escisión por UV.

La figura 30 muestra una reacción de extensión de ADN usando análogos de nucleótido modificados con fosfato.

La figura 31 muestra una reacción de extensión de ADN usando análogos de nucleótido marcados con etiqueta en base.

10 La figura 32 muestra una reacción de extensión de ADN usando análogos de nucleótido marcados en 2' o 3'-OH.

La figura 33 muestra un esquema de secuenciación de ADN mediante nanoporo con nucleótidos modificados, particularmente aplicable a la secuenciación en tiempo real de moléculas individuales que implica la adición de los 4 nucleótidos y polimerasa al mismo tiempo para entrar en contacto con una única molécula de molde.

15 La figura 34 muestra fosfatos de nucleósido modificados en fosfato, base, 2' y 3' con posibles ligadores y etiquetas; BASE = adenina, guanina, timina, citosina, uracilo, 5-metil-C, 7-desaza-A, 7-desaza-G o derivados de los mismos; R₁ y R₂ = H, OH, F, NH₂, N₃ u OR'; n = 1-5; A = O, S, CH₂, CHF, CFF, NH; Z = O, S, BH₃; X = ligador que une fosfato o el 2'-O o 3'-O o la base al resto detectable y puede contener O, N o S, átomos de P (el ligador también puede ser un resto detectable, directa o indirectamente, tal como aminoácidos, péptidos, proteínas, hidratos de carbono, PEG de diferentes longitudes y pesos moleculares, colorantes orgánicos o inorgánicos, colorantes fluorescentes y fluorogénicos, fármacos, oligonucleótidos, etiquetas de masa, etiquetas quimioluminiscentes y pueden contener cargas positivas o negativas); Y = etiquetas o resto detectable, tal como compuestos alifáticos o aromáticos orgánicos con uno o más anillos, colorantes, proteínas, hidratos de carbono, PEG de diferentes longitudes y pesos moleculares, fármacos, oligonucleótidos, etiquetas de masa, etiquetas fluorescentes, etiquetas quimioluminiscentes y pueden contener carga positiva o negativa.

20

25 La figura 35 muestra estructuras de nucleótidos marcados con PEG-fosfato y ejemplos de PEG posibles con diferentes grupos reactivos para reaccionar con grupos funcionales.

La figura 36 muestra ejemplos específicos, no limitativos, de grupos reactivos en los fosfatos terminales, que también pueden unirse con cambios apropiados a un resto de base de nucleósido, y grupos con los que pueden reaccionar los grupos para formar etiquetas.

30 La figura 37 muestra un esquema de matriz de nanoporos para secuenciación de ADN en paralelo masiva mediante síntesis.

La figura 38 muestra la síntesis de nucleótidos marcados con PEG-fosfato.

35 La figura 39 muestra espectros de masa de MALDI-TOF de los productos de extensión de ADN generados mediante incorporación de análogos de nucleótido marcados con PEG-fosfato (dG4P-PEG); los productos individuales mostrados en los espectros indican que dG4P-PEG24 y dG4P-PEG37 se incorporan con una eficiencia de casi el 100%.

La figura 40 muestra las distribuciones de profundidad de bloqueo relativas para nanoporo de α -hemolisina en presencia de PEG que contienen 49, 37, 24 ó 16 monómeros de óxido de etileno a un potencial aplicado de +40 mV; las cuatro especies se identifican fácilmente.

40 La figura 41 muestra (A) separación y distribución de masa de unidades de polietilenglicol (PEG) mixtas a través de un único nanoporo; y (B) selección de 4 unidades de PEG diferenciadas con separación de línea base como etiquetas para las 4 bases, A, C, G y T; también se muestran las estructuras de PEG lineales y ramificados.

La figura 42 muestra la síntesis de PEG-trifosfatos cargados (la carga puede ajustarse basándose en los requisitos).

La figura 43 muestra la síntesis de 5'-trifosfatos de nucleósido marcados con etiqueta en fosfato.

45 La figura 44 muestra la síntesis de 5'-tetrafosfatos de nucleósido marcados con etiqueta en fosfato.

La figura 45 muestra la síntesis de 5'-pentafosfatos de nucleósido marcados con etiqueta en fosfato terminal.

La figura 46 muestra una plataforma de medición de nanoporos con CMOS integrado: (A) una microfotografía del chip preamplificador de CMOS de ocho canales con una imagen de un canal de amplificador con el electrodo de lado cis integrado; (B) diagrama que muestra la integración de dos chips con un nanoporo de estado sólido; (C)

diagrama que muestra la sección transversal del chip y cómo se graba el nanoporo directamente en el chip en la implementación de un chip; el empaquetamiento se produce con un pocillo independiente en el lado cis; y una imagen de TEM de un nanoporo de 3,5 nm de diámetro.

5 La figura 47 muestra prestaciones eléctricas de la electrónica de nanoporos con CMOS integrado (A) espectro de ruido de corriente de línea base referenciado a la entrada para $C_F = 0,15$ pF, filtro de Bessel de 4 polos de 1 MHz, $f_s = 4$ MS/s. También se muestra el cabezal abierto medido de un instrumento Axopatch 200B en modo de célula completa con $\beta = 1$, filtro de Bessel de 4 polos de 100 kHz, $f_s = 250$ kS/s (B) umbral mínimo de ruido del nuevo amplificador con un nanoporo unido en comparación con el mismo nanoporo medido mediante el instrumento Axopatch 200B.

10 La figura 48 muestra el anclaje de la polimerasa en las inmediaciones del nanoporo; un pocillo ayuda a restringir la difusión; L designa la distancia crítica desde la abertura de poro a la que los movimientos moleculares debidos a difusión y electroforesis son iguales.

La figura 49 muestra la síntesis de 5'-polifosfatos de nucleósido marcados con etiqueta.

La figura 50 muestra la síntesis de PEG-nucleótidos bloqueados en 3'-O.

15 La figura 51 muestra la secuenciación mediante síntesis con PEG-nucleótidos y detección en nanoporos (muchas copias de la misma molécula de ADN inmovilizadas sobre una perla y adición de un PEG-nucleótido cada vez); uso del mismo PEG unido a los cuatro nucleótidos; adición de un PEG-nucleótido cada vez, lecturas de al menos una base por ciclo si se incorpora el nucleótido correcto.

20 La figura 52 muestra la secuenciación mediante síntesis con PEG-nucleótidos bloqueados en 3'-O y detección en nanoporos (muchas copias de la misma molécula de ADN inmovilizada en una perla y adición de los cuatro PEG-nucleótidos bloqueados en 3'-O a la vez); adición de los cuatro nucleótidos unidos a PEG de tamaño diferente, bloqueados en 3' (dNTP-PEG bloqueados en 3') juntos; detección del nucleótido incorporado basándose en la señal de bloqueo de los PEG liberados; el grupo de bloqueo en 3' se retira mediante tratamiento con TECP y se continúa el ciclo para secuenciar correctamente el molde incluyendo regiones homopoliméricas.

25 La figura 53 muestra un esquema para una matriz de alta densidad en paralelo masiva de micropocillos para realizar el procedimiento de secuenciación. Cada pocillo puede contener un dispositivo de nanoporo y molde de ADN diferente.

La figura 54 muestra estructuras de cuatro moléculas de cumarina-PEG-dG4P.

30 La figura 55 muestra la caracterización de cuatro nucleótidos de cumarina-PEG_n-dG4P (n = 16, 20, 24, 36) mediante MALDI-TOF-EM; además de los picos de producto de longitud completa y picos de sales relacionadas, hay un segundo pico principal a la izquierda en cada espectro (cumarina-PEG_n-NH₂) que representa la escisión del enlace N-P entre el polifosfato y el ligador de aminoheptano debido a la naturaleza ácida de la matriz usada para el análisis de MALDI-TOF-EM.

35 La figura 56 muestra la estructura de molde-bucle-cebador usada para reacciones de SBS; la C en la "cadena de molde" permite la adición mediante polimerasa de dGMP a la "cadena de cebador" y la liberación de las etiquetas de cumarina-PEG-trifosfato a partir de los nucleótidos de cumarina-PEG-dG4P.

40 La figura 57 muestra dos métodos para generar etiquetas de cumarina-PEG-NH₂; el tratamiento directo de los análogos de nucleótido (cumarina-PEG_n-dG4P, n = 16, 20, 24, 36) (parte superior) con ácido acético al 10% produce las etiquetas de cumarina-PEG-NH₂ que son idénticas a las etiquetas liberadas por polimerasa (parte inferior derecha) tras el tratamiento con fosfatasa alcalina; después se caracterizan las etiquetas de cumarina-PEG-NH₂ (parte inferior izquierda) a nivel de moléculas individuales mediante firmas de bloqueo de corriente de nanoporos.

45 La figura 58 muestra la medición de MALDI-TOF-EM de los productos de extensión obtenidos con los cuatro nucleótidos de cumarina-PEG-dG4P; se usó un molde-bucle-cebador, en el que el molde contenía una C en la siguiente posición, junto con uno de los cuatro nucleótidos modificados con etiqueta de PEG para la reacción de polimerasa; en cada caso, se incorpora 2'-dGMP en el ADN, conduciendo a un producto de extensión de cebador de base individual.

Descripción detallada

50 Aunque en el presente documento se han mostrado y descrito diversas realizaciones de la invención, resultará evidente para los expertos en la técnica que tales realizaciones sólo se proporcionan a modo de ejemplo. A los expertos en la técnica se les pueden ocurrir numerosas variaciones, cambios y sustituciones sin apartarse de la invención. Debe entenderse que pueden emplearse diversas alternativas a las realizaciones de la invención descritas en el presente documento.

El término "nanoporo", tal como se usa en el presente documento, se refiere de manera general a un poro, canal o paso formado o proporcionado de otra manera en una membrana. Un nanoporo puede definirse mediante 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 por un material polimérico. El nanoporo puede estar dispuesto adyacente o en proximidad de un circuito de detección, tal como, por ejemplo, un circuito de semiconductor de óxido de metal complementario (CMOS) o 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. La alfa-hemolisina es un ejemplo de un nanoporo de proteína.
- El término “ácido nucleico”, tal como se usa en el presente documento, se refiere de manera general a una molécula que comprende una o más subunidades de ácido nucleico. Un ácido nucleico puede incluir una o más subunidades seleccionadas de adenosina (A), citosina (C), guanina (G), timina (T) y uracilo (U). 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.
- Los artículos “un”, “una” y “el/la” no son limitativos. Por ejemplo, “el método” incluye la definición más amplia del significado de la frase, que puede ser más de un método.
- Un “derivado” de adenina, guanina, citosina, timina o uracilo, incluye una 7-desaza-purina y una 5-metil-pirimidina. Otros ejemplos incluyen 7-desaza-adenina, 7-desaza-guanina y 5-metil-citosina.
- Tal como se usa en el presente documento, “alquilo” incluye grupos de hidrocarburo alifático saturado tanto ramificado como de cadena lineal que tienen el número especificado de átomos de carbono y pueden no estar sustituidos o estar sustituidos. Por tanto, se define que C1-Cn tal como en “alquilo C1-Cn” incluye grupos que tienen 1, 2... n-1, o n carbonos en una disposición lineal o ramificada. Por ejemplo, se define que un “alquilo C1-C5” incluye grupos que tienen 1, 2, 3, 4 ó 5 carbonos en una disposición lineal o ramificada, y específicamente incluye metilo, etilo, n-propilo, isopropilo, n-butilo, t-butilo y pentilo.
- Tal como se usa en el presente documento, “alqueno” se refiere a un radical de hidrocarburo no aromático, lineal o ramificado, que contiene al menos 1 doble enlace de carbono a carbono, y pueden estar presentes hasta el máximo número posible de dobles enlaces carbono-carbono no aromáticos, y puede no estar sustituido o estar sustituido. Por ejemplo, “alqueno C2-C5” significa un radical alqueno que tiene 2, 3, 4 ó 5 átomos de carbono, y hasta 1, 2, 3 ó 4 dobles enlaces carbono-carbono respectivamente. Los grupos alqueno incluyen etenilo, propenilo y butenilo.
- El término “alquino” se refiere a un radical de hidrocarburo lineal o ramificado, que contiene al menos 1 triple enlace de carbono a carbono, y pueden estar presentes hasta el máximo número posible de triples enlaces carbono-carbono no aromáticos, y puede no estar sustituido o estar sustituido. Por tanto, “alquino C2-C5” significa un radical alquino que tiene 2 ó 3 átomos de carbono y 1 triple enlace carbono-carbono, o que tiene 4 ó 5 átomos de carbono y hasta 2 triples enlaces carbono-carbono. Los grupos alquino incluyen etinilo, propinilo y butinilo.
- El término “sustituido” se refiere a un grupo funcional tal como se describió anteriormente tal como un alquilo, o un hidrocarbilo, en el que al menos un enlace con un átomo de hidrógeno contenido en el mismo se sustituye por un enlace con un átomo distinto de hidrógeno o distinto de carbono, siempre que se mantengan las valencias normales y que la(s) sustitución/sustituciones de(n) como resultado un compuesto estable. Los grupos sustituidos también incluyen grupos en los que uno o más enlaces con un átomo de carbono o hidrógeno se sustituyen por uno o más enlaces, incluyendo dobles o triples enlaces, con un heteroátomo. Los ejemplos no limitativos de sustituyentes incluyen los grupos funcionales descritos anteriormente, y por ejemplo, N, por ejemplo para formar -CN.
- Se entiende que un experto habitual en la técnica puede seleccionar sustituyentes y patrones de sustitución en los compuestos de la presente invención para proporcionar compuestos que son químicamente estables y que pueden sintetizarse fácilmente usando, por ejemplo, los métodos expuestos a continuación, a partir de materiales de partida fácilmente disponibles. Si un sustituyente está a su vez sustituido con más de un grupo, se entiende que esos múltiples grupos pueden estar en el mismo carbono o en carbonos diferentes, siempre que se obtenga como resultado una estructura estable.
- Al elegir los compuestos de la presente invención, un experto habitual en la técnica reconocerá que los diversos sustituyentes, es decir R₁, R₂, etc., deben elegirse de conformidad con los principios de la conectividad de estructuras químicas.
- En las estructuras de compuestos representadas en el presente documento, generalmente no se muestran los átomos de hidrógeno, excepto en azúcares de ribosa y desoxirribosa. Sin embargo, se entiende que existen suficientes átomos de hidrógeno en los átomos de carbono representados como para satisfacer la regla del octeto.
- Tal como se usan en el presente documento, y a menos que se mencione lo contrario, cada uno de los siguientes términos tendrá la definición expuesta: A - adenina; C - citosina; ADN - ácido desoxirribonucleico; G - guanina; ARN - ácido ribonucleico; T - timina; U - uracilo; dNPP - polifosfato de desoxirribonucleótido; y rNPP - polifosfato de ribonucleótido.
- Un ácido nucleico puede incluir cualquier molécula de ácido nucleico, incluyendo, sin limitación, ADN, ARN e híbridos o variantes de los mismos. Un ácido nucleico puede ser monocatenario o bicatenario. En una realización,

las bases de ácido nucleico que forman moléculas de ácido nucleico pueden ser las bases A, C, G, T y U, así como derivados de las mismas. Se muestran a modo de ejemplo derivados de estas bases en PCR Systems, Reagents and Consumables (Perkin Elmer Catalogue 1996-1997, Roche Molecular Systems, Inc., Branchburg, Nueva Jersey, EE.UU.).

5 Un polifosfato de nucleótido, tal como un polifosfato de desoxirribonucleótido (“dNPP”) o un polifosfato de ribonucleótido (“rNPP”), es un nucleótido que comprende múltiples, es decir tres, cuatro, cinco, seis o más, fosfatos unidos de una manera lineal a su átomo de carbono de azúcar 5'. Un análogo de polifosfato de nucleótido es un análogo de un polifosfato de desoxirribonucleótido de este tipo o de un polifosfato de ribonucleótido de este tipo tal como se define en el presente documento, que se diferencia de los mismos por tener una etiqueta unida al mismo en una posición especificada. Tales análogos pueden incorporarse en un cebador o una cadena de extensión de ácido nucleico, tal como una cadena de extensión de ADN, mediante contacto con una ácido nucleico polimerasa apropiada en las condiciones de polimerización de ácido nucleico apropiadas.

En una realización, el dNPP es un trifosfato de desoxinucleótido.

15 Tal como se usa en el presente documento, un tetranucleótido, un pentanucleótido o un hexanucleótido abarcan 4, 5 ó 6, respectivamente, residuos de monómero de ácido nucleico unidos mediante enlaces fosfodiéster, en los que el residuo terminal libre puede ser un nucleótido o un nucleósido. En una realización, el residuo terminal libre es un nucleósido y los demás residuos son nucleótidos.

20 “Sustrato sólido” significará cualquier medio adecuado presente en la fase sólida al que puede fijarse un ácido nucleico. Los ejemplos no limitativos incluyen chips, pocillos, perlas, estructuras de nanoporos y columnas. En una realización no limitativa el sustrato sólido puede estar presente en una disolución, incluyendo una disolución acuosa de electrolito.

25 “Hibridar” significará el apareamiento de un ácido nucleico monocatenario a otro ácido nucleico (tal como cebador) basándose en el principio bien entendido de complementariedad de secuencias. En una realización el otro ácido nucleico es un ácido nucleico monocatenario. La propensión a la hibridación entre ácidos nucleicos depende de la temperatura y la fuerza iónica del medio, la longitud de los ácidos nucleicos y el grado de complementariedad. El efecto de estos parámetros sobre la hibridación se describe, por ejemplo, en Sambrook J, Fritsch EF, Maniatis T., Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press, Nueva York (1989). Tal como se usa en el presente documento, hibridación de una secuencia de cebador, o de un producto de extensión de ADN, a otro ácido nucleico significará apareamiento suficiente de tal manera que el cebador, o producto de extensión de ADN, respectivamente, se extiende mediante la creación de un enlace fosfodiéster con un nucleótido o análogo de nucleótido disponible que puede formar un enlace fosfodiéster, con el mismo.

35 Tal como se usa en el presente documento, a menos que se especifique lo contrario, una base que es “diferente de” otra base o una lista mencionada de bases significará que la base tiene una estructura diferente de la otra base o bases. Por ejemplo, una base que es “diferente de” adenina, timina y citosina puede incluir una base que es guanina o una base que es uracilo.

40 “Cebador” tal como se usa en el presente documento (una secuencia de cebador) es un oligonucleótido corto, de manera habitual químicamente sintetizado, de longitud apropiada, por ejemplo de aproximadamente 18-24 bases, suficiente para hibridarse a un ADN diana (por ejemplo un ADN monocatenario) y permitir la adición de un residuo de nucleótido al mismo, o la síntesis de oligonucleótido o polinucleótido a partir del mismo, en condiciones adecuadas. En una realización el cebador es un cebador de ADN, es decir un cebador que consiste en, o que consiste en gran medida en, residuos de desoxirribonucleótido. Los cebadores se diseñan para tener una secuencia que es el complemento inverso de una región de ADN de molde/diana al que se hibrida el cebador. La adición de un residuo de nucleótido en el extremo 3' de un cebador mediante formación de un enlace fosfodiéster da como resultado un producto de extensión de ADN. La adición de un residuo de nucleótido en el extremo 3' del producto de extensión de ADN mediante formación de un enlace fosfodiéster da como resultado un producto de extensión de ADN adicional.

Métodos y sistemas para la identificación y secuenciación de ácido nucleico

50 En el presente documento se describen métodos, dispositivos y sistemas para secuenciar ácidos nucleicos usando un nanoporo. Los métodos pueden detectar con precisión acontecimientos de incorporación de nucleótido individuales, tal como tras la incorporación de un nucleótido en una cadena en crecimiento que es complementario a un molde. Una enzima (por ejemplo, ADN polimerasa) puede incorporar nucleótidos en una cadena de polinucleótido en crecimiento, en la que el nucleótido añadido es complementario a la cadena de ácido nucleico de molde correspondiente, que se hibrida a la cadena en crecimiento (por ejemplo, reacción en cadena de la polimerasa (PCR)). Estos acontecimientos de incorporación de nucleótido liberan etiquetas a partir de los nucleótidos, que pasan a través de un nanoporo y se detectan. De esta manera, puede identificarse la base incorporada (es decir, A, C, G, T o U) porque se libera una etiqueta única a partir de cada tipo de nucleótido (es decir, A, C, G, T o U).

Pueden detectarse acontecimientos de incorporación de nucleótido en tiempo real (es decir, a medida que se producen) y con la ayuda de un nanoporo. En algunos casos, una enzima (por ejemplo, ADN polimerasa) unida al, o

en proximidad del, nanoporo puede facilitar el flujo de una molécula de ácido nucleico a través de, o adyacente a, un nanoporo. Un acontecimiento de incorporación de nucleótido, o la incorporación de una pluralidad de nucleótidos, puede liberar una o más moléculas de etiqueta (también “etiquetas” en el presente documento), que pueden detectarse mediante un nanoporo a medida que las etiquetas fluyen a través del, o adyacentes al, nanoporo. En algunos casos, una enzima unida al, o en proximidad del, nanoporo puede ayudar en la detección de etiquetas u otros subproductos liberados tras la incorporación de uno o más nucleótidos.

Los métodos descritos en el presente documento pueden ser métodos de moléculas individuales. Es decir, la señal que se detecta se genera por una única molécula (es decir, incorporación de nucleótidos individuales) y no se genera a partir de una pluralidad de moléculas clonales. El método puede no requerir amplificación de ADN.

Pueden producirse acontecimientos de incorporación de nucleótido a partir de una mezcla que comprende una pluralidad de nucleótidos (por ejemplo, trifosfato de desoxirribonucleótido (dNTP en el que N es adenosina (A), citidina (C), timidina (T), guanosina (G) o uridina (U)). Los acontecimientos de incorporación de nucleótido no se producen necesariamente a partir de una disolución que comprende un único tipo de nucleótido (por ejemplo, dATP). Los acontecimientos de incorporación de nucleótido no se producen necesariamente a partir de disoluciones alternantes de una pluralidad de nucleótidos (por ejemplo, dATP, seguido por dCTP, seguido por dGTP, seguido por dTTP, seguido por dATP).

La secuenciación de ADN es una tecnología fundamental para la biología. Se han desarrollado varios métodos analíticos para detectar ADN o ARN a nivel de moléculas individuales usando tecnologías microscópicas químicas o físicas (Perkins *et al.* 1994, Rief *et al.* 1999, Smith *et al.* 1996, y Vercoutere *et al.* 2001).

En los últimos años, se han explorado tecnologías de detección de iones tales como canal iónico, que se basa en la detección de ion hidrógeno (H⁺) liberado cuando se incorpora un nucleótido en una cadena de ADN mediante una polimerasa (Rothberg *et al.* 2011), para detectar cadenas de ADN o ARN individuales (Kasianowicz 2003 y 2004, Chandler *et al.* 2004, Deamer *et al.* 2002, Berzakov *et al.* 2001, y Henrickson *et al.* 2000).

En algunos casos, puede usarse un canal de α -hemolisina, una exotoxina secretada por una bacteria, para detectar ácidos nucleicos al nivel de moléculas individuales (Kasianowicz *et al.* 1996). Una proteína α -hemolisina es un polipéptido monomérico que se autoensambla en una membrana de bicapa lipídica para formar un poro heptamérico, con un vestíbulo de 2,6 nm de diámetro y una abertura limitante de 1,5 nm de diámetro (el punto más estrecho del poro) (Meller *et al.* 2000, Akeson *et al.* 1999, y Deamer *et al.* 2002). La abertura limitante del nanoporo permite que moléculas de ácido nucleico (diámetro de ~2,0 nm) lineales monocatenarias, pero no bicatenarias, pasen a través. En una disolución acuosa de sal iónica tal como KCl, cuando se aplica una tensión apropiada a través de la membrana, el poro formado por un canal de α -hemolisina transporta una corriente iónica suficientemente intensa y estable. Los ácidos nucleicos polianiónicos se impulsan a través del poro mediante el campo eléctrico aplicado, bloqueando o reduciendo de ese modo la corriente iónica que de otro modo puede no estar impedida. Este procedimiento de paso genera una firma electrónica (figura 21) (Vercoutere *et al.* 2001 y Deamer *et al.* 2002). Una molécula de ácido nucleico particular, cuando entra y pasa a través del nanoporo, genera una firma característica que la distingue de otras moléculas de ácido nucleico. La duración del bloqueo es proporcional a la longitud de ácido nucleico, y la intensidad de señal está relacionada con las propiedades estéricas y electrónicas de los nucleótidos, concretamente la identidad de las cuatro bases (A, C, G y T). Por tanto, se obtiene un diagrama de acontecimiento específico, que es una representación gráfica del tiempo de translocación frente a la corriente de bloqueo, y se usa para distinguir la longitud y la composición de polinucleótidos mediante técnicas de registro de canales individuales basándose en parámetros característicos tales como corriente de translocación, duración de translocación y su dispersión correspondiente en el diagrama (Meller *et al.* 2000).

También se ha mostrado que un nanoporo de proteína con un adaptador unido covalentemente puede identificar con precisión 5'-monofosfatos de nucleósido no marcados (dAMP, dGMP, dCMP y dTMP) con alta precisión (Clarke *et al.* 2009). Por ejemplo, se ha unido covalentemente de manera satisfactoria adaptador de aminociclodextrina dentro del poro de α -hemolisina. Cuando se captura un dNMP y se impulsa a través del poro en una membrana de bicapa lipídica, la corriente iónica a través del poro se reduce a uno de cuatro niveles, que representan, cada uno de los cuatro dNMP (A, G, C o T). Además, Robertson *et al.* (2007) han demostrado recientemente que cuando una molécula de polietilenglicol (PEG) entra en un único poro de α -hemolisina, provoca estados de conductancia dependientes de la masa diferenciados con tiempos de residencia medios característicos. El espectro de masas basado en conductancia resuelve claramente las unidades de repetición de etilenglicol, y el tiempo de residencia aumenta con la masa del PEG.

Aunque el enfoque de nanoporos actual resulta prometedor como método de detección de ADN, aún no se ha alcanzado el objetivo más desafiante de secuenciación precisa entre bases.

Los métodos para secuenciar ácidos nucleicos pueden incluir recuperar una muestra biológica que tiene el ácido nucleico que va a secuenciarse, extraer o aislar de otro modo la muestra de ácido nucleico a partir de la muestra biológica, y en algunos casos preparar la muestra de ácido nucleico para la secuenciación.

La figura 1 ilustra esquemáticamente un método para secuenciar una muestra de ácido nucleico. El método

comprende aislar la molécula de ácido nucleico a partir de una muestra biológica (por ejemplo, muestra de tejido, muestra de líquido), y preparar la muestra de ácido nucleico para la secuenciación. En algunos casos, la muestra de ácido nucleico se extrae a partir de una célula. Algunas técnicas a modo de ejemplo para extraer ácidos nucleicos son usar lisozima, sonicación, extracción, altas presiones o cualquier combinación de las mismas. El ácido nucleico es ácido nucleico libre de células en algunos casos y no requiere extracción a partir de una célula.

En algunos casos, puede prepararse una muestra de ácido nucleico para la secuenciación mediante un procedimiento que implica retirar proteínas, residuos de pared celular y otros componentes a partir de la muestra de ácido nucleico. Hay muchos productos comerciales disponibles para lograr esto, tales como, por ejemplo, columnas de centrifugación. También puede usarse precipitación en etanol y centrifugación.

La muestra de ácido nucleico puede someterse a reparto (o fraccionamiento) para dar una pluralidad de fragmentos, lo que puede facilitar la secuenciación de ácidos nucleicos, tal como con la ayuda de un dispositivo que incluye una pluralidad de nanoporos en una matriz. Sin embargo, el fraccionamiento la(s) molécula(s) de ácido nucleico que va(n) a secuenciarse puede no ser necesario.

En algunos casos, se determinan secuencias largas (es decir, pueden no requerirse métodos de "secuenciación al azar"). Puede determinarse cualquier longitud adecuada de secuencia de ácido nucleico. Por ejemplo, pueden secuenciarse al menos aproximadamente 400, aproximadamente 500, aproximadamente 600, aproximadamente 700, aproximadamente 800, aproximadamente 800, aproximadamente 1000, aproximadamente 1500, aproximadamente 2000, aproximadamente 2500, aproximadamente 3000, aproximadamente 3500, aproximadamente 4000, aproximadamente 4500, aproximadamente 5000, aproximadamente 6000, aproximadamente 7000, aproximadamente 8000, aproximadamente 9000, aproximadamente 10000, aproximadamente 20000, aproximadamente 40000, aproximadamente 60000, aproximadamente 80000 o aproximadamente 100000 bases, y similares. En algunos casos, se secuencian al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 800, al menos 1000, al menos 1500, al menos 2000, al menos 2500, al menos 3000, al menos 3500, al menos 4000, al menos 4500, al menos 5000, al menos 6000, al menos 7000, al menos 8000, al menos 9000, al menos 10000, al menos 20000, al menos 40000, al menos 60000, al menos 80000, al menos 100000 bases, y similares. En algunos casos las bases secuenciadas son contiguas. En algunos casos, la muestra de ácido nucleico puede someterse a reparto antes de la secuenciación.

Secuenciación en nanoporos y detección molecular

En el presente documento se proporcionan sistemas y métodos para secuenciar una molécula de ácido nucleico con la ayuda de un nanoporo. El nanoporo puede formarse o incorporarse de otro modo en una membrana dispuesta adyacente a un electrodo de detección de un circuito de detección, tal como un circuito integrado. El circuito integrado puede ser un circuito integrado específico de aplicación (ASIC). En algunos ejemplos, el circuito integrado es un transistor de efecto de campo o un semiconductor de óxido de metal complementario (CMOS). El circuito de detección puede estar situado en un chip u otro dispositivo que tiene el nanoporo, o fuera del chip o dispositivo, tal como en una configuración fuera del 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).

En algunos casos, a medida que un ácido nucleico o etiqueta fluye a través del nanoporo, el circuito de detección detecta una señal eléctrica asociada con el ácido nucleico o la etiqueta. El ácido nucleico puede ser una subunidad de una cadena más grande. La etiqueta puede ser un subproducto de un acontecimiento de incorporación de nucleótido. Una señal detectada puede recopilarse y almacenarse en una ubicación de memoria, y usarse posteriormente para construir una secuencia del ácido nucleico. La señal recopilada puede procesarse para justificar cualquier anomalía en la señal detectada, tal como errores.

La figura 2 muestra un ejemplo de un detector (o sensor) de nanoporos que tiene control de temperatura, tal como puede prepararse según métodos descritos en la publicación de solicitud de patente estadounidense n.º 2011/0193570. Con referencia a la figura 2A, el detector de nanoporos comprende un electrodo superior 201 en contacto con una disolución conductora (por ejemplo, disolución de sal) 207. Un electrodo conductor inferior 202 está cerca de, adyacente a, o en proximidad de, un nanoporo 206, que está insertado en una membrana 205. En algunos casos, el electrodo conductor inferior 202 está incorporado en un semiconductor 203 en el que está incorporado un conjunto de circuitos eléctrico en un sustrato de semiconductor 204. Una superficie del semiconductor 203 puede tratarse para que sea hidrófoba. Una muestra que está detectándose pasa a través del poro en el nanoporo 206. El sensor de chip de semiconductor se coloca en un empaquetamiento 208 y éste, a su vez, está en las inmediaciones de un elemento de control de temperatura 209. El elemento de control de temperatura 209 puede ser un dispositivo de calentamiento y/o enfriamiento termoelectrónico (por ejemplo, dispositivo de Peltier). Múltiples detectores de nanoporos pueden formar una matriz de nanoporos.

Con referencia a la figura 2B, en la que números iguales representan elementos iguales, la membrana 205 puede estar dispuesta sobre un pocillo 210, en el que el sensor 202 forma parte de la superficie del pocillo. La figura 2C muestra un ejemplo en el que el electrodo 202 sobresale a partir de la superficie de semiconductor tratada 203.

En algunos ejemplos, la membrana 205 se forma en el electrodo conductor inferior 202 y no en el semiconductor

203. En tal caso, la membrana 205 puede formar interacciones de acoplamiento con el electrodo conductor inferior 202. Sin embargo, en algunos casos, la membrana 205 se forma en el electrodo conductor inferior 202 y el semiconductor 203. Como alternativa, la membrana 205 puede formarse en el semiconductor 203 y no en el electrodo conductor inferior 202, pero puede extenderse sobre el electrodo conductor inferior 202.

5 Secuenciación indirecta con nanoporos

Pueden usarse nanoporos para secuenciar moléculas de ácido nucleico de manera indirecta, opcionalmente con detección eléctrica. La secuenciación indirecta puede ser cualquier método en el que una molécula de ácido nucleico polimerizada tal como ADN o ARN no pasa a través del nanoporo. La molécula de ácido nucleico puede ubicarse al menos parcialmente en el vestíbulo del nanoporo, pero no en el poro (es decir, porción más estrecha) del nanoporo. La molécula de ácido nucleico puede pasar dentro de cualquier distancia adecuada desde el y/o en proximidad del nanoporo, opcionalmente dentro de una distancia tal que etiquetas liberadas a partir de acontecimientos de incorporación de nucleótido se detectan en el nanoporo.

Pueden detectarse subproductos de acontecimientos de incorporación de nucleótido mediante el nanoporo. "acontecimientos de incorporación de nucleótido" son la incorporación de un nucleótido en una cadena de polinucleótido en crecimiento. Puede correlacionarse un subproducto con la incorporación de un nucleótido de tipo dado. Los acontecimientos de incorporación de nucleótido se catalizan generalmente mediante una enzima, tal como ADN polimerasa, y usan interacciones de pares de bases con una molécula de molde para elegir entre los nucleótidos disponibles para la incorporación en cada ubicación.

En algunos casos, el subproducto pasa a través del nanoporo y/o genera una señal detectable en el nanoporo. Las moléculas de etiqueta liberadas son un ejemplo de subproductos. En algunos casos, los subproductos son protones (es decir, un cambio de pH). En otros casos, los subproductos son fosfatos (por ejemplo, fosfatos liberados durante acontecimientos de incorporación de nucleótido). Por ejemplo, cada uno de los diferentes tipos de nucleótidos puede comprender un número diferente de fosfatos, y la detección de los fosfatos liberados permite determinar la identidad del nucleótido incorporado.

En la figura 3 se representa un ejemplo del método. En este caso, la cadena de ácido nucleico 300 pasa a través o en proximidad (pero no a través tal como se indica mediante la flecha en 301) del nanoporo 302. Una enzima 303 (por ejemplo, ADN polimerasa) extiende una cadena de ácido nucleico en crecimiento 304 mediante incorporación de un nucleótido cada vez usando una primera molécula de ácido nucleico como molde 300 (es decir, la enzima cataliza acontecimientos de incorporación de nucleótido).

La enzima 303 puede estar unida al nanoporo 302. Los métodos adecuados para unir la enzima al nanoporo incluyen reticulación tal como la formación de enlaces disulfuro intramoleculares. El nanoporo y la enzima también pueden ser una proteína de fusión, que se codifica por una cadena de polipéptido individual. Los métodos para producir proteínas de fusión pueden incluir fusionar la secuencia codificante para la enzima en marco y adyacente a la secuencia codificante para el nanoporo (sin un codón de terminación entremedias) y expresar esta secuencia de fusión a partir de un único promotor. En algunos casos, también se unen enzimas fosfatasa al nanoporo.

En algunos casos, la ADN polimerasa es polimerasa 9°N o una variante de la misma, ADN polimerasa I de *E. Coli*, ADN polimerasa de bacteriófago T4, Sequenase, ADN polimerasa Taq, polimerasa 9°N (exo-)A485L/Y409V o ADN polimerasa de Phi29 (ADN polimerasa de ϕ 29).

Secuenciación en nanoporos de moléculas de etiqueta

Una muestra de ácido nucleico puede secuenciarse usando nucleótidos o análogos de nucleótido marcados con etiqueta. En algunos ejemplos, un método para secuenciar una molécula de ácido nucleico comprende (a) polimerizar nucleótidos marcados con etiqueta, en el que una etiqueta asociada con un nucleótido individual se libera tras la polimerización, y (b) detectar la etiqueta liberada con la ayuda de un nanoporo.

En algunos casos, el método comprende además dirigir la etiqueta liberada a partir de un nucleótido individual a través del nanoporo. La etiqueta liberada puede dirigirse mediante cualquier técnica adecuada, en algunos casos con la ayuda de una enzima (o motor molecular). Alternativamente, la etiqueta liberada puede dirigirse a través del nanoporo sin el uso de una enzima. Por ejemplo, la etiqueta puede dirigirse mediante una diferencia de tensión a través del nanoporo tal como se describe en el presente documento.

Siguiendo haciendo referencia a la figura 3, la enzima extrae a partir de una combinación de nucleótidos (círculos negros en la indicación 305) unidos a moléculas de etiqueta (círculos blancos en la indicación 305). Cada tipo de nucleótido está unido a una molécula de etiqueta diferente de modo que cuando se liberan las etiquetas y pasan a través del nanoporo 306, pueden distinguirse unas de otras basándose en la señal que se genera en el nanoporo.

La figura 4 muestra un ejemplo de diferentes señales que se generan por diferentes etiquetas a medida que se detectan mediante el nanoporo. Se detectan cuatro intensidades de señal diferentes (401, 402, 403 y 404). Estas pueden corresponder a cuatro etiquetas diferentes. Por ejemplo, la etiqueta presentada al nanoporo y/o liberada mediante incorporación de adenosina (A) puede generar una señal con una amplitud 401. Una etiqueta presentada

5 al nanoporo y/o liberada mediante incorporación de citosina (C) puede generar una señal con una amplitud superior 403; una etiqueta presentada al nanoporo y/o liberada mediante incorporación de guanina (G) puede generar una señal con una amplitud aún superior 404; y una etiqueta presentada al nanoporo y/o liberada mediante incorporación de timina (T) puede generar una señal con una amplitud todavía superior 402. La señal puede volver a un nivel de línea base 405 entre detecciones en algunos casos.

10 La tasa de acontecimientos de incorporación de nucleótido es generalmente más lenta que (o igual a) la tasa a la que moléculas de etiquetas liberadas durante los acontecimientos de incorporación de nucleótido pasan a través del, y/o se detectan mediante el, nanoporo. Generalmente, la tasa de acontecimientos de incorporación de nucleótido no es mayor que la tasa a la que moléculas de etiquetas liberadas durante los acontecimientos de incorporación de nucleótido pasan a través del, y/o se detectan mediante el, nanoporo (es decir, de lo contrario los acontecimientos de incorporación de nucleótido no se detectan con precisión y/o en la secuencia correcta).

Matrices de nanoporos para secuenciación

15 La figura 5 muestra que puede secuenciarse una pluralidad de moléculas de ácido nucleico en una matriz de detectores de nanoporos. En este caso, cada ubicación de nanoporo (por ejemplo, 501) comprende un nanoporo, opcionalmente unido a una enzima polimerasa y/o enzimas fosfatasa. Generalmente también hay un sensor en cada ubicación de matriz tal como se describe en otra parte en el presente documento.

20 En algunos ejemplos, se proporciona una matriz de nanoporos unidos a una ácido nucleico polimerasa, y se polimerizan nucleótidos marcados con etiqueta con la polimerasa. Durante la polimerización, se libera una etiqueta y se detecta mediante el nanoporo. 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 nanoporos, y similares. En algunos casos, la matriz comprende al menos 200, al menos 400, al menos 600, al menos 800, al menos 1000, al menos 1500, al menos 2000, al menos 3000, al menos 4000, al menos 5000, al menos 10000, al menos 15000, al menos 20000, al menos 40000, al menos 60000, al menos 80000, al menos 100000, al menos 200000, al menos 400000, al menos 600000, al menos 800000, al menos 1000000 nanoporos, y similares. Los nanoporos pueden ser individualmente direccionables. En algunos casos, la matriz puede incluir nanoporos individualmente direccionables a una densidad de al menos aproximadamente 500, 600, 700, 800, 900, 1000, 10.000, 100.000 ó 1.000.000 de nanoporos individualmente direccionables por mm².

35 En algunos casos, se libera una única etiqueta tras la incorporación de un único nucleótido y se detecta mediante un nanoporo. En otros casos, se libera una pluralidad de etiquetas tras la incorporación de una pluralidad de nucleótidos. Un sensor de nanoporos adyacente a un nanoporo puede detectar una etiqueta liberada individual, o una pluralidad de etiquetas liberadas. Una o más señales asociadas con una pluralidad de etiquetas liberadas pueden detectarse y procesarse para producir una señal promediada.

40 Pueden detectarse etiquetas mediante el sensor en función del tiempo. Pueden usarse etiquetas detectadas con el tiempo para determinar la secuencia de ácido nucleico de la muestra de ácido nucleico, tal como con la ayuda de un sistema informático (véase, por ejemplo, la figura 18) que está programado para registrar datos de sensor y generar información de secuencia a partir de los datos.

Precisión de secuenciación

45 Los métodos proporcionados en el presente documento pueden distinguir con precisión entre acontecimientos de incorporación de nucleótido individuales (por ejemplo, acontecimientos de moléculas individuales). Los métodos pueden distinguir con precisión entre acontecimientos de incorporación de nucleótido individuales en un único pase, es decir, sin tener que volver a secuenciar una molécula de ácido nucleico dada.

50 Un método para la secuenciación de ácidos nucleicos comprende distinguir entre acontecimientos de incorporación de nucleótido individuales con una precisión de más de aproximadamente 4σ . En algunos casos, los acontecimientos de incorporación de nucleótido se detectan con la ayuda de un nanoporo. Pueden liberarse etiquetas asociadas con los nucleótidos tras la incorporación y las etiquetas pasan a través del nanoporo. Una etiqueta diferente puede estar asociada con, y/o liberarse a partir de, cada tipo de nucleótido (por ejemplo, A, C, T, G) y se detecta a medida que pasa a través del nanoporo. Los errores incluyen, pero no se limitan a, (a) no lograr detectar una etiqueta, (b) identificar erróneamente una etiqueta, (c) detectar una etiqueta cuando no hay ninguna etiqueta, (d) detectar etiquetas en el orden incorrecto (por ejemplo, dos etiquetas se liberan en un primer orden, pero se adelantan entre sí y se detectan en un segundo orden), (e) se detecta como si se hubiera liberado una etiqueta que no se ha liberado a partir de un nucleótido, o cualquier combinación de los mismos. En algunas realizaciones, la precisión de distinción entre acontecimientos de incorporación de nucleótido individuales es del 100% menos la tasa a la que se producen errores (es decir, tasa de error).

La precisión de distinción entre acontecimientos de incorporación de nucleótido individuales es cualquier porcentaje adecuado. La precisión de distinción entre acontecimientos de incorporación de nucleótido individuales puede ser de aproximadamente el 93%, aproximadamente el 94%, aproximadamente el 95%, aproximadamente el 96%, aproximadamente el 97%, aproximadamente el 98%, aproximadamente el 99%, aproximadamente el 99,5%, aproximadamente el 99,9%, aproximadamente el 99,99%, aproximadamente el 99,999%, aproximadamente el 99,9999%, y similares. En algunos casos, la precisión de distinción entre acontecimientos de incorporación de nucleótido individuales es de al menos el 93%, al menos el 94%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99%, al menos el 99,5%, al menos el 99,9%, al menos el 99,99%, al menos el 99,999%, al menos el 99,9999%, y similares. En algunos casos, la precisión de distinción entre acontecimientos de incorporación de nucleótido individuales se notifica en unidades de sigma (σ). Sigma es una variable estadística que se usa algunas veces en gestión de empresas y estrategia de fabricación para notificar tasas de error tales como el porcentaje de productos libres de defectos. En este caso, los valores de sigma pueden usarse de manera intercambiable con la precisión según la siguiente relación: 4σ es una precisión del 99,38%, 5σ es una precisión del 99,977%, y 6σ es una precisión del 99,99966%.

La distinción entre acontecimientos de incorporación de nucleótido individuales, según métodos descritos en el presente documento, puede usarse para determinar con precisión una secuencia de ácido nucleico. En algunos casos, la determinación de la secuencia de ácido nucleico de un ácido nucleico (por ejemplo, ADN y ARN) incluye errores. Los errores a modo de ejemplo incluyen, pero no se limitan a, deleciones (no lograr detectar un ácido nucleico), inserciones (detectar un ácido nucleico donde no hay ninguno realmente presente) y sustituciones (detectar el ácido nucleico incorrecto). La precisión de secuenciación de ácidos nucleicos puede determinarse alineando la secuencia de ácido nucleico medida con la secuencia de ácido nucleico auténtica (por ejemplo, según técnicas bioinformáticas) y determinando el porcentaje de posiciones de ácido nucleico que son deleciones, inserciones y/o sustituciones. Los errores son cualquier combinación de deleciones, inserciones y sustituciones. La precisión oscila entre el 0% y el 100%, siendo el 100% una determinación completamente correcta de la secuencia del ácido nucleico. De manera similar, la tasa de error es del 100% - la precisión y oscila entre el 0% y el 100%, siendo una tasa de error del 0% una determinación completamente correcta de la secuencia del ácido nucleico.

La precisión de secuenciación de ácidos nucleicos tal como se realiza según los métodos y/o usando los dispositivos descritos en el presente documento es alta. La precisión es cualquier valor adecuadamente alto. En algunos casos, la precisión es de aproximadamente el 95%, aproximadamente el 95,5%, aproximadamente el 96%, aproximadamente el 96,5%, aproximadamente el 97%, aproximadamente el 97,5%, aproximadamente el 98%, aproximadamente el 98,5%, aproximadamente el 99%, aproximadamente el 99,5%, aproximadamente el 99,9%, aproximadamente el 99,99%, aproximadamente el 99,999%, aproximadamente el 99,9999%, y similares. En algunos casos, la precisión es de al menos el 95%, al menos el 95,5%, al menos el 96%, al menos el 96,5%, al menos el 97%, al menos el 97,5%, al menos el 98%, al menos el 98,5%, al menos el 99%, al menos el 99,5%, al menos el 99,9%, al menos el 99,99%, al menos el 99,999%, al menos el 99,9999%, y similares. En algunos casos, la precisión es de entre aproximadamente el 95% y el 99,9999%, entre aproximadamente el 97% y el 99,9999%, entre aproximadamente el 99% y el 99,9999%, entre aproximadamente el 99,5% y el 99,9999%, entre aproximadamente el 99,9% y el 99,9999%, y similares.

Puede lograrse una alta precisión realizando múltiples pases (es decir, secuenciando una molécula de ácido nucleico una pluralidad de veces, por ejemplo, haciendo pasar el ácido nucleico a través o en proximidad de un nanoporo y secuenciando bases de ácido nucleico de la molécula de ácido nucleico). Los datos a partir de múltiples pases pueden combinarse (por ejemplo, deleciones, inserciones y/o sustituciones en un primer pase se corrigen usando datos de otros pases repetidos). El método proporciona una alta precisión con pocos pases (también denominados lecturas, multiplicidad de cobertura de secuenciación). El número de pases es cualquier número adecuado, y no se necesita que sea un número entero. En algunas realizaciones, la molécula de ácido nucleico se secuencía 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces, 10 veces, 12 veces, 14 veces, 16 veces, 18 veces, 20 veces, 25 veces, 30 veces, 35 veces, 40 veces, 45 veces, 50 veces, y similares. En algunas realizaciones, la molécula de ácido nucleico se secuencía como máximo 1 vez, como máximo 2 veces, como máximo 3 veces, como máximo 4 veces, como máximo 5 veces, como máximo 6 veces, como máximo 7 veces, como máximo 8 veces, como máximo 9 veces, como máximo 10 veces, como máximo 12 veces, como máximo 14 veces, como máximo 16 veces, como máximo 18 veces, como máximo 20 veces, como máximo 25 veces, como máximo 30 veces, como máximo 35 veces, como máximo 40 veces, como máximo 45 veces, como máximo 50 veces, y similares. En algunas realizaciones, la molécula de ácido nucleico se secuencía entre aproximadamente 1 vez y 10 veces, entre aproximadamente 1 vez y 5 veces, entre aproximadamente 1 vez y 3 veces, y similares. El nivel de precisión puede lograrse combinando datos recopilados a partir de como máximo 20 pases. En algunas realizaciones, el nivel de precisión se logra combinando datos recopilados a partir de como máximo 10 pases. En algunas realizaciones, el nivel de precisión se logra combinando datos recopilados a partir de como máximo 5 pases. En algunos casos, el nivel de precisión se logra en un único pase.

La tasa de error es cualquier tasa adecuadamente baja. En algunos casos, la tasa de error es de aproximadamente el 10%, aproximadamente el 5%, aproximadamente el 4%, aproximadamente el 3%, aproximadamente el 2%, aproximadamente el 1%, aproximadamente el 0,5%, aproximadamente el 0,1%, aproximadamente el 0,01%, aproximadamente el 0,001%, aproximadamente el 0,0001%, y similares. En algunos casos, la tasa de error es de

como máximo el 10%, como máximo el 5%, como máximo el 4%, como máximo el 3%, como máximo el 2%, como máximo el 1%, como máximo el 0,5%, como máximo el 0,1%, como máximo el 0,01%, como máximo el 0,001%, como máximo el 0,0001%, y similares. En algunos casos, la tasa de error es de entre el 10% y el 0,0001%, entre el 3% y el 0,0001%, entre el 1% y el 0,0001%, entre el 0,01% y el 0,0001%, y similares.

5 Preparación de molde

El método puede implicar secuenciar una cadena de ácido nucleico de molde añadiendo nucleótidos marcados con etiqueta a una cadena complementaria a la cadena de molde y detectando moléculas de etiqueta liberadas en un nanoporo.

10 La figura 10 muestra un método para preparar el molde de ácido nucleico. En este caso, la molécula de ácido nucleico que va a secuenciarse es bicatenaria y comprende una cadena sentido 1001 y una cadena antisentido 1002. Una horquilla de ácido nucleico 1003 puede ligarse en el extremo 3' de la cadena sentido y el extremo 5' de la cadena antisentido.

15 Las cadenas de la molécula de ácido nucleico pueden disociarse para formar una molécula de molde monocatenaria que comprende la cadena sentido 1004, la horquilla 1005 y la cadena antisentido 1006. Este ácido nucleico monocatenario puede secuenciarse tal como se describe en el presente documento.

20 En algunos casos, el presente método para preparar el molde de ácido nucleico permite secuenciar tanto la cadena sentido como la cadena antisentido en una única serie de secuenciación. Esto puede producir dos conjuntos de datos redundantes (es decir, cada posición de par de bases de ácido nucleico se secuencian dos veces) que pueden dar como resultado una determinación más precisa de la secuencia que secuenciar únicamente una cadena de la molécula de ácido nucleico bicatenario original.

Configuración de dispositivo

25 La figura 11 es un diagrama esquemático de un dispositivo de nanoporos 100 (o sensor) que puede usarse para secuenciar un ácido nucleico y/o detectar una molécula de etiqueta tal como se describe en el presente documento. La bicapa lipídica que contiene nanoporos puede estar caracterizada por una resistencia y capacitancia. El dispositivo de nanoporos 100 incluye una bicapa lipídica 102 formada en una superficie compatible con bicapa lipídica 104 de un sustrato sólido conductor 106, en el que la superficie compatible con bicapa lipídica 104 puede estar aislada mediante superficies incompatibles con bicapa lipídica 105 y el sustrato sólido conductor 106 puede estar eléctricamente aislado mediante materiales aislantes 107, y en el que la bicapa lipídica 102 puede estar rodeada por lípido amorfo 103 formado en la superficie incompatible con bicapa lipídica 105. La bicapa lipídica 102 puede tener incorporada una estructura de nanoporo individual 108 que tiene un nanoporo 110 lo suficientemente grande como para el paso de las moléculas de etiqueta que están caracterizándose y/o pequeños iones (por ejemplo, Na⁺, K⁺, Ca²⁺, Cl⁻) entre los dos lados de la bicapa lipídica 102. Una capa de moléculas de agua 114 puede adsorberse en la superficie compatible con bicapa lipídica 104 e intercalarse entre la bicapa lipídica 102 y la superficie compatible con bicapa lipídica 104. La película acuosa 114 adsorbida en la superficie compatible con bicapa lipídica hidrófila 104 puede fomentar el ordenamiento de moléculas lipídicas y facilitar la formación de bicapa lipídica en la superficie compatible con bicapa lipídica 104. Puede proporcionarse una cámara de muestra 116 que contiene una disolución de la molécula de ácido nucleico 112 y nucleótidos marcados con etiqueta sobre la bicapa lipídica 102. La disolución puede ser una disolución acuosa que contiene electrolitos y tamponada a una concentración de iones óptima y mantenida a un pH óptimo para mantener el nanoporo 110 abierto. El dispositivo incluye un par de electrodos 118 (que incluyen un nodo negativo 118a y un nodo positivo 118b) acoplados a una fuente de tensión variable 120 para proporcionar estímulo eléctrico (por ejemplo, polarización de tensión) a través de la bicapa lipídica y para detectar características eléctricas de la bicapa lipídica (por ejemplo, resistencia, capacitancia y flujo de corriente iónica). La superficie del electrodo positivo 118b es, o forma parte de, la superficie compatible con bicapa lipídica 104. El sustrato sólido conductor 106 puede estar acoplado a, o formar parte de, uno de los electrodos 118. El dispositivo 100 también puede incluir un circuito eléctrico 122 para controlar la estimulación eléctrica y para procesar la señal detectada. En algunas realizaciones, la fuente de tensión variable 120 se incluye como parte del circuito eléctrico 122. El conjunto de circuitos eléctrico 122 puede incluir amplificador, integrador, filtro de ruido, lógica de control de realimentación y/o diversos otros componentes. El conjunto de circuitos eléctrico 122 puede ser un conjunto de circuitos eléctrico integrado, integrado dentro de un sustrato de silicio 128 y puede estar acoplado además a un procesador informático 124 acoplado a una memoria 126.

55 La superficie compatible con bicapa lipídica 104 puede formarse a partir de diversos materiales que son adecuados para la transducción de iones y formación de gas para facilitar la formación de bicapa lipídica. En algunas realizaciones, pueden usarse materiales hidrófilos conductores o semiconductores porque pueden permitir una mejor detección de un cambio en las características eléctricas de bicapa lipídica. Los materiales de ejemplo incluyen Ag-AgCl, Au, Pt o silicio dopado u otros materiales de semiconductor. En algunos casos, el electrodo no es un electrodo de sacrificio.

La superficie incompatible con bicapa lipídica 105 puede formarse a partir de diversos materiales que no son adecuados para la formación de bicapa lipídica y normalmente son hidrófobos. En algunas realizaciones, se

prefieren materiales hidrófobos no conductores, dado que esto aísla eléctricamente las regiones de bicapa lipídica además de separar las regiones bicapa lipídica unas de otras. Los materiales incompatibles con bicapa lipídica de ejemplo incluyen por ejemplo nitruro de silicio (por ejemplo, Si_3N_4) y Teflon.

5 En un ejemplo, el dispositivo de nanoporos 100 de la figura 11 es un dispositivo de nanoporos de alfa-hemolisina (aHL) que tiene una única proteína alfa-hemolisina (aHL) 108 incorporada en una bicapa lipídica de difitanoilfosfatidilcolina (DPhPC) 102 formada sobre una superficie de Pt compatible con bicapa lipídica 104 recubierta en un material de aluminio 106. La superficie de Pt compatible con bicapa lipídica 104 está aislada mediante superficies de nitruro de silicio incompatibles con bicapa lipídica 105, y el material de aluminio 106 está eléctricamente aislado mediante materiales de nitruro de silicio 107. El aluminio 106 está acoplado a un conjunto de circuitos eléctrico 122 que está integrado en un sustrato de silicio 128. Un electrodo de plata-cloruro de plata colocado en el chip o que se extiende hacia abajo desde una placa de cubierta 128 entra en contacto con una disolución acuosa que contiene moléculas de ácido nucleico.

15 El nanoporo de aHL es un ensamblaje de siete péptidos individuales. La entrada o el vestíbulo del nanoporo de aHL tiene aproximadamente 26 Angstroms de diámetro, lo que es suficientemente ancho como para albergar una porción de una molécula de ADNbc. A partir del vestíbulo, el nanoporo de aHL en primer lugar se ensancha y después se estrecha hasta un cilindro que tiene un diámetro de aproximadamente 15 Angstroms, lo que es suficientemente ancho como para permitir que una única molécula de ADNmc (o las moléculas de etiqueta liberadas) pase a través pero no lo suficientemente ancho como para permitir que una molécula de ADNbc pase a través.

20 Además de DPhPC, la bicapa lipídica del dispositivo de nanoporos puede ensamblarse a partir de diversos otros materiales anfífilos adecuados, seleccionados basándose en diversas consideraciones, tales como el tipo de nanoporo usado, el tipo de molécula que está caracterizándose y diversas características físicas, químicas y/o eléctricas de la bicapa lipídica formada, tales como estabilidad y permeabilidad, resistencia y capacitancia de la bicapa lipídica formada. Los materiales anfífilos de ejemplo incluyen diversos fosfolípidos tales como palmitoil-oleoil-fosfatidilcolina (POPC) y éster metílico de dioleoil-fosfatidilo (DOPME), difitanoilfosfatidilcolina (DPhPC) 25 dipalmitoilfosfatidilcolina (DPPC), fosfatidilcolina, fosfatidiletanolamina, fosfatidilserina, ácido fosfatídico, fosfatidilinositol, fosfatidilglicerol y esfingomielina.

Además del nanoporo de aHL mostrado anteriormente, el nanoporo puede ser de diversos otros tipos de nanoporos. Los ejemplos incluyen γ -hemolisina, leucocidina, melitina y diversos otros nanoporos que se producen de manera natural, naturales modificados y sintéticos. Un nanoporo adecuado puede seleccionarse basándose en diversas características de la molécula de analito tales como el tamaño de la molécula de analito con respecto al tamaño de poro del nanoporo. Por ejemplo, el nanoporo de aHL que tiene un tamaño de poro restrictivo de aproximadamente 15 Angstroms.

Altas densidades de matriz

35 La matriz de detectores de nanoporos puede tener una alta densidad de sitios diferenciados. Por ejemplo, un gran número de sitios por área unitaria (es decir, densidad) permite la construcción de dispositivos más pequeños, que son portátiles, de bajo coste o tienen otras características ventajosas. Un gran número de sitios que comprenden un nanoporo y un circuito de detección puede permitir secuenciar un gran número de moléculas de ácido nucleico de una vez. Un sistema de este tipo puede aumentar el rendimiento y/o reducir el coste de secuenciación de una muestra de ácido nucleico.

40 Puede secuenciarse una muestra de ácido nucleico usando un sensor (o detector) que tiene un sustrato con una superficie que comprende sitios diferenciados, teniendo cada sitio individual un nanoporo, una polimerasa y opcionalmente al menos una enzima fosfatasa unida al nanoporo y un circuito de detección adyacente al nanoporo. El sistema puede comprender además una célula de flujo en comunicación de fluido con el sustrato, estando la célula de flujo adaptada para suministrar uno o más reactivos a dicho sustrato.

45 La superficie comprende cualquier densidad adecuada de sitios diferenciados (por ejemplo, una densidad adecuada para secuenciar una muestra de ácido nucleico en una cantidad dada de tiempo o para un coste dado). La superficie puede tener una densidad de sitios diferenciados mayor de o igual a aproximadamente 500 sitios por 1 mm^2 . En algunas realizaciones, la superficie tiene una densidad de sitios diferenciados de aproximadamente 200, aproximadamente 300, aproximadamente 400, aproximadamente 500, aproximadamente 600, aproximadamente 700, aproximadamente 800, aproximadamente 900, aproximadamente 1000, aproximadamente 2000, aproximadamente 3000, aproximadamente 4000, aproximadamente 5000, aproximadamente 6000, aproximadamente 7000, aproximadamente 8000, aproximadamente 9000, aproximadamente 10000, aproximadamente 20000, aproximadamente 40000, aproximadamente 60000, aproximadamente 80000, aproximadamente 100000 o aproximadamente 500000 sitios por 1 mm^2 . En algunos casos, la superficie tiene una densidad de sitios diferenciados de al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1000, al menos 2000, al menos 3000, al menos 4000, al menos 5000, al menos 6000, al menos 7000, al menos 8000, al menos 9000, al menos 10000, al menos 20000, al menos 40000, al menos 60000, al menos 80000, al menos 100000 o al menos 500000 sitios por 1 mm^2 .

Medición de corriente

Un chip de secuenciación basado en nanoporos puede incorporar un gran número de células que funcionan de manera autónoma o individualmente direccionables configuradas como una matriz. Un dispositivo de nanoporos puede incluir una matriz de nanoporos individualmente direccionables. Cada nanoporo individualmente direccionable puede incluir un electrodo individualmente direccionable. Por ejemplo una matriz de un millón de células puede estar construida por 1000 filas de células y 1000 columnas de células. Esta matriz puede permitir la secuenciación en paralelo de moléculas de ácido nucleico midiendo la diferencia de conductancia cuando etiquetas liberadas tras acontecimientos de incorporación de nucleótido pasan a través del nanoporo, por ejemplo. Además esa implementación de conjunto de circuitos permite determinar las características de conductancia del complejo poro-molécula lo que puede ser extremadamente valioso para distinguir etiquetas específicas.

En algunos casos, puede medirse la corriente a una tensión aplicada. Con el fin de lograr esto, puede aplicarse un potencial deseado al electrodo, y posteriormente puede mantenerse el potencial aplicado a lo largo de toda la medición. En una implementación, puede usarse una topología de integrador de op-amp para este fin tal como se describe en el presente documento. El integrador mantiene el potencial de tensión en el electrodo por medio de realimentación capacitiva. El circuito de integrador puede proporcionar características espectaculares de linealidad, coincidencia de célula a célula y desviación. El integrador de op-amp requiere normalmente un gran tamaño con el fin de lograr las prestaciones requeridas. En el presente documento se describe una topología de integrador más compacto.

En algunos casos, puede aplicarse un potencial de tensión "líquido" a la cámara lo que proporciona un potencial eléctrico común (por ejemplo, 350 mV) para todas las células en el chip. El circuito de integrador puede inicializar el electrodo (que es eléctricamente la placa superior del condensador de integración) hasta un potencial mayor que el potencial de líquido común. Por ejemplo, polarizar a 450 mV puede dar un potencial positivo de 100 mV entre electrodo y líquido. Este potencial de tensión positivo puede provocar que fluya una corriente desde el electrodo hasta el contacto de cámara de líquido. En este caso, los portadores son: (a) iones K^+ que fluyen a través del poro desde el lado de electrodo (trans) de la bicapa hasta el lado de depósito de líquido (cis) de la bicapa y (b) iones de cloro (Cl^-) en el lado trans que reaccionan con el electrodo de plata según la siguiente reacción electroquímica: $Ag + Cl^- \rightarrow AgCl + e^-$.

En algunos casos, K^+ fluye fuera de la célula encerrada (desde el lado trans hasta el cis de la bicapa) mientras que Cl^- se convierte en cloruro de plata. El lado de electrodo de la bicapa puede llegar a desalarse como resultado del flujo de corriente. En algunos casos, una matriz o material esponjoso de líquido de plata/cloruro de plata puede servir como depósito para suministrar iones Cl^- en la reacción inversa que se produce en el contacto de cámara eléctrica para completar el circuito.

En algunos casos, los electrones fluyen en última instancia sobre el lado superior del condensador de integración lo que crea la corriente eléctrica que se mide. La reacción electroquímica convierte plata en cloruro de plata y la corriente seguirá fluyendo únicamente mientras haya plata disponible para convertirse. El suministro limitado de plata conduce a una vida de electrodo dependiente de la corriente en algunos casos. En algunas realizaciones, se usan materiales de electrodo que no se agotan (por ejemplo, platino).

Conjunto de circuitos de célula

En la figura 14B se muestra un ejemplo de conjunto de circuitos de célula. Se aplica una tensión aplicada V_a a un op-amp 1400 previo a una compuerta de transporte de corriente de MOSFET 1401. También se muestran en este caso un electrodo 1402 y la resistencia del ácido nucleico y/o la etiqueta detectada mediante el dispositivo 1403.

Una tensión aplicada V_a puede impulsar la compuerta de transporte de corriente 1401. La tensión resultante en el electrodo es entonces $V_a - V_t$ donde V_t es la tensión umbral del MOSFET. En algunos casos, esto da como resultado un control limitado de la tensión real aplicada al electrodo dado que una tensión de umbral de MOSFET puede variar considerablemente a lo largo del procedimiento, tensión, temperatura e incluso entre dispositivos dentro de un chip. Esta variación de V_t puede ser mayor a niveles de corriente bajos en los que pueden intervenir efectos de fugas inferiores al umbral. Por tanto, con el fin de proporcionar un mejor control de la tensión aplicada, puede usarse un op-amp en una configuración de realimentación de seguidor con el dispositivo de transporte de corriente. Esto garantiza que la tensión aplicada al electrodo es V_a , independientemente de la variación de la tensión umbral de MOSFET.

En la figura 13 se muestra otro ejemplo de conjunto de circuitos de célula e incluye un integrador, comparador y lógica digital para desplazar hacia dentro bits de control y simultáneamente desplazar hacia fuera el estado de la salida de comparador. El conjunto de circuitos de célula puede estar adaptado para su uso con sistemas y métodos proporcionados en el presente documento. Las líneas B0 a B1 pueden salir del registro de desplazamiento. Las señales analógicas se comparten por todas las células dentro de un banco, mientras que las líneas digitales pueden encadenarse a modo de margarita de una célula a otra.

La lógica digital de célula comprende el registro de desplazamiento de datos (DSR) de 5 bits, registros de carga paralela (PLR) de 5 bits, lógica de control y circuito de integrador analógico. Usando la señal de LIN, los datos de

control desplazados en el DSR se cargan en paralelo en el PLR. Estos 5 bits controlan la lógica de sincronismo de "interrupción previa a la conexión" digital que controla las conmutaciones en la célula. Además, la lógica digital tiene un enclavamiento de establecer-restablecer (SR) para registrar la conmutación de la salida de comparador.

5 La arquitectura suministra una tasa de muestras variable que es proporcional a la corriente de célula individual. Una corriente superior puede dar como resultado más muestras por segundo que una corriente inferior. La resolución de la medición de corriente está relacionada con la corriente que está midiéndose. Una corriente pequeña puede medirse con una resolución más fina que una corriente grande, lo cual puede ser un beneficio con respecto a sistemas de medición de resolución fijada. Hay una entrada analógica que permite al usuario ajustar tasas de muestras cambiando la oscilación de tensión del integrador. Puede ser posible aumentar la tasa de muestras con el fin de analizar procesos biológicamente rápidos o ralentizar la tasa de muestras (y de ese modo ganar precisión) con el fin de analizar procesos biológicamente lentos.

15 La salida del integrador se inicializa a la tensión LVB (polarización de tensión baja) y se integra hasta la tensión CMP. Se genera una muestra cada vez que la salida de integrador oscila entre estos dos niveles. Por tanto, cuanto mayor es la corriente más rápido oscila la salida de integrador y por tanto más rápida es la tasa de muestras. De manera similar, si se reduce la tensión de CMP se reduce la oscilación de salida del integrador necesaria para generar una nueva muestra y por tanto se aumenta la tasa de muestras. Por tanto, simplemente reducir la diferencia de tensión entre LVB y CMP proporciona un mecanismo para aumentar la tasa de muestras.

20 Un chip de secuenciación basado en nanoporos puede incorporar un gran número de células que funcionan de manera autónoma o individualmente direccionables configuradas como matriz. Por ejemplo una matriz de un millón de células puede estar construida por 1000 filas de células y 1000 columnas de células. Esta matriz permite la secuenciación en paralelo de moléculas de ácido nucleico midiendo la diferencia de conductancia cuando se detectan etiquetas liberadas tras acontecimientos de incorporación de nucleótido mediante el nanoporo, por ejemplo. Además, esta implementación de conjunto de circuitos permite determinar las características de conductancia del complejo poro-molécula lo cual puede ser valioso para distinguir entre etiquetas.

25 Las estructuras de células electrónicas de nanoporo/bicapa integradas pueden aplicar tensiones apropiadas con el fin de realizar mediciones de corriente. Por ejemplo, puede ser necesario tanto (a) controlar el potencial de tensión de electrodo como (b) monitorizar la corriente de electrodo simultáneamente con el fin de funcionar correctamente.

30 Además, puede ser necesario controlar células independientemente unas de otras. Este control independiente de una célula puede requerirse con el fin de gestionar un gran número de células que pueden estar en diferentes estados físicos. Puede usarse un control preciso del estímulo de forma de onda de tensión lineal por tramos aplicado al electrodo para la transición entre los estados físicos de la célula.

35 Con el fin de reducir el tamaño y la complejidad de circuito puede ser suficiente proporcionar lógica para aplicar dos tensiones independientes. Esto permite dos agrupamientos independientes de células y aplicar estímulos de transición de estado correspondientes. Las transiciones de estado son de naturaleza estocástica con una probabilidad de aparición relativamente baja. Por tanto, puede ser altamente útil poder determinar la tensión de control apropiada y posteriormente realizar una medición para determinar si se ha producido la transición de estado deseada. Por ejemplo, puede aplicarse la tensión apropiada a una célula y después medirse la corriente para determinar si se ha formado una bicapa. Las células se dividen en dos grupos: (a) aquellas en las que se ha formado una bicapa y ya no necesitan que se les aplique la tensión. A esas células se les puede aplicar una polarización de 0 V con el fin de realizar la operación nula (NOP), es decir, permanecer en el mismo estado, y (b) aquellas que no tienen una bicapa formada. A esas células se les volverá a aplicar la tensión eléctrica de formación de bicapa.

45 Puede lograrse una simplificación y reducción de tamaño de circuito sustanciales restringiendo las tensiones aplicadas permitidas a dos y realizando la transición iterativa de células en lotes entre los estados físicos. Por ejemplo, puede lograrse una reducción en al menos un factor de 1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50 ó 100 restringiendo las tensiones aplicadas permitidas.

En la figura 14A se muestra aún otra implementación de la invención usando un circuito de medición compacto. En algunos casos, el circuito de medición compacto puede usarse para lograr las altas densidades de matriz descritas en el presente documento. Este circuito también está diseñado para aplicar una tensión al electrodo al tiempo que simultáneamente se miden corrientes de nivel bajo.

50 La célula funciona como integrador ultracompacto (UCI) y el funcionamiento básico se describe en el presente documento. Se conecta eléctricamente la célula a un electrodo electroquímicamente activo (por ejemplo, AgCl) a través de la conexión de electrodo-detección (ELSNS). El transistor de NMOS M11 realiza dos funciones independientes: (1) funciona como seguidor de fuente para aplicar una tensión al nodo de ELSNS dada por ($V_{g1} - V_{t1}$) y (2) funciona como transportador de corriente para mover electrones desde el condensador C1 hasta el nodo de ELSNS (y viceversa).

55 En algunos casos, puede aplicarse un potencial de tensión controlado al electrodo de ELSNS y puede hacerse variar simplemente cambiando la tensión en la compuerta del seguidor de fuente de electrodo M11. Además cualquier corriente desde el pin de fuente de M11 se propaga directamente y con precisión hasta el pin de sumidero de M11

en el que puede acumularse en el condensador C0. Por tanto M11 y C0 actúan juntos como integrador ultracompacto. Este integrador puede usarse para determinar la corriente emitida/extraída hacia/desde el electrodo midiendo el cambio en tensión integrado en el condensador según lo siguiente: $I \cdot t = C \cdot V$, donde I es la corriente, t es el tiempo, C es la capacitancia y V es el cambio de tensión.

5 En algunos casos, el cambio de tensión se mide a un intervalo fijo t (por ejemplo, cada 1 ms).

El transistor M2 puede estar configurado como seguidor de fuente con el fin de amortiguar la tensión de condensador y proporcionar una representación de baja impedancia de la tensión integrada. Esto impide que la compartición de carga cambie la tensión en el condensador.

10 El transistor M3 puede usarse como dispositivo de acceso a fila con la salida de tensión analógica AOUT conectada como columna compartida con muchas otras células. Sólo se permite una única fila de la señal AOUT conectada a columna de modo que se mide la tensión de una única célula.

En una implementación alternativa el transistor M3 puede omitirse conectando el sumidero del transistor M2 a una "carril conmutado" seleccionable de fila.

15 El transistor M4 puede usarse para restablecer la célula a una tensión de partida predeterminada a partir de la cual se integra la tensión. Por ejemplo, aplicar una alta tensión (por ejemplo: hasta $V_{DD} = 1,8 \text{ V}$) tanto a RST como a RV aumentará el condensador hasta un valor previamente cargado de $(V_{DD} - V_{t5})$. El valor de partida exacto puede variar tanto de célula a célula (debido a la variación de V_t de M4 y M2) como también de medición a medición debido al ruido térmico de conmutación de restablecimiento (ruido \sqrt{KTC}). Como resultado, se usa una técnica de muestreo doble correlacionado (CDS) para medir la tensión de partida de integrador y la tensión final para
20 determinar el cambio de tensión real durante este periodo de integración.

Obsérvese también que el sumidero del transistor M4 puede conectarse a una tensión controlada RV (tensión de restablecimiento). En funcionamiento normal esto puede impulsarse hasta V_{DD} , sin embargo también puede impulsarse hasta una tensión baja. Si el "sumidero" de M4 se impulsa de hecho hasta la tierra, entonces el flujo de corriente puede invertirse (es decir, la corriente puede fluir desde el electrodo al interior del circuito a través de M1 y
25 M4 y puede intercambiarse la noción de fuente y sumidero). En algunos casos, cuando se hace funcionar el circuito en este modo, la tensión negativa aplicada al electrodo (con respecto a la referencia de líquido) se controla mediante esta tensión RV (suponiendo que V_{g1} y V_{g5} son al menos un umbral mayor que RV). Por tanto, puede usarse una tensión de tierra en RV para aplicar una tensión negativa al electrodo (por ejemplo, para lograr una electroporación o formación de bicapa).

30 Un convertidor de analógico a digital (ADC, no mostrado) mide la tensión de AOUT inmediatamente después del restablecimiento y de nuevo después del periodo de integración (realiza la medición de CDS) con el fin de determinar la corriente integrada durante un periodo de tiempo fijo. Y puede implementarse un ADC por columna o usarse un transistor independiente para cada columna como multiplexador analógico para compartir un único ADC entre múltiples columnas. Este factor de multiplexación de columna puede hacerse variar dependiendo de los
35 requisitos para ruido, precisión y rendimiento.

En cualquier momento dado, cada célula puede estar en uno de cuatro estados físicos diferentes: (1) cortocircuito a líquido (2) bicapa formada (3) bicapa + poro (4) bicapa + poro + ácido nucleico y/o moléculas de etiqueta.

40 En algunos casos, se aplica una tensión con el fin de mover células entre estados. La operación NOP se usa para dejar una célula en un estado deseado particular mientras que otras células se estimulan con un potencial aplicado para moverse de un estado a otro.

45 Esto puede lograrse teniendo dos (o más) tensiones diferentes que pueden aplicarse a la tensión de compuerta del seguidor de fuente de M1 que se usa indirectamente para controlar la tensión aplicada al electrodo con respecto al potencial de líquido. Por tanto, el transistor M5 se usa para aplicar una tensión A mientras que el transistor M6 se usa para aplicar una tensión B. Por tanto, en conjunto M5 y M6 funcionan como un multiplexador analógico impulsándose o bien SELA o bien SELB altas para seleccionar la tensión.

Dado que cada célula puede estar en un estado diferente posible y dado que SELA y SELB son complementarias, puede usarse un elemento de memoria en cada célula para seleccionar entre la tensión A o B. Este elemento de memoria puede ser un elemento dinámico (condensador) que se refrescó en cada ciclo o un simple elemento de memoria de eliminador de enclavamiento-enclavamiento (inversor con acoplamiento cruzado).

50 Estructura de chip de prueba de op-amp

55 En algunos ejemplos, un chip de prueba incluye una matriz de 264 sensores dispuestos en cuatro grupos independientes (también conocidos como bancos) de 66 células de sensor cada uno. Cada grupo está dividido a su vez en tres "columnas" con 22 "células" de sensores en cada columna. El nombre de "célula" es intencionado dado que de manera ideal se forma una célula virtual que consiste en una bicapa lipídica y nanoporo insertado por encima de cada uno de los 264 sensores en la matriz (aunque el dispositivo puede funcionar satisfactoriamente con tan sólo

una fracción de las células de sensor así pobladas).

Hay un único panel de I/O analógico que aplica un potencial de tensión al líquido contenido dentro de un cilindro conductor montado en la superficie de la pastilla. Este potencial de "líquido" se aplica al lado superior del poro y es común para todas las células en una matriz de detectores. El lado inferior del poro tiene un electrodo expuesto y cada célula de sensor puede aplicar un potencial de lado inferior diferenciado a su electrodo. Después se mide la corriente entre la conexión de líquido superior y la conexión de electrodo de cada célula en el lado inferior del poro. La célula de sensor mide la corriente que se desplaza a través del poro según se modula mediante la molécula de etiqueta que pasa dentro del poro.

En algunos casos, cinco bits controlan el modo de cada célula de sensor. Siguiendo haciendo referencia a la figura 12, cada una de las 264 células en la matriz puede controlarse individualmente. Se aplican valores de manera independiente a un grupo de 66 células. El modo de cada una de las 66 células en un grupo se controla desplazando en serie 330 ($66 * 5$ bits/célula) valores digitales en un registro de desplazamiento de datos (DSR). Estos valores se desplazan en la matriz usando los pines KIN (reloj) y DIN (entrada de datos) con un par de pines independiente para cada grupo de 66 células.

Por tanto se usan 330 relojes para desplazar 330 bits en el registro de desplazamiento DSR. Un segundo registro de carga paralela (PLR) de 330 bits se carga en paralelo a partir de este registro de desplazamiento cuando se determina que la LIN<i> (entrada de carga) correspondiente es alta. Al mismo tiempo que se carga en paralelo el PLR, se carga el valor de estado de la célula en el DSR.

Una operación completa puede consistir en 330 relojes para desplazar 330 bits de datos en el DSR, un único ciclo de reloj con señal de LIN determinada como alta, seguido por 330 ciclos de reloj para leer los datos de estado captados desplazados fuera del DSR. La operación se optimiza de modo que pueden desplazarse 330 nuevos bits en el DSR simultáneamente mientras están leyéndose los 330 bits a partir de la matriz. Por tanto, a una frecuencia de reloj de 50 MHz el tiempo de ciclo para una lectura es de $331/50$ MHz = 6,62 us.

Nucleótidos marcados con etiqueta

En algunos casos un nucleótido marcado con etiqueta comprende una etiqueta que puede escindirse en un acontecimiento de polimerización de nucleótido y detectarse con la ayuda de un nanoporo. La etiqueta puede estar unida al 5'-fosfato del nucleótido. En algunos casos, la etiqueta no es un fluoróforo. La etiqueta puede ser detectable mediante su carga, forma, tamaño o cualquier combinación de los mismos. Las etiquetas a modo de ejemplo incluyen diversos polímeros. Cada tipo de nucleótido (es decir, A, C, G, T) comprende generalmente una etiqueta única.

Las etiquetas pueden estar ubicadas en cualquier posición adecuada en el nucleótido. La figura 6 proporciona un ejemplo de un nucleótido marcado con etiqueta. En este caso, R_1 es generalmente OH y R_2 es H (es decir, para ADN) u OH (es decir, para ARN), aunque otras modificaciones son aceptables. En la figura 6, X es cualquier ligador adecuado. En algunos casos, el ligador puede escindirse. Los ejemplos de ligadores incluyen, sin limitación, O, NH, S o CH_2 . Los ejemplos de grupos químicos adecuados para la posición Z incluyen O, S o BH_3 . La base es cualquier base adecuada para su incorporación en un ácido nucleico incluyendo adenina, guanina, citosina, timina, uracilo o un derivado de las mismas. En algunos casos también son aceptables bases universales.

El número de fosfatos (n) es cualquier valor de número entero adecuado (por ejemplo, un número de fosfatos tal que el nucleótido puede incorporarse en una molécula de ácido nucleico). En algunos casos, todos los tipos de nucleótidos marcados con etiqueta tienen el mismo número de fosfatos, pero esto no se requiere. En algunas aplicaciones, hay una etiqueta diferente para cada tipo de nucleótido y el número de fosfatos no se usa necesariamente para distinguir las diversas etiquetas. Sin embargo, en algunos casos más de un tipo de nucleótido (por ejemplo, A, C, T, G o U) tienen la misma molécula de etiqueta y la capacidad para distinguir un nucleótido de otro está determinada al menos en parte por el número de fosfatos (teniendo diversos tipos de nucleótidos un valor diferente para n). En diversas realizaciones, el valor para n es 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o mayor.

A continuación se describen etiquetas adecuadas. En algunos casos, la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto. Cuando se une la etiqueta, la carga en el compuesto global puede ser neutra. La liberación de la etiqueta puede dar como resultado dos moléculas, una etiqueta cargada y un nucleótido cargado. En algunos casos la etiqueta cargada pasa a través de un nanoporo y se detecta.

En la figura 7 se muestran más ejemplos de nucleótidos marcados con etiqueta adecuados. La etiqueta puede estar unida a la molécula de azúcar, la molécula de base o cualquier combinación de las mismas. Con referencia a la figura 7, Y es una etiqueta y X es un ligador escindible. Además, R_1 , si está presente, es generalmente OH, $-OCH_2N_3$ u $-O-2$ -nitrobenzilo, y R_2 , si está presente, es generalmente H. Además, Z es generalmente O, S o BH_3 , y n es cualquier número entero incluyendo 1, 2, 3 ó 4. En algunos casos, A es O, S, CH_2 , CHF, CFF o NH.

Siguiendo haciendo referencia a la figura 7, el tipo de base en cada análogo de dNPP es generalmente diferente del tipo de base en cada uno de los otros tres análogos de dNPP, y el tipo de etiqueta en cada análogo de dNPP es generalmente diferente del tipo de etiqueta en cada uno de los otros tres análogos de dNPP. Las bases adecuadas

incluyen, pero no se limitan a, adenina, guanina, citosina, uracilo o timina o un derivado de cada una de las mismas. En algunos casos, la base es una de 7-desazaguanina, 7-desaza-adenina o 5-metilcitosina.

En casos en los que R_1 es $-O-CH_2N_3$, los métodos comprenden opcionalmente además tratar el análogo de dNPP incorporado para retirar el $-CH_2N_3$ y dar como resultado un grupo OH unido a la posición 3' permitiendo así la incorporación de un análogo de dNPP adicional.

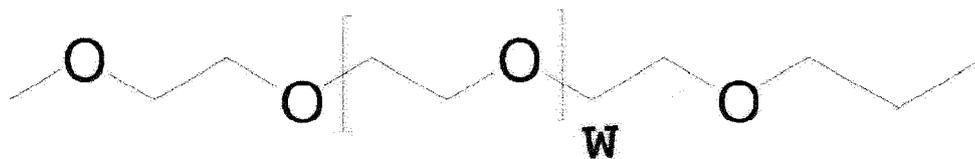
En casos en los que R_1 es $-O-2$ -nitrobencilo, los métodos comprenden opcionalmente además tratar el análogo de nucleótido incorporado para retirar el -2 -nitrobencilo y dar como resultado un grupo OH unido a la posición 3' permitiendo así la incorporación de un análogo de dNPP adicional.

Etiquetas a modo de ejemplo

Una etiqueta puede ser cualquier grupo químico o molécula que puede detectarse en un nanoporo. En algunos casos, una etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alquenilo, un grupo alquinilo, un grupo azido o una combinación de los mismos.

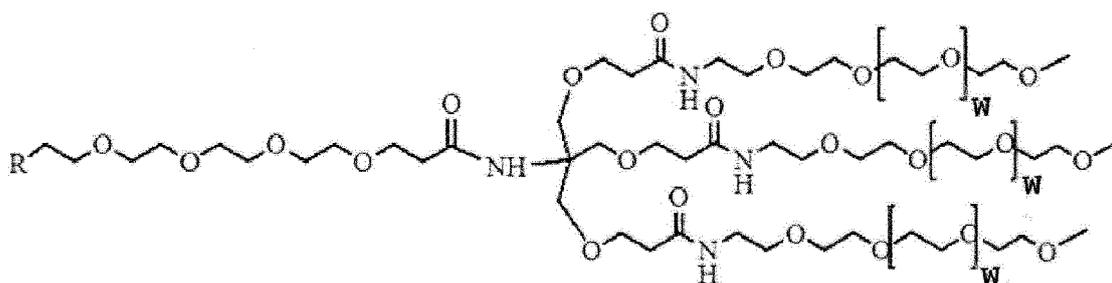
También se contempla que la etiqueta comprenda además un número apropiado de lisinas o argininas para equilibrar el número de fosfatos en el compuesto.

En algunos casos, la etiqueta es un polímero. Polietilenglicol (PEG) es un ejemplo de un polímero y tiene la siguiente estructura:



Puede usarse cualquier número de unidades de etilenglicol (W). En algunos casos, W es un número entero entre 0 y 100. En algunos casos, el número de unidades de etilenglicol es diferente para cada tipo de nucleótido. En una realización, los cuatro tipos de nucleótidos comprenden etiquetas que tienen 16, 20, 24 ó 36 unidades de etilenglicol. En algunos casos, la etiqueta comprende además un resto identificable adicional, tal como un colorante basado en cumarina.

En algunos casos, una etiqueta comprende múltiples cadenas de PEG. En un ejemplo, una etiqueta tiene la siguiente estructura:



en la que R es NH_2 , OH, COOH, CHO, SH o N_3 , y W es un número entero de desde 0 hasta 100.

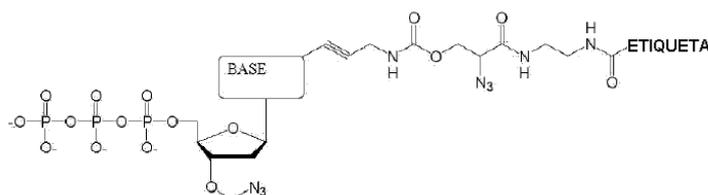
En algunos casos se elige una etiqueta de las moléculas (dCp)m, (dGp)m, (dAp)m, y (dTp)m. La figura 8 muestra estas moléculas unidas a un nucleótido. En este caso, "m" es, independientemente, un número entero de desde 0 hasta 100, y en las que cuando m es 0 el fosfato terminal del dNPP está unido directamente al átomo de 3'-O del nucleósido mostrado en el lado izquierdo de la estructura. En algunos casos, el valor de n es diferente para cada tipo de base.

En algunos casos, una etiqueta es un hidrocarbilo, sustituido o no sustituido, tal como un alquilo, alquenilo, alquinilo, y que tiene una masa de 3000 Dalton o menos.

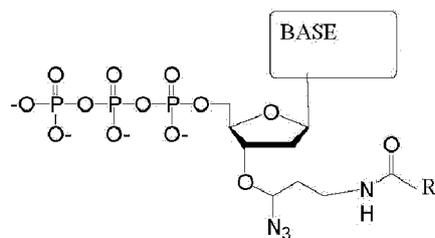
Tal como se usa en el presente documento, el término "alquilo" incluye grupos de hidrocarburo alifático saturado tanto ramificado como de cadena lineal que tienen el número especificado de átomos de carbono y pueden no estar sustituidos o estar sustituidos. Tal como se usa en el presente documento, "alquenilo" se refiere a un radical de hidrocarburo no aromático, lineal o ramificado, que contiene al menos 1 doble enlace de carbono a carbono, y

5 pueden estar presentes hasta el máximo número posible de dobles enlaces carbono-carbono no aromáticos, y puede estar no sustituido o estar sustituido. El término "alquilo" se refiere a un radical de hidrocarburo lineal o ramificado, que contiene al menos 1 triple enlace de carbono a carbono, y pueden estar presentes hasta el máximo número posible de triples enlaces carbono-carbono no aromáticos, y puede estar no sustituido o estar sustituido. El término "sustituido" se refiere a un grupo funcional tal como se describió anteriormente tal como un alquilo, o un hidrocarbilo, en el que al menos un enlace con un átomo de hidrógeno contenido en el mismo se sustituye por un enlace con un átomo distinto de hidrógeno o distinto de carbono, siempre que se mantengan las valencias normales y que la(s) sustitución/sustituciones de(n) como resultado un compuesto estable. Los grupos sustituidos también incluyen grupos en los que uno o más enlaces con un átomo de carbono o hidrógeno se sustituyen por uno o más enlaces, incluyendo dobles o triples enlaces, con un heteroátomo.

Los ejemplos no limitativos de nucleótidos marcados con etiqueta incluyen compuestos que tienen la estructura:

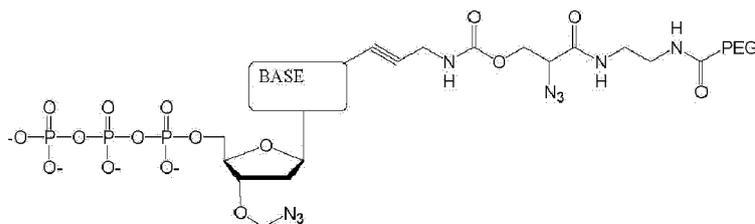


o



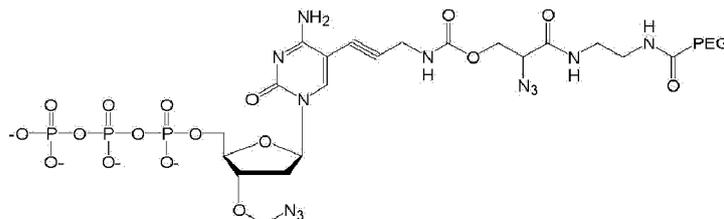
15 en las que "R" es un hidrocarbilo sustituido o no sustituido, de hasta 3000 Dalton, y en las que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

Ejemplos no limitativos adicionales de nucleótidos marcados con etiqueta incluyen compuestos que tienen la estructura:

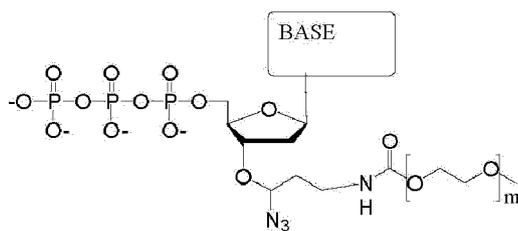


20 en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

Ejemplos no limitativos adicionales de nucleótidos marcados con etiqueta incluyen compuestos que tienen la estructura:



25 Ejemplos no limitativos adicionales de nucleótidos marcados con etiqueta incluyen compuestos que tienen la estructura:

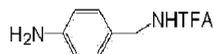


en la que m es un número entero de desde 1-50, y en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

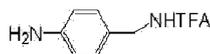
Métodos para unir etiquetas

- 5 Puede usarse cualquier método adecuado para unir las etiquetas. En un ejemplo, pueden unirse etiquetas al fosfato terminal mediante (a) poner en contacto un trifosfato de nucleótido con diciclohexilcarbodiimida/dimetilformamida en condiciones que permiten producción de un trimetafosfato cíclico; (b) poner en contacto el producto resultante de la etapa a) con un nucleófilo para formar un compuesto funcionalizado con -OH o -NH₂; y (c) hacer reaccionar el producto de la etapa b) con una etiqueta que tiene un grupo -COR unido a la misma en condiciones que permiten que la etiqueta se una indirectamente a un fosfato terminal formando así el análogo de trifosfato de nucleótido.

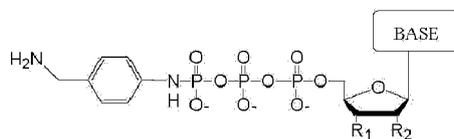
En algunos casos, el nucleófilo es H₂N-R-OH, H₂N-R-NH₂, R'S-R-OH, R'S-R-NH₂ o



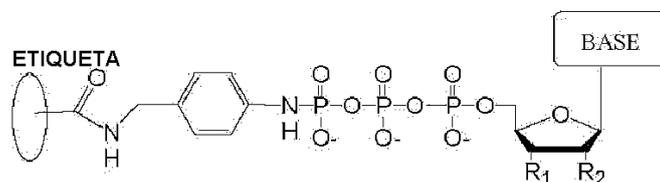
En algunos casos, el método comprende, en la etapa b), poner en contacto el producto resultante de la etapa a) con un compuesto que tiene la estructura:



- 15 y posteriormente o de manera simultánea poner en contacto el producto con NH₄OH para formar un compuesto que tiene la estructura:



- 20 Después puede hacerse reaccionar el producto de la etapa b) con una etiqueta que tiene un grupo -COR unido a la misma en condiciones que permiten que la etiqueta se una indirectamente a un fosfato terminal formando así el análogo de trifosfato de nucleótido que tiene la estructura:



en la que R₁ es OH, en la que R₂ es H u OH, en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

- 25 Liberación de etiquetas

Una etiqueta puede liberarse de cualquier manera. En algunos casos, la etiqueta está unida a polifosfato (por ejemplo, figura 6) y la incorporación del nucleótido en una molécula de ácido nucleico da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo. La incorporación puede catalizarse mediante al menos una polimerasa, opcionalmente unida al nanoporo. En algunos casos, al menos una enzima fosfatasa también está unida al poro. La enzima fosfatasa puede escindir la etiqueta a partir del polifosfato para liberar la etiqueta. En algunos casos, las enzimas fosfatasa están posicionadas de tal manera que pirofosfato producido mediante la polimerasa en una reacción de polimerasa interacciona con las enzimas fosfatasa antes de entrar en el poro.

En algunos casos, la etiqueta no está unida a polifosfato (véase, por ejemplo, la figura 7). En estos casos, la etiqueta

está unida mediante un ligador escindible (X). En la patente estadounidense n.º 6.664.079 se dan a conocer métodos para la producción de análogos de nucleótido con grupos reactivos ocupados de manera escindible y/o unidos de manera escindible.

5 El ligador puede ser cualquier ligador adecuado y escindirse de cualquier manera adecuada. Los ligadores pueden ser fotoescindibles. En una realización se usa luz UV para escindir de manera fotoquímica los restos y ligadores escindibles de manera fotoquímica. En una realización, el ligador fotoescindible es un resto 2-nitrobenzilo.

El grupo $-CH_2N_3$ puede tratarse con TCEP (tris(2-carboxietil)fosfina) para retirarlo a partir del átomo de 3'-O de un análogo de dNPP, o análogo de rNPP, creando así un grupo 3'-OH.

Detección de etiquetas

10 Las etiquetas pueden fluir a través de un nanoporo después de liberarse a partir del nucleótido. En algunos casos, se aplica una tensión para tirar de las etiquetas a través del nanoporo. Al menos aproximadamente el 85%, al menos el 90%, al menos el 95%, al menos el 99,9 o al menos el 99,99% de las etiquetas liberadas pueden translocarse a través del nanoporo.

15 En algunos casos del método, una polimerasa extrae a partir de una combinación de nucleótidos marcados con etiqueta que comprende una pluralidad de bases diferentes (por ejemplo, A, C, G, T y/o U). También es posible poner en contacto de manera iterativa la polimerasa con los diversos tipos de bases marcadas con etiqueta. En este caso, puede no ser necesario que cada tipo de nucleótido tenga una base única, pero la realización de ciclos entre diferentes tipos de base añade coste y complejidad al procedimiento en algunos casos, no obstante esta realización queda abarcada en la presente invención.

20 La figura 9 muestra que la incorporación del nucleótido marcado con etiqueta en una molécula de ácido nucleico (por ejemplo, usando una polimerasa para extender una base de cebador emparejada a un molde) libera una etiqueta detectable-polifosfato. En algunos casos, el polifosfato de etiqueta se detecta a medida que pasa a través del nanoporo. Incluso es posible distinguir el nucleótido basándose en el número de fosfatos que comprende el polifosfato (por ejemplo, aunque las etiquetas sean idénticas). No obstante, cada tipo de nucleótido tiene
25 generalmente una etiqueta única.

Con referencia a la figura 9, el compuesto de polifosfato de etiqueta puede tratarse con fosfatasa (por ejemplo, fosfatasa alcalina) antes de hacer pasar la etiqueta a través de un nanoporo y medir la corriente iónica.

30 La etiqueta puede detectarse en el nanoporo (al menos en parte) debido a su carga. En algunos casos, el compuesto de etiqueta es un compuesto cargado de manera alternativa que tiene una primera carga neta y, después de una reacción química, física o biológica, una segunda carga neta diferente. En algún caso, la magnitud de la carga en la etiqueta es la misma que la magnitud de la carga en el resto del compuesto. En una realización, la etiqueta tiene una carga positiva y la retirada de la etiqueta cambia la carga del compuesto.

35 En algunos casos, a medida que la etiqueta pasa a través del nanoporo, puede generar un cambio electrónico. En algunos casos el cambio electrónico es un cambio en la amplitud de corriente, un cambio en la conductancia del nanoporo o cualquier combinación de los mismos.

El nanoporo puede ser biológico o sintético. También se contempla que el poro sea proteico, por ejemplo en el que el poro es una proteína alfa-hemolisina. Un ejemplo de un nanoporo sintético es un poro de estado sólido o grafeno.

40 En algunos casos, se unen enzimas polimerasa y/o enzimas fosfatasa al nanoporo. Las proteínas de fusión o reticulaciones de disulfuro son ejemplos de métodos para unirse a un nanoporo proteico. En el caso de un nanoporo de estado sólido, la unión a la superficie cerca del nanoporo puede realizarse mediante uniones de biotina-estreptavidina. En un ejemplo la ADN polimerasa se une a una superficie sólida a través de una superficie de oro modificada con una monocapa autoensamblada de alcanotiol funcionalizada con grupos amino, en la que los grupos amino están modificados para dar ésteres de NHS para su unión a grupos amino en la ADN polimerasa.

45 Un aspecto de la presente divulgación proporciona un método para secuenciar una molécula de ácido nucleico. En algunas realizaciones, dicho procesador informático está en una estación de trabajo que está en proximidad de dicho chip. En algunos casos, la etiqueta pasa a través del nanoporo. En algunas realizaciones, la etiqueta pasa adyacente al nanoporo. En algunas realizaciones, la tasa de polimerización es menor que la tasa de paso de etiqueta a través del, o adyacente al, nanoporo. En algunos casos, dicho electrodo está adaptado para suministrar un estímulo eléctrico a través de dicha membrana. El chip puede tener características y propiedades dadas a
50 conocer, por ejemplo, en la patente estadounidense n.º 8.324.914.

55 En algunas realizaciones, dicha membrana tiene una capacitancia mayor de aproximadamente $5 \text{ fF}/\mu\text{m}^2$ según se mide a través de dicha membrana. En algunos casos, dicha membrana tiene una resistencia mayor de o igual a aproximadamente $500 \text{ M}\Omega$ según se mide a través de dicha membrana. En algunas realizaciones, dicha membrana tiene una resistencia menor de o igual a aproximadamente $1 \text{ G}\Omega$ a través de dicha membrana. En algunos casos, dicha resistencia se mide con la ayuda de electrodos opuestos dispuestos adyacentes a dicha membrana. En

algunas realizaciones, dicha resistencia se mide con la ayuda de electrodos opuestos dispuestos adyacentes a dicha membrana.

5 En algunos casos, cada nanoporo individualmente direccionable está adaptado para regular el flujo molecular. En algunos casos, dicho nanoporo individualmente direccionable está adaptado para detectar dicha etiqueta tras el flujo molecular de dicha etiqueta del mismo a través de, o adyacente a, dicho nanoporo. En algunos casos, dicho electrodo es individualmente direccionable. En algunos casos, dicho electrodo está acoplado a un circuito integrado que procesa una señal detectada con la ayuda de dicho electrodo.

En algunos casos, dicho circuito integrado comprende un controlador lógico. En algunos casos, dicho electrodo forma parte de un circuito integrado que procesa una señal detectada con la ayuda de dicho electrodo.

10 En algunos casos, dicha membrana es una bicapa lipídica. En algunos casos, la membrana es una bicapa lipídica de difitanoilfosfatidilcolina (DPhPC). En algunos casos, el nanoporo es un nanoporo de alfa-hemolisina. En algunos casos, dicha membrana muestra (i) una capacitancia mayor de aproximadamente $5 \text{ fF}/\mu\text{m}^2$ o una resistencia menor de o igual a aproximadamente $1 \text{ G}\Omega$ a través de dicha membrana, o (ii) una capacitancia mayor de aproximadamente $5 \text{ fF}/\mu\text{m}^2$ y una resistencia menor de o igual a aproximadamente $1 \text{ G}\Omega$ a través de dicha membrana. En algunos casos, dicha membrana está dispuesta adyacente a una superficie compatible con membrana. En algunos casos, dicha pluralidad de nanoporos individualmente direccionables están a una densidad de al menos aproximadamente 500, 600, 700, 800, 900, 1000, 10.000, 100.000 ó 1.000.000 nanoporos individualmente direccionables por mm^2 .

20 En algunos casos, cada tipo de nucleótido comprende una etiqueta única. En algunos casos, la etiqueta está inicialmente unida al 5'-fosfato del nucleótido individual. En algunos casos, el cebador se aparea a una posición específica en el molde de ácido nucleico monocatenario.

25 Otro aspecto de la presente divulgación proporciona un método para la secuenciación de ácidos nucleicos, que comprende detectar, con la ayuda de un nanoporo, la incorporación de un nucleótido en una molécula de ácido nucleico, en el que la molécula de ácido nucleico no pasa a través del nanoporo. En algunos casos, etiquetas asociadas con los nucleótidos se liberan tras la incorporación, y en los que después de liberarse las etiquetas pasan a través del nanoporo.

En algunos casos, se detectan acontecimientos de incorporación de nucleótido con una precisión de al menos 4σ . En algunos casos, se detectan acontecimientos de incorporación de nucleótido con una precisión de al menos 5σ . En algunos casos, se detectan acontecimientos de incorporación de nucleótido con una precisión de al menos 6σ .

30 Otro aspecto de la presente divulgación proporciona un método para la secuenciación de ácidos nucleicos, que comprende detectar un subproducto de un acontecimiento de incorporación de nucleótido individual con la ayuda de un nanoporo. En algunos casos, el nucleótido no se detecta directamente mediante dicho nanoporo. En algunos casos, el subproducto del acontecimiento de incorporación de nucleótido es una molécula de etiqueta que se libera tras dicho acontecimiento de incorporación de nucleótido individual. En algunos casos, la molécula de etiqueta pasa a través del nanoporo. Un método para secuenciar una molécula de ácido nucleico, que comprende distinguir entre acontecimientos de incorporación de nucleótido individuales con una precisión de más de 4σ . En algunos casos, la precisión es mayor de 5σ . En algunos casos, la precisión es mayor de 6σ . En algunos casos, los acontecimientos de incorporación de nucleótido se detectan con la ayuda de un nanoporo. En algunos casos, dicho nanoporo es un nanoporo individualmente direccionable. En algunos casos, dicho nanoporo está en una membrana que está dispuesta adyacente a un electrodo. En algunos casos, dicho electrodo está en una matriz de electrodos a una densidad de al menos aproximadamente 500 electrodos por mm^2 . En algunos casos, dichos acontecimientos de incorporación de nucleótido individuales comprenden la incorporación de un nucleótido en una cadena de ácido nucleico que es complementaria a dicha molécula de ácido nucleico, en el que dicho nucleótido comprende una etiqueta que se libera tras la incorporación de dicho nucleótido en dicha cadena de ácido nucleico, y en el que dicha etiqueta pasa a través de, o adyacente a, dicho nanoporo después de liberarse a partir de dicho nucleótido. En algunos casos, dicha etiqueta se detecta con la ayuda de dicho electrodo después de liberarse a partir de dicho nucleótido.

35 Otro aspecto de la presente divulgación proporciona un método para la secuenciación de ácidos nucleicos, comprendiendo el método: (a) proporcionar una matriz de nanoporos, en el que un nanoporo individual en dicha matriz está acoplado a una ácido nucleico polimerasa; y (b) polimerizar nucleótidos marcados con etiqueta con la polimerasa, en el que un nucleótido individual marcado con etiqueta comprende una etiqueta, y en el que la etiqueta se libera y se detecta con la ayuda del nanoporo. En algunos casos, la etiqueta pasa adyacente al nanoporo después de liberarse. En algunos casos, la etiqueta pasa a través del nanoporo después de liberarse. En algunos casos, la tasa de polimerización es menor que la tasa de paso de etiqueta a través del nanoporo. En algunos casos, los nanoporos son individualmente direccionables.

Otro aspecto de la presente divulgación proporciona un nucleótido marcado con etiqueta, en el que el nucleótido comprende una etiqueta que puede escindirse en un acontecimiento de polimerización de nucleótido y detectarse con la ayuda de un nanoporo en un chip que comprende una matriz de nanoporos. En algunos casos, la etiqueta

está unida al 5'-fosfato del nucleótido. En algunos casos, la etiqueta no es un fluoróforo. En algunos casos, la etiqueta puede detectarse mediante su carga, forma, tamaño o cualquier combinación de los mismos.

- Otro aspecto de la presente divulgación proporciona un sistema para secuenciar una molécula de ácido nucleico. En algunos casos, dicho electrodo está adaptado para suministrar un estímulo eléctrico a través de dicha membrana.
- 5 En algunos casos, dicha membrana tiene una capacitancia mayor de aproximadamente $5 \text{ fF}/\mu\text{m}^2$ según se mide a través de dicha membrana. En algunos casos, dicha membrana tiene una resistencia mayor de o igual a aproximadamente $500 \text{ M}\Omega$ según se mide a través de dicha membrana. En algunos casos, dicha membrana tiene una resistencia menor de o igual a aproximadamente $1 \text{ G}\Omega$ a través de dicha membrana. En algunos casos, dicha resistencia se mide mediante electrodos opuestos dispuestos adyacentes a dicha membrana. En algunos casos, cada nanoporo individualmente direccionable está adaptado para regular flujo molecular. En algunos casos, cada nanoporo individualmente direccionable está adaptado para regular flujo molecular con la ayuda de un estímulo eléctrico aplicado a dicho nanoporo. En algunos casos, dicho procesador informático está en una estación de trabajo que está en proximidad de dicho chip. En algunos casos, dicho procesador informático está comprendido en dicho chip. En algunos casos, dicho nanoporo individualmente direccionable está adaptado para detectar dicha etiqueta tras el flujo molecular de dicha etiqueta del mismo a través de, o adyacente a, dicho nanoporo. En algunos casos, dicho electrodo es individualmente direccionable. En algunos casos, dicho electrodo está acoplado a un circuito integrado que procesa una señal detectada con la ayuda de dicho electrodo. En algunos casos, dicho circuito integrado comprende un controlador lógico. En algunos casos, dicho electrodo forma parte de un circuito integrado que procesa una señal detectada con la ayuda de dicho electrodo.
- 10
- 15
- 20 En algunos casos, dicha membrana es una bicapa lipídica. En algunos casos, el nanoporo es un nanoporo de alfa-hemolisina. En algunos casos, la membrana es una bicapa lipídica de difitanoiolfosfatidilcolina (DPhPC). En algunos casos, dicha membrana está dispuesta adyacente a una superficie compatible con membrana. En algunos casos, dicha pluralidad de nanoporos individualmente direccionables están a una densidad de al menos aproximadamente 500 nanoporos individualmente direccionables por mm^2 . En algunos casos, dicha densidad es de al menos
- 25 aproximadamente 1000 nanoporos individualmente direccionables por mm^2 .

- Otro aspecto de la presente divulgación proporciona un método para secuenciar una molécula de ácido nucleico, comprendiendo el método proporcionar una matriz de sitios individualmente direccionables a una densidad de al menos aproximadamente 500 sitios por mm^2 , teniendo cada sitio un nanoporo unido a una ácido nucleico polimerasa, y, en un sitio dado de la matriz, polimerizar nucleótidos marcados con etiqueta con una polimerasa, en el que tras la polimerización se libera una etiqueta y se detecta mediante un nanoporo en el sitio dado. En algunos casos, el método comprende además dirigir generar, con la ayuda de un procesador, una secuencia de ácido nucleico de la molécula de ácido nucleico basándose en las etiquetas detectadas. En algunos casos, la etiqueta pasa a través del nanoporo. En algunos casos, la etiqueta pasa adyacente al nanoporo. En algunos casos, la tasa de polimerización es menor que la tasa de paso de etiqueta a través del, o adyacente al, nanoporo.
- 30

35 Métodos y sistemas para secuenciación de etiquetas

- Otro aspecto de la presente divulgación proporciona un sistema de medición de conductancia que comprende unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica. El sistema puede incluir además unos medios para aplicar un campo eléctrico a través de la barrera, unos medios para medir un cambio en el campo eléctrico, al menos una polimerasa unida al poro, y una o más enzimas fosfatasa unidas al poro.
- 40

En una realización del sistema, el poro tiene un diámetro de desde aproximadamente 1 hasta 10 nm . En otra realización, la polimerasa y las enzimas fosfatasa están unidas covalentemente al poro. En una realización adicional, más enzimas fosfatasa que polimerasas están unidas al poro.

- 45 En una realización del sistema, las enzimas fosfatasa están posicionadas de tal manera que el polifosfato producido mediante la polimerasa en una reacción de polimerasa interacciona con las enzimas fosfatasa antes de entrar en el poro.

En otra realización, la tasa de interacción entre las enzimas fosfatasa y el polifosfato es más rápida que, o igual a, la tasa de la polimerasa que produce el polifosfato.

- 50 En otra realización, cada uno de los compartimentos primero y segundo tiene una carga eléctrica. También se contempla que el interior del poro tenga una carga negativa.

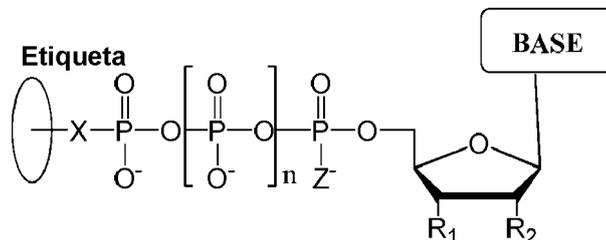
En aún otra realización del sistema, el poro es biológico o sintético. También se contempla que el poro sea proteico, por ejemplo en el que el poro es una proteína alfa-hemolisina.

En una realización adicional del sistema, el poro es un poro de estado sólido o grafeno.

- 55 También se contempla que el sistema comprenda una matriz de poros que tienen cada uno características sustancialmente idénticas, o una matriz de poros de diferentes diámetros, o una matriz de poros en la que se aplican campos eléctricos diferentes a través de la barrera.

Se contempla además que el sistema de medición de conductancia esté integrado con electrónica de CMOS, o que el poro o la matriz de poros esté integrado directamente en una pastilla de CMOS tal como se muestra en la figura 46.

Un compuesto que tiene la estructura:



5 en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alquenilo, un grupo alquinilo, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que la base es adenina, guanina, citosina, timina, uracilo o un derivado de una de esas bases, en la que n es 1, 2, 3 ó 4, y en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto.

10 En una realización del compuesto, la magnitud de la carga en la etiqueta es la misma que la magnitud de la carga en el resto del compuesto.

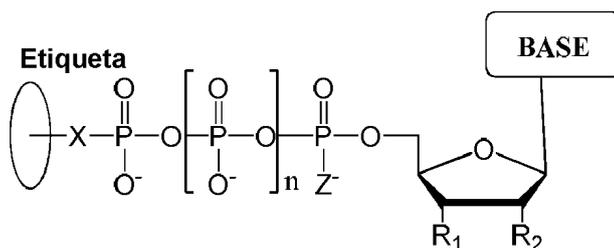
En otra realización del compuesto, la etiqueta que comprende múltiples unidades de etilenglicol, preferiblemente 16, 20, 24 ó 36 unidades de etilenglicol.

En una realización adicional del compuesto, la etiqueta comprende además un resto identificable adicional, tal como un colorante basado en cumarina.

15 En una realización, la etiqueta tiene una carga positiva. En otra realización, la retirada de la etiqueta cambia la carga del compuesto.

También se contempla que la etiqueta comprende además un número apropiado de lisinas o argininas para equilibrar el número de fosfatos en el compuesto.

Una composición que comprende cuatro tipos diferentes de un compuesto que tiene la estructura:



25 en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alquenilo, un grupo alquinilo, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que n es 1, 2, 3 ó 4, en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto, en la que la base de un primer tipo de compuesto es adenina o un derivado de la misma, la base de un segundo tipo de compuesto es guanina o un derivado de la misma, la base de un tercer tipo de compuesto es citosina o un derivado de la misma, y la base de un cuarto tipo de compuesto es timina o un derivado de la misma o uracilo o un derivado del mismo, y en la que la etiqueta en cada tipo de compuesto es diferente de la etiqueta en cada uno de los otros tres tipos de compuesto.

La invención se define en las reivindicaciones adjuntas. Aunque en el presente documento se han mostrado y descrito diversas realizaciones, resultará evidente para los expertos en la técnica que tales realizaciones sólo se proporcionan a modo de ejemplo.

40 Un método para la secuenciación de ácidos nucleicos, comprendiendo el método proporcionar una matriz de sitios

individualmente direccionables, teniendo cada sitio un nanoporo unido a una ácido nucleico polimerasa, y, en un sitio dado de dicha matriz, polimerizar nucleótidos marcados con etiqueta con una polimerasa, en el que se libera una etiqueta y se detecta mediante un nanoporo en dicho sitio dado.

5 En una realización, la etiqueta pasa a través del nanoporo. En otra realización, la tasa de polimerización es más lenta que la tasa de paso de etiqueta a través del nanoporo.

Un método para la secuenciación de ácidos nucleicos, comprendiendo el método: (a) polimerizar nucleótidos marcados con etiqueta a una primera tasa, en el que una etiqueta asociada con un nucleótido individual se libera tras la polimerización; y (b) detectar la etiqueta liberada haciéndola pasar a través de un nanoporo a una segunda tasa, en el que la segunda tasa es más rápida o igual que la primera tasa.

10 En una realización, cada tipo de nucleótido comprende una etiqueta única. En otra realización, la etiqueta está inicialmente unida al 5'-fosfato del nucleótido individual.

Un método para la secuenciación de ácidos nucleicos, comprendiendo el método: (a) polimerizar nucleótidos marcados con etiqueta, en el que una etiqueta asociada con un nucleótido individual se libera tras la polimerización; y (b) detectar la etiqueta liberada con la ayuda de un nanoporo.

15 En una realización, el método comprende además dirigir la etiqueta liberada a partir de un nucleótido individual a través del nanoporo. En otra realización, cada tipo de nucleótido comprende una etiqueta única. En una realización adicional, la etiqueta está inicialmente unida al 5'-fosfato del nucleótido individual.

20 Un método para la secuenciación de ácidos nucleicos, que comprende detectar, con la ayuda de un nanoporo, la incorporación de un nucleótido en una molécula de ácido nucleico, en el que dicha molécula de ácido nucleico no pasa a través del nanoporo.

En una realización, etiquetas asociadas con los nucleótidos se liberan tras la incorporación y las etiquetas pasan a través del nanoporo. En otra realización, se detectan acontecimientos de incorporación de nucleótido con una precisión de al menos 4σ .

25 Un método para la secuenciación de ácidos nucleicos, que comprende detectar un subproducto de un acontecimiento de incorporación de nucleótido individual con la ayuda de un nanoporo.

En una realización, el nucleótido no se detecta directamente. En otra realización, el subproducto del acontecimiento de incorporación de nucleótido es una molécula de etiqueta liberada. En una realización adicional, la molécula de etiqueta pasa a través del nanoporo.

30 Un método para la secuenciación de ácidos nucleicos, que comprende distinguir entre acontecimientos de incorporación de nucleótido individuales con una precisión de más de 4σ , 5σ o 6σ .

En una realización, los acontecimientos de incorporación de nucleótido se detectan con la ayuda de un nanoporo. En otra realización, etiquetas asociadas con los nucleótidos se liberan tras la incorporación y las etiquetas pasan a través del nanoporo.

35 Un método para la secuenciación de ácidos nucleicos, comprendiendo el método proporcionar una matriz de nanoporos unidos a una ácido nucleico polimerasa y polimerizar nucleótidos marcados con etiqueta con la polimerasa, en el que la etiqueta se libera y se detecta mediante el nanoporo.

En una realización, la etiqueta pasa a través del nanoporo. En otra realización, la tasa de polimerización es más lenta que la tasa de paso de etiqueta a través del nanoporo.

40 Un nucleótido marcado con etiqueta, en el que el nucleótido comprende una etiqueta que puede escindirse en un acontecimiento de polimerización de nucleótido y detectarse con la ayuda de un nanoporo.

En una realización, la etiqueta está unida al 5'-fosfato del nucleótido. En otra realización, la etiqueta no es un fluoróforo. En una realización adicional, la etiqueta puede detectarse mediante su carga, forma, tamaño o cualquier combinación de los mismos.

45 Un método para determinar la identidad de un compuesto que comprende: (a) poner en contacto el compuesto con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; (b) registrar el cambio en el campo eléctrico cuando el compuesto se transloca a través del poro en el que el cambio en el campo eléctrico es el resultado de interacción entre el compuesto, el electrolito y el poro, y es indicativo del tamaño, la carga y la composición del compuesto, permitiendo así la correlación entre el cambio y valores predeterminados para determinar la identidad del compuesto.

50

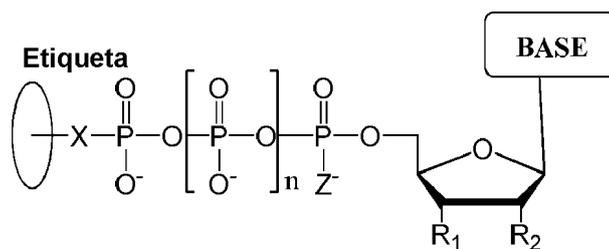
En una realización, el método comprende además una etapa de tratar el compuesto con una enzima fosfatasa antes de la etapa (a).

Un método para determinar si un compuesto es una etiqueta o un precursor de la etiqueta que comprende: (a) poner en contacto el compuesto con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; registrar el cambio en el campo eléctrico cuando el compuesto se transloca a través del poro; y comparar el cambio en el campo eléctrico con valores predeterminados correspondientes a la etiqueta y al precursor de la etiqueta, determinando así si el compuesto es la etiqueta o el precursor de la misma.

En una realización, el método comprende además una etapa de ajustar la polarización de corriente del campo eléctrico en la etapa (a).

Un método para determinar la secuencia de nucleótidos de un ADN monocatenario, comprendiendo el método:

(a) poner en contacto el ADN monocatenario con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; (iv) al menos una polimerasa unida al poro; y (v) más de una enzima fosfatasa unida al poro, y una composición que comprende cuatro tipos diferentes de un compuesto que tiene la estructura:



en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alquenilo, un grupo alquinilo, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que n es 1, 2, 3 ó 4, en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto, en la que la base de un primer tipo de compuesto es adenina o un derivado de la misma, la base de un segundo tipo de compuesto es guanina o un derivado de la misma, la base de un tercer tipo de compuesto es citosina o un derivado de la misma, y la base de un cuarto tipo de compuesto es timina o un derivado de la misma, y en la que la etiqueta en cada tipo de compuesto es diferente de la etiqueta en cada uno de los otros tres tipos de compuesto, en el que el ADN monocatenario está en una disolución de electrolito en contacto con la polimerasa unida al poro y en el que el ADN monocatenario tiene un cebador hibridado a una porción del mismo, en condiciones que permiten que la polimerasa catalice la incorporación de uno de los compuestos en el cebador si el compuesto es complementario al residuo de nucleótido del ADN monocatenario inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ADN, en el que la incorporación del compuesto da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo, en el que la enzima fosfatasa unida al poro escinde la etiqueta a partir del polifosfato para liberar la etiqueta;

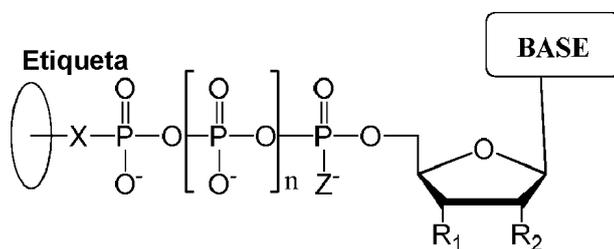
(b) determinar qué compuesto se ha incorporado en el cebador para formar el producto de extensión de ADN en la etapa (a) aplicando un campo eléctrico a través de la barrera y midiendo un cambio electrónico a través del poro resultante de que la etiqueta generada en la etapa (a) se transloca a través del poro, en el que el cambio electrónico es diferente para cada tipo de etiqueta, identificando así el residuo de nucleótido en el ADN monocatenario complementario al compuesto incorporado; y

(c) realizar de manera repetida la etapa (b) para cada residuo de nucleótido del ADN monocatenario que está secuenciándose, determinando así la secuencia de nucleótidos del ADN monocatenario.

Un método para determinar la secuencia de nucleótidos de un ADN monocatenario, comprendiendo el método:

(a) poner en contacto el ADN monocatenario con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; (iv) al

menos una polimerasa unida al poro; y (v) más de una enzima fosfatasa unida al poro, y un compuesto que tiene la estructura:



5 en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alquenoilo, un grupo alquinilo, un grupo azido o una combinación de los mismos, en la que R₁ es OH, en la que R₂ es H u OH, en la que X es O, NH, S o CH₂, en la que Z es O, S o BH₃, en la que la base es adenina, guanina, citosina, timina o un derivado de una de esas bases, en la que n es 1, 2, 3 ó 4, y en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto,

15 en el que el ADN monocatenario está en una disolución de electrolito en contacto con la polimerasa unida al poro y en el que el ADN monocatenario tiene un cebador hibridado a una porción del mismo, en condiciones que permiten que la polimerasa catalice la incorporación del compuesto en el cebador si es complementario al residuo de nucleótido del ADN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ADN,

20 en el que si no se incorpora el compuesto, repetir de manera iterativa la puesta en contacto con diferentes compuestos hasta que se incorpora un compuesto, con la condición de que (1) el tipo de base en el compuesto es diferente del tipo de base en cada uno de los compuestos anteriores, y (2) el tipo de etiqueta en el compuesto es diferente del tipo de etiqueta en cada uno de los compuestos anteriores,

en el que la incorporación del compuesto da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo,

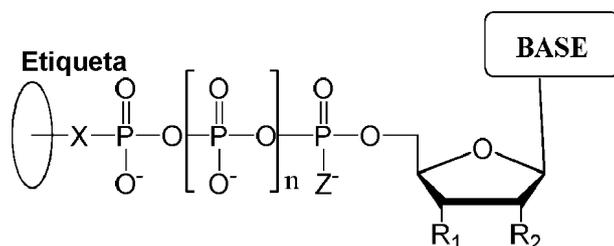
en el que la enzima fosfatasa unida al poro escinde la etiqueta a partir del polifosfato para liberar la etiqueta;

25 (b) determinar qué compuesto se ha incorporado en el cebador para formar el producto de extensión de ADN en la etapa (a) aplicando un campo eléctrico a través de la barrera y midiendo un cambio electrónico a través del poro resultante de que la etiqueta generada en la etapa (a) se transloca a través del poro, en el que el cambio electrónico es diferente para cada tipo de etiqueta, identificando así el residuo de nucleótido en el ADN monocatenario complementario al compuesto incorporado; y

30 (c) realizar de manera iterativa las etapas (a) y (b) para cada residuo de nucleótido del ADN monocatenario que está secuenciándose, determinando así la secuencia de nucleótidos del ADN monocatenario.

Un método para determinar la secuencia de nucleótidos de un ARN monocatenario, comprendiendo el método:

35 (a) poner en contacto el ARN monocatenario con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; (iv) al menos una polimerasa unida al poro; y (v) más de una enzima fosfatasa unida al poro, y una composición que comprende cuatro tipos diferentes de un compuesto que tiene la estructura:



40 en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un

5 colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alquenilo, un grupo alquinilo, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que n es 1, 2, 3 ó 4, en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto, en la que la base de un primer tipo de compuesto es adenina o un derivado de la misma, la base de un segundo tipo de compuesto es guanina o un derivado de la misma, la base de un tercer tipo de compuesto es citosina o un derivado de la misma, y la base de un cuarto tipo del compuesto es uracilo o un derivado del mismo, y en la que la etiqueta en cada tipo de compuesto es diferente de la etiqueta en cada uno de los otros tres tipos de compuesto,

10 en el que el ARN monocatenario está en una disolución de electrolito en contacto con la polimerasa unida al poro y en el que el ARN monocatenario tiene un cebador hibridado a una porción del mismo, en condiciones que permiten que la polimerasa catalice la incorporación de uno de los compuestos en el cebador si el compuesto es complementario al residuo de nucleótido del ARN monocatenario inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ARN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ARN,

15 en el que la incorporación del compuesto da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo,

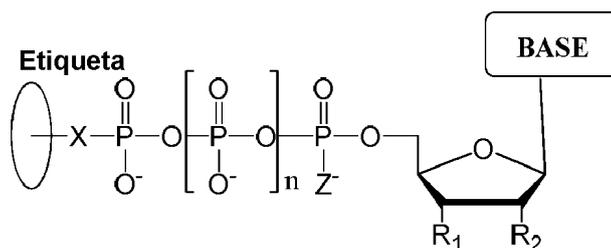
en el que la enzima fosfatasa unida al poro escinde la etiqueta a partir del polifosfato para liberar la etiqueta;

20 (b) determinar qué compuesto se ha incorporado en el cebador para formar el producto de extensión de ARN en la etapa (a) aplicando un campo eléctrico a través de la barrera y midiendo un cambio electrónico a través del poro resultante de que la etiqueta generada en la etapa (a) se transloca a través del poro, en el que el cambio electrónico es diferente para cada tipo de etiqueta, identificando así el residuo de nucleótido en el ARN monocatenario complementario al compuesto incorporado; y

25 (c) realizar de manera repetida la etapa (b) para cada residuo de nucleótido del ARN monocatenario que está secuenciándose, determinando así la secuencia de nucleótidos del ARN monocatenario.

Un método para determinar la secuencia de nucleótidos de un ARN monocatenario, comprendiendo el método:

30 (a) poner en contacto el ARN monocatenario con un sistema de medición de conductancia que comprende: (i) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (ii) unos medios para aplicar un campo eléctrico a través de la barrera; (iii) unos medios para medir un cambio en el campo eléctrico; (iv) al menos una polimerasa unida al poro; y (v) más de una enzima fosfatasa unida al poro, y un compuesto que tiene la estructura:



35 en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alquenilo, un grupo alquinilo, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que la base es adenina, guanina, citosina, uracilo o un derivado de una de esas bases, en la que n es 1, 2, 3 ó 4, y en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto,

40 en el que el ARN monocatenario está en una disolución de electrolito en contacto con la polimerasa unida al poro y en el que el ARN monocatenario tiene un cebador hibridado a una porción del mismo, en condiciones que permiten que la polimerasa catalice la incorporación del compuesto en el cebador si es complementario al residuo de nucleótido del ARN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ARN,

45 en el que si no se incorpora el compuesto, repetir de manera iterativa la puesta en contacto con diferentes

compuestos hasta que se incorpora un compuesto, con la condición de que (1) el tipo de base en el compuesto es diferente del tipo de base en cada uno de los compuestos anteriores, y (2) el tipo de etiqueta en el compuesto es diferente del tipo de etiqueta en cada uno de los compuestos anteriores,

5 en el que la incorporación del compuesto da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo,

en el que la enzima fosfatasa unida al poro escinde la etiqueta a partir del polifosfato para liberar la etiqueta;

10 (b) determinar qué compuesto se ha incorporado en el cebador para formar el producto de extensión de ARN en la etapa (a) aplicando un campo eléctrico a través de la barrera y midiendo un cambio electrónico a través del poro resultante de que la etiqueta generada en la etapa (a) se transloca a través del poro, en el que el cambio electrónico es diferente para cada tipo de etiqueta, identificando así el residuo de nucleótido en el ARN monocatenario complementario al compuesto incorporado; y

(c) realizar de manera iterativa las etapas (a) y (b) para cada residuo de nucleótido del ARN monocatenario que está secuenciándose, determinando así la secuencia de nucleótidos del ARN monocatenario.

15 En una realización del método, más enzimas fosfatasa que polimerasas están unidas al poro. En una realización, el ARN o ADN monocatenario se obtiene desnaturalizando un ARN o ADN bicatenario, lo que sea aplicable. En otra realización, múltiples copias del mismo ARN o ADN monocatenario están inmovilizadas en una perla. También se contempla que la secuencia de nucleótidos del ARN o ADN monocatenario se determine usando múltiples copias del mismo ARN o ADN monocatenario.

20 En otra realización, se realiza una etapa de lavado tras cada iteración de la etapa (b) para retirar compuesto no incorporado del contacto con el ARN o ADN monocatenario. En una realización adicional, se contempla una etapa tras cada iteración de la etapa (b) para determinar la identidad de un resto identificable adicional unido a la etiqueta.

En una realización del método, al menos el 85%, el 90%, el 95% o el 99% de las etiquetas liberadas se translocan a través del nanoporo.

25 En una realización, el compuesto comprende además un terminador reversible, opcionalmente, el método comprende además una etapa de retirar el terminador reversible tras cada iteración de la etapa (b), en el que el terminador reversible se retira mediante medios biológicos, medios químicos, medios físicos o mediante irradiación de luz.

En otra realización, el interior del poro tiene una carga que es de signo inverso con respecto a la carga de la etiqueta o del polifosfato que tiene la etiqueta unida al mismo.

30 En aún otra realización, cada uno de los compartimentos primero y segundo del sistema de medición de conductancia tiene una carga, opcionalmente, las cargas de los compartimentos primero y segundo son de polaridad opuesta. También se contempla que las cargas de los compartimentos primero y segundo sean ajustables.

En una realización adicional, la tasa de translocación de etiqueta a través del poro en la etapa (b) se determina basándose en la carga de etiqueta y las cargas de los compartimentos primero y segundo.

35 En aún una realización adicional, cada uno de los compartimentos primero y segundo tiene una carga de tal manera que en la etapa (b) la etiqueta se transloca a través del poro a una tasa que es más rápida o igual que la tasa a la que la etiqueta o el polifosfato que tiene la etiqueta unida al mismo se libera en la etapa (a).

Un sistema de medición de conductancia que comprende:

una barrera eléctricamente resistiva que separa al menos unas disoluciones de electrolito primera y segunda;

40 dicha barrera eléctricamente resistiva comprende al menos un poro con un diámetro a escala nanométrica;

al menos un compuesto con una etiqueta en al menos una de dichas disoluciones de electrolito primera y segunda;

estando dicho al menos un poro configurado para permitir que se impulse una corriente iónica a través de dichas disoluciones de electrolito primera y segunda mediante un potencial aplicado;

45 comprendiendo dicho al menos un poro una característica configurada para escindir la etiqueta a partir del compuesto para liberar la etiqueta; y

unos medios de medición de la corriente iónica y unos medios de registro de su transcurso en el tiempo como serie temporal, incluyendo periodos de tiempo en los que el al menos un poro no está obstruido por la etiqueta y también periodos de tiempo en los que la etiqueta provoca pulsos de conductancia reducida.

50 En una realización del sistema, la etiqueta tiene un tiempo de residencia en el poro que es mayor que las limitaciones de ancho de banda de corriente iónica y el ruido de disparo de corriente de dichos medios de medición

de la corriente iónica.

- 5 Un método para delinear segmentos de una serie temporal de conductancia para dar regiones estadísticamente compatibles con el nivel de conductancia de poro no obstruido, y pulsos de conductancia reducida, y también segmentos estadísticamente estacionarios dentro de pulsos de conductancia reducida individuales, generándose dicha serie temporal de conductancia con un sistema de medición de conductancia que comprende:
- una barrera eléctricamente resistiva que separa al menos unas disoluciones de electrolito primera y segunda;
- dicha barrera eléctricamente resistiva comprende al menos un poro con un diámetro a escala nanométrica;
- al menos un compuesto con una etiqueta en al menos una de dichas disoluciones de electrolito primera y segunda;
- 10 estando dicho al menos un poro configurado para permitir que se impulse una corriente iónica a través de dichas disoluciones de electrolito primera y segunda mediante un potencial aplicado;
- comprendiendo dicho al menos un poro una característica configurada para escindir la etiqueta a partir del compuesto para liberar la etiqueta; y
- 15 unos medios de medición de la corriente iónica y unos medios de registro de dicha serie temporal de conductancia, incluyendo periodos de tiempo en los que el al menos un poro no está obstruido por dicha etiqueta y también periodos de tiempo en los que dicha etiqueta provoca pulsos de conductancia reducida;
- seleccionándose dicho método para delinear segmentos de una serie temporal de conductancia del grupo que consiste en:
- (a) una decodificación de Viterbi de la secuencia de estados de probabilidad máxima de una densidad continua de un modelo de Markov oculto estimada a partir de la serie temporal de conductancia sin procesar;
- 20 (b) una delineación de las regiones de pulsos de conductancia reducida mediante comparación con un umbral para la desviación con respecto al nivel de conductancia de poro abierto; y
- (c) unos medios para caracterizar pulsos de conductancia reducida estimando las tendencias centrales de los niveles de corriente iónica para cada segmento, o mediante medición de tendencias centrales y duración de segmento en conjunto, seleccionándose la medición de tendencia central de segmento del grupo que consiste en: (i) un parámetro medio de una componente gaussiana de un primer GMM estimado a partir de la serie temporal de conductancia como parte de un modelo de Markov oculto de densidad continua; (ii) una media aritmética; (iii) una media recortada; (iv) una mediana; y (v) un estimador de máximo a posteriori de ubicación de muestra, o un estimador de probabilidad máxima de ubicación de muestra.
- 25 En otra realización, el método comprende además al menos uno: (a) una estimación de probabilidad máxima de un segundo modelo mixto gaussiano basándose en las medidas de tendencia central de segmentos de conductancia; (b) un hallazgo de picos por medio de interpolación y suavizado de la densidad de probabilidad empírica de las estimaciones de tendencias centrales de segmentos de la serie temporal de conductancia y hallazgo de raíces de las derivadas de las funciones de interpolación; y (c) otros medios de localización de los modos de estimador de distribución multimodal.
- 30 Un método para determinar al menos un parámetro de un compuesto en una disolución que comprende las etapas de:
- colocar un primer fluido en un primer depósito;
- colocar un segundo fluido en un segundo depósito; comprendiendo al menos uno de dichos fluidos primero y segundo al menos un compuesto, en el que el compuesto es un nucleótido marcado con etiqueta o una etiqueta escindida a partir de un nucleótido marcado con etiqueta; estando dicho primer fluido en dicho primer depósito separado de dicho segundo fluido en dicho segundo depósito con una barrera eléctricamente resistiva; comprendiendo dicha barrera eléctricamente resistiva al menos un poro;
- 40 hacer pasar una corriente iónica a través de dicho primer fluido, dicho al menos un poro, y dicho segundo fluido con un potencial eléctrico entre dichos fluidos primero y segundo;
- 45 medir la corriente iónica que pasa a través de dicho al menos un poro y la duración de cambios en la corriente iónica; llevándose a cabo la medición de la corriente iónica durante un periodo de tiempo suficiente para medir una reducción en la corriente iónica provocada por el compuesto que interacciona con dicho al menos un poro; y
- determinar al menos un parámetro del compuesto analizando matemáticamente los cambios en la corriente iónica y la duración de los cambios en la corriente iónica a lo largo del periodo de tiempo; comprendiendo dicho análisis matemático al menos una etapa seleccionada del grupo que consiste en: (a) un parámetro medio de una componente gaussiana de un primer GMM estimado a partir de la serie temporal de conductancia como parte de un
- 50

modelo de Markov oculto de densidad continua; (b) una extracción de media de acontecimientos; (c) asignación de estado de acontecimiento de probabilidad máxima; (d) detección y cálculo de promedio de umbral; (e) análisis de ventana deslizante; (f) una media aritmética; (g) una media recortada; (h) una mediana; y (i) un estimador de máximo a posteriori de ubicación de muestra, o un estimador de probabilidad máxima de ubicación de muestra.

- 5 En una realización del método, el compuesto se trata con fosfatasa antes de medir la reducción en la corriente iónica. En otra realización, el compuesto es un compuesto cargado de manera alternativa que tiene una primera carga neta y, después de una reacción química, física o biológica, una segunda carga neta diferente.

En otra realización, el análisis matemático se selecciona del grupo que consiste en GMM, detección y cálculo de promedio de umbral, y análisis de ventana deslizante.

- 10 En una realización adicional, el al menos un parámetro se selecciona del grupo que consiste en la concentración, el tamaño, la carga y la composición del compuesto.

Se contempla que una realización del método comprenda una etapa de calibrar el sistema de medición de conductancia.

En una realización, la precisión del método es mayor de 4σ , 5σ o 6σ .

- 15 Un nucleótido marcado con etiqueta, en el que el nucleótido comprende una etiqueta que puede escindirse en un acontecimiento de polimerización de nucleótido y detectarse con la ayuda de un nanoporo.

En una realización, la etiqueta está unida al 5'-fosfato del nucleótido. En otra realización, la etiqueta no es un fluoróforo. En una realización adicional, la etiqueta puede detectarse mediante su carga, forma, tamaño o cualquier combinación de los mismos.

- 20 En una realización del sistema de medición de conductancia, las disoluciones de electrolito primera y segunda son la misma.

En una realización del método, las disoluciones de electrolito primera y segunda son la misma.

Un nucleótido marcado con etiqueta, en el que el nucleótido comprende una etiqueta que puede escindirse en un acontecimiento de polimerización de nucleótido y detectarse con la ayuda de un nanoporo.

- 25 En una realización, la etiqueta está unida al 5'-fosfato del nucleótido. En otra realización, la etiqueta no es un fluoróforo. En una realización adicional, la etiqueta puede detectarse mediante su carga, forma, tamaño o cualquier combinación de los mismos.

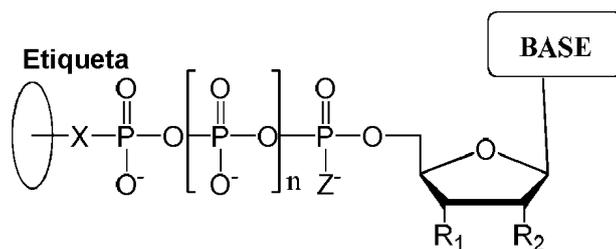
Un método para la secuenciación de ácidos nucleicos, comprendiendo el método proporcionar una matriz de sitios individualmente direccionables, teniendo cada sitio un nanoporo unido a una ácido nucleico polimerasa, y, en un sitio dado de dicha matriz, polimerizar nucleótidos marcados con etiqueta con una polimerasa, en el que se libera una etiqueta y se detecta mediante un nanoporo en dicho sitio dado.

- 30

Un método para determinar la secuencia de nucleótidos de un ADN monocatenario que comprende:

(a) poner en contacto el ADN monocatenario, en el que el ADN monocatenario está en una disolución de electrolito en contacto con un nanoporo en una membrana y en el que el ADN monocatenario tiene un cebador hibridado a una porción del mismo, con una ADN polimerasa y cuatro análogos de polifosfato de desoxirribonucleótido (dNPP) al menos uno de los cuales puede hibridarse a cada uno de un nucleótido de A, T, G o C en el ADN que está secuenciándose en condiciones que permiten que la ADN polimerasa catalice la incorporación de uno de los análogos de dNPP en el cebador si es complementario al residuo de nucleótido del ADN monocatenario que está inmediateamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ADN, en el que cada uno de los cuatro análogos de dNPP tiene la estructura:

- 35
- 40



en la que la base es adenina, guanina, citosina, timina o uracilo o un derivado de una o más de estas bases, en la que R_1 es OH, en la que R_2 es H, en la que X es O, NH, S o CH_2 , en la que n es 1, 2, 3 ó 4, en la que Z es O, S o BH_3 , y

- 45

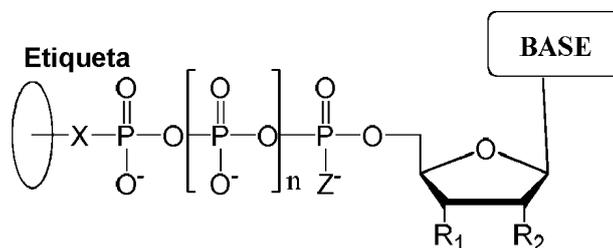
con la condición de que (i) el tipo de base en cada análogo de dNPP es diferente del tipo de base en cada uno de los otros tres análogos de dNPP, y (ii) o bien el valor de n de cada análogo de dNPP es diferente del valor de n de cada uno de los otros tres análogos de dNPP, o bien el valor de n de cada uno de los cuatro análogos de dNPP es el mismo y el tipo de etiqueta en cada análogo de dNPP es diferente del tipo de etiqueta en cada uno de los otros tres análogos de dNPP, en el que la incorporación del análogo de dNPP da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo; y

(b) identificar qué análogo de dNPP se ha incorporado en el cebador para formar el producto de extensión de ADN en la etapa (a) aplicando una tensión a través de la membrana y midiendo un cambio electrónico a través del nanoporo resultante de que el polifosfato que tiene la etiqueta unida al mismo generado en la etapa (a) se transloca a través del nanoporo, en el que el cambio electrónico es diferente para cada valor de n, o para cada tipo de etiqueta diferente, lo que sea aplicable, permitiendo así identificar el residuo de nucleótido en el ADN monocatenario complementario al análogo de dNPP incorporado; y

(c) realizar de manera repetida la etapa (b) para cada residuo de nucleótido del ADN monocatenario que está secuenciándose, en el que en cada iteración de la etapa (b) se identifica qué análogo de dNPP se ha incorporado en el producto de extensión de ADN en la etapa (a), en el que el análogo de dNPP se ubica inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del producto de extensión de ADN, determinando así la secuencia de nucleótidos del ADN monocatenario.

Un método para determinar la secuencia de nucleótidos de un ADN monocatenario que comprende:

(a) poner en contacto el ADN monocatenario, en el que el ADN monocatenario está en una disolución de electrolito en contacto con un nanoporo en una membrana y en el que el ADN monocatenario tiene un cebador hibridado a una porción del mismo, con una ADN polimerasa y un análogo de polifosfato de desoxirribonucleótido (dNPP) que puede hibridarse a un nucleótido de A, T, G o C en el ADN que está secuenciándose en condiciones que permiten que la ADN polimerasa catalice la incorporación del análogo de dNPP en el cebador si es complementario al residuo de nucleótido del ADN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ADN, en el que el análogo de dNPP tiene la estructura:



en la que la base es adenina, guanina, citosina, uracilo o timina o un derivado de una de esas bases, en la que R₁ es -OH, -O-CH₂N₃ u -O-2-nitrobencilo, en la que R₂ es H, en la que X es O, NH, S o CH₂, en la que n es 1, 2, 3 ó 4, en la que Z es O, S o BH₃, y

en el que si no se incorpora el análogo de dNPP, repetir de manera iterativa la puesta en contacto con un análogo de dNPP diferente hasta que se incorpora un análogo de dNPP, con la condición de que (i) el tipo de base en cada análogo de dNPP es diferente del tipo de base en cada uno de los otros análogos de dNPP, y (ii) o bien el valor de n de cada análogo de dNPP es diferente del valor de n de cada uno de los otros análogos de dNPP, o bien el valor de n de cada uno de los análogos de dNPP es el mismo y el tipo de etiqueta en cada análogo de dNPP es diferente del tipo de etiqueta en cada uno de los otros análogos de dNPP, en el que la incorporación del análogo de dNPP da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo;

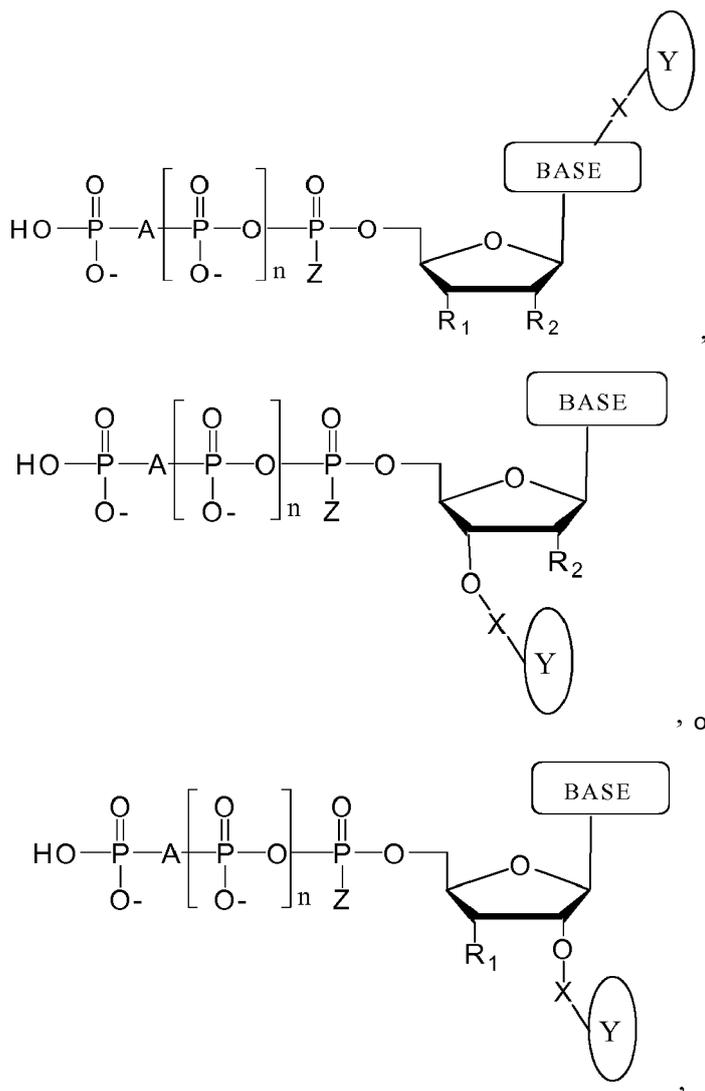
(b) identificar qué análogo de dNPP se ha incorporado en el cebador para formar el producto de extensión de ADN en la etapa (a) aplicando una tensión a través de la membrana y midiendo un cambio electrónico a través del nanoporo resultante de que el polifosfato que tiene la etiqueta unida al mismo generado en la etapa (a) se transloca a través del nanoporo, en el que el cambio electrónico es diferente para cada valor de n, o para cada tipo de etiqueta diferente, lo que sea aplicable, permitiendo así identificar el residuo de nucleótido en el ADN monocatenario complementario al análogo de dNPP incorporado;

(c) realizar de manera repetida las etapas (a) y (b) para cada residuo de nucleótido del ADN monocatenario que está secuenciándose, en el que en cada iteración de la etapa (a) el análogo de dNPP se incorpora en el producto de extensión de ADN si es complementario al residuo de nucleótido del ADN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del producto de extensión de ADN, determinando así la secuencia de nucleótidos del ADN monocatenario.

Un método para determinar la secuencia de nucleótidos de un ADN monocatenario que comprende:

- 5 (a) poner en contacto el ADN monocatenario, en el que el ADN monocatenario está en una disolución de electrolito en contacto con un nanoporo en una membrana y en el que el ADN monocatenario tiene un cebador hibridado a una porción del mismo, con una ADN polimerasa y cuatro análogos de polifosfato de desoxirribonucleótido (dNPP) al menos uno de los cuales puede hibridarse a cada uno de un nucleótido de A, T, G o C en el ADN que está secuenciándose en condiciones que permiten que la ADN polimerasa catalice la incorporación de uno de los análogos de dNPP en el cebador si es complementario al residuo de nucleótido del ADN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ADN, en el que cada uno de los cuatro análogos de dNPP tiene una estructura elegida de las siguientes:

10



15

en la que la base es adenina, guanina, citosina, uracilo o timina o un derivado de una o más de estas bases, en el que Y es una etiqueta, en el que R₁, si está presente, es OH, en el que R₂, si está presente, es H, en la que X es un ligador escindible, en la que Z es O, S o BH₃, en la que n es 1, 2, 3 ó 4, en el que A es O, S, CH₂, CHF, CFF o NH, y

20

con la condición de que (i) el tipo de base en cada análogo de dNPP es diferente del tipo de base en cada uno de los otros tres análogos de dNPP, y (ii) el tipo de etiqueta en cada análogo de dNPP es diferente del tipo de etiqueta en cada uno de los otros tres análogos de dNPP;

(b) escindir la etiqueta a partir del análogo de dNPP incorporado en la etapa (a);

(c) identificar qué análogo de dNPP se ha incorporado en el cebador para formar el producto de extensión de ADN en la etapa (a) aplicando una tensión a través de la membrana y midiendo un cambio electrónico a través del nanoporo resultante de que la etiqueta escindida en la etapa (b) se transloca a través del nanoporo, en el que el cambio electrónico es diferente para cada tipo de etiqueta diferente, permitiendo así identificar el residuo de nucleótido en el ADN monocatenario complementario al análogo de dNPP incorporado; y

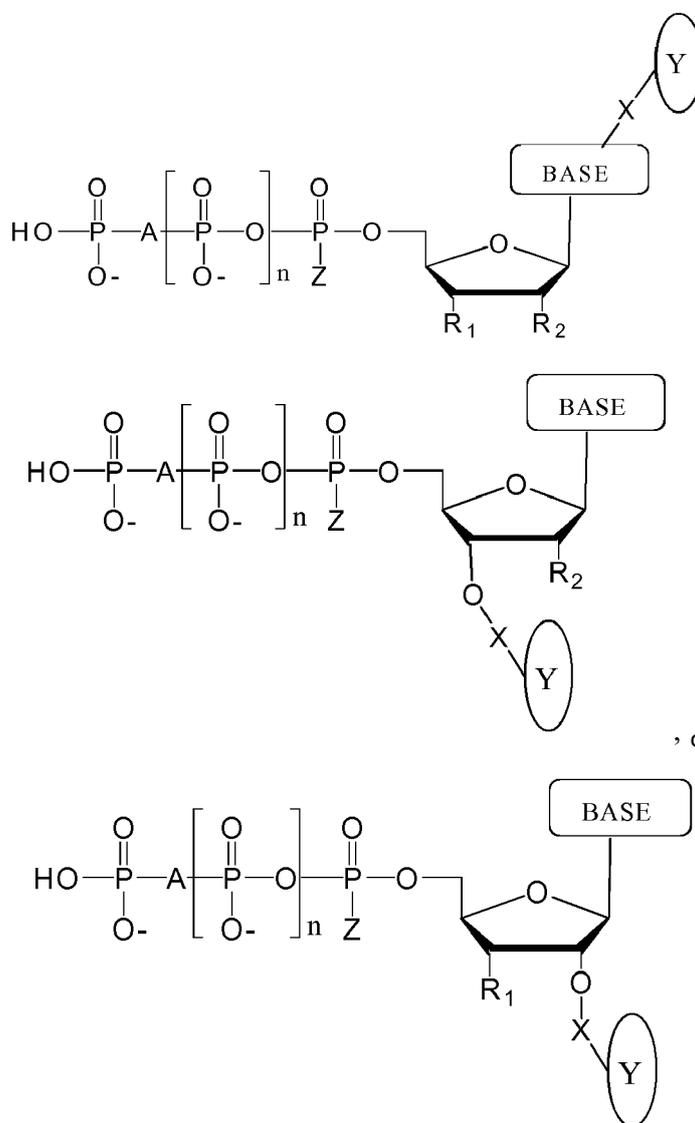
(d) realizar de manera repetida las etapas (b) y (c) para cada residuo de nucleótido del ADN monocatenario que está secuenciándose, en el que en cada iteración de la etapa (c) se identifica qué análogo de dNPP se ha incorporado en el producto de extensión de ADN en la etapa (a) inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del producto de extensión de ADN, determinando así la secuencia de nucleótidos del ADN monocatenario.

5

Un método para determinar la secuencia de nucleótidos de un ADN monocatenario que comprende:

(a) poner en contacto el ADN monocatenario, en el que el ADN monocatenario está en una disolución de electrolito en contacto con un nanoporo en una membrana y en el que el ADN monocatenario tiene un cebador hibridado a una porción del mismo, con una ADN polimerasa y un análogo de polifosfato de desoxirribonucleótido (dNPP) que puede hibridarse a un nucleótido de A, T, G o C en el ADN que está secuenciándose en condiciones que permiten que la ADN polimerasa catalice la incorporación del análogo de dNPP en el cebador si es complementario al residuo de nucleótido del ADN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ADN, en el que el análogo de dNPP tiene la estructura:

10



15

en la que la base es adenina, guanina, citosina, uracilo o timina, o derivado de una de esas bases, en el que Y es una etiqueta, y en el que R₁, si está presente, es OH, -OCH₂N₃, u -O-2-nitrobencilo, en el que R₂, si está presente, es H, en la que X es un ligador escindible, en la que Z es O, S o BH₃, en la que n es 1, 2, 3 ó 4, en el que A es O, S, CH₂, CHF, CFF o NH, y

20

en el que si no se incorpora el análogo de dNPP, repetir de manera iterativa la puesta en contacto con un análogo de dNPP diferente hasta que se incorpora un análogo de dNPP,

con la condición de que (i) el tipo de base en cada análogo de dNPP es diferente del tipo de base en cada uno de los otros análogos de dNPP, y (ii) el tipo de etiqueta en cada análogo de dNPP es diferente del tipo de etiqueta en cada uno de los otros análogos de dNPP;

(b) escindir la etiqueta a partir del análogo de dNPP incorporado en la etapa (a); y

5 (c) identificar qué análogo de dNPP se ha incorporado en el cebador para formar el producto de extensión de ADN en la etapa (a) aplicando una tensión a través de la membrana y midiendo un cambio electrónico a través del nanoporo resultante de que la etiqueta escindida en la etapa (b) se transloca a través del nanoporo, en el que el cambio electrónico es diferente para cada tipo de etiqueta, permitiendo así identificar el residuo de nucleótido en el ADN monocatenario complementario al análogo de dNPP incorporado;

10 (d) realizar de manera repetida las etapas (a) a (c) para cada residuo de nucleótido del ADN monocatenario que está secuenciándose, en el que en cada iteración de la etapa (a) el análogo de dNPP se incorpora en el producto de extensión de ADN si es complementario al residuo de nucleótido del ADN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario hibridado al residuo de nucleótido 3'-terminal del producto de extensión de ADN, determinando así la secuencia de nucleótidos del ADN monocatenario.

15 En algunos casos un nucleótido marcado con etiqueta comprende una etiqueta que puede escindirse en un acontecimiento de polimerización de nucleótido y detectarse con la ayuda de un nanoporo. La etiqueta puede estar unida al 5'-fosfato del nucleótido. En algunos casos, la etiqueta no es un fluoróforo. La etiqueta puede ser detectable mediante su carga, forma, tamaño o cualquier combinación de los mismos. Las etiquetas a modo de ejemplo incluyen diversos polímeros. Cada tipo de nucleótido (es decir, A, C, G, T) comprende generalmente una etiqueta
20 única.

Las etiquetas pueden estar ubicadas en cualquier posición adecuada en el nucleótido. La figura 34 proporciona algunos ejemplos no limitativos de un nucleótido marcado con etiqueta. En el primer diagrama, R₁ es generalmente OH y R₂ es H (es decir, para ADN) u OH (es decir, para ARN), aunque otras modificaciones son aceptables. En la figura 34, X es cualquier ligador adecuado. En algunos casos, el ligador puede escindirse. Los ejemplos de ligadores
25 incluyen, sin limitación, O, NH, S o CH₂. Los ejemplos de grupos químicos adecuados para la posición Z incluyen O, S o BH₃. La base es cualquier base adecuada para su incorporación en un ácido nucleico incluyendo adenina, guanina, citosina, timina, uracilo o un derivado de las mismas. En algunos casos también son aceptables bases universales.

El número de fosfatos (n) es cualquier valor de número entero adecuado (por ejemplo, un número de fosfatos de tal manera que el nucleótido puede incorporarse en una molécula de ácido nucleico). En algunos casos, todos los tipos de nucleótidos marcados con etiqueta tienen el mismo número de fosfatos, pero esto no se requiere. En algunas aplicaciones, hay una etiqueta diferente para cada tipo de nucleótido y el número de fosfatos no se usa necesariamente para distinguir las diversas etiquetas. Sin embargo, en algunos casos más de un tipo de nucleótido (por ejemplo, A, C, T, G o U) tienen la misma molécula de etiqueta y la capacidad para distinguir un nucleótido de otro está determinada al menos en parte por el número de fosfatos (teniendo diversos tipos de nucleótidos un valor diferente para n). En diversas realizaciones, el valor para n es 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o mayor.

A continuación se describen etiquetas adecuadas. En algunos casos, la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto. Cuando se une la etiqueta, la carga en el compuesto global puede ser neutra. La liberación de la etiqueta puede dar como resultado dos moléculas, una etiqueta cargada y un nucleótido cargado. En algunos casos la etiqueta cargada pasa a través de un nanoporo y se detecta.
40

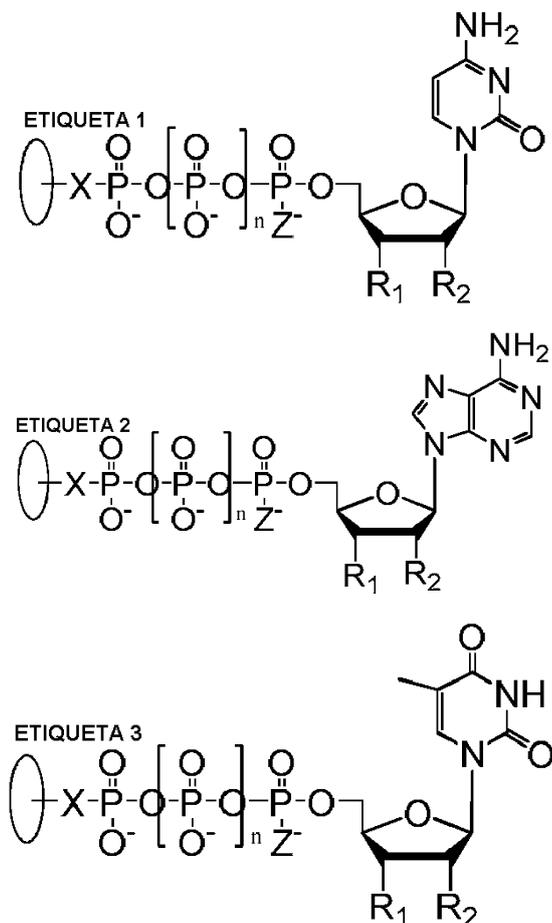
En la figura 34 se muestran ejemplos adicionales nucleótidos marcados con etiqueta adecuados. Tal como se muestra en los diagramas segundo a cuarto, la etiqueta también puede estar unida a la molécula de azúcar, la molécula de base o cualquier combinación de las mismas. Con referencia a la figura 34, Y es una etiqueta y X es un ligador escindible. Además, R₁, si está presente, es generalmente OH, -OCH₂N₃ u -O-2-nitrobencilo, y R₂, si está presente, es generalmente H. Además, Z es generalmente O, S o BH₃, y n es cualquier número entero incluyendo 1, 2, 3 ó 4. En algunos casos, A es O, S, CH₂, CHF, CFF o NH.
45

Siguiendo haciendo referencia a la figura 34, el tipo de base en cada análogo de dNPP es generalmente diferente del tipo de base en cada uno de los otros tres análogos de dNPP, y el tipo de etiqueta en cada análogo de dNPP es generalmente diferente del tipo de etiqueta en cada uno de los otros tres análogos de dNPP. Las bases adecuadas incluyen, pero no se limitan a, adenina, guanina, citosina, uracilo o timina o un derivado de cada una de las mismas. En algunos casos, la base es una de 7-desazaguanina, 7-desaza-adenina o 5-metilcitosina.
50

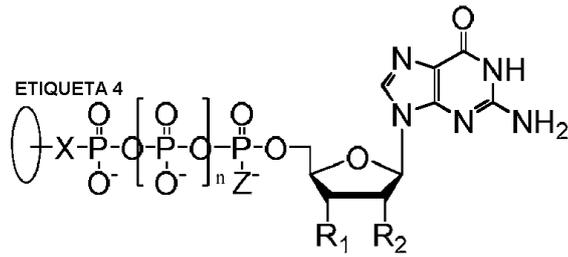
En casos en los que R₁ es -O-CH₂N₃, los métodos comprenden opcionalmente además tratar el análogo de dNPP incorporado para retirar el -CH₂N₃ y dar como resultado un grupo -OH unido a la posición 3' permitiendo así la incorporación de un análogo de dNPP adicional.

55 En casos en los que R₁ es -O-2-nitrobencilo, los métodos comprenden opcionalmente además tratar el análogo de nucleótido incorporado para retirar el -2-nitrobencilo y dar como resultado un grupo -OH unido a la posición 3' permitiendo así la incorporación de un análogo de dNPP adicional

- Una etiqueta puede ser cualquier grupo químico o molécula que puede detectarse en un nanoporo. En una realización de los métodos la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un colorante, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un colorante fluorescente, un compuesto quimioluminiscente, un aminoácido, un péptido, un hidrato de carbono, un monofosfato de nucleótido, un difosfato de nucleótido, un ácido alifático, un ácido aromático, un alcohol, un tiol no sustituido o sustituido con uno o más halógenos, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alqueno, un grupo alquino, un grupo azido o una combinación de los mismos.
- 5 En una realización de los métodos la base se selecciona del grupo que consiste en adenina, guanina, citosina, timina, 7-desazaguanina, 7-desaza-adenina o 5-metilcitosina.
- 10 En una realización los métodos comprenden además una etapa de lavado tras cada iteración de la etapa (b) para retirar análogos de dNPP no incorporados del contacto con el ADN monocatenario.
- En una realización los métodos comprenden además una etapa de lavado tras cada iteración de la etapa (c) para retirar análogos de dNPP no incorporados del contacto con el ADN monocatenario.
- 15 En una realización los métodos el ADN monocatenario, disolución de electrolito y nanoporo en la membrana están ubicados dentro de un único recipiente.
- En una realización de los métodos en la que R_1 es $-O-CH_2N_3$, los métodos comprenden opcionalmente además tratar el análogo de dNPP incorporado para retirar el $-CH_2N_3$ y dar como resultado un grupo $-OH$ unido a la posición 3' permitiendo así la incorporación de un análogo de dNPP adicional.
- 20 En una realización de los métodos en la que R_1 es $-O-2$ -nitrobencilo, los métodos comprenden opcionalmente además tratar el análogo de nucleótido incorporado para retirar el -2 -nitrobencilo y dar como resultado un grupo $-OH$ unido a la posición 3' permitiendo así la incorporación de un análogo de dNPP adicional.
- En una realización de los métodos los análogos de dNPP tienen las siguientes estructuras:



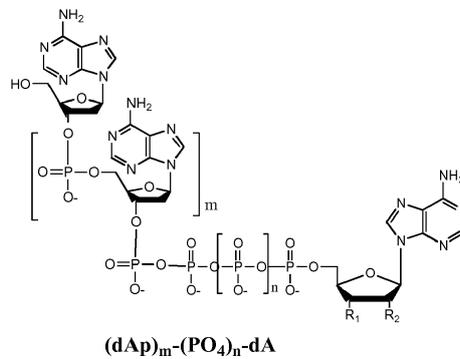
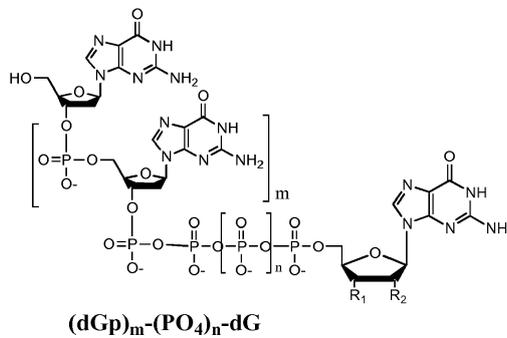
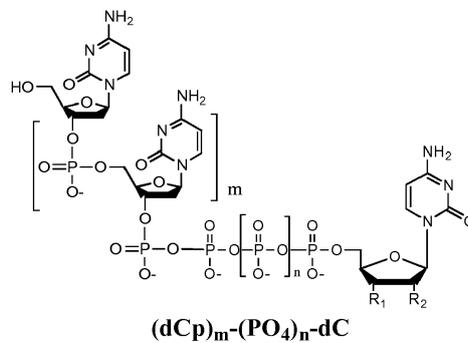
25



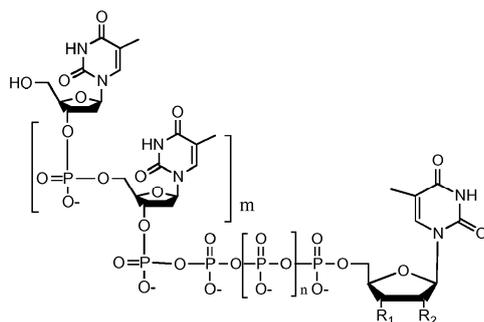
en las que R₁ es OH, en las que R₂ es H u OH, en las que Z es O, S o BH₃, y en las que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

- 5 En una realización de los métodos la etiqueta es un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, o un hexanucleótido, en los que la base del mononucleótido, el dinucleótido, el trinucleótido, el tetranucleótido, el pentanucleótido o el hexanucleótido es el mismo tipo de base que la base del análogo de dNPP.

En una realización de los métodos la etiqueta se elige de las siguientes:



10

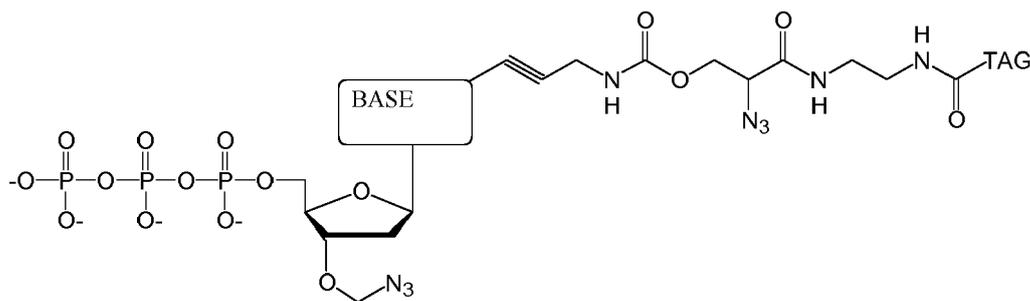


$(dTp)_m-(PO_4)_n-dT$

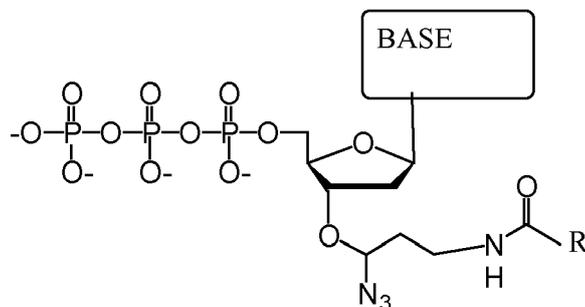
5 en las que en cada estructura n es, independientemente, 1, 2, 3 ó 4, y m es, independientemente, un número entero de desde 0 hasta 100, y en las que cuando m es 0 el fosfato terminal del dNPP está unido directamente al átomo de 3'-O del nucleósido mostrado en el lado izquierdo de la estructura, y en las que el valor de n es diferente para cada tipo de base.

En una realización de los métodos m es un número entero de desde 0 hasta 50. En una realización de los métodos m es un número entero de desde 0 hasta 10.

Se proporcionan diversos ejemplos no limitativos de nucleótidos marcados con etiqueta. En una realización de los métodos el análogo de dNPP tiene la estructura:

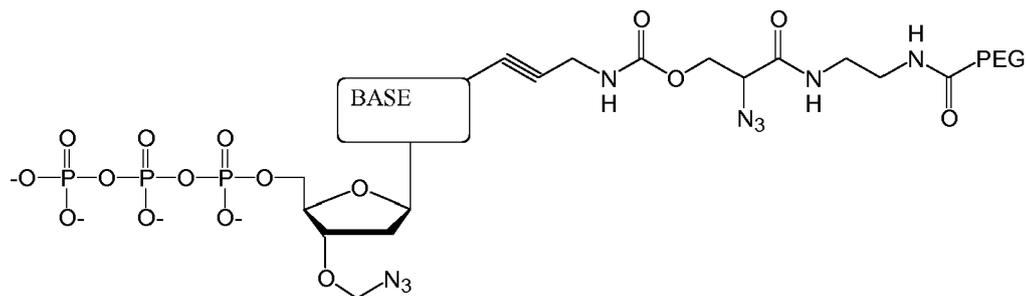


10



en la que R es un hidrocarbilo sustituido o no sustituido, de hasta 3000 Dalton, y en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

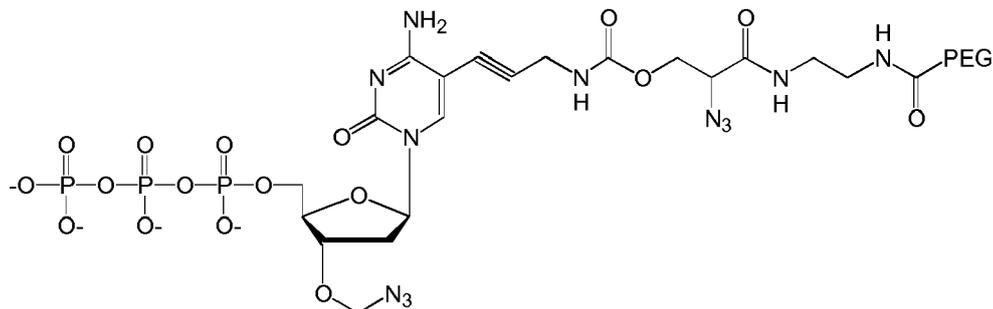
En una realización de los métodos el análogo de dNPP tiene la estructura:



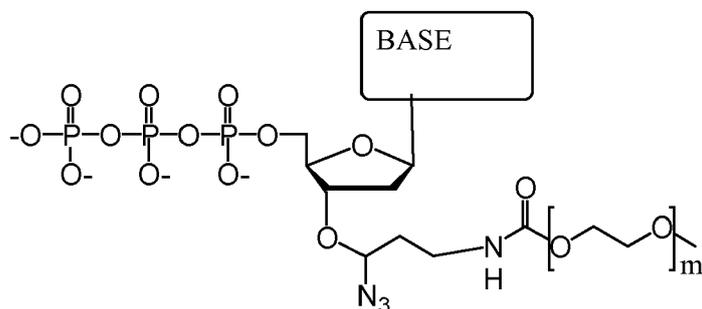
15

en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

En una realización de los métodos el análogo de dNPP tiene la estructura:



En una realización de los métodos el análogo de dNPP tiene la estructura:



- 5 en la que m es un número entero de desde 1-50, y en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

En una realización de los métodos el cambio electrónico es un cambio en la amplitud de corriente.

En una realización de los métodos el cambio electrónico es un cambio en la conductancia del nanoporo.

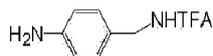
- 10 En una realización de los métodos el nanoporo es biológico. En una realización de los métodos el nanoporo es proteico. En una realización de los métodos el nanoporo comprende alfa-hemolisina. En una realización de los métodos el nanoporo es grafeno. En una realización de los métodos el nanoporo es un nanoporo de estado sólido. En una realización de los métodos el nanoporo está en una membrana de estado sólido.

En una realización de los métodos el ADN monocatenario, el cebador o la ADN polimerasa está unido a una superficie sólida.

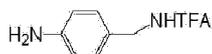
- 15 En otra realización de los métodos el nanoporo forma parte de una matriz de nanoporos.

Puede usarse cualquier método adecuado para unir las etiquetas. En un ejemplo, pueden unirse etiquetas al fosfato terminal mediante (a) poner en contacto un trifosfato de nucleótido con diciclohexilcarbodiimida/dimetilformamida en condiciones que permiten producción de un trimetafosfato cíclico; (b) poner en contacto el producto resultante de la etapa a) con un nucleófilo para formar un compuesto funcionalizado con -OH o -NH₂; y (c) hacer reaccionar el producto de la etapa b) con una etiqueta que tiene un grupo -COR unido a la misma en condiciones que permiten que la etiqueta se una indirectamente a un fosfato terminal formando así el análogo de trifosfato de nucleótido.

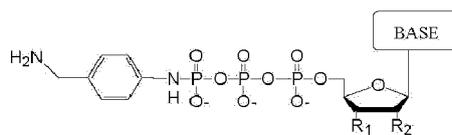
- 20 En algunos casos, el nucleófilo es H₂N-R-OH, H₂N-R-NH₂, R'S-R-OH, R'S-R-NH₂ o



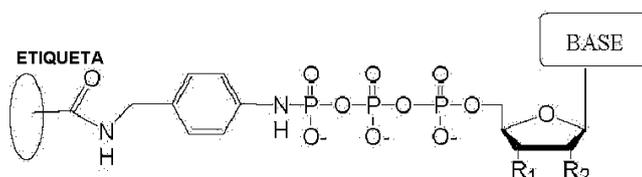
- 25 En algunos casos, el método comprende, en la etapa b), poner en contacto el producto resultante de la etapa a) con un compuesto que tiene la estructura:



y posteriormente, o de manera simultánea, poner en contacto el producto con NH₄OH para formar un compuesto que tiene la estructura:



Después puede hacerse reaccionar el producto de la etapa b) con una etiqueta que tiene un grupo -COR unido a la misma en condiciones que permiten que la etiqueta se una indirectamente a un fosfato terminal formando así el análogo de trifosfato de nucleótido que tiene la estructura:

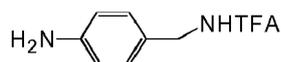


5 en la que R₁ es OH, en la que R₂ es H u OH, en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

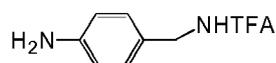
10 En particular es un procedimiento para producir un análogo de trifosfato de nucleótido, en el que el análogo de trifosfato de nucleótido difiere de un trifosfato de nucleótido por tener una etiqueta unida al fosfato terminal del mismo, que comprende: (a) poner en contacto un trifosfato de nucleótido con dicitohexilcarbodiimida/dimetilformamida en condiciones que permiten producción de un trimetafosfato cíclico; y (b) poner en contacto el producto resultante de la etapa a) con una etiqueta que tiene un grupo hidroxilo o amino unido a la misma en condiciones que permiten la apertura nucleófila del trimetafosfato cíclico para unir la etiqueta a un fosfato terminal formando así el análogo de trifosfato de nucleótido.

15 Un procedimiento para producir un análogo de trifosfato de nucleótido, en el que el análogo de trifosfato de nucleótido difiere de un trifosfato de nucleótido por tener una etiqueta unida al fosfato terminal del mismo, que comprende: (a) poner en contacto un trifosfato de nucleótido con dicitohexilcarbodiimida/dimetilformamida en condiciones que permiten producción de un trimetafosfato cíclico; (b) poner en contacto el producto resultante de la etapa a) con un nucleófilo para formar un compuesto funcionalizado con -OH o -NH₂; y (c) hacer reaccionar el producto de la etapa b) con una etiqueta que tiene un grupo -COR unido a la misma en condiciones que permiten que la etiqueta se una indirectamente a un fosfato terminal formando así el análogo de trifosfato de nucleótido.

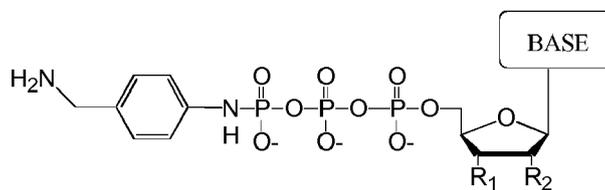
20 En una realización del presente procedimiento el nucleófilo es H₂N-R-OH, H₂N-R-NH₂, R'S-R-OH, R'S-R-NH₂ o



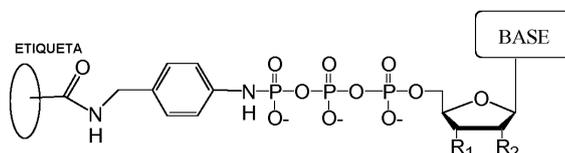
25 En una realización el presente procedimiento comprende en la etapa b) poner en contacto el producto resultante de la etapa a) con un compuesto que tiene la estructura:



y después NH₄OH para formar un compuesto que tiene la estructura:



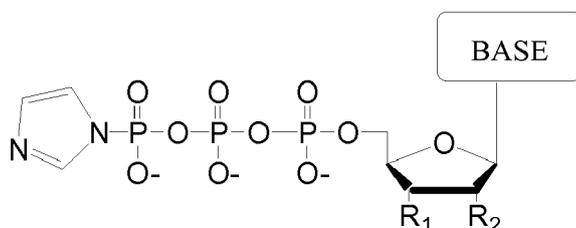
30 y hacer reaccionar el producto de la etapa b) con una etiqueta que tiene un grupo -COR unido a la misma en condiciones que permiten que la etiqueta se una indirectamente a un fosfato terminal formando así el análogo de trifosfato de nucleótido que tiene la estructura:



en la que R₁ es OH, en la que R₂ es H u OH, en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

5 Un procedimiento para producir un análogo de tetrafosfato de nucleótido, en el que el análogo de tetrafosfato de nucleótido difiere de un tetrafosfato de nucleótido por tener una etiqueta unida al fosfato terminal del mismo, que comprende:

(a) poner en contacto un trifosfato de nucleótido con 1,1'-carbonildiimidazol/dimetilformamida en condiciones que permiten la formación de la siguiente estructura:

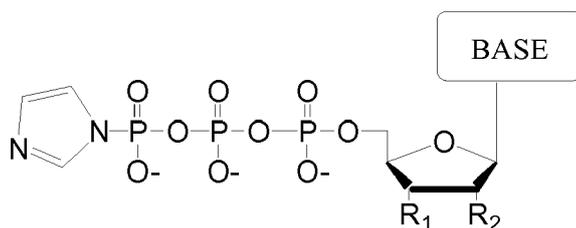


10 en la que R₁ es OH, en la que R₂ es H u OH, en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina; y

(b) poner en contacto el producto resultante de la etapa a) con una etiqueta que tiene un grupo monofosfato unido a la misma en condiciones que permiten la formación del análogo de tetrafosfato de nucleótido.

15 Un procedimiento para producir un análogo de tetrafosfato de nucleótido, en el que el análogo de tetrafosfato de nucleótido difiere de un tetrafosfato de nucleótido por tener una etiqueta unida al fosfato terminal del mismo, que comprende:

(a) poner en contacto un trifosfato de nucleótido con 1,1'-carbonildiimidazol/dimetilformamida en condiciones que permiten la formación de la siguiente estructura:



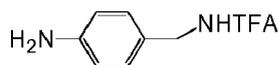
20 en la que R₁ es OH, en la que R₂ es H u OH, en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina;

(b) poner en contacto el producto resultante de la etapa a) con ácido fosfórico en condiciones que permiten la formación de un tetrafosfato de nucleótido;

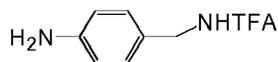
(c) poner en contacto el tetrafosfato de nucleótido con 1) carbonildiimidazol/dimetilformamida; 2) un nucleófilo y después 3) NH₄OH para formar un compuesto funcionalizado con -OH o -NH₂; y

25 (d) poner en contacto el producto de la etapa c) con una etiqueta que tiene un grupo -COR unido a la misma en condiciones que permiten que la etiqueta se una indirectamente a un fosfato terminal formando así el análogo de tetrafosfato de nucleótido.

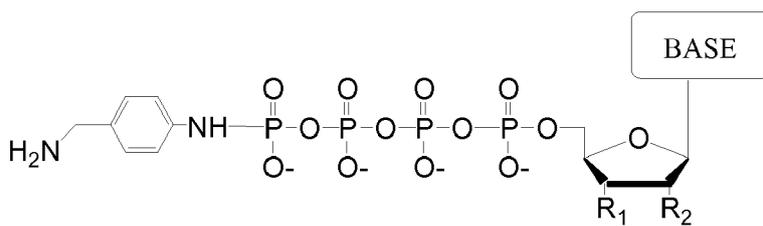
En una realización del presente procedimiento el nucleófilo es H₂N-R-OH, H₂N-R-NH₂, R'S-R-OH, R'S-R-NH₂ o



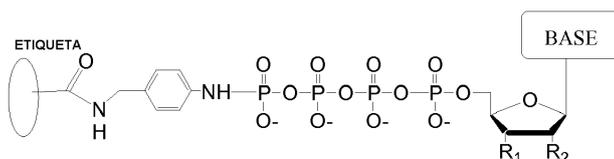
30 En una realización el presente procedimiento comprende en la etapa b) poner en contacto el tetrafosfato de nucleótido con 1) carbonildiimidazol/dimetilformamida; 2) un compuesto que tiene la estructura:



y después 3) NH₄OH para formar un compuesto que tiene la estructura:



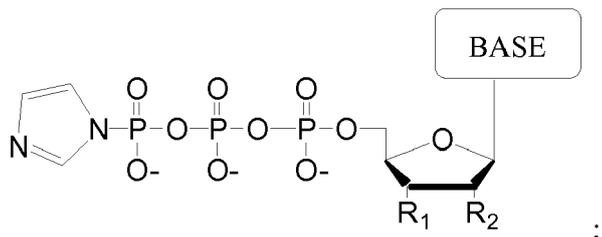
y poner en contacto el producto de la etapa b) con una etiqueta que tiene un grupo -COR unido a la misma en condiciones que permiten que la etiqueta se una indirectamente a un fosfato terminal formando así el análogo de trifosfato de nucleótido que tiene la estructura:



5 en la que R₁ es OH, en la que R₂ es H u OH, en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

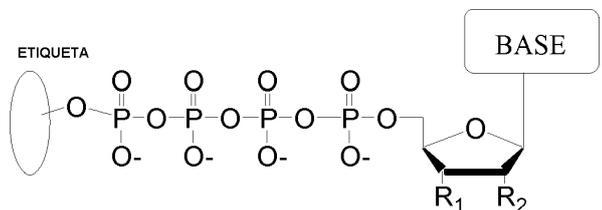
10 Un procedimiento para producir un análogo de tetrafosfato de nucleótido, en el que el análogo de tetrafosfato de nucleótido difiere de un tetrafosfato de nucleótido por tener una etiqueta unida al fosfato terminal del mismo, que comprende:

(a) poner en contacto un trifosfato de nucleótido con 1,1'-carbonildiimidazol/dimetilformamida en condiciones que permiten la formación de la siguiente estructura:



15 (b) poner en contacto el producto resultante de la etapa a) con ácido fosfórico en condiciones que permiten la formación de un tetrafosfato de nucleótido; y

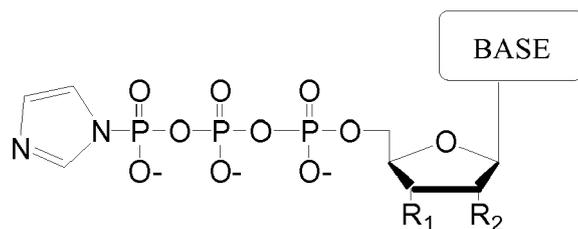
(c) poner en contacto el tetrafosfato de nucleótido con carbonildiimidazol/dimetilformamida y una etiqueta que tiene un grupo hidroxilo o amino unido a la misma para formar un compuesto que tiene la estructura:



20 en la que R₁ es OH, en la que R₂ es H u OH, en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

Un procedimiento para producir un análogo de pentafofosfato de nucleótido, en el que el análogo de pentafofosfato de nucleótido difiere de un pentafofosfato de nucleótido por tener una etiqueta unida al fosfato terminal del mismo, que comprende:

25 (a) poner en contacto un trifosfato de nucleótido con 1,1'-carbonildiimidazol/dimetilformamida en condiciones que permiten la formación de la siguiente estructura:

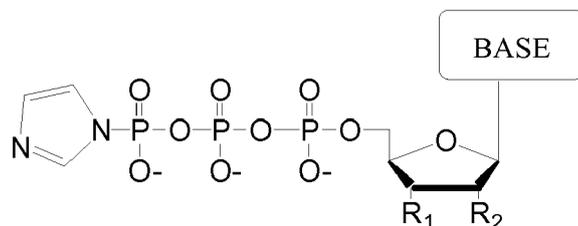


en la que R₁ es OH, en la que R₂ es H u OH, en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina; y

- 5 (b) poner en contacto el producto resultante de la etapa a) con una etiqueta que tiene un grupo pirofosfato unido a la misma en condiciones que permiten la formación del análogo de pentafofosfato de nucleótido.

Un procedimiento para producir un análogo de pentafofosfato de nucleótido, en el que el análogo de pentafofosfato de nucleótido difiere de un pentafofosfato de nucleótido por tener una etiqueta unida al fosfato terminal del mismo, que comprende:

- 10 (a) poner en contacto un trifosfato de nucleótido con 1,1'-carbonildiimidazol/dimetilformamida en condiciones que permiten la formación de la siguiente estructura:



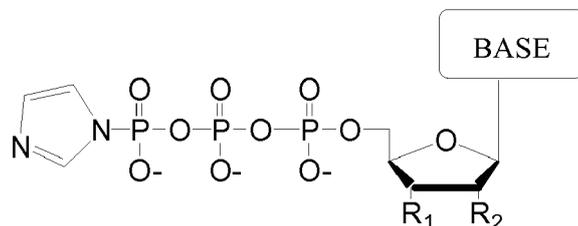
en la que R₁ es OH, en la que R₂ es H u OH, en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina;

- 15 (b) poner en contacto el producto resultante de la etapa a) con un grupo pirofosfato en condiciones que permiten la formación de un pentafofosfato de nucleótido; y

(c) poner en contacto el pentafofosfato de nucleótido con carbonildiimidazol/dimetilformamida y una etiqueta que tiene un grupo hidroxilo o amino unido a la misma para formar el análogo de pentafofosfato de nucleótido.

- 20 Un procedimiento para producir un análogo de hexafofosfato de nucleótido, en el que el análogo de hexafofosfato de nucleótido difiere de un hexafofosfato de nucleótido por tener una etiqueta unida al fosfato terminal del mismo, que comprende:

(a) poner en contacto un trifosfato de nucleótido con 1,1'-carbonildiimidazol/dimetilformamida en condiciones que permiten la formación de la siguiente estructura:

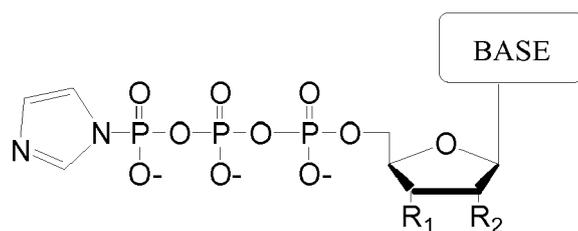


- 25 en la que R₁ es OH, en la que R₂ es H u OH, en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina; y

(b) poner en contacto el producto resultante de la etapa a) con una etiqueta que tiene un grupo trifosfato unido a la misma en condiciones que permiten la formación del análogo de hexafofosfato de nucleótido.

- 30 Un procedimiento para producir un análogo de hexafofosfato de nucleótido, en el que el análogo de hexafofosfato de nucleótido difiere de un hexafofosfato de nucleótido por tener una etiqueta unida al fosfato terminal del mismo, que comprende:

(a) poner en contacto un trifosfato de nucleótido con 1,1'-carbonildiimidazol/dimetilformamida en condiciones que permiten la formación de la siguiente estructura:

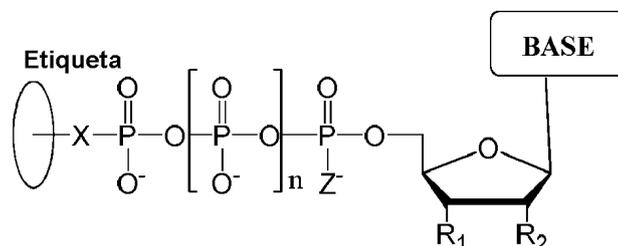


en la que R_1 es OH, en la que R_2 es H u OH, en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina;

5 (b) poner en contacto el producto resultante de la etapa a) con un grupo trifosfato en condiciones que permiten la formación de un hexafosfato de nucleótido; y

(c) poner en contacto el hexafosfato de nucleótido con carbonildiimidazol/dimetilformamida y una etiqueta que tiene un grupo hidroxilo o amino unido a la misma para formar el análogo de hexafosfato de nucleótido.

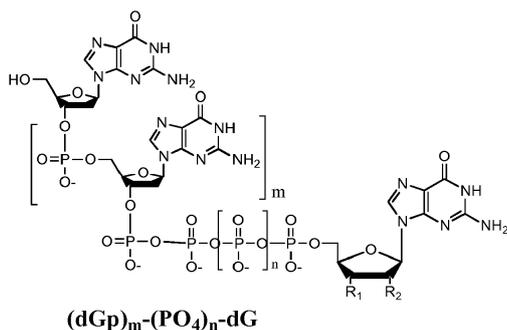
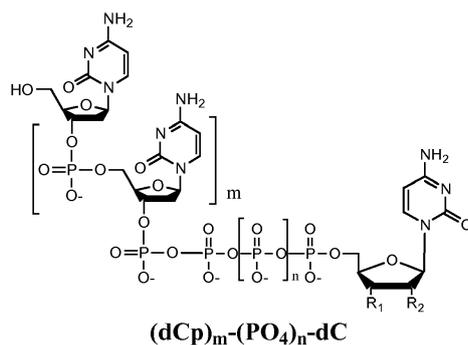
Un compuesto que tiene la estructura:

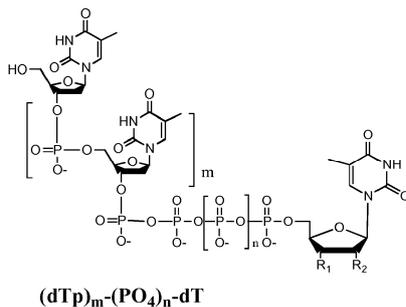
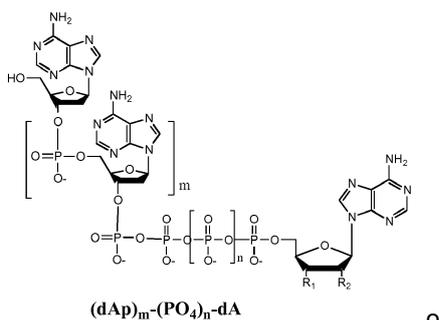


10 en la que la etiqueta es etilenglicol, un aminoácido, un hidrato de carbono, un colorante, mononucleótido, dinucleótido, trinucleótido, tetranucleótido, pentanucleótido o hexanucleótido, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina, y en la que n es 1, 2, 3 ó 4.

En una realización R_2 es H. En una realización R_2 es OH.

15 En algunos casos una etiqueta se elige de las moléculas $(dCp)_m$, $(dGp)_m$, $(dAp)_m$, y $(dTp)_m$. La figura 27(a) muestra algunos ejemplos de estas moléculas unidas a un nucleótido. Por ejemplo, un compuesto que tiene la estructura:

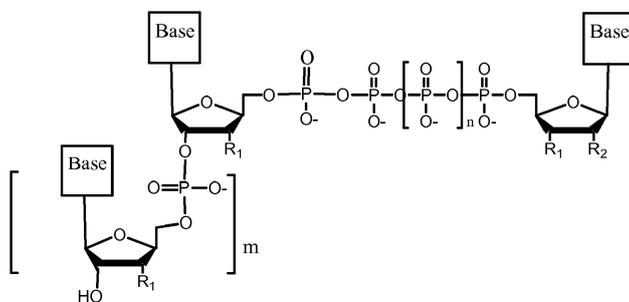




5 en las que en cada estructura n es, independientemente, 1, 2, 3 ó 4, y m es, independientemente, un número entero de desde 0 hasta 100, y en las que cuando m es 0 el fosfato terminal del dNTP está unido directamente al átomo de 3'-O del nucleósido mostrado en el lado izquierdo de la estructura, en las que R₁ es -OH u -O-CH₂N₃, y R₂ es H u OH. En algunos casos, el valor de n es diferente para cada tipo de base.

En una realización m es desde 0 hasta 50. En una realización m es desde 0 hasta 10. En una realización R₁ es -OH. En una realización R₂ es -H. En una realización R₂ es -OH.

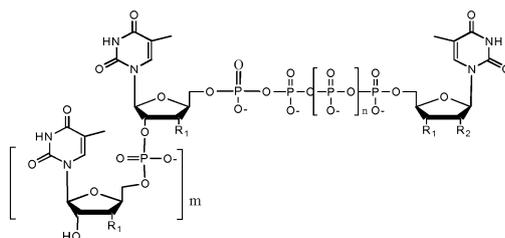
Un compuesto que tiene la estructura:



10 en la que m es un número entero de desde 0 hasta 100, y en el que el compuesto comprende un único tipo de base, y en el que la base es adenina, guanina, citosina, uracilo o timina o un derivado de una de esas bases.

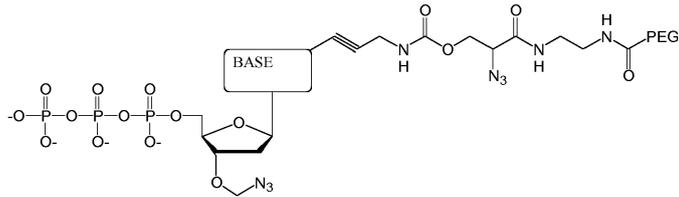
En una realización m es desde 0 hasta 50. En una realización m es desde 0 hasta 10.

En una realización el compuesto tiene la estructura:



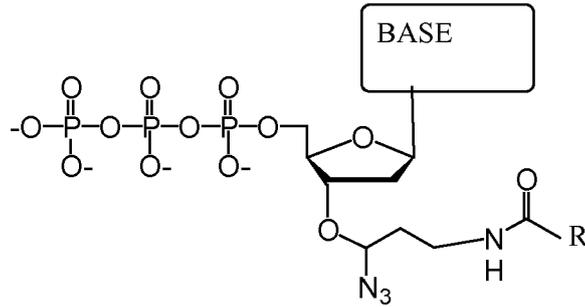
15 en la que m es un número entero de desde 0 hasta 100.

Un compuesto que tiene la estructura:



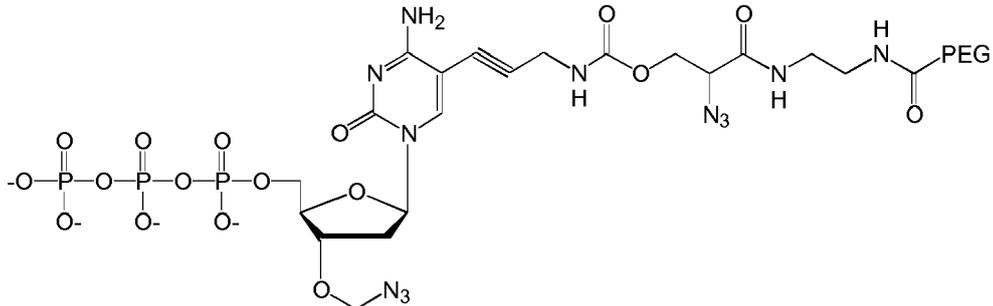
en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

Un compuesto que tiene la estructura:

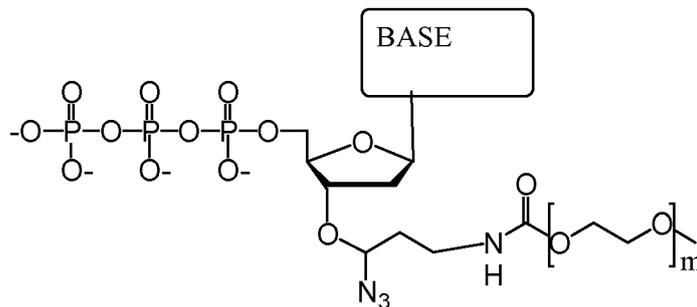


- 5 en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina, y R es un hidrocarbilo sustituido o no sustituido, de hasta 3000 Dalton.

Un compuesto que tiene la estructura:



Un compuesto que tiene la estructura:



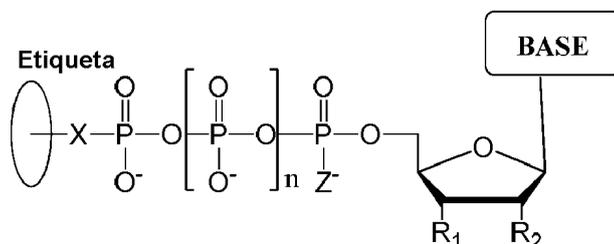
10

en la que la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina, y m es un número entero de desde 1-50.

Un compuesto que tiene la estructura:

porción del mismo, con una ARN polimerasa y cuatro análogos de polifosfato de ribonucleótido (rNPP) al menos uno de los cuales puede hibridarse a cada uno de un nucleótido de A, U, G o C en el ARN que está secuenciándose en condiciones que permiten que la ARN polimerasa catalice la incorporación de uno de los análogos de rNPP en el cebador si es complementario al residuo de nucleótido del ARN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ARN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ARN, en el que cada uno de los cuatro análogos de rNPP tiene la estructura:

5



en la que la base es adenina, guanina, citosina, timina o uracilo o un derivado de una o más de estas bases, en la que R_1 es OH, en la que R_2 es OH, en la que X es O, NH, S o CH_2 , en la que n es 1, 2, 3 ó 4, en la que Z es O, S o BH_3 , y con la condición de que (i) el tipo de base en cada análogo de rNPP es diferente del tipo de base en cada uno de los otros tres análogos de rNPP, y (ii) o bien el valor de n de cada análogo de rNPP es diferente del valor de n de cada uno de los otros tres análogos de rNPP, o bien el valor de n de cada uno de los cuatro análogos de rNPP es el mismo y el tipo de etiqueta en cada análogo de rNPP es diferente del tipo de etiqueta en cada uno de los otros tres análogos de rNPP, en el que la incorporación del análogo de rNPP da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo; y

10

15

(b) identificar qué análogo de rNPP se ha incorporado en el cebador para formar el producto de extensión de ARN en la etapa (a) aplicando una tensión a través de la membrana y midiendo un cambio electrónico a través del nanoporo resultante de que el polifosfato que tiene la etiqueta unida al mismo generado en la etapa (a) se transloca a través del nanoporo, en el que el cambio electrónico es diferente para cada valor de n, o para cada tipo de etiqueta diferente, lo que sea aplicable, permitiendo así identificar el residuo de nucleótido en el ARN monocatenario complementario al análogo de rNPP incorporado; y

20

(c) realizar de manera repetida la etapa (b) para cada residuo de nucleótido del ARN monocatenario que está secuenciándose, en el que en cada iteración de la etapa (b) se identifica qué análogo de rNPP se ha incorporado en el producto de extensión de ARN en la etapa (a), en el que el análogo de rNPP se ubica inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ARN monocatenario hibridado al residuo de nucleótido 3'-terminal del producto de extensión de ARN, determinando así la secuencia de nucleótidos del ARN monocatenario.

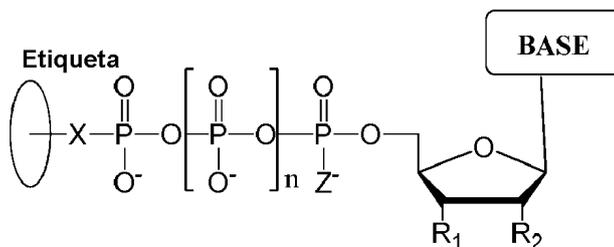
25

Un método para determinar la secuencia de nucleótidos de un ARN monocatenario que comprende:

(a) poner en contacto el ARN monocatenario, en el que el ARN monocatenario está en una disolución de electrolito en contacto con un nanoporo en una membrana y en el que el ARN monocatenario tiene un cebador hibridado a una porción del mismo, con una ARN polimerasa y un análogo de polifosfato de ribonucleótido (rNPP) que puede hibridarse a un nucleótido de A, U, G o C en el ARN que está secuenciándose en condiciones que permiten que la ARN polimerasa catalice la incorporación del análogo de rNPP en el cebador si es complementario al residuo de nucleótido del ARN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ARN monocatenario hibridado al residuo de nucleótido 3'-terminal del cebador, para formar un producto de extensión de ARN, en el que el análogo de rNPP tiene la estructura:

30

35



en la que la base es adenina, guanina, citosina, uracilo o un derivado de una de esas bases, en la que R_1 es -OH, -O- CH_2N_3 u -O-2-nitrobencilo, en la que R_2 es -OH, en la que X es O, NH, S o CH_2 , en la que n es 1, 2, 3 ó 4, en la que Z es O, S o BH_3 , y en el que si no se incorpora el análogo de rNPP, repetir de manera iterativa la puesta en contacto con un análogo de rNPP diferente hasta que se incorpora un análogo de rNPP, con la condición de que (i) el tipo de base en cada análogo de rNPP es diferente del tipo de base en cada uno de los otros análogos de rNPP, y (ii) o bien el valor de n de cada análogo de rNPP es diferente del valor de n de cada uno de los otros análogos de rNPP, o bien el valor de n de cada uno de los análogos de rNPP es el mismo y el tipo de etiqueta en cada análogo

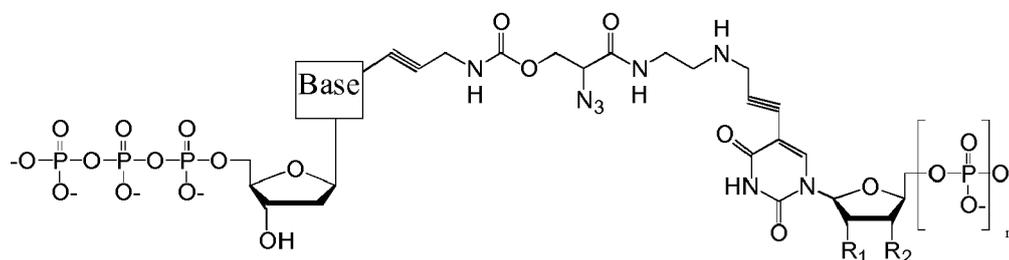
40

de rNPP es diferente del tipo de etiqueta en cada uno de los tres análogos de rNPP, en el que la incorporación del análogo de rNPP da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo;

5 (b) identificar qué análogo de rNPP se ha incorporado en el cebador para formar el producto de extensión de ARN en la etapa (a) aplicando una tensión a través de la membrana y midiendo un cambio electrónico a través del nanoporo resultante de que el polifosfato que tiene la etiqueta unida al mismo generado en la etapa (a) se transloca a través del nanoporo, en el que el cambio electrónico es diferente para cada valor de n, o diferente para cada tipo de etiqueta, lo que sea aplicable, permitiendo así identificar el residuo de nucleótido en el ARN monocatenario complementario al análogo de rNPP incorporado;

10 (c) realizar de manera repetida las etapas (a) y (b) para cada residuo de nucleótido del ARN monocatenario que está secuenciándose, en el que en cada iteración de la etapa (a) el análogo de rNPP se incorpora en el producto de extensión de ARN si es complementario al residuo de nucleótido del ARN monocatenario que está inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ARN monocatenario hibridado al residuo de nucleótido 3'-terminal del producto de extensión de ARN, determinando así la secuencia de nucleótidos del ARN monocatenario.

En una realización el análogo de dNPP tiene la estructura:



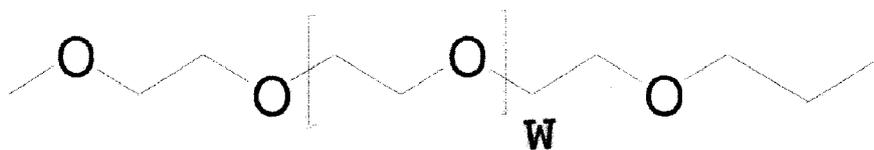
en la que n es 1 ó 2 y la base es adenina, guanina, citosina, timina, uracilo, una 7-desazapurina o una 5-metilpirimidina.

En una realización el nanoporo biológico está integrado con electrónica de CMOS. En otra realización el nanoporo de estado sólido está integrado con electrónica de CMOS.

20 En una realización la unión a la superficie sólida se realiza mediante uniones de biotina-estreptavidina. En otra realización la ADN polimerasa se une a la superficie sólida a través de una superficie de oro modificada con una monocapa autoensamblada de alcanotiol funcionalizada con grupos amino, en la que los grupos amino están modificados para dar ésteres de NHS para su unión a grupos amino en la ADN polimerasa.

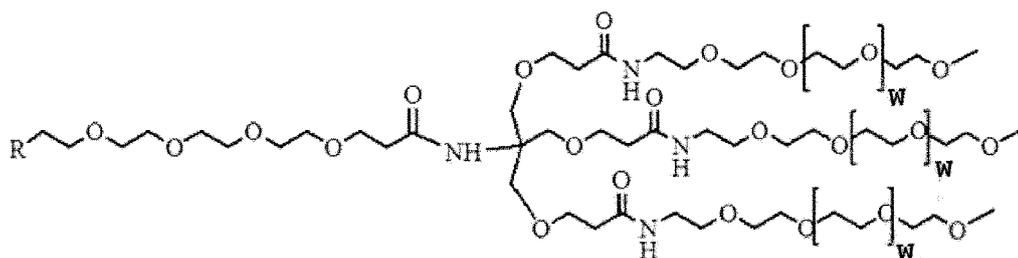
25 En una realización el análogo de dNPP es un polifosfato de nucleósido marcado con etiqueta en fosfato terminal. En una realización adicional cada tipo de análogo de dNPP tiene una etiqueta de polietilenglicol que difiere en cuanto al tamaño con respecto a las etiquetas de polietilenglicol de cada uno de los otros tres tipos de análogos de dNPP.

En algunos casos, la etiqueta es un polímero. El polietilenglicol (PEG) es un ejemplo de un polímero. En una realización la etiqueta tiene la siguiente estructura:



30 en la que W es un número entero entre 0 y 100. Puede usarse cualquier número de unidades de etilenglicol (W). En algunos casos, W es un número entero entre 0 y 100. En algunos casos, el número de unidades de etilenglicol es diferente para cada tipo de nucleótido. En una realización, los cuatro tipos de nucleótidos comprenden etiquetas que tienen 16, 20, 24 ó 36 unidades de etilenglicol. En algunos casos, la etiqueta comprende además un resto identificable adicional, tal como un colorante basado en cumarina.

35 En algunos casos, una etiqueta comprende múltiples cadenas de PEG. Un ejemplo de una etiqueta de este tipo tiene la siguiente estructura:



en la que R es NH₂, OH, COOH, CHO, SH o N₃, y W es un número entero de desde 0 hasta 100.

Una composición que comprende al menos cuatro análogos de polifosfato de desoxinucleótido (dNPP), que tienen cada uno una estructura seleccionada de las estructuras expuestas en las reivindicaciones 74 y 75, en la que cada uno de los cuatro análogos de dNPP comprende un tipo de base que es diferente del tipo de base de los otros tres análogos de dNPP.

En una realización, cada uno de los cuatro análogos de dNPP tiene una etiqueta de polietilenglicol que es diferente en cuanto al tamaño con respecto a las etiquetas de polietilenglicol de cada uno de los otros tres análogos de dNPP.

En una realización, la carga neta en el polifosfato de nucleósido marcado con etiqueta es neutra. En otra realización, la etiqueta liberada tiene una carga positiva.

En una realización, el método comprende además una etapa de tratar con fosfatasa alcalina después de la etapa b), en el que la fosfatasa alcalina hidroliza grupos fosfato libres en el pirofosfato de etiqueta liberada.

En una realización múltiples copias del ADN monocatenario están inmovilizadas en una perla.

Un método tal como se muestra en la figura 51 para determinar la secuencia de nucleótidos de múltiples copias de la misma molécula de ADN monocatenario que comprende:

(a) tratar el ADN monocatenario en una disolución de electrolito en contacto con un nanoporo en una membrana y en el que el ADN tiene un cebador hibridado a una porción del mismo, con una ADN polimerasa y sucesivamente que cada uno de cuatro análogos de desoxirribonucleótido marcados con etiqueta que pueden hibridarse a un nucleótido de A, T, G o C en el ADN que está secuenciándose en condiciones que permiten que la ADN polimerasa catalice la incorporación del análogo en el extremo de un producto de extensión del cebador si es complementario al residuo de nucleótido del ADN que está secuenciándose inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario que está secuenciándose hibridado al residuo de nucleótido 3'-terminal del cebador, en el que si no se incorpora el análogo, repetir de manera iterativa la puesta en contacto con un análogo diferente hasta que se incorpora un análogo, con la condición de que (i) el tipo de base en cada análogo es diferente del tipo de base en cada uno de los otros análogos, y (ii) el tipo de etiqueta en cada análogo es diferente del tipo de etiqueta en cada uno de los otros análogos, en el que la incorporación del análogo da como resultado la liberación de la etiqueta;

(b) identificar el análogo que se ha incorporado en el producto de extensión en la etapa (a) aplicando una tensión a través de la membrana y midiendo una carga eléctrica a través del nanoporo resultante de la etiqueta unida al análogo; y

(c) realizar de manera repetida las etapas (a) y (b), obteniendo así la secuencia de nucleótidos del ADN monocatenario.

Un método tal como se muestra en la figura 52 para determinar la secuencia de nucleótidos de múltiples copias de la misma molécula de ADN monocatenario que comprende:

(a) tratar el ADN monocatenario en una disolución de electrolito en contacto con un nanoporo en una membrana y en el que el ADN tiene un cebador hibridado a una porción del mismo, con una ADN polimerasa y cuatro análogos de desoxirribonucleótido bloqueados en 3' al menos uno de los cuales puede hibridarse a cada uno de un nucleótido de A, T, G o C en el ADN que está secuenciándose en condiciones que permiten que la ADN polimerasa catalice la incorporación del análogo en el extremo de un producto de extensión del cebador si es complementario al residuo de nucleótido del ADN que está secuenciándose inmediatamente en sentido 5' con respecto a un residuo de nucleótido del ADN monocatenario que está secuenciándose hibridado al residuo de nucleótido 3'-terminal del cebador, en el que cada análogo comprende un terminador reversible, con la condición de que (i) el tipo de base en cada análogo es diferente del tipo de base en cada uno de los otros tres análogos, y (ii) la etiqueta en cada análogo es diferente de la etiqueta en cada uno de los otros tres análogos, en el que la incorporación del análogo da como resultado la liberación de la etiqueta;

(b) identificar el análogo que se ha incorporado en el producto de extensión en la etapa (a) aplicando una tensión a través de la membrana y midiendo una carga eléctrica a través del nanoporo resultante de la etiqueta unida al

análogo;

(c) retirar el terminador reversible a partir del análogo que se ha incorporado en el producto de extensión en la etapa (a); y

5 (d) realizar de manera repetida las etapas (b) y (c), obteniendo así la secuencia de nucleótidos del ADN monocatenario.

En una realización, el nanoporo está integrado directamente en una pastilla de CMOS tal como se muestra en la figura 46.

Se contempla que el nanoporo tenga una carga negativa o, alternativamente, tenga una carga que es de signo opuesto a la carga de la etiqueta o del polifosfato que tiene la etiqueta unida al mismo.

10 En una realización, la tasa de incorporación del análogo de nucleótido mediante la polimerasa es menor que, o alternativamente es igual a, la tasa de translocación de la etiqueta o el polifosfato que tiene la etiqueta unida al mismo a través del nanoporo.

También se da a conocer un método que comprende obtener el ARN o ADN monocatenario que va a secuenciarse a partir de un ARN o ADN bicatenario antes de la etapa (a).

15 En una realización, la polimerasa está unida al nanoporo.

Se contempla que la etiqueta pueda detectarse basándose en el tamaño, longitud, forma, masa, carga o cualquier combinación de los mismos.

Se contempla que diversas realizaciones del sistema de medición de conductancia también puedan aplicarse al método para determinar una secuencia de nucleótidos, y viceversa.

20 También se proporciona un compuesto que tiene la estructura de cualquiera de los compuestos expuestos en las figuras y/o los esquemas de la presente solicitud.

La presente divulgación también proporciona un análogo de dNPP que comprende una etiqueta que tiene la estructura de cualquiera de las etiquetas expuestas en las figuras y/o los esquemas de la presente solicitud.

25 En una realización, la etiqueta es un hidrocarbilo, sustituido o no sustituido, tal como un alquilo, alquenilo, alquinilo, y que tiene una masa de 3000 Dalton o menos.

30 En una realización el ADN monocatenario, ARN, cebador o sonda se une a un sustrato sólido mediante química de cicloadición de azida-alquino 1,3-dipolar. En una realización el ADN, ARN, cebador o sonda se une a un sustrato sólido mediante una molécula de polietilenglicol. En una realización el ADN, ARN, cebador o sonda está marcado con alquino. En una realización el ADN, ARN, cebador o sonda se une a un sustrato sólido mediante una molécula de polietilenglicol y un sustrato sólido está funcionalizado con azida. En una realización el ADN, ARN, cebador o sonda se inmoviliza sobre el sustrato sólido mediante una unión azido, una unión alquinilo o interacción de biotina-estreptavidina. La inmovilización de ácidos nucleicos se describe en Immobilization of DNA on Chips II, editado por Christine Wittmann (2005), Springer Verlag, Berlín, que se incorpora en el presente documento como referencia. En una realización el ADN es ADN monocatenario. En una realización el ARN es ARN monocatenario.

35 En una realización el sustrato sólido está en forma de un chip, una perla, un pocillo, un tubo capilar, un portaobjetos, una oblea, un filtro, una fibra, unos medios porosos, un nanotubo poroso o una columna. Esta divulgación también proporciona el presente método, en el que el sustrato sólido es un metal, oro, plata, cuarzo, sílice, un plástico, polipropileno, un vidrio o diamante. Esta divulgación también proporciona el presente método, en el que el sustrato sólido es una sustancia no metálica porosa a la que se une o en la que se impregna un metal o una combinación de metales. La superficie sólida puede estar en diferentes formas incluyendo los ejemplos no limitativos de un chip, una perla, un tubo, una matriz, un nanotubo. La superficie sólida puede fabricarse a partir de materiales habituales para micromatrices de ADN, incluyendo los ejemplos no limitativos de vidrio o nailon. La superficie sólida, por ejemplo perlas/microperlas, puede a su vez inmovilizarse en otra superficie sólida tal como un chip.

45 En una realización se separan muestras de ácido nucleico, ADN, ARN, cebador o sonda en compartimentos, pocillos o depresiones diferenciados en una superficie o en un recipiente.

También se da a conocer el presente método, en el que aproximadamente 1000 o menos copias de la muestra de ácido nucleico, ADN, ARN, cebador o sonda, se unen a la superficie sólida. También se da a conocer cuando se unen 2×10^7 , 1×10^7 , 1×10^6 ó 1×10^4 o menos copias de la muestra de ácido nucleico, ADN, ARN, cebador o sonda a la superficie sólida.

50 En una realización la muestra de ácido nucleico, ADN, ARN, cebador o sonda inmovilizada está inmovilizada a una alta densidad. Esta invención también proporciona la presente invención en la que más de, o hasta, 1×10^7 , 1×10^8 , 1×10^9 copias de la muestra de ácido nucleico, ADN, ARN, cebador o sonda, se inmovilizan en el sustrato sólido.

En una realización la ADN polimerasa es polimerasa 9°N o una variante de la misma, ADN polimerasa I de *E. Coli*, ADN polimerasa de bacteriófago T4, Sequenase, ADN polimerasa Taq o polimerasa 9°N (exo-)A485L/Y409V.

5 En una realización de los métodos o de las composiciones descritos en el presente documento, el ADN es monocatenario. En una realización de los métodos o de las composiciones descritos en el presente documento, el ARN es monocatenario, Phi29 o variantes de los mismos.

En una realización de los métodos descritos para la secuenciación de ARN, la polimerasa es una ARN polimerasa, transcriptasa inversa o polimerasa apropiada para la polimerización de ARN.

10 Los ligadores pueden ser fotoescindibles. En una realización se usa luz UV para escindir de manera fotoquímica los restos y ligadores fotoquímicamente escindibles. En una realización, el ligador fotoescindible es un resto 2-nitrobencilo.

El grupo $-CH_2N_3$ puede tratarse con TCEP (tris(2-carboxietil)fosfina) para retirarlo a partir del átomo de 3'-O de un análogo de dNPP, o análogo de rNPP, creando así un grupo 3'-OH.

15 Una etiqueta puede liberarse de cualquier manera. En algunos casos, la etiqueta está unida a polifosfato (por ejemplo, figura 34) y la incorporación del nucleótido en una molécula de ácido nucleico da como resultado la liberación de un polifosfato que tiene la etiqueta unida al mismo. La incorporación puede catalizarse por al menos una polimerasa, opcionalmente unida al nanoporo. En algunos casos, al menos una enzima fosfatasa también está unida al poro. La enzima fosfatasa puede escindir la etiqueta a partir del polifosfato para liberar la etiqueta. En algunos casos, las enzimas fosfatasa están posicionadas de tal manera que el pirofosfato producido mediante la polimerasa en una reacción de polimerasa interacciona con las enzimas fosfatasa antes de entrar en el poro.

20 En algunos casos, la etiqueta no está unida a polifosfato (véase, por ejemplo, la figura 34). En estos casos, la etiqueta está unida mediante un ligador escindible (X). Se dan a conocer métodos para la producción de análogos de nucleótido con grupos reactivos ocupados de manera escindible y/o unidos de manera escindible en la patente estadounidense n.º 6.664.079, que se incorpora en el presente documento como referencia.

25 El ligador puede ser cualquier ligador adecuado y escindirse de cualquier manera adecuada. Por ejemplo, los ligadores pueden ser fotoescindibles. En una realización se usa luz que no daña al ADN para escindir de manera fotoquímica restos y ligadores escindibles de manera fotoquímica. En una realización, el ligador fotoescindible es un resto 2-nitrobencilo. En otra realización, el grupo $-CH_2N_3$ puede tratarse con TCEP (tris(2-carboxietil)fosfina) para retirarlo a partir del átomo de 3'-O de un análogo de dNPP, o análogo de rNPP, creando así un grupo 3'-OH.

30 Un "residuo de nucleótido" es un único nucleótido en el estado en el que existe después de incorporarse en, y convertirse de ese modo en un monómero de, un polinucleótido. Por tanto, un residuo de nucleótido es un monómero de nucleótido de un polinucleótido, por ejemplo ADN, que está unido a un monómero de nucleótido adyacente del polinucleótido a través de un enlace fosfodiéster en la posición 3' de su azúcar y está unido a un segundo monómero de nucleótido adyacente a través de su grupo fosfato, con las excepciones de que (i) un residuo de nucleótido 3'-terminal sólo está unido a un monómero de nucleótido adyacente del polinucleótido mediante un enlace fosfodiéster a partir de su grupo fosfato, y (ii) un residuo de nucleótido 5'-terminal sólo está unido a un monómero de nucleótido adyacente del polinucleótido mediante un enlace fosfodiéster a partir de la posición 3' de su azúcar.

35 Debido a reglas de apareamiento de bases bien entendidas, determinar la identidad (de la base) de análogo de dNPP (o análogo de rNPP) incorporado en un cebador o producto de extensión de ADN (o producto de extensión de ARN) midiendo la señal eléctrica única de la etiqueta que se transloca a través del nanoporo, y de ese modo la identidad del análogo de dNPP (o análogo de rNPP) que se incorporó, permite la identificación del residuo de nucleótido complementario en el polinucleótido monocatenario al que se hibrida el cebador o producto de extensión de ADN (o producto de extensión de ARN). Por tanto, si el análogo de dNPP que se incorporó comprende una adenina, una timina, una citosina o una guanina, entonces el residuo de nucleótido complementario en el ADN monocatenario se identifica como una timina, una adenina, una guanina o una citosina, respectivamente. La purina adenina (A) se empareja con la pirimidina timina (T). La pirimidina citosina (C) se empareja con la purina guanina (G). De manera similar, con respecto al ARN, si el análogo de rNPP que se incorporó comprende una adenina, un uracilo, una citosina o una guanina, entonces el residuo de nucleótido complementario en el ARN monocatenario se identifica como un uracilo, una adenina, una guanina o una citosina, respectivamente.

40 La incorporación en un oligonucleótido o polinucleótido (tal como un cebador o cadena de extensión de ADN) de un análogo de dNPP o rNPP significa la formación de un enlace fosfodiéster entre el átomo de carbono en 3' del residuo de nucleótido 3'-terminal del polinucleótido y el átomo de carbono en 5' del análogo de dNPP o análogo de rNPP, respectivamente.

45 Tal como se usa en el presente documento, a menos que se especifique lo contrario, una base (por ejemplo de un análogo de polifosfato de nucleótido) que es diferente del tipo de base de una molécula a la que se hace referencia, por ejemplo otro análogo de polifosfato de nucleótido, significa que la base tiene una estructura química diferente de la base o bases distintas/de referencia. Por ejemplo, una base que es diferente de adenina puede incluir una base

que es guanina, una base que es uracilo, una base que es citosina y una base que es timina. Por ejemplo, una base que es diferente de adenina, timina y citosina puede incluir una base que es guanina y una base que es uracilo.

5 Tal como se usa en el presente documento, a menos que se especifique lo contrario, una etiqueta (por ejemplo de un análogo de polifosfato de nucleótido) que es diferente del tipo de etiqueta de una molécula a la que se hace referencia, por ejemplo otro análogo de polifosfato de nucleótido, significa que la etiqueta tiene una estructura química diferente de la estructura química de la etiqueta o etiquetas diferentes/a las que se hace referencia.

10 Las etiquetas pueden fluir a través de un nanoporo después de liberarse a partir del nucleótido. En algunos casos, se aplica una tensión para tirar de las etiquetas a través del nanoporo. Al menos aproximadamente el 85%, al menos el 90%, al menos el 95%, al menos el 99%, al menos el 99,9 o al menos el 99,99% de las etiquetas liberadas pueden translocarse a través del nanoporo.

15 En algunos casos del método, una polimerasa extrae a partir de una combinación de nucleótidos marcados con etiqueta que comprende una pluralidad de bases diferentes (por ejemplo, A, C, G, T y/o U). También es posible poner en contacto de manera iterativa la polimerasa con los diversos tipos de bases marcadas con etiqueta. En este caso, puede no ser necesario que cada tipo de nucleótido tenga una base única, pero la realización de ciclos entre diferentes tipos de base añade coste y complejidad al procedimiento en algunos casos, no obstante esta realización queda abarcada en la presente invención.

20 La figura 9 muestra que la incorporación del nucleótido marcado con etiqueta en una molécula de ácido nucleico (por ejemplo, usando una polimerasa para extender una base de cebador emparejada a un molde) libera una etiqueta detectable-polifosfato. En algunos casos, el polifosfato de etiqueta se detecta a medida que pasa a través del nanoporo. Incluso es posible distinguir el nucleótido basándose en el número de fosfatos que comprende el polifosfato (por ejemplo, aunque las etiquetas sean idénticas). No obstante, cada tipo de nucleótido tiene generalmente una etiqueta única.

Con referencia a la figura 28, el compuesto de polifosfato de etiqueta puede tratarse con fosfatasa (por ejemplo, fosfatasa alcalina) antes de hacer pasar la etiqueta a través de un nanoporo y medir la corriente iónica.

25 La etiqueta puede detectarse en el nanoporo (al menos en parte) debido a su carga. En algunos casos, el compuesto de etiqueta es un compuesto cargado de manera alternativa que tiene una primera carga neta y, después de una reacción química, física o biológica, una segunda carga neta diferente. En algún caso, la magnitud de la carga en la etiqueta es la misma que la magnitud de la carga en el resto del compuesto. En una realización, la etiqueta tiene una carga positiva y la retirada de la etiqueta cambia la carga del compuesto.

30 En algunos casos, a medida que la etiqueta pasa a través del nanoporo, puede generar un cambio electrónico. En algunos casos el cambio electrónico es un cambio en la amplitud de corriente, un cambio en la conductancia del nanoporo o cualquier combinación de los mismos.

35 “Nanoporo” tal como se usa en el presente documento, se refiere de manera general a un poro, canal o paso formado o proporcionado de otra manera en una barrera/membrana. “Nanoporo” incluye, por ejemplo, una estructura que comprende (a) unos compartimentos primero y segundo separados por una barrera física, barrera que tiene al menos un poro con un diámetro, por ejemplo, de desde aproximadamente 1 hasta 10 nm, y (b) unos medios para aplicar un campo eléctrico a través de la barrera de modo que una molécula cargada tal como ADN, nucleótido, análogo de nucleótido o etiqueta, puede pasar desde el primer compartimento a través del poro hasta el segundo compartimento. De manera ideal el nanoporo comprende además unos medios para medir la firma electrónica de una molécula que pasa a través de su barrera. La barrera de nanoporo puede ser sintética o producirse en parte de manera natural. Una barrera/membrana puede ser una membrana orgánica, tal como una bicapa lipídica, o una membrana sintética, tal como una membrana formada por un material polimérico. Las barreras pueden incluir, por ejemplo, bicapas lipídicas que tienen en las mismas alfa-hemolisina, canales de proteína oligomérica, tales como porinas, y péptidos sintéticos y similares. Las barreras también pueden incluir placas inorgánicas que tienen uno o más orificios de un tamaño adecuado. El nanoporo puede estar dispuesto adyacente a, o en proximidad de, un circuito de detección, tal como, por ejemplo, un circuito de semiconductor de óxido de metal complementario (CMOS) o 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. El nanoporo puede ser biológico o sintético. También se contempla que el poro sea proteico, alfa-hemolisina es un ejemplo de un nanoporo de proteína. Un ejemplo de un nanoporo sintético es un poro de estado sólido o grafeno. En el presente documento “nanoporo”, “barrera de nanoporo” y “poro” en la barrera de nanoporo se usan algunas veces de manera equivalente.

45 En algunos casos, se unen enzimas polimerasa y/o enzimas fosfatasa al nanoporo. Las proteínas de fusión o reticulaciones de disulfuro son ejemplos de métodos para unirse a un nanoporo proteico. En el caso de un nanoporo de estado sólido, la unión a la superficie cerca del nanoporo puede realizarse mediante uniones de biotina-estreptavidina. En un ejemplo la ADN polimerasa se une a una superficie sólida a través de una superficie de oro modificada con una monocapa autoensamblada de alcanotiol funcionalizada con grupos amino, en la que los grupos amino están modificados para dar ésteres de NHS para su unión a grupos amino en la ADN polimerasa.

55 En el presente documento se describen métodos, dispositivos y sistemas para secuenciar ácidos nucleicos usando

un nanoporo. Los métodos pueden detectar con precisión acontecimientos de incorporación de nucleótido individuales, tal como tras la incorporación de un nucleótido en una cadena en crecimiento que es complementario a un molde. Una enzima (por ejemplo, ADN polimerasa) puede incorporar nucleótidos a una cadena de polinucleótido en crecimiento, en la que el nucleótido añadido es complementario a la cadena de ácido nucleico de molde correspondiente que se hibrida a la cadena en crecimiento (por ejemplo, reacción en cadena de la polimerasa (PCR)). Estos acontecimientos de incorporación de nucleótido liberan etiquetas a partir de los nucleótidos que pasan a través de un nanoporo y se detectan. De esta manera, puede identificarse la base incorporada (es decir, A, C, G, T o U) porque se libera una etiqueta única a partir de cada tipo de nucleótido (es decir, A, C, G, T o U).

Pueden detectarse acontecimientos de incorporación de nucleótido en tiempo real (es decir, a medida que se producen) y con la ayuda de un nanoporo. En algunos casos, una enzima (por ejemplo, ADN polimerasa) unida al, o en proximidad del, nanoporo puede facilitar el flujo de una molécula de ácido nucleico a través de un nanoporo. Un acontecimiento de incorporación de nucleótido, o la incorporación de una pluralidad de nucleótidos, puede liberar una o más moléculas de etiqueta (también "etiquetas" en el presente documento), que pueden detectarse mediante un nanoporo a medida que las etiquetas fluyen a través del nanoporo. En algunos casos, una enzima unida al, en proximidad del, nanoporo puede ayudar en la detección de etiquetas u otros subproductos liberados tras la incorporación de uno o más nucleótidos.

Los métodos descritos en el presente documento pueden ser métodos de moléculas individuales. Es decir, la señal que se detecta se genera mediante una única molécula (es decir, incorporación de nucleótidos individuales) y no se genera a partir de una pluralidad de moléculas clonales. El método puede no requerir amplificación de ADN.

Pueden producirse acontecimientos de incorporación de nucleótido a partir de una mezcla que comprende una pluralidad de nucleótidos (por ejemplo, trifosfato de desoxirribonucleótido (dNTP en el que N es adenosina (A), citidina (C), timidina (T), guanosina (G) o uridina (U)). Los acontecimientos de incorporación de nucleótido no se producen necesariamente a partir de una disolución que comprende un único tipo de nucleótido (por ejemplo, dATP). Los acontecimientos de incorporación de nucleótido no se producen necesariamente a partir de disoluciones alternantes de una pluralidad de nucleótidos (por ejemplo, dATP, seguido por dCTP, seguido por dGTP, seguido por dTTP, seguido por dATP).

Los sistemas y dispositivos de nanoporos de la presente divulgación pueden combinarse con, o modificarse mediante, otros dispositivos de nanoporos, tales como los descritos en las patentes estadounidenses n.ºs 7.005.264 B2; 7.846.738; 6.617.113; 6.746.594; 6.673.615; 6.627.067; 6.464.842; 6.362.002; 6.267.872; 6.015.714; 5.795.782; y las publicaciones de patente estadounidense n.ºs 2004/0121525, 2003/0104428 y 2003/0104428.

En una realización de las moléculas y los métodos dados a conocer en el presente documento la etiqueta está unida al resto de la molécula mediante un ligador químico que puede escindirse.

En una realización el nanoporo está en una membrana de estado sólido. En una realización la membrana es una membrana de nitruro de silicio. En una realización el nanoporo es un bioporo. En una realización el poro es proteico. En una realización el poro es un poro de alfa-hemolisina. En una realización el poro es un poro de grafeno.

En una realización el ADN, ARN o ácido nucleico monocatenario está ubicado en un lado de la membrana en el que está ubicado el nanoporo y la membrana está ubicada en una disolución de electrolito conductor.

Cuando se proporciona un intervalo de valores, a menos que el contexto indique claramente lo contrario, se entiende que cada número entero intermedio del valor, y cada décima de cada número entero intermedio del valor, a menos que el contexto indique claramente lo contrario, entre el límite superior e inferior de ese intervalo, y cualquier otro valor mencionado o intermedio en ese intervalo mencionado, queda abarcado dentro de la invención. Los límites superior e inferior de estos intervalos más pequeños pueden incluirse independientemente en los intervalos más pequeños, y también quedan abarcados dentro de la invención, sujetos a cualquier límite específicamente excluido en el intervalo mencionado. Cuando el intervalo mencionado incluye uno o ambos de los límites, los intervalos que excluyen (i) cualquiera o (ii) ambos de esos límites incluidos también quedan incluidos en la invención.

Todas las combinaciones de los diversos elementos descritos en el presente documento están dentro del alcance de la invención. Todas las subcombinaciones de los diversos elementos descritos en el presente documento también están dentro del alcance de la invención.

Los métodos para secuenciar ácidos nucleicos pueden incluir recuperar una muestra biológica que tiene el ácido nucleico que va a secuenciarse, extraer o aislar de otro modo la muestra de ácido nucleico a partir de la muestra biológica, y en algunos casos preparar la muestra de ácido nucleico para la secuenciación.

En el presente documento se proporcionan sistemas y métodos para secuenciar una molécula de ácido nucleico con la ayuda de un nanoporo. El nanoporo puede formarse o incorporarse de otra forma en una membrana dispuesta adyacente a un circuito de detección, tal como un transistor de efecto de campo o un semiconductor de óxido de metal complementario (CMOS). En algunos casos, a medida que un ácido nucleico o etiqueta fluye a través del nanoporo, el circuito de detección detecta una señal eléctrica asociada con el ácido nucleico o etiqueta. El ácido nucleico puede ser una subunidad de una cadena más grande. La etiqueta puede ser un subproducto de un

acontecimiento de incorporación de nucleótido u otra interacción entre un ácido nucleico marcado con etiqueta y el nanoporo o una especie adyacente al nanoporo, tal como una enzima que escinde una etiqueta a partir de un ácido nucleico.

5 Los subproductos de acontecimientos de incorporación de nucleótido pueden detectarse mediante el nanoporo. "Acontecimientos de incorporación de nucleótido" son la incorporación de un nucleótido en una cadena de polinucleótido en crecimiento. Un subproducto puede correlacionarse con la incorporación de un nucleótido de tipo dado. Los acontecimientos de incorporación de nucleótido se catalizan generalmente mediante una enzima, tal como ADN polimerasa, y usan interacciones de pares de bases con una molécula de molde para elegir entre los nucleótidos disponibles para su incorporación en cada ubicación.

10 En algunos casos, la ADN polimerasa es polimerasa 9°N o una variante de la misma, ADN polimerasa I de *E. Coli*, ADN polimerasa de bacteriófago T4, Sequenase, ADN polimerasa Taq, polimerasa 9°N (exo-)A485L/Y409V o ADN polimerasa de Phi29 (ADN polimerasa de ϕ 29).

15 Una muestra de ácido nucleico puede secuenciarse usando nucleótidos o análogos de nucleótido marcados con etiqueta. En algunos ejemplos, un método para secuenciar una molécula de ácido nucleico comprende (a) polimerizar nucleótidos marcados con etiqueta, en el que una etiqueta asociada con un nucleótido individual se libera tras la polimerización, y (b) detectar la etiqueta liberada con la ayuda de un nanoporo.

20 La tasa de acontecimientos de incorporación de nucleótido es generalmente más lenta que (o igual a) la tasa a la que moléculas de etiquetas liberadas durante los acontecimientos de incorporación de nucleótido pasan a través del, y/o se detectan mediante el, nanoporo. Generalmente, la tasa de acontecimientos de incorporación de nucleótido no es mayor que la tasa a la que moléculas de etiquetas liberadas durante los acontecimientos de incorporación de nucleótido pasan a través del, y/o se detectan mediante el, nanoporo (es decir, de lo contrario los acontecimientos de incorporación de nucleótido no se detectan con precisión y/o en la secuencia correcta).

25 En algunos casos, se libera una única etiqueta tras la incorporación de un único nucleótido y se detecta mediante un nanoporo. En otros casos, se libera una pluralidad de etiquetas tras la incorporación de una pluralidad de nucleótidos. Un sensor de nanoporo adyacente a un nanoporo puede detectar una etiqueta liberada individual, o una pluralidad de etiquetas liberadas. Una o más señales asociadas con una pluralidad de etiquetas liberadas pueden detectarse y procesarse para producir una señal promediada.

30 Los métodos proporcionados en el presente documento pueden distinguir con precisión entre acontecimientos de incorporación de nucleótido individuales (por ejemplo, acontecimientos de moléculas individuales). Los métodos pueden distinguir con precisión entre acontecimientos de incorporación de nucleótido individuales en un único pase, sin tener que volver a secuenciar una molécula de ácido nucleico dada.

35 Un método para la secuenciación de ácidos nucleicos comprende distinguir entre acontecimientos de incorporación de nucleótido individuales con una precisión de más de aproximadamente 4σ . En algunos casos, los acontecimientos de incorporación de nucleótido se detectan con la ayuda de un nanoporo. Pueden liberarse etiquetas asociadas con los nucleótidos tras la incorporación y las etiquetas pasan a través del nanoporo. Una etiqueta diferente puede estar asociada con, y/o liberarse a partir de, cada tipo de nucleótido (por ejemplo, A, C, T, G) y se detecta a medida que pasa a través del nanoporo. Los errores incluyen, pero no se limitan a, (a) no lograr detectar una etiqueta, (b) identificar erróneamente una etiqueta, (c) detectar una etiqueta cuando no hay ninguna etiqueta, (d) detectar etiquetas en el orden incorrecto (por ejemplo, dos etiquetas se liberan en un primer orden, pero se adelantan entre sí y se detectan en un segundo orden), (e) se detecta como si se hubiera liberado una etiqueta que no se ha liberado a partir de un nucleótido, o cualquier combinación de los mismos. En algunas realizaciones, la precisión de distinción entre acontecimientos de incorporación de nucleótido individuales es del 100% menos la tasa a la que se producen errores (es decir, tasa de error).

45 La precisión de distinción entre acontecimientos de incorporación de nucleótido individuales es cualquier porcentaje adecuado. En algunos casos, la precisión de distinción entre acontecimientos de incorporación de nucleótido individuales se notifica en unidades sigma (σ). Sigma es una variable estadística que se usa algunas veces en gestión de empresas y estrategia de fabricación para notificar tasas de error tales como el porcentaje de productos libres de defectos. En este caso, los valores de sigma pueden usarse de manera intercambiable con la precisión según la siguiente relación: 4σ es una precisión del 99,38%, 5σ es una precisión del 99,977%, y 6σ es una precisión del 99,99966%.

El método puede implicar secuenciar una cadena de ácido nucleico de molde añadiendo nucleótidos marcados con etiqueta a una cadena complementaria a la cadena de molde y detectando moléculas de etiqueta liberadas en un nanoporo. Los métodos dados a conocer en el presente documento pueden combinarse con otros métodos de secuenciación, tales como, por ejemplo, los descritos en la patente estadounidense n.º 5.470.724.

55 Esta invención se entenderá mejor mediante referencia a los siguientes detalles experimentales, pero los expertos en la técnica apreciarán fácilmente que los experimentos específicos detallados sólo son ilustrativos de la invención tal como se describe más completamente en las siguientes reivindicaciones más adelante.

Otro aspecto de la presente divulgación proporciona un sistema de medición de conductancia que comprende: (a) unos compartimentos primero y segundo con unas disoluciones de electrolito primera y segunda separados por una barrera física, barrera que tiene al menos un poro con un diámetro a escala nanométrica; (b) unos medios para aplicar un campo eléctrico a través de la barrera; (c) unos medios para medir un cambio en el campo eléctrico; (d) al menos una polimerasa unida al poro; y (e) más de una enzima fosfatasa unida al poro.

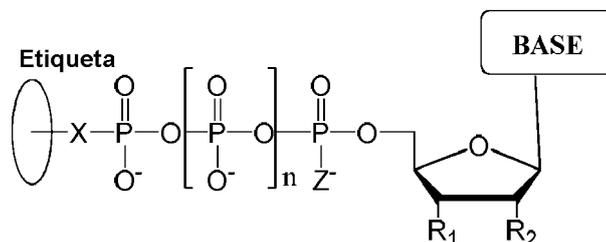
En algunos casos, el poro tiene un diámetro de desde aproximadamente 1 hasta 10 nm. En algunos casos, la polimerasa y las enzimas fosfatasa están unidas covalentemente al poro. En algunos casos, más enzimas fosfatasa que polimerasas están unidas al poro. En algunos casos, las enzimas fosfatasa están posicionadas de tal manera que polifosfato producido mediante la polimerasa en una reacción de polimerasa interacciona con las enzimas fosfatasa antes de entrar en el poro. En algunos casos, la tasa de interacción entre las enzimas fosfatasa y el polifosfato es más rápida que, o igual a, la tasa de la polimerasa que produce el polifosfato.

En algunos casos, cada uno de los compartimentos primero y segundo tiene una carga eléctrica. En algunos casos, el interior del poro tiene una carga negativa. En algunos casos, el poro es biológico o sintético. En algunos casos, el poro es proteico. En algunos casos, el poro es una proteína de alfa-hemolisina. En algunos casos, el poro es un poro de estado sólido. En algunos casos, el poro está formado por grafito.

En algunos casos, el sistema de medición de conductancia comprende además una matriz de poros que tienen cada uno características sustancialmente idénticas. En algunos casos, el sistema de medición de conductancia comprende además una matriz de poros de diámetros diferentes. En algunos casos, el sistema de medición de conductancia comprende además una matriz de poros, en el que los poros están configurados para aplicar campos eléctricos diferentes a través de la barrera.

En algunos casos, el sistema de medición de conductancia está integrado con electrónica de CMOS. En algunos casos, el poro o la matriz de poros está integrado directamente en una pastilla de CMOS tal como se muestra en la figura 46.

Otro aspecto de la presente divulgación proporciona un compuesto que tiene la estructura:



en la que la etiqueta comprende uno o más de etilenglicol, un aminoácido, un hidrato de carbono, un péptido, un colorante, un compuesto quimioluminiscente, un mononucleótido, un dinucleótido, un trinucleótido, un tetranucleótido, un pentanucleótido, un hexanucleótido, un ácido alifático, un ácido aromático, un alcohol, un grupo tiol, un grupo ciano, un grupo nitro, un grupo alquilo, un grupo alquenilo, un grupo alquinilo, un grupo azido o una combinación de los mismos, en la que R_1 es OH, en la que R_2 es H u OH, en la que X es O, NH, S o CH_2 , en la que Z es O, S o BH_3 , en la que la base es adenina, guanina, citosina, timina, uracilo o un derivado de una de esas bases, en la que n es 1, 2, 3 ó 4, y en la que la etiqueta tiene una carga que es de signo inverso con respecto a la carga en el resto del compuesto.

En algunos casos, la magnitud de la carga en la etiqueta es la misma que la magnitud de la carga en el resto del compuesto. En algunos casos, la etiqueta comprende múltiples unidades de etilenglicol. En algunos casos, la etiqueta comprende 16, 20, 24 ó 36 unidades de etilenglicol. En algunos casos, la etiqueta comprende un resto identificable adicional. En algunos casos, el resto identificable adicional es un colorante basado en cumarina. En algunos casos, la etiqueta tiene una carga positiva. En algunos casos, la retirada de la etiqueta cambia la carga del compuesto. En algunos casos, la etiqueta comprende además un número apropiado de lisinas o argininas para equilibrar el número de fosfatos.

Otro aspecto de la presente divulgación proporciona un método para determinar la secuencia de nucleótidos de un ARN o ADN monocatenario. En algunos casos, más enzimas fosfatasa que polimerasas están unidas al poro. En algunos casos, el ARN o ADN monocatenario se obtiene desnaturalizando un ARN o ADN bicatenario, lo que sea aplicable. En algunos casos, múltiples copias del mismo ARN o ADN monocatenario están inmovilizadas en una perla. En algunos casos, la secuencia de nucleótidos del ARN o ADN monocatenario se determina usando múltiples copias del mismo ARN o ADN monocatenario. En algunos casos, el método comprende además una etapa de lavado tras cada iteración de la etapa (b) para retirar compuesto no incorporado del contacto con el ARN o ADN monocatenario. En algunos casos, el método comprende además una etapa tras cada iteración de la etapa (b) para determinar la identidad de un resto identificable adicional unido a la etiqueta. En algunos casos, al menos el 85-99% de las etiquetas liberadas se translocan a través del poro. En algunos casos, el compuesto comprende además un terminador reversible. En algunos casos, el método comprende además una etapa de retirar el terminador reversible

tras cada iteración de la etapa (b). En algunos casos, el terminador reversible se retira mediante medios biológicos, medios químicos, medios físicos o mediante irradiación de luz. En algunos casos, el interior del poro tiene una carga que es de signo inverso con respecto a la carga de la etiqueta o del polifosfato que tiene la etiqueta unida al mismo. En algunos casos, cada uno de los compartimentos primero y segundo del sistema de medición de conductancia tiene una carga. En algunos casos, las cargas de los compartimentos primero y segundo son de polaridad opuesta. En algunos casos, las cargas de los compartimentos primero y segundo son ajustables. En algunos casos, la tasa de translocación de etiqueta a través del poro en la etapa (b) se determina basándose en la carga de etiqueta y las cargas de los compartimentos primero y segundo. En algunos casos, cada uno de los compartimentos primero y segundo tiene una carga de tal manera que en la etapa (b) la etiqueta se transloca a través del poro a una tasa que es más rápida que, o igual a, la tasa a la que la etiqueta o el polifosfato que tiene la etiqueta unida al mismo está liberándose en la etapa (a).

Otro aspecto de la presente divulgación proporciona un método para determinar al menos un parámetro de un compuesto en una disolución. En algunos casos, el compuesto se trata con fosfatasa antes de medir la reducción en la corriente iónica. En algunos casos, el compuesto es un compuesto cargado de manera alternativa que tiene una primera carga neta y, después de una reacción química, física o biológica, una segunda carga neta diferente. En algunos casos, el análisis matemático se selecciona del grupo que consiste en GMM, detección y cálculo de promedio de umbral, y análisis de ventana deslizante. En algunos casos, el al menos un parámetro se selecciona del grupo que consiste en la concentración, el tamaño, la carga y la composición del compuesto.

En algunos casos, el método comprende además una etapa de calibrar el sistema de medición de conductancia. En algunos casos, la precisión es mayor de 4σ . En algunos casos, la precisión es mayor de 5σ . En algunos casos, la precisión es mayor de 6σ .

Aspectos describen métodos y sistemas de medición de conductancia que implican disoluciones de electrolito primera y segunda y/o fluidos primero y segundo. En algunos casos, las disoluciones de electrolito primera y segunda son la misma. En algunos casos, el primer fluido y el segundo fluido son el mismo.

25 Detección en nanoporos y etiquetas

La invención dada a conocer en el presente documento se refiere a nucleótidos modificados para el análisis de moléculas individuales de ADN (o ARN, cambiando lo que sea necesario) usando nanoporos. Pueden realizarse modificaciones en diversas posiciones de un nucleótido, es decir el fosfato terminal, la base y/o el 2' o 3'-OH para formar un análogo de nucleótido. Tras una reacción de extensión de polimerasa en un complejo de molde-cebador, el pirofosfato unido a etiqueta liberada pasa a través de un nanoporo y el bloqueo de corriente resultante se monitoriza para determinar la base de nucleótido añadida. Si la modificación o etiqueta está en el resto de base, o el 2'/3'-OH del resto de azúcar del nucleótido, entonces tras la incorporación mediante ADN/ARN polimerasa, el ligador-etiqueta se escinde a partir de la base/azúcar mediante medios químicos o fotoquímicos y el ligador-etiqueta liberado pasa a través de un nanoporo para identificar el nucleótido añadido.

Se diseñan y sintetizan 5'-polifosfatos de nucleósido que portan un número diferente de grupos fosfato como ligadores y modificados con etiquetas unidas al fosfato terminal de los nucleótidos. Tras la incorporación mediante ADN/ARN polimerasa en una reacción de extensión de molde-cebador, el polifosfato unido a etiqueta liberada (di, tri, tetra, penta, etc.) puede detectarse usando un nanoporo para producir datos de secuencia. Opcionalmente, los polifosfatos de etiqueta liberada también pueden tratarse con fosfatasa alcalina para proporcionar etiquetas libres. Usando cuatro etiquetas diferentes que son distintas y específicas para cada base de nucleótido, puede determinarse la secuencia del ADN o ARN de molde.

Se proporcionan nucleótidos que portan un número diferente de grupos fosfato o etiquetas para la síntesis de nucleótidos modificados, que son sustratos eficientes en reacciones de polimerasa. El polifosfato unido a etiqueta liberada se detecta usando un nanoporo para determinar condiciones para el diseño y modificación de los nucleótidos para lograr señales de bloqueo diferenciadas.

También se proporcionan nucleótidos que portan un ligador-etiqueta unido al resto de base de nucleótido, y/o el 2'/3'-OH del resto de azúcar, para la reacción de ADN polimerasa para generar un producto de extensión de ADN de una única base marcado con ligador-etiqueta. Estos nucleótidos son buenos sustratos para ADN/ARN polimerasas habitualmente usadas. El ligador-etiqueta unido al producto de ADN extendido se escinde mediante medios químicos o fotoquímicos para generar el cebador listo para su extensión adicional usando los nucleótidos modificados. El ligador-etiqueta liberado se hace pasar a través de un nanoporo y se identifica basándose en la diferencia de tamaño, forma y carga en la etiqueta para producir datos de secuencia.

Tal como se da a conocer en el presente documento, estas herramientas moleculares facilitan la secuenciación de moléculas individuales usando nanoporos a una resolución de base individual.

Aquí se dan a conocer diversas mejoras para el enfoque de nanoporos: 1) lograr una distinción precisa y evidente de las cuatro bases (A, C, G y T) que constituyen las moléculas de ácido nucleico; 2) potenciar y diferenciar la intensidad de las señales de detección; 3) desarrollar un método eficaz para distinguir y procesar las señales de bloqueo electrónico generadas; 4) controlar la tasa de translocación de ácidos nucleicos a través del poro, tal como

ralentizar el movimiento de etiquetas para mejorar la capacidad de distinción entre bases; y 5) diseñar y realizar nanoporos sintéticos nuevos y más eficaces para diferenciar los cuatro nucleótidos diferentes en ADN.

Las estructuras de cuatro nucleótidos se muestran en la figura 22. A y G son purinas, mientras que C y T son pirimidinas. El tamaño molecular global de A y G es muy similar, mientras que el tamaño de C y T es similar. Se ha mostrado que los nanoporos pueden diferenciar entre purinas y pirimidinas (Akeson *et al.* 1999 y Meller *et al.* 2000), pero no pueden distinguir entre purinas individuales, A y G, o entre pirimidinas individuales, C y T.

Estudios anteriores han mostrado modificaciones de 5'-trifosfatos de nucleósido, incluyendo introducir más grupos fosfato para producir tetra, penta o hexa-fosfatos, introducir colorante directamente en el fosfato terminal, o unir un ligador entre el fosfato terminal y el colorante (Kumar *et al.*, 2006 y 2008). Los tetra y penta-fosfatos son mejores sustratos de ADN polimerasa, y se han desarrollado nucleótidos de hexa-fosfato marcados con colorante (Kumar *et al.* 2005; Sood *et al.* 2005; Eid *et al.* 2009).

En el presente documento se dan a conocer análogos de nucleótido que están diseñados para potenciar la distinción de cada nucleótido mediante modificación de los nucleótidos en el resto de fosfato terminal. Se sintetizan 5'-polifosfatos de nucleósido y se unen diferentes etiquetas (tales como polietilenglicoles (PEG) de diferente longitud/masa, aminoácidos, hidratos de carbono, oligonucleótidos, colorantes o moléculas orgánicas/inorgánicas) al grupo fosfato terminal. Después de reacciones de extensión de polimerasa, se generan restos de polifosfato unido a etiqueta (figura 9) y se produce una señal diferente específica para cada base cuando los restos de polifosfato unido a etiqueta pasan a través del nanoporo. Estas modificaciones aumentan la distinción de las bases mediante nanoporo debido al aumento de diferencias de tamaño, masa o carga de unidades de polifosfato marcado con etiqueta liberadas entre los cuatro nucleótidos (A, G, C y T).

La tasa de translocación de ADN a través del nanoporo se reduce debido a la voluminosidad de los polifosfatos unidos a etiqueta liberada, aunque no se necesita reducir la tasa de translocación de las etiquetas a través del nanoporo siempre que puedan diferenciarse las etiquetas. Por tanto, se pueden lograr la precisión y fiabilidad requeridas para la secuenciación entre bases. Otros parámetros analíticos en la secuenciación en nanoporos, tales como concentración del polinucleótido, magnitud de la tensión aplicada, temperatura y valor de pH de la disolución, se optimizan con el fin de obtener los resultados más precisos y fiables para la detección y el análisis de cadena de ADN.

Los enfoques de moléculas individuales para secuenciación permiten la posibilidad de derivar haplotipos para estudios genéticos y permitir la secuenciación directa de ARNm. Entre los posibles enfoques de moléculas individuales para decodificar la secuencia de moléculas de ADN o ARN se encuentra el uso de nanoporos biológicos o sintéticos como detectores de las bases de ADN individuales.

El enfoque de secuenciación mediante síntesis (SBS) existente usa terminadores reversibles de nucleótidos fluorescentes escindibles (C_F-NRT) (Guo *et al.* 2010). El método de SBS se basa en la capacidad de detenerse tras cada adición de nucleótido durante la reacción de polimerasa y el uso de fluoróforos específicos para distinguir entre las 4 bases. Sin embargo, una limitación principal de SBS para la secuenciación de moléculas individuales es el requisito de detectores de fluorescencia caros y software de obtención de imágenes rápido. El método y procedimiento dados a conocer en el presente documento aprovechan las ventajas de SBS, especialmente su alta precisión, con la velocidad y sensibilidad del nanoporo como detector de impedancia de corriente iónica.

Aunque se ha investigado mucho sobre la introducción de ADN a través de nanoporos, con la esperanza de distinguir cada base a medida que pasa a través debido a su efecto variable sobre la corriente iónica, esto ha sido muy difícil de lograr, debido tanto a la velocidad de transmisión como al efecto de bases circundantes que pueden contribuir con sus propios efectos sobre iones y contraiones que pasan a través de los poros (Timp *et al.* 2010). El uso de ciclodextrinas u otras estructuras en forma de anillo en la luz de poros de proteína ayuda a proporcionar un mecanismo de trinquete para ralentizar el tiempo de tránsito (Astier *et al.* 2006), pero la capacidad de reconocer de manera absoluta cada base para su secuenciación a medida que pasa sigue suponiendo un desafío. Una estrategia alternativa que usa exonucleasa para permitir que un nucleótido cada vez atraviese el poro ha conducido a la distinción de bases individuales (Clarke *et al.* 2009). Sin embargo, todavía hay dificultades en el control del tiempo de reacción de la exonucleasa para diferentes longitudes de ADN y nucleótido y la velocidad a la que los iones liberados llegan al poro con este enfoque.

La propia reacción de polimerasa presenta una alta procesividad y tasas estables de incorporación de bases. De hecho, se han usado reacciones de polimerasa para controlar el movimiento de cadenas de ADN a través de nanoporos para dirigir la distinción de bases (Benner *et al.* 2007, Cockroft *et al.* 2008, Hurt *et al.* 2009). Durante la reacción de polimerasa, hay una liberación de un resto pirofosfato (PPi). Por tanto, si se une una etiqueta diferente al trifosfato para cada uno de los cuatro nucleótidos, estas pueden distinguirse a medida que se liberan y pasan a través de un nanoporo apropiado para la determinación de secuencia de ADN. Estos análogos de pirofosfato relativamente pequeños, o moléculas equivalentes con grupos cargados de manera positiva adicionales, pueden alcanzar el poro de manera extremadamente rápida. La tasa de incorporación de nucleótidos mediante polimerasas es de aproximadamente 1000 nucleótidos por segundo, es decir un milisegundo por adición de base, mientras que la tasa de transporte a través del nanoporo es de 1 molécula por microsegundo. Por tanto, con ingeniería y fluidica

apropiadas, no hay problemas de desfase para secuenciar ADN con el presente enfoque, ni tampoco hay dificultades con la decodificación de tramos de homopolímeros. Se ha mostrado que puede distinguirse entre un amplio intervalo de tamaño de polietilenglicoles que difieren en tan sólo una o dos unidades de carbono mediante el efecto que tienen sobre corrientes de bloqueo en nanoporos (Reiner *et al.* 2010, Robertson *et al.* 2007), una resolución esencialmente equivalente a la obtenida mediante un espectrómetro de masas. Por tanto, tal como se describe a continuación, se unen cadenas de PEG de longitud diferente al fosfato terminal de dATP, dCTP, dGTP y dTTP. A medida que se incorpora cada nucleótido durante la reacción de polimerasa, un grupo fosfato marcado con etiqueta de manera específica se libera en el nanoporo, produciendo una señal de bloqueo de corriente diferenciada para indicar qué nucleótido se incorpora. La velocidad de secuenciación es extremadamente rápida, limitada únicamente por la tasa de la reacción de polimerasa. Como enfoque alternativo para marcar con etiqueta los nucleótidos, también se usan diferentes longitudes de cadena de fosfato (por ejemplo, tri, tetra y penta-fosfatos).

Adicionalmente, también se usan nanoporos de estado sólido que tienen ventajas en cuanto a un mejor control sobre, y flexibilidad de, fabricación, por tanto garantizan un transporte vectorial rápido de polifosfatos marcados con etiqueta pero no de los precursores de nucleótido o el ADN hacia y a través de los nanoporos o nanocanales. Para lograr esto, se incorporan dos características de diseño importantes. En primer lugar, los precursores (polifosfatos de nucleótido marcado con etiqueta) se sintetizan con una carga neutra global, mientras que los fosfatos marcados con etiqueta escindidos tienen una carga positiva global. Usando una corriente que atrae a iones positivos, los nanoporos sólo necesitan distinguir las cuatro moléculas marcadas con etiqueta liberadas alternativas. Se logra una carga diferencial en precursores y productos incorporando en las etiquetas un número de lisinas o argininas (cargadas de manera positiva) que equilibra exactamente el número de fosfatos (cargados de manera negativa). Tras la incorporación del α -fosfato en el cebador en crecimiento, hay una lisina más que fosfato en el producto liberado. Opcionalmente, puede usarse fosfatasa alcalina para escindir todos los fosfatos para producir una etiqueta de PEG con una carga positiva más fuerte. En segundo lugar, para garantizar que los fosfatos liberados se mueven inmediatamente a través del poro más cercano, la ADN polimerasa se inmoviliza en la entrada del poro, por ejemplo mediante una unión de biotina-estreptavidina. A medida que la cadena de ADN se inserta a través de la polimerasa, los productos marcados con etiqueta liberados sólo tienen que difundirse la misma distancia corta para alcanzar el nanoporo.

También es importante reconocer las ventajas del mecanismo de transducción bioelectrónica con respecto a enfoques ópticos. Para técnicas de transducción óptica de moléculas individuales, la señal a partir de un único fluoróforo es normalmente < 2500 fotones/s (correspondiente a niveles de corriente detectados del orden de 50 fA) a niveles de ruido corto altos, requiriendo una óptica compleja para intentar recopilar cada fotón emitido, haciendo que el ajuste a escala de las plataformas hasta densidades superiores sea difícil. Las reacciones de síntesis deben ralentizarse hasta 1 Hz para permitir tiempos de integración suficientes para estas señales ópticas débiles y con ruido. Los desafíos para las técnicas ópticas han abierto la posibilidad para enfoques de detección bioelectrónica, que tienen niveles de señal significativamente superiores (normalmente más de tres órdenes de magnitud superiores), permitiendo la posibilidad de detección con alto ancho de banda con el diseño conjunto apropiado de transductor, detector y amplificador. Los niveles de señal para nanoporos pueden ser de hasta 100 pA a partir de alfa-hemolisina (Kasianowicz *et al.* 1996), 300 pA para MspA (Derrington *et al.* 2010) y más de 4 nA a partir de nanoporos de estado sólido (Wanunu *et al.* 2010).

Se ha dirigido un esfuerzo significativo hacia el desarrollo de tecnología de nanoporos como mecanismo de transducción bioelectrónica (Benner *et al.* 2007, Deamer *et al.* 2002, Kasianowicz *et al.* 1996, Branton 2008, Branton *et al.* 2008, Chen 2004, Gershow *et al.* 2007, Nealy 2007, Matysiak *et al.* 2006). Dos atributos esenciales de este sensor electrónico le proporcionan sensibilidad de moléculas individuales. El primero es la geometría muy localizada (a escala nanométrica) de la sensibilidad de carga en el propio poro. El diámetro de un poro puede ser de 2-3 nm, y debido a la exploración de carga de electrolito la corriente medida es altamente sensible a fuentes de carga a más de unos pocos nanómetros a partir del poro. En segundo lugar, el sensor de nanoporo proporciona una ganancia mediante el efecto de la carga de movimiento comparativamente lento que tiene un biopolímero sobre una concentración cercana de iones de sal de movilidad superior. Sin embargo, los nanoporos están extremadamente limitados mediante el tiempo relativamente corto que pasan las biomoléculas en la región sensible a carga del poro. Esto se aborda directamente mediante el uso de etiquetas, que pueden optimizarse para producir altos niveles de señal y acontecimientos de translocación más largos. Al mismo tiempo, la integración conjunta de CMOS de estos poros se aprovecha para mejorar drásticamente los anchos de banda de ruido limitado para la detección en un dispositivo de nanoporos. Mediante esta plataforma se soportan poros tanto de estado sólido como biológicos. Esta integración de estado sólido, junto con microfluídica asociada, también permite de manera única el aumento a escala de este diseño para dar matrices más grandes con electrónica integrada para la detección.

Sistemas informáticos

Los métodos y sistemas de secuenciación de ácidos nucleicos de la divulgación pueden regularse con la ayuda de sistemas informáticos. La figura 18 muestra un sistema 1800 que comprende un sistema informático 1801 acoplado a un sistema de secuenciación de ácidos nucleicos 1802. El sistema informático 1801 puede ser un servidor o una pluralidad de servidores. El sistema informático 1801 puede estar programado para regular la preparación y el procesamiento de muestras, y la secuenciación de ácidos nucleicos mediante el sistema de secuenciación 1802. El sistema de secuenciación 1802 puede ser un secuenciador (o detector) basado en nanoporos, tal como se describe

en otra parte en el presente documento.

El sistema informático puede estar programado para implementar los métodos de la invención. El sistema informático 1801 incluye una unidad central de procesamiento (CPU, también “procesador” en el presente documento) 1805, que puede ser un procesador de un único núcleo o de múltiples núcleos, o una pluralidad de procesadores para procesamiento en paralelo. El sistema informático 1801 también incluye una memoria 1810 (por ejemplo, memoria de acceso aleatorio, memoria de sólo lectura, memoria flash), unidad de almacenamiento electrónica 1815 (por ejemplo, disco duro), interfaz de comunicaciones 1820 (por ejemplo, adaptador de red) para comunicarse con uno o más de otros sistemas, y dispositivos periféricos 1825, tales como memoria caché, otra memoria, almacenamiento de datos y/o adaptadores de visualización electrónicos. La memoria 1810, la unidad de almacenamiento 1815, la interfaz 1820 y los dispositivos periféricos 1825 están en comunicación con la CPU 1805 a través de un bus de comunicaciones (líneas continuas), tal como una placa madre. La unidad de almacenamiento 1815 puede ser una unidad de almacenamiento de datos (o repositorio de datos) para almacenar datos. El sistema informático 1801 puede estar operativamente acoplado a una red informática (“red”) con la ayuda de la interfaz de comunicaciones 1820. La red puede ser Internet, una internet y/o extranet, o una intranet y/o extranet que está en comunicación con Internet. La red puede incluir uno o más servidores informáticos, lo cual puede permitir el cálculo distribuido.

Pueden implementarse métodos de la invención a modo de código (o software) ejecutable por máquina (o procesador informático) almacenado en una ubicación de almacenamiento electrónico del sistema informático 1801, tal como, por ejemplo, en la memoria 1810 o la unidad de almacenamiento electrónico 1815. Durante el uso, el código puede ejecutarse por el procesador 1805. En algunos casos, el código puede recuperarse a partir de la unidad de almacenamiento 1815 y almacenarse en la memoria 1810 para su fácil acceso por el procesador 1805. En algunas situaciones, puede excluirse la unidad de almacenamiento electrónico 1815 y se almacenan instrucciones ejecutables por máquina en la memoria 1810.

El código puede compilarse previamente y configurarse para su uso con una máquina que tiene un procesador adaptado para ejecutar el código, o puede compilarse durante el tiempo de ejecución. El código puede suministrarse en un lenguaje de programación que puede seleccionarse para permitir que se ejecute el código de una manera previamente compilada o tal como se compila.

El sistema informático 1801 puede estar adaptado para almacenar información de perfil de usuario, tal como, por ejemplo, un nombre, dirección física, dirección de correo electrónico, número de teléfono, alias de mensajería instantánea (IM), información de educación, información de trabajo, gustos y/o aversiones sociales, y otra información de posible relevancia para el usuario u otros usuarios. Tal información de perfil puede almacenarse en la unidad de almacenamiento 1815 del sistema informático 1801.

Aspectos de los sistemas y métodos proporcionados en el presente documento, tales como el sistema informático 1801, pueden implementarse en la programación. Diversos aspectos de la tecnología pueden considerarse como “productos” o “artículos de fabricación” normalmente en forma de código ejecutable por máquina (o procesador) y/o datos asociados que se portan o implementan en un tipo de medio legible por máquina. Puede almacenarse código ejecutable por máquina en una unidad de almacenamiento electrónico, tal como memoria (por ejemplo, ROM, RAM) o un disco duro. Los medios de tipo de “almacenamiento” pueden incluir cualquiera o la totalidad de la memoria tangible de los ordenadores, procesadores o similares, o módulos asociados de la misma, tales como diversas memorias de semiconductor, unidades de cinta, unidades de disco y similares, lo cual puede proporcionar almacenamiento no transitorio en cualquier momento para la programación de software. La totalidad o partes del software pueden comunicarse a veces a través de Internet o diversas otras redes de telecomunicación. Tales comunicaciones, por ejemplo, pueden permitir la carga del software de un ordenador o procesador a otro, por ejemplo, de un servidor de gestión u ordenador anfitrión a la plataforma informática de un servidor de aplicación. Por tanto, otro tipo de medios que pueden portar los elementos de software incluye ondas ópticas, eléctricas y electromagnéticas, tal como se usan en interfaces físicas entre dispositivos locales, a través de redes cableadas y de línea terrestre óptica y a través de diversos enlaces aéreos. Los elementos físicos que portan tales ondas, tales como enlaces cableados o inalámbricos, enlaces ópticos o similares, también pueden considerarse como medios que portan el software. Tal como se usa en el presente documento, a menos que se restrinja a medios de “almacenamiento” tangibles, no transitorios, términos tales como “medio legible” por ordenador o máquina se refieren a cualquier medio que participa en proporcionar instrucciones a un procesador para su ejecución.

Por tanto, un medio legible por máquina, tal como código ejecutable por ordenador, puede adoptar muchas formas, incluyendo, pero sin limitarse a, un medio de almacenamiento tangible, un medio de onda portadora o medio de transmisión física. Los medios de almacenamiento no volátiles incluyen, por ejemplo, discos ópticos o magnéticos, tales como cualquiera de los dispositivos de almacenamiento en cualquier ordenador o similar, tal como puede usarse para implementar las bases de datos, etc., mostradas en los dibujos. Los medios de almacenamiento volátiles incluyen memoria dinámica, tal como memoria principal de una plataforma informática de este tipo. Los medios de transmisión tangibles incluyen cables coaxiales; hilo de cobre y fibra óptica, incluyendo los hilos que comprenden un bus dentro de un sistema informático. Los medios de transmisión de ondas portadoras pueden adoptar la forma de señales eléctricas o electromagnéticas, u ondas acústicas o luminosas tales como las generadas durante comunicaciones de datos por radiofrecuencia (RF) e infrarrojos (IR). Por tanto, las formas

habituales de medios legibles por ordenador incluyen por ejemplo: un disquete, un disco flexible, disco duro, cinta magnética, cualquier otro medio magnético, un CD-ROM, DVD o DVD-ROM, cualquier otro medio óptico, cinta de papel de tarjetas perforadas, cualquier otro medio de almacenamiento físico con patrones de orificios, una RAM, una ROM, una PROM y EPROM, una FLASH-EPROM, cualquier otro cartucho o chip de memoria, una onda portadora que transporta datos o instrucciones, cables o enlaces que transportan una onda portadora de este tipo, o cualquier otro medio a partir del cual un ordenador puede leer código de programación y/o datos. Muchas de estas formas de medios legibles por ordenador pueden estar implicadas en transportar una o más secuencias de una o más instrucciones a un procesador para su ejecución.

Los sistemas y métodos proporcionados en el presente documento pueden combinarse con, o modificarse por, otros sistemas y métodos, tales como, por ejemplo, los sistemas y métodos descritos en la publicación de patente PCT n.º WO/2012/083249.

Ejemplos

EJEMPLO 1

I. Diseño y síntesis de nucleótidos modificados

El efecto de la voluminosidad del polifosfato marcado con etiqueta sobre señales de bloqueo electrónico generadas mediante un nanoporo se determina usando diversos nucleótidos unidos a fosfato con etiquetas o grupos de diferente tamaño unidos al fosfato terminal del nucleótido. En la figura 23 se muestran estructuras de cuatro 5'-polifosfatos de nucleósido marcados con etiqueta en fosfato. En primer lugar, se sintetiza una serie de 5'-tri, tetra, penta y hexa-fosfatos de nucleósido. En estos nucleótidos, el fosfato terminal está unido a un ligador a través del cual se unen diferentes etiquetas, por ejemplo etilenglicoles de diferente longitud y masa u otras moléculas que aumentan la voluminosidad o carga del polifosfato liberado. Estos nucleótidos se someten a prueba con nanoporos para determinar qué etiquetas o grupos voluminosos unidos al fosfato terminal se correlacionan con una diferencia más drástica en la señal de bloqueo electrónico entre las diferentes bases.

1) Polifosfatos de nucleósido modificados en fosfato terminal

a. 5'-Trifosfatos de nucleósido marcados con etiqueta en fosfato terminal

Tal como se muestra en la figura 24, pueden sintetizarse 5'-trifosfatos de nucleósido marcados con etiqueta en fosfato terminal haciendo reaccionar el dNTP correspondiente con DCC/DMF para dar trimetafosfato cíclico que puede abrirse con nucleófilos apropiados para dar 5'-trifosfato de nucleósido unido a etiqueta o ligador. Esto puede usarse en una reacción de extensión de molde-cebador y el pirofosfato unido a etiqueta liberada puede leerse usando nanoporos. Alternativamente, el ligador unido al fosfato puede hacerse reaccionar con éster de etiqueta-NHS para proporcionar un 5'-trifosfato de nucleósido unido a etiqueta alternativo.

b. 5'-Tetrafosfatos de nucleósido marcados con etiqueta en fosfato terminal

Para la síntesis de 5'-tetrafosfatos de nucleósido marcados con etiqueta en fosfato terminal, en primer lugar se hace reaccionar el trifosfato correspondiente con CDI en DMF para activar el grupo fosfato terminal que después se hace reaccionar con ácido fosfórico o monofosfato de etiqueta para dar el tetrafosfato (figura 25). El fosfato terminal en el tetrafosfato puede activarse adicionalmente con CDI seguido por reacción con nucleófilos apropiados para proporcionar un tetrafosfato unido a ligador que puede usarse adicionalmente para unir etiquetas de diferente masa, longitud o voluminosidad, tales como éster de m-dPEG-NHS, también mostrado en la figura 25.

c. 5'-Penta y hexafosfatos de nucleósido marcados con etiqueta en fosfato terminal

La síntesis de 5'-penta y hexafosfatos de nucleósido marcados con etiqueta en fosfato terminal sigue el mismo principio tal como se muestra en la figura 26. Pueden prepararse a partir de cualquiera de trifosfatos o tetrafosfatos activados mediante reacción con ácido fosfórico, pirofosfato o fosfatos unidos a etiqueta. Alternativamente, puede unirse un ligador a penta o hexa-fosfato seguido por reacción con ésteres de NHS activados.

d. Polifosfatos de nucleósido unidos a oligo-etiqueta

Hay varios problemas con el enfoque actual para la secuenciación en nanoporos tales como el reconocimiento de las bases a medida que pasan a través del nanoporo y la velocidad o tasa de transporte para permitir que se registre el reconocimiento de la nucleobase. El ADN pasa a través de un nanoporo de alfa-hemolisina a una tasa de 1-5 μ s, lo cual es demasiado rápido como para registrarse para experimentos de secuenciación de moléculas individuales. Se ha realizado algún avance para superar estos problemas mediante una variedad de estrategias de ingeniería de proteínas incluyendo el uso de frenos moleculares (oligonucleótidos cortos unidos covalentemente) (Bailey, H. 2006).

Tal como se da a conocer en el presente documento, pueden unirse oligonucleótidos cortos al fosfato terminal de un polifosfato de nucleósido mediante reacción del fosfato terminal activado con el 3'-OH o el 5'-OH del oligonucleótido. Alternativamente, puede activarse el 3' o 5'-fosfato del oligonucleótido con CDI o imidazol/DCC y hacerse reaccionar

con 5'-polifosfatos de nucleósido. En las figuras 27(a) y 27(b), respectivamente, se muestran estructuras de fosfatos de nucleósido unidos a oligonucleótido (oligo-3' a 5'-fosfato; oligo-5' a 5'-fosfato). El subproducto de reacción de polimerasa que se monitoriza mediante pase a través del nanoporo se muestra en la figura 27(c).

5 La tasa de migración a través del nanoporo del subproducto de reacción de polimerasa puede controlarse uniendo oligonucleótidos de diferente longitud a diferentes 5'-polifosfatos de nucleósido. Por ejemplo, si el nucleósido dA tiene 1 ó 2 unidades de oligo-dA unidas, dT puede tener 3 unidades de oligo-dT, dC puede tener 4 unidades de oligo-dC, y dG puede tener 5 unidades de oligo-dG. Pueden usarse diferentes combinaciones del número de oligonucleótidos para cada nucleótido para controlar el tiempo de retención y transporte en un nanoporo.

10 El tiempo de retención y transporte en un nanoporo también puede controlarse añadiendo un número diferente de grupos fosfato a los nucleótidos. Por tanto, la carga y masa pueden variar para cada polifosfato de nucleótido.

Ejemplos de estructura de ligador-etiqueta

En la tabla 1 se proporcionan ejemplos específicos de grupos reactivos en los fosfatos terminales o el resto de base de nucleósido y grupos con los que pueden reaccionar los grupos . Los grupos reactivos con los que pueden reaccionar pueden estar presentes o bien en el ligador o bien en la etiqueta.

15 TABLA 1

Posibles sustituyentes reactivos y grupos funcionales reactivos con los mismos	
Grupos reactivos	Grupos funcionales
Ésteres de succinimidilo	Amino primario, amino secundario
Anhídridos, haluros de ácido	Grupos amino e hidroxilo
Carboxilo	Amino, hidroxilo, tioles
Aldehído, isotiocianato e isocianatos	Grupos amino
Vinilsulfona y diclorotriazina	Grupos amino
Haloacetamidas	Tioles, imidazoles
Maleimidas	Tioles, hidroxilo, amino
Tioles	Tioles, maleimida, haloacetamida
Fosforamidas, P activado	Grupos hidroxilo, amino, tiol
Azido	Alquino

Las etiquetas que pueden detectarse mediante nanoporos se incluyen en este caso, pero no se limitan de ningún modo a este grupo de compuestos. Un experto en la técnica puede cambiar el/los grupo(s) funcional(es) para encontrar una etiqueta adecuada.

20 Las etiquetas incluyen compuestos alifáticos, aromáticos, arilo, heteroarilo con uno o más anillos de 4-8 miembros y pueden estar opcionalmente sustituidas con halo, hidroxilo, amino, nitro, alcoxilo, ciano, alquilo, arilo, heteroarilo, ácido, aldehído, azido, alquenilo, alquinilo u otros grupos. Estos incluye, polietilenglicoles (PEG), hidratos de carbono, aminoácidos, péptidos, colorantes fluorescentes, fluorogénicos (no fluorescentes pero que se vuelven fluorescentes tras la retirada de un grupo protector) cromogénicos (incolores pero que se vuelven de color tras la retirada de un grupo protector), compuestos quimioluminiscentes, nucleósidos, mono, di o polifosfatos de nucleósido, oligonucleótidos, arilo, heteroarilo o compuestos alifáticos. En la figura 35 se facilitan algunos ejemplos.

En la figura 36 se muestra a modo de ejemplo la estructura de nucleótidos marcados con PEG-fosfato y algunos ejemplos de posibles PEG con diferentes grupos reactivos para reaccionar con grupos funcionales.

30 En este caso se proporcionan algunos otros ejemplos de los colorantes o compuestos que pueden usarse para unirse al fosfato terminal o al resto de base de los nucleótidos. Estos no son de ningún modo los únicos compuestos que pueden usarse. Se indican en este caso como ejemplos y un experto en la técnica puede encontrar fácilmente un ligador-etiqueta adecuado que puede unirse al nucleótido y detectarse mediante nanoporos.

Otros ejemplos de etiquetas adecuadas son:

35 Colorantes fluorescentes: colorantes de xantina, colorantes de bodipy, colorantes de cianina. Compuestos quimioluminiscentes: compuestos de 1,2-dioxetano (Tropix Inc., Bedford, MA). Aminoácidos y péptidos: aminoácidos que se producen de manera natural o modificados y polímeros de los mismos. Hidratos de carbono: glucosa, fructosa, galactosa, manosa, etc. NMP y NDP: monofosfatos de nucleósido, difosfatos de nucleósido. Ácidos alifáticos o aromáticos, alcoholes, tioles, sustituidos con halógenos, ciano, nitro, alquilo, alquenilo, alquinilo, azido u otros grupos de este tipo.

2) 5'-Trifosfatos de nucleósido de base modificada

40 Se sintetiza una variedad de terminadores reversibles de nucleótido (NRT) para secuenciación mediante síntesis (SBS) de ADN en los que un ligador escindible une un colorante fluorescente a la base de nucleótido y el 3'-OH del

nucleótido se bloquea con un grupo de terminación reversible pequeño (Ju *et al.* 2006, Guo *et al.* 2008 y 2010). Usando estos NRT, se detiene de manera reversible la síntesis de ADN en cada posición. Tras registrar la señal fluorescente a partir de la base incorporada, se retiran los restos escindibles de los nucleótidos incorporados y se repite el ciclo.

- 5 También puede usarse el mismo tipo de nucleótidos para la secuenciación de ADN en nanoporos. Tal como se muestra en la figura 28(A), puede sintetizarse un grupo de bloqueo pequeño en 3'-OH y una etiqueta unida a la base unida a través de un ligador escindible. Tras la reacción de extensión de polimerasa, se escinden tanto el grupo de bloqueo en 3'-O como la etiqueta de la base y puede usarse la etiqueta liberada para pasar a través del nanoporo y monitorizarse la señal de bloqueo. Pueden usarse cuatro etiquetas diferentes (por ejemplo polietilenglicoles (PEG) de diferente longitud y peso molecular, tal como se muestra en la figura 28(A)), una para cada una de las cuatro bases, diferenciando por tanto las señales de bloqueo.

15 Alternativamente, el grupo de bloqueo en 3'-O no se usa porque se ha mostrado que un grupo o base de nucleótido voluminoso puede impedir que la ADN polimerasa añada más de un nucleótido cada vez (Harris *et al.* 2008). Tal como se muestra en la figura 28(B), se introduce un dNMP voluminoso mediante un ligador escindible. Por tanto, se introducen diferentes dNMP mediante un ligador según el dNTP original. Por ejemplo, con nucleótido dTTP, se introduce un dTMP (para dATP, se introduce un dAMP; para dGTP, un dGMP y para dCTP, un dCMP). Tras la incorporación mediante polimerasa y escisión con TCEP, se generan dNMP modificados que se hacen pasar a través del canal de nanoporo y se detectan mediante métodos apropiados.

3) 5'-Trifosfatos de nucleósido modificado en 2' o 3'-OH

- 20 Puede llevarse a cabo la síntesis de los cuatro 5'-trifosfatos de nucleósido modificados en 3' (Guo *et al.* 2008, Li *et al.* 2003, Seo *et al.* 2004). dNTP unidos a 3'-O-2-nitrobencilo y 3'-O-azidometilo (figura 29A y 29B, respectivamente) son buenos sustratos para ADN polimerasas. Tras la incorporación mediante ADN/ARN polimerasa en una reacción de secuenciación, estos nucleótidos marcados con etiqueta en 3'-O terminan la síntesis tras la extensión de una única base debido al grupo de bloqueo en 3'-OH. La extensión adicional sólo es posible tras la escisión del grupo de bloqueo a partir de la posición 3'-O. El grupo 3'-O-2-nitrobencilo puede escindirse eficazmente mediante luz UV y 2'-O-azidometilo mediante tratamiento con TCEP para generar el grupo OH libre para una extensión adicional. El producto escindido a partir de la reacción (figura 29C o 29D) se monitoriza para detectar el bloqueo electrónico mediante pase a través del nanoporo y registro de la señal. Se sintetizan cuatro dNTP protegidos con nitrobencilo sustituido diferentes y cuatro dNTP sustituidos con azidometilo diferentes, uno para cada una de las cuatro bases de ADN.

II. Extensión de ADN usando nucleótidos modificados

1) Nucleótidos marcados con etiqueta en fosfato

35 Los polifosfatos de nucleósido marcados con etiqueta en fosfato terminal descritos anteriormente se usan en reacciones de polimerasa para generar productos de extensión. Tal como se muestra en la figura 30, tras una reacción de polimerasa, se obtiene el subproducto liberado del nucleótido marcado con etiqueta en fosfato, polifosfato de etiqueta, y el ADN extendido está libre de cualquier modificación. Después se usa el polifosfato de etiqueta liberada en un nanoporo modificado por ingeniería mediante técnicas de registro de canales individuales para análisis de secuenciación. Los polifosfatos de etiqueta liberada también pueden tratarse con fosfatasa alcalina para proporcionar etiquetas libres que también pueden detectarse. Usando cuatro etiquetas diferentes para los cuatro nucleótidos (A, T, G y C) para generar cuatro polifosfatos marcados con etiqueta diferentes que difieren en cuanto a masa, carga o voluminosidad, puede determinarse la secuencia del ADN.

2) Nucleótidos marcados con etiqueta en base con ligadores escindibles

45 Se sintetizan trifosfatos de nucleótido marcado con etiqueta en base para la secuenciación mediante síntesis (SBS) de ADN y secuenciación de moléculas individuales (Guo *et al.* 2008 y 2010). La adición de grandes grupos voluminosos en la posición 5 de pirimidinas (C u T) y la posición 7 de 7-desazapurinas (G y A) puede bloquear la adición de más de un nucleótido en una reacción de ADN polimerasa. Se sintetizan nucleótidos modificados con un ligador escindible, un grupo voluminoso y diferentes cargas unidos a la base de nucleótido. Los nucleótidos modificados también pueden tener un pequeño grupo de bloqueo en el 3'-OH de los nucleótidos. Estos nucleótidos modificados se usan en una reacción de extensión de polimerasa. Tal como se muestra en la figura 31, tras la extensión con el nucleótido apropiado, el ligador y la etiqueta de la base de nucleótido y de 3'-O de azúcar, si está bloqueado, se escinden mediante medios químicos o fotoquímicos y se usa el ligador-etiqueta liberado en un nanoporo modificado por ingeniería mediante técnicas de registro de canales individuales para análisis de secuenciación.

3) Nucleótidos marcados con etiqueta en 2' o 3' con ligadores escindibles

55 También pueden unirse un ligador y una etiqueta al 2' o 3'-OH de nucleótidos. Tras una reacción de extensión de polimerasa, el ligador-etiqueta se escinde a partir del producto extendido mediante reacción química, fotoquímica o enzimática para liberar el 3'-OH libre para su extensión adicional. Tal como se muestra en la figura 32, después se

usa el ligador-etiqueta liberado en un nanoporo modificado por ingeniería mediante técnicas de registro de canales individuales para análisis de secuenciación.

III. Estudios de secuenciación de ADN usando nanoporos

Se evalúa la distinción de diferentes nucleótidos en la secuenciación de ADN usando nanoporos siguiendo la estrategia mostrada en las figuras 30-32. Para validar la capacidad de un nanoporo para distinguir los cuatro ligadores-etiquetas diferentes en ADN, se realiza una serie de experimentos tal como se muestra en la figura 33. La ADN/ARN polimerasa puede unirse al nanoporo y se añade un molde que va a secuenciarse junto con el cebador. O bien el molde de ADN o bien el cebador también puede inmovilizarse encima del nanoporo y después formar posteriormente un complejo de molde-cebador tras la adición de una ADN polimerasa. A este complejo de molde-cebador se le añaden cuatro nucleótidos marcados con etiqueta de manera diferente juntos o de manera secuencial. Tras la incorporación catalizada por polimerasa del nucleótido correcto, el nucleótido añadido libera el polifosfato unido a etiqueta (en el caso de nucleótidos marcados en fosfato terminal) que después pasa a través del nanoporo para generar la señal eléctrica que va a registrarse y usarse para identificar la base añadida. Opcionalmente, el polifosfato de etiqueta liberada también puede tratarse con fosfatasa alcalina para proporcionar una etiqueta libre que también puede detectarse mediante pase a través del nanoporo. Cada etiqueta genera una firma de bloqueo electrónico diferente debido a la diferencia de tamaño, masa o carga. En el caso de nucleótidos modificados en base o modificados en 2'/3', tras la extensión de ADN/ARN polimerasa, la etiqueta del cebador extendido se escinde mediante medios químicos, fotoquímicos o enzimáticos y se monitoriza la firma electrónica de la etiqueta liberada. La forma, tamaño, masa, carga u otras propiedades de la etiqueta pueden ajustarse según los requisitos.

Tal como se da a conocer en el presente documento, se distinguen y caracterizan señales a partir de cada uno de los nucleótidos (figura 34) y las transiciones entre nucleótidos de diferentes identidades. La magnitud y duración de las firmas de bloqueo en el diagrama de acontecimientos se analizan y comparan con diagramas conocidos. Por tanto, con estos diseños químicos racionales y modificaciones de los elementos estructurales de ADN, se optimiza el uso de nanoporos para descifrar la secuencia de ADN a nivel de moléculas individuales con resolución de base individual.

Para implementar esta estrategia novedosa para la secuenciación de ADN, puede construirse una matriz de nanoporos en una superficie plana para llevar a cabo secuenciación de ADN en paralelo masiva tal como se muestra en la figura 37. La matriz de nanoporos también puede construirse en un chip de silicio u otras superficies de este tipo. El nanoporo puede construirse a partir de la proteína con bicapas lipídicas u otras capas de este tipo (poro de alfa-hemolisina, porina A de *Mycobacterium smegmatis*, MspA) (Derrington *et al.* 2010) o pueden ser nanoporos de estado sólido sintéticos fabricados en nitruro de silicio, óxido de silicio u óxidos de metales (Storm *et al.* 2005; Wanunu *et al.* 2008) o un híbrido entre un poro de estado sólido y α -hemolisina (Hall *et al.* 2010).

La figura 37 muestra un esquema de matriz de nanoporos para la secuenciación de ADN en paralelo masiva mediante síntesis. Los nanoporos pueden detectar cada subproducto de adición de nucleótido catalizada por ADN/ARN polimerasa (etiqueta unida al fosfato o la base y/o 2', 3'-OH del resto de azúcar) a medida que pasa a través del nanoporo. Las propiedades eléctricas de diferentes etiquetas distinguirán las bases basándose en su propiedad de bloqueo en los nanoporos. La matriz de nanoporos mostrada en la figura 37 pueden leer cada uno la misma secuencia o secuencias diferentes. Aumentar el número de veces que se lee cada secuencia dará como resultado una mejor calidad de los datos de secuencia resultantes.

40 EJEMPLO 2

I. Síntesis de 5'-tetrafosfatos de desoxiguanosina marcados con PEG (dG4P-PEG):

Se sintetizan 5'-tetrafosfatos de desoxiguanosina marcados con PEG (dG4P-PEG) según la figura 38. En primer lugar, trifosfato de 2'-desoxiguanosina (dGTP) reacciona con CDI en DMF para activar el grupo fosfato terminal que después se hace reaccionar con fosfato de dibutilamonio para dar el tetrafosfato. El fosfato terminal en este tetrafosfato se activa adicionalmente con EDAC en tampón de imidazol 0,1 M seguido por reacción con diaminoheptano para proporcionar un tetrafosfato unido a amino que se hace reaccionar adicionalmente con ésteres de mPEG-NHS para proporcionar los cuatro PEG-dG4P requeridos. Tras la incorporación mediante polimerasa, la carga neta sobre el PEG liberado es de -3 (PEG-NH-trifosfato).

II. Pruebas de nucleótidos modificados en reacciones de extensión de una única base.

50 Los dG4P-PEG se caracterizan mediante espectroscopía de masas de MALDI-TOF tal como se muestra en la tabla II.

Tabla II. Resultados de MALDI-TOF-EM para dG4P-PEG		
	P.M. calculado	P.M. medido
dG4P-PEG24	1798	1798
dG4P-PEG37	2371	2374

Los dG4P-PEG son sustratos excelentes para ADN polimerasa en la extensión de cebador. Los espectros de masas

de MALDI-TOF de los productos de extensión de ADN se muestran en la figura 39.

EJEMPLO 3

Detección de moléculas individuales mediante nanoporos de los PEG usados para marcar los nucleótidos

5 El polietilenglicol es un polímero no electrolítico que se une de manera débil a cationes (por ejemplo, se une a iones K^+ a $K_d \sim 2$ M). Por tanto, la carga neta en el polímero depende de la concentración de cationes móviles y de la presencia de otros restos que están químicamente unidos al mismo. Se ha demostrado que un único nanoporo de α -hemolisina puede distinguir fácilmente entre polímeros de PEG de tamaño diferente a una resolución mejor que monomérica, es decir, mejor que 44 g/mol (Reiner *et al.* 2010; Robertson *et al.* 2007). Ese nivel de distinción se vuelve posible porque el polímero reduce la conductancia del poro debido a la exclusión de volumen (la conductancia de poro disminuye al aumentar el tamaño de polímero) y uniéndose a cationes móviles que de lo contrario pueden fluir libremente a través del poro (Reiner *et al.* 2010). Además, el tiempo de residencia del polímero en el poro es altamente sensible a la carga del polímero, que para PEG varía a escala en proporción a la longitud del polímero. Un nanoporo debe poder distinguir entre PEG de tamaño diferente que están químicamente unidos a otros restos. Se someten a prueba en nanoporos PEG (PEG 16, 24, 37 y 49) para marcar nucleótidos y generan firmas de bloqueo electrónico diferenciadas a nivel de moléculas individuales tal como se muestra en la figura 40.

Para investigar el efecto de la voluminosidad de los polifosfatos marcados con etiqueta de manera diversa sobre señales de bloqueo electrónico generadas en el nanoporo, se sintetizan diversos nucleótidos unidos a fosfato con etiquetas de polietilenglicol (PEG) de diferente tamaño unidas al fosfato terminal del nucleótido. En primer lugar, tal como se muestra en la figura 23, se sintetiza una serie de 5'-tri, tetra y penta-fosfatos de nucleósido con el fosfato terminal unido a través de un ligador al que pueden unirse diferentes etiquetas, por ejemplo PEG de diferente longitud y masa u otras moléculas para aumentar el tamaño molecular o modificar la carga del polifosfato liberado. Después se someten a prueba estos nucleótidos en reacciones de polimerasa acopladas con detección mediante nanoporos para ver qué etiquetas o grupos voluminosos unidos al fosfato terminal producen diferencias más drásticas en las señales de bloqueo electrónico de entre diferentes bases.

25 I. Examen y selección de 4 etiquetas de PEG con señales de bloqueo en nanoporos diferenciadas

Recientemente se ha mostrado que cuando una molécula de polietilenglicol (PEG) entra en un único poro de α -hemolisina, provoca estados de conductancia dependientes de masa diferenciados con tiempos de residencia medios característicos (Robertson *et al.* 2007). La figura 41A muestra que el espectro de masas basado en conductancia resuelve claramente las unidades de repetición de etilenglicol, y el tiempo de residencia aumenta con la masa de PEG.

I.a Pruebas de PEG para determinar firmas de bloqueo en nanoporos.

Se seleccionan PEG de diferente longitud y peso molecular (comercialmente disponibles de Quanta Biodesign Ltd u otros proveedores) y se monitorizan las señales de bloqueo en nanoporos, tal como se describe en el ejemplo 2. Tal como se muestra en la figura 41A, los PEG de 28-48 unidades de etilenglicol se distinguen claramente mediante nanoporos. Por tanto, se seleccionan PEG con un amplio intervalo de unidades de etilenglicol que presentan señales de bloqueo en nanoporos muy diferenciadas como etiquetas para marcar los nucleótidos A, C, G y T. En la figura 41B se muestran ejemplos. También se evalúan PEG ramificados como etiquetas ya que pueden modificarse con cargas positivas de una manera más directa. En la parte inferior de la figura 41 se muestran estructuras de algunos PEG lineales y ramificados.

40 I.b Diseño y síntesis de PEG marcados con fosfato seleccionados en I.a

En la secuenciación en nanoporos, las señales de bloqueo de corriente en el nanoporo se generan mediante los PEG-fosfatos liberados durante la reacción de polimerasa. Por tanto, se diseñan y sintetizan PEG marcados con fosfato con ligadores cargados de manera positiva y se someten a prueba estas moléculas con nanoporos orgánicos (por ejemplo, α -hemolisina) y sintéticos (fase sólida) para evaluar sus señales de bloqueo de corriente. Los PEG seleccionados se convierten en sus trifosfatos tal como se muestra en la figura 42. Por ejemplo, puede convertirse amino-butanol protegido con Fmoc en el trifosfato correspondiente mediante reacción en primer lugar con oxiclورو de fósforo seguido por reacción con pirofosfato de tributilamonio en una reacción en un solo recipiente. Tras la purificación, se activa el trifosfato con DCC/DMF o CDI/DMF para proporcionar trifosfato activado que reacciona con el grupo OH de los PEG para generar PEG-trifosfatos. El mismo esquema puede aplicarse para PEG tanto lineales como ramificados. Estos fosfatos de PEG se someten a prueba en nanoporos para optimizar las condiciones para generar señales de bloqueo de corriente diferenciadas.

Los ligadores de poliaminoácido (polilisina, poliarginina, polilisina interrumpida) se sintetizan mediante estrategias de síntesis de péptidos convencionales; si está incorporada una unión éster con la cadena de polifosfato, debe ser posible usar fosfatasa alcalina para escindirlos, dando como resultado etiquetas más fuertemente positivas para la interrogación mediante nanoporos. También pueden incorporarse cargas positivas en las cadenas de PEG.

I.c Diseño y síntesis de una biblioteca de 5'-trifosfatos de nucleósido marcados con etiqueta en fosfato terminal.

Se diseñan y sintetizan 5'-tri, tetra y penta-fosfatos de nucleósido marcados con etiqueta en fosfato terminal. Se someten a prueba estas moléculas en la reacción de polimerasa y se seleccionan las óptimas para la detección en nanoporos. Se ha mostrado que 5'-tri, tetra y penta-fosfatos de nucleósido marcados con etiqueta en fosfato terminal con una variedad de etiquetas, incluyendo polilisinas pequeñas o grandes, aminoácidos, una variedad de colorantes cargados de manera negativa o positiva, tales como colorantes de transferencia de energía, y unidades de etilenglicol, se aceptan por ADN polimerasas como sustratos excelentes para la extensión de cebador (Kumar *et al.* 2006 y 2008; Sood *et al.* 2005; y Eid *et al.* 2009).

I.c.1 Diseño y síntesis de 5'-trifosfatos de nucleósido marcados con etiqueta en fosfato terminal.

Tal como se muestra en la figura 43, se sintetizan 5'-trifosfatos de nucleósido marcados con etiqueta en fosfato terminal haciendo reaccionar el dNTP correspondiente con DCC/DMF para producir un trimetafosfato cíclico que puede abrirse con nucleófilos para generar un 5'-trifosfato de nucleósido unido a etiqueta o ligador. Además, el ligador unido al fosfato puede hacerse reaccionar con ésteres de PEG-NHS para proporcionar 5'-trifosfatos de nucleósido unidos a PEG alternativos. El 5'-trifosfato de nucleósido marcado con etiqueta en fosfato terminal resultante se usa en la reacción de extensión de molde-cebador y el pirofosfato unido a etiqueta liberada se detecta y se diferencia mediante sus parámetros de bloqueo de corriente en nanoporo específicos.

I.c.2 Diseño y síntesis de 5'-tetrafosfatos de nucleósido marcados con etiqueta en fosfato terminal.

Para la síntesis de 5'-tetrafosfatos de nucleósido marcados con etiqueta en fosfato terminal, en primer lugar se hace reaccionar el trifosfato correspondiente con CDI en DMF para activar el grupo fosfato terminal que después se hace reaccionar con ácido fosfórico o monofosfato de etiqueta para dar el tetrafosfato tal como se muestra en la figura 44. El fosfato terminal en el tetrafosfato puede activarse adicionalmente con EDAC en tampón de imidazol 0,1 M seguido por reacción con un nucleófilo apropiado para proporcionar un tetrafosfato unido a ligador que puede usarse para unir etiquetas de diferente masa, longitud o carga, tales como ésteres de m-PEG-NHS. En este caso, se usan cuatro trimetil-lisinas para neutralizar la carga de cuatro fosfatos. Tras la incorporación mediante polimerasa, la carga neta en el PEG liberado es de +1 o, si se trata con fosfatasa alcalina, +4, lo cual puede detectarse mediante el nanoporo.

I.c.3 Diseño y síntesis de 5'-penta-fosfatos de nucleósido marcados con etiqueta en fosfato terminal.

La síntesis de 5'-penta-fosfatos de nucleósido marcados con etiqueta en fosfato terminal sigue el mismo principio tal como se muestra en la figura 45. Pueden prepararse a partir o bien de trifosfatos o bien de tetrafosfatos activados mediante reacción con ácido fosfórico, pirofosfato o fosfatos unidos a etiqueta. Alternativamente, puede unirse un ligador al penta-fosfato seguido por reacción con ésteres de NHS activados.

Los polifosfatos de nucleósido marcados con etiqueta en fosfato terminal descritos anteriormente se usan en la reacción de polimerasa para generar productos de extensión. Siguiendo el esquema mostrado en la figura 30, se evalúa el rendimiento de los polifosfatos de nucleósido marcados con etiqueta en fosfato terminal en la extensión de polimerasa. En primer lugar se realiza una reacción de extensión de una única base y se caracteriza el producto de extensión de ADN mediante espectroscopía de masas de MALDI-TOF para evaluar la eficacia de incorporación. Tras establecer condiciones de reacción optimizadas, se inmoviliza el molde sobre perlas magnéticas y se repite la reacción de extensión de una única base, tras lo cual se aíslan las etiquetas de polifosfato liberadas a partir de la disolución para su detección usando un único nanoporo. Esta reacción se realiza de manera continua para evaluar los 4 nucleótidos (A, C, G y T) y se detectan sus etiquetas liberadas correspondientes mediante el nanoporo. La reacción de polimerasa continua con los nucleótidos con etiqueta de polifosfato y la clara distinción de la etiqueta de polifosfato liberada mediante nanoporos establecen la viabilidad del enfoque.

Tal como se muestra en la figura 30, tras la reacción de polimerasa, se obtiene el subproducto liberado del nucleótido marcado con etiqueta en fosfato (polifosfato de etiqueta) y la cadena de ADN extendida está libre de cualquier modificación. Esto es ventajoso porque cualquier cicatriz que quede en la cadena de ADN en crecimiento puede afectar a su capacidad para reconocerse por polimerasa con adiciones de nucleótidos crecientes, terminando eventualmente la síntesis de ADN adicional. El polifosfato unido a etiqueta liberada se somete a ensayo en el nanoporo para evaluar la sensibilidad y precisión de la secuenciación. En experimentos iniciales se someten a prueba las etiquetas para determinar sus señales de bloqueo antes de llevar a cabo reacciones de SBS. Puede determinarse la secuencia de ADN si se usan etiquetas diferentes para los cuatro nucleótidos para generar cuatro polifosfatos unidos con etiquetas diferentes que difieren en cuanto a la masa, carga o voluminosidad, y producir 4 señales de bloqueo diferenciadas.

II. Detección de los fosfatos marcados con etiqueta liberados mediante nanoporos de proteína

Se usa un único nanoporo de α -hemolisina para detectar PEG que están unidos a nucleótidos unidos a través de un ligador multifosfato y el mismo polímero después de haberse escindido el resto de nucleótido/ribose mediante la reacción de ADN polimerasa. Cada una de las cuatro bases de ADN diferentes se une a un polímero de PEG con una longitud única. Por tanto, se identifica cada base que se retira a partir del PEG mediante la polimerasa. Dado que los nucleótidos sin reaccionar no pueden separarse a partir de los polifosfatos marcados con etiqueta liberados, especialmente en situaciones en tiempo real, se aprovecha la extremada sensibilidad del método con respecto a la carga molecular para distinguir entre el producto de reacción liberado y el material de partida. Se mide la

conductancia de α -hemolisina individual usando soportes de vidrio cónicos (White *et al.*, 2006 y 2007) que permiten la recopilación de datos a 100 kHz y ruido de RMS de ~ 4 pA. Se miden las distribuciones de profundidad de bloqueo y tiempo de residencia tanto de los nucleótidos marcados con etiqueta como de los productos marcados con etiqueta a lo largo de un amplio intervalo de potenciales transmembrana para determinar condiciones óptimas para la distinción de nucleótidos y para extender la presente comprensión teórica de interacciones de PEG-nanoporo (Robertson *et al.* 2007) a moléculas con cargas fijas. La caracterización y comprensión teórica permiten la identificación sin ambigüedad de los nucleótidos incorporados en polinucleótidos mediante polimerasa. Por tanto, con estos diseños y modificaciones químicos racionales de los elementos estructurales de ADN, se optimiza el uso de nanoporos para descifrar el ADN a nivel de moléculas individuales con resolución de base individual en nanoporos de proteína o sintéticos.

EJEMPLO 4

Fabricación de un único nanoporo de estado sólido para secuenciación de moléculas individuales

La transición de un nanoporo de proteína a un nanoporo de estado sólido hace posible la fabricación de matrices de nanoporos de alta densidad, una etapa clave para producir un secuenciador de ADN electrónico de moléculas individuales de alto rendimiento. En este caso, se desarrolla una plataforma integrada de un único nanoporo de estado sólido para caracterizar los nucleótidos marcados con etiqueta en la reacción de polimerasa basándose en el conocimiento obtenido a partir del nanoporo de proteína.

Plataforma de nanoporos integrada.

Se desarrolló electrónica de CMOS de bajo ruido integrada especializada, que cuando se integra con nanoporos de estado sólido, suministra ventajas de rendimiento significativas con respecto a técnicas de medición "convencionales" que emplean amplificadores electrofisiológicos externos, tales como el instrumento Axopatch 200B. Estas ventajas proceden de aprovechar la realimentación capacitiva (en vez de resistiva) en un diseño de amplificador de integración personalizado. Se retira corriente CC, que es característica de esta y otras interfaces bioelectrónicas, con una fuente de corriente de bajo ruido que funciona en un servobucle de CC. Capacitancias de entrada de amplificador reducidas y capacitancias parásitas reducidas asociadas con la integración conjunta mejoran el rendimiento de ruido a altas frecuencias, permitiendo anchos de banda que se aproximan a 1 MHz para poros de estado sólido. Tal alta resolución temporal, en combinación con las etiquetas desarrolladas, proporcionará una alta flexibilidad para sintonizar esta plataforma para alta sensibilidad y rendimiento en tiempo real.

El uso de este nanoporo integrado con CMOS (CNP) integró el circuito en cualquiera de una configuración de dos chips o de un chip tal como se muestra en la figura 46. En el primer caso, el poro se empaqueta junto con el CNP tal como se muestra en la figura 46B. En el segundo, el poro se fabrica directamente en el CNP tal como se muestra en la figura 46C con fluidica a ambos lados del chip. En ambos casos, el electrodo en cis, que se conecta a la entrada del amplificador, se integra directamente en la superficie del CNP. La configuración de un chip tiene la ventaja de poder ajustarse fácilmente a escala hasta una plataforma multiplexada a costa de complejidad de fabricación adicional. La capacidad de realizar procesamiento posterior de un dado de CMOS fabricado (que no tiene más de 5 mm en un lado) es una capacidad única establecida a lo largo de los últimos cinco años (Huang *et al.* 2011, Lei *et al.* 2008, y Levine *et al.* 2009). Este enfoque aprovecha completamente flujos de procedimiento de fundición existentes en vez de requerir el desarrollo de nuevos procedimientos.

El enfoque de fabricación en un chip procede adaptando técnicas de fabricación de nanoporos de estado sólido convencionales (Rosenstein *et al.* 2011). En zonas de la pastilla reservadas para los sensores, se han bloqueado todos los metales, dejando una pila gruesa de capas de relleno de vidrio y cubierta de nitruro de silicio alternantes. La mayor parte de la pila dieléctrica se graba usando un plasma de CHF_3 acoplado por inducción. Tras depositar y disponer en patrones una máscara de grabado de Si_3N_4 de PECVD en la parte posterior de la pastilla, se realizan aberturas localizadas en el sustrato de silicio usando un grabado con hidróxido de potasio anisotrópico. Después se usa una corta inmersión en ácido fluorhídrico tamponado para aislar una única capa de nitruro de silicio de 50 nm a partir de la pila dieléctrica original como membrana suspendida. Finalmente, se perforan nanoporos a través de estas membranas de nitruro con un microscopio electrónico de transmisión de alta resolución.

El ruido medido de este sistema se muestra en la figura 47A, junto con una medida del ruido de referencia para una configuración comparable del instrumento Axopatch 200B. Para el ancho de banda más alto soportado por el instrumento Axopatch ($B = 100$ kHz), el amplificador integrado tiene un umbral mínimo de ruido de $3,2$ pA_{RMS}, en comparación con 9 pA_{RMS} para el instrumento Axopatch. Al ancho de banda más alto caracterizado para el amplificador integrado ($B = 1$ MHz), el nivel de ruido es de 27 pA_{RMS}, en contraposición a 247 pA_{RMS} modelado mediante extrapolación de la respuesta del instrumento Axopatch más allá de su intervalo soportado (aproximadamente un ruido inferior en un factor de diez). Como punto de comparación, para una señal de 1 nA, sólo se transportan aproximadamente 6250 iones a través del poro en 1 us. Un nivel de ruido con referencia a la entrada de 27 pA_{V RMS} para el amplificador integrado permite una resolución de tan sólo 150 iones en este intervalo.

También es importante indicar que estas prestaciones eléctricas superiores se obtienen con un amplificador integrado que consume un área de tan sólo $0,2$ mm² en un chip de CMOS en comparación con un amplificador

Axopatch montado en bastidor, lo que demuestra la importancia de la electrónica innovadora. Cuando se conecta un nanoporo a la entrada de amplificador, la introducción de ruido de $1/f$ y capacitancia de membrana aumenta el espectro de ruido por encima de la línea base de cabezal abierto. La figura 47b muestra un espectro de ruido típico en este caso, demostrando umbrales mínimos de ruido de tan sólo $10 \text{ pA}_{\text{RMS}}$ y $163 \text{ pA}_{\text{RMS}}$ para anchos de banda de 100 kHz y 1 MHz, respectivamente. Se muestran comparaciones medidas con el instrumento Axopatch hasta 100 kHz para el mismo nanoporo. A 100 kHz, hay una reducción de más de un factor de dos en la potencia de ruido con referencia a la entrada para el CNP. Si el instrumento Axopatch puede medirse a anchos de banda superiores, puede haber una diferencia de potencia de ruido de un factor de seis a 1 MHz.

Esta plataforma también permite la integración de nanoporos biológicos, proporcionando incluso más flexibilidad. Se crean nanoporos biológicos en membranas lipídicas (normalmente 1,2-dioleoil-sn-glicero-3-fosfolina (DOPC)) formadas sobre un orificio en una membrana de teflón entre dos células de fluido. La superficie debe ser lo suficientemente hidrófila para que se forme la membrana a partir de vesículas unilamelares. La conductancia entre las dos cámaras de la célula se monitoriza mientras se añade la proteína de membrana a una de las células, que se purga inmediatamente una vez que se detecta la incorporación. Las membranas usadas para fabricar los nanoporos también pueden usarse como soportes sólidos para bicapas lipídicas con la perforación de orificios más grandes en las membranas, sobre los que se forma la bicapa lipídica (Clarke *et al.* 2009; Benner *et al.*, 2007; Hou *et al.*, 2009; y Wang *et al.* 2011). Se han diseñado por ingeniería membranas lipídicas de bicapa (BLM) planas con diferentes canales de proteínas en soportes sólidos en patrones con orificios en patrones nanométricos ($\sim 100 \text{ nm}$ de diámetro), así como anclándolas directamente sobre oro mediante un ensamblaje de monocapa autoensamblado (Axelrod *et al.*, 1976, Bultmann *et al.* 1991, Dutta *et al.* 2010, Jenkins *et al.* 2001, Nam *et al.* 2006, Palegrosdemange *et al.* 1991, Shen *et al.* 2009, Srinivasan *et al.* 2001, Yang *et al.* 2003, Yin *et al.* 2005). Además, se ha mostrado que la formación de BLM contiguas con un coeficiente de difusión de $4 \mu\text{m}^2/\text{s}$ sobre sustratos con patrones nanométricos; BLM formadas sobre ensamblajes de SAM-oro produjeron un coeficiente de $0,8 \mu\text{m}^2/\text{s}$. Ambos se encuentran dentro del intervalo de difusión ideal de $0,1\text{-}10 \mu\text{m}^2/\text{s}$ representativo de BLM bien formadas (Axelrod *et al.* 1976, Bultmann *et al.* 1991). Las caracterizaciones eléctricas de estas BLM indican una membrana de alta impedancia con una resistencia de $1,4 \text{ GW}\cdot\text{mm}^2$, haciéndolas adecuadas para análisis eléctrico adicional de nanoporos biológicos formados en la membrana (Oliver *et al.* 1994, Shi *et al.* 2000, Wieheman 1988).

Inmovilización de polimerasa en superficies que portan nanoporos

El tamaño de la polimerasa es de aproximadamente $5 \text{ nm} \times 5 \text{ nm}$. Se posiciona una polimerasa cerca de la entrada de cada nanoporo. Para lograr esto para los nanoporos de estado sólido, es necesario que (1) se modifique una posición única en la superficie con grupos funcionales durante la fabricación de CMOS para unirse a la polimerasa; (2) que los sitios sean lo suficientemente pequeños como para que sólo pueda unirse una molécula de polimerasa; (3) que estén lo suficientemente alejados como para que haya poca posibilidad de difusión de los polifosfatos marcados con etiqueta liberados a un canal cercano; y (4) que el agente de reticulación sea lo suficientemente flexible como para que la enzima esté funcionalmente intacta. El anclaje de polimerasa se logra combinando un punto de unión en patrones con el uso de una concentración apropiada de disolución de polimerasa durante la incubación de tal manera que como máximo se une una molécula de enzima.

El establecimiento del punto de anclaje apropiado para la polimerasa se logra aprovechando enfoques de fabricación existentes para nanoporos de estado sólido. Normalmente, para maximizar las señales de transducción, estos poros se crean adelgazando una membrana de Si_3N_4 soportada usando litografía por haz de electrones para definir una ventana que posteriormente se adelgaza con un grabado por plasma (por ejemplo SF_6). Después se perfora el nanoporo en la región adelgazada usando ablación por haz de electrones. El pocillo creado mediante esta ventana (figura 48) crea un lugar natural para anclar la polimerasa, garantizando una estrecha proximidad a la entrada de nanoporo. Antes de grabar la ventana adelgazada, puede aumentarse la membrana original con una capa epitaxial enterrada de material de unión. Una vez grabada la ventana, puede convertirse en una región de pared lateral selectiva para la unión de polimerasa. Los materiales de unión incluyen dióxido de silicio u oro. Sin embargo, puede haber una selectividad limitada con dióxido de silicio porque también puede formarse un óxido sobre la superficie de nitruro de silicio en condiciones apropiadas.

En principio, con superficies de dióxido de silicio, pueden usarse uniones de biotina-estreptavidina (Korlach *et al.* 2008 y 2010), usando moléculas de PEG biotiniladas sobre los parches de sílice e incubando polimerasa marcada en el extremo con biotina en presencia de estreptavidina. El resto de la superficie se pasiva con poli(ácido vinilfosfónico). Debido a las preocupaciones planteadas anteriormente, en vez de eso es preferible modificar la superficie de oro con una monocapa autoensamblada de alcanotiol (SAM) funcionalizada con grupos amino (Love *et al.* 2005). Estos pueden modificarse fácilmente para dar ésteres de NHS para la unión a grupos amino en la polimerasa. El grosor y la homogeneidad de la capa se determinan mediante elipsometría o microscopía de fuerza atómica.

Desarrollo de nucleótidos modificados en 5' con ligadores cargados de manera positiva

Se genera un sistema para la difusión rápida de las etiquetas liberadas hacia los poros mientras que los nucleótidos precursores y el ADN se repelen por los poros. Los nucleótidos marcados con etiqueta se modifican por ingeniería de modo que tras la incorporación en el ADN, la etiqueta liberada a partir del nucleósido tiene una carga positiva

acumulativa mientras que los nucleótidos con etiqueta intactos permanecen neutros. Esto permite regular activamente la etiqueta liberada específicamente a través del canal de detección, si el canal está cargado de manera negativa según los métodos (Wanunu *et al.* 2007). Dado que todas las demás moléculas libres presentes en la mezcla de reacción (cebadores, nucleótidos sin reaccionar, molde), distintos de la etiqueta, están cargados de manera negativa, sólo la etiqueta liberada que porta carga positiva se atrae al interior del canal, aumentando la especificidad de detección y reduciendo el ruido. Puede usarse un número diferente de grupos cargados en etiquetas diferentes, dependiendo de la base de nucleótido específica. Por tanto, puede usarse la carga acumulativa de la etiqueta junto con su tamaño para la distinción de bases. Tras la incorporación y liberación de la etiqueta, si se considera que el polifosfato enmascara la carga positiva, puede retirarse usando reacciones secundarias (por ejemplo, fosfatasa alcalina inmovilizada en un segundo sitio aguas abajo en el poro). La etiqueta cargada de manera positiva puede regularse al interior del canal cargado de manera negativa para su detección y reconocimiento.

Difusión y deriva

Un aspecto crítico de este sistema de secuenciación es la captura fiable y sincronizada de la etiqueta liberada de cada nucleótido mediante el nanoporo adyacente. Pueden diseñarse condiciones por ingeniería de tal manera que se capturan etiquetas rápidamente y en el orden correcto. Adicionalmente, debe minimizarse la tasa de captura de etiquetas no incorporadas, y la interferencia a partir de canales adyacentes debe ser despreciable. Crear el pocillo en la entrada del poro (tal como se muestra en la figura 48) ayuda en este procedimiento, que también depende de la estrecha proximidad de la polimerasa a la abertura de nanoporo. El análisis de procedimientos de captura en nanoporos considera generalmente un procedimiento radialmente simétrico que rodea al poro. La geometría dicta que, en ausencia de un campo eléctrico, una molécula tiende a difundir más lejos de un poro, oponiéndose a la atracción electrostática. Con un gradiente de tensión, existe una distancia crítica L a la que el movimiento molecular debido a la difusión y a la electroforesis son iguales (Gershow *et al.* 2007). Esta distancia crítica es una función de la corriente iónica (I) y la conductividad de electrolito (σ), así como la constante de difusión (D) y la movilidad (μ) de la

molécula de analito, $L = \frac{|I| \mu}{2\pi\sigma D}$. La captura es un procedimiento estadístico, pero se captura aproximadamente el 50% de las moléculas a una distancia L. Esta probabilidad aumenta para distancias más cortas, y supera el 90% para $d < L/3$. Durante este procedimiento, normalmente se capturan moléculas en una escala temporal del orden de $t_{captura} = \frac{L^2}{2D}$. Colocando la polimerasa dentro de L/3 del nanoporo, se capturan casi todas las moléculas. Esto también garantiza que $t_{captura}$ es significativamente más rápido que la tasa de incorporación por polimerasa, para capturar bases en el orden correcto.

Un valor aproximado para el coeficiente de difusión de moléculas de PEG de 25 unidades en agua es $D = 3e-10 \text{ m}^2/\text{s}$ (Shimada *et al.* 2005), que es del mismo orden de magnitud que un fragmento de ADNmc de longitud similar (Nkodo *et al.* 2001). Suponiendo la validez de la relación de Nernst-Einstein (aunque esto no siempre es cierto para polímeros), puede estimarse la movilidad como función de la constante de difusión y la carga neta (Q),

$\mu \approx \frac{QD}{k_B T}$. Entonces, para estas estimaciones, con $I = 5 \text{ nA}$ en KCl 1 M, véase la siguiente tabla.

	+1e	+4e
captura del 50%	2,1 nm	5,8 nm
captura del 90%	0,7 nm	1,9 nm
$t_{captura}$	7,1 ns	114 ns

EJEMPLO 5

Fabricar una matriz de nanoporos de estado sólido

Además de las prestaciones mejoradas, sólo con la electrónica integrada es posible producir matrices de nanoporos en paralelo de manera masiva. Esto implica la topología de un chip mostrada en la figura 46C en la que se integran nanoporos directamente en la pastilla de CMOS con fluidica a ambos lados del chip. El enfoque para integrar múltiples poros también se muestra en la figura 46C. En este caso, se usan pocillos de material fotosensible SU-8 para aislar nanoporos individuales unos de otros. Este es un enfoque similar al de Rothberg *et al.* 2011. Sin embargo, en Rothberg *et al.* los pocillos todavía pueden permanecer “conectados” mediante el depósito de disolución por encima del chip. En el presente caso, dado que se necesita aislamiento eléctrico entre los depósitos en cis, se usa un recubrimiento de PDMS para sellar los pocillos para la medición después de la introducción de reactivos tal como se muestra en la figura 46C. Se integran 64 nanoporos de estado sólido en la misma pastilla de 5 mm por 5 mm. El diseño de amplificador de integración de corriente, que puede tener que duplicarse en cada sitio de poro, sólo tiene 250 μm por 150 μm , pero tiene que dejarse espacio adicional para la fabricación del propio poro. A medida que se desarrollan adicionalmente técnicas de fabricación para reducir el área de chip, esto puede ajustarse fácilmente a escala a una matriz de 16 por 16 electrodos.

EJEMPLO 6

Pirosecuenciación usando nucleótido marcado con etiqueta en fosfato y detección en nanoporos

La pirosecuenciación es el método de secuenciación mediante síntesis (SBS) que se basa en la detección de pirofosfato que se libera cuando se incorpora un nucleótido en la cadena de ADN en crecimiento en la reacción de polimerasa (Ronaghi *et al.* 1998). En este enfoque, cada uno de los cuatro dNTP se añade secuencialmente con un cóctel de enzimas, sustratos y los componentes de reacción de polimerasa habituales. Si el nucleótido añadido es complementario a la primera base disponible en el molde, el nucleótido se incorporará y se liberará un pirofosfato. A través de una cascada de enzimas, el pirofosfato liberado se convierte en ATP, y después se convierte en una señal de luz visible mediante luciferasa de luciérnaga. Por otro lado, si el nucleótido añadido no se incorpora, no se producirá ninguna luz y el nucleótido simplemente se degradará mediante la enzima apirasa. La pirosecuenciación se ha aplicado satisfactoriamente a la detección de polimorfismos de un solo nucleótido (SNP) y la secuenciación de ADN. Se desarrolló una plataforma de secuenciación comercial combinando la pirosecuenciación y la amplificación de molde de ADN en micropelotas individuales para secuenciación de ADN de alto rendimiento (Margulies *et al.* 2005). Sin embargo, hay dificultades inherentes en la pirosecuenciación para determinar el número de nucleótidos incorporados en regiones homopoliméricas (por ejemplo, una concatenación de varias T en una fila) del molde. Además de esto, hay otros aspectos de la pirosecuenciación que todavía necesitan mejorarse. Por ejemplo, cada uno de los cuatro nucleótidos tiene que añadirse y detectarse por separado. La acumulación de nucleótidos no degradados y otros componentes también puede reducir la precisión del método cuando se secuencia un molde de ADN largo.

Este es un enfoque de pirosecuenciación modificado que se basa en la detección de fosfatos de etiqueta o etiqueta liberados durante la reacción de polimerasa. En este enfoque, se usan nucleótidos marcados con etiqueta en fosfato en reacción catalizada por polimerasa en un complejo de molde-cebador. Tras la incorporación de los nucleótidos marcados con etiqueta, se libera el resto fosfato-etiqueta, que puede detectarse mediante pase a través de un nanoporo. Puede usarse la misma etiqueta en cada nucleótido o puede usarse una etiqueta de peso molecular y longitud diferentes (tal como PEG). Se ha mostrado que polietilenglicoles (PEG) de longitud y masa diferentes pueden resolverse a una sensibilidad de moléculas individuales cuando pasan a través de nanoporos de hemolisina (Robertson *et al.* 2009).

Puede usarse un canal de α -hemolisina para detectar ácidos nucleicos a nivel de moléculas individuales (Kasianowicz *et al.* 1996). El polipéptido monomérico se autoensambla en una bicapa lipídica para formar un poro heptamérico, con una abertura limitante de 1,5 nm de diámetro. En una disolución acuosa de sal iónica, el poro formado por el canal de α -hemolisina transporta una corriente iónica intensa y estable cuando se aplica una tensión apropiada a través de la membrana. La abertura limitante del nanoporo permite que pasen a través moléculas de ácido nucleico monocatenario lineal pero no bicatenario (diámetro $\sim 2,0$ nm). Los ácidos nucleicos polianiónicos se impulsan a través del poro mediante el campo eléctrico aplicado, que bloquea o reduce la corriente iónica. Este paso genera una firma electrónica única. Por tanto, se obtendrá un diagrama de acontecimientos específico, que es la representación gráfica del tiempo de translocación frente a la corriente de bloqueo, y se usará para distinguir la longitud y la composición de polinucleótidos mediante técnicas de registro de canales individuales basándose en parámetros característicos tales como corriente de translocación, duración de translocación y sus dispersiones correspondientes en el diagrama. Se seleccionan cuatro etiquetas de PEG, que se ha mostrado que producen señales de bloqueo de corriente diferenciadas en nanoporos, para acoplarse con cuatro nucleótidos (A, C, G, T) en el fosfato terminal. Estos análogos de nucleótido novedosos se usan en una reacción de polimerasa y se usan nanoporos para detectar las etiquetas liberadas para decodificar las bases incorporadas tal como se muestra en la figura 33.

Hay varias ventajas en este enfoque:

- 1) Evitar el uso de muchas enzimas diferentes (ahorra coste y complejidad).
- 2) Adición de polifosfato de nucleótido unido a etiqueta individual de manera secuencial o los cuatro nucleótidos con diferentes etiquetas unidas a cada nucleótido.
- 3) Uso de PEG como etiquetas que pueden detectarse mediante nanoporos a una resolución de unidades individuales.
- 4) Secuenciación con detección de moléculas individuales en tiempo real a medida que la etiqueta pasa a través del nanoporo.
- 5) Secuenciación en paralelo de manera masiva, bajo coste y alto rendimiento.

Tal como se muestra en la figura 33, se inmoviliza ADN polimerasa en el nanoporo y se añade el molde-cebador junto con los nucleótidos marcados con etiqueta de PEG. Tras la incorporación del nucleótido marcado con etiqueta de PEG correcto, los fosfatos de PEG liberados pasan a través del nanoporo y se mide la señal de bloqueo electrónico. Los PEG de longitud diferente tienen señales de bloqueo diferentes, por tanto, pueden usarse 4 PEG diferentes para 4 nucleótidos diferentes.

Los nucleótidos pueden añadirse uno cada vez, si se añade el nucleótido correcto proporciona una señal de bloqueo

diferenciada. Sin embargo, si el nucleótido no es complementario a la base de ácido nucleico de molde, no se incorporará y por tanto no se detectará ninguna señal. De una manera en paralelo masiva, puede construirse una matriz de alta densidad de micro/nanopocillos para realizar el procedimiento bioquímico. Cada micro/nanopocillo contiene un molde de ADN diferente y dispositivo de nanoporos. Los PEG liberados se detectan a una sensibilidad de moléculas individuales.

Los métodos generales para la síntesis de 5'-polifosfato de nucleósido marcado con etiqueta se muestran en la figura 49. Pueden sintetizarse 5'-tri, tetra, penta o hexa-fosfatos de nucleósido marcados en fosfato terminal partiendo a partir de los 5'-trifosfatos de nucleósido (NTP) correspondientes. Por tanto, en primer lugar se activa el trifosfato con DCC/DMF que puede hacerse reaccionar directamente con el nucleófilo de etiqueta para dar NTP unido a etiqueta o puede hacerse reaccionar con un nucleófilo de ligador con el que puede hacerse reaccionar un NHS de etiqueta o una etiqueta activada de manera apropiada para proporcionar NTP unido a ligador con etiqueta. Para la síntesis de tetrafosfatos (N4P) o pentafofosfatos (N5P) de nucleósido unidos a etiqueta, en primer lugar se hace reaccionar el trifosfato activado con ácido fosfórico o pirofosfato para dar tetra y penta-fosfato, respectivamente, que puede hacerse reaccionar con nucleófilo de ligador seguido por la reacción con etiquetas activadas de manera apropiada.

La síntesis de nucleótidos marcados con PEG se comentó anteriormente en los ejemplos 2 y 3. Los nucleótidos marcados con PEG tienen -3, -4, -5 o -6 cargas basándose en el uso de tri, tetra, penta o hexa-fosfatos. Tras la reacción de extensión de cebador catalizada por polimerasa, la carga neta en las etiquetas de PEG liberadas será uno menos (-1) que el PEG-nucleótido de partida, lo cual es suficiente para distinguir mediante la señal de bloqueo iónico de nanoporo (además el PEG-nucleótido sin reaccionar es más voluminoso que los fosfatos de PEG liberados, de ahí una señal de bloqueo iónico diferente). Alternativamente, si está presente fosfatasa alcalina en la mezcla de reacción, el PEG liberado será neutro (los grupos fosfato libres se hidrolizan mediante fosfatasa alcalina). Las etiquetas de PEG liberadas también pueden cargarse de manera positiva tal como se muestra a continuación de modo que pueden detectarse fácilmente mediante nanoporos. De manera similar, también pueden cargarse altamente de manera negativa.

Síntesis de polifosfatos de nucleósido unidos a etiqueta cargada de manera positiva:

Los polifosfatos de nucleósido unidos a etiqueta cargada de manera positiva se sintetizan tal como se muestra en la figura 44. En primer lugar, se hace reaccionar un aminoácido de trimetil-(lisina)_n-glicina cargado de manera positiva (K((Me)₃)_n-Gly) con el éster de PEG-NHS y después se activa para formar el éster de PEG-K((Me)₃)_n-Gly-NHS. Este éster activado se hace reaccionar con el polifosfato de nucleósido terminado en amino tal como se muestra en las figuras 38 y 44. La carga neta en el tetrafosfato de nucleósido es neutra, pero tras la incorporación por polimerasa, el PEG liberado tiene una carga positiva de +1 y si se añade fosfatasa alcalina al cóctel de reacción, la carga neta en el PEG liberado es de +4. Por tanto la etiqueta liberada puede separarse fácilmente e identificarse pasando a través del nanoporo.

Síntesis de polifosfatos de nucleósido unidos a PEG bloqueados en 3' para secuenciación mediante síntesis con detección en nanoporos.

La síntesis de polifosfatos de nucleósido bloqueados en 3' sigue esencialmente la misma ruta tal como se muestra para polifosfatos de nucleósido unidos a etiqueta, excepto porque el 5'-trifosfato de nucleósido de partida es dNTP bloqueado en 3'-O. Tal como se muestra en la figura 50, en primer lugar se hace reaccionar 3'-O-azidometil-dNTP (6) con CDI o DCC/DMF seguido por reacción con ácido fosfórico (tetrafosfato) o pirofosfato (pentafofosfatos). Esto se hace reaccionar tras la purificación con el nucleófilo apropiado para proporcionar fosfato terminado en amino que después se hace reaccionar con el éster de PEG-NHS apropiado (neutro, cargado de manera positiva o cargado de manera negativa) para proporcionar el polifosfato de nucleósido unido a PEG bloqueado en 3'-O requerido.

Esquema de secuenciación con PEG-nucleótidos y detección en nanoporos (se inmovilizan muchas copias de una molécula de ADN en una perla y adición secuencial de un PEG-nucleótido cada vez).

Tal como se muestra en la figura 51, se inmovilizan las moléculas de ADN sobre una perla. Por tanto cada perla tiene muchas copias de la misma molécula de ADN. Se añade la perla a un micro/nanopocillo que está unido a un nanoporo. El ADN forma el complejo con la ADN polimerasa que o bien está unida al nanoporo o bien se añade al micro/nanopocillo junto con el nucleótido unido a PEG. Los nucleótidos pueden añadirse uno cada vez, si se añade el nucleótido correcto se incorpora y libera una etiqueta de PEG lo cual da una señal de bloqueo diferenciada cuando se hace pasar a través de un nanoporo. Sin embargo, si el nucleótido no es complementario a la base de ácido nucleico de molde, no se incorporará y por tanto no se detectará ninguna señal. En este caso, puede usarse el PEG de igual longitud y peso molecular en los cuatro nucleótidos, o, si se desea, también pueden usarse cuatro PEG diferentes. Por tanto, la adición de base de ácido nucleico puede detectarse fácilmente mediante la señal de bloqueo de nanoporo a una sensibilidad de moléculas individuales.

Secuenciación mediante síntesis con PEG-nucleótidos bloqueados en 3'-O y detección en nanoporos (se inmovilizan muchas copias de una molécula de ADN en una perla y adición simultánea de los cuatro PEG-nucleótidos bloqueados en 3'-O).

Las regiones homopoliméricas del ADN pueden secuenciarse de manera correcta usando este enfoque. Por tanto, si el grupo 3'-OH del nucleótido se bloquea mediante un resto reversible, la síntesis de ADN se detendrá tras la adición de tan sólo un nucleótido. La síntesis puede continuar tras la retirada del grupo de bloqueo para generar un grupo 3'-OH libre. Tal como se muestra en la figura 52, los cuatro 3'-O-azidometil-nucleótidos unidos a PEG de diferente tamaño pueden añadirse al micro/nanopocillo de reacción y, siempre que se incorpore un nucleótido correcto, se lee la etiqueta de PEG liberada mediante pase a través del nanoporo y se detecta la señal iónica. Dado que el grupo 3'-OH está bloqueado sólo se añade un nucleótido cada vez. Este grupo bloqueado en 3'-O puede escindirise mediante tratamiento con TECP y por tanto un grupo OH libre está listo para la incorporación de nucleótido adicional. Mediante adición de nucleótido y escisión repetidas, puede secuenciarse la región homopolimérica de manera correcta y fácil.

Pirosecuenciación en paralelo de manera masiva usando nanoporos:

Tal como se muestra en la figura 53, de una manera en paralelo masiva, puede construirse una matriz de alta densidad de micropocillos para realizar el procedimiento bioquímico. Cada micro/nanopocillo contiene un molde de ADN diferente y dispositivo de nanoporos. Los PEG liberados se detectan a una sensibilidad de moléculas individuales.

Resumen del experimento:

- 1) Cualquier etiqueta de diferente tamaño, longitud, peso molecular, carga, unida al fosfato terminal del nucleótido que puede detectarse mediante nanoporos tras incorporación por polimerasa.
- 2) Etiqueta unida a los tri, tetra, penta, hexa-fosfatos.
- 3) Detección electrónica
- 4) Grupo de moléculas de ADN unidas a la perla o superficie sólida y sensibilidad de detección de moléculas individuales (alta densidad y alta sensibilidad).
- 5) Región homopolimérica fácilmente secuenciada usando nucleótidos bloqueados en 3'-O unidos a etiqueta.
- 6) Añadir un nucleótido con etiqueta por ciclo.
- 7) Añadir los cuatro nucleótidos marcados con etiqueta de manera reversible juntos para secuenciar regiones homopoliméricas.
- 8) Alta sensibilidad, precisión y velocidad.
- 9) Secuenciación en paralelo masiva.

EJEMPLO 7

Espectrometría de masas/tamaño de moléculas individuales en disolución usando nanoporos

Método

Se formaron membranas de bicapa lipídica planas libres de disolvente a partir de difitanoil-fosfatidilcolina (1,2-difitanoil-sn-glicero-3-fosfolina; Avanti Polar Lipids, Alabaster, Ala.) en pentano (J. T. Baker, Phillipsburg, N.J.) sobre un orificio de $\sim 70 \mu\text{m}$ de diámetro en una división de Teflon de $25 \mu\text{m}$ de grosor que separa dos cámaras de Teflon idénticas. Se trató previamente el orificio con una disolución de 1:400 vol/vol de hexadecano (Aldrich, St. Louis, Mo.) en pentano. Ambas cámaras contenían KCl 4 M (Mallinckrodt, Paris, Ky.), 2-amino-2-hidroximetil-1,3-propanodiol 5 mM (Tris; Schwarz/Mann Biotech, Cleveland, Ohio), ajustado a pH 7,5 con ácido cítrico concentrado (Fluka, Buchs, Suiza).

Se formaron canales individuales añadiendo $\sim 0,25 \mu\text{g}$ de α -hemolisina (List Biological Laboratories, Campbell, Calif.) a la disolución en un lado de la división. Tras formarse un canal individual, se purgó rápidamente la primera cámara con tampón nuevo para impedir la incorporación de canales adicionales. A menos que se mencione lo contrario, los datos se obtuvieron con un potencial aplicado de -40 mV con dos electrodos de Ag/AgCl separados por el electrolito en volumen mediante puentes de sal de Vycor (KCl 3 M). Se midió la corriente usando un amplificador de registro fisiológico de fijación de voltaje Axopatch 200B (Molecular Devices, Sunnyvale, Calif.) y se filtró a 10 kHz con un filtro de Bessel de cuatro polos antes de la digitalización a 50 kHz.

La toxina α -hemolisina puede formar al menos dos conformeros que tienen propiedades de regulación y niveles de conductancia diferentes. En este caso sólo se usó el conformero de conductancia superior, que tiene una conductancia aproximadamente óhmica de $3,75 \text{ nS}$ entre $\pm 50 \text{ mV}$ (datos no mostrados). Se añadió PEG (PEG 1500 polidispersado; Fluka; o PEG 1294 monodispersado; Polypure, Oslo, Noruega) a la segunda cámara a partir de disoluciones madre de 12 mg/ml en electrolito hasta una concentración final de $0,045 \text{ mg/ml}$.

Se obtuvieron espectros de masas de MALDI-TOF de las muestras de PEG con un dispositivo Voyager DE-STR (PerSeptive Biosystems, Framingham, Masa.) usando el modo de reflectrón. Se produjo la desorción/ionización mediante irradiación con luz UV pulsada (337 nm) a partir de un láser de nitrógeno. Se hizo funcionar el instrumento a 25 kV en el modo de iones positivos usando un tiempo de retardo de extracción establecido a 600 ns. Se calculó el promedio de los espectros finales a partir de 100 disparos mientras se movía el láser sobre la superficie de la muestra con la potencia de láser establecida ligeramente por encima del umbral para la aparición de cada espectro. Se prepararon las muestras a partir de las disoluciones de PEG al 1% p/p en agua destilada. La disolución de matriz era 1:1 de acetonitrilo:agua saturada con ácido retinoico todo trans (Sigma, St. Louis, Mo.) con ácido trifluoroacético al 0,1% (Matheson, Joliet, Ill.) añadido. Se mezclaron la muestra y la matriz a 1:1 hasta un volumen total de 2 μ l antes de secarse.

Discusión

En ausencia de analito, la corriente iónica provocada por un potencial de CC está bien definida. El ruido intrínseco en la corriente iónica puede estar provocado en parte por el movimiento browniano de iones en el nanoporo y la capacitancia de barrera resistiva. La adición de analito (por ejemplo, polietilenglicol) provoca disminuciones transitorias bien definidas en la conductancia. Cada pulso puede corresponder a la presencia de una única molécula de PEG en el nanoporo. Las reducciones de corriente cubren un intervalo de tan sólo \sim 50 picoamperios para una muestra de PEG-1500 polidispersado (masa molecular promedio de \sim 1500 g/mol).

Los polímeros no electrolíticos provocan reducciones bien definidas en la corriente iónica a medida que se dividen en el interior de un nanoporo solitario en una membrana de bicapa lipídica. La corriente iónica, a través de un canal de α -hemolisina bañado por una disolución libre de polímero, es quiescente. La adición de PEG polidispersado ($M_r = 1.500$ g/mol) provoca pulsos de conductancia reducida de corriente persistente.

Un único nanoporo distingue entre polímeros con masas moleculares diferentes. La diferencia entre los estados de conductancia provocados por PEG polidispersado ($M_r = 1.500$) y monodispersado ($M = 1.294$ g/mol, $n = 29$) resulta fácilmente evidente. Los datos de serie temporal contenían \sim 500 y \sim 700 acontecimientos para las muestras de PEG poli y monodispersado, respectivamente. Los histogramas con todos los puntos de la corriente iónica reflejan las naturalezas diferenciadas de las dos muestras de polímero. Los histogramas de corriente iónica para cada muestra se calcularon a partir de $>10^5$ acontecimientos de pulso de conductancia reducida. Los pulsos de conductancia reducida de corriente iónica pequeña, de larga duración, cerca de cero en la serie temporal de PEG monodispersado están provocados lo más probablemente por impurezas en las muestras de PEG. Estos acontecimientos tienen larga duración pero son pocos en número.

La calibración del espectro de masas o tamaño puede lograrse mediante varias técnicas. Por ejemplo, repetir el experimento basado en conductancia usando un analito de tamaño convencional permite la asignación del pico de PEG 1294 g/mol en la muestra polidispersada indicado como datos de muestra polidispersada. Los picos circundantes en el histograma basado en conductancia están provocados por moléculas de PEG que difieren en una única unidad de etilenglicol (es decir, $\text{CH}_2\text{-CH}_2\text{-O}$). Una comparación de la distribución de tamaños basada en conductancia con un espectro de masas de MALDI-TOF de la misma muestra de PEG polidispersado demuestra la precisión de este método.

EJEMPLO 8

Secuenciación de moléculas individuales usando nucleótidos de polifosfatos marcados con etiqueta y nanoporos

Hay una necesidad significativa de secuenciar con precisión moléculas de ADN y ARN individuales para medicina personalizada. En el presente documento se describe una estrategia novedosa de secuenciación mediante síntesis (SBS) basada en nanoporos que diferencia con precisión a nivel de moléculas individuales cuatro etiquetas de tamaño diferente que están inicialmente unidas al 5'-fosfato de cada nucleótido. A medida que se incorpora cada nucleótido en la cadena de ADN en crecimiento durante la reacción de polimerasa, su etiqueta se libera mediante formación de enlace fosfodiéster entre el α -fosfato del nucleótido marcado con etiqueta y el grupo 3'-OH del nucleótido anterior. Las etiquetas liberadas entran en un nanoporo en el orden en el que se liberaron y producen una firma de bloqueo de corriente iónica única debido a su tamaño, forma y carga, determinando así la secuencia de ADN de manera electrónica a nivel de moléculas individuales con una resolución de bases individuales. Como ejemplo no limitativo, se unen cuatro etiquetas de PEG-cumarina de longitud diferente al fosfato terminal de 5'-tetrafosfato de 2'-desoxiguanosina. Se observa la incorporación eficaz de estos nucleótidos modificados durante la reacción de polimerasa, y una distinción de etiqueta mejor que 6σ entre las cuatro etiquetas basándose en el grado al que diferentes etiquetas reducen la corriente iónica de nanoporo. El enfoque molecular descrito en este caso, acoplado con polimerasa unida de manera covalente a los nanoporos en un formato de matriz, produce una plataforma de SBS basada en nanoporos de moléculas individuales.

MÉTODOS

Síntesis de análogos de nucleótido de cumarina-PEG-dG4P

Todos los nucleótidos se purifican mediante HPLC de fase inversa en una columna de 150 x 4,6 mm (Supelco), fase

móvil: A, Et₃N 8,6 mM/1,1,1,3,3,3-hexafluoro-2-propanol 100 mM en agua (pH 8,1); B, metanol. Se realiza la elución a partir del 100% de A isocrático a lo largo de 10 min seguido por un gradiente lineal del 0-50% de B durante 20 min y después el 50% de B isocrático a lo largo de otros 30 min.

A. Síntesis de cumarina-PEG_n-dG4P:

5 La síntesis de cumarina-PEG_n-dG4P implica tres etapas tal como se muestra en el esquema en la figura 15.

A.1 Síntesis de 5'-tetrafosfato de 2'-desoxiguanosina (dG4P):

10 En primer lugar, se lleva a cabo la síntesis de 2'-dG4P partiendo de 2'-dGTP. Se convierten 300 umoles de 2'-dGTP (sal de trietilamonio) en la sal de tributilamonio usando 1,5 mmol (5 eq) de tributilamina en piridina anhidra (5 ml). Se concentra la disolución resultante hasta sequedad y se evapora conjuntamente con 5 ml de DMF anhidra (x2). Se disuelve el dGTP (sal de tributilamonio) en 5 ml de DMF anhidra, y se añaden 1,5 mmol de 1,1-carbonildiimidazol. Se agita la reacción durante 6 h, tras lo cual se añaden 12 ul de metanol y se continúa agitando durante 30 min. A esta disolución se le añaden 1,5 mmol de ácido fosfórico (sal de tributilamonio, en DMF) y se agita la mezcla de reacción durante la noche a temperatura ambiente.

15 Se diluye la mezcla de reacción con agua y se purifica en una columna Sephadex-A25 usando un gradiente de TEAB de 0,1 M a 1 M (pH 7,5). El dG4P eluye al final del gradiente. Se combinan las fracciones apropiadas y se purifican adicionalmente mediante HPLC de fase inversa para proporcionar 175 umol del tetrafosfato puro (dG4P). ³¹P-RMN: δ, -10,7 (d, IP, α-P), -11,32 (d, IP, δ-P), -23,23 (dd, 2P, β, γ-P); ESI-EM (modo positivo): calc. 587,2; hallado 585,9 (M-2).

A.2 Síntesis de dG4P-heptil-NH₂:

20 A 80 umol de dG4P en 2 ml de agua y 3,5 ml 1-metilimidazol-HCl 0,2 M (pH 6) se le añaden 154 mg de EDAC y 260 mg de diaminoheptano. Se ajusta el pH de la disolución resultante a 6 con HCl conc. y se agita a temperatura ambiente durante la noche. Se diluye esta disolución con agua y se purifica mediante cromatografía de intercambio iónico Sephadex-A25 seguida por HPLC de fase inversa para dar ~20 μmol de dG4P-NH₂. Esto se confirma mediante datos de ESI-EM (modo positivo): calc. 699,1; hallado (698,1, M-1).

25 B. Síntesis de cumarina-PEG-ácidos y ésteres de NHS:

30 Se hacen reaccionar los amino-dPEG-ácidos comercialmente disponibles (amino-d(PEG)₁₆, 20, 24, 36-ácidos; Quanta Biodesign) con éster de 6-metoxi-cumarina-NHS para proporcionar la cumarina-(PEG)_n-ácido correspondiente. Se disuelve amino-PEG-ácido (1 eq) en tampón de carbonato-bicarbonato (pH 8,6), seguido por adición de cumarina-NHS (1 eq) en DMF, y se agita la mezcla de reacción durante la noche. Se purifica la cumarina-PEG-ácido mediante cromatografía en gel de sílice usando una mezcla de CH₂Cl₂-MeOH (5-15%) y se combinan las fracciones apropiadas. Estos compuestos se analizan mediante análisis por ¹H-RMN y MALDI-TOF-EM.

Datos de MALDI-TOF-EM:

	Cumarina-PEG16-ácido	Cumarina-PEG20-ácido	Cumarina-PEG24-ácido	Cumarina-PEG36-ácido
PM esperado	996	1.172	1.348	1.877
PM observado*	1.016	1.192	1.368	1.899

*Diferencia en valores observados debido a la presencia de sal de sodio.

35 Se convierten los cumarina-PEG-ácidos en los ésteres de NHS correspondientes mediante reacción con 1,5 eq de carbonato de disuccinimidilo (DSC) y 2 eq de trietilamina en DMF anhidra durante 2 h. El éster de NHS resultante, que se mueve ligeramente más rápido que el ácido en placas de gel de sílice, se purifica mediante cromatografía en gel de sílice usando una mezcla de CH₂Cl₂-MeOH (5-15%) y se usa en la siguiente etapa.

C. Cumarina-PEG_n-dG4P:

40 Se lleva dG4P-heptil-NH₂ de la etapa A anterior a tampón de carbonato-bicarbonato 0,1 M (pH 8,6) y a esta disolución con agitación se le añade uno de los compuestos de cumarina-PEG-NHS (en DMF). Se agita la mezcla resultante durante la noche a temperatura ambiente y después se purifica sobre un cartucho de gel de sílice (el 15-25% de MeOH en CH₂Cl₂ para retirar la cumarina-ácido o NHS sin reaccionar y después isopropanol/NH₄OH/H₂O 6:4:1). Esto se purifica adicionalmente dos veces mediante HPLC de fase inversa para proporcionar cumarina-PEG-dG4P puro. La estructura se confirma mediante análisis en MALDI-TOF-EM. Cumarina-PEG16-dG4P: tiempo de retención, 31,7 min; cumarina-PEG20-dG4P: tiempo de retención, 32,2 min; cumarina-PEG24-dG4P: tiempo de retención, 33,0 min; cumarina-PEG36-dG4P: tiempo de retención, 34,3 min.

Datos de MALDI-TOF-EM:

	Cumarina-PEG16-	Cumarina-PEG20-	Cumarina-PEG24-	Cumarina-PEG36-
--	-----------------	-----------------	-----------------	-----------------

	dG4P	dG4P	dG4P	dG4P
PM esperado	1.673	1.850	2.025	2.554
PM observado	1.682	1.858	2.036	2.569

Reacciones de extensión de ADN polimerasa usando cumarina-PEG_n-dG4P:

Se realizan reacciones de extensión usando un molde-cebador en bucle (5'-GATCGCGCCGCGCCTTGGCGCGGCGC-3', PM 7966), en el que la siguiente base complementaria en el molde es una C, permitiendo la extensión mediante una única G (figura 56). Cada reacción de extensión se lleva a cabo en un ciclador térmico GeneAmp PCR System 9700 (Applied Biosystems) a 65°C durante 25 minutos en reacciones de 20 µl que consisten en molde-cebador en bucle 3 mM, 1X tampón Therminator γ (KCl 50 mM, Tris-HCl 20 mM, MgSO₄ 5 mM, IGEPAL CA-630 al 0,02% (pH 9,2 a 25°C)), 2 unidades de ADN polimerasa de Therminator γ (New England Biolabs), y 15 µM de uno de los nucleótidos de cumarina-PEG-dG4P. Se precipitan los productos de extensión de ADN con etanol, se purifican a través de columnas C18 ZipTip (Millipore) y se caracterizan mediante MALDI-TOF-EM usando un instrumento ABI Voyager. Tal como se muestra en la figura 58, se obtienen cuatro productos idénticos (peso molecular esperado de 8.295).

Se repiten reacciones de extensión de polimerasa para cada cumarina-PEG_n-dG4P y se tratan los productos (cumarina-PEG_n-trifosfato, figura 57) con fosfatasa alcalina (1 U a 37°C durante 15 min) para producir las etiquetas de cumarina-PEG_n-NH₂. Se extraen en diclorometano y se caracterizan mediante análisis de MALDI-TOF-EM.

Hidrólisis ácida de cumarina-PEG-dG4P (figura 57):

Se añade ácido acético a los nucleótidos de la cumarina-PEG-dG4P hasta una concentración final del 10% y se agita vigorosamente la disolución durante la noche para garantizar la hidrólisis del enlace N-P entre el fosfato δ y la heptilamina. Se seca la disolución usando un dispositivo CentriVap y vuelve a suspenderse en un volumen apropiado de agua. Se recoge una alícuota de 1 µl para caracterización mediante espectrometría de masas de MALDI-TOF y se mide una segunda alícuota a 260 nm y 350 nm usando un espectrofotómetro NanoDrop ND-1000.

Los compuestos de cumarina-PEG-amina resultantes tienen el tamaño esperado según se mide mediante MALDI-TOF-EM (véase la figura 16 y la siguiente tabla).

	Cumarina-PEG16-NH ₂	Cumarina-PEG20-NH ₂	Cumarina-PEG24-NH ₂	Cumarina-PEG36-NH ₂
PM esperado	1.107	1.284	1.460	1.988
PM observado	1.115	1.289	1.465	1.991

Mediciones en nanoporos:

Formación de membrana y canal

Se insertan canales de α-hemolisina individuales en membranas de bicapa lipídica (BLM) planas libres de disolvente (Montal *et al.*, 1972) fabricados a través de un orificio de ~80 µm de diámetro en una división de Teflon de 25 µm de grosor que separa dos pocillos de disolución de electrolito tal como se describió anteriormente (Reiner *et al.* 2010). Se usa KCl 4 M, Tris 10 mM ajustado a pH 7,2 con ácido cítrico a lo largo de todo el experimento. Se forman membranas humedeciendo en primer lugar la división con el 1% v/v de hexadecano/pentano. Se dispersa diftanoilfosfatidilcolina (DPhypC) 10 mg/ml en pentano en ambas superficies de contacto de aire-disolución de electrolito con los niveles de disolución muy por debajo del orificio en la división de Teflon. Después de 10 min, se aumentan los niveles de disolución por encima del orificio de manera espontánea para formar una membrana. Se inyectan aproximadamente 0,5 µl de α-hemolisina 0,5 mg/ml en la disolución de manera inmediatamente adyacente a la membrana y se observa la corriente iónica hasta que se inserta un canal individual en la membrana. Después se intercambia el contenido en la cámara cis por disolución de electrolito libre de proteína para mantener un canal individual.

Se añaden moléculas de cumarina-PEG_n-NH₂ (n = 16, 20, 24 y 36) al lado trans del poro (definido como el lado de cilindro β del canal) hasta una concentración final de entre 0,4 µmol/l y 1 µmol/l de cada componente. Se registra la corriente iónica entre dos Ag/AgCl coincidentes (KCl 3 M) a un potencial fijado (-40 mV) durante aproximadamente 15 min para lograr datos estadísticos de recuento suficientes. Se registran los datos con un filtro de Bessel de 4 polos a 10 kHz sobremuestreado a 50 kHz.

Análisis de datos

Se analizan los datos fuera de línea con un programa interno escrito en LabVIEW (National Instruments) tal como se describió anteriormente (Rodrigues *et al.* 2008). En resumen, se localizan bloqueos con un detector de acontecimientos basándose en un algoritmo de umbral simple establecido a 5 σ del ruido de corriente en el estado abierto. Cuando se detecta un acontecimiento, los puntos en el tiempo de aumento y el tiempo de disminución se descartan (~60 µs y 20 µs, respectivamente). La profundidad de bloqueo media se calcula a partir de los puntos

restantes y la corriente de canal abierto se calcula partir de la media de 0,8 ms de datos de canal abierto separados 0,2 ms a partir del umbral. Los datos se notifican como razón de las medias ($\langle i \rangle / \langle i_{abierto} \rangle$) y los espectros de nanoporos se calculan como histograma de estos valores.

Discusión

5 En 1996, Kasianowicz *et al.* (Kasianowicz *et al.* 1996) demostraron por primera vez que puede usarse el canal de α -hemolisina (α HL) para detectar ácidos nucleicos al nivel de moléculas individuales. El canal de α HL tiene una
 10 abertura limitante de 1,5 nm de diámetro (Song *et al.* 1996; Bezrukov *et al.* 1996; Krasilnikov 2002; Kasianowicz 1995) y su regulación dependiente de tensión puede controlarse, de tal manera que el poro permanece abierto de
 15 manera indefinida (Kasianowicz 1995) lo que hace que sea un candidato ideal para la detección y distinción basadas en nanoporos. Se impulsan ácidos nucleicos polianiónicos monocatenarios individuales a través del poro mediante el
 campo eléctrico aplicado, y los polinucleótidos provocan reducciones transitorias bien definidas en la conductancia de poro (Kasianowicz *et al.* 1996; Vercoutere *et al.* 2001; Deamer *et al.* 2002; Kasianowicz 2004). Dado que el
 tiempo de residencia del polinucleótido en el poro es proporcional a la longitud de contorno de ARN o ADN, se sugiere que un nanoporo puede poder secuenciar ADN a modo de cinta de teletipo si las cuatro bases pueden
 20 distinguirse unas de otras (Kasianowicz *et al.* 1996). Hacia este objetivo (Kasianowicz 1996; Kasianowicz *et al.* 2008; Kasianowicz *et al.* 2002), se usa un canal de α HL con un adaptador unido de manera covalente en el poro para
 identificar 5'-monofosfatos de nucleósido sin marcar (Clarke *et al.* 2009). Sin embargo, no se ha documentado un sistema de exonucleasa-nanoporo completo basado en este concepto para secuenciar ADN.

25 A pesar de la capacidad de nanoporos para detectar y caracterizar algunas propiedades físicas del ADN a nivel de moléculas individuales, aún no se ha realizado el objetivo más demandante de una secuenciación entre bases
 precisa haciendo pasar un ADN monocatenario a través del nanoporo. Oxford Nanopore Technologies anunció recientemente la capacidad de lograr la secuenciación de cadenas en un nanoporo a una resolución de 3 bases con
 una tasa de error del 4% (AGBT Meeting, 2012). Otro grupo notificó una secuenciación de cadenas con resolución de bases individuales con un nanoporo, pero tuvo dificultades para determinar correctamente secuencias de
 homopolímeros (Manrao *et al.* 2012).

El canal de α HL nativo tiene la capacidad inherente de distinción molecular a alta resolución. Por ejemplo, puede distinguir entre iones H^+ y D^+ acuosos, (Kasianowicz *et al.* 1995) y Robertson *et al.* (2007) demostraron
 30 recientemente que el canal puede separar fácilmente moléculas de polietilenglicol (PEG) a nivel monomérico. En este último estudio, un espectro de tamaño o masas moleculares estimado a partir de la corriente media provocada
 por moléculas de PEG individuales resuelve fácilmente las unidades de repetición de etilenglicol. Además, el tiempo de residencia medio del polímero en el poro aumenta con la masa de PEG (Robertson *et al.* 2007; Reiner *et al.*
 2010). Basándose en estas observaciones y el hecho de que la ADN polimerasa puede reconocer análogos de nucleótido con extensa modificación en el grupo fosfato terminal en 5' como sustratos eficaces (Kumar 2005, 2006,
 2008; Sood *et al.* 2005; Eid *et al.* 2009), se desarrolla un enfoque de secuenciación de ADN de moléculas
 35 individuales novedoso que puede identificar bases individuales mediante la detección y diferenciación de un subproducto liberado (por ejemplo, etiquetas de PEG de diferente longitud a partir de la reacción de ADN
 polimerasa, figura 9) en vez de los propios nucleótidos.

En este enfoque, durante la formación de enlace fosfodiéster en la reacción de polimerasa, la escisión del enlace fosfato α - β en el nucleótido incorporado libera la etiqueta. En la figura 30 se muestra un ejemplo de una secuencia
 40 de reacción de cuatro bases con diferentes etiquetas para cada base. Una matriz de tales nanoporos, uno de los cuales se muestra gráficamente en la figura 33, cada uno con una polimerasa unida de manera covalente adyacente
 a una de las entradas de poro, permite la secuenciación mediante síntesis (SBS) de moléculas individuales. La adición de cada nucleótido se determina mediante el efecto de la etiqueta liberada sobre la conductancia de
 nanoporo.

45 Este sistema de SBS basado en etiqueta de 5'-fosfato ofrece una ventaja con respecto a la secuenciación de cadenas a través de nanoporos en cuanto a que la velocidad de tránsito a través del poro ya no es un problema,
 porque la tasa de extensión de polimerasa y liberación es más lenta que el tiempo de tránsito de etiqueta a través del poro. Esto también puede eliminar problemas de fase inherentes a métodos de secuenciación de cadenas. Se
 describe la síntesis y la incorporación eficaz de nucleótidos con etiquetas unidas a 5'-fosfato que presentan cuatro
 50 PEG de diferente longitud y un resto cumarina. Se demuestran cuatro patrones de bloqueo de corriente diferenciados de las etiquetas liberadas en un poro de α -hemolisina al nivel de moléculas individuales, estableciendo
 la viabilidad del enfoque de SBS electrónico de moléculas individuales.

Diseño, síntesis y caracterización de nucleótidos marcados con PEG

55 Los cuatro 5'-tetrafosfatos de 2'-desoxiguanosina marcados con etiqueta en 5'-fosfato (figura 54) se sintetizan según el esquema de síntesis generalizado mostrado en la figura 15 (véanse los métodos anteriores para detalles). En
 primer lugar, se convierte 5'-trifosfato de 2'-desoxiguanosina (dGTP) en 5'-tetrafosfato de 2'-desoxiguanosina (dG4P) y después se añade un ligador de diaminoheptano al fosfato terminal del tetrafosfato, con el fin de unir etiquetas de
 PEG de diferente longitud. En un conjunto separado de reacciones, se hace reaccionar éster N-hidroxisuccinimídico de 6-metoxi-cumarina con una de cuatro moléculas de amino-PEG-COOH con 16, 20, 24 ó 36 unidades de

etilenglicol, para producir moléculas de cumarina-PEG_n-COOH, que posteriormente se convierten en los ésteres de NHS correspondientes. Estos se hacen reaccionar con la dG4P-heptilamina (dG4P-NH₂) para generar los cuatro análogos de nucleótido finales, abreviados cumarina-PEG_n-dG4P (figura 54). El resto de cumarina se usa para rastrear la purificación de productos intermedios y los análogos de nucleótido finales. La síntesis de las moléculas esperadas se confirma mediante espectroscopía de masas de MALDI-TOF (figura 55).

Los nucleótidos de cumarina-PEG-dG4P se emplean en reacciones de extensión de polimerasa usando la variante de Therminator γ de ADN polimerasa. Se diseña un cebador-bucle-molde en el que la siguiente base complementaria es una C, permitiendo que se añada dGMP al cebador de ADN (figura 56). Se libera cumarina-PEG-trifosfato durante la reacción (figura 57). MALDI-TOF-EM confirmó que de hecho cada uno de los cuatro análogos de dG4P dio el producto de extensión correcto con una eficacia de incorporación del 100%, tal como se muestra mediante la aparición de un pico a 8.290 Dalton (figura 58). La ausencia de un pico de cebador a 7.966 Dalton sugiere que la reacción avanzó esencialmente hasta completarse.

Toda la incorporación representó los análogos de cumarina-PEG-dG4P, y no el posible dGTP o dG4P residual, dado que las moléculas se purifican dos veces en un sistema de HPLC que separa estas moléculas eficazmente con una diferencia de tiempo de retención de más de 10 min entre los dos compuestos. Para excluir adicionalmente esta posibilidad, se tratan los análogos de cumarina-PEG-dG4P purificados con fosfatasa alcalina, que puede degradar cualquier tri o tetra-fosfato contaminante para dar el nucleósido libre, y se usan los cumarina-PEG-nucleótidos purificados de nuevo mediante HPLC resultantes en reacciones de extensión. Resulta importante que las cadenas extendidas contienen nucleótidos naturales sin ninguna modificación, permitiendo que la SBS continúe a lo largo de longitudes extensas.

Las etiquetas liberadas a partir de reacciones de polimerasa son cumarina-PEG-trifosfato (cumarina-PEG-P₃, figura 57). Para reducir la complejidad de la carga en las etiquetas, se tratan las etiquetas liberadas con fosfatasa alcalina, produciendo etiquetas de cumarina-PEG-NH₂, que se analizan para determinar sus efectos de bloqueo de corriente en nanoporos. En el desarrollo adicional del sistema de SBS basado en nanoporos, se trata la cumarina-PEG-P₃ liberada con fosfatasa alcalina, que, como la polimerasa, puede unirse a una entrada de los nanoporos, para generar etiquetas de cumarina-PEG-NH₂, u optimizar las condiciones para usar nanoporos para detectar directamente las etiquetas cargadas liberadas. En el estudio objeto, con el fin de obtener grandes cantidades de material para pruebas mediante MALDI-TOF-EM y nanoporos de proteína, se producen versiones sintéticas de las etiquetas liberadas esperadas (cumarina-PEG-NH₂) mediante hidrólisis ácida de los cuatro análogos de cumarina-PEG-nucleótido para escindir el enlace P-N entre el polifosfato y el resto heptilamina (figura 57).

EJEMPLO 9

Caracterización de las etiquetas liberadas mediante MALDI-TOF-EM

Las moléculas de cumarina-PEG-NH₂ esperadas se confirman mediante análisis de MALDI-TOF-EM, tras la purificación mediante HPLC (figura 16). Los resultados de MALDI-TOF-EM indican que las etiquetas de cumarina-PEG-NH₂ generadas mediante hidrólisis ácida son idénticas a las etiquetas liberadas producidas durante la reacción de polimerasa tras el tratamiento con fosfatasa alcalina.

Con referencia a la figura 16, las etiquetas de cumarina-PEG-NH₂ generadas mediante hidrólisis ácida de cumarina-PEG16-dG4P que produce cumarina-PEG16-NH₂, cumarina-PEG20-dG4P que produce cumarina-PEG20-NH₂, cumarina-PEG24-dG4P que produce cumarina-PEG24-NH₂ y cumarina-PEG36-dG4P que produce cumarina-PEG36-NH₂, son idénticas a las etiquetas liberadas correspondientes generadas en reacciones de extensión de polimerasa tras el tratamiento con fosfatasa alcalina, tal como se muestra mediante análisis de MALDI-TOF-EM. Se muestra una imagen compuesta de cuatro espectros de EM obtenidos por separado. Las estructuras de las etiquetas de cumarina-PEG-NH₂ se muestran a continuación.

EJEMPLO 10

Distinción de etiquetas liberadas en nanoporos de proteína a moléculas individuales

Con referencia a la figura 17, se combinaron cuatro compuestos de cumarina-PEG_n-NH₂ (n = 16, 20, 24 y 36), derivados a partir de los cuatro nucleótidos comparables mediante hidrólisis ácida, y se diluyeron en KCl 4 M, Tris 10 mM, pH 7,2 para la medición en nanoporos. Los datos de serie temporal a la izquierda indican que cuando estas etiquetas de PEG entran en un canal iónico de α -hemolisina individual, provocan bloqueos de corriente que son característicos de su tamaño. Un histograma a la derecha muestra el bloqueo de corriente medio provocado por moléculas individuales que muestra una resolución de línea base con un ancho de banda de medición de 10 kHz. Las barras coloreadas en la parte superior representan la distribución de 6 σ de los datos (suponiendo distribuciones gaussianas para cada una de las cuatro etiquetas de PEG que pueden representar cada uno de los cuatro nucleótidos de ADN), lo cual sugiere que puede distinguirse una base individual con una precisión mejor que 1 de cada 300.000 acontecimientos, representada en esta figura usando las designaciones A, C, G y T, lo cual puede producirse cuando se usan cuatro nucleótidos diferentes con cuatro PEG de longitud diferente para la secuenciación de ADN.

Para demostrar el enfoque de SBS de moléculas individuales electrónico, se someten a prueba cuatro etiquetas de cumarina-PEG_n-NH₂ liberadas para determinar sus efectos de bloqueo de corriente sobre un nanoporo de α HL (figura 17). La distribución de frecuencia relativa del histograma de acontecimientos de bloqueo ($\langle i \rangle / \langle i_{abierto} \rangle$) muestra cuatro picos diferenciados y bien separados para las cuatro etiquetas de cumarina-PEG_n-NH₂ (n = 36, 24, 20 y 16 de izquierda a derecha respectivamente en la figura 17, parte inferior derecha).

Para destacar la amplia separación de los picos, y ofrecer una clara evidencia de que puede lograrse la detección de un nucleótido específico mediante la señal de bloqueo única proporcionada por su PEG liberado, los picos se ajustan con funciones gaussianas individuales y se muestran las distribuciones de error de 6σ correspondientes (rectángulos coloreados en la parte superior en la figura 17, parte inferior derecha). Las moléculas de cumarina-PEG-NH₂ aplicadas por separado se caracterizan con el poro (datos no mostrados), lo cual confirmó la identidad de los picos relacionados con PEG mostrados en esta figura.

Tal como se describe en este caso, un canal iónico de α HL individual puede separar moléculas individuales basándose en su tamaño y resuelve fácilmente una mezcla de PEG hasta mejor que el tamaño de una unidad de monómero individual (es decir, <44 g/mol). Esta alta resolución surge de las interacciones entre el polímero de PEG, el electrolito (cationes móviles) y las cadenas laterales de aminoácido que revisten la luz del canal de α HL. Estas interacciones permiten usar el poro como sensor a escala nanométrica que es específico para el tamaño, la carga y la propiedad química de un analito.

En este caso, tal análisis se extiende a PEG con diferentes grupos químicos en cualquier extremo terminal. El registro de corriente iónica de canales individuales en la figura 17, parte superior y parte inferior izquierda, ilustra los bloqueos provocados por las cuatro moléculas de cumarina-PEG_n-NH₂ de tamaño diferente, una cada vez. Como con PEG sin modificar, cada uno de los bloqueos de corriente es unimodal (es decir, se describe bien con distribuciones gaussianas y valores medios bien definidos).

Para distinguir con precisión entre las cuatro bases (A, C, G y T) para la secuenciación en nanoporos, se necesita adoptar una o más de las siguientes estrategias: 1) potenciar y diferenciar la intensidad de las señales de detección; 2) desarrollar un método eficaz para distinguir y procesar las señales de bloqueo electrónico generadas; 3) controlar la tasa de translocación de ácidos nucleicos a través del poro, por ejemplo, ralentizando el movimiento de ADN para la secuenciación de cadenas; y 4) diseñar y realizar nanoporos sintéticos nuevos y más eficaces. Tal como se demuestra en este caso, transformar el problema de resolver las bases individuales en el de distinguir entre cuatro etiquetas únicas resuelve esencialmente los tres primeros problemas.

En este caso, se demuestra un enfoque novedoso para potenciar la distinción de cuatro nucleótidos modificándolos en el resto de fosfato terminal. Kumar *et al.* notificaron por primera vez sobre la modificación de 5'-trifosfatos de nucleósido, o bien introduciendo más grupos fosfato para producir tetra y penta-fosfatos e introduciendo colorante directamente en el fosfato terminal o bien uniendo un ligador entre el fosfato terminal y el colorante (Kumar *et al.* 2006, 2008). Se mostró que los tetra y penta-fosfatos eran mejores sustratos de ADN polimerasa y se han usado ampliamente nucleótidos de fosfato marcados con fluoróforo para la secuenciación de ADN (Kumar *et al.* 2005; Sood *et al.* 2005; Eid *et al.* 2009; Sims *et al.* 2011).

El sistema de SBS de nanoporos de moléculas individuales, que se muestra esquemáticamente en la figura 33, representa la ADN polimerasa unida en estrecha proximidad a la entrada de nanoporo y se añade un molde que va a secuenciarse junto con el cebador. A este complejo de molde-cebador se le añaden juntos cuatro nucleótidos marcados con etiqueta de manera diferente. Después de que la polimerasa catalice la incorporación del nucleótido correcto, se libera el polifosfato unido a etiqueta y pasa a través del nanoporo para generar una señal de bloqueo de corriente, identificando así la base añadida. Cada etiqueta genera una firma de bloqueo electrónica diferente debido al tamaño, masa o carga diferentes.

Las propiedades físicas y químicas de la etiqueta pueden ajustarse adicionalmente para optimizar la eficacia de captura y la precisión de medición. Por ejemplo, la inserción de un ligador cargado de manera positiva que consiste en cuatro lisinas o argininas entre el polifosfato y el PEG produce precursores con una carga neutra y etiquetas liberadas con una carga positiva neta. Usando la magnitud y signo apropiados del potencial, las etiquetas liberadas, pero no los sustratos de nucleótido, se transportan a través del poro.

Puede lograrse una distinción adicional de sustrato y producto mediante la inclusión de varias moléculas de fosfatasa alcalina unidas de manera covalente adyacentes a la polimerasa en el borde de cada nanoporo, lo cual garantiza una carga positiva incluso superior en las etiquetas. Es importante que cada etiqueta liberada en una reacción de polimerasa se mantenga en el orden apropiado. Por tanto, se necesitan varias enzimas fosfatasa para cada molécula de polimerasa debido a las tasas de renovación similares para las dos enzimas.

A pesar de todas estas precauciones, algunos nucleótidos sin reaccionar pueden entrar en el poro. Por tanto, la capacidad para distinguir entre etiquetas escindidas y nucleótidos sin reaccionar es importante; deben diferenciarse fácilmente debido a sus diferencias significativas de tamaño y carga, una capacidad inherente del sistema de nanoporos.

El método descrito en el presente documento puede aplicarse o bien a nanoporos de proteínas (por ejemplo α HL, porina A de *Mycobacterium smegmatis*, MspA) (Derrington *et al.* 2010) o bien a nanoporos de estado sólido (Garaj *et al.* 2010; Hall *et al.* 2010; Merchant *et al.* 2010; Schneider *et al.* 2010; Storm *et al.* 2005; Wanunu *et al.* 2008). Estas estrategias proporcionan nanoporos con diferentes propiedades que son apropiados para detectar una biblioteca de etiquetas. Para implementar esta estrategia novedosa para la secuenciación de ADN, puede construirse una matriz de nanoporos³⁷ en una superficie plana para facilitar la secuenciación de ADN en paralelo de manera masiva.

En conclusión, se demuestra una plataforma de secuenciación de ADN de moléculas individuales basada en SBS y en nanoporos que aprovecha etiquetas liberables novedosas en los sustratos de nucleótido para la reacción de polimerasa. Tal plataforma puede realizar lecturas largas y precisas, y una secuenciación de ADN de moléculas individuales electrónicas de rendimiento muy alto.

EJEMPLO 11

Síntesis de análogos de nucleótido de cumarina-PEG-dG4P

Todos los nucleótidos se purifican mediante HPLC de fase inversa en una columna de 150 x 4,6 mm (Supelco), fase móvil: A, Et₃N 8,6 mM/1,1,1,3,3,3-hexafluoro-2-propanol 100 mM en agua (pH 8,1); B, metanol. Se realiza la elución a partir del 100% de A isocrático a lo largo de 10 min seguido por un gradiente lineal del 0-50% de B durante 20 min y después el 50% de B isocrático a lo largo de otros 30 min.

Síntesis de cumarina-PEG_n-dG4P:

La síntesis de cumarina-PEG_n-dG4P implica tres etapas tal como se muestra en el esquema en la figura 15.

A.1 Síntesis de 5'-tetrafosfato de 2'-desoxiguanosina (dG4P):

En primer lugar se lleva a cabo la síntesis de 2'-dG4P partiendo de 2'-dGTP. Se convierten 300 umoles de 2'-dGTP (sal de trietilamonio) en la sal de tributilamonio usando 1,5 mmol (5 eq) de tributilamina en piridina anhidra (5 ml). Se concentra la disolución resultante hasta sequedad y se evapora conjuntamente con 5 ml de DMF anhidra (x2). Se disuelve el dGTP (sal de tributilamonio) en 5 ml de DMF anhidra y se añaden 1,5 mmol de 1,1-carbonildimidazol. Se agita la reacción durante 6 h, tras lo cual se añaden 12 ul de metanol y se continúa agitando durante 30 min. A esta disolución, se le añaden 1,5 mmol de ácido fosfórico (sal de tributilamonio, en DMF) y se agita la mezcla de reacción durante la noche a temperatura ambiente.

Se diluye la mezcla de reacción con agua y se purifica en una columna Sephadex-A25 usando un gradiente de TEAB de 0,1 M a 1 M (pH 7,5). El dG4P eluye al final del gradiente. Se combinan las fracciones apropiadas y se purifican adicionalmente mediante HPLC de fase inversa para proporcionar 175 umol del tetrafosfato puro (dG4P). ³¹P-RMN: δ , -10,7 (d, IP, α -P), -11,32 (d, IP, δ -P), -23,23 (dd, 2P, β , γ -P); ESI-EM (modo positivo): calc. 587,2; hallado 585,9 (M-2).

A.2 Síntesis de dG4P-heptil-NH₂:

A 80 umol de dG4P en 2 ml de agua y 3,5 ml de 1-metilimidazol-HCl 0,2 M (pH 6) se le añaden 154 mg de EDAC y 260 mg de diaminoheptano. Se ajusta el pH de la disolución resultante a 6 con HCl conc. y se agita a temperatura ambiente durante la noche. Se diluye esta disolución con agua y se purifica mediante cromatografía de intercambio iónico Sephadex-A25 seguida por HPLC de fase inversa para dar ~20 μ mol de dG4P-NH₂. Esto se confirma mediante datos de ESI-EM (modo positivo): calc. 699,1; hallado (698,1, M-1).

B) Síntesis de cumarina-PEG-ácidos y ésteres de NHS:

Se hacen reaccionar los amino-dPEG-ácidos comercialmente disponibles (amino-d(PEG)16, 20, 24, 36-ácidos; Quanta Biodesign) con éster de 6-metoxi-cumarina-NHS para proporcionar la cumarina-(PEG)_n-ácido correspondiente. Se disuelve amino-PEG-ácido (1 eq) en tampón de carbonato-bicarbonato (pH 8,6), seguido por adición de cumarina-NHS (1 eq) en DMF, y se agita la mezcla de reacción durante la noche. Se purifica la cumarina-PEG-ácido mediante cromatografía en gel de sílice usando una mezcla de CH₂Cl₂-MeOH (5-15%) y se combinan las fracciones apropiadas. Estos compuestos se analizan mediante análisis por ¹H-RMN y MALDI-TOF-EM.

Datos de MALDI-TOF-EM:

	Cumarina-PEG16-ácido	Cumarina-PEG20-ácido	Cumarina-PEG24-ácido	Cumarina-PEG36-ácido
PM esperado	996	1.172	1.348	1.877
PM observado*	1.016	1.192	1.368	1.899

*Diferencia en valores observados debido a la presencia de sal de sodio.

Se convierten los cumarina-PEG-ácidos en los ésteres de NHS correspondientes mediante reacción con 1,5 eq de carbonato de disuccinimidilo (DSC) y 2 eq de trietilamina en DMF anhidra durante 2 h. El éster de NHS resultante,

que se mueve ligeramente más rápido que el ácido en placas de gel de sílice, se purifica mediante cromatografía en gel de sílice usando una mezcla de CH_2Cl_2 -MeOH (5-15%) y se usa en la siguiente etapa.

C) Cumarina-PEG_n-dG4P:

5 Se lleva dG4P-heptil-NH₂ de la etapa A) anterior a tampón de carbonato-bicarbonato 0,1 M (pH 8,6) y a esta disolución con agitación se le añade uno de los compuestos de cumarina-PEG-NHS (en DMF). Se agita la mezcla resultante durante la noche a temperatura ambiente y después se purifica sobre un cartucho de gel de sílice (el 15-25% de MeOH en CH_2Cl_2 para retirar la cumarina-ácido o NHS sin reaccionar y después isopropanol/ $\text{NH}_4\text{OH}/\text{H}_2\text{O}$ 6:4:1). Esto se purifica adicionalmente dos veces mediante HPLC de fase inversa para proporcionar cumarina-PEG-dG4P puro. La estructura se confirma mediante análisis en MALDI-TOF-EM. Cumarina-PEG16-dG4P: tiempo de retención, 31,7 min; cumarina-PEG20-dG4P: tiempo de retención, 32,2 min; cumarina-PEG24-dG4P: tiempo de retención, 33,0 min; cumarina-PEG36-dG4P: tiempo de retención, 34,3 min.

Datos de MALDI-TOF-EM:

	Cumarina-PEG16-dG4P	Cumarina-PEG20-dG4P	Cumarina-PEG24-dG4P	Cumarina-PEG36-dG4P
PM esperado	1.673	1.850	2.025	2.554
PM observado	1.682	1.858	2.036	2.569

EJEMPLO 12

Caracterización de las etiquetas liberadas mediante MALDI-TOF-EM

15 Las moléculas de cumarina-PEG-NH₂ esperadas se confirman mediante análisis de MALDI-TOF-EM, tras la purificación mediante HPLC (figura 16). Los resultados de MALDI-TOF-EM indican que las etiquetas de cumarina-PEG-NH₂ generadas mediante hidrólisis ácida son idénticas a las etiquetas liberadas producidas durante la reacción de polimerasa tras el tratamiento con fosfatasa alcalina.

20 Con referencia a la figura 16, las etiquetas de cumarina-PEG-NH₂ generadas mediante hidrólisis ácida de cumarina-PEG16-dG4P que produce cumarina-PEG16-NH₂, cumarina-PEG20-dG4P que produce cumarina-PEG20-NH₂, cumarina-PEG24-dG4P que produce cumarina-PEG24-NH₂ y cumarina-PEG36-dG4P que produce cumarina-PEG36-NH₂, son idénticas a las etiquetas liberadas correspondientes generadas en reacciones de extensión de polimerasa tras el tratamiento con fosfatasa alcalina, tal como se muestra mediante análisis de MALDI-TOF-EM. Se muestra una imagen compuesta de cuatro espectros de EM obtenidos por separado. Las estructuras de las etiquetas de cumarina-PEG-NH₂ se muestran a la derecha.

EJEMPLO 13

Distinción de etiquetas liberadas en nanoporos de proteína a moléculas individuales

30 Con referencia a la figura 17, se combinaron cuatro compuestos de cumarina-PEG_n-NH₂ (n = 16, 20, 24 y 36), derivados a partir de los cuatro nucleótidos comparables mediante hidrólisis ácida, y se diluyeron en KCl 4 M, Tris 10 mM, pH 7,2 para la medición en nanoporos. Los datos de serie temporal a la izquierda indican que cuando estas etiquetas de PEG entran en un canal iónico de α -hemolisina individual, provocan bloqueos de corriente que son característicos de su tamaño. Un histograma a la derecha muestra el bloqueo de corriente medio provocado por moléculas individuales que muestra una resolución de línea base con un ancho de banda de medición de 10 kHz. Las barras coloreadas en la parte superior representan la distribución de 6 σ de los datos (suponiendo distribuciones gaussianas para cada una de las cuatro etiquetas de PEG que pueden representar cada uno de los cuatro nucleótidos de ADN), lo cual sugiere que puede distinguirse una base individual con una precisión mejor que 1 de cada 300.000 acontecimientos, representada en esta figura usando las designaciones A, C, G y T, lo cual puede producirse cuando se usan cuatro nucleótidos diferentes con cuatro PEG de longitud diferente para la secuenciación de ADN.

40 Para demostrar el enfoque de SBS de moléculas individuales electrónico, se someten a prueba cuatro etiquetas de cumarina-PEG_n-NH₂ liberadas para determinar sus efectos de bloqueo de corriente sobre un nanoporo de α HL (figura 17). La distribución de frecuencia relativa del histograma de acontecimientos de bloqueo ($\langle i \rangle / \langle i_{\text{abierto}} \rangle$), muestra cuatro picos diferenciados y bien separados para las cuatro etiquetas de cumarina-PEG_n-NH₂ (n = 36, 24, 20, y 16 de izquierda a derecha respectivamente en la figura 17, parte inferior derecha).

45 Para destacar la amplia separación de los picos y ofrecer una clara evidencia de que puede lograrse la detección de un nucleótido específico mediante la señal de bloqueo única proporcionada por su PEG liberado, los picos se ajustan con funciones gaussianas individuales y se muestran las distribuciones de error de 6 σ correspondientes (rectángulos coloreados en la parte superior en la figura 17, parte inferior derecha). Las moléculas de cumarina-PEG-NH₂ aplicadas por separado se caracterizan con el poro (datos no mostrados), lo cual confirmó la identidad de los picos relacionados con PEG mostrados en esta figura.

50

Tal como se describe en este caso, un canal iónico de α HL individual puede separar moléculas individuales basándose en su tamaño, y resuelve fácilmente una mezcla de PEG hasta mejor que el tamaño de una unidad de monómero individual (es decir, <44 g/mol). Esta alta resolución surge de las interacciones entre el polímero de PEG, el electrolito (cationes móviles) y cadenas laterales de aminoácido que revisten la luz del canal de α HL. Estas interacciones permiten usar el poro como sensor a escala nanométrica que es específico para el tamaño, la carga y la propiedad química de un analito.

En este caso, tal análisis se extiende a PEG con diferentes grupos químicos en cualquier extremo terminal. El registro de corriente iónica de canales individuales en la figura 17, parte superior y parte inferior izquierda, ilustra los bloqueos provocados por las cuatro moléculas de cumarina-PEG_n-NH₂ de tamaño diferente, una cada vez. Como con PEG sin modificar, cada uno de los bloqueos de corriente es unimodal (es decir, se describe bien con distribuciones gaussianas y valores medios bien definidos).

EJEMPLO 14

Detección de etiquetas

Se usa el dispositivo para detectar 4 niveles de corriente diferenciados para 4 moléculas de etiqueta diferentes. Tal como se observa en la figura 19, cada una de las etiquetas puede distinguirse de cualquiera de las otras tres (es decir, el histograma muestra cuatro picos diferenciados marcados en el gráfico con la etiqueta correspondiente).

Cada molécula de etiqueta es un homopolímero "T" de aproximadamente 30 bases de longitud, biotilado en el extremo 3' con 2 regiones en la cadena posiblemente modificadas. En cada molécula de 30 bases de longitud, las regiones modificadas son; desde el extremo en 3', las posiciones de base 11, 12 y 13 y las posiciones 17, 18 y 19. Tal como se usa en este caso, "x" es un sitio abásico (no base) y "T" es timina. Las cuatro etiquetas son:

(a) etiqueta falsa XXX-XXX que tiene una secuencia; estreptavidina-biotina-10T-xxx-3T-xxx-11T

(b) etiqueta falsa TTT-XXX que tiene una secuencia; estreptavidina-biotina-10T-TTT-3T-xxx-11T

(c) etiqueta de 30T que tiene una secuencia; estreptavidina-biotina-30T

(d) etiqueta falsa iFluorT que tiene una secuencia; estreptavidina-biotina-10T-TTT-3T-T-FluorT-T-11T, en la que la T en la posición 18 está marcada con Fluoroscene.

Los resultados son para un poro en una matriz que captura múltiples moléculas a partir de la disolución a lo largo del tiempo. Las condiciones de detección son KCl 1 M, tamponado con HEPES 20 mM, pH 7,5 a temperatura ambiente. Cada molécula se captura y se contiene en el poro mientras se aplica una tensión. Se aumenta la tensión aplicada hasta +160 mV, se captura una nueva molécula y se reduce la tensión por debajo de 0 V y la molécula marcada con etiqueta sale del poro. Después se repite el ciclo. Cuatro moléculas de etiqueta diferentes están en la mezcla de muestra a la vez.

Tal como se muestra en la figura 20, el eje horizontal de la representación gráfica es el tiempo (medido en segundos) frente a la corriente (medida en picoamperios (pA)) en el eje vertical. No se muestra la forma de onda de tensión aplicada. La forma de onda de tensión aplicada comienza por debajo de 0 V y aumenta rápidamente hasta +160 mV y se mantiene ahí durante aproximadamente 2,3 segundos. Después se reduce la tensión gradualmente hasta por debajo de 0 V. Las lecturas de corriente siguen a la tensión aplicada siendo la corriente de una molécula capturada plana mientras la tensión aplicada está a +160 mV y después se reduce gradualmente a medida que se reduce gradualmente la tensión.

Tal como se muestra en la figura 19, las bandas claras observadas durante la aplicación de 160 mV se conectan o se difuminan ligeramente en el histograma porque también se representa gráficamente la corriente durante la reducción gradual. A pesar de esto, pueden observarse bandas de captura diferenciadas y repetibles para cada molécula de etiqueta.

A partir de lo anterior debe entenderse que, aunque se han ilustrado y descrito implementaciones particulares, pueden realizarse diversas modificaciones a las mismas y se contemplan en el presente documento. Tampoco se pretende que la invención se limite mediante los ejemplos específicos proporcionados dentro de la memoria descriptiva. Aunque se ha descrito la invención con referencia a la memoria descriptiva anteriormente mencionada, no se pretende que las descripciones e ilustraciones de las realizaciones preferibles en el presente documento se interpreten en un sentido limitativo. Además, se entenderá que ninguno de los aspectos de la invención está limitado a las representaciones, configuraciones o proporciones relativas específicas expuestas en el presente documento que dependen de una variedad de condiciones y variables. Diversas modificaciones en cuanto a la forma y detalle de las realizaciones de la invención resultarán evidentes para un experto en la técnica. Por tanto, se contempla que la invención también cubrirá cualquiera de tales modificaciones, variaciones y equivalentes. Se pretende que las siguientes reivindicaciones definan el alcance de la invención y que métodos y estructuras dentro del alcance de estas reivindicaciones y sus equivalentes queden cubiertos por las mismas.

Bibliografía:

1. Reunión de AGBT, Marco Island, FL, 2012 y comunicado de prensa del 17 de febrero de Oxford Nanopore Technologies (<http://www.nanoporetech.com/news/press-releases/view/39>)
- 5 2. Akeson, M., Branton, D., Kasianowicz, J.J., Brandin, E. y Deamer, D.W. Microsecond time-scale discrimination between polycytidylic acid and polyadenylic acid segments within single RNA molecules. *Biophys. J.* 1999, 77, 3227-3233.
3. Aksimentiev, A. *et al.*, Microscopic Kinetics of DNA Translocation through Synthetic Nanopores. *Biophysical Journal* 2004 87, 2086-2097.
- 10 4. Astier, Y., Braha, O. y Bayley, H. Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. *J Am Chem Soc* 128, 1705-10 (2006).
5. Axelrod, D., Koppel, D.E., Schlessinger, J., Elson, E. y Webb, W.W. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophysical Journal* 16, 1055-1069 (1976).
6. Bailey, H. Sequencing single molecules of DNA. *Curr. Opinion Chem Biol.* 2006, 10, 628-637.
- 15 7. Benner, S. *et al.* Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nat Nanotechnol* 2, 718-24 (2007).
8. Bezrukov S.M., Vodyanoy, I., Brutyan, R.A. y Kasianowicz, J.J. Dynamics and free energy of polymers partitioning into a nanoscale pore. *Macromolecules* 29, 8517-22 (1996).
- 20 9. Bezrukov, S.M., y Kasianowicz, J.J. Neutral polymers in the nanopore of alamethicin and alpha-hemolysin. *Biologicheskie Membrany* 2001, 18, 453-457.
10. Bokhari, S.H. y Sauer, J.R., A Parallel Graph Decomposition Algorithm for DNA Sequencing with Nanopores. *Bioinformatics* 2005 21(7), 889-896.
11. Branton, D. Nanopore sequencing. *Nat. Biotechnol.* 26, 1146-1153 (2008).
12. Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nat Biotechnol* 26, 1146-53 (2008).
- 25 13. Bultmann, T., Vaz, W.L., Melo, E.C., Sisk, R.B. y Thompson, T.E. Fluid-phase connectivity and translational diffusion in a eutectic, two-component, two-phase phosphatidylcholine bilayer. *Biochemistry* 30, 5573-9 (1991).
14. Chandler, E.L., Smith, A.L., Burden, L.M., Kasianowicz y Burden, D.L. Membrane Surface Dynamics of DNA-Threaded Nanopores Revealed by Simultaneous Single-Molecule Optical and Ensemble Electrical Recording. *Langmuir* 2004, 20, 898-905.
- 30 15. Chen, P. Probing single DNA molecule transport using fabricated nanopores. *Nano Lett.* 4, 2293-2298 (2004).
16. Clarke, J., Wu, H., Jayasinghe, L., Patel, A., Reid, S. y Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Biotech.* 2009, 1-6.
17. Cockroft, S.L., Chu, J., Amorin, M. y Ghadiri, M.R. A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution. *J Am Chem Soc* 130, 818-20 (2008).
- 35 18. Deamer, D.W. y Branton, D. Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.* 2002, 35(10), 817-825.
19. Deamer, D.W. Nanopore Analysis of Nucleic Acids Bound to Exonucleases and Polymerases, *Annual Review of Biophysics*, Vol. 39: 79-90 (fecha de publicación del volumen, junio de 2010).
20. Derrington, I.M. *et al.* Nanopore DNA sequencing with MspA. *Proc Natl Acad Sci USA* 107, 16060-5 (2010).
- 40 21. Dutta, D., Pulsipher, A. y Yousaf, M.N. Selective Tethering of Ligands and Proteins to a Microfluidically Patterned Electroactive Fluid Lipid Bilayer Array. *Langmuir* 26, 9835-9841 (2010).
22. Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* 2009, 323, 133-138.
23. Fologea, D. *et al.*, Slowing DNA Translocation in a Solid State Nanopore. *Nano Letters* 2005 5(9), 1734-1737.
- 45 24. Fologea, D. *et al.*, Detecting Single Stranded DNA with a Solid State Nanopore. *Nano Letters* 2005 5(10), 1905-1909.

25. Garaj, S. *et al.* Graphene as a subnanometre trans-electrode membrane. *Nature* 467, 190-3 (2010).
26. Gershow, M. y Golovchenko, J.A. Recapturing and trapping single molecules with a solid state nanopore. *Nature Nanotechnology* 2, 775-779 (2007).
- 5 27. Guo, J.; Xu, N., Li, Z., Zhang, S.; Wu, J., Kim, D. H., Marma, M. S., Meng, Q., Cao, H., Li, X., Shi, S., Yu, L., Kalachikov, S., Russo, J. J., Turro, N. J., Ju, J.. Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc. Natl. Acad. Sci. USA* 2008, 105(27), 9145-9150.
28. Guo, J., Yu, L., Turro, N. J., y Ju, J. An integrated system for DNA sequencing by synthesis using novel nucleotide analogues. *Accounts of Chemical Research* 2010, 43(4), 551-563.
- 10 29. Hall, A.R. *et al.* Hybrid pore formation by directed insertion of alpha-haemolysin into solid-state nanopores. *Nat Nanotechnol* 5, 874-7 (2010).
30. Harris, T. D., Buzby, P. J. *et al.* Single-molecule DNA sequencing of a viral genome. *Science* 2008, 320, 106-109.
31. Healy, K. Nanopore-based single-molecule DNA analysis. *Nanomedicine (Lond)* 2, 459-81 (2007).
- 15 32. Heng, J.B. *et al.*, Stretching DNA Using the Electric Field in a Synthetic Nanopore. *Nano Letters* 2005 5(10), 1883-1888.
33. Heng, J.B. *et al.*, The Electromechanics of DNA in a synthetic Nanopore. *Biophysical Journal* 2006, 90, 1098-1106.
34. Henrickson, S.E., Misakian, M., Robertson, B. y Kasianowicz, J.J. Driven asymmetric DNA transport in a nanometer-scale pore. *Physical Review Letters* 2000, 85, 3057-3060.
- 20 35. Hou, X. *et al.* A biomimetic potassium responsive nanochannel: G-quadruplex DNA conformational switching in a synthetic nanopore. *J Am Chem Soc* 131, 7800-5 (2009).
36. Huang, T.C. *et al.* Gene expression analysis with an integrated CMOS microarray by time-resolved fluorescence detection. *Biosens Bioelectron* 26, 2660-5 (2011).
- 25 37. Hurt, N., Wang, H., Akeson, M. y Lieberman, K.R. Specific nucleotide binding and rebinding to individual DNA polymerase complexes captured on a nanopore. *J Am Chem Soc* 131, 3772-8 (2009).
38. Jenkins, A.T.A., Neumann, T. y Offenhausser, A. Surface plasmon microscopy measurements of lipid vesicle adsorption on a micropatterned self-assembled monolayer. *Langmuir* 17, 265-267 (2001).
39. Ju, J., Kim, D. H., Bi, L., Meng, Q., Bi, X., Li, Z., Li, X., Marma, M. S., Shundi, S., Wu, J., Edwards, J. R., Romu, A., y Turro, N. J. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl. Acad. Sci. USA* 2006, 103(52), 19635-19640.
- 30 40. Kasianowicz, J.J. y Bezrukov, S.M. Protonation dynamics of the alpha-toxin ion channel from spectral analysis of pH-dependent current fluctuations. *Biophys J* 69, 94-105 (1995).
41. Kasianowicz, J.J., Brandin, E., Branton, D. y Deamer, D.W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. USA* 1996, 93, 13770-13773.
- 35 42. Kasianowicz, J.J., Henrickson, S.E., Weetall, H.H. y Robertson, B. Simultaneous multianalyte detection with a nanometer-scale pore. *Anal Chem* 73, 2268-72 (2001).
43. Kasianowicz, J.J., *et al.* Physics of DNA threading through a nanometer pore and applications to simultaneous multianalyte sensing. in *Structure and Dynamics of Confined Polymers*, ed. Kasianowicz, J.J., Kellmayer, M.S.Z. & Deamer, D.W., Nato Science Series, Kluwer Academic Publishers, The Netherlands 87, 141-64 (2002).
- 40 44. Kasianowicz, J.J. Nanometer-scale pores: potential applications for DNA characterization and analyte detection. *Disease Markers* 2003, 18, 185-191.
- 45 45. Kasianowicz, J.J. Nanopore. Flossing with DNA. *Nature Materials* 2004, 3, 355-356.
46. Korlach, J. *et al.* Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proc Natl Acad Sci USA* 105, 1176-81 (2008).
47. Korlach, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Methods Enzymol* 472, 431-55 (2010).
48. Krasilnikov, O.V. Sizing channels with neutral polymers. in *Structure and Dynamics of Confined Polymers*, ed.

- Kasianowicz, J.J., Kellmayer, M.S.Z. & Deamer, D.W., Nato Science Series, Kluwer Academic Publishers, The Netherlands 87, 97-116 (2002).
49. Kumar, S, y Sood, A. Labeled Nucleoside Polyphosphates. Patente estadounidense 7.041.812 (2006).
50. Kumar, S., McDougall, M., Sood, A., Nelson, J., Fuller, C., Macklin, J. y Mitsis, P. Terminal-Phosphate-Labeled Nucleotides with New Linkers. Patente estadounidense 7.393.640 (2008).
5
51. Kumar, S., Sood, A., Wegener, J., Finn, P., Nampalli, S., Nelson, J., Sekher, A., Mitsis, P., Macklin, J. y Fuller, C. W. Terminal Phosphate Labeled Nucleotides: Synthesis, Applications and Linker effect on incorporation by DNA Polymerases. *Nucleosides, Nucleotides & Nucleic Acids* (2005) 24 (5-7), 401-408.
52. Lee S.E., Sidorov A., Goullain T., Mignet N., Thorpe S.J., Brazier J.A., Dickman M.J., Hornby D.P., Grasby, J.A. y Williams, D.M. Enhancing the catalytic repertoire of nucleic acids: a systematic study of linker length and rigidity. *Nucleic Acids Research* 2001, 29(7), 1565-1573.
10
53. Lei, N., Watson, B.O., MacLean, J.N., Yuste, R. y Shepard, K.L. A 256-by-56 CMOS Microelectrode Array for Extracellular Neural Stimulation of Acute Brain Slices. in *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International* 148-603 (2008).
54. Levine, P.M., Gong, P., Levicky, R. y Shepard, K.L. Real-time, multiplexed electrochemical DNA detection using an active complementary metal-oxide-semiconductor biosensor array with integrated sensor electronics. *Biosens Bioelectron* 24, 1995-2001 (2009).
15
55. Li, Z., Bai, X., Ruparel, H., Kim, S., Turro, N.J. y Ju, J. A photocleavable fluorescent nucleotide for DNA sequencing and analysis. *Proc. Natl. Acad. Sci. USA* 2003, 100, 414-419.
56. Love, J.C., Estroff, L.A., Kriebel, J.K., Nuzzo, R.G. y Whitesides, G.M. Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chem Rev* 105, 1103-69 (2005).
20
57. Manrao, E.A., *et al.* Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol*, Avance de publicación en línea (2012).
58. Margulies, M. *et al.* Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature* 437: 376-380 (2005).
25
59. Mathe, J. *et al.*, Nanopore Unzipping of Individual Hairpin Molecules. *Biophysical Journal* 2004 87, 3205-3212.
60. Matysiak, S., Montesi, A., Pasquali, M., Kolomeisky, A.B. y Clementi, C. Dynamics of polymer translocation through nanopores: theory meets experiment. *Phys Rev Lett* 96, 118103 (2006).
61. McNally, B. *et al.* Optical recognition of converted DNA nucleotides for single-molecule DNA sequencing using nanopore arrays. *Nano Lett* 10, 2237-44 (2010).
30
62. Meller, A., Nivon, L., Brandin, E., Golovchenko, J. y Branton, D. Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl. Acad. Sci. USA* 2000, 97, 1079-1084.
63. Meller, A. *et al.*, Single Molecule Measurements of DNA Transport Through a Nanopore. *Electrophoresis* 2002 23, 2583-2591.
64. Merchant, C.A. *et al.* DNA translocation through graphene nanopores. *Nano Lett* 10, 2915-21 (2010).
35
65. Montal, M. y Mueller, P. Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. *Proc Natl Acad Sci USA* 69, 3561-3566 (1972).
66. Nam, J.M., Nair, P.M., Neve, R.M., Gray, J.W. y Groves, J.T. A fluid membrane-based soluble ligand-display system for live-cell assays. *ChemBioChem* 7, 436-440 (2006).
67. Nkodo, A.E. *et al.* Diffusion coefficient of DNA molecules during free solution electrophoresis. *ELECTROPHORESIS* 22, 2424-2432 (2001).
40
68. Oliver, A.E. y Deamer, D.W. Alpha-Helical Hydrophobic Polypeptides Form Proton-Selective Channels in Lipid Bilayers. *Biophysical Journal* 66, 1364-1379 (1994).
69. Palegrosdemange, C., Simon, E.S., Prime, K.L. y Whitesides, G.M. Formation of Self-Assembled Monolayers by Chemisorption of Derivatives of Oligo(Ethylene Glycol) of Structure Hs(Ch₂)₁₁(Och₂ch₂)₂Meta-Oh on Gold. *Journal of the American Chemical Society* 113, 12-20 (1991).
45
70. Perkins, T.T., Quake, S.R., Smith, D.E. y Chu, S. Relaxation of a single DNA molecule observed by optical microscopy. *Science* 1994, 264, 822-826.

71. Reiner, J.E., Kasianowicz, J.J., Nablo, B.J. y Robertson, J.W. Theory for polymer analysis using nanopore-based single-molecule mass spectrometry. *Proc Natl Acad Sci USA* 107, 12080-5 (2010).
72. Rief, M., Clausen-Schaumann, H. y Gaub, H.E. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.* 1999, 6, 346-349.
- 5 73. Robertson, J. W. F., Rodrigues, C. W., Stanford, V. M., Rubinson, K. A., Krasilnikov, O. V. y Kasianowicz, J. J. Single-molecule mass spectrometry in solution using a solitary nanopore. *Proc. Natl. Acad. Sci. USA* 2007, 104, 8207-8211.
74. Rodrigues, C.G. *et al.* Mechanism of KCl enhancement in detection of nonionic polymers by nanopore sensors. *Biophys J* 95, 5186-5192 (2008).
- 10 75. Ronaghi M., Uhlen, M. y Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* 281, 363-365 (1998).
76. Rosenstein, J., V. Ray, M. Drndic, y K.L. Shepard. Solid-state nanopores integrated with low-noise preamplifiers for high-bandwidth DNA analysis. in *Life Science Systems and Applications Workshop (LiSSA)*, 2011 IEEE/NIH (2011).
- 15 77. Rosenstein, J., Ray, V., Drndic, M. y Shepard, K.L. Nanopore DNA sensors in CMOS with on-chip low-noise preamplifiers. in *Solid-State Sensors, Actuators and Microsystems Conference (TRANSDUCERS)*, 2011 16th International 874-877 (2011).
78. Rothberg, J.M. *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348-352 (2011).
- 20 79. Sauer-Budge, A.F. *et al.*, Unzipping Kinetics of Double Stranded DNA in a Nanopore. *Physical Review Letters* 2003 90(23), 238101-1 - 238101-4.
80. Schneider, G.F. *et al.* DNA translocation through graphene nanopores. *Nano Lett* 10, 3163-7 (2010).
81. Seo, T.S., Bai, X., Ruparel, H., Li, Z., Turro, N.J. y Ju, J. Photocleavable fluorescent nucleotides for DNA sequencing on a chip constructed by site-specific coupling chemistry. *Proc. Natl. Acad. Sci. USA* 2004, 101, 5488-5493.
- 25 82. Shapovalov, G. y Lester, H. A. Getting transitions in bacterial ion channels measured at 3 microseconds resolution. *J. Gen. Physiol.* 2004, 124, 151-161.
83. Shen, K., Tsai, J., Shi, P. y Kam, L.C. Self-aligned supported lipid bilayers for patterning the cell-substrate interface. *J Am Chem Soc* 131, 13204-5 (2009).
- 30 84. Shi, H. y Ratner, B.D. Template recognition of protein-imprinted polymer surfaces. *Journal of Biomedical Materials Research* 49, 1-11 (2000).
85. Shimada, K., Kato, H., Saito, T., Matsuyama, S. y Kinugasa, S. Precise measurement of the self-diffusion coefficient for poly(ethylene glycol) in aqueous solution using uniform oligomers. *Journal of Chemical Physics* 122(2005).
- 35 86. Sims, P.A., Greenleaf, W.J., Duan, H. y Xie, X.S. Fluorogenic DNA sequencing in PDMS microreactors. *Nat Methods* 8, 575-80 (2011).
87. Smith, S.B., Cui, Y. y Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* 1996, 271, 795-799.
- 40 88. Song, L., *et al.* Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* 274, 1859-66 (1996).
89. Sood, A., Kumar, S., Wegener, J., Nampalli, S., Nelson, J., Macklin, J. y Fuller, C. W. Terminal Phosphate Labeled Nucleotides with improved substrate properties for Nucleic Acid Assays. *J. Am. Chem. Soc.* 2005, 127(8), 2394-2395.
90. Srinivasan, M.P., Ratto, T.V., Stroeve, P. y Longo, M.L. Patterned supported bilayers on self-assembled monolayers: Confinement of adjacent mobile bilayers. *Langmuir* 17, 7951-7954 (2001).
- 45 91. Storm, A.J. *et al.* Fast DNA translocation through a solid-state nanopore. *Nano Lett* 5, 1193-1197 (2005).
92. Timp, W. *et al.* Nanopore Sequencing: Electrical Measurements of the Code of Life. *IEEE Trans Nanotechnol* 9, 281-294 (2010).

93. Wang, H. *et al.*, DNA heterogeneity and Phosphorylation unveiled by Single-Molecule Electrophoresis. PNAS 2004 101(37), 13472-13477.
94. Wang, Y., Zheng, D., Tan, Q., Wang, M.X. y Gu, L.Q. Nanopore-based detection of circulating microRNAs in lung cancer patients. Nat Nanotechnol 6, 668-74 (2011).
- 5 95. Wanunu, M. y Meller, A. Chemically modified solid-state nanopores. Nano Lett 7, 1580-5 (2007).
96. Wanunu, M., Sutin, J., McNally, B., Chow, A. y Meller, A. DNA translocation governed by interactions with solid-state nanopores. Biophys J 95, 4716-25 (2008).
97. Wanunu, M. *et al.* Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. Nat Nanotechnol 5, 807-14 (2010).
- 10 98. White, R.J. *et al.* Ionic conductivity of the aqueous layer separating a lipid bilayer membrane and a glass support. Langmuir 22, 10777-83 (2006).
99. White, R.J. *et al.* Single ion-channel recordings using glass nanopore membranes. J Am Chem Soc 129, 11766-75 (2007).
- 15 100. Wieheman, K. Investigation of the bicinchoninic acid protein assay: identification of the groups responsible for color formation. Analytical Biochemistry 175(1988).
101. Yang, T.L., Baryshnikova, O.K., Mao, H.B., Holden, M.A. y Cremer, P.S. Investigations of bivalent antibody binding on fluid-supported phospholipid membranes: The effect of hapten density. Journal of the American Chemical Society 125, 4779-4784 (2003).
- 20 102. Yin, P. Tethered Bilayer Membrane Sensors with Small Transmembrane Peptide Ion Channels - Recent Developments, Future Research and Potential Applications. in Advances in Planar Lipid Bilayers and Liposomes, Vol. volumen 2 (ed. Ottova-Leitmannova, A.) 49-76 (Academic Press, 2005).
103. Vercoutere, W., Winters-Hilt, S., Olsen, H., Deamer, D., Hausler, D. y Akeson, M. Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. Nat. Biotech 2001, 19, 248-252.
- 25

REIVINDICACIONES

1. Método para la secuenciación de ácidos nucleicos, comprendiendo el método:
 - a) proporcionar una molécula de ácido nucleico y un chip que comprende una pluralidad de nanoporos individualmente direccionables, en el que dicha pluralidad de nanoporos individualmente direccionables están a una densidad de al menos 500 nanoporos individualmente direccionables por mm², en el que dicha pluralidad de nanoporos individualmente direccionables comprende un nanoporo en una bicapa lipídica, teniendo dicho nanoporo una ácido nucleico polimerasa unida al nanoporo adyacente a la entrada de dicho nanoporo, estando dicha bicapa lipídica dispuesta adyacente a al menos un electrodo y separando al menos unas disoluciones de electrolito primera y segunda, en el que dicho al menos un electrodo detecta características eléctricas de la bicapa lipídica incluyendo flujo de corriente iónica a través del nanoporo individualmente direccionable, y en el que la etiqueta liberada se transloca a través de dicho nanoporo;
 - b) proporcionar un flujo de corriente iónica a través del nanoporo individualmente direccionable entre las disoluciones de electrolito primera y segunda mediante un potencial aplicado;
 - c) dirigir dicha molécula de ácido nucleico a la ácido nucleico polimerasa unida al nanoporo individualmente direccionable;
 - d) extender un cebador hibridado a la molécula de ácido nucleico usando nucleótidos marcados con etiqueta, en el que cada nucleótido marcado con etiqueta tiene una etiqueta unida al 5'-fosfato del nucleótido marcado con etiqueta, y en el que una etiqueta asociada con un nucleótido individual se libera a partir del nucleótido marcado con etiqueta con al menos un fosfato unido a la etiqueta tras la incorporación del nucleótido marcado con etiqueta en el cebador mediante la polimerasa, y en el que la etiqueta liberada se transloca a través de dicho nanoporo; y
 - e) detectar el cambio en el flujo de corriente iónica a través del nanoporo individualmente direccionable cuando la etiqueta liberada se transloca a través del nanoporo, determinando así la secuencia de ácido nucleico de la molécula de ácido nucleico.
2. Método según la reivindicación 1, en el que el nucleótido marcado con etiqueta comprende una etiqueta que no es un fluoróforo.
3. Método según la reivindicación 1, en el que el nucleótido marcado con etiqueta comprende una etiqueta que puede detectarse mediante su firma de bloqueo.
4. Método según una cualquiera de las reivindicaciones 1-3, en el que la bicapa lipídica es una bicapa lipídica de difitanoilfosfatidilcolina (DPhPC).
5. Método según una cualquiera de las reivindicaciones 1-3, en el que la bicapa lipídica tiene una capacitancia mayor de aproximadamente 5 fF/μm² según se mide a través de dicha bicapa lipídica.
6. Método según una cualquiera de las reivindicaciones 1-3, en el que la bicapa lipídica tiene una resistencia mayor de o igual a 500 MΩ según se mide a través de dicha bicapa lipídica.
7. Método según una cualquiera de las reivindicaciones 1-3, en el que la bicapa lipídica tiene una resistencia menor de o igual a aproximadamente 1 GΩ a través de dicha bicapa lipídica.
8. Método según una cualquiera de las reivindicaciones 1-7, en el que la bicapa lipídica se forma sobre una superficie compatible con bicapa lipídica.
9. Método según la reivindicación 8, en el que el al menos un electrodo es o forma parte de la superficie compatible con bicapa lipídica.
10. Método según una cualquiera de las reivindicaciones 1-9, en el que se proporciona un par de electrodos en la etapa a).
11. Método según la reivindicación 10, en el que el par de electrodos están enfrentados y dispuestos adyacentes a la bicapa lipídica.
12. Método según la reivindicación 10, en el que el par de electrodos están acoplados a una fuente de tensión variable para proporcionar el potencial aplicado en la etapa d).
13. Método según la reivindicación 11 ó 12, en el que el par de electrodos detectan las características eléctricas de la bicapa lipídica.
14. Método según la reivindicación 13, en el que las características eléctricas de la bicapa lipídica incluyen resistencia, capacitancia y flujo de corriente iónica.

15. Método según una cualquiera de las reivindicaciones 1-14, en el que la pluralidad de nanoporos individualmente direccionables están a una densidad de al menos 1000 nanoporos individualmente direccionables por mm^2 .

Figura 1

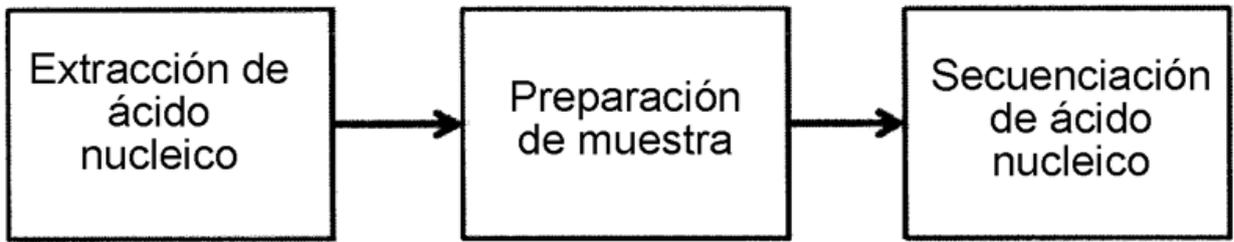


Figura 2A

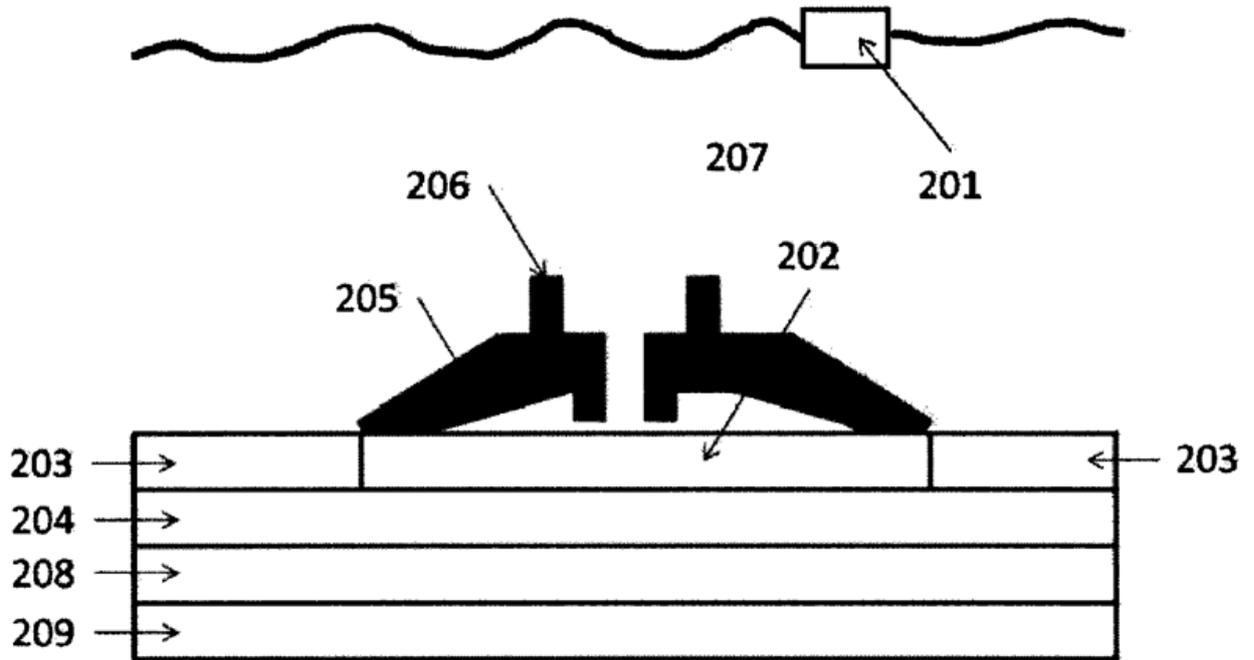


Figura 2B

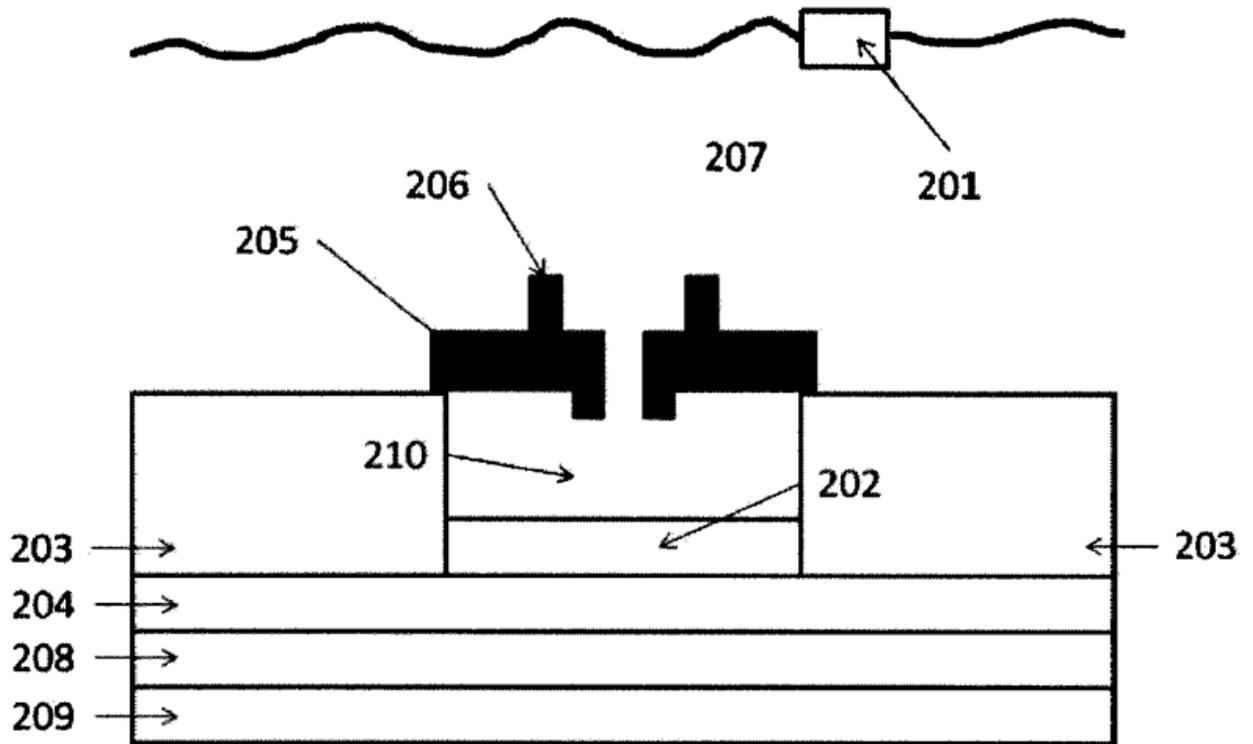


Figura 2C

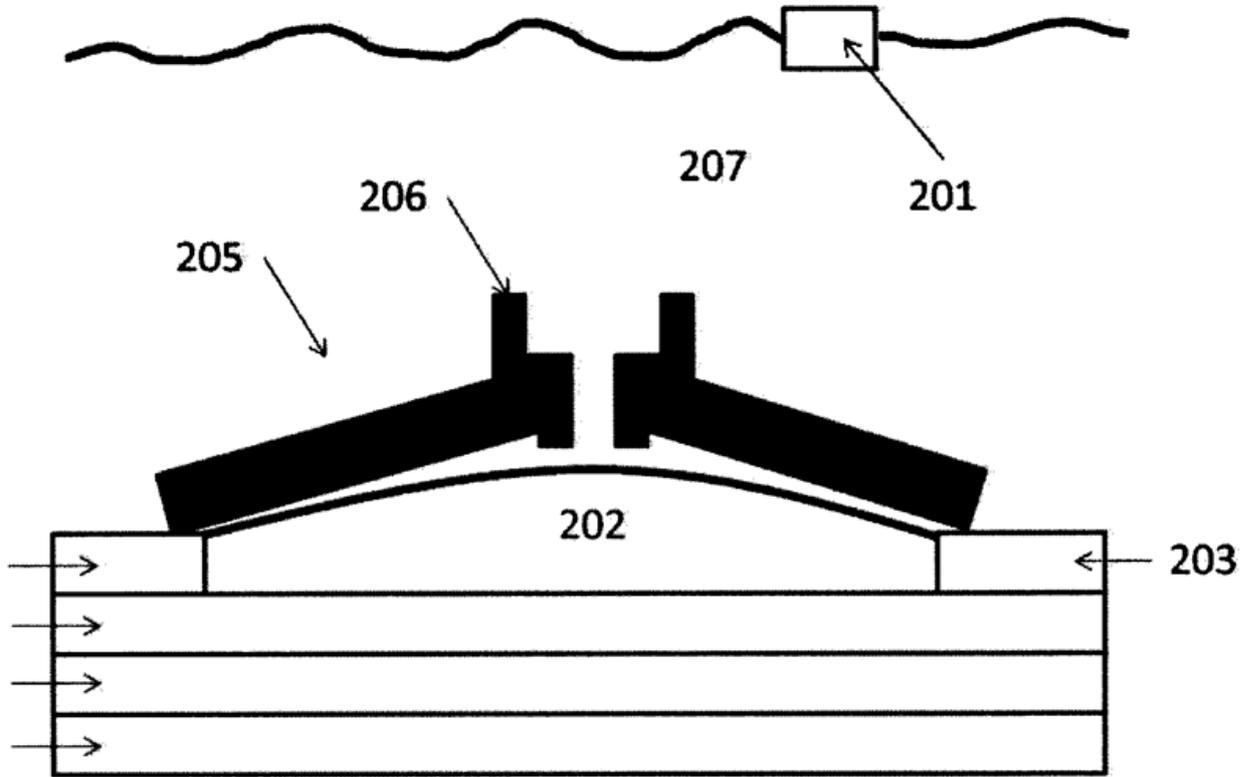


Figura 3

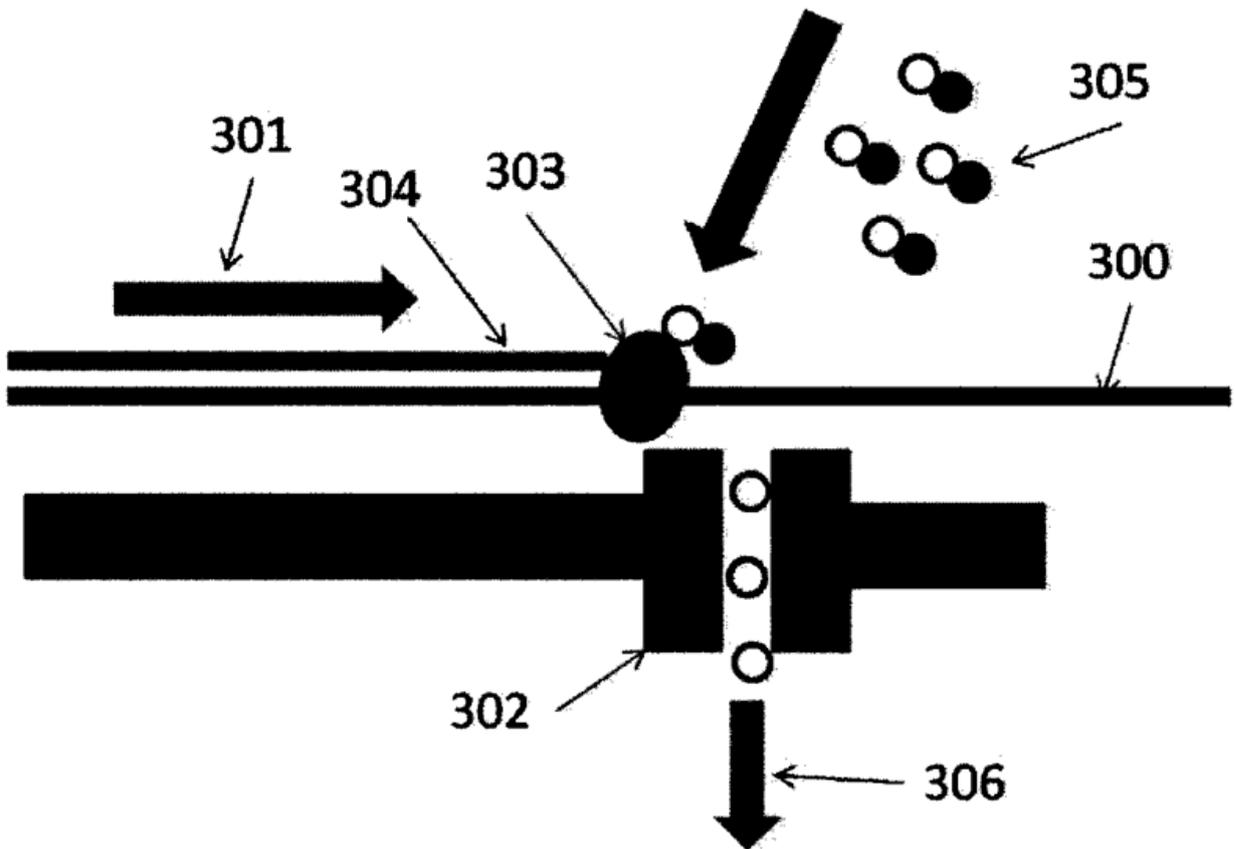


Figura 4

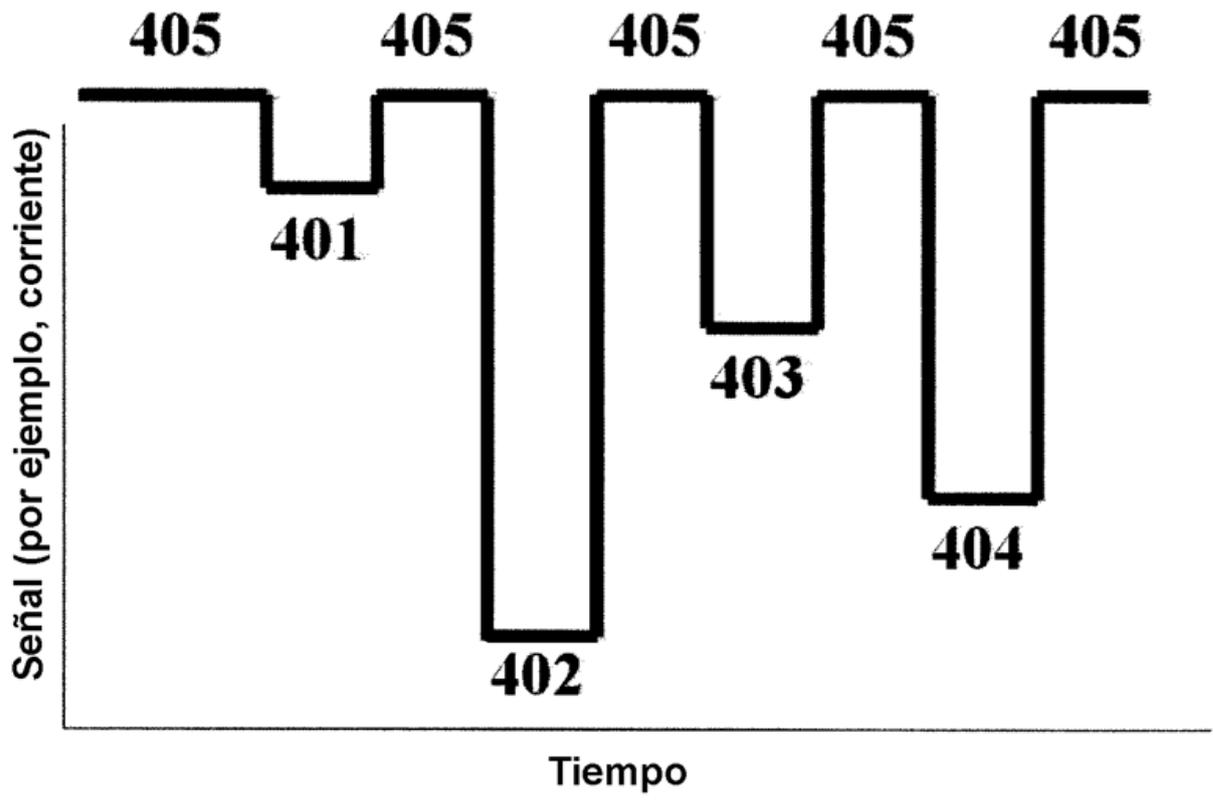


Figura 5

501

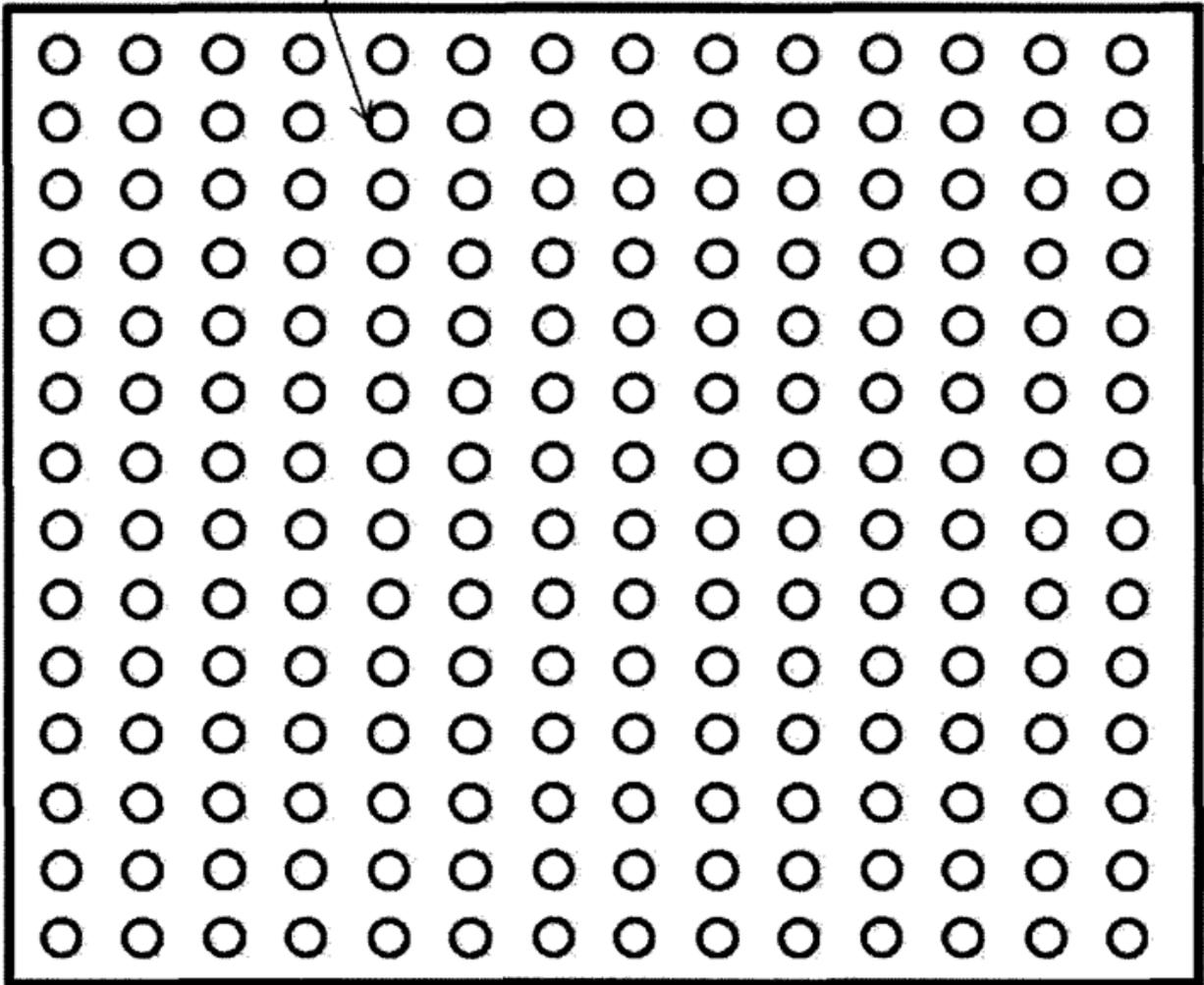


Figura 6

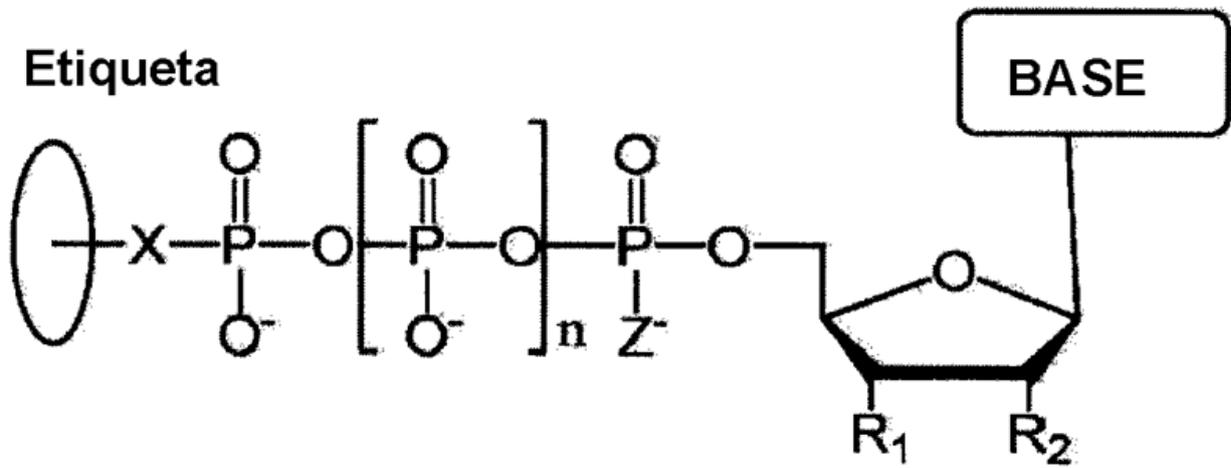


Figura 7

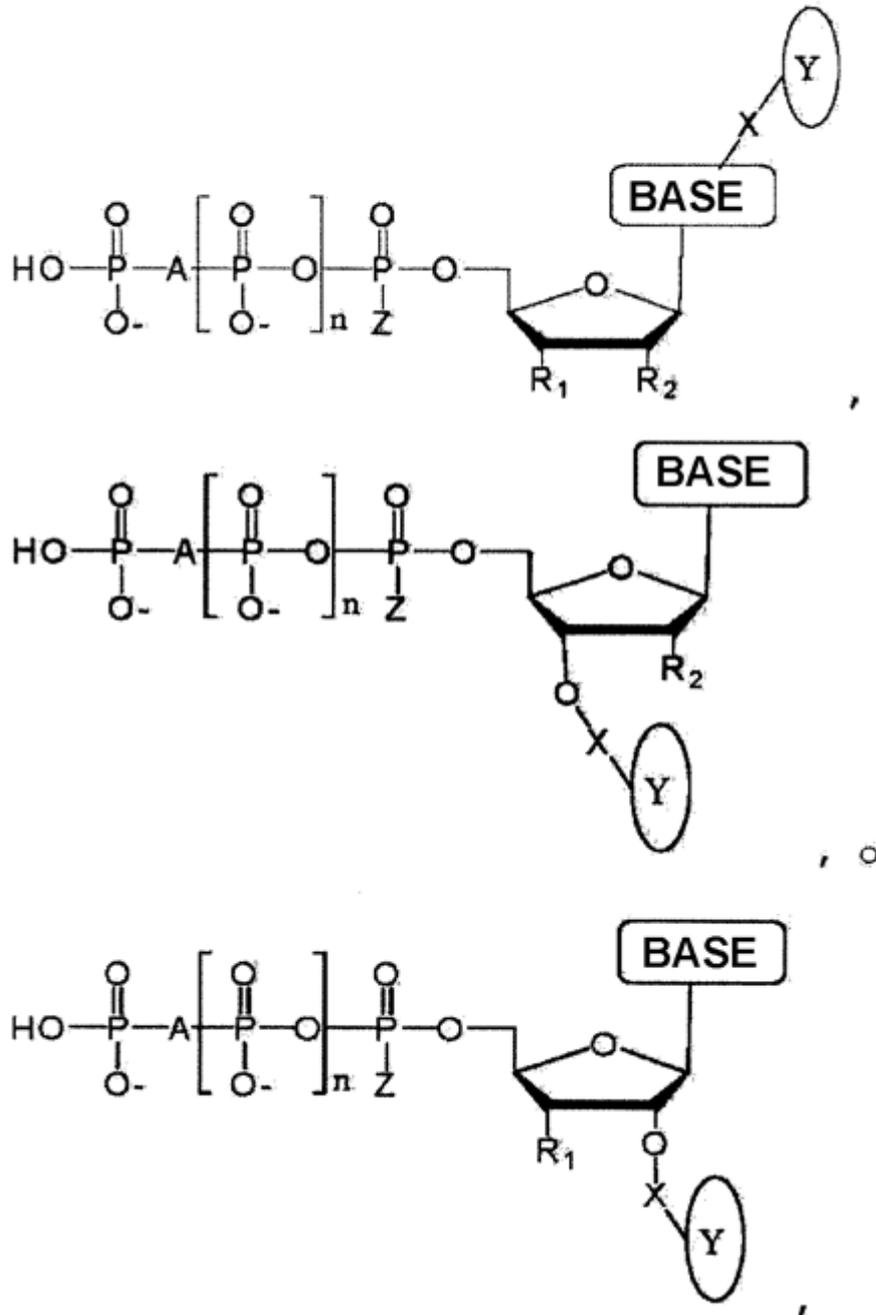


Figura 8

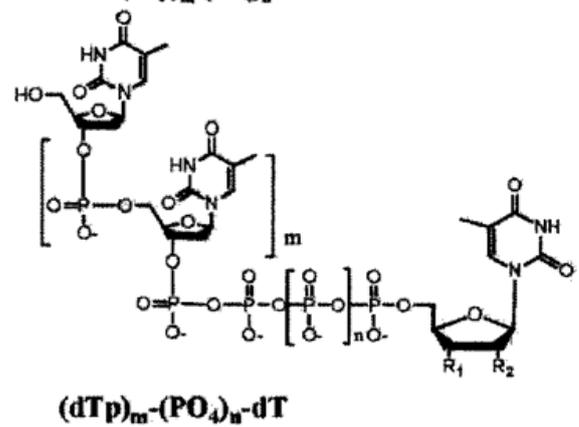
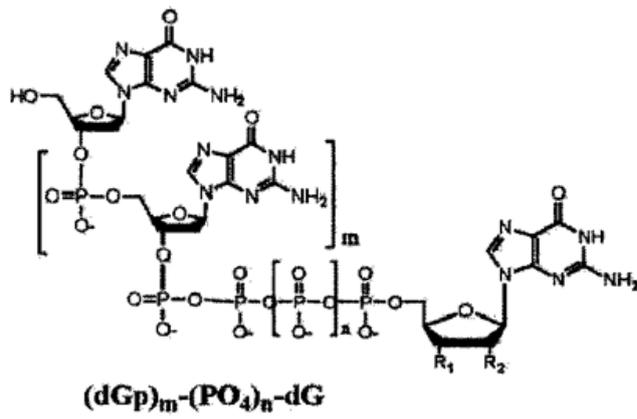
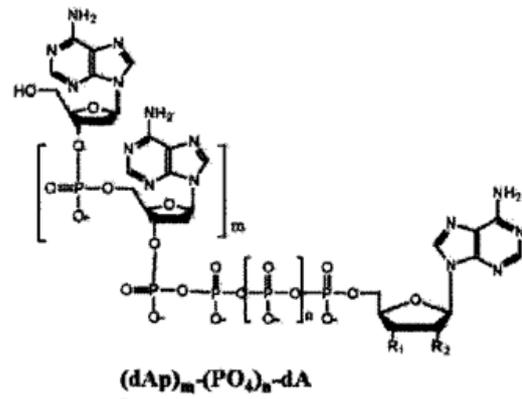
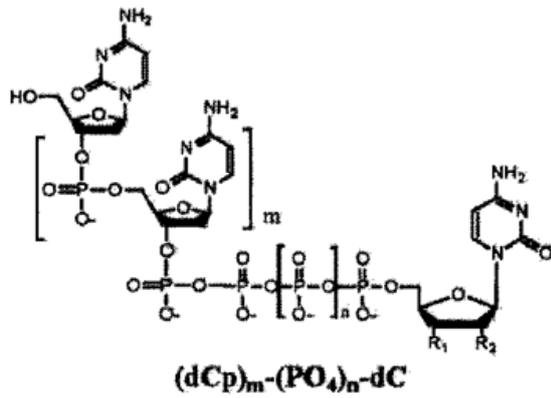


Figura 9

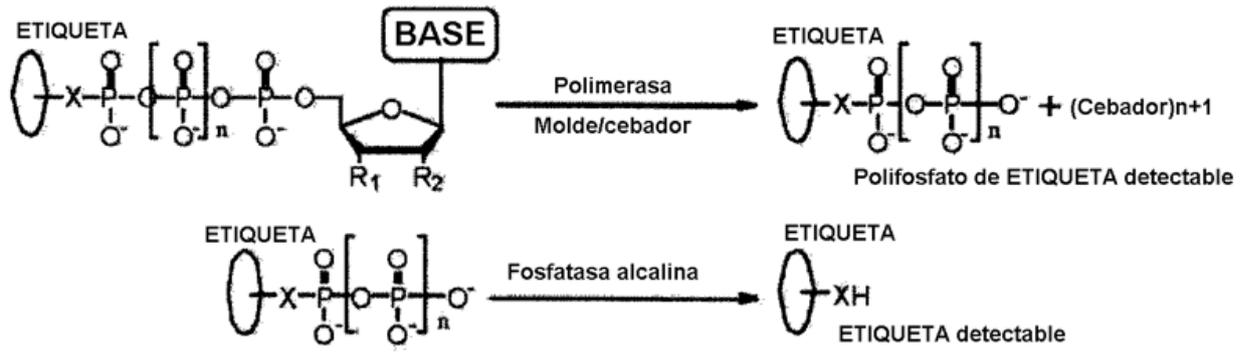


Figura 10

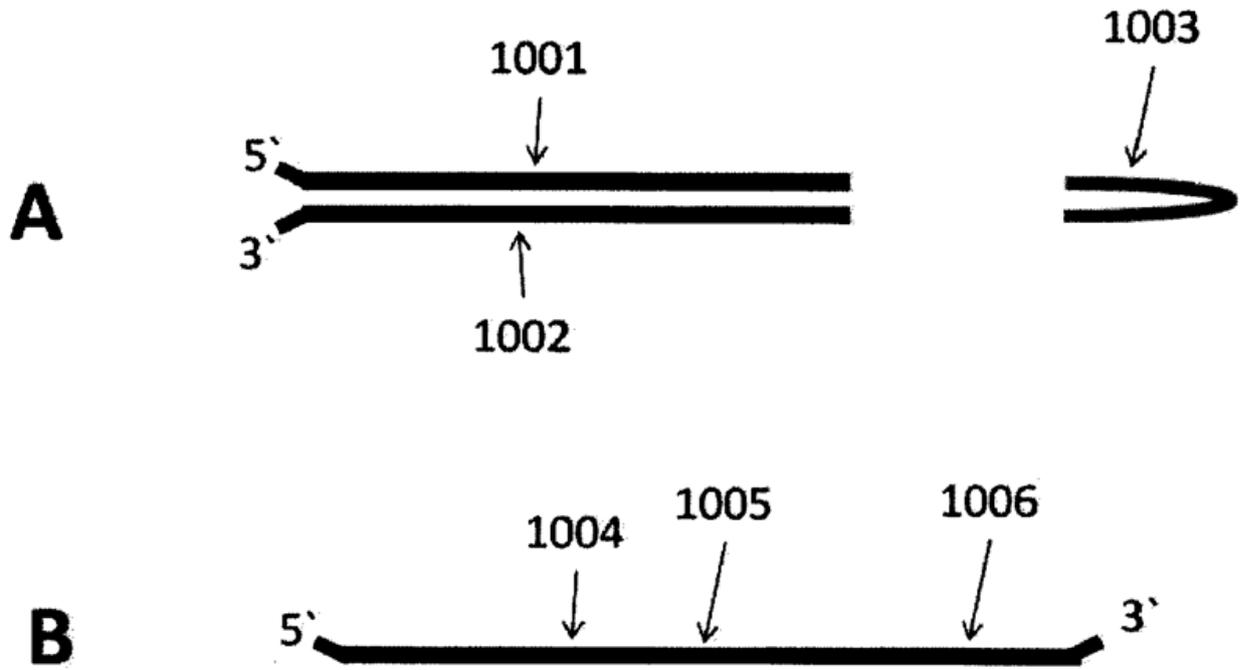


Figura 11

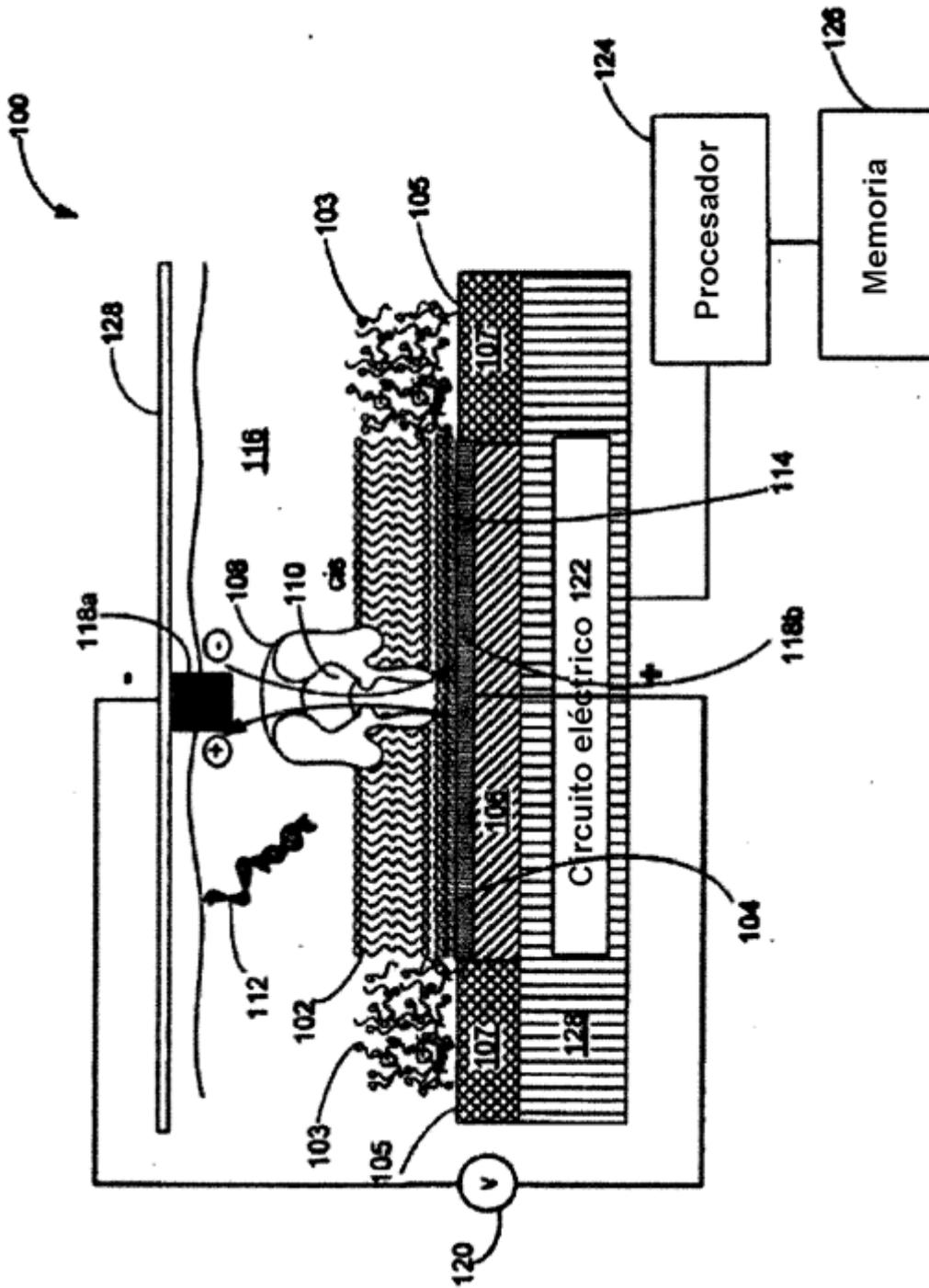


Figura 13

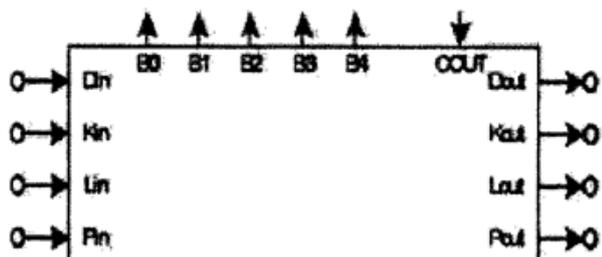
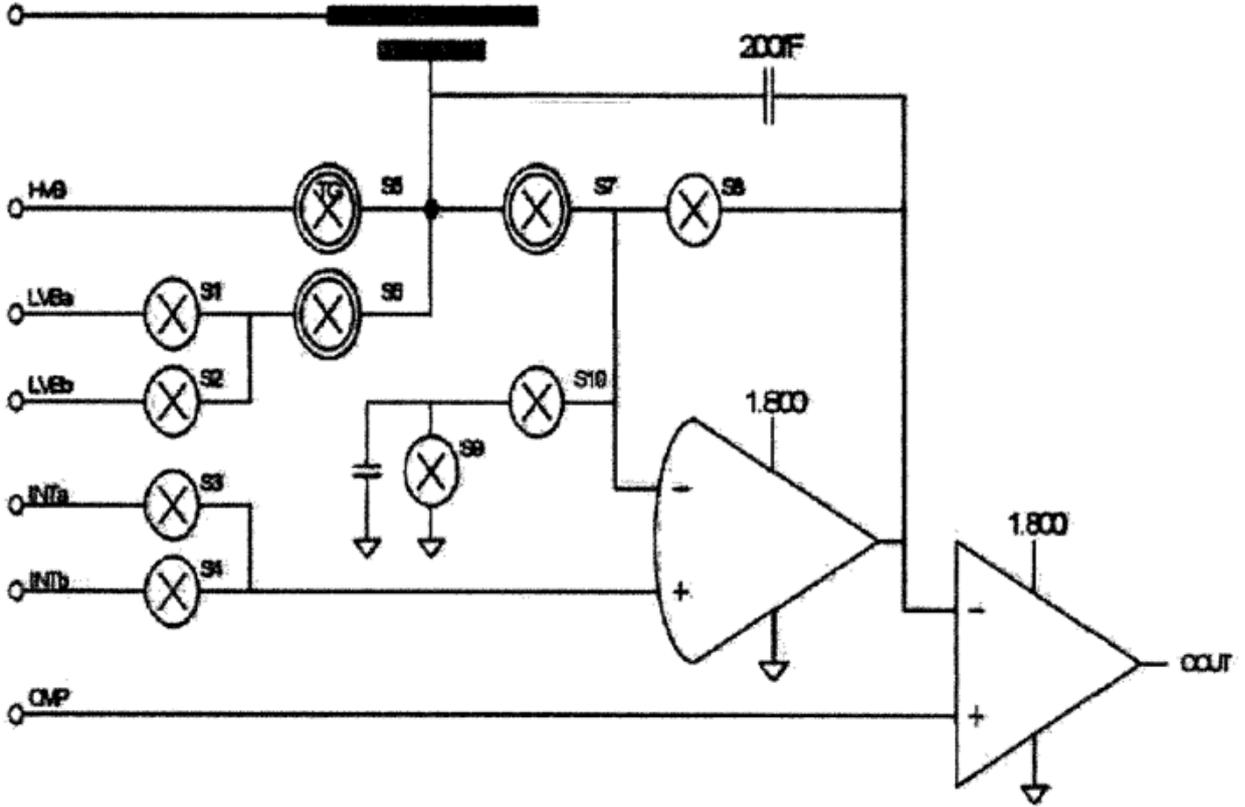


Figura 14A

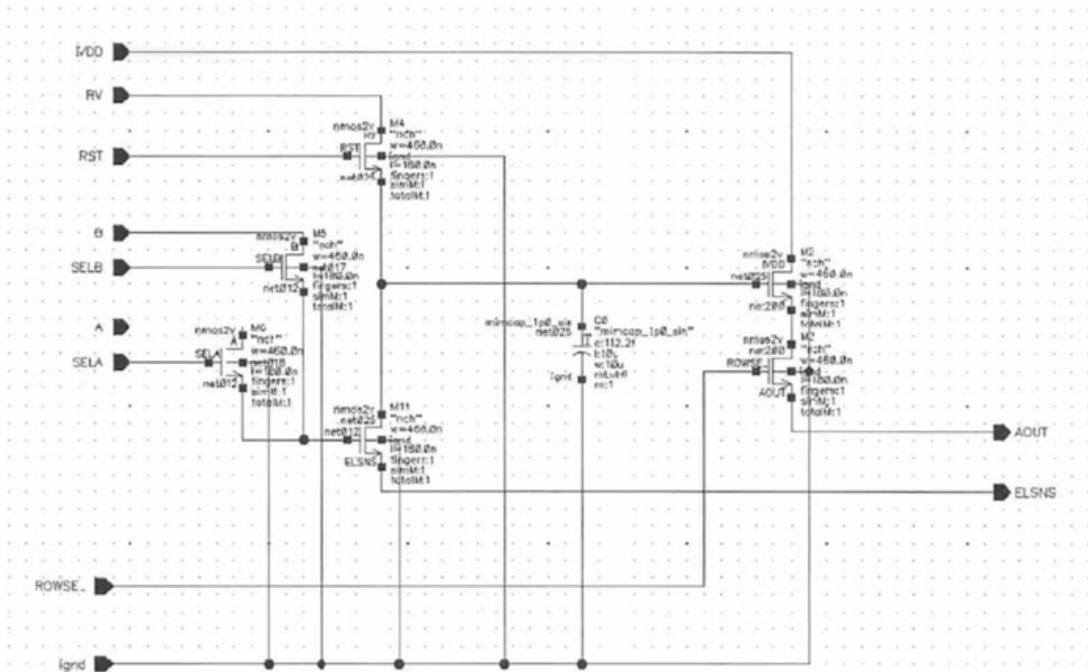


Figura 14B

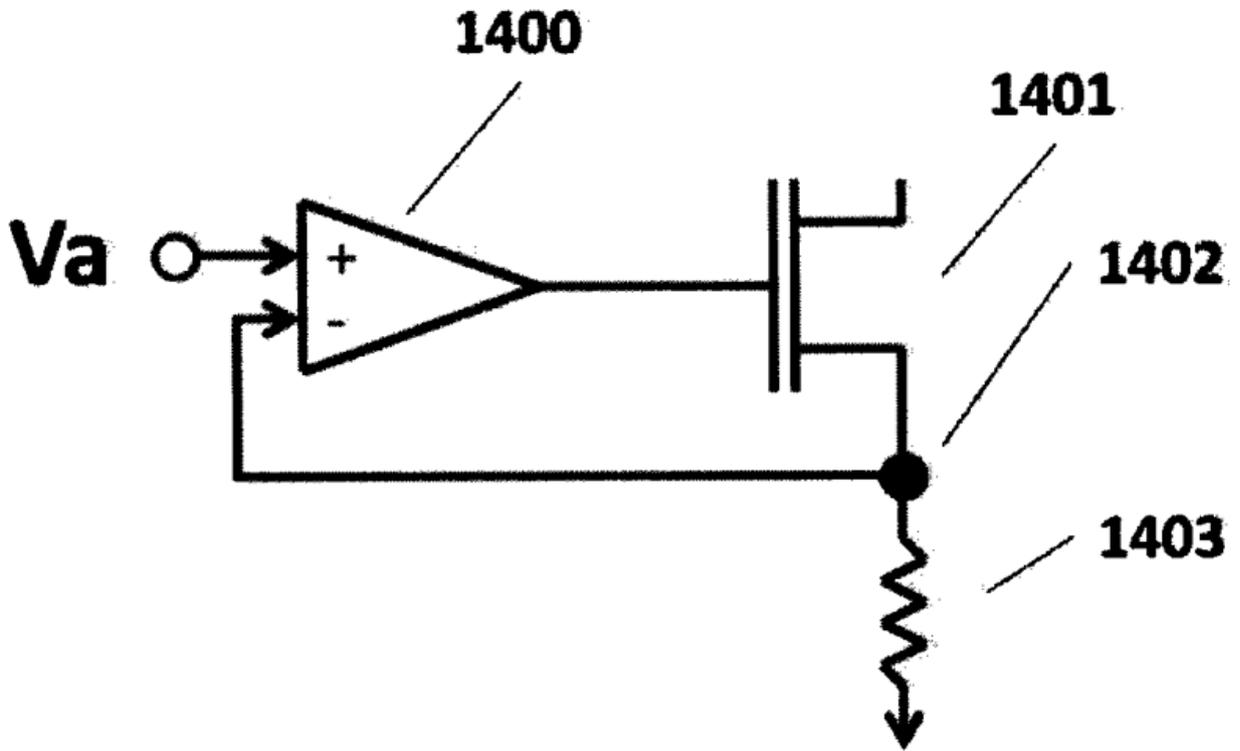
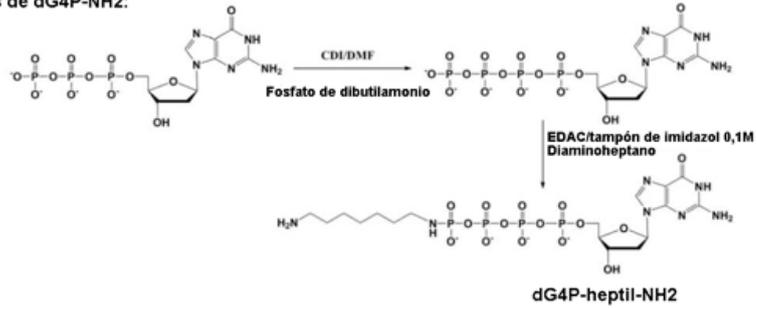
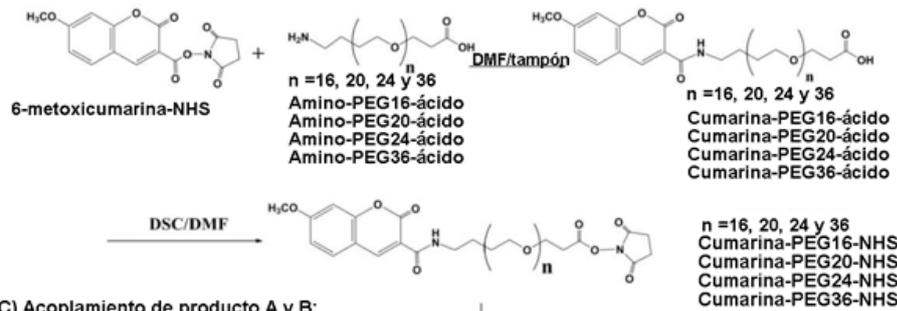


Figura 15

A) Síntesis de dG4P-NH₂:



B) Síntesis de cumarina-PEG-NHS:



C) Acoplamiento de producto A y B:

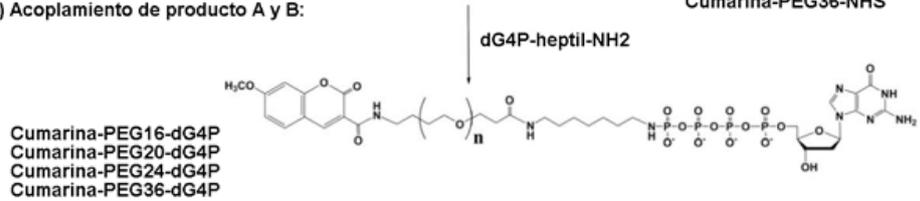


Figura 17

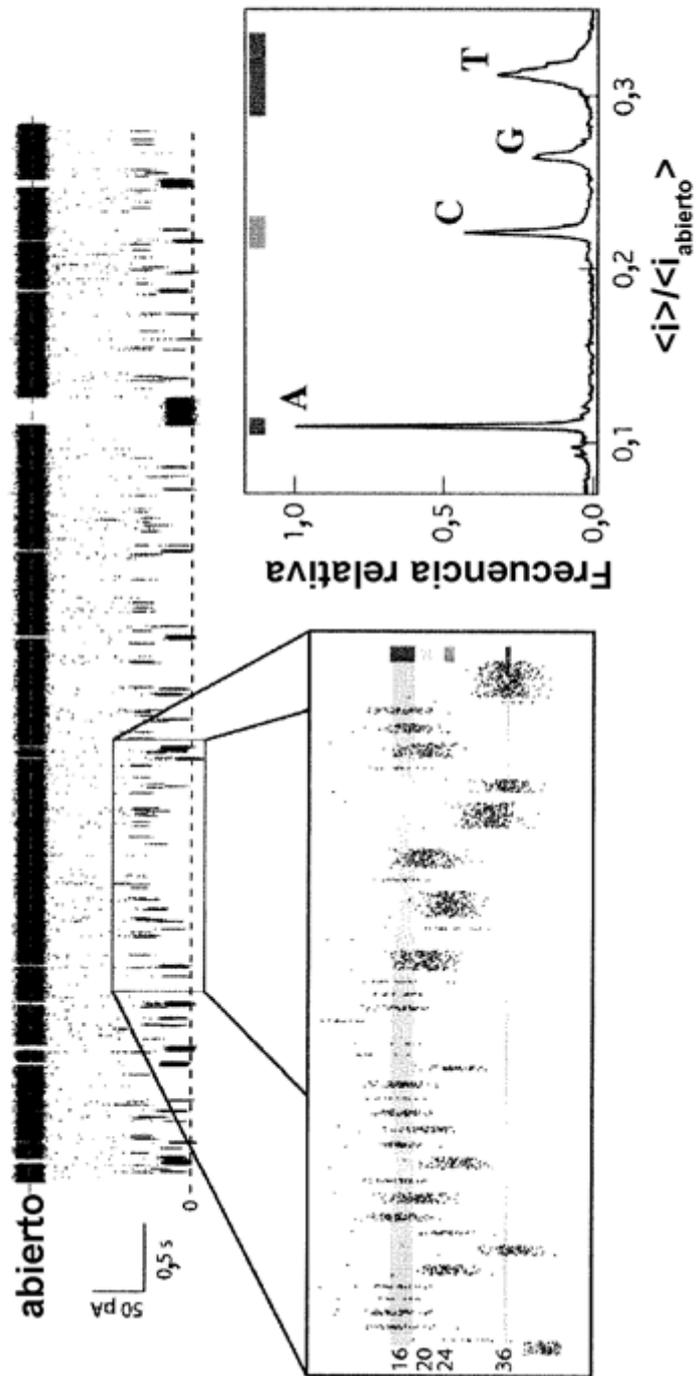


Figura 18

1800

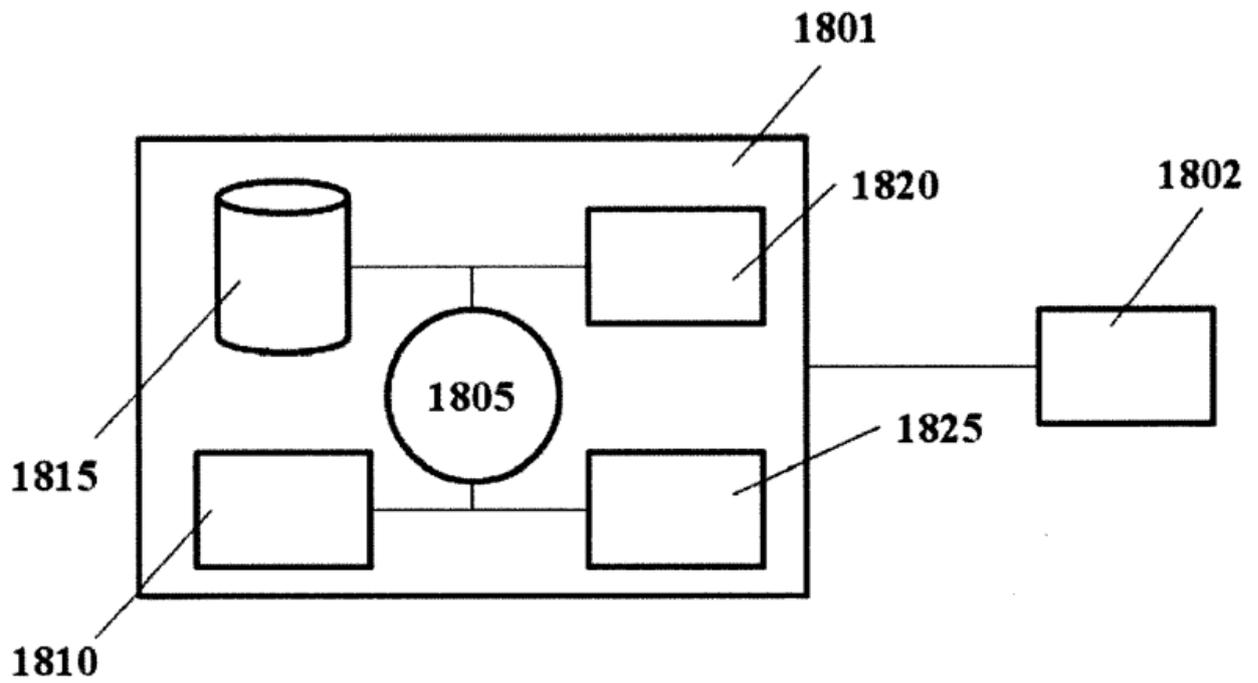


Figura 19

Lecturas de corriente de célula

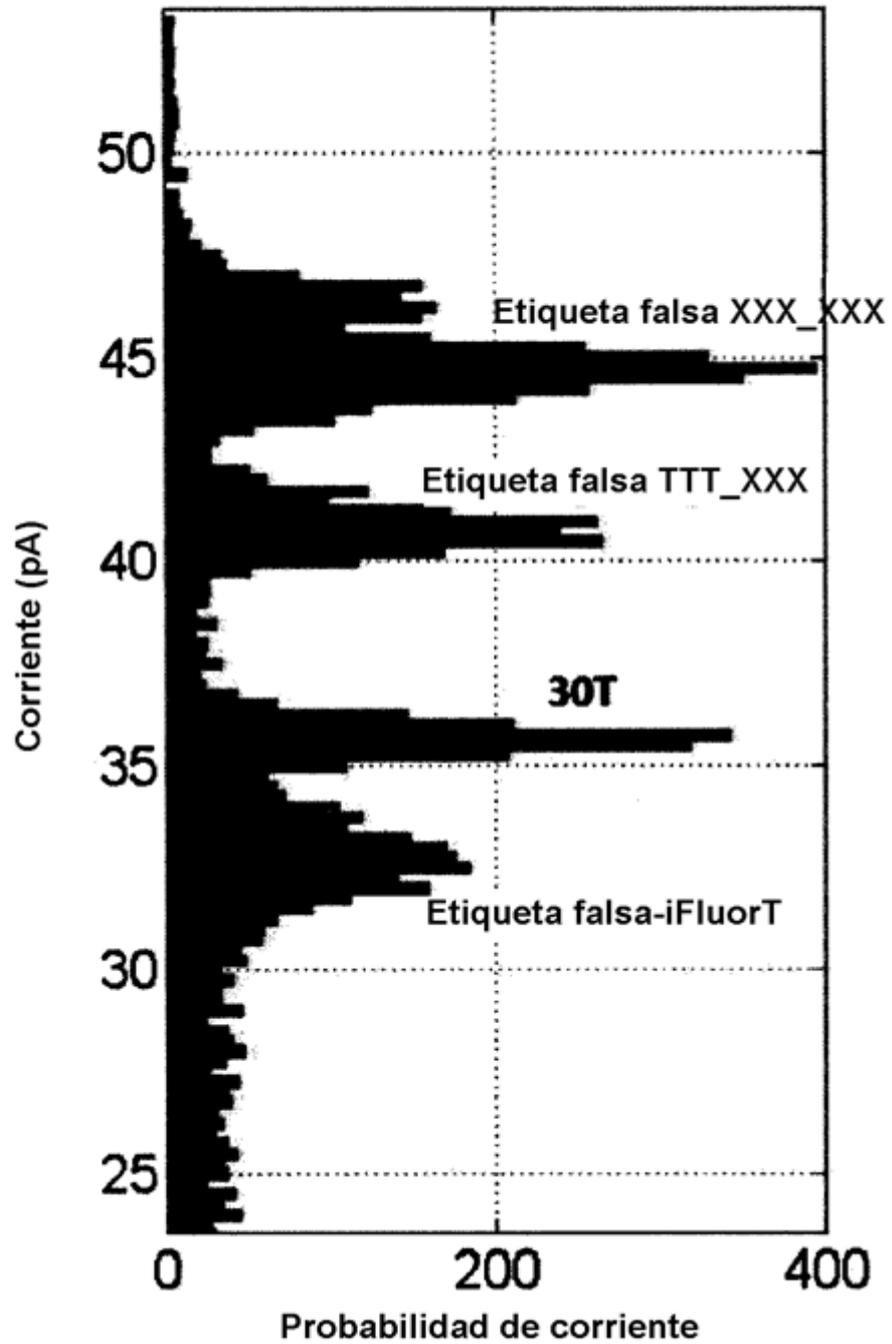


Figura 20
Corriente para B0C41

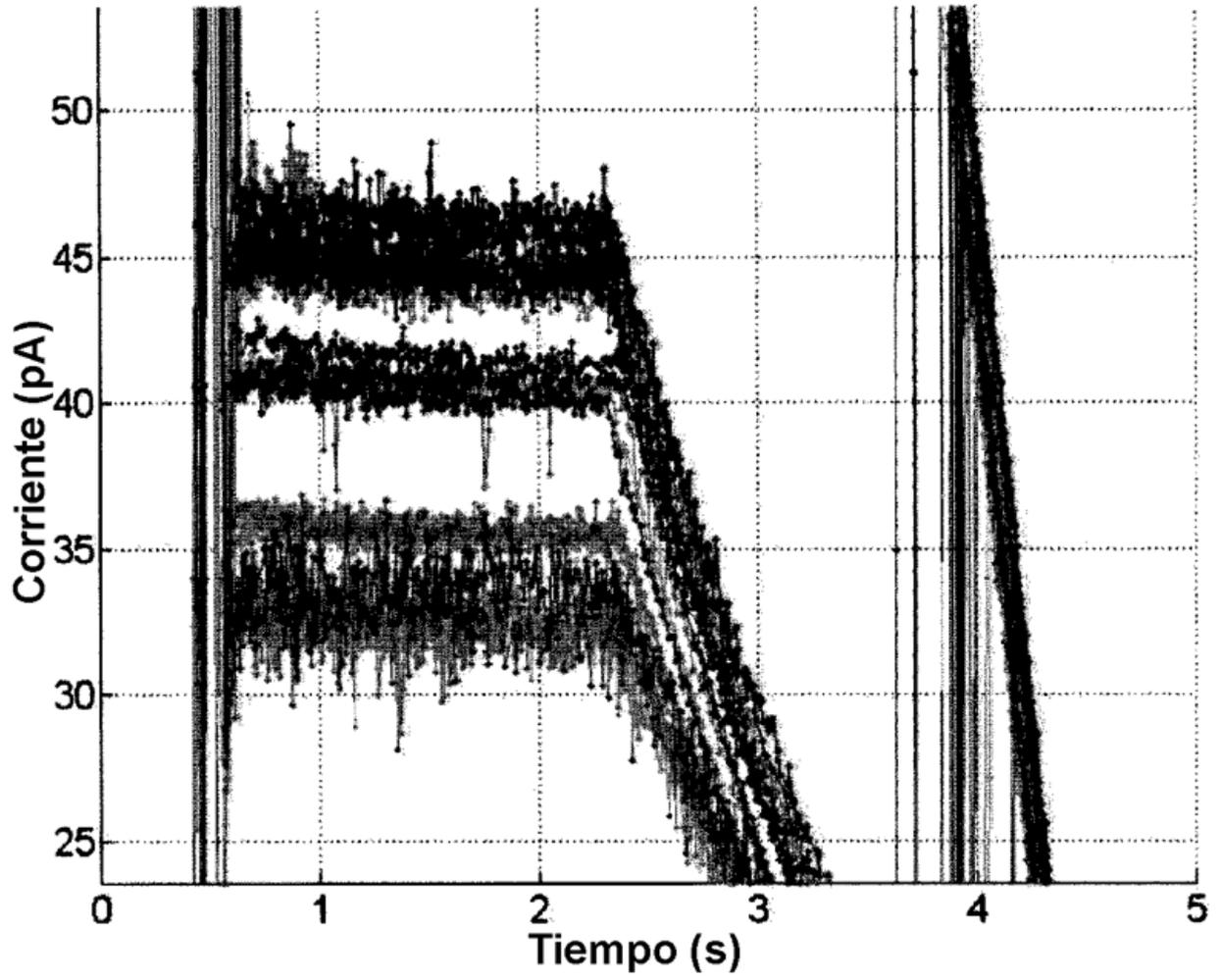


Figura 21

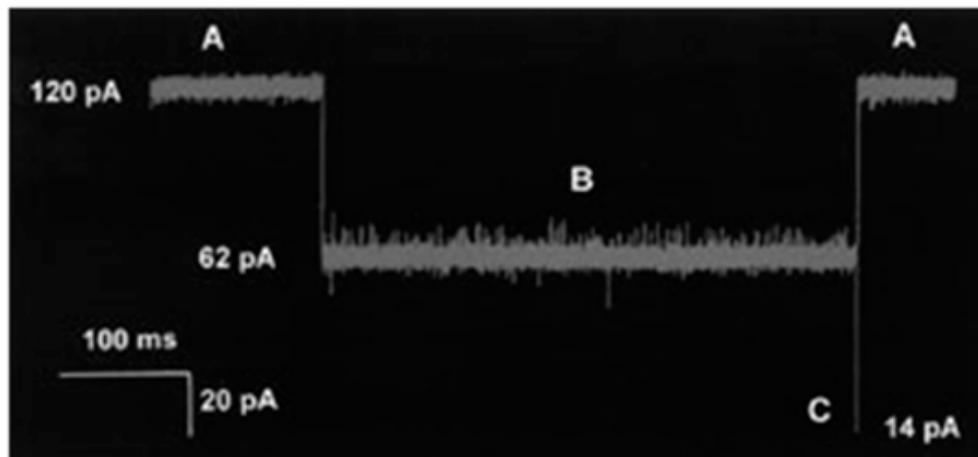
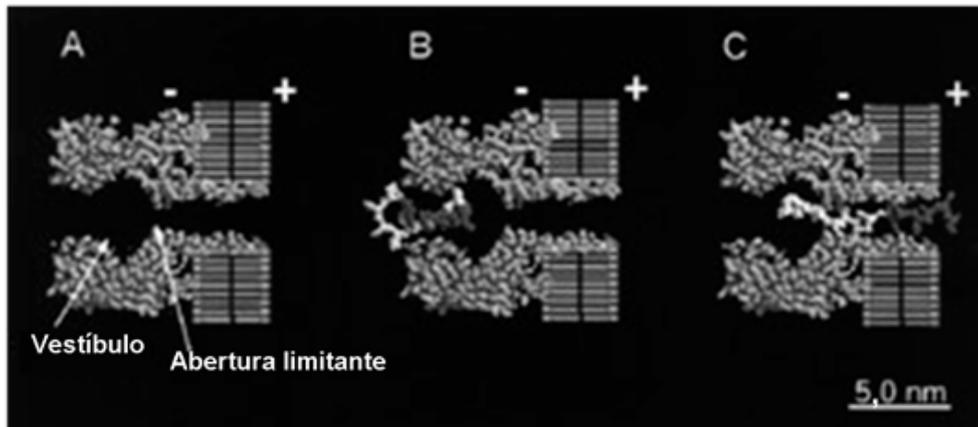


Figura 22

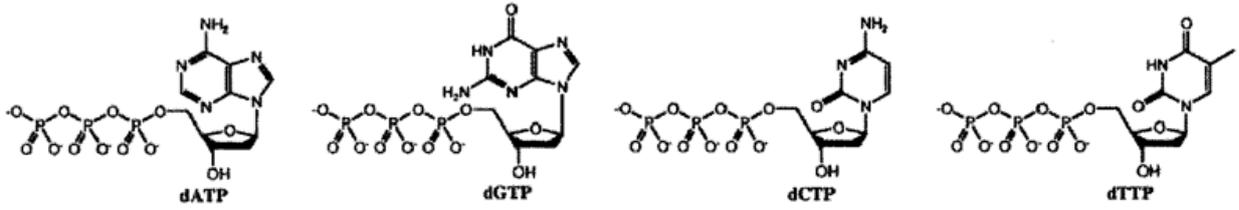
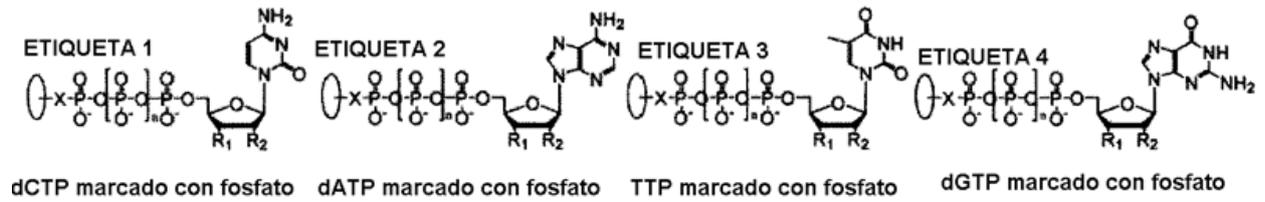


Figura 23



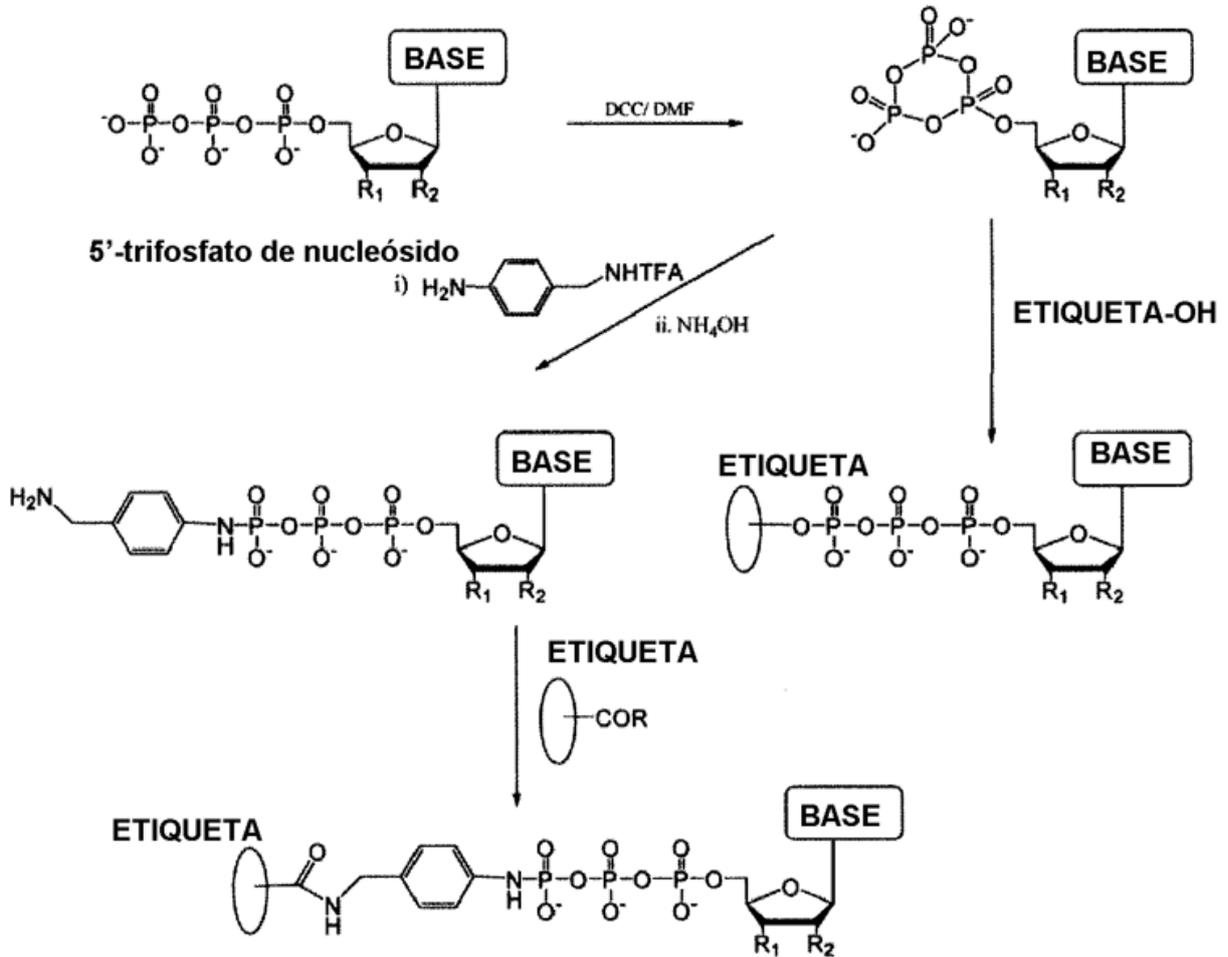
R_1 y $R_2 = H$ u OH

$n = 1-4$

$X = O, NH, S$ o CH_2

ETIQUETA = resto detectable diferente para cada nucleótido

Figura 24

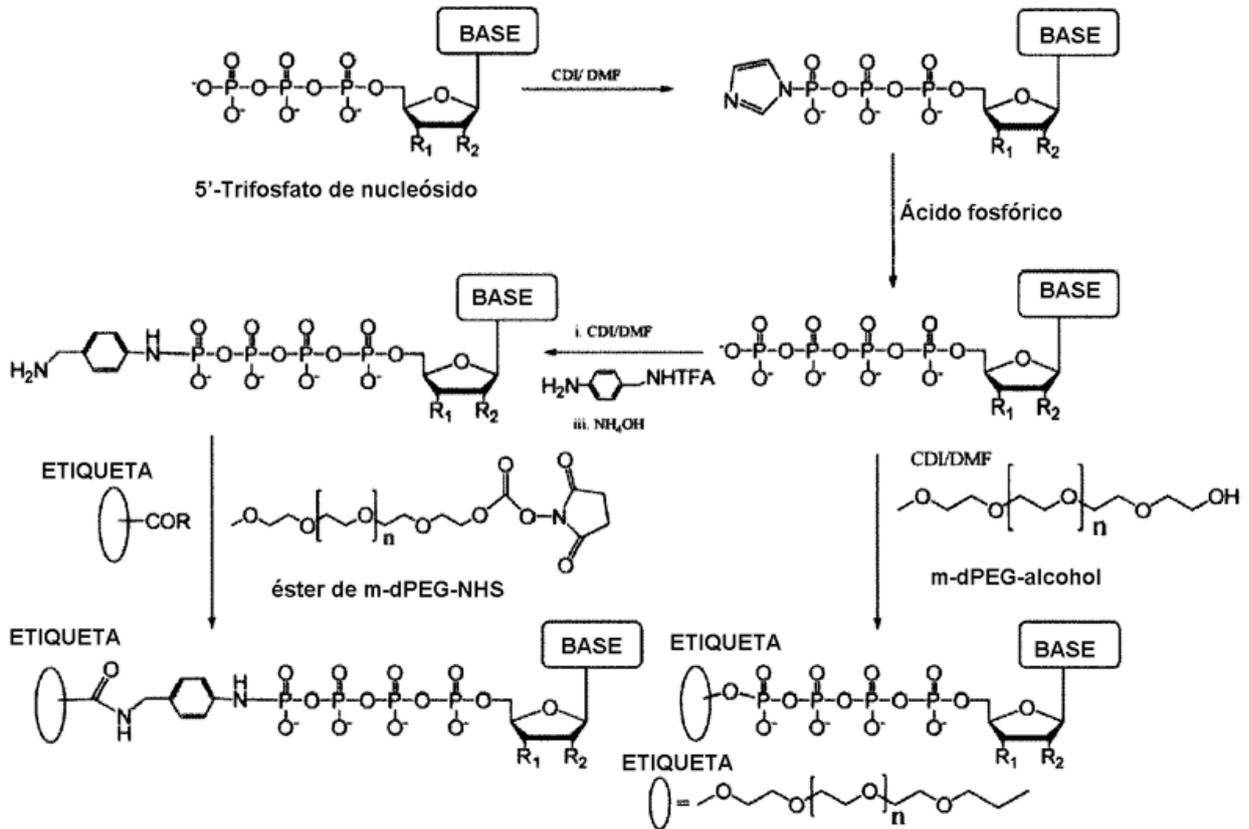


BASE = A, C, T, U, G, 7-desaza-A, 7-desaza-G o derivados de las mismas
 R_1 y R_2 = H u OH

ETIQUETA-OH = resto detectable en forma hidroxilo tal como alcoholes alifáticos o aromáticos

ETIQUETA-COR = resto detectable para reaccionar con el grupo amino, tal como ésteres de NHS, cloruro de ácido, etc.

Figura 25



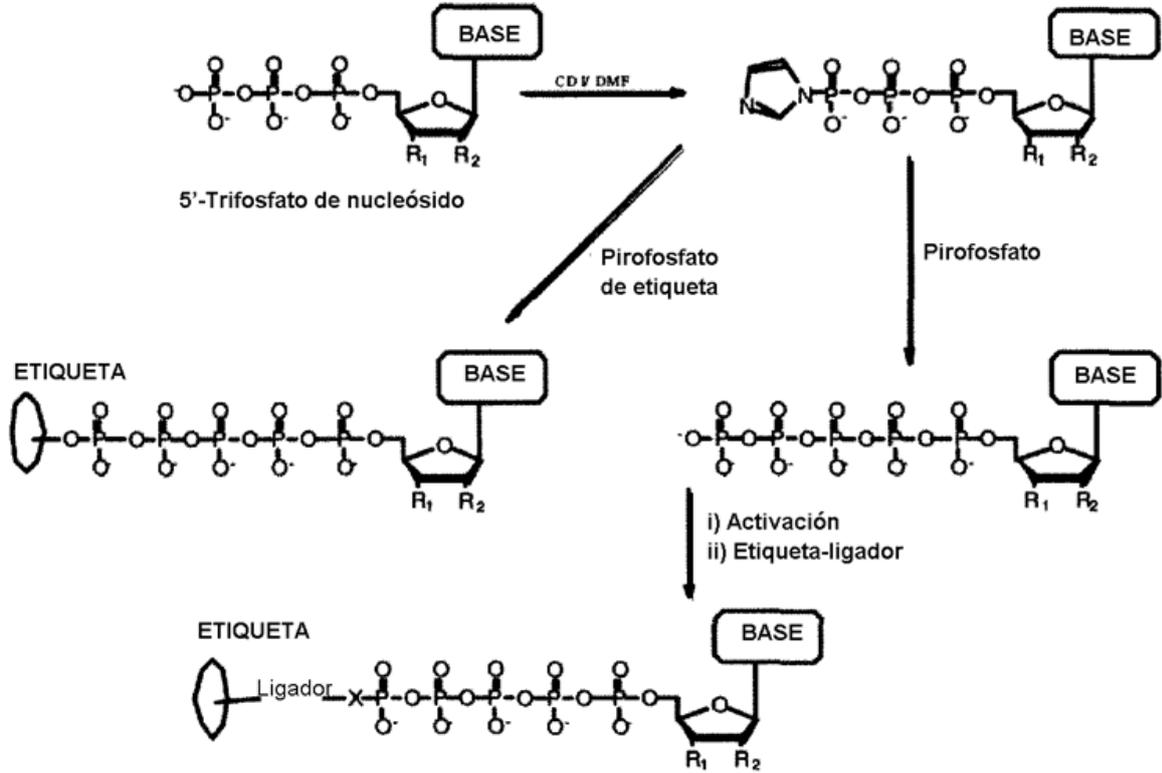
BASE = A, C, T, U, G, 7-desaza-A, 7-desaza-G o derivados de las mismas

R₁ y R₂ = H u OH

ETIQUETA = resto detectable en forma hidroxilo tal como alcoholes alifáticos o aromáticos, aminoácidos, colorantes, hidratos de carbono, oligonucleótidos, etc.

ETIQUETA-COR = resto detectable para reaccionar con el grupo amino, tal como ésteres de NHS, cloruro de ácido, etc.

Figura 26



BASE = A, C, T, U, G, 7-desaza-A, 7-desaza-G o derivados de las mismas
 R₁ y R₂ = H u OH

ETIQUETA = resto detectable en forma hidroxilo tal como alcoholes alifáticos o aromáticos, aminoácidos, colorantes, hidratos de carbono, oligonucleótidos, etc.

ETIQUETA-COR = resto detectable para reaccionar con el grupo amino, tal como ésteres de NHS, cloruro de ácido, etc.

Figura 27

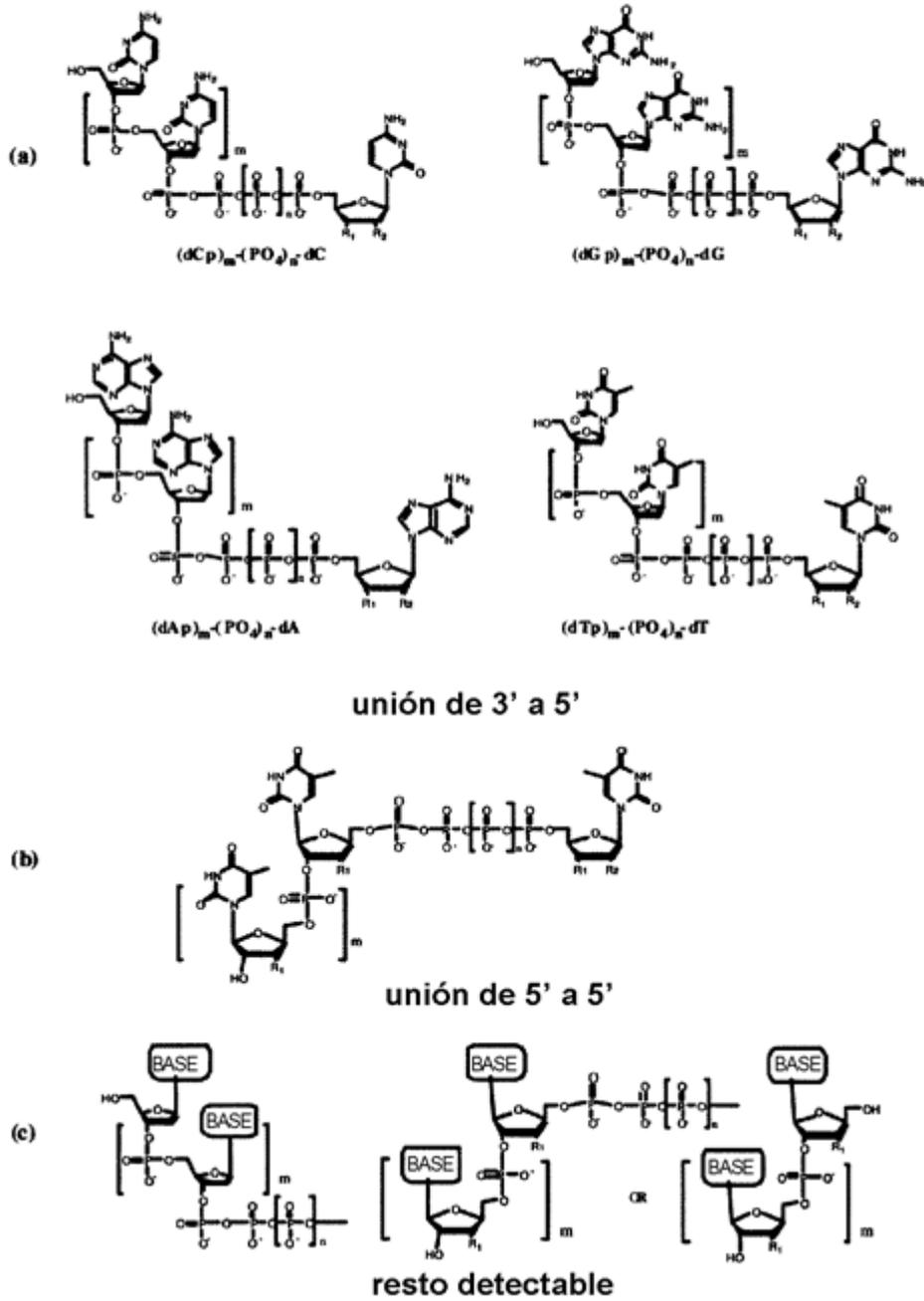


Figura 28(A)

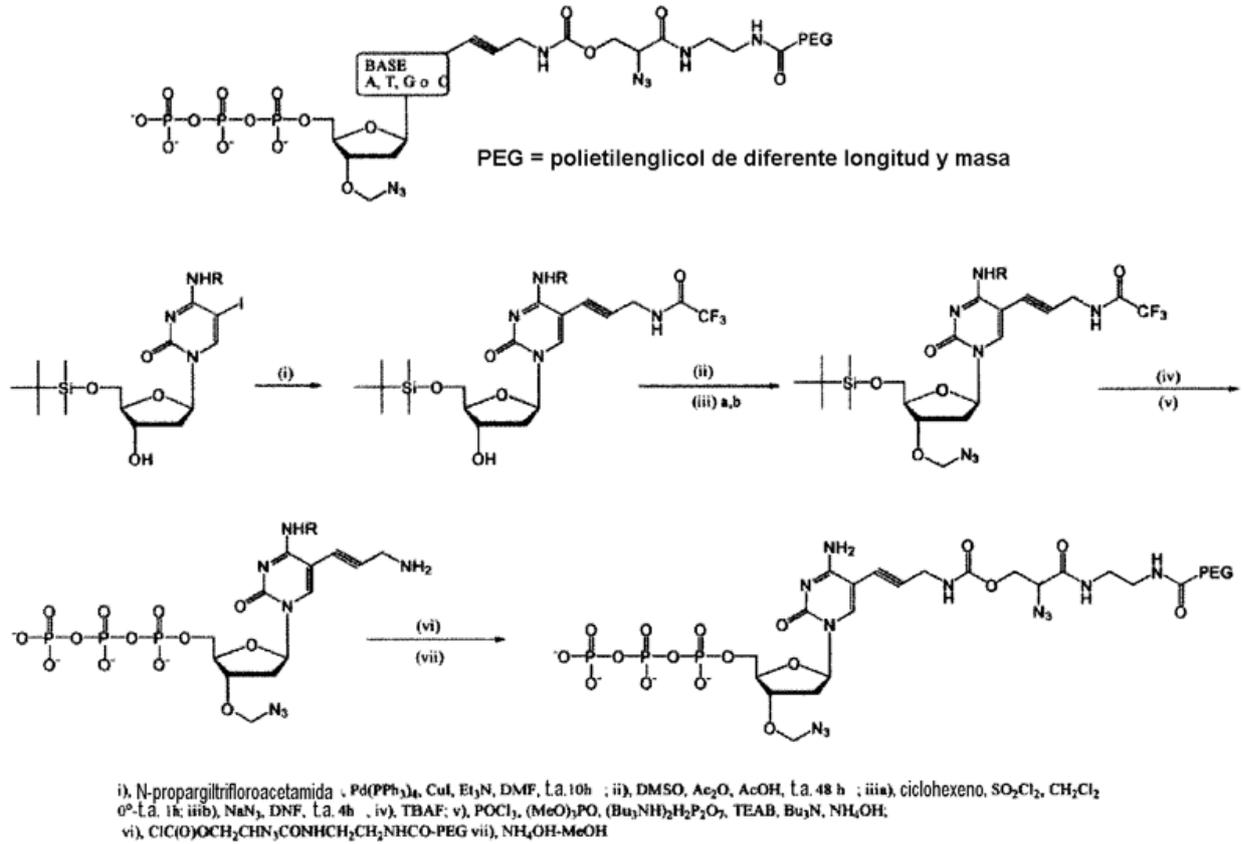


Figura 28(B)

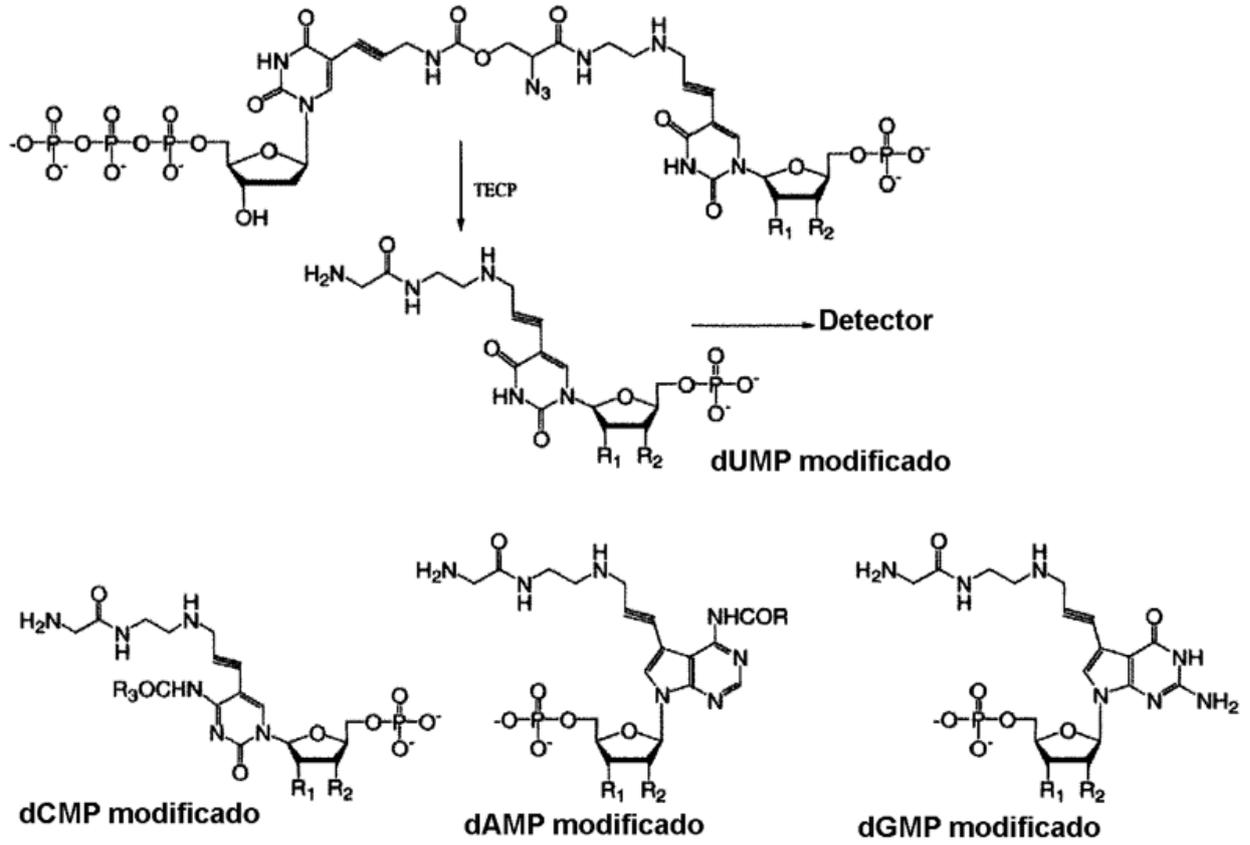


Figura 29

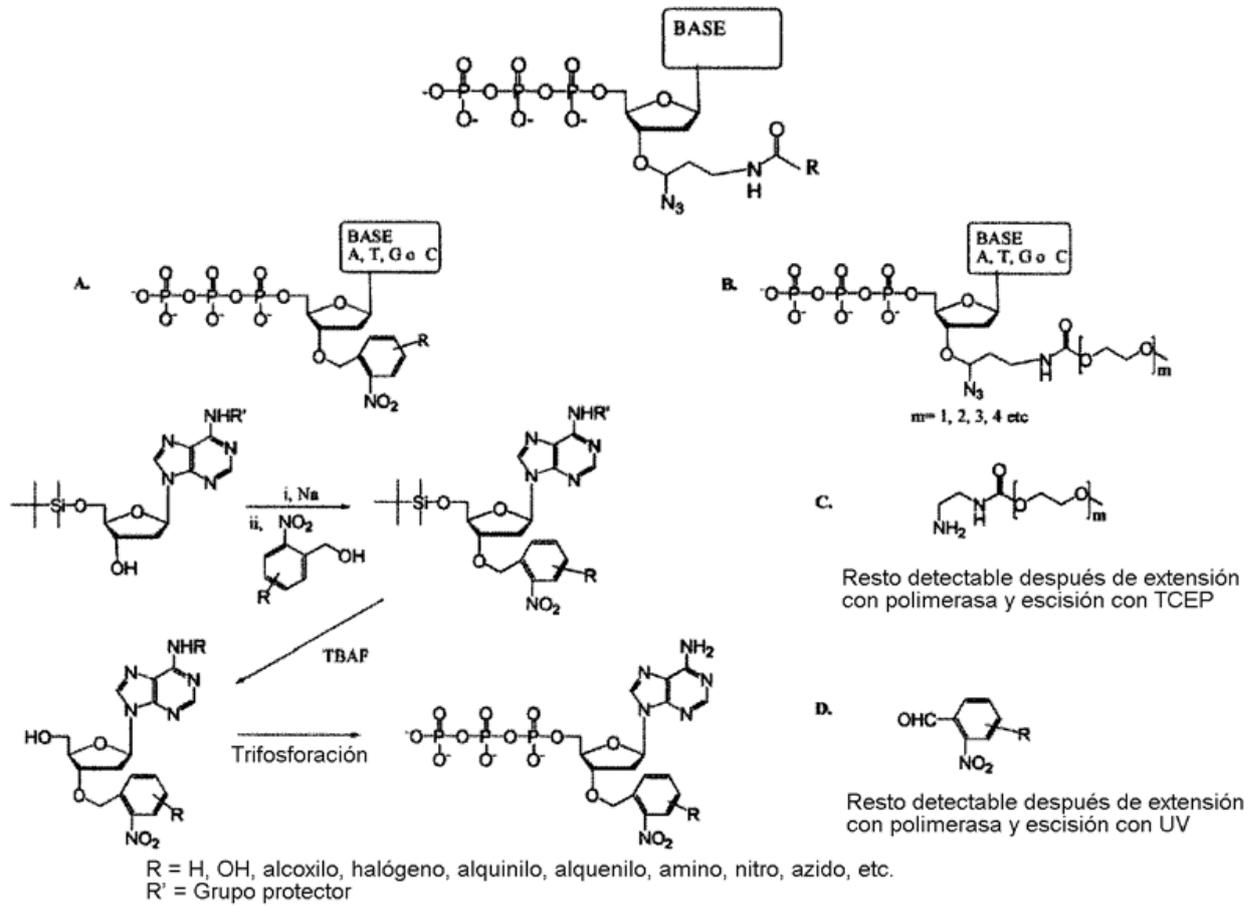


Figura 30

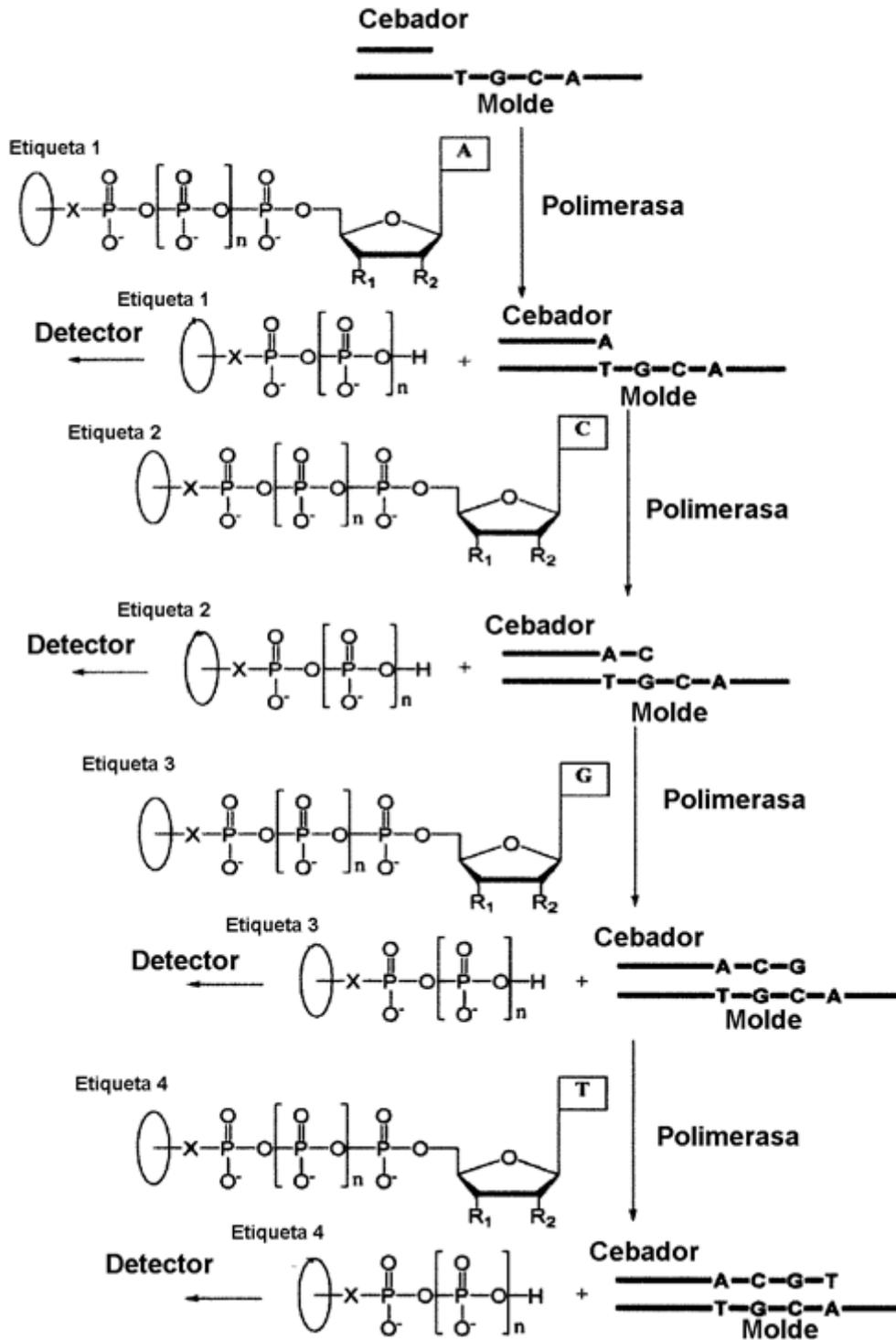


Figura 31

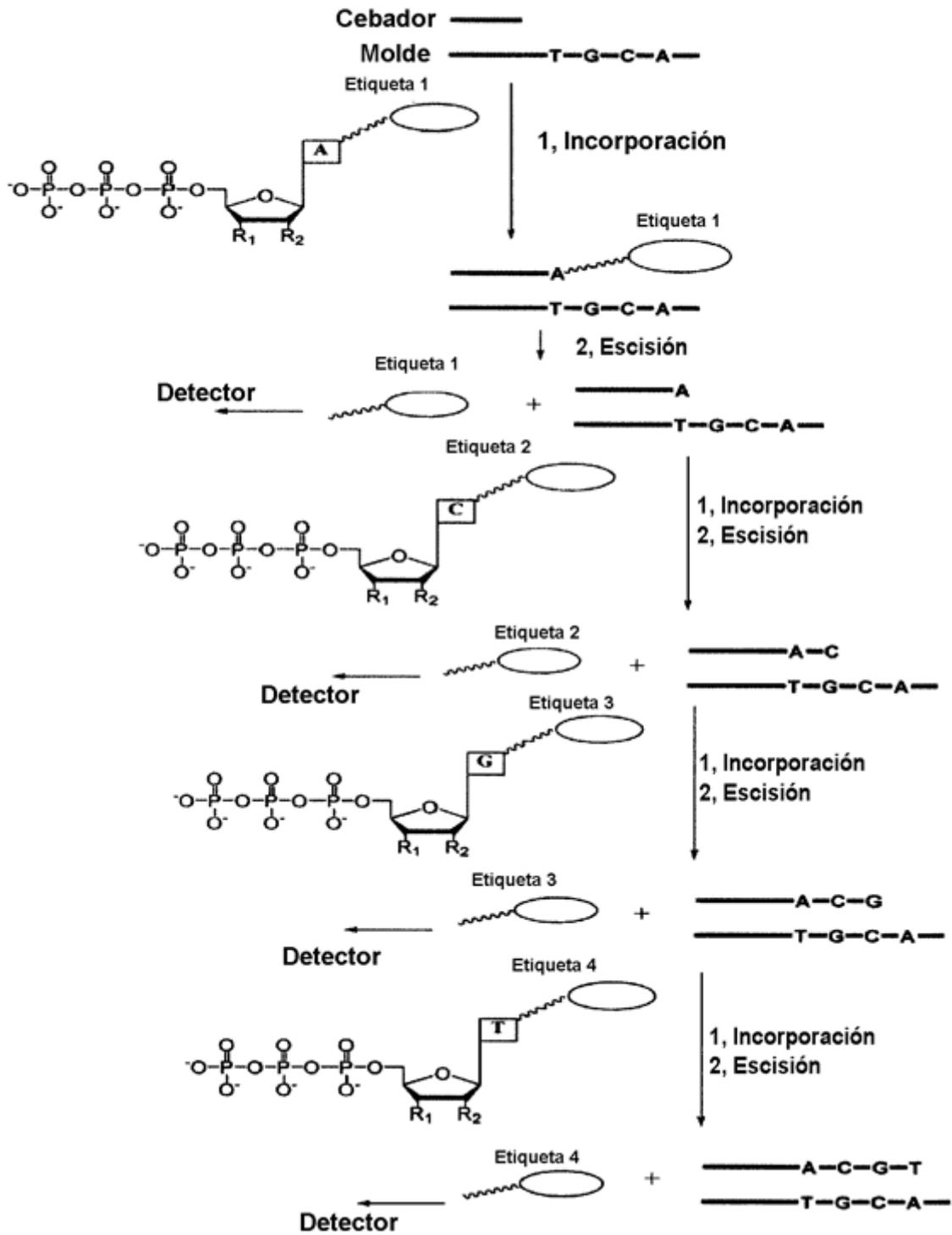


Figura 32

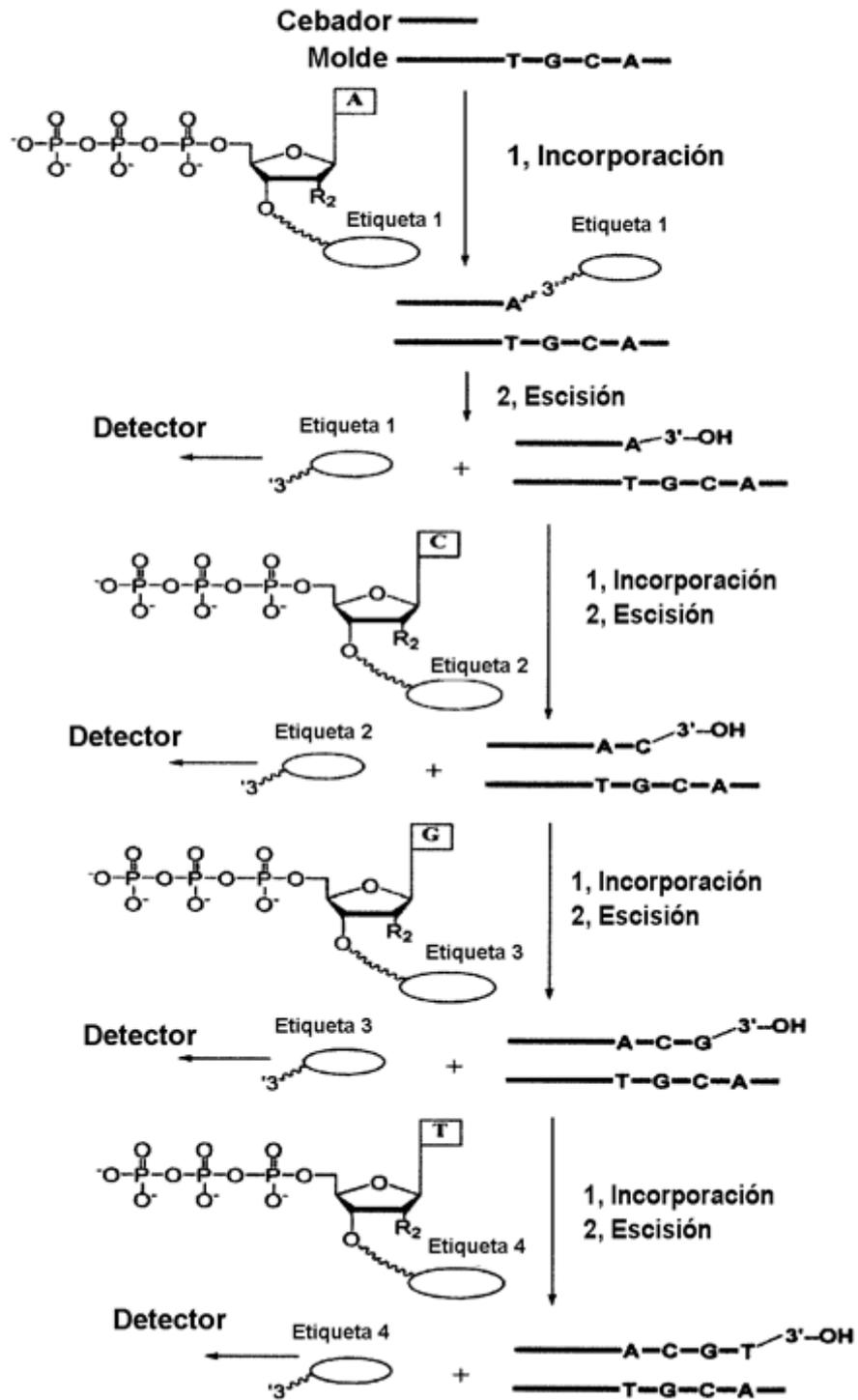


Figura 33

Añadir: A-etiqueta 1
C-etiqueta 2
G-etiqueta 3
T-etiqueta 4

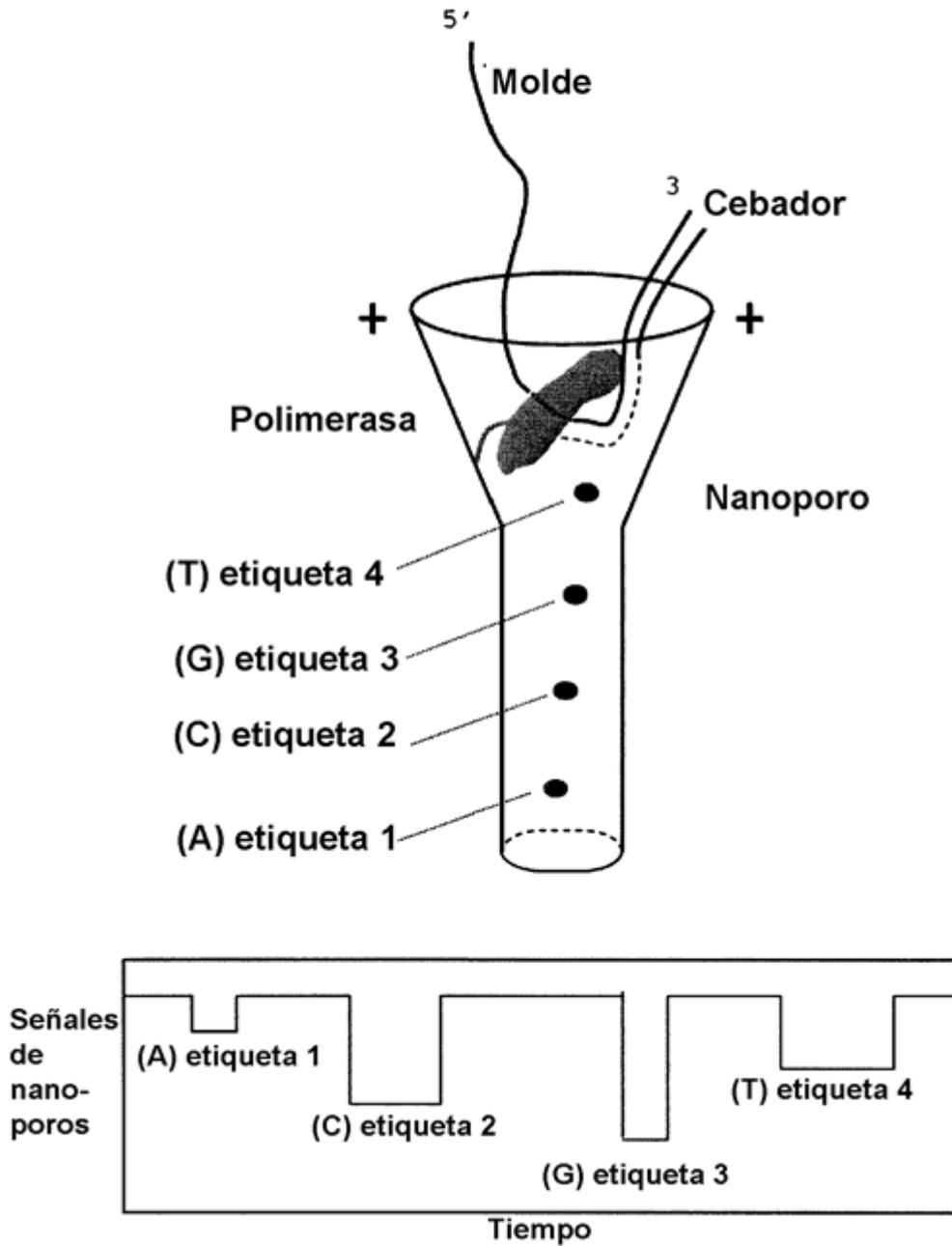
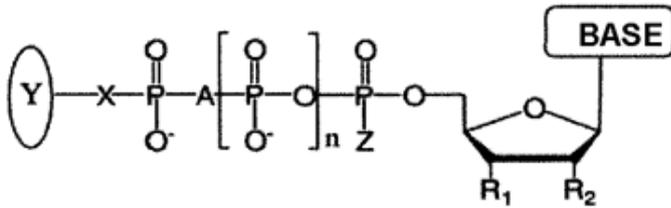
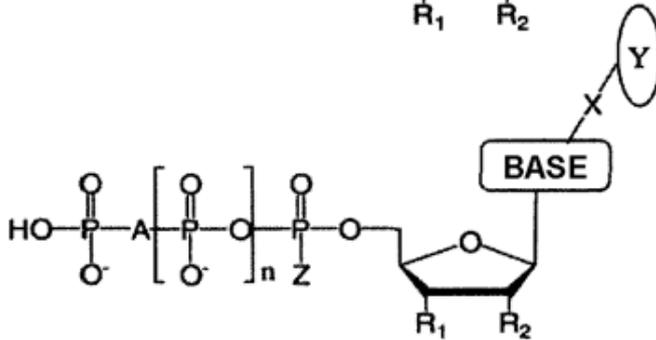


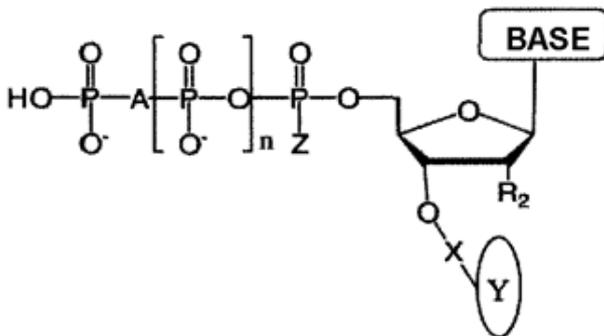
Figura 34



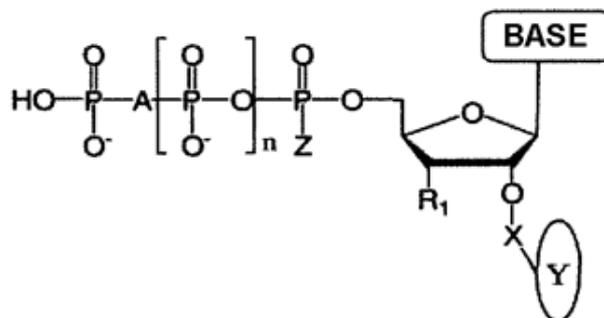
Nucleótidos marcados con etiqueta en fosfato



Nucleótidos marcados con etiqueta en base

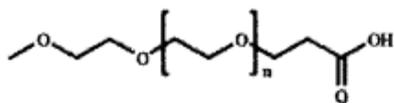
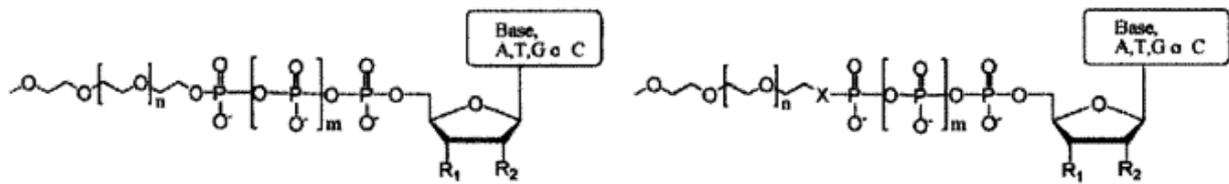


Nucleótidos marcados con etiqueta en 3'-O

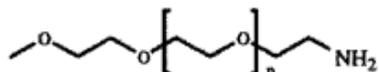


Nucleótidos marcados con etiqueta en 2'-O

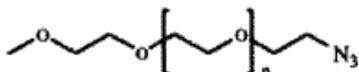
Figura 35



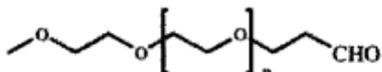
Reacciona con grupo amino



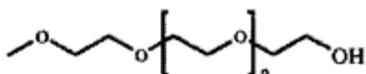
Reacciona con ácidos o aldehidos activados



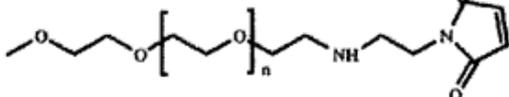
Reacciona con alquinos



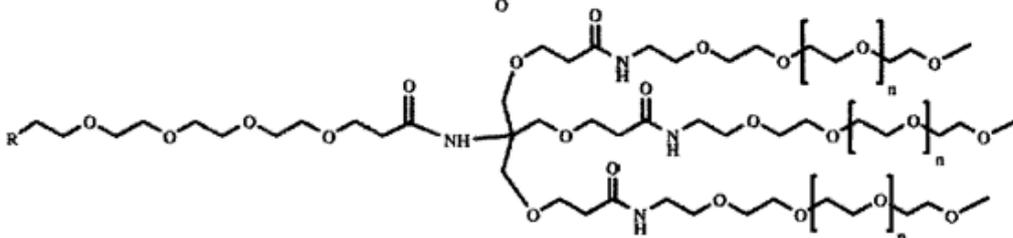
Reacciona con grupo amino



Reacciona con fosfatos activados



Reacciona con tioles



R = NH₂, OH, COOH, COOR', CHO, SH, N₃

Figura 36

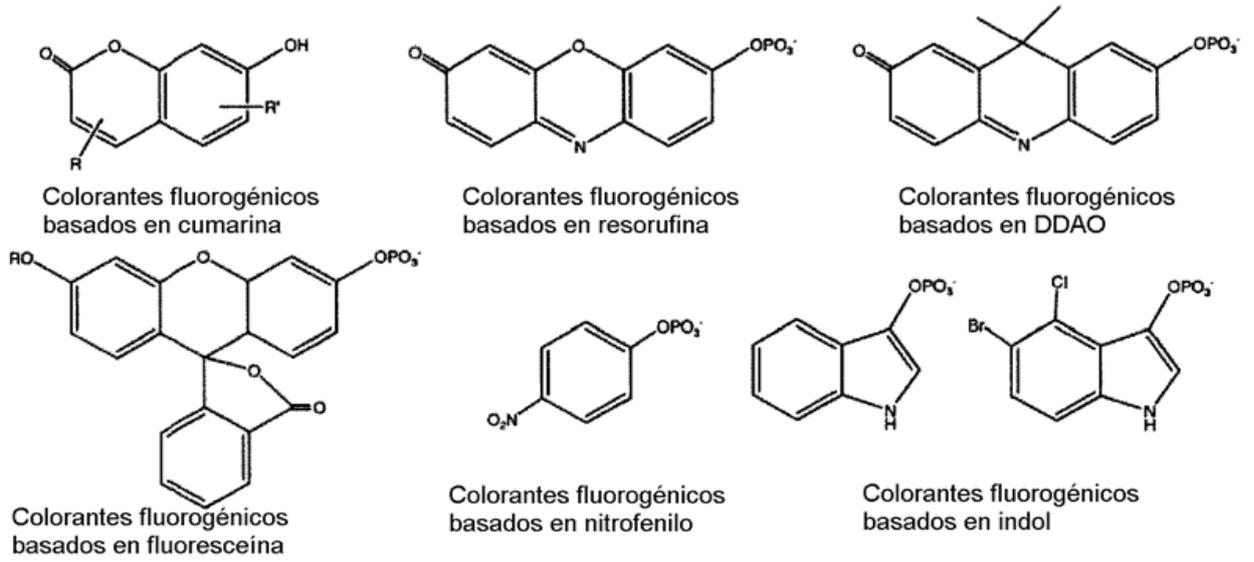
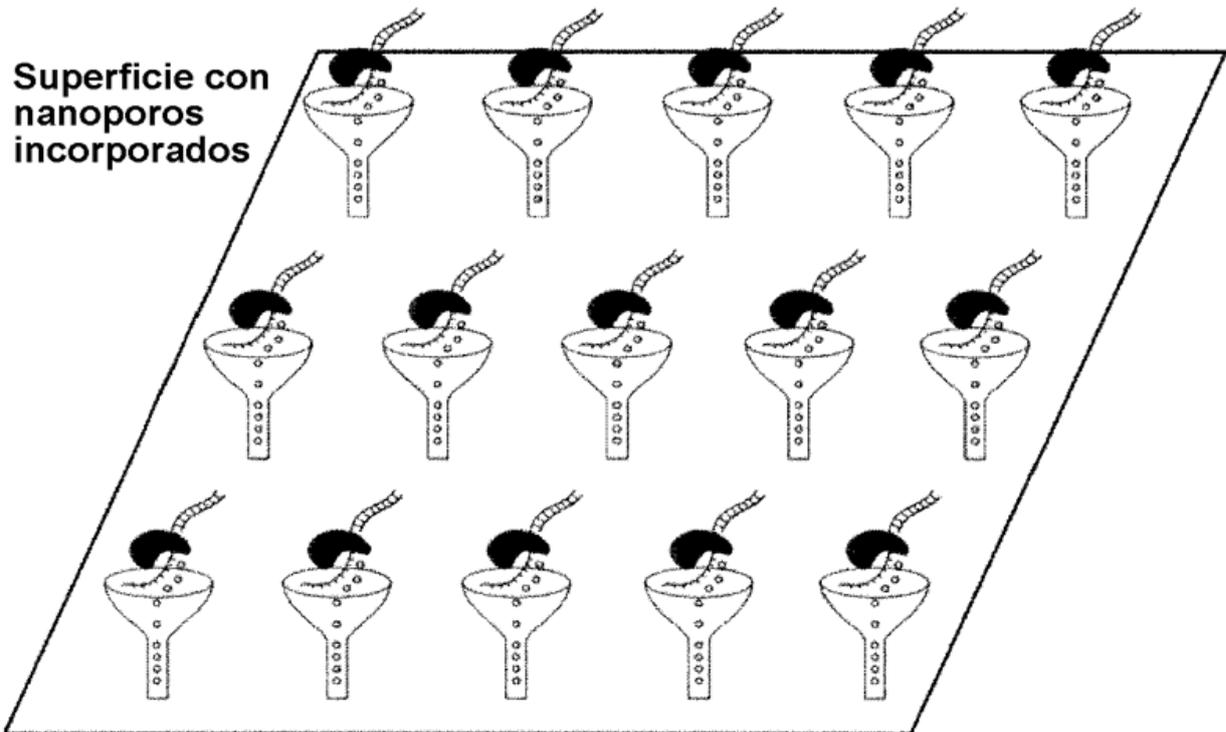
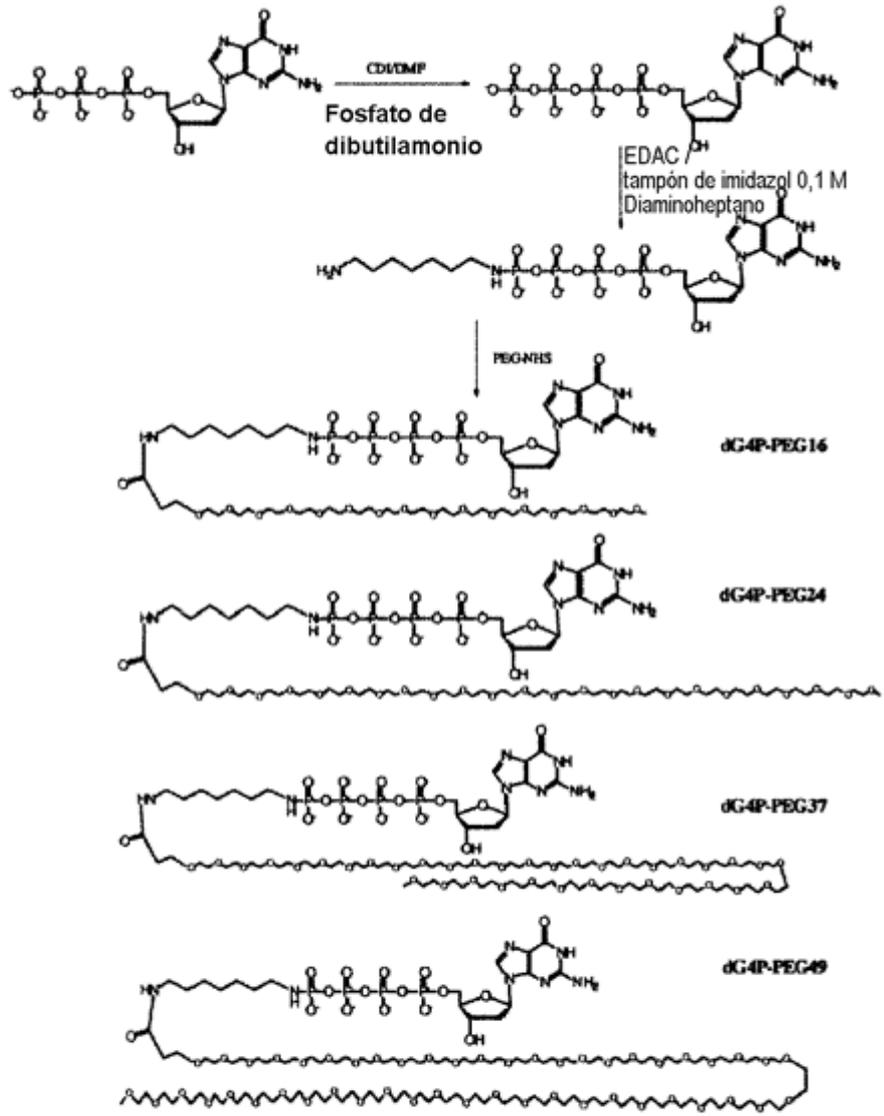


Figura 37

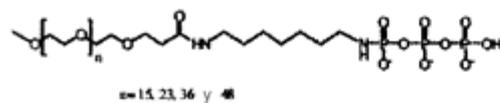


Matriz de nanoporos para secuenciación mediante síntesis

Figura 38

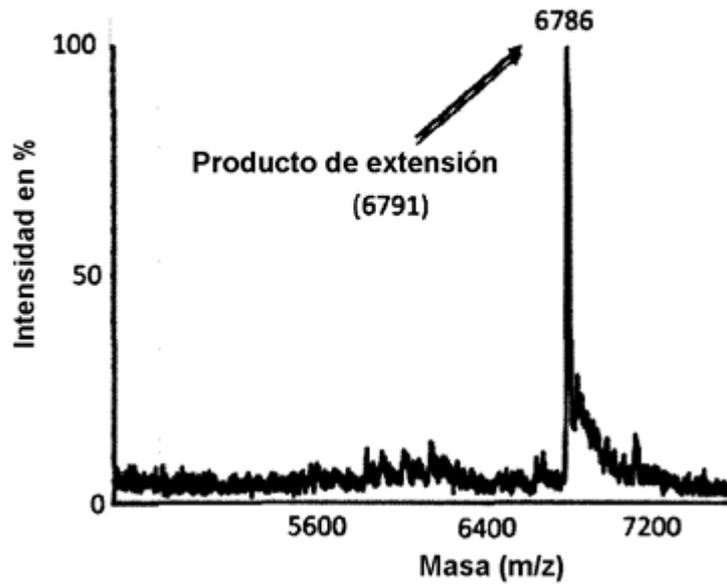


PEG o fosfatos de PEG liberados tras incorporación con polimerasa

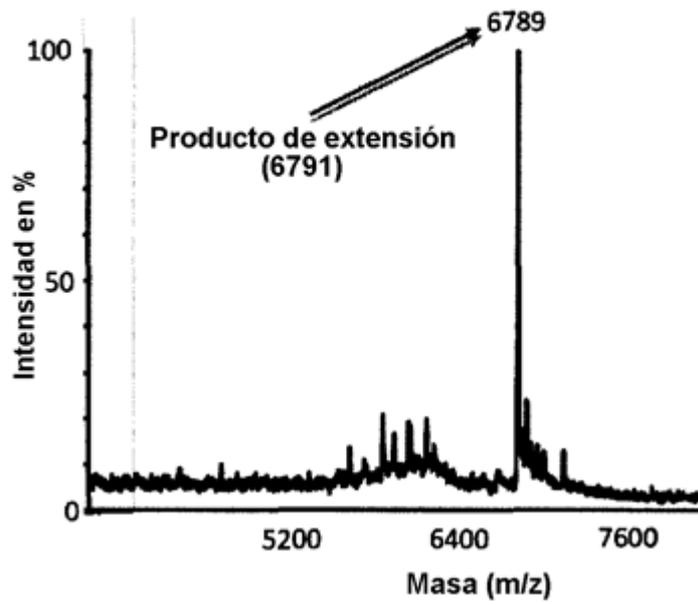


Síntesis de nucleótidos con etiqueta de PEG de primera generación

Figura 39



Reacción de extensión con dG4P-PEG24



Reacción de extensión con dG4P-PEG37

Figura 40

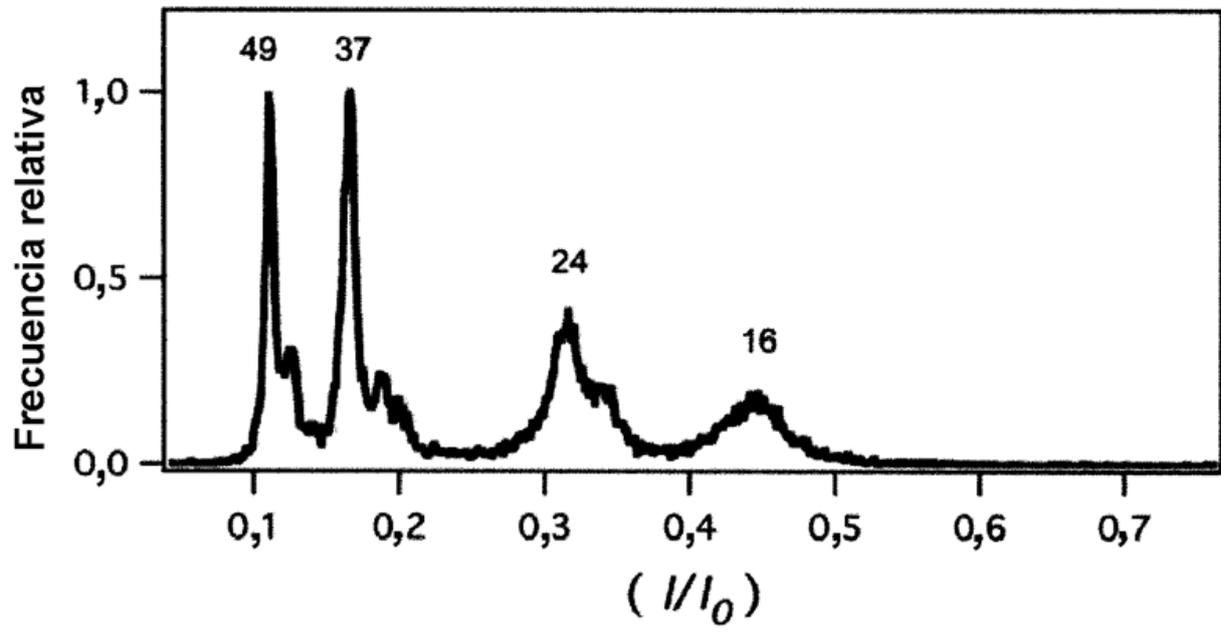


Figura 42

Síntesis de trifosfatos de PEG cargados

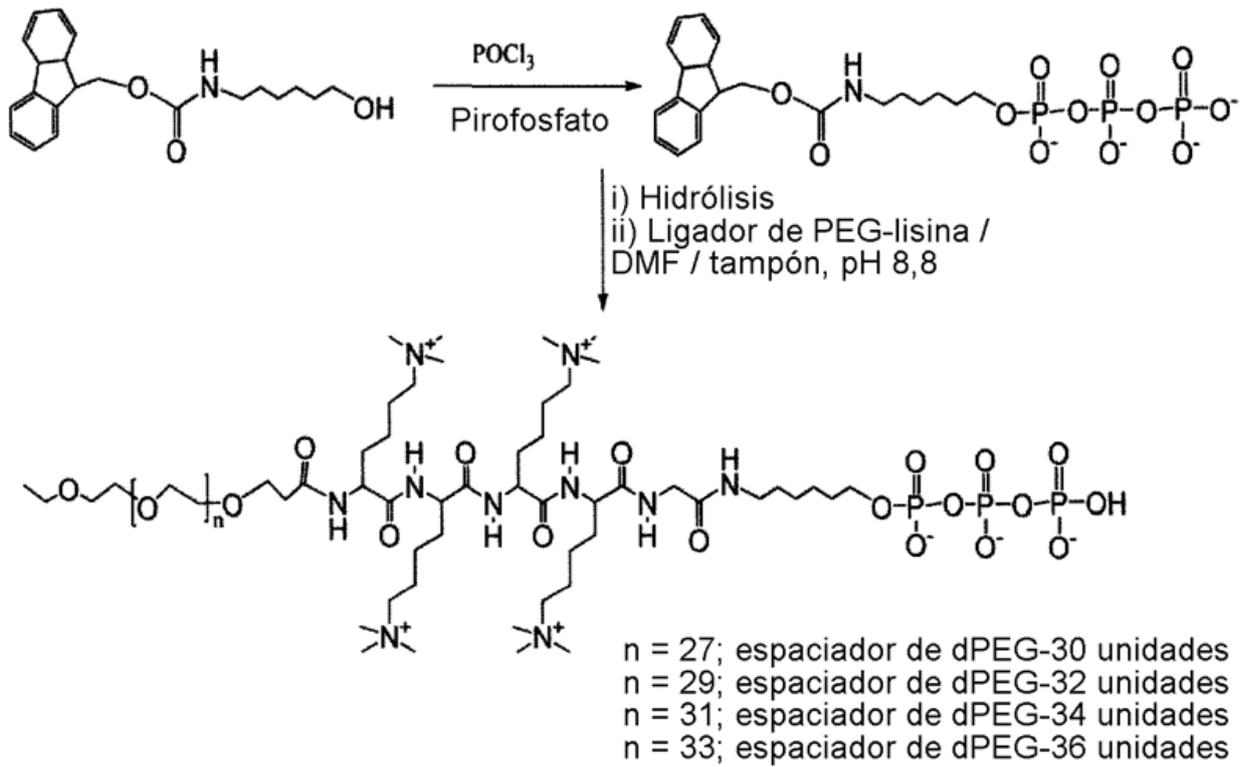


Figura 43

Síntesis de 5'-trifosfatos de nucleósido marcados con etiqueta en fosfato

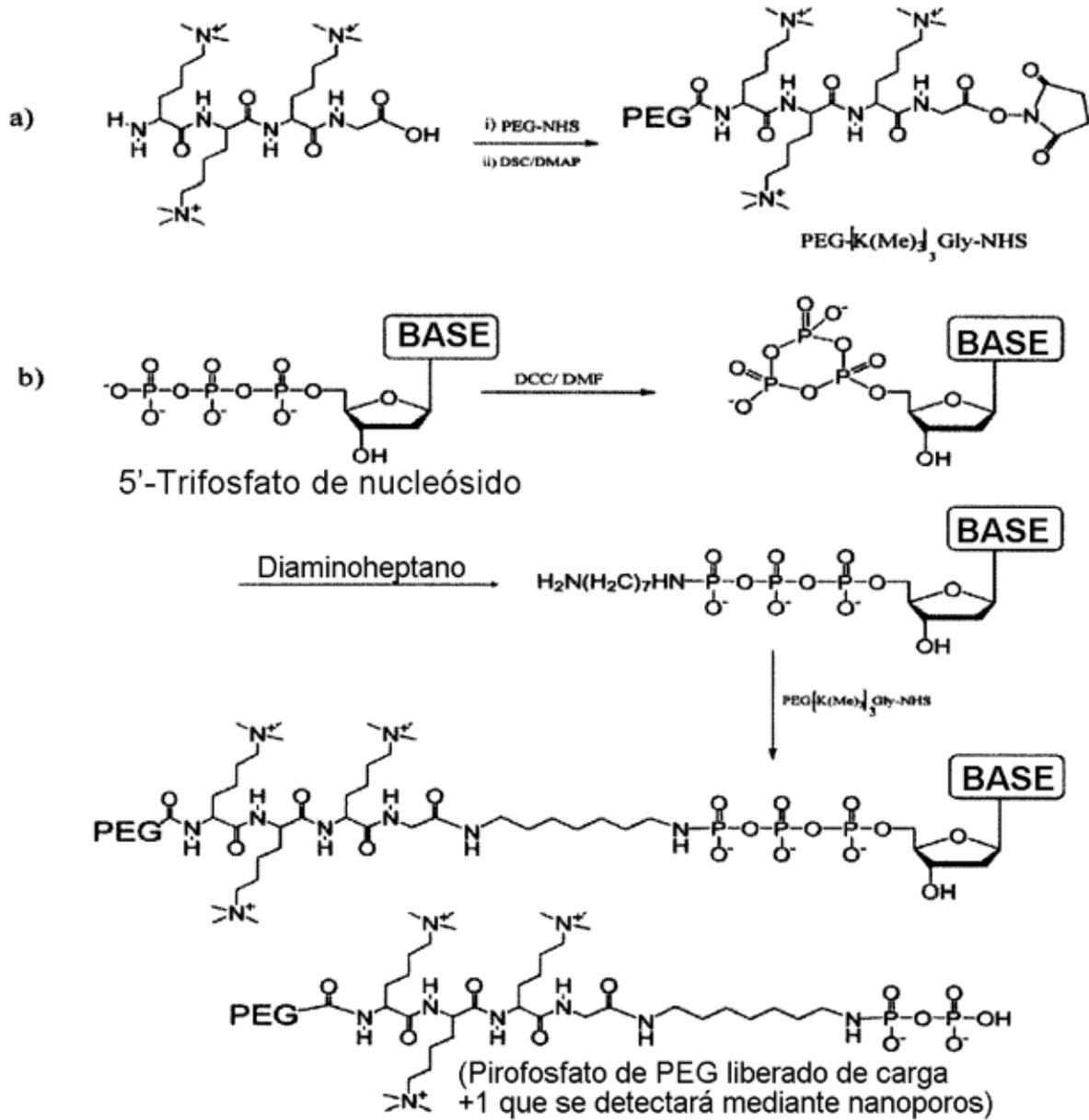


Figura 44

Síntesis de 5'-tetrafosfatos de nucleósido marcados con etiqueta en fosfato

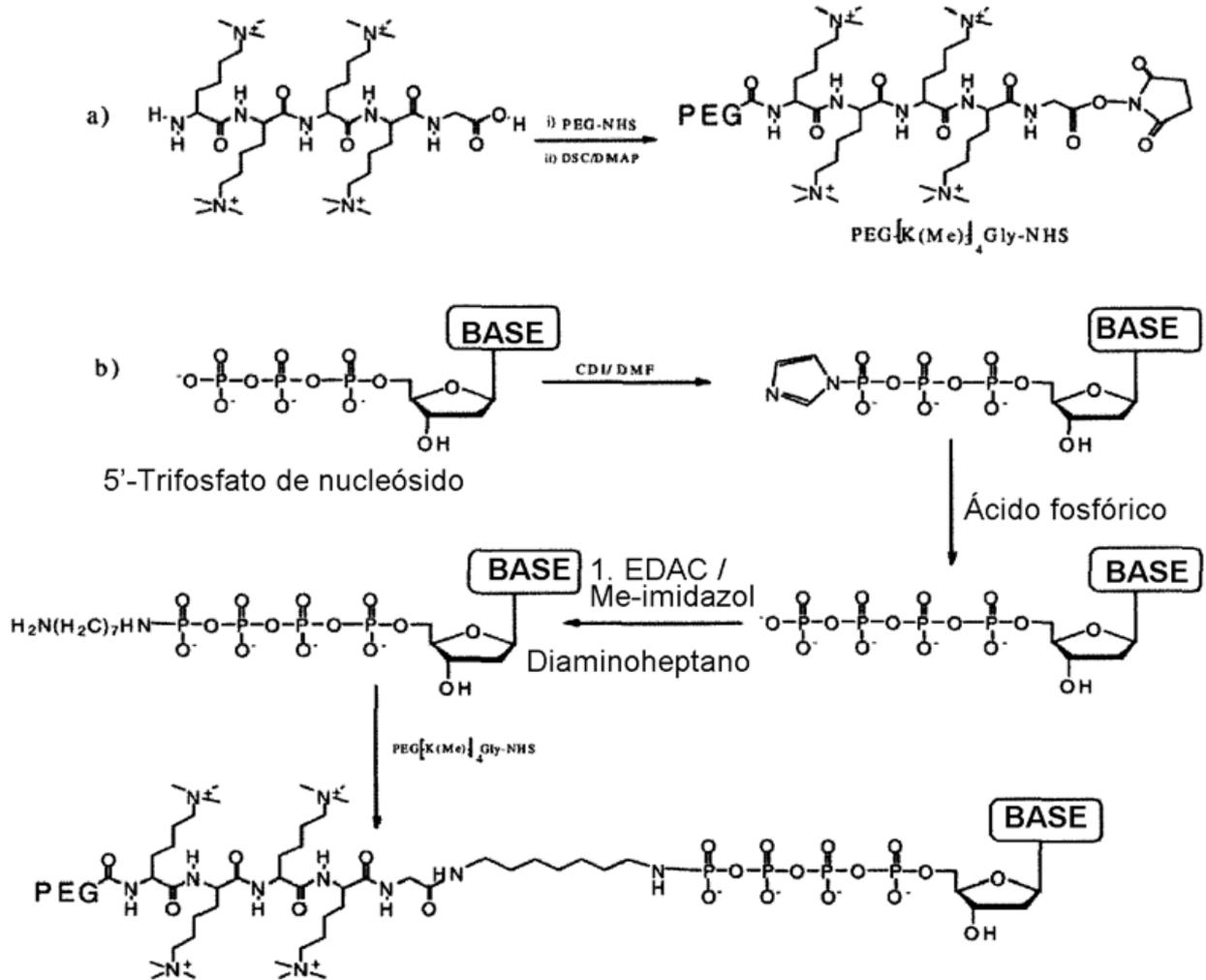


Figura 45

Síntesis de 5'-pentafosfatos de nucleósido marcados con etiqueta en fosfato terminal

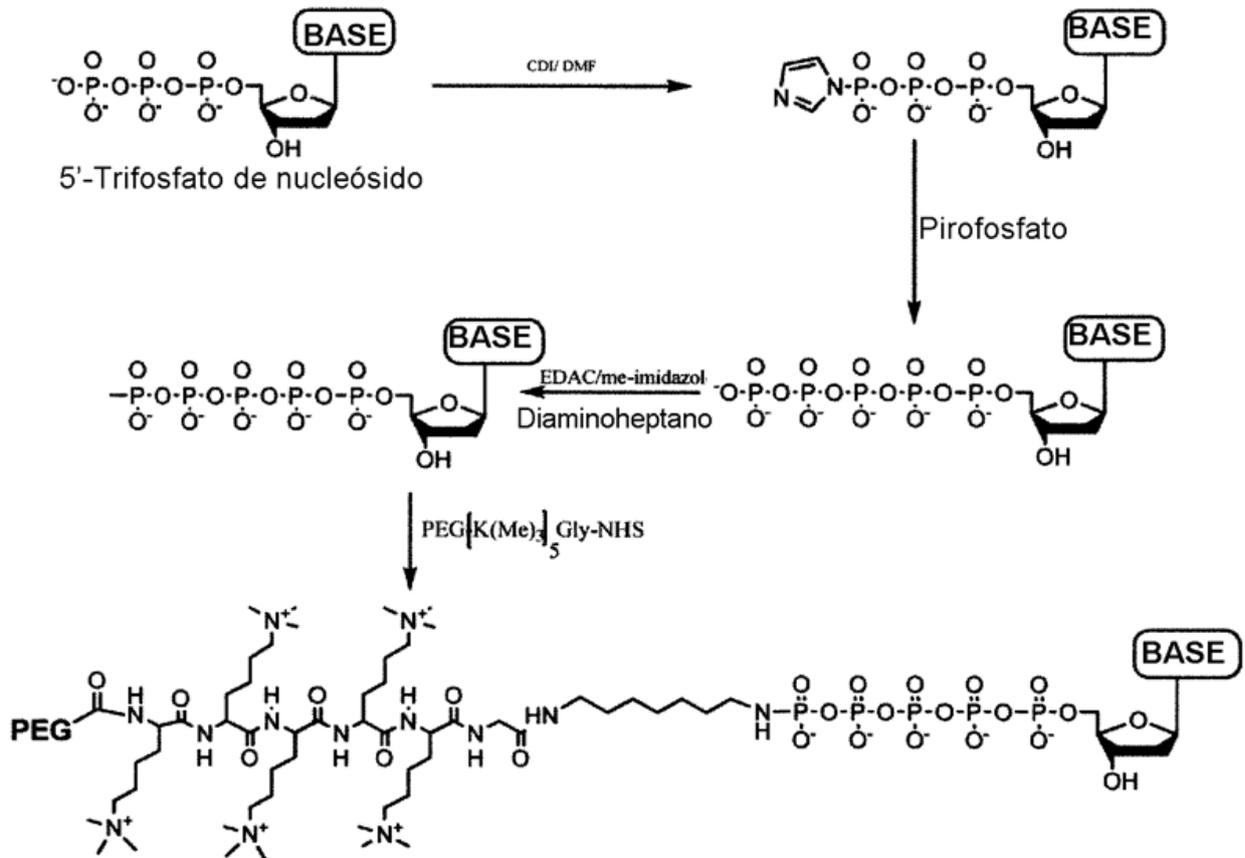


Figura 47

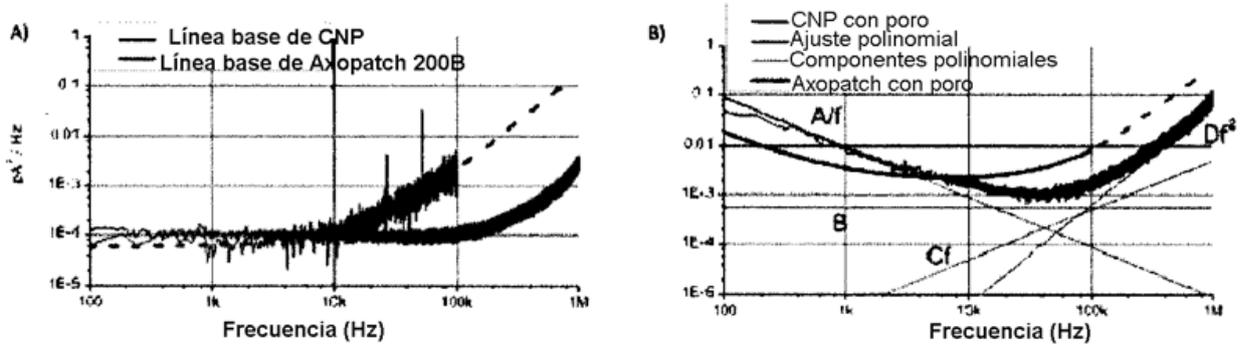


Figura 48

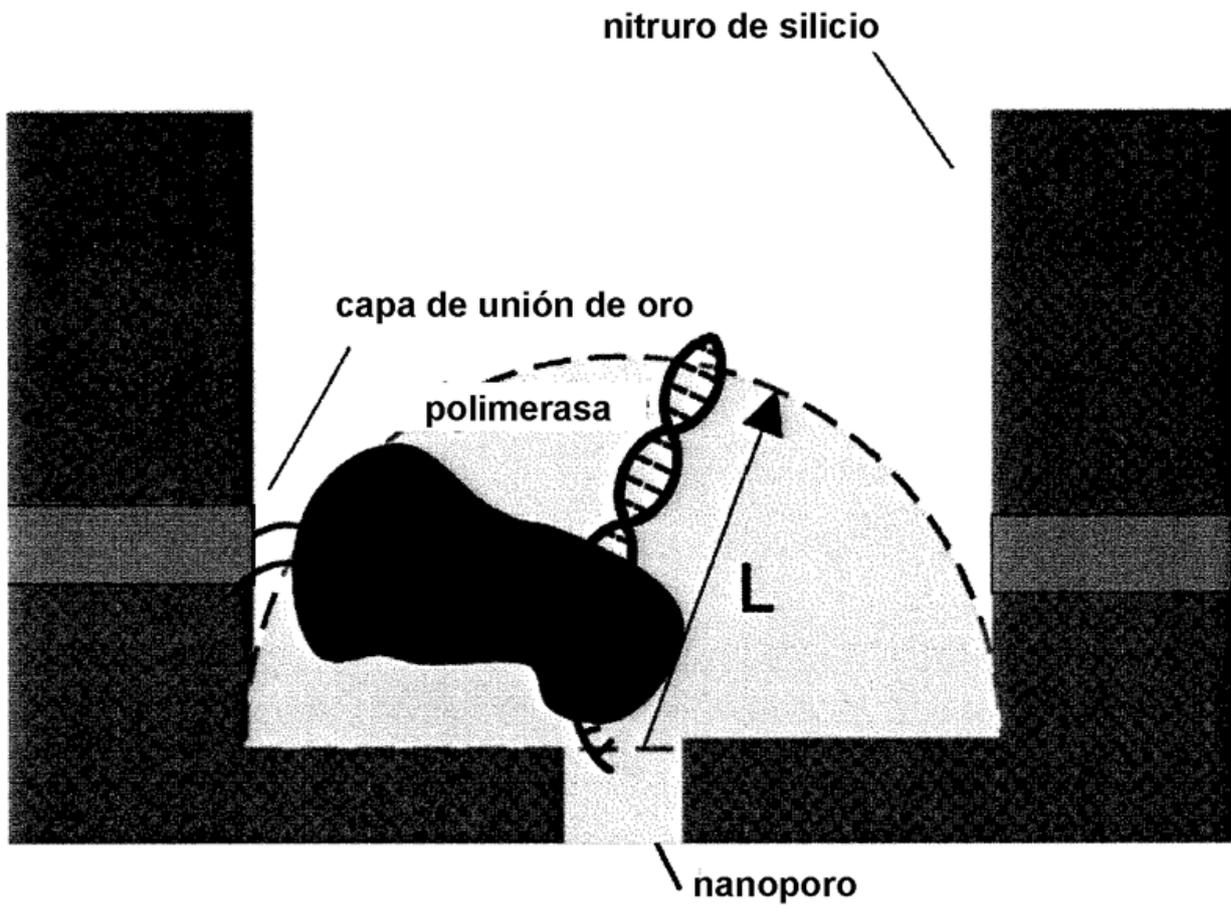
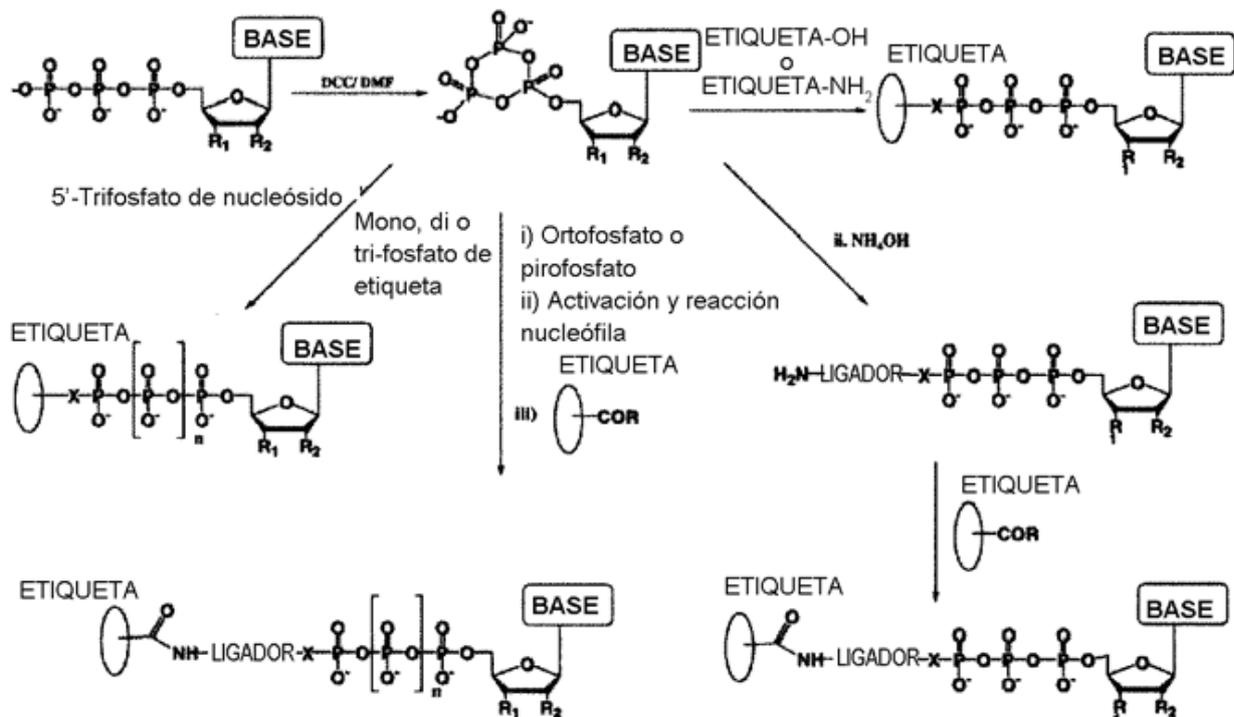


Figura 49



BASE = A, C, T, U, G, 7-desaza-A, 7-desaza-G o derivados de las mismas

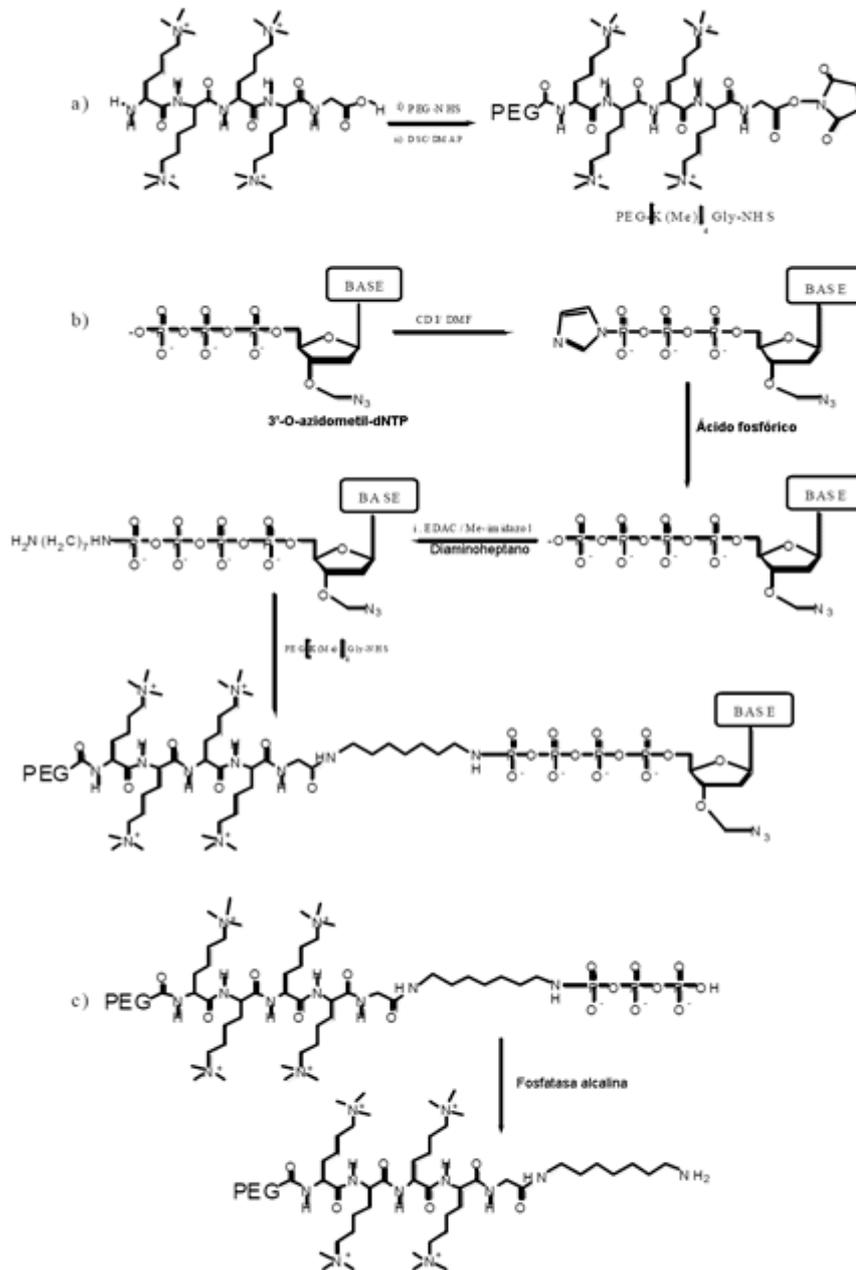
R₁ y R₂ = H u OH

X = O, S, NH; n = 2, 3, 4

ETIQUETA-OH/NH₂ = resto detectable tal como aminas o alcoholes alifáticos o aromáticos

ETIQUETA-COR = resto detectable para reaccionar con el grupo amino, tal como ésteres de NHS, cloruro de ácido, etc.

Figura 50



Síntesis de PEG-nucleótidos bloqueados en 3'-O

Figura 51

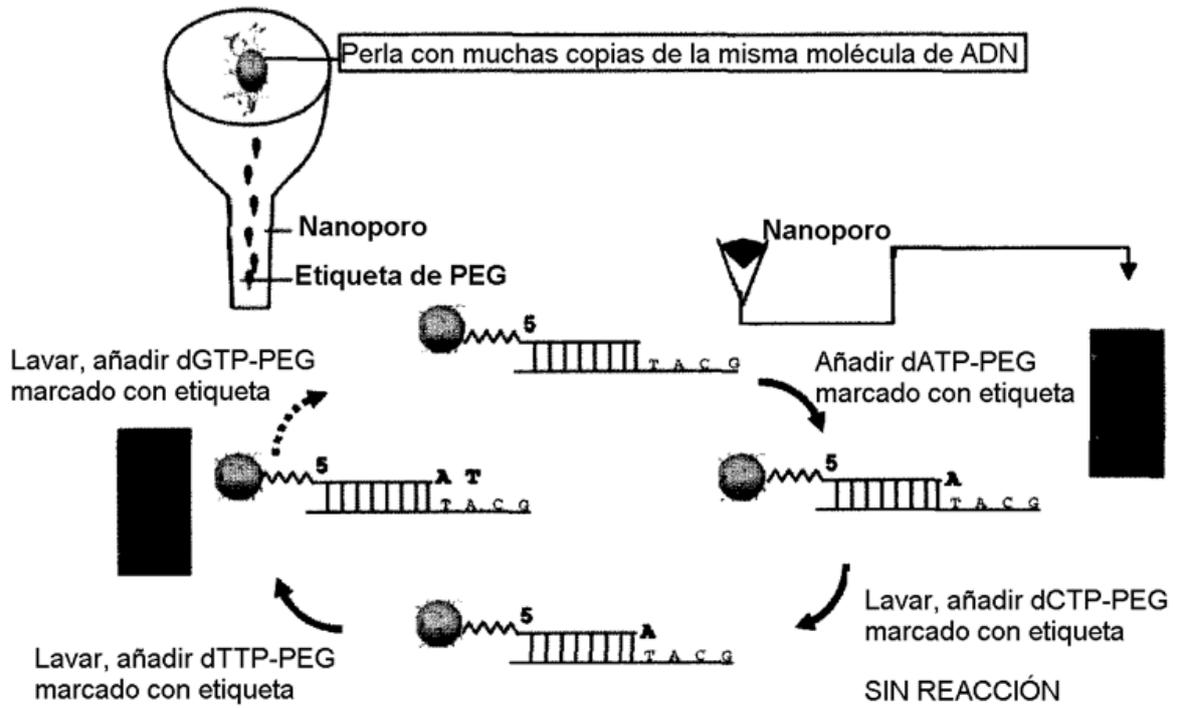


Figura 52

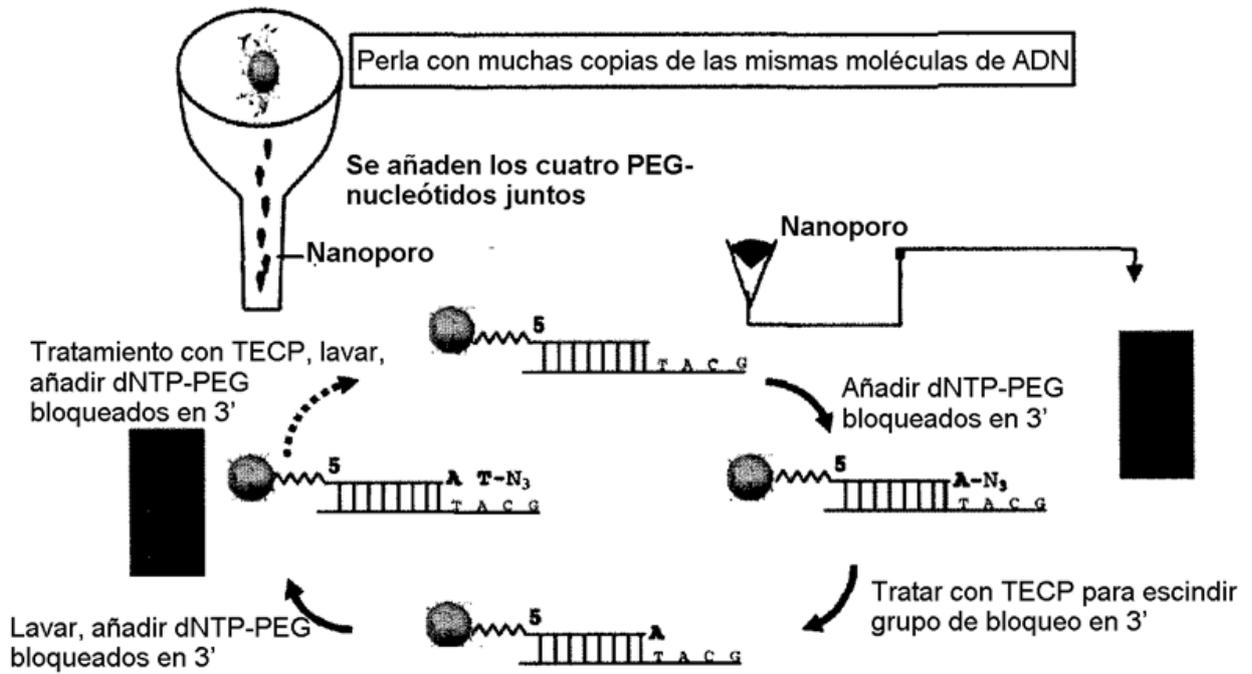


Figura 53

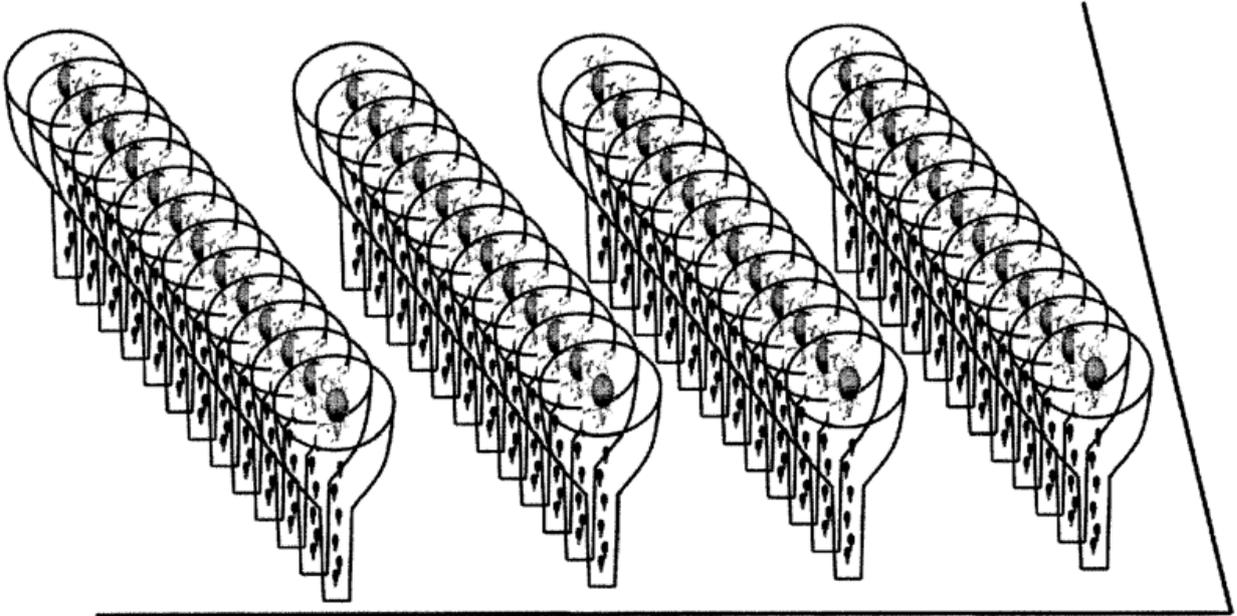


Figura 54

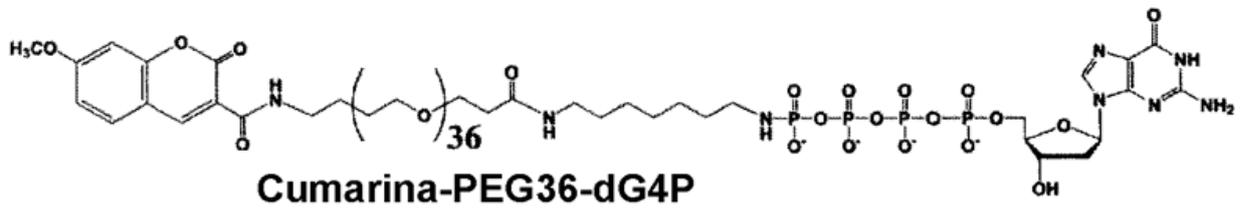
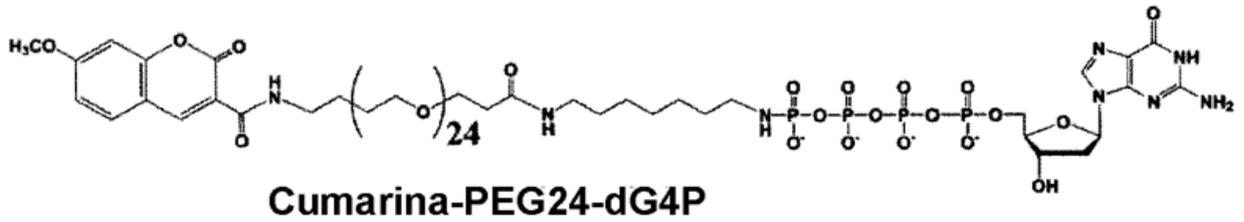
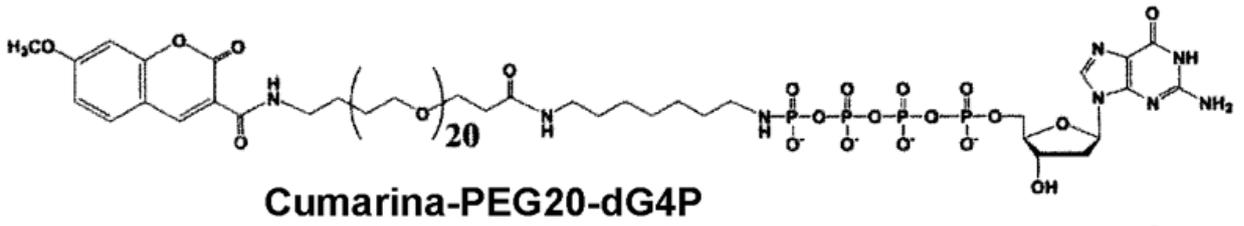
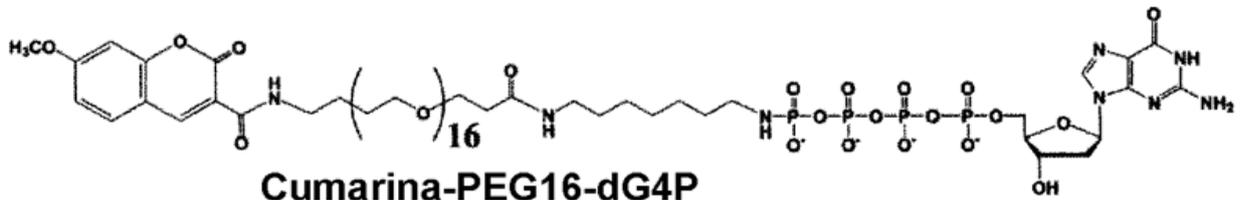
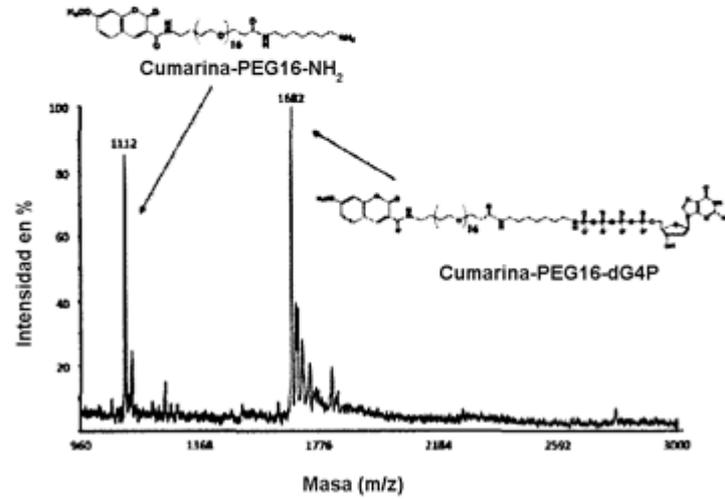


Figura 55

A. Espectro de masas de MALDI-TOF de cumarina-PEG16-dG4P



B. Espectro de masas de MALDI-TOF de cumarina-PEG20-dG4P

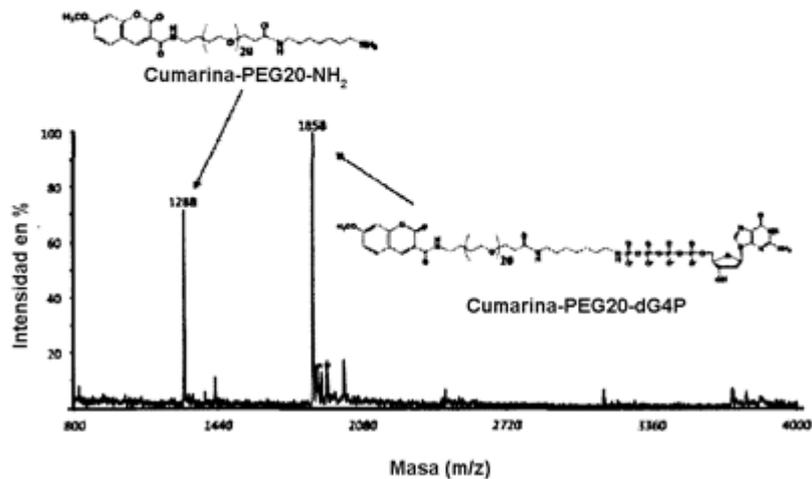
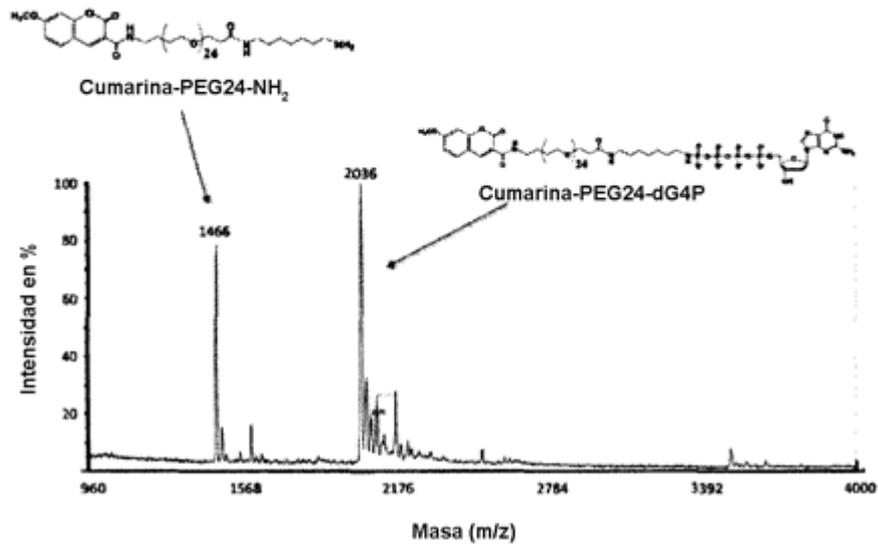


Figura 55 (continuación)

C. Espectro de masas de MALDI-TOF de cumarina-PEG24-dG4P



D. Espectro de masas de MALDI-TOF de cumarina-PEG36-dG4P

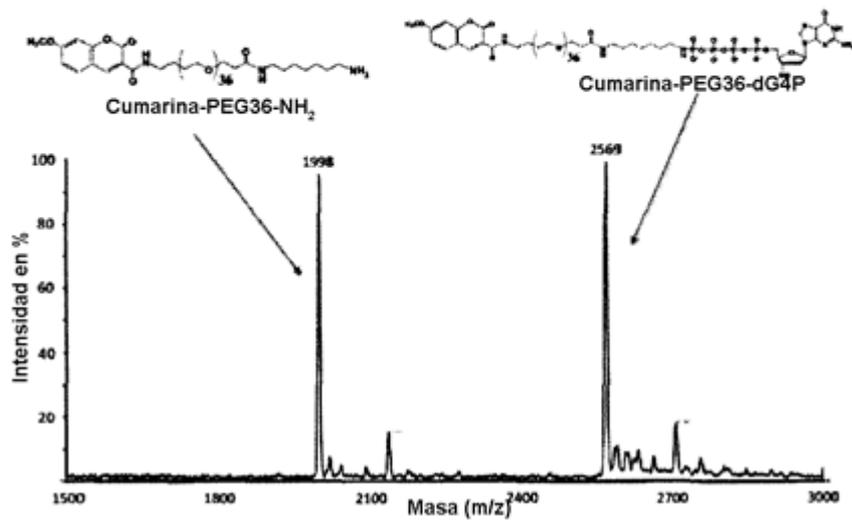


Figura 56

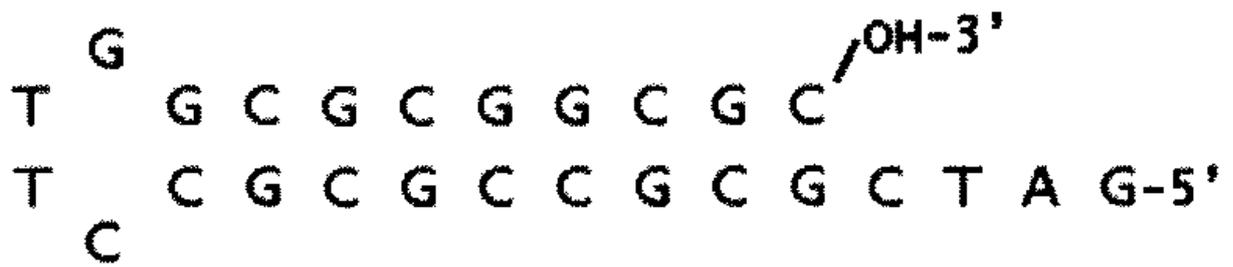


Figura 58

