

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 713 653**

51 Int. Cl.:

C07H 21/04 (2006.01)

C07H 19/00 (2006.01)

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **12.03.2014** **PCT/US2014/024391**

87 Fecha y número de publicación internacional: **25.09.2014** **WO14150851**

96 Fecha de presentación y número de la solicitud europea: **12.03.2014** **E 14769128 (1)**

97 Fecha y número de publicación de la concesión europea: **16.01.2019** **EP 2970366**

54 Título: **Análogos nucleotídicos para secuenciación**

30 Prioridad:

15.03.2013 US 201361791730 P

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

23.05.2019

73 Titular/es:

IBIS BIOSCIENCES, INC. (100.0%)
2251 Faraday Avenue Suite 150
Carlsbad, CA 92008, US

72 Inventor/es:

HAYDEN, MARK A.;
HUFF, JEFFREY;
ESHOO, MARK W. y
PICURI, JOHN

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 713 653 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

Análogos nucleotídicos para secuenciación

5 Campo de la invención

En el presente documento se proporciona tecnología relacionada con la secuenciación de ácidos nucleicos, pero no exclusivamente, con composiciones, métodos, sistemas y kits relacionados con nucleótidos que comprenden un resto detectable electroquímicamente y uno o más restos inhibidores de la síntesis fotolábiles.

10 Antecedentes

La secuenciación del ADN está impulsando la investigación y el descubrimiento de la genómica. La finalización del Proyecto Genoma Humano fue un logro monumental con una increíble cantidad de esfuerzos combinados entre los centros del genoma y los científicos de todo el mundo. Este proyecto de una década se completó con el método de secuenciación de Sanger, que sigue siendo la metodología básica de secuenciación del genoma en los centros de secuenciación de genoma de alto rendimiento. La razón principal detrás del éxito prolongado de este método es su método básico y eficiente, aunque elegante, método de terminación de cadena dideoxi. Con mejoras crecientes en la secuenciación de Sanger, que incluyen el uso de la excitación fluorescente inducida por láser de los colorantes de transferencia de energía, las ADN polimerasas diseñadas, la electroforesis capilar, la preparación de muestras, la informática y el software de análisis de secuencias, la plataforma de secuenciación de Sanger ha podido mantener su estado. Los actuales secuenciadores de ADN basados en Sanger de última generación pueden producir más de 700 bases de secuencia claramente legible en una sola ejecución desde los moldes hasta 30 kb de longitud. Sin embargo, como ocurre con la mayoría de las invenciones tecnológicas, las mejoras continuas en esta plataforma de secuenciación han llegado a un estancamiento, con una estimación del coste actual para producir una secuencia borrador de genoma microbiano de alta calidad en aproximadamente 10.000 dólares americanos por par de megabases. Los secuenciadores de ADN actuales basados en el método de Sanger permiten analizar en paralelo hasta 384 muestras.

Es evidente que la explotación de la secuencia completa del genoma humano para la medicina clínica y la atención médica requiere métodos precisos de secuenciación de ADN de bajo costo y alto rendimiento. De hecho, tanto el sector público (National Human Genome Research Institute, NHGRI) como en el sector privado de ciencias genómicas (The J. Craig Venter Science Foundation and Archon X prize for genomics) han emitido una llamada al desarrollo de tecnología de secuenciación de "próxima generación" que reduzca el coste de la secuenciación a una décima parte de su coste actual en los próximos diez años. En consecuencia, para superar las limitaciones de las tecnologías de secuenciación convencionales actuales, se han investigado diversos nuevos métodos de secuenciación de ADN, incluidos los enfoques de secuenciación por síntesis (SBS), tales como pirosecuenciación (Ronaghi et al. (1998) Science 281: 363-365), secuenciación de moléculas de ADN individuales (Braslaysky et al. (2003) Proc. Natl. Acad. Sci. USA 100: 3960-3964) y colonias de polimerasa (secuenciación "polonia") (Mitra et al. (2003) Anal. Biochem. 320: 55-65).

Algunas tecnologías de secuenciación de próxima generación convencionales incluyen métodos de detección óptica de una sola molécula, por ejemplo, como se usan en tecnologías desarrolladas por PacBio; métodos ópticos (clonales), por ejemplo, como se usan en las tecnologías desarrolladas por Illumina; y métodos basados en nucleótidos marcados con fluorescencia (incluidos los que utilizan la fotodesprotección), por ejemplo, como se usa en la tecnología desarrollada por Lasergen. Dichos métodos tienen diversos grados de ventajas y desventajas, pero el desafío importante hasta ahora ha sido el problema de realizar tales análisis de secuencia con sistemas de instrumentación de coste ultrabajo con reactivos desechables y de bajo coste.

El concepto de secuenciación por síntesis de ADN (SBS) se reveló en 1988 con un intento de secuenciar el ADN detectando el grupo pirofosfato que se genera cuando se incorpora un nucleótido mediante una reacción de ADN polimerasa (Hyman (1999) Anal. Biochem. 174: 423-436). Las tecnologías SBS posteriores se basaron en formas adicionales para detectar la incorporación de un nucleótido a una cadena de ADN en crecimiento. En general, la SBS convencional utiliza un cebador oligonucleotídico diseñado para hibridar con una posición predeterminada de la molécula del molde de muestra que se va a secuenciar. El complejo cebador-molde se presenta con un nucleótido en presencia de una enzima polimerasa. Si el nucleótido es complementario a la posición en la molécula del molde de muestra que está directamente en 3' del extremo del cebador oligonucleotídico, la ADN polimerasa extenderá el cebador con el nucleótido. La incorporación del nucleótido y la identidad del nucleótido insertado se pueden detectar, por ejemplo, mediante la emisión de luz, un cambio en la fluorescencia, un cambio en el pH (véase, por ejemplo, la patente de Estados Unidos N.º 7.932.034), un cambio en la conformación de la enzima, o algún otro cambio físico o químico en la reacción (véase, por ejemplo, el documento WO 1993/023564 y el documento WO 1989/009283; Seo et al. (2005) "Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides," PNAS 102: 5926-59). Tras cada incorporación con éxito de un nucleótido, se detecta una señal que refleja la ocurrencia, la identidad y el número de incorporaciones de nucleótidos. Los nucleótidos no incorporados se pueden eliminar (por ejemplo, por degradación química o por lavado) y la siguiente posición en el molde-cebador se puede consultar con otra especie de nucleótido.

Si bien se ha hecho evidente que la secuenciación de próxima generación tiene una amplia aplicación en los diagnósticos que incluyen cáncer, enfermedades infecciosas, medicamentos complementarios y enfermedades hereditarias, los sistemas de secuenciación de la próxima generación existentes están diseñados para secuenciar genomas completos y, por lo tanto, los sistemas tienen un alto coste por prueba (por ejemplo, aproximadamente de 100 a 500 dólares americanos por prueba). Además, la implementación de los sistemas comerciales actuales también es costosa (por ejemplo, de 75.000 a 700.000 dólares americanos) y el flujo de trabajo de muestra a secuencia es laborioso. Como tales, las tecnologías existentes no proporcionan una muestra al sistema de secuencia que es deseable para aplicaciones de diagnóstico.

Como tal, es un objetivo generar datos de alta calidad a un coste razonable y entregar datos de secuenciación de próxima generación de manera precisa y rápida en un sistema fácil de usar. Empresas como PacBio han desarrollado procesos químicos específicos para la implementación en sus sistemas. Al mismo tiempo, otras empresas como VisiGen y Life Technologies han optado por métodos químicos alternativos para abordar la secuenciación de bajo coste.

En particular, LaserGen ha desarrollado enfoques que utilizan sistemas de detección óptica y ciertos métodos químicos de reacción para producir y polimerizar nucleótidos foto-desprotegibles que puedan emplearse en aplicaciones de secuenciación de próxima generación, por ejemplo, como se describe en las patentes de Estados Unidos n.º 7.893.227; 7.897.737; 7.964.352; y 8.148.503. Los nucleótidos de LaserGen tienen un resto terminador fluorescente y fotoescindible unido a la base nucleotídica y un hidroxilo 3' no bloqueado en el azúcar ribosa. El terminador fluorescente fotoescindible proporciona un sustrato para la polimerización, por ejemplo, una polimerasa añade el análogo nucleotídico al hidroxilo en 3' de la cadena sintetizada. Mientras está unido al nucleótido en el extremo 3', el terminador fluorescente fotoescindible evita la adición de nucleótidos adicionales por la polimerasa. Asimismo, el resto fluorescente proporciona la identificación del nucleótido añadido usando una fuente de luz de excitación y un detector de emisión de fluorescencia. Al exponerse a una fuente de luz de la longitud de onda apropiada, la luz escinde el terminador fluorescente, fotoescindible del extremo 3' de la cadena, eliminando así el bloque a la síntesis y se agrega otro análogo nucleotídico para comenzar el ciclo nuevamente. Cuando se usan en una reacción de secuenciación por síntesis, los compuestos de nucleótidos marcados con fluorescencia LaserGen ofrecen una forma de fotodesprotección y al mismo tiempo permiten la extensión, por ejemplo, desbloqueando estéricamente la región en la enzima para permitir la extensión. Sin embargo, estos compuestos tienen que usar la fluorescencia para detectar la presencia o ausencia de un evento de incorporación particular. Como tal, la necesidad de utilizar la detección óptica en el contexto de la utilización de la escisión óptica de marcadores fotodesprotegibles en estas tecnologías es prohibitiva para lograr el coste óptimo más bajo al crear un sistema de secuenciación de bajo coste. El sistema requiere una coordinación cuidadosa de la excitación y la desprotección de las fuentes de luz para que la excitación no desproteja los nucleótidos libres o incorporados. En consecuencia, los métodos existentes y las ideas anteriores son inadecuados para lograr el coste óptimo deseado para la secuenciación.

Sumario

Por consiguiente, en el presente documento se proporciona tecnología para la secuenciación de ácidos nucleicos que utiliza una química de desprotección óptica para controlar la secuenciación gradual de una cadena de ácido nucleico y un medio electroquímico para detectar e identificar cada base en particular. Una ventaja de la tecnología es que el coste y la complejidad de los componentes de detección óptica en un secuenciador se reducen al tiempo que se mantiene el uso ventajoso de la desprotección fotoquímica como el enfoque fundamental para controlar la reacción de secuenciación. En consecuencia, las tecnologías descritas en el presente documento ofrecen un costo más bajo para la secuenciación que las tecnologías convencionales que dependen tanto de la desprotección fotoquímica como de la detección óptica.

En un aspecto, en el presente documento se proporciona una tecnología que comprende el uso de un elemento de detección eléctrica junto con la desprotección fotoactivada de análogos nucleotídicos. Además, la tecnología implementa estos componentes en un enfoque de guía de onda en modo cero para la secuenciación de próxima generación. La combinación de la desprotección eléctrica con la detección eléctrica permite que los sistemas de secuenciación implementen la iluminación global, en lugar de barrido óptico, para desproteger las reacciones de secuenciación masivamente paralelas. Además, el uso de un detector eléctrico en lugar de los costosos elementos de barrido de resolución ultraalta que encuentran uso en los sistemas convencionales de detección óptica existentes reduce la complejidad y el coste del banco de óptica. Como resultado, aumenta la velocidad de detección y disminuye el coste por lectura.

La tecnología encuentra uso en los métodos de secuenciación por síntesis, por ejemplo, según lo provisto por la plataforma Solexa/Illumina (véase, por ejemplo, Voelkerding et al., *Clinical Chem.* 55: 641-658, 2009; MacLean et al., *Nature Rev. Microbiol.* 7: 287-296; y en la patente de Estados Unidos n.º 6.833.246; 7.115.400; y 6.969.488;).

Además, el uso del enfoque de guía de onda en modo cero, por ejemplo, el confinamiento estocástico/óptico de moléculas individuales, y los nucleótidos ópticamente desbloqueables permiten métodos de secuenciación por síntesis de próxima generación utilizando un único fluido de secuenciación en lugar de una serie de etapas de

lavado ineficientes. La secuenciación por síntesis convencional de próxima generación requiere lavar la reacción después de cada adición de base. Estos lavados aumentan el tiempo de secuenciación y el coste de la secuenciación al requerir fluidos más complejos y grandes volúmenes de reactivos costosos. Tal como se proporciona en la tecnología en el presente documento, los nucleótidos ópticamente electrobloqueables, detectables electroquímicamente, se utilizan en un sistema de secuenciación de guía de onda en modo cero, que permite desbloquear el nucleótido en el sitio de la extensión de polimerasa (en el volumen de iluminación de la guía de onda en modo cero) sin desbloquear los otros nucleótidos en la mezcla de reacción. Como resultado, se utiliza un único fluido sin requerir el lavado de la mezcla de reacción después de cada adición de base. En algunas realizaciones, la detección de los nucleótidos incorporados es por medios electroquímicos en lugar de por medios ópticos.

En un aspecto, la tecnología proporciona clases de estructuras químicas (por ejemplo, análogos nucleotídicos) que mejoran la precisión de la secuenciación. En algunas realizaciones, los análogos nucleotídicos se sintetizan utilizando enfoques químicos existentes modificados para incluir un elemento de detección electroquímico, en lugar de fluorescente. En algunas realizaciones, la tecnología proporciona análogos nucleotídicos para la detección electroquímica y la desprotección fotoquímica.

En algunas realizaciones, la tecnología se refiere a una nueva clase química de análogos nucleotídicos fotodesprotegibles que comprenden un resto de terminación detectable electroquímicamente en la base nucleotídica. En algunas realizaciones, la tecnología se refiere a una nueva clase química de análogos nucleotídicos fotodesprotegibles que comprenden un resto de terminación detectable electroquímicamente en el nucleótido en el extremo 3'. En algunas realizaciones, la tecnología se refiere a una nueva clase química de análogos nucleotídicos fotodesprotegibles que comprenden un resto de terminación detectable electroquímicamente o por fluorescencia en la base nucleotídica y un resto terminador en el nucleótido en el extremo 3'. Como tales, los aspectos relacionados proporcionan una tecnología para la secuenciación eliminando secuencialmente el resto terminal detectable electroquímicamente o por fluorescencia de la base nucleotídica y eliminando el resto terminal del extremo 3' del nucleótido, por ejemplo, mediante exposición a una fuente de luz o fuentes que proporcionan una luz (por ejemplo, un fotón o más de un fotón) que escinde los restos de terminación del análogo nucleotídico. A diferencia de las tecnologías convencionales, las realizaciones de la tecnología proporcionan un nucleótido que se desbloquea con la luz y permite la extensión del polímero de ácido nucleico por una polimerasa de ácido nucleico y la detección del nucleótido se realiza por medios electroquímicos en lugar de por medios ópticos.

Por ejemplo, en algunas realizaciones, la secuenciación se realiza mediante la siguiente secuencia de eventos con el uso a modo de ejemplo de un nucleótido que comprende al menos dos restos de terminación fotoquímicos diferentes. Primero, un análogo nucleotídico que comprende un resto de terminación detectable electroquímicamente o por fluorescencia (por ejemplo, unido a la base nucleotídica) y un segundo resto de terminación (por ejemplo, unido al hidroxilo en 3') está orientado en el sitio activo de la polimerasa (por ejemplo, mediante una polimerasa ubicado en el volumen de iluminación de una guía de onda en modo cero) para emparejarse a una base complementaria de la cadena molde y estar adyacente al hidroxilo en 3' libre de la cadena sintetizada en crecimiento. A continuación, el análogo nucleotídico se añade al extremo 3' de una cadena en crecimiento mediante la polimerasa, por ejemplo, mediante el ataque catalizado por enzimas del hidroxilo en 3' en el alfa-fosfato del análogo nucleotídico. La extensión adicional de la cadena por la polimerasa está bloqueada por el grupo de terminación en 3' en el análogo nucleotídico incorporado. El detector electroquímico o fluorescente (por ejemplo, después de la excitación por una fuente de emisión) consulta el resto detectable electroquímicamente o por fluorescencia en el nucleótido incorporado para identificar la base agregada a la cadena sintetizada.

A continuación, el resto de terminación detectable electroquímicamente o por fluorescencia (por ejemplo, unido a la base nucleotídica) se elimina mediante exposición (por ejemplo, en el volumen de iluminación de una guía de onda en modo cero) a una longitud de onda de luz que escinde el resto de terminación detectable electroquímicamente o por fluorescencia del análogo nucleotídico. Aunque se ha liberado el resto de terminación detectable electroquímicamente o por fluorescencia, la extensión adicional de la cadena por la polimerasa permanece bloqueada por el grupo de terminación en 3' (segundo resto de terminación) en el análogo nucleotídico incorporado. A continuación, el segundo resto de terminación se elimina mediante exposición (por ejemplo, en el volumen de iluminación de una guía de onda en modo cero) a una longitud de onda de luz que escinde el segundo resto de terminación del análogo nucleotídico. En algunas realizaciones, la longitud de onda de la luz que escinde el resto de terminación detectable electroquímicamente o por fluorescencia es diferente de la longitud de onda de la luz que escinde el segundo resto de terminación, por ejemplo, bloqueando el hidroxilo en 3'. Después de que se libera el segundo resto de terminación, el hidroxilo en 3' de la cadena en crecimiento queda libre para una polimerización adicional: la siguiente base se incorpora para continuar otro ciclo, por ejemplo, un análogo nucleotídico está orientado en el sitio activo de la polimerasa, se agrega el análogo nucleotídico hasta el extremo 3' de la cadena en crecimiento por la polimerasa, se consulta el análogo nucleotídico para identificar la base agregada, y se desprotege el análogo nucleotídico.

En algunas realizaciones, la secuenciación se realiza mediante la siguiente secuencia de eventos con el uso a modo de ejemplo de un nucleótido que comprende un resto de terminación fotoquímico. Primero, un análogo nucleotídico que comprende un resto de terminación detectable electroquímicamente (por ejemplo, unido a la base nucleotídica) está orientado en el sitio activo de la polimerasa (por ejemplo, por una polimerasa ubicada en el volumen de

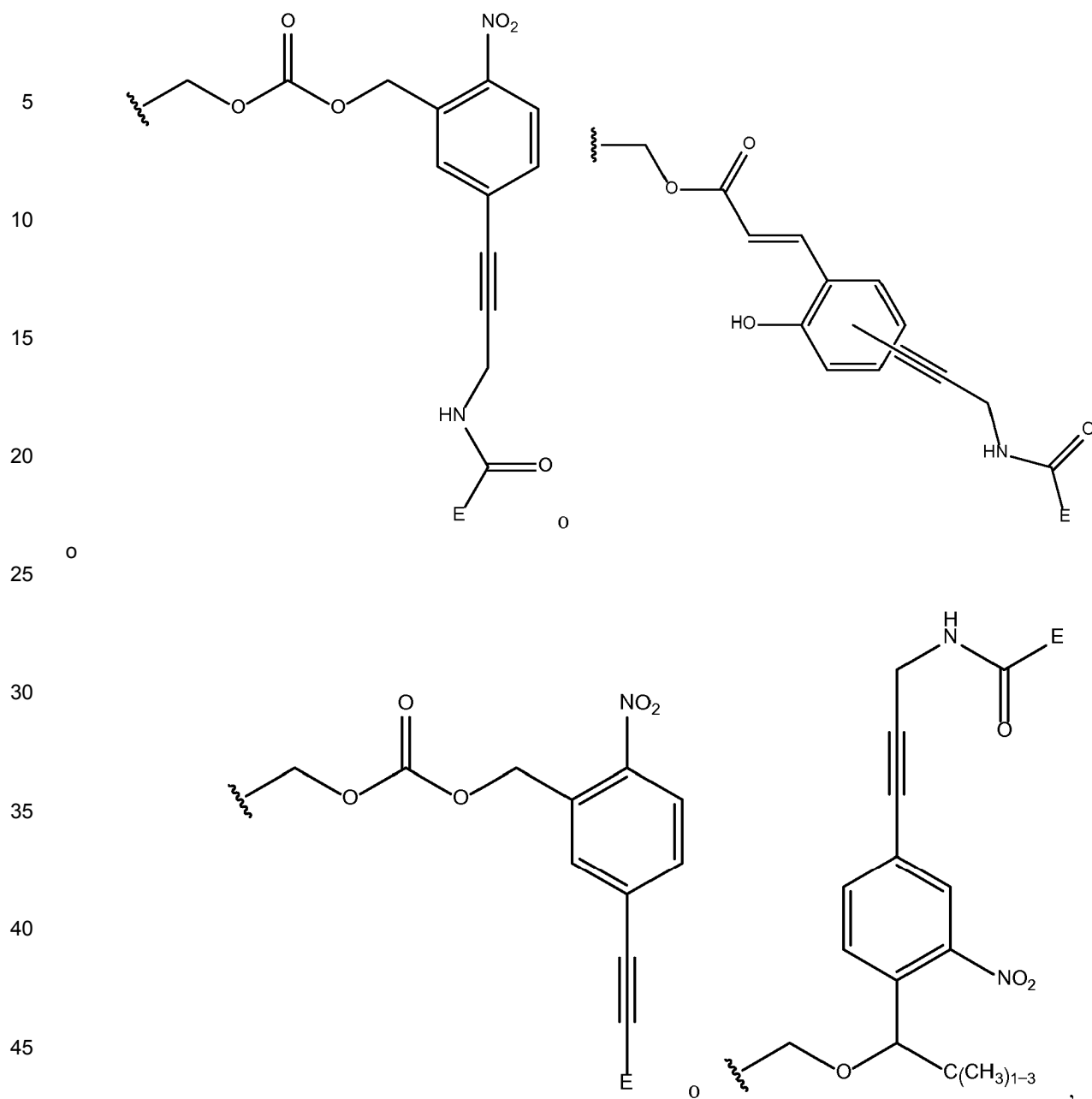
iluminación de una guía de onda en modo cero) para aparearse con una base complementaria de la cadena molde y para ser adyacente al hidroxilo 3' libre de la cadena sintetizada en crecimiento. A continuación, el análogo nucleotídico se añade al extremo 3' de una cadena en crecimiento mediante la polimerasa, por ejemplo, mediante el ataque catalizado por enzimas del hidroxilo en 3' en el alfa-fosfato del análogo nucleotídico. La extensión adicional de la cadena por la polimerasa está bloqueada por el resto de terminación detectable electroquímicamente del análogo nucleotídico incorporado, por ejemplo, por el impedimento estérico de la catálisis, la deformación de los constituyentes del sitio activo, etc. El resto detectable electroquímicamente en el nucleótido incorporado es consultado por un detector electroquímico para identificar la base agregada a la cadena sintetizada.

A continuación, el resto de terminación detectable electroquímicamente (por ejemplo, unido a la base nucleotídica) se elimina mediante exposición (por ejemplo, en el volumen de iluminación de una guía de onda en modo cero) a una longitud de onda de luz que escinde el resto de terminación detectable electroquímicamente del análogo nucleotídico. Después de que se libera el resto de terminación detectable electroquímicamente, el hidroxilo en 3' de la cadena en crecimiento queda libre para una polimerización adicional: la siguiente base se incorpora para continuar otro ciclo, por ejemplo, un análogo nucleotídico se orienta en el sitio activo de la polimerasa, se añade el análogo nucleotídico al extremo 3' de la cadena en crecimiento por la polimerasa, se consulta el análogo nucleotídico para identificar la base añadida y se desprotege el análogo nucleotídico.

Por consiguiente, en algunas realizaciones, la tecnología proporciona que los nucleótidos en el extremo de la cadena están protegidos en 3', por lo que no pueden reaccionar con el trifosfato en 5' de los nuevos nucleótidos que entran en el sitio activo hasta después de desbloquear el extremo 3' (ya sea eliminando un resto de terminación del hidroxilo en 3' o eliminando el impedimento estérico a la polimerización). Por lo tanto, uno puede leer toda o sustancialmente toda la señal de fluorescencia o electroquímica sin los problemas asociados con la reincorporación y la lectura de la siguiente base. En realizaciones relacionadas con los reactivos basados en la detección electroquímica, dichos reactivos se desarrollan para la secuenciación utilizando desprotección fotoquímica utilizando un solo grupo de desprotección fotoquímica y un elemento electroquímicamente activo que sirve para la detección. Este enfoque mejora las tecnologías existentes (por ejemplo, como la proporciona la química de LaserGen) y, por lo tanto, encuentra uso en las aplicaciones de secuenciación por síntesis que tienen mayor precisión y menor coste al incluir estos enfoques en sistemas ópticos de bajo coste, tal como los sistemas basados en guías de onda en modo cero y similares.

Por consiguiente, en el presente documento se proporciona tecnología relacionada con un análogo nucleotídico que comprende un nucleótido que comprende un resto fosfato, una base y un azúcar; y un resto de terminación fotoescindible detectable electroquímicamente, unido al nucleótido. En algunas realizaciones, el análogo nucleotídico comprende un resto de terminación fotoescindible, detectable electroquímicamente, que está unido al resto de fosfato del nucleótido. En algunas realizaciones, el análogo nucleotídico comprende un resto de terminación fotoescindible, detectable electroquímicamente, que está unido a la base del nucleótido. En algunas realizaciones, el análogo nucleotídico comprende un resto de terminación fotoescindible, detectable electroquímicamente, que está unido al azúcar del nucleótido.

Las realizaciones proporcionan que los restos de terminación fotoescindibles son fotoescindibles, por ejemplo, se pueden escindir del análogo nucleotídico tras la exposición a una fuente de luz de la longitud de onda y la intensidad apropiadas para efectuar la escisión. Como tal, en algunas realizaciones, el análogo nucleotídico comprende un resto de terminación fotoescindible, detectable electroquímicamente, que está unido al nucleótido por un enlazador fotoescindible. En algunas realizaciones, el resto de terminación fotoescindible detectable electroquímicamente tiene una estructura que es:

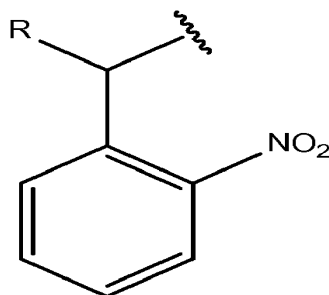


donde E representa un resto o marcador detectable electroquímicamente, por ejemplo, un marcador electroquímico que es una especie fácilmente oxidable o reducible, tal como un grupo redox orgánico, un grupo organometálico, una nanopartícula metálica y un punto cuántico. Los grupos redox orgánicos de ejemplo incluyen, pero no se limitan a, azul de metileno, antraquinona y tionina; los ejemplos de grupos organometálicos incluyen, pero no se limitan a los mismos, ferroceno, derivados de ferroceno, complejos de bipyrideno de rutenio y complejos de bipyrideno de osmio; las nanopartículas metálicas de ejemplo incluyen, pero no se limitan a los mismos, oro y plata; los puntos cuánticos de ejemplo incluyen, pero no se limitan a los mismos, CdS, CdSe y ZnS. En algunas realizaciones, se contempla que los análogos nucleotídicos comprenden una estructura como se proporciona en las patentes de Estados Unidos Nos. 7.893.227; 7.897.737; 7.964.352; y 8.148.503, con el resto fluorescente descrito en las mismas reemplazado por un resto detectable electroquímicamente como se describe en el presente documento o como se conoce en la técnica.

En algunas realizaciones, el análogo nucleotídico comprende un resto de terminación fotoescindible detectable electroquímicamente, que está unido al nucleótido mediante un enlazador fotoescindible seleccionado del grupo que consiste en éter, éster, diéster, etc. La presente tecnología abarca, en algunas realizaciones, enlazadores

fotoescindibles incluidos, pero sin limitación a los mismos, restos 2-nitrobencilo, restos 2-nitrobencilo alfa-sustituidos (por ejemplo, restos 1-(2-nitrofenil)etilo), restos 3,5-dimetoxibencilo, ácido tihidroxámico, restos 7-nitroindolina, restos 9-fenilxantilo, restos benzoínicos, restos hidroxifenacilo y restos ácido N-hidroxi-succinimidil-4-azidosalicílico (NHS-ASA). La presente tecnología también contempla enlazadores fotoescindibles que comprenden restos 2-nitrobencilo y "brazos reticuladores" (o "brazos espaciadores" que separan adicionalmente un enlazador fotoescindible de un nucleótido al que está unido operativamente. Los ejemplos de tales "brazos reticuladores" incluyen, pero no se limitan a los mismos, cadenas alquílicas largas o unidades de repetición de restos de caproilo unidos por enlaces amida.

En algunas realizaciones, el análogo nucleotídico comprende un segundo resto de terminación fotoescindible unido al nucleótido, por ejemplo, unido al azúcar del nucleótido, por ejemplo, en la posición 3' del azúcar. Los ejemplos de restos de terminación fotoescindibles que encuentran uso como un segundo resto de terminación fotoescindible unido al nucleótido incluyen, pero no se limitan a los mismos



La tecnología no está limitada en el resto usado para bloquear el extremo 3'- Por ejemplo, un método de secuenciación que usa análogos nucleotídicos modificados con 3'-O-alilo se describe en Ruparel et al. (2005) Proc. Natl. Acad. Sci. 102: 5932-5937. Otros terminadores reversibles se describen, por ejemplo, en las patentes de Estados Unidos N.º 5.872.244; 6.232.465; 6.214.987; 5.808.045; 5.763.594, y 5.302.509; y la publicación de solicitud de patente de Estados Unidos n.º 20030215862. En algunas realizaciones, un nucleósido 5'-trifosfato modificado en 2' (por ejemplo, 2'-fosfato) encuentra uso en la tecnología, por ejemplo, como se ha descrito como un sustrato para ciertas enzimas de polimerización de ácidos nucleicos (véase, por ejemplo, la publicación de solicitud de patente de Estados Unidos n.º 2005/00373898 y 2005/0037991). Otros bloqueadores en 3' se proporcionan, por ejemplo, en Guo et al. (2008) "Four-color DNA Sequencing with 3'-O-modified Nucleotide Reversible Terminators and Chemically Cleavable Fluorescent Dideoxynucleotides" Proc Natl Acad Sci USA 105: 9145- 9150 y en Ju et al (2006) "Four-Color DNA Sequencing by Synthesis Using Cleavable Fluorescent Nucleotide Reversible Terminators" Proc Natl Acad Sci USA 103: 19635-19640.

En algunas realizaciones, el análogo nucleotídico comprende además un segundo resto de terminación fotoescindible unido al nucleótido, en el que el segundo resto de terminación fotoescindible está unido al nucleótido mediante un enlazador fotoescindible. En algunas realizaciones, el análogo nucleotídico comprende además un segundo resto de terminación fotoescindible unido al nucleótido, en el que el segundo resto de terminación fotoescindible está unido al nucleótido mediante un enlazador fotoescindible. En algunas realizaciones, el análogo nucleotídico comprende además un segundo resto de terminación fotoescindible unido al nucleótido, en el que el segundo resto de terminación fotoescindible está unido al nucleótido mediante un enlazador fotoescindible que es diferente del enlazador fotoescindible unido al resto de terminación fotoescindible detectable electroquímicamente.

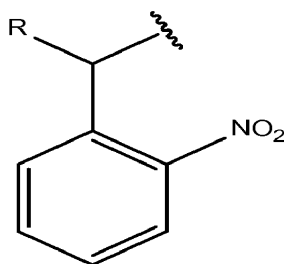
La tecnología no está limitada en los enlazadores fotoescindibles utilizados para unir un resto detectable (por ejemplo, un resto detectable electroquímicamente y/o por fluorescencia). Por ejemplo, la presente tecnología abarca enlazadores fotoescindibles que incluyen, pero sin limitaciones, restos 2-nitrobencilo, restos 2-nitrobencilo alfa-sustituidos (por ejemplo, restos 1-(2-nitrofenil)etilo), restos 3,5-dimetoxibencilo, ácido tihidroxámico, restos 7-nitroindolina, restos 9-fenilxantilo, restos de benzoína, restos de hidroxifenacilo y restos de ácido N-hidroxi-succinimidil-4-azidosalicílico (NHS-ASA). La presente tecnología también contempla enlazadores fotoescindibles que comprenden restos 2-nitrobencilo y "brazos reticuladores" (o "brazos espaciadores" que separan adicionalmente un enlazador fotoescindible de un nucleótido al que está unido operativamente. Los ejemplos de tales "brazos reticuladores" incluyen, pero no se limitan a los mismos, cadenas alquílicas largas o unidades de repetición de restos de caproilo unidos por enlaces amida.

Es preferible que la luz requerida para activar la escisión no afecte a los otros componentes de los nucleótidos modificados. Por ejemplo, si se usa un fluoróforo como marcador, es preferible que absorba la luz de una longitud de onda diferente a la requerida para escindir la molécula enlazadora. Los enlazadores adecuados incluyen aquellos basados en compuestos de O-nitrobencilo y compuestos de nitroveratrilo. También se pueden utilizar enlazadores basados en la química de la benzoína. Véase, por ejemplo, Lee et al. (1999) J. Org. Chem. 64: 3454-3460.

En el presente documento se proporcionan análogos nucleotídicos que comprenden un nucleótido (por ejemplo, que comprenden un resto fosfato, una base y un azúcar); un resto de terminación fotoescindible, detectable electroquímicamente, unido a la base del nucleótido; y un segundo resto de terminación fotoescindible unido al azúcar del nucleótido. Además, también se proporcionan análogos nucleotídicos que comprenden un nucleótido (por ejemplo, que comprende un resto fosfato, una base y un azúcar); un resto de terminación fotoescindible, detectable por fluorescencia, unido al nucleótido; y un segundo resto de terminación fotoescindible unido al nucleótido.

En algunas realizaciones del análogo nucleotídico, el resto de terminación fotoescindible, detectable por fluorescencia, está unido al resto fosfato del nucleótido y el segundo resto de terminación fotoescindible está unido al azúcar del nucleótido. En algunas realizaciones del análogo nucleotídico de la reivindicación, el resto de terminación fotoescindible, detectable por fluorescencia, está unido a la base del nucleótido y el segundo resto de terminación fotoescindible está unido al azúcar del nucleótido.

La tecnología no está limitada en los restos fluorescentes y/o grupos fotoescindibles que están unidos a los análogos nucleotídicos. Por ejemplo, en algunas realizaciones del análogo nucleotídico, el resto de terminación fotoescindible detectable por fluorescencia comprende un resto detectable por fluorescencia que se basa en un colorante, en el que el colorante es xanteno, fluoresceína, rodamina, BODIPY, cianina, cumarina, pirina, ftalocianina, fitocianina, fitocianina, ficobiliproteína, ALEXA FLUOR® 350, ALEXA FLUOR® 405, ALEXA FLUOR® 430, ALEXA FLUOR® 488, ALEXA FLUOR® 514, ALEXA FLUOR® 532, ALEXA FLUOR® 546, ALEXA FLUOR® 558, ALEXA FLUOR® 568, ALEXA FLUOR® 594, ALEXA FLUOR® 610, ALEXA FLUOR® 633, ALEXA FLUOR® 647, ALEXA FLUOR® 660, ALEXA FLUOR® 680, ALEXA FLUOR® 700, ALEXA FLUOR® 750, o un colorante de escuarina; y el segundo resto de terminación fotoescindible es



En algunas realizaciones, el resto de terminación fotoescindible, detectable por fluorescencia, está unido al nucleótido por un primer enlazador fotoescindible y el segundo resto de terminación fotoescindible está unido al nucleótido por un segundo enlazador fotoescindible. En algunas realizaciones, se proporciona un análogo nucleotídico que comprende un nucleótido (por ejemplo, que comprende un resto fosfato, una base y un azúcar); un resto de terminación fotoescindible, detectable por fluorescencia unido a la base del nucleótido; y un segundo resto de terminación fotoescindible unido al azúcar del nucleótido.

La tecnología abarca nucleótidos y bases naturales y sintéticos. Como tales, en algunas realizaciones, el análogo nucleotídico comprende un nucleótido que es adenina, citosina, guanina, timina o uracilo. En algunas realizaciones, el nucleótido comprende un resto fosfato que es un trifosfato, un difosfato o un monofosfato.

La tecnología se refiere a nucleótidos que encuentran uso en la secuenciación por síntesis. De acuerdo con esto, las realizaciones de nucleótidos comprenden un resto de terminación fotoescindible que es un grupo que imparte propiedades de terminación de polimerasa al análogo nucleotídico.

Realizaciones relacionadas proporcionan composiciones que comprenden un análogo nucleotídico y una polimerasa. La tecnología no está limitada en la polimerasa que encuentra uso en la tecnología. Las polimerasas que encuentran uso incluyen, pero no se limitan a los mismos, Taq ADN polimerasa, Klenow (exo⁻) ADN polimerasa, Bst ADN polimerasa, ADN polimerasa VENT® (exo⁻) (por ejemplo, una ADN polimerasa A clonada de *Thermococcus litoralis* y que contiene las mutaciones D141A y E143A), Pfu (exo⁻) ADN polimerasa y ADN polimerasa DEEPVENT™ (exo⁻) (ADN polimerasa A clonada de la especie de *Pyrococcus* GB-D y que contiene las mutaciones D141A y E143A), ADN polimerasa AMPLITAQ®, FS (ADN polimerasa Taq que contiene las mutaciones G46D y F667Y), ADN polimerasa THERMOSEQUENASE™ (ADN polimerasa Taq que contiene la mutación F667Y), ADN polimerasa THERMOSEQUENASE™ II (mezcla de ADN polimerasa THERMOSEQUENASE™ y *T. acidophilum* pirofosfatasa), ADN polimerasa THERMINATOR™ (ADN polimerasa A clonada de la especie de *Thermococcus* 9°N-7 y que contiene las mutaciones D141A, E143A y A485L), THERMINATOR™ II ADN polimerasa (ADN polimerasa THERMINATOR™ que contiene la mutación Y409V adicional) y ADN polimerasa VENT® (exo⁻) A488L (VENT® (exo⁻) que contiene la mutación A488L). En algunas realizaciones, se proporciona una composición que comprende además un ácido nucleico, por ejemplo, un molde que se va a secuenciar (por ejemplo, para tener su secuencia de nucleótidos determinada).

En algunas realizaciones, una polimerasa se modifica para mejorar la incorporación de los análogos nucleotídicos descritos en el presente documento. Las polimerasas modificadas de ejemplo se desvelan en las patentes de Estados Unidos N.º 4.889.818; 5.374.553; 5.420.029; 5.455.170; 5.466.591; 5.618.711; 5.624.833; 5.674.738; 5.789.224; 5.795.762; 5.939.292; y las publicaciones de patente de Estados Unidos n.º 2002/0012970 y 2004/0005599. Un ejemplo no limitante de una polimerasa modificada incluye la ADN polimerasa G46E E678G CS5, la ADN polimerasa G46E E678G CS5, la Taq ADN polimerasa E615G, la polimerasa Δ ZO5R y la ADN polimerasa G46E L329A E678G CS5 descritas en la publicación de Patente de Estados Unidos n.º 2005/0037398. La producción de polimerasas modificadas se puede lograr utilizando muchas técnicas convencionales en biología molecular y ADN recombinante descritas en el presente documento y conocidas en la técnica. En algunas realizaciones, en las que el desbloqueo de 2'-fosfato genera un 2'-hidroxilo, por ejemplo, se pueden usar un ribonucleótido, mutantes de polimerasa, tales como los descritos en la patente de Estados Unidos n.º 5.939.292, que incorporan NTP así como dNTP.

Además, la tecnología proporciona realizaciones de métodos para secuenciar un ácido nucleico, comprendiendo los métodos hibridar un cebador con un ácido nucleico para formar un complejo de cebador/ácido nucleico hibridado; proporcionar una pluralidad de análogos nucleotídicos, comprendiendo los análogos nucleotídicos un nucleótido y un resto de terminación fotoescindible, detectable electroquímicamente, unido al nucleótido; hacer reaccionar el complejo de cebador/ácido nucleico hibridado y el análogo nucleotídico con una polimerasa para agregar el análogo nucleotídico al cebador mediante una reacción de polimerasa para formar un producto extendido que comprende un análogo nucleotídico incorporado; consultar el producto extendido para identificar el análogo nucleotídico incorporado; y exponer el producto extendido a una fuente de luz que proporciona una longitud de onda de luz para eliminar un resto de terminación fotoescindible y detectable electroquímicamente del análogo nucleotídico incorporado. En algunas realizaciones, los métodos comprenden el uso de un análogo nucleotídico que comprende además un segundo resto de terminación fotoescindible unido al nucleótido y el método comprende además exponer el producto extendido a una fuente de luz que proporciona una segunda longitud de onda de luz para eliminar un segundo resto de terminación fotoescindible del análogo nucleotídico incorporado.

Las realizaciones de ejemplo para secuenciar un ácido nucleico comprenden hibridar un cebador con un ácido nucleico para formar un complejo de cebador/ácido nucleico hibridado; proporcionar una pluralidad de análogos nucleotídicos, comprendiendo cada análogo nucleotídico un nucleótido y un resto de terminación fotoescindible detectable por fluorescencia unido al nucleótido y un segundo resto de terminación fotoescindible unido al nucleótido; hacer reaccionar el complejo de cebador/ácido nucleico hibridado y el análogo nucleotídico con una polimerasa para añadir el análogo nucleotídico al cebador mediante una reacción de polimerasa para formar un producto extendido que comprende un análogo nucleotídico incorporado; consultar el producto extendido para identificar el análogo nucleotídico incorporado; exponer el producto extendido a una fuente de luz que proporciona una primera longitud de onda de luz para eliminar un resto de terminación fotoescindible detectable por fluorescencia del análogo nucleotídico incorporado; y exponer el producto extendido a una fuente de luz que proporciona una segunda longitud de onda de luz para eliminar un segundo resto de terminación fotoescindible del análogo nucleotídico incorporado.

Las realizaciones de los métodos proporcionan la determinación de la secuencia de múltiples bases (nucleótidos) de un ácido nucleico. Por consiguiente, se proporcionan realizaciones en las que el método comprende además repetir las etapas de reacción, consulta y exposición en uno o más ciclos adicionales para identificar una pluralidad de bases en el ácido nucleico, en el que la etapa de reacción del ciclo $n + 1$ comprende reaccionar el producto extendido de ciclo n y un análogo nucleotídico con la polimerasa para añadir el análogo nucleotídico al producto extendido de ciclo n por una reacción de polimerasa para formar un producto extendido de ciclo $n + 1$ que comprende un análogo nucleotídico incorporado.

Algunas realizaciones proporcionan el uso de la guía de onda en modo cero en un método de secuenciación. Como tal, en algunas realizaciones, los métodos proporcionan además un sustrato que comprende una guía de onda de modo cero y que localiza la polimerasa dentro de un volumen de observación de la guía de onda en modo cero. Los análogos nucleotídicos proporcionan la identificación de un nucleótido por el resto detectable por fluorescencia o detectable eléctricamente unido al análogo nucleotídico. Por consiguiente, la tecnología proporciona en algunas realizaciones que la etapa de consulta se realiza con un elemento de detección eléctrica, por ejemplo, que comprende el uso de voltimetría cíclica, voltimetría de decaído por adsorción, voltimetría de corriente alterna, voltimetría de pulso diferencial y quimioluminiscencia. En algunas realizaciones, la consulta se realiza mediante un elemento de detección de fluorescencia, por ejemplo, exponiendo un resto fluorescente a una longitud de onda de emisión desde una fuente de luz y detectando una emisión desde el resto fluorescente. En algunas realizaciones, la exposición del producto extendido a la primera longitud de onda de la luz es simultánea a la exposición del producto extendido a la segunda longitud de onda de la luz; en algunas realizaciones, la exposición del producto extendido a la primera longitud de onda de la luz es antes o después de exponer el producto extendido a la segunda longitud de onda de la luz.

En algunas realizaciones, la tecnología proporciona análogos nucleotídicos que comprenden dos restos fotoescindibles y realizaciones relacionadas de métodos en los que exponer el producto extendido a la primera longitud de onda de la luz no elimina el segundo resto de terminación fotoescindible del análogo nucleotídico

incorporado y exponer el producto extendido a la segunda la longitud de onda de la luz no elimina el resto de terminación fotoescindible, detectable electroquímicamente, ni el resto de terminación fotoescindible detectable por fluorescencia del análogo nucleotídico.

- 5 Algunos grupos fotoescindibles se escinden por exposición a un fotón de luz en la longitud de onda apropiada y algunos grupos fotoescindibles se escinden por exposición a dos o más fotones de luz. Como tal, en algunas realizaciones, exponer el producto extendido a una fuente de luz que proporciona una primera longitud de onda de luz para eliminar un resto de terminación fotoescindible, detectable electroquímicamente, o un resto de terminación fotoescindible, detectable por fluorescencia del análogo nucleotídico incorporado comprende exponer el producto extendido a uno el fotón de la primera longitud de onda y la exposición del producto extendido a una fuente de luz que proporciona una segunda longitud de onda de luz para eliminar un segundo resto de terminación fotoescindible del análogo nucleotídico incorporado comprende exponer el producto extendido a más de un fotón de la segunda longitud de onda; como alternativa, algunas realizaciones proporcionan que exponer el producto extendido a una fuente de luz que proporciona una primera longitud de onda de luz para eliminar un resto de terminación fotoescindible, detectable electroquímicamente, o un resto de terminación fotoescindible, detectable por fluorescencia del análogo nucleotídico incorporado comprende exponer el producto extendido a más de un fotón de la primera longitud de onda y la exposición del producto extendido a una fuente de luz que proporciona una segunda longitud de onda de la luz para eliminar un segundo resto de terminación fotoescindible del análogo nucleotídico incorporado comprende exponer el producto extendido a un fotón de la segunda longitud de onda.

- 20 Algunas realizaciones proporcionan que la longitud de onda de la luz utilizada para escindir un resto de terminación fotoescindible es 354 nm, 766 nm o 520 nm. Los métodos de secuenciación proporcionados en el presente documento incorporan análogos naturales y sintéticos de nucleótidos; por ejemplo, las realizaciones proporcionan que el nucleótido sea adenina, citosina, guanina, timina o uracilo. Los métodos se refieren a métodos de secuenciación por síntesis, por ejemplo, los métodos que comprenden el uso de un resto de terminación fotoescindible que es un grupo que imparte propiedades de terminación de la polimerasa al análogo nucleotídico y el método comprende la inhibición de una etapa de reacción cuando un nucleótido incorporado comprende un resto de terminación fotoescindible.

- 30 Por consiguiente, la tecnología también proporciona composiciones para la secuenciación de un ácido nucleico. Por ejemplo, las realizaciones proporcionan una composición para secuenciar un ácido nucleico que comprende cuatro análogos nucleotídicos como se describe en el presente documento, en el que un primer análogo nucleotídico comprende un nucleótido adenina, un segundo análogo nucleotídico comprende un nucleótido citosina, un tercer análogo nucleotídico comprende un nucleótido guanina y un cuarto análogo nucleotídico comprende un nucleótido timina o uracilo. El resto detectable electroquímicamente o por fluorescencia unido a los nucleótidos se usa para identificar el nucleótido que se incorpora a un ácido nucleico durante una reacción de secuenciación por síntesis. Por lo tanto, la tecnología se refiere a una composición para secuenciar un ácido nucleico, comprendiendo la composición cuatro análogos nucleotídicos como se describe en el presente documento, en el que un primer análogo nucleotídico comprende un primer resto detectable electroquímicamente o un primer resto detectable por fluorescencia, un segundo análogo nucleotídico comprende un segundo resto detectable electroquímicamente o un segundo resto detectable por fluorescencia, un tercer análogo nucleotídico comprende un tercer resto detectable electroquímicamente o un tercer resto detectable por fluorescencia, y un cuarto análogo nucleotídico comprende un cuarto resto detectable electroquímicamente o un cuarto restodetectable por fluorescencia.

- 45 En el presente documento también se describen realizaciones de kit de la tecnología. Como tales, las realizaciones del kit para secuenciar un ácido nucleico comprenden cuatro análogos nucleotídicos como se describe en el presente documento, en las que un primer análogo nucleotídico comprende un nucleótido adenina, un segundo análogo nucleotídico comprende un nucleótido citosina, un tercer análogo nucleotídico comprende un nucleótido guanina y un cuarto nucleótido análogo comprende un nucleótido timina o un nucleótido uracilo. Las realizaciones de kit adicionales para secuenciar un ácido nucleico comprenden cuatro análogos nucleotídicos como se describe en el presente documento, en el que un primer análogo nucleotídico comprende un primer resto detectable electroquímicamente o un primer resto detectable por fluorescencia, un segundo análogo nucleotídico comprende un segundo resto detectable electroquímicamente o un segundo resto detectable por fluorescencia, un tercer análogo nucleotídico comprende un tercer resto detectable electroquímicamente o un tercer resto detectable por fluorescencia, y un cuarto análogo nucleotídico comprende un cuarto resto detectable electroquímicamente o un cuarto resto detectable por fluorescencia. Algunas realizaciones de kits comprenden cuatro análogos nucleotídicos como se describe en el presente documento, en el que un primer análogo nucleotídico comprende un nucleótido adenina, un segundo análogo nucleotídico comprende un nucleótido citosina, un tercer análogo nucleotídico comprende un nucleótido guanina y un cuarto análogo nucleotídico comprende un nucleótido timina o un nucleótido uracilo; y una polimerasa. Además, algunas realizaciones de kit comprenden cuatro análogos nucleotídicos como se describe en el presente documento, en el que un primer análogo nucleotídico comprende un primer resto detectable electroquímicamente o un primer resto detectable por fluorescencia, un segundo análogo nucleotídico comprende un segundo resto detectable electroquímicamente o un segundo resto detectable por fluorescencia, un tercer análogo nucleotídico comprende un tercer resto detectable electroquímicamente o un tercer resto detectable por fluorescencia, y un cuarto análogo nucleotídico comprende un cuarto resto detectable electroquímicamente o un cuarto resto detectable por fluorescencia; y una polimerasa.

Los sistemas para secuenciar un ácido nucleico también están abarcados por la tecnología proporcionada en el presente documento. En algunas realizaciones, la tecnología proporciona un sistema para secuenciar un ácido nucleico, comprendiendo el sistema un análogo nucleotídico como se describe en el presente documento, una composición como se describe en el presente documento, o un kit como se describe en el presente documento; una o más fuentes de luz coherente; y un elemento de detección de fluorescencia o electroquímica. Las realizaciones del sistema comprenden además una o más de una polimerasa, una guía de onda en modo cero, un ordenador para identificar bases y recopilar información de secuencias, y/o un componente de software para adquirir y analizar datos de secuencias.

Las realizaciones adicionales serán evidentes para los expertos en la técnica relevante basadas en las enseñanzas contenidas en el presente documento.

Breve descripción de los dibujos

Estas y otras características, aspectos y ventajas de la tecnología actual se entenderán mejor con respecto a los siguientes dibujos:

La figura 1 es un diagrama que muestra una realización del método de la tecnología proporcionada en el presente documento. El dibujo muestra la polimerización de un análogo nucleotídico al hidroxilo (-OH) de una cadena de ADN; la identificación de la base incorporada ("base de consulta") a través de su resto detectable electroquímicamente o detectable por fluorescencia; fotoescisión de un resto de terminación detectable electroquímicamente o un resto de terminación detectable por fluorescencia (ET/FT) mediante una primera longitud de onda de luz ($h\nu_1$); fotoescisión de un segundo resto terminador (XT) por una segunda longitud de onda de la luz ($h\nu_2$); que proporciona un grupo hidroxilo libre en la cadena de ADN en crecimiento para que el ciclo se repita con el siguiente análogo nucleotídico que se incorporará.

Debe entenderse que las figuras no están necesariamente dibujadas a escala, ni tampoco los objetos en las figuras están necesariamente dibujados a escala en relación unos con otros. Las figuras son representaciones que pretenden brindar claridad y comprensión a diversas realizaciones de aparatos, sistemas y métodos desvelados en el presente documento. Siempre que sea posible, se utilizarán los mismos números de referencia en todos los dibujos para hacer referencia a partes iguales o similares. Además, debe apreciarse que los dibujos no pretenden limitar el alcance de las presentes enseñanzas de ninguna manera.

Descripción detallada

En el presente documento se proporciona tecnología relacionada con la secuenciación de ácidos nucleicos, pero no exclusivamente, con composiciones, métodos, sistemas y kits relacionados con nucleótidos que comprenden un resto detectable electroquímicamente y uno o más restos inhibidores de la síntesis fotolábiles.

Los encabezados de sección usados en el presente documento son únicamente con motivos organizativos y no deben interpretarse de ningún modo como limitantes de la materia objeto descrita.

En esta descripción detallada de las diversas realizaciones, a efectos de explicación, se exponen numerosos detalles específicos para proporcionar una comprensión completa de las realizaciones desveladas. Un experto en la materia apreciará, sin embargo, que estas diversas realizaciones se pueden practicar con o sin estos detalles específicos. En otros casos, las estructuras y dispositivos se muestran en forma de diagrama de bloques. Además, un experto en la técnica puede apreciar fácilmente que las secuencias específicas en las que se presentan y realizan los métodos son ilustrativas y se contempla que las secuencias pueden variarse y aún permanecer dentro del alcance de las diversas realizaciones desveladas en el presente documento.

A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que entiende habitualmente un experto en la técnica a la que pertenecen las diversas realizaciones descritas en el presente documento. Cuando las definiciones de términos en las referencias incorporadas parecen diferir de las definiciones proporcionadas en las presentes enseñanzas, prevalecerá la definición provista en las presentes enseñanzas.

Definiciones

Para facilitar la comprensión de la tecnología actual, a continuación se definen una serie de términos y frases. A lo largo de la descripción detallada se exponen definiciones adicionales.

A lo largo de la especificación y las reivindicaciones, los siguientes términos toman los significados explícitamente asociados en el presente documento, a menos que el contexto indique claramente lo contrario. La expresión "en una realización" como se usa en el presente documento no se refiere necesariamente a la misma realización, aunque puede hacerlo. Además, la frase "en otra realización" como se usa en el presente documento no se refiere

necesariamente a una realización diferente, aunque puede ser así. Por lo tanto, como se describe a continuación, varias realizaciones de la invención se pueden combinar fácilmente, sin apartarse del alcance o espíritu de la invención.

Además, como se usa en el presente documento, el término "o" es un operador "o" inclusivo y es equivalente al término "y/o" a menos que el contexto indique claramente lo contrario. La expresión "a base de" no es exclusiva y permite basarse en factores adicionales no descritos, a menos que el contexto indique claramente lo contrario. Además, a lo largo de la especificación, el significado de "un", "uno/a" y "el/la" incluye las referencias en plural. El significado de "en" incluye "en" y "sobre".

Como se usa en el presente documento, un "nucleótido" comprende una "base" (como alternativa, una "nucleobase" o "base nitrogenada"), un "azúcar" (en particular, un azúcar de cinco carbonos, por ejemplo, ribosa o 2-desoxirribosa), y un "resto fosfato" de uno o más grupos fosfato (por ejemplo, un monofosfato, un difosfato o un trifosfato que consiste en uno, dos o tres fosfatos unidos, respectivamente). Sin el resto fosfato, la nucleobase y el azúcar forman un "nucleósido". Por lo tanto, un nucleótido también se puede llamar un monofosfato de nucleósido o un difosfato de nucleósido o un trifosfato de nucleósido, dependiendo del número de grupos de fosfato unidos. El resto de fosfato se une generalmente al carbono 5 del azúcar, aunque algunos nucleótidos comprenden restos de fosfato unidos al carbono 2 o al carbono 3 del azúcar. Los nucleótidos contienen una purina (en los nucleótidos adenina y guanina) o una base pirimidínica (en los nucleótidos citosina, timina y uracilo). Los ribonucleótidos son nucleótidos en los que el azúcar es ribosa. Los desoxirribonucleótidos son nucleótidos en los que el azúcar es desoxirribosa.

Como se usa en el presente documento, un "ácido nucleico" significará cualquier molécula de ácido nucleico, incluyendo, sin limitación, ADN, ARN e híbridos de los mismos. Las bases de ácido nucleico que forman moléculas de ácido nucleico pueden ser las bases A, C, G, T y U, así como sus derivados. Los derivados de estas bases son bien conocidos en la técnica. Se debe entender que el término incluye, como equivalentes, análogos de ADN o ARN hechos de análogos nucleotídicos. El término como se usa en el presente documento también abarca el ADNc, que es complementario, o copia, ADN producido a partir de un molde de ARN, por ejemplo, mediante la acción de una transcriptasa inversa. Es bien sabido que el ADN (ácido desoxirribonucleico) es una cadena de nucleótidos que consiste en 4 tipos de nucleótidos, A (adenina), T (timina), C (citosina) y G (guanina), y ese ARN (ácido ribonucleico) es una cadena de nucleótidos que consiste en 4 tipos de nucleótidos A, U (uracilo), G y C. También se sabe que todos estos 5 tipos de nucleótidos se unen específicamente entre sí en combinaciones llamadas pares de bases complementarias. Es decir, la adenina (A) se aparea con timina (T) (en el caso del ARN, sin embargo, la adenina (A) se aparea con uracilo (U)) y la citosina (C) se aparea con guanina (G), de modo que cada uno de estos pares de bases forman una doble cadena. Como se usa en el presente documento, "datos de secuenciación de ácido nucleico", "información de secuenciación de ácido nucleico", "secuencia de ácido nucleico", "secuencia genómica", "secuencia genética", "secuencia de fragmentos" o "lectura de secuenciación de ácido nucleico" denota cualquier información o datos que son indicativos del orden de las bases nucleotídicas (por ejemplo, adenina, guanina, citosina y timina/uracilo) en una molécula (por ejemplo, un genoma completo, un transcriptoma completo, un exoma, oligonucleótido, polinucleótido, fragmento, etc.) de ADN o ARN.

Es bien sabido que el ADN (ácido desoxirribonucleico) es una cadena de nucleótidos que consiste en 4 tipos de nucleótidos; A (adenina), T (timina), C (citosina) y G (guanina), y ese ARN (ácido ribonucleico) está compuesto por 4 tipos de nucleótidos; A, U (uracilo), G y C. También se sabe que todos estos 5 tipos de nucleótidos se unen específicamente entre sí en combinaciones llamadas pares de bases complementarias. Es decir, la adenina (A) se aparea con timina (T) (en el caso del ARN, sin embargo, la adenina (A) se aparea con uracilo (U)) y la citosina (C) se aparea con guanina (G), de modo que cada uno de estos pares de bases forman una doble cadena. Como se usa en el presente documento, "datos de secuenciación de ácido nucleico", "información de secuenciación de ácido nucleico", "secuencia de ácido nucleico", "secuencia genómica", "secuencia genética", "secuencia de fragmentos" o "lectura de secuenciación de ácido nucleico" denota cualquier información o datos que son indicativos del orden de las bases nucleotídicas (por ejemplo, adenina, guanina, citosina y timina/uracilo) en una molécula (por ejemplo, genoma completo, transcriptoma completo, exoma, oligonucleótido, polinucleótido, fragmento, etc.) de ADN o ARN. Debe entenderse que las presentes enseñanzas contemplan la información de secuencia obtenida utilizando todas las variedades disponibles de técnicas, plataformas o tecnologías, entre las que se incluyen: electroforesis capilar, micromatrices, sistemas basados en ligado, sistemas basados en la polimerasa, sistemas basados en la hibridación, sistemas de identificación directa o indirecta de nucleótidos, pirosecuenciación, sistemas de detección basados en iones o pH, sistemas basados en firmas electrónicas, etc.

La referencia a una base, un nucleótido u otra molécula puede estar en singular o en plural. Es decir, "una base" puede hacer referencia a una sola molécula de esa base o a una pluralidad de la base, por ejemplo, en una solución.

Un "polinucleótido", "ácido nucleico" u "oligonucleótido" se refiere a un polímero lineal de nucleósidos (incluidos desoxirribonucleósidos, ribonucleósidos o análogos de los mismos) unidos por enlaces internucleosídicos. Típicamente, un polinucleótido comprende al menos tres nucleósidos. Por lo general, los oligonucleótidos varían en tamaño desde unas pocas unidades monoméricas, por ejemplo, de 3-4 a varios cientos de unidades monoméricas. Cuando un polinucleótido como un oligonucleótido esté representado por una secuencia de letras, tal como

"ATGCCTG", se entenderá que los nucleótidos están en el orden 5' -> 3' de izquierda a derecha y que "A" denota desoxiadenosina, "C" denota desoxicitidina, "G" denota desoxiguanosina y "T" denota timidina a menos que se indique lo contrario. Las letras A, C, G y T pueden usarse para hacer referencia a las propias bases, a los nucleósidos o a los nucleótidos que comprenden las bases, como es estándar en la técnica.

5 Como se usa en el presente documento, la frase "dNTP" significa desoxinucleótido-trifosfato, en el que el nucleótido comprende una base nucleotídica, tal como A, T, C, G o U.

10 El término "monómero" como se usa en el presente documento significa cualquier compuesto que pueda incorporarse en una cadena molecular en crecimiento por una polimerasa dada. Dichos monómeros incluyen, sin limitaciones, los nucleótidos naturales (por ejemplo, ATP, GTP, TTP, UTP, CTP, dATP, dGTP, dTTP, dUTP, dCTP, análogos sintéticos), precursores para cada nucleótido, nucleótidos de origen no natural y sus precursores o cualquier otra molécula que pueda incorporarse a una cadena polimérica en crecimiento por una polimerasa dada.

15 Como se usa en el presente documento, "complementario" generalmente se refiere a la duplicación de nucleótidos específicos para formar pares de bases canónicos de Watson-Crick, como entienden los expertos en la técnica. Sin embargo, el complementario también incluye apareamiento de bases de análogos nucleotídicos que son capaces de aparearse bases universales con los nucleótidos A, T, G o C y ácidos nucleicos bloqueados que mejoran la estabilidad térmica de los dúplex. Un experto en la materia reconocerá que la rigurosidad de la hibridación es un factor
20 determinante en el grado de apareamiento o apareamiento erróneo en el dúplex formado por la hibridación.

Como se usa en el presente documento, "resto" se refiere a una de dos o más partes en las que se puede dividir algo, como, por ejemplo, las diversas partes de un hilo, una molécula o una sonda.

25 La frase "resto detectable electroquímicamente", como se usa en el presente documento, se refiere a una sustancia que puede aceptar o donar al menos un electrón durante una reacción electroquímica, típicamente oxidación y/o reducción (redox). Cada evento electroquímico, es decir, una transferencia de electrones a o desde el resto detectable electroquímicamente, contribuye a una corriente o voltaje eléctrico que el sistema puede detectar y
30 registrar.

Como se usa en el presente documento, un "enlazador" es una molécula o resto que une dos moléculas o restos y proporciona un espaciado entre las dos moléculas o restos de manera que puedan funcionar de la manera deseada. Por ejemplo, un enlazador puede comprender una cadena de hidrocarburo de diamina que está unida covalentemente a través de un grupo reactivo en un extremo a una molécula análoga de oligonucleótido y a través
35 de un grupo reactivo en otro extremo a un soporte sólido, tal como, por ejemplo, una superficie de perlas. El acoplamiento de enlazadores a nucleótidos y construcciones de sustrato de interés se puede lograr mediante el uso de reactivos de acoplamiento que son conocidos en la técnica (véase, por ejemplo, Efimov et al., Nucleic Acids Res. 27: 4416-4426, 1999). Los métodos de derivación y acoplamiento de moléculas orgánicas son bien conocidos en las técnicas de química orgánica y bioorgánica. Un enlazador también puede ser escindible (por ejemplo, foto-
40 escindible) o reversible.

Tal como se usa en el presente documento, el término "enlazador fotoescindible" se refiere a un enlazador que se puede eliminar de un nucleótido, polinucleótido, grupo químico o ácido nucleico, al cual está unido o unido operativamente, por exposición a radiación electromagnética (por ejemplo, luz visible, luz ultravioleta, etc.). La
45 longitud de onda de la luz utilizada para fotoescindir el enlazador depende de la estructura del enlazador fotoescindible

Una "polimerasa" es una enzima generalmente para unir nucleótidos 3'-OH 5'-trifosfato, oligómeros y sus análogos. Las polimerasas incluyen, pero no se limitan a, ADN polimerasas dependientes de ADN, ARN polimerasas dependientes de ADN, ADN polimerasas dependientes de ARN, ARN polimerasas dependientes de ARN, ADN polimerasa de T7, ADN polimerasa de T3, ADN polimerasa de T4, ARN polimerasa de T7, ARN polimerasa de T3, ARN polimerasa de SP6, ADN polimerasa 1, fragmento Klenow, ADN polimerasa de Thermophilus aquaticus, Tth
50 ADN polimerasa, ADN polimerasa Vent (New England Biolabs), ADN polimerasa Deep Vent (New England Biolabs), fragmento grande de ADN polimerasa Bst, fragmento de Stoffel, ADN Polimerasa de 9 ° N, ADN polimerasa de pfu, ADN polimerasa de Tfl, polimerasa RepliPhi Phi29, ADN polimerasa Tli, ADN polimerasa beta eucariótica, telomerasa, polimerasa Terminator (New England Biolabs), KOD HiFi. ADN polimerasa (Novagen), ADN polimerasa KOD1, Q-beta replicasa, transferasa terminal, transcriptasa inversa AMV, transcriptasa inversa M-MLV, transcriptasa inversa Phi6, transcriptasa inversa VIH-1, nuevas polimerasas descubiertas mediante bioprospección y polimerasas citadas en la publicación de solicitud de patente de Estados Unidos n.º 2007/0048748 y en las patentes de Estados
55 Unidos n.º 6.329.178; 6.602.695 y 6.395.524. Estas polimerasas incluyen isoformas mutantes de tipo salvaje y variantes genéticamente modificadas, tales como exo-polimerasas y otras mutantes, por ejemplo, que toleran los nucleótidos marcados y los incorporan en una cadena de ácido nucleico.

El término "cebador" se refiere a un oligonucleótido, ya sea de origen natural como en un producto de digestión de restricción purificado o que se produzca sintéticamente, que sea capaz de actuar como un punto de inicio de la
65 síntesis cuando se coloque en condiciones en las que se induzca la síntesis de un producto de extensión del

cebador que sea complementario a una cadena de ácido nucleico (por ejemplo, en presencia de nucleótidos y un agente inductor tal como ADN polimerasa y a una temperatura y pH adecuados). Preferentemente, el cebador es monocatenario para una máxima eficiencia en la amplificación, pero, como alternativa, puede ser bicatenario. Si es bicatenario, el cebador se trata primero para separar las cadenas antes de usar para preparar productos de extensión. Preferentemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser lo bastante largo como para cebar la síntesis de los productos de extensión en presencia del agente inductor. Las longitudes exactas de los cebadores dependerán de muchos factores, incluidos la temperatura, la fuente del cebador y el uso del procedimiento.

Como se usa en el presente documento, un "sistema" denota un conjunto de componentes, reales o abstractos, que comprende un todo en el que cada componente interacciona con o está relacionado con al menos otro componente dentro del todo.

Se describen varias plataformas de secuenciación de ácidos nucleicos, ensamblaje de ácidos nucleicos y/o sistemas de mapeo (por ejemplo, software de y/o hardware de ordenador), por ejemplo, en la publicación de solicitud de patente de Estados Unidos n.º 2011/0270533.

Como se usa en el presente documento, los términos "alquilo" y el prefijo "alq" incluyen tanto los grupos saturados o insaturados de cadena lineal como los de cadena ramificada y los grupos cíclicos, por ejemplo, grupos cicloalquilo y cicloalqueno. A menos que se especifique lo contrario, los grupos alquilo acíclicos son de 1 a 6 carbonos. Los grupos cíclicos pueden ser monocíclicos o policíclicos y, preferentemente, tienen de 3 a 8 átomos de carbono en el anillo. Los grupos cíclicos de ejemplo incluyen grupos ciclopentilo, ciclopentilo, ciclohexilo y adamantilo. Los grupos alquilo pueden estar sustituidos con uno o más sustituyentes o no sustituidos. Los sustituyentes de ejemplo incluyen alcoxi, arilo, sulfhidrido, alquiltio, ariltio, halógeno, alquilsililo, hidroxilo, fluoroalquilo, perfluoroalquilo, amino, aminoalquilo, amino disustituido, amino cuaternario, hidroxialquilo, carboxialquilo y carboxilo. Cuando se usa el prefijo "alq", el número de carbonos contenidos en la cadena alquilo viene dado por el intervalo que precede directamente a este término, con el número de carbonos contenidos en el resto del grupo que incluye este prefijo definido en otra parte en el presente documento. Por ejemplo, el término "alcarilo C₁-C₄" ilustra un grupo arilo de 6 a 18 carbonos (por ejemplo, véase más adelante) unido a un grupo alquilo de 1 a 4 carbonos.

Como se usa en el presente documento, el término "arilo" se refiere a un anillo o sistema de anillo aromático carbocíclico. A menos que se especifique lo contrario, los grupos arilo son de 6 a 18 carbonos. Los ejemplos de grupos arilo incluyen grupos fenilo, naftilo, bifenilo, fluorenilo e indenilo.

Como se usa en el presente documento, el término "heteroarilo" se refiere a un anillo aromático o sistema de anillo que contiene al menos un heteroátomo de anillo (por ejemplo, O, S, Se, N o P). A menos que se especifique lo contrario, los grupos heteroarilo son de 1 a 9 carbonos. Los grupos heteroarilo incluyen grupos furanilo, tienilo, pirrolilo, imidazolilo, pirazolilo, oxazolilo, isoxazolilo, tiazolilo, isotiazolilo, triazolilo, tetrazolilo, oxadiazolilo, oxatriazolilo, piridilo, piridazilo, pirimidilo, pirazilo, triazolo, benzofuranilo, isobenzofuranilo, benzotienilo, indol, indazolilo, indolizililo, benzisoxazolilo, quinolinilo, isoquinolinilo, cinnolinilo, quinazolinilo, naftirililo, ftalazinilo, fenantrolinilo, purinilo y carbazolilo.

Como se usa en el presente documento, el término "heterociclo" se refiere a un anillo o sistema de anillo no aromático que contiene al menos un heteroátomo de anillo (por ejemplo, O, S, Se, N o P). A menos que se especifique lo contrario, los grupos heterocíclicos son de 2 a 9 carbonos. Los grupos heterocíclicos incluyen, por ejemplo, grupos dihidropirrolilo, tetrahidropirrolilo, piperazinilo, piranilo, dihidropiranilo, tetrahidropiranilo, dihidrofuranilo, tetrahidrofuranilo, dihidrotiofeno, tetrahidrotiofeno y morfolinilo.

Los grupos arilo, heteroarilo o heterocíclicos pueden estar sin sustituir o sustituidos con uno o más sustituyentes seleccionados del grupo que consiste en alquilo C₁₋₆, hidroxi, halo, nitro, alcoxi C₁₋₆, alquiltio C₁₋₆, trifluorometilo, acilo C₁₋₆, arilcarbonilo, heteroarilcarbonilo, nitrilo, alcocarbonilo C₁₋₆, alcarilo (donde el grupo alquilo tiene de 1 a 4 átomos de carbono), y alcheteroarilo (donde el grupo alquilo tiene de 1 a 4 átomos de carbono).

Como se usa en el presente documento, el término "alcoxi" se refiere a un sustituyente químico de la fórmula -OR, en la que R es un grupo alquilo. Por "arilo" se entiende un sustituyente químico de la fórmula -OR', en la que R' es un grupo arilo.

Como se usa en el presente documento, el término "alcarilo C_x-C_y" se refiere a un sustituyente químico de fórmula -RR', en la que R es un grupo alquilo de carbonos x a y y R' es un grupo arilo como se define en otra parte en el presente documento.

Como se usa en el presente documento, el término "alcheteroarilo C_x-C_y" se refiere a un sustituyente químico de fórmula -RR', en la que R es un grupo alquilo de x a y carbonos, y R' es un grupo heteroarilo como se define en otra parte en el presente documento.

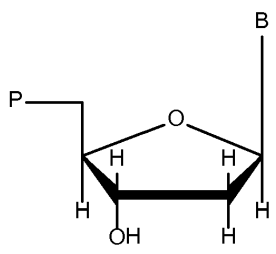
Como se usa en el presente documento, el término "O, S o N no vecinal" se refiere a un sustituyente heteroátomo de oxígeno, azufre o nitrógeno en un enlace, en el que el sustituyente heteroátomo no forma un enlace a un carbono saturado que está enlazado a otro heteroátomo.

5 Realizaciones de la tecnología

La tecnología descrita en el presente documento se refiere a análogos nucleotídicos y métodos relacionados, composiciones (por ejemplo, mezclas de reacción), kits y sistemas para la secuenciación de ácidos nucleicos. En particular, algunas realizaciones de los análogos nucleotídicos comprenden un terminador fotoescindible para controlar la secuenciación gradual de una cadena de ácido nucleico y un resto detectable electroquímicamente para la detección e identificación de cada base particular, por ejemplo, para determinar una secuencia de ácido nucleico. La tecnología ofrece ventajas sobre los métodos convencionales, tal como un menor coste y una menor complejidad.

Análogos nucleotídicos

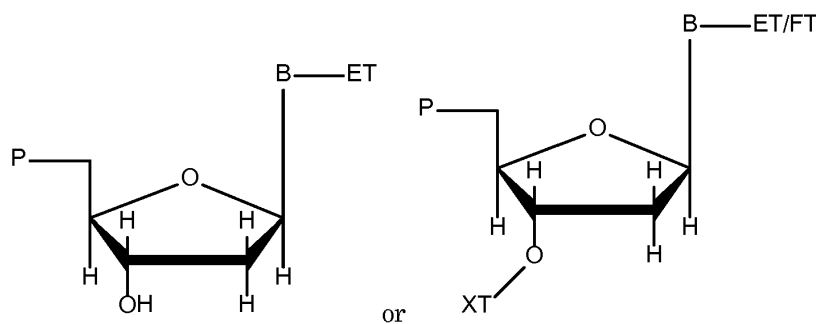
En el presente documento se proporcionan análogos nucleotídicos. Los análogos nucleotídicos comprenden uno o más restos de terminación fotoescindibles, por ejemplo, que comprenden además, en algunas realizaciones, un resto detectable por fluorescencia o un resto detectable electroquímicamente. Por ejemplo, la tecnología proporciona un nucleótido tal como:



en el que P es el resto fosfato (por ejemplo, un monofosfato, un difosfato o un trifosfato) como se define en el presente documento y B es una base, por ejemplo, adenina, timina, citosina, guanina o uracilo.

Los análogos nucleotídicos no están limitados a un grupo fosfato específico. En una realización, el grupo fosfato es un grupo monofosfato o un polifosfato, tal como un grupo difosfato o un grupo trifosfato. En algunas realizaciones, el grupo fosfato es un pirofosfato. Además, la base de los análogos nucleotídicos no se limita a una base específica. En algunas realizaciones, la base es una adenina, citosina, guanina, timina, uracilo y análogos de las mismas, tales como, por ejemplo, bases acíclicas. Los análogos nucleotídicos no están limitados a un resto de azúcar específico. En algunas realizaciones, dicho resto de azúcar es una ribosa, desoxirribosa, didesoxirribosa y sus análogos.

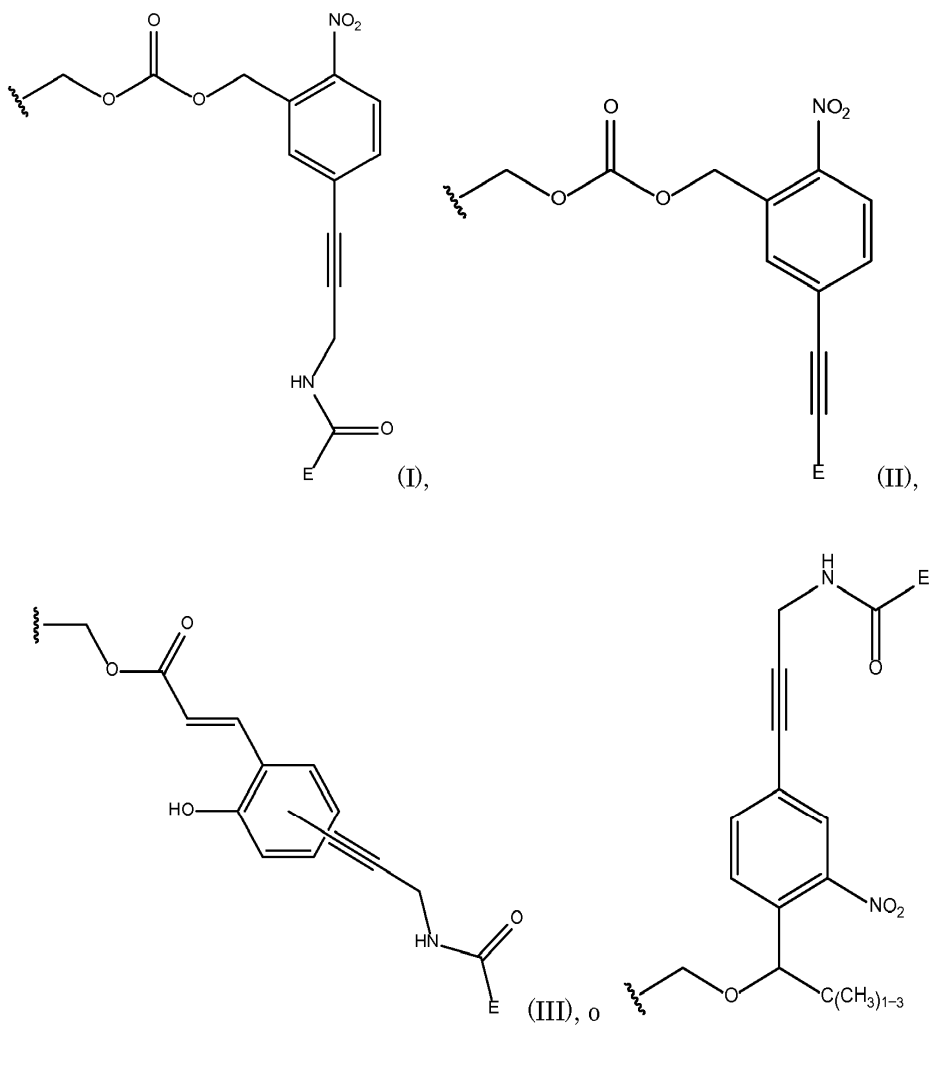
En algunas realizaciones, el análogo nucleotídico tiene una estructura que es:



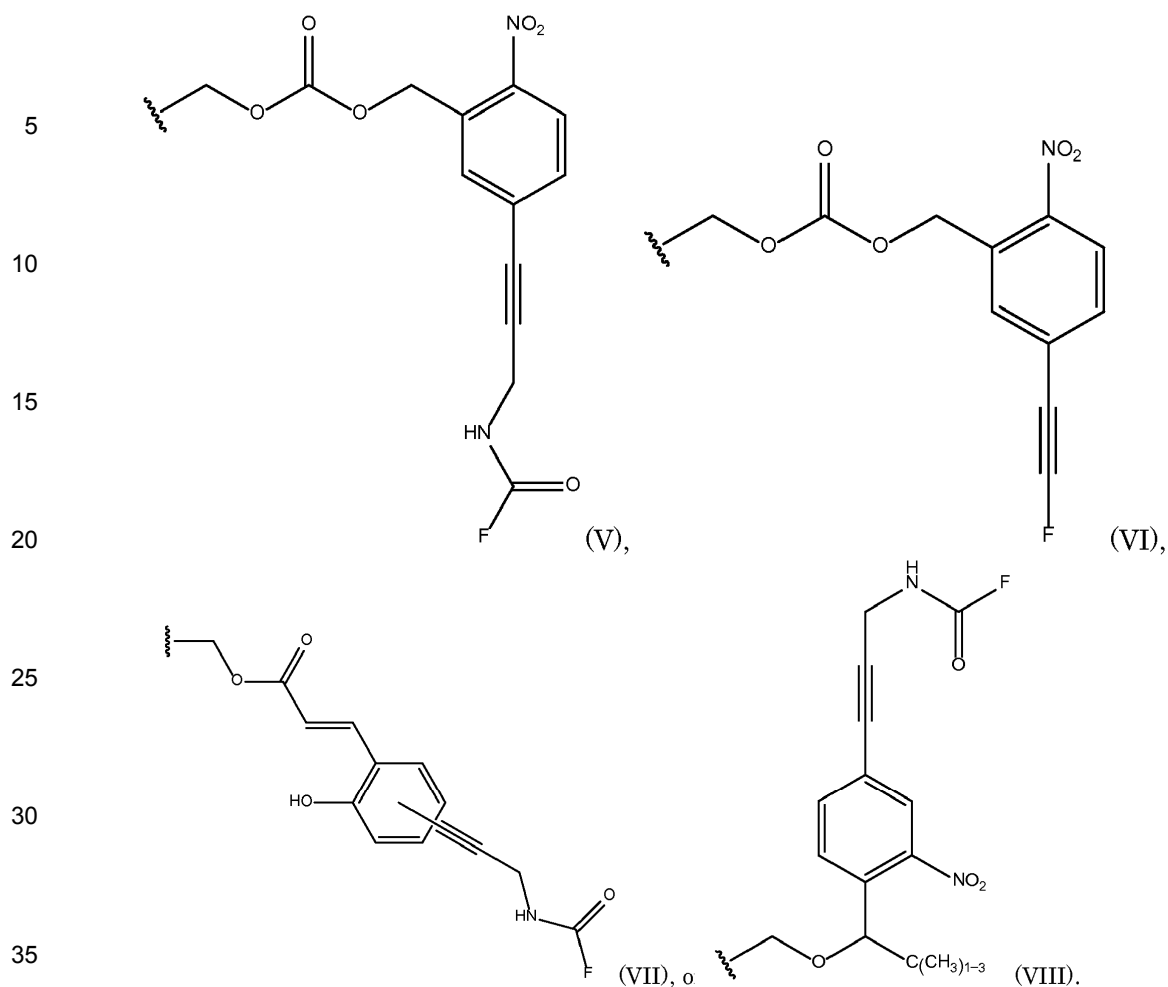
en la que ET es un resto terminador detectable electroquímicamente (por ejemplo, que comprende un resto E detectable electroquímicamente como se describe en el presente documento o como se conoce en la técnica), ET/FT es un resto terminador detectable por fluorescencia (por ejemplo, que comprende un resto detectable por fluorescencia como se describe en el presente documento, por ejemplo, "F" en las estructuras proporcionadas en el presente documento) o un resto terminador detectable electroquímicamente (por ejemplo, que comprende un resto detectable electroquímicamente como se describe en el presente documento, por ejemplo, "E" en las estructuras proporcionadas en el presente documento, o como se conoce en la técnica), y XT es un segundo resto de terminación (por ejemplo, como se proporciona en el presente documento y/o como se conoce en la técnica). En

algunas realizaciones, se proporcionan cuatro análogos nucleotídicos, cada uno de los cuales comprende un ET o FT diferente y el mismo XT.

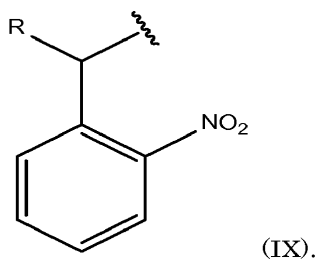
Por ejemplo, en algunas realizaciones ET es



En algunas realizaciones, FT es

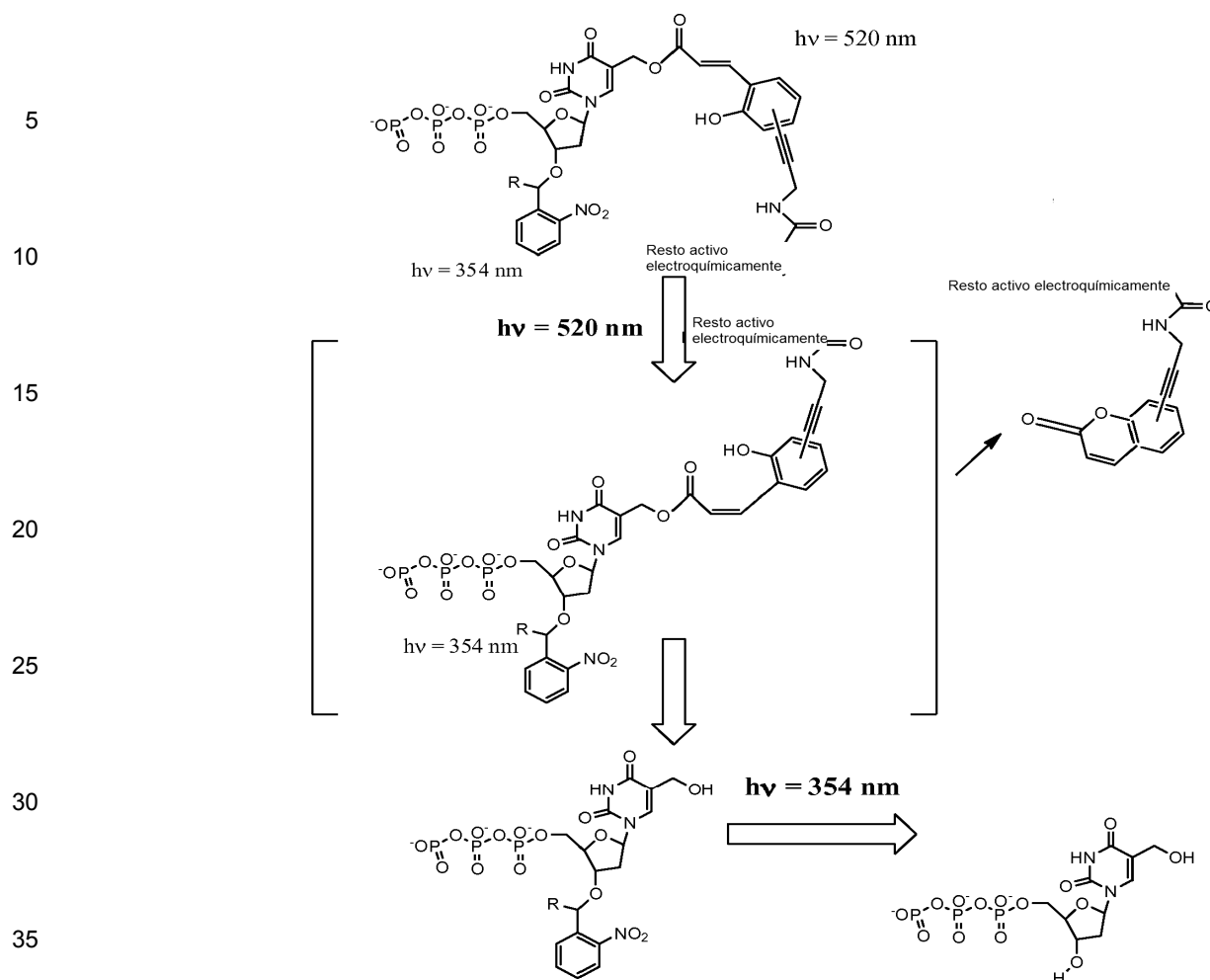


En algunas realizaciones, XT es



La síntesis de compuestos que se proporciona en el presente documento se realiza como se describe, por ejemplo, en las patentes de Estados Unidos N.º 7.893.227; 7.897.737; 7.964.352; y 8.148.503, con las modificaciones necesarias para reemplazar los restos fluorescentes con restos detectables electroquímicamente para proporcionar los diversos análogos nucleotídicos descritos en el presente documento. En el siguiente ejemplo se proporcionan esquemas sintéticos adicionales.

En algunas realizaciones, los terminadores III y VII se escinden por un fotón de luz que tiene una longitud de onda de 500-550 nm (por ejemplo, 520 nm). En algunas realizaciones, el terminador XT se escinde por un fotón de luz que tiene una longitud de onda de 325-375 nm (por ejemplo, 354 nm). En algunas realizaciones, los terminadores II y VI son escindidos por fotones de luz (por ejemplo, dos o más fotones) que tienen una longitud de onda de 740-780 nm (por ejemplo, 766 nm). Ejemplos de los productos de escisión son:



En algunas realizaciones, la tecnología permite desproteger el extremo 3' y la base de forma selectiva y/o secuencial en el mismo sistema mediante el ajuste fino de las longitudes de onda utilizadas para desproteger, por ejemplo, para que un enlace pueda romperse con un color de luz antes de que el otro enlace se rompa por el otro color de la luz. El color de la luz se usa como un "interruptor" específico que rompe los enlaces (por ejemplo, para liberar los terminadores) pero sin impactar o dañar el enlace no deseado. Esto permite el control y ofrece diversos beneficios descritos en el presente documento. En algunas realizaciones, los análogos nucleotídicos comprenden un elemento de detección eléctrica, por ejemplo, como se indica con los nuevos compuestos proporcionados en el presente documento y para compuestos mejorados que comprenden un marcador fluorescente en uso convencional. Los compuestos mejorados eliminan o minimizan las características fluorescentes indeseables y los inconvenientes asociados sufridos por esa técnica anterior.

En este esquema anterior, una vez que se ha consultado al resto electroquímico para identificar la base que se ha incorporado, se retira de la base mediante el procedimiento de fotoescisión y se difunde fuera de la ubicación de la reacción de secuenciación (por ejemplo, una reacción de secuenciación en un espacio estocásticamente confinado, tal como en una guía de onda de modo cero o nanoestructura similar). La cadena en crecimiento es libre de comenzar el ciclo nuevamente e identificar la siguiente base en la secuencia.

En realizaciones similares, un análogo nucleotídico como los descritos en, por ejemplo, las patentes de Estados Unidos N.º 7.893.227; 7.897.737; 7.964.352; y 8.148.503, y en Litosh et al (2010) "Improved nucleotide selectivity and termination of 3'-OH unblocked reversible terminators by molecular tuning of 2-nitrobenzyl alkylated HOMedU triphosphates" Nucleic Acids Res. 39 (6): e39, y otros, se modifican (por ejemplo, un análogo nucleotídico terminador en el que un resto detectable por fluorescencia se reemplaza por un resto detectable electroquímicamente) y se usa sin el enfoque de doble marcaje fotoquímico. Por consiguiente, debe entenderse que la tecnología también proporciona una mejora de los análogos nucleotídicos existentes, tales como los compuestos desprotegibles fotoquímicos en general y no se limita al uso en el modo de longitud de onda dual como se ha descrito anteriormente.

Guías de onda de modo cero

En algunos ensayos, las moléculas están confinadas en una serie, matriz u otra disposición de pequeños orificios, poros o pocillos, por ejemplo, una guía de onda de modo cero (ZMW), por ejemplo, como se describe en la publicación de solicitud de patente de Estados Unidos n.º 2011/0117637. Las matrices de ZMW se han aplicado a diversos análisis bioquímicos y han encontrado una utilidad particular para el análisis genético. Los ZMW comprenden típicamente un núcleo, pocillo o abertura a nanoescala dispuestos en una capa de revestimiento opaco que se dispone sobre un sustrato transparente, por ejemplo, un orificio circular en una película de revestimiento de aluminio depositada sobre un sustrato de sílice transparente. Véase, por ejemplo, J. Koriach et al, "Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures", 105 PNAS 1176–81 (2008). Un agujero ZMW típico tiene un diámetro de ~70 nm y una profundidad de ~100 nm. La tecnología ZMW permite el análisis sensible de moléculas individuales porque, a medida que la luz viaja a través de una pequeña abertura, el campo óptico decae exponencialmente dentro de la cámara. Es decir, debido a las estrechas dimensiones del pocillo, se evitará que la radiación electromagnética que tenga una frecuencia por encima de una frecuencia de corte particular se propague por todo el núcleo. A pesar de lo anterior, la radiación penetrará una distancia limitada en el núcleo, proporcionando un volumen iluminado muy pequeño dentro del núcleo. Al iluminar un volumen muy pequeño, se pueden interrogar cantidades muy pequeñas de reactivos, que incluyen, por ejemplo, reacciones de una sola molécula. El volumen de observación dentro de un ZMW iluminado es de ~20 zeptolitros (20×10^{-21} litros). Dentro de este volumen, la actividad de la ADN polimerasa que incorpora un solo nucleótido puede detectarse fácilmente.

Al monitorizar las reacciones a nivel de una sola molécula, uno puede identificar y/o monitorizar con precisión una reacción dada. En particular, en algunas realizaciones, la tecnología de guía de onda en modo cero es la base para un campo de secuenciación de ADN de molécula única que controla la síntesis de molécula por molécula (por ejemplo, nucleótido por nucleótido) de una cadena de ADN de una manera dependiente del molde mediante una única enzima polimerasa (por ejemplo, secuenciación del ADN en tiempo real de una sola molécula (SMRT), como la realizada por un secuenciador RS de Pacific Biosciences (Pacific Biosciences, Menlo Park, CA)). Véanse, por ejemplo, las patentes de Estados Unidos n.º 7.476.503; 7.486.865; 7.907.800; 7.170.050; 8.795.961; 8.501.405; 9.063.156 y 8.153.375. Véanse también, Eid, J. et al. 2009. "Real-time DNA sequencing from single polymerase molecules", 323 Science: 133–38 (2009); Koriach, J. et al. "Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides", 27 Nucleosides, Nucleotides & Nucleic Acids: 1072–82 (2008); Lundquist, P. M. et al., "Parallel confocal detection of single molecules in real time", 33 Optics Letters: 1026–28 (2008); Koriach, J. et al., "Selective aluminum passivation for targeted immobilization of single dna polymerase molecules in zero-mode waveguide nanostructures", 105 Proc Natl Acad Sci USA: 1176–81 (2008); Foquet, M. et al., "Improved fabrication of zero-mode waveguides for single-molecule detection", 103 Journal of Applied Physics (2008); and Levene, M. J. et al. "Zero-mode waveguides for single-molecule analysis at high concentrations", 299 Science: 682–86 (2003).

En algunas tecnologías, la colocación de componentes en los pocillos de ZMW se basa en la difusión simple para entregar componentes (por ejemplo, macromoléculas tales como ADN polimerasa y/o ADN y/o complejos de ADN/ADN polimerasa) en el sitio deseado (por ejemplo, la parte inferior del pocillo de ZMW) en las guías de onda de modo cero. Algunas tecnologías colocan componentes en los pocillos de un ZMW utilizando tecnología basada en el transporte activo de componentes del ensayo (por ejemplo, una macromolécula como un ADN, una ADN polimerasa, un complejo ADN/ADN polimerasa, una proteína, etc.) para obtener un sitio para un ensayo (por ejemplo, el fondo de un pocillo de ZMW). Las variaciones particulares de las tecnologías de transporte activo y liberación utilizan filamentos de actina o microtúbulos que se unen al fondo de una guía de onda de modo cero. Los filamentos o microtúbulos de actina sirven como guías de transporte para las macromoléculas (por ejemplo, la ADN polimerasa o el complejo de ADN polimerasa/ADN). Véase, por ejemplo, la publicación internacional número WO–2013/102081.

En ciertas realizaciones, la tecnología encuentra uso para los métodos de secuenciación de ADN que utilizan guías de onda de modo cero (ZMW), por ejemplo, según lo desarrollado por Pacific Biosciences o métodos similares. En algunas realizaciones de esta tecnología, la secuenciación de ADN se realiza en un chip, cada uno de los cuales contiene miles de guías de onda de modo cero (ZMW). En algunas realizaciones, un ZMW es un agujero, de decenas de nanómetros de diámetro, fabricado en una película metálica de 100 nm depositada sobre un sustrato de dióxido de silicio. Cada ZMW se convierte en una cámara de visualización nanofotónica que proporciona un volumen de detección de solo 20 zeptolitros (10^{-21} l). En este volumen, la actividad de una sola molécula puede detectarse entre un fondo de miles de nucleótidos marcados. La ZMW proporciona una ventana para observar la ADN polimerasa a medida que realiza la secuenciación por síntesis. Dentro de cada cámara, una única molécula de ADN polimerasa está unida a la superficie inferior de manera que reside permanentemente dentro del volumen de detección. Los análogos nucleotídicos como se proporcionan en el presente documento, cada tipo (por ejemplo, A, T, C y G) marcados con un fluoróforo de color diferente o un grupo diferente detectable electroquímicamente, se introducen en la solución de reacción a altas concentraciones que promueven la velocidad, precisión y capacidad de procesamiento de la enzima. Debido al pequeño tamaño de la ZMW, incluso a estas altas concentraciones biológicamente relevantes, el volumen de detección está ocupado por los nucleótidos solo una pequeña fracción del

tiempo. Además, las visitas al volumen de detección son rápidas y duran solo unos pocos microsegundos, debido a la pequeña distancia que la difusión tiene para transportar los nucleótidos. El resultado es un fondo muy bajo.

Restos fluorescentes

En algunas realizaciones, el resto detectable es un colorante fluorogénico. Se conocen varias clases de colorantes fluorogénicos y compuestos específicos que son apropiados para realizaciones particulares de la tecnología: derivados de xanteno, tales como fluoresceína, rodamina, verde de Oregón, eosina y rojo Texas; derivados de cianina, tales como cianina, indocarbocianina, oxacarbocianina, tiacarbocianina y merocianina; derivados de naftaleno (derivados de dansilo y prodan); derivados de cumarina; derivados de oxadiazol tales como piridiloxazol, nitrobenzoxadiazol y benzoxadiazol; derivados de pireno tales como azul en cascada; derivados de oxazina tales como rojo Nilo, azul Nilo, violeta de cresilo y oxazina 170; derivados de acridina tales como proflavina, naranja de acridina y amarillo de acridina; derivados de arilmetina tales como auramina, cristal violeta y verde malaquita; y derivados de tetrapírol tales como porfina, ftalocianina, bilirrubina. En algunas realizaciones, el resto fluorescente es un colorante que es xanteno, fluoresceína, rodamina, BODIPY, cianina, cumarina, pireno, ftalocianina, ficobiliproteína, ALEXA FLUOR® 350, ALEXA FLUOR® 405, ALEXA FLUOR® 430, ALEXA FLUOR® 488, ALEXA FLUOR® 514, ALEXA FLUOR® 532, ALEXA FLUOR® 546, ALEXA FLUOR® 555, ALEXA FLUOR® 568, ALEXA FLUOR® 568, ALEXA FLUOR® 594, ALEXA FLUOR® 610, ALEXA FLUOR® 633, ALEXA FLUOR® 647, ALEXA FLUOR® 660, ALEXA FLUOR® 680, ALEXA FLUOR® 700, ALEXA FLUOR® 750, o un colorante de escuaraina.

Además, los ejemplos no limitantes de fluoróforos incluyen colorantes que pueden sintetizarse u obtenerse comercialmente (por ejemplo, Operon Biotechnologies, Huntsville, Alabama). Un gran número de colorantes (más de 50) están disponibles para su aplicación en aplicaciones de excitación por fluorescencia. Estos colorantes incluyen los de las familias de colorantes de fluoresceína, rodamina Alexa Fluor, Bodipy, cumarina y cianina. Los ejemplos específicos de fluoróforos incluyen, pero no se limitan a los mismos, FAM, TET, HEX, Cy3, TMR, ROX, rojo Texas, rojo LC 640, Cy5 y rojo LC 705. En algunas realizaciones, los colorantes con máximos de emisión de 410 nm (por ejemplo, Cascade Blue) a 775 nm (por ejemplo, Alexa Fluor 750) están disponibles y se pueden usar. Por supuesto, un experto en la técnica reconocerá que también se pueden usar colorantes que tienen máximos de emisión fuera de estos intervalos. En algunos casos, los colorantes que oscilan entre 500 nm y 700 nm tienen la ventaja de estar en el espectro visible y pueden detectarse utilizando tubos fotomultiplicadores convencionales. En algunas realizaciones, la amplia gama de colorantes disponibles permite la selección de conjuntos de colorantes que tienen longitudes de onda de emisión que se extienden a través del rango de detección. Los sistemas de detección capaces de distinguir muchos colorantes son conocidos en la técnica.

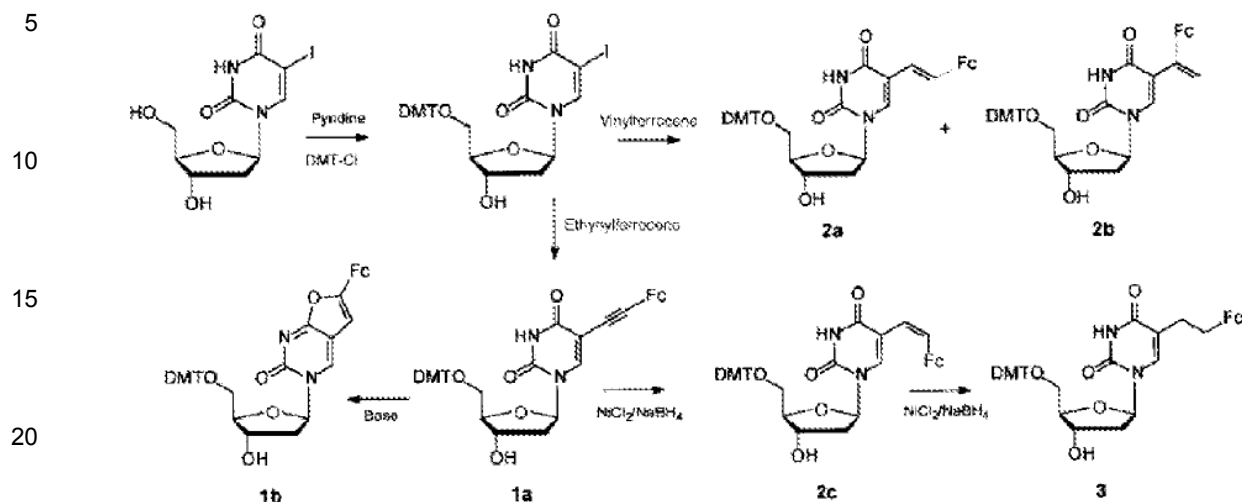
Restos detectables electroquímicamente

En algunas realizaciones, los análogos nucleotídicos proporcionados comprenden un resto electroquímico que es una especie fácilmente oxidable o reducible. El resto electroquímicamente activo puede ser un número de moléculas y/o restos directamente electroquímicamente activos, mediadores electroquímicos y/u otros marcadores tales como enzimas u otras moléculas que inducen o influyen en la generación de señales electrónicas o electroquímicas. Cuando se une a un análogo nucleotídico, el resto electroquímicamente activo (por ejemplo, detectable electroquímicamente) identifica la base nucleotídica incorporada en el extremo 3' de la cadena en crecimiento que se está sintetizando y, por lo tanto, permite determinar la secuencia de la cadena de ADN o ARN que se está secuenciando. Los restos detectables electroquímicamente abarcados por la tecnología incluyen, entre otros, un grupo redox orgánico, un grupo organometálico, una nanopartícula metálica y un punto cuántico. Los grupos redox orgánicos de ejemplo incluyen, pero no se limitan a, azul de metileno, antraquinona y tionina; los ejemplos de grupos organometálicos incluyen, pero no se limitan a los mismos, ferroceno, derivados de ferroceno, complejos de bipyrideno de rutenio y complejos de bipyrideno de osmio; las nanopartículas metálicas de ejemplo incluyen, pero no se limitan a los mismos, oro y plata; los puntos cuánticos de ejemplo incluyen, pero no se limitan a los mismos, CdS, CdSe y ZnS. En algunas realizaciones, se contempla que los análogos nucleotídicos comprenden una estructura como se proporciona en las patentes de Estados Unidos Nos. 7.893.227; 7.897.737; 7.964.352; y 8.148.503, con el resto fluorescente descrito en las mismas reemplazado por un resto detectable electroquímicamente como se describe en el presente documento o como se conoce en la técnica.

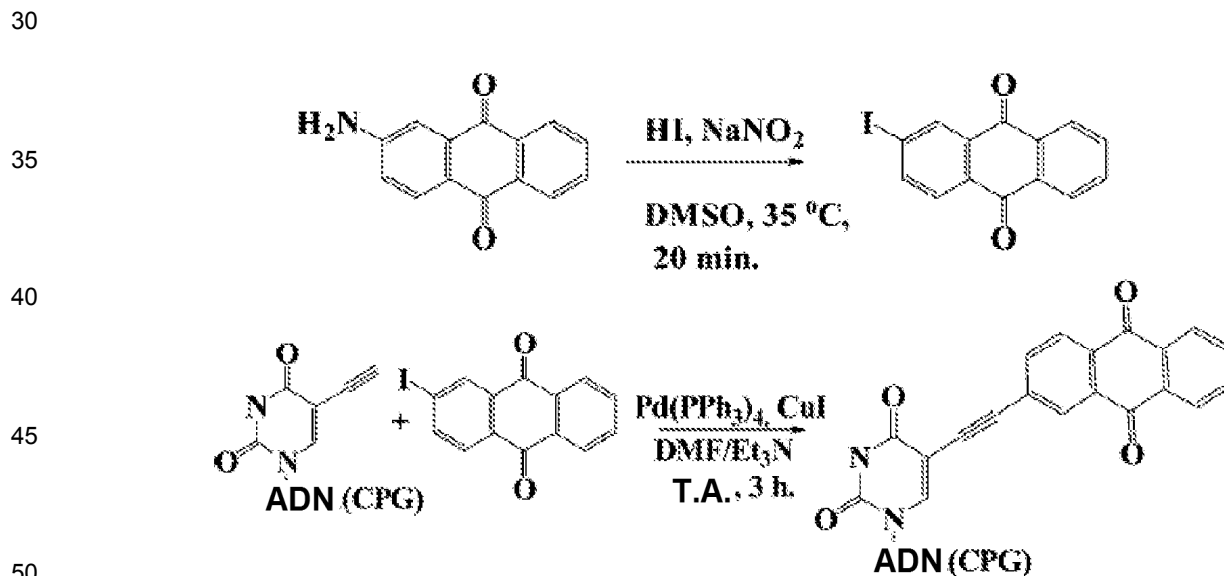
En algunas realizaciones, el resto detectable electroquímicamente comprende un metal de transición, tal como rutenio, osmio, hierro, rodio o cobre. Los ácidos nucleicos modificados que comprenden estos metales se describen, por ejemplo, en *Angew. Chem., Int. Ed. Engl.* 1995, 34, 352–354; Meade, T. J. *Metal Ions in Biological Systems*; Sigel, A., Sigel, H., Eds.; Marcel Dekker: New York, 1996; pág. 32, 453–478; *Helv. Chim. Acta* 1997, 80, 640–652; *J. Am. Chem. Soc.* 1998, 120, 2194–2195; *J. Am. Chem. Soc.* 2000, 122, 6287–6288; *Inorg. Chem.* 1999, 38, 174–189; *Chem. Commun.* 1997, 1609–1610; *Nucleic Acids Res.* 1996, 24, 4273–4280; *Chem. Commun.* 1996, 555–557; *J. Am. Chem. Soc.* 1997, 119, 5045–5046; *Nature* 1996, 382, 731–735; and *J. Am. Chem. Soc.* 1994, 116, 5981–5982.

Un método sintético de ejemplo para proporcionar un resto detectable electroquímicamente unido a un nucleótido es una reacción para unir un grupo ferroceno al carbono C-5 de timidina, por ejemplo, como se describe en Pike et al. (2002) "Metallocene – DNA: Synthesis, Molecular and Electronic Structure and DNA Incorporation of C5–

Ferrocenylthymidine Derivatives" Chem. Eur. J. 8(13): 2891-2899. Se pueden sintetizar diversos derivados de ferrocenil timidina de la siguiente manera:



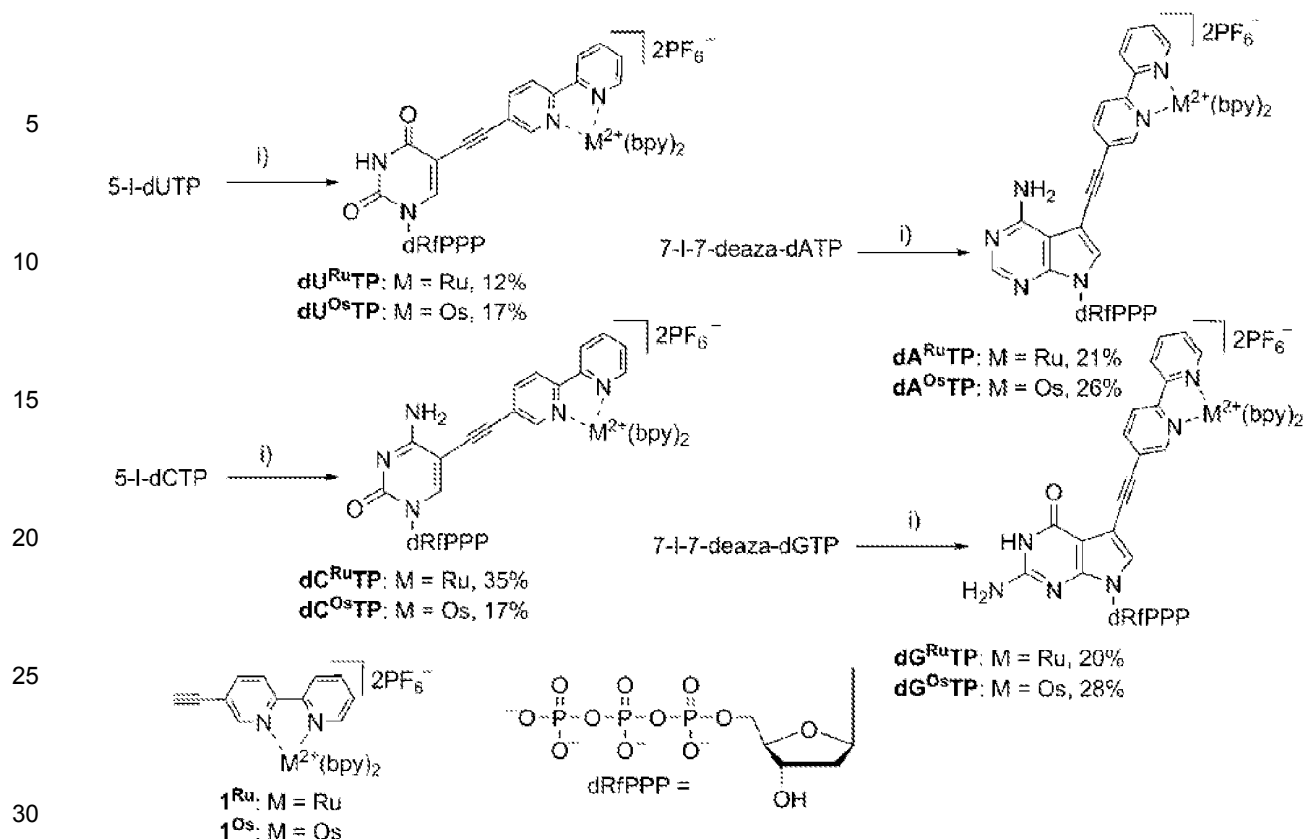
25 Otro método sintético de ejemplo para proporcionar un resto detectable electroquímicamente unido a un nucleótido es una reacción para unir una base modificada con etiniluracilo a 2-yodoantraquinona, por ejemplo, como se describe en Gorodetsky et al. (2007) "Coupling into the Base Pair Stack is Necessary for DNA-Mediated Electrochemistry" Bioconjugate Chem.18: 1434-1441.



55 Otro método sintético de ejemplo para proporcionar un resto detectable electroquímicamente unido a un nucleótido es una reacción para unir complejos de bipyridina de rutenio, osmio u otros metales de transición a bases de nucleótidos usando intermedios de análogos nucleotídicos yodados, por ejemplo, como se describe en Vrabel et al. (2009) "Base-Modified DNA Labeled by [Ru(bpy)3]2+ and [Os(bpy)3]2+ Complexes" Chem. Eur. J. 15: 1144-1154.

60

65



En este esquema, 1^{Ru} o 1^{Os} , $\text{Pd}(\text{OAc})_2$, TPPTS, CuI, Et_3N , $\text{H}_2\text{O}/\text{CH}_3\text{CN}$ 2:1, 1 hora, 70 °C.

Otro método sintético de ejemplo para proporcionar un resto detectable electroquímicamente unido a un análogo nucleotídico es proporcionado por la patente de Estados Unidos N.º 7.893.227; 7.897.737; 7.964.352; y 8.148.503, y en Litosh et al (2010) "Improved nucleotide selectivity and termination of 3'-OH unblocked reversible terminators by molecular tuning of 2-nitrobenzyl alkylated HOMedU triphosphates" Nucleic Acids Res. 39 (6): e39, en el que un resto detectable electroquímicamente se sustituye por un resto detectable por fluorescencia. Los intermediarios químicos se fabrican utilizando síntesis conocidas en la técnica, tal como se describe en la patente de Estados Unidos N.º 7.893.227; 7.897.737; 7.964.352; y 8.148.503; en Litosh et al, citado anteriormente; y según lo proporcionado por los esquemas sintéticos descritos en el presente documento. Se utilizan esquemas sintéticos similares para unir grupos redox a otras partes del nucleótido, tal como el azúcar o grupo trifosfato.

Otro ejemplo de nucleótido que comprende un resto detectable electroquímicamente es, por ejemplo, un nucleótido que comprende un grupo ferrocenilo unido a la posición 2' del anillo de ribosa. Véase, por ejemplo, Yu (2001) "2'-Ribose-Ferrocene Oligonucleotides for Electronic Detection of Nucleic Acids" J. Org. Chem. 66: 2937-2942.

Métodos de secuenciación

La tecnología se refiere, en algunas realizaciones, a métodos para secuenciar un ácido nucleico. En algunas realizaciones, la secuenciación se realiza mediante la siguiente secuencia de eventos con el uso de ejemplo de un análogo nucleotídico que comprende al menos dos restos de terminación fotoquímicos diferentes. Primero, un análogo nucleotídico que comprende un resto de terminación detectable electroquímicamente o por fluorescencia (por ejemplo, unido a la base nucleotídica) y un segundo resto de terminación (por ejemplo, unido al hidroxilo en 3') está orientado en el sitio activo de la polimerasa (por ejemplo, mediante una polimerasa ubicado en el volumen de iluminación de una guía de onda en modo cero) para emparejarse a una base complementaria de la cadena molde y estar adyacente al hidroxilo en 3' libre de la cadena sintetizada en crecimiento. A continuación, el análogo nucleotídico se añade al extremo 3' de una cadena en crecimiento mediante la polimerasa, por ejemplo, mediante el ataque catalizado por enzimas del hidroxilo en 3' en el alfa-fosfato del análogo nucleotídico. La extensión adicional de la cadena por la polimerasa está bloqueada por el grupo de terminación en 3' en el análogo nucleotídico incorporado. El detector electroquímico o fluorescente (por ejemplo, después de la excitación por una fuente de

emisión) consulta el resto detectable electroquímicamente o por fluorescencia en el nucleótido incorporado para identificar la base agregada a la cadena sintetizada.

A continuación, el resto de terminación detectable electroquímicamente o por fluorescencia (por ejemplo, unido a la base nucleotídica) se elimina mediante exposición (por ejemplo, en el volumen de iluminación de una guía de onda en modo cero) a una longitud de onda de luz que escinde el resto de terminación detectable electroquímicamente o por fluorescencia del análogo nucleotídico. Aunque se ha liberado el resto de terminación detectable electroquímicamente o por fluorescencia, la extensión adicional de la cadena por la polimerasa permanece bloqueada por el grupo de terminación en 3' (segundo resto de terminación) en el análogo nucleotídico incorporado.

A continuación, el segundo resto de terminación se elimina mediante exposición (por ejemplo, en el volumen de iluminación de una guía de onda en modo cero) a una longitud de onda de luz que escinde el segundo resto de terminación del análogo nucleotídico. En algunas realizaciones, la longitud de onda de la luz que escinde el resto de terminación detectable electroquímicamente o por fluorescencia es diferente de la longitud de onda de la luz que escinde el segundo resto de terminación, por ejemplo, bloqueando el hidroxilo en 3'. Después de que se libera el segundo resto de terminación, el hidroxilo en 3' de la cadena en crecimiento queda libre para una polimerización adicional: la siguiente base se incorpora para continuar otro ciclo, por ejemplo, un análogo nucleotídico se orienta en el sitio activo de la polimerasa, se añade el análogo nucleotídico al extremo 3' de la cadena en crecimiento por la polimerasa, se consulta el análogo nucleotídico para identificar la base añadida y se desprotege el análogo nucleotídico.

Algunas realizaciones se refieren a la secuenciación paralela (por ejemplo, masivamente paralela). Por ejemplo, en algunas realizaciones, todas las reacciones de secuenciación se exponen a la luz en paralelo para desproteger los análogos nucleotídicos en paralelo. La identificación de la base incorporada se realiza mediante detección electroquímica del resto detectable electroquímicamente (por ejemplo, que proporciona una "firma" de cada base) antes de que se escinda.

En algunas realizaciones, la secuenciación se realiza mediante la siguiente secuencia de eventos con el uso a modo de ejemplo de un nucleótido que comprende un resto de terminación fotoquímico. Primero, un análogo nucleotídico que comprende un resto de terminación detectable electroquímicamente (por ejemplo, unido a la base nucleotídica) está orientado en el sitio activo de la polimerasa (por ejemplo, por una polimerasa ubicada en el volumen de iluminación de una guía de onda en modo cero) para aparearse con una base complementaria de la cadena molde y para ser adyacente al hidroxilo 3' libre de la cadena sintetizada en crecimiento. A continuación, el análogo nucleotídico se añade al extremo 3' de una cadena en crecimiento mediante la polimerasa, por ejemplo, mediante el ataque catalizado por enzimas del hidroxilo en 3' en el alfa-fosfato del análogo nucleotídico. La extensión adicional de la cadena por la polimerasa está bloqueada por el resto de terminación detectable electroquímicamente del análogo nucleotídico incorporado, por ejemplo, por el impedimento estérico de la catálisis, la deformación de los constituyentes del sitio activo, etc. El resto detectable electroquímicamente en el nucleótido incorporado es consultado por un detector electroquímico para identificar la base agregada a la cadena sintetizada.

A continuación, el resto de terminación detectable electroquímicamente (por ejemplo, unido a la base nucleotídica) se elimina mediante exposición (por ejemplo, en el volumen de iluminación de una guía de onda en modo cero) a una longitud de onda de luz que escinde el resto de terminación detectable electroquímicamente del análogo nucleotídico. Después de que se libera el resto de terminación detectable electroquímicamente, el hidroxilo en 3' de la cadena en crecimiento queda libre para una polimerización adicional: la siguiente base se incorpora para continuar otro ciclo, por ejemplo, un análogo nucleotídico se orienta en el sitio activo de la polimerasa, se añade el análogo nucleotídico al extremo 3' de la cadena en crecimiento por la polimerasa, se consulta el análogo nucleotídico para identificar la base añadida y se desprotege el análogo nucleotídico.

Por consiguiente, en algunas realizaciones, la tecnología proporciona que los nucleótidos en el extremo de la cadena están protegidos en 3', por lo que no pueden reaccionar con el trifosfato en 5' de los nuevos nucleótidos que entran en el sitio activo hasta después de desbloquear el extremo 3' (ya sea eliminando un resto de terminación del hidroxilo en 3' o eliminando el impedimento estérico a la polimerización). Por lo tanto, uno puede leer toda o sustancialmente toda la señal de fluorescencia o electroquímica sin los problemas asociados con la reincorporación y la lectura de la siguiente base. En realizaciones relacionadas con los reactivos basados en la detección electroquímica, dichos reactivos se desarrollan para la secuenciación utilizando desprotección fotoquímica utilizando un solo grupo de desprotección fotoquímica y un elemento electroquímicamente activo que sirve para la detección. Este enfoque mejora las tecnologías existentes (por ejemplo, como lo proporciona la química de LaserGen) y, por lo tanto, encuentra uso en las aplicaciones de secuenciación por síntesis que tienen mayor precisión y menor coste al incluir estos enfoques en sistemas ópticos de bajo coste, tal como los sistemas basados en guías de onda en modo cero y similares.

Plataformas de secuenciación de ácidos nucleicos

En algunas realizaciones de la tecnología, se generan datos de secuencia de ácido nucleico. Varias realizaciones de plataformas de secuenciación de ácidos nucleicos (por ejemplo, un secuenciador de ácidos nucleicos) incluyen componentes como se describe a continuación. Según diversas realizaciones, un instrumento de secuenciación

incluye una unidad de control y liberación de fluidos, una unidad de procesamiento de muestras, una unidad de detección de señales y una unidad de adquisición, análisis y control de datos. Varias realizaciones del instrumento proporcionan una secuenciación automatizada que se usa para recopilar información de secuencia de una pluralidad de secuencias en paralelo y/o sustancialmente de forma de simultánea.

5 En algunas realizaciones, la unidad de administración y control de fluidos incluye un sistema de administración de reactivos. El sistema de liberación de reactivos incluye un reservorio de reactivos para el almacenamiento de diversos reactivos. Los reactivos pueden incluir cebadores basados en ARN, cebadores de ADN directos/inversos, mezclas de nucleótidos (por ejemplo, composiciones que comprenden análogos nucleotídicos como se proporcionan
10 en el presente documento) para secuenciación por síntesis, tampones, reactivos de lavado, reactivos de bloqueo, reactivos de separación y similares. Además, el sistema de liberación de reactivos puede incluir un sistema de pipeteado o un sistema de flujo continuo que conecta la unidad de procesamiento de muestras con el depósito de reactivos.

15 En algunas realizaciones, la unidad de procesamiento de muestras incluye una cámara de muestras, tal como una celda de flujo, un sustrato, una micromatriz, una bandeja de múltiples pocillos o similares. La unidad de procesamiento de muestras puede incluir varios carriles, varios canales, varios pocillos u otros medios para procesar múltiples conjuntos de muestras de manera sustancialmente simultánea. Además, la unidad de procesamiento de muestras puede incluir múltiples cámaras de muestras para permitir el procesamiento de múltiples ciclos
20 simultáneamente. En realizaciones particulares, el sistema puede realizar detección de señal en una cámara de muestras mientras procesa sustancialmente simultáneamente otra cámara de muestras. Además, la unidad de procesamiento de muestras puede incluir un sistema de automatización para mover o manipular la cámara de muestras. En algunas realizaciones, la unidad de detección de señal puede incluir un sensor de imagen o detección. Por ejemplo, el sensor de imágenes o detección (por ejemplo, un detector de fluorescencia o un detector eléctrico)
25 puede incluir un CCD, un CMOS, un sensor de iones, tal como una capa sensible a los iones que recubre un CMOS, un detector de corriente, o similar. La unidad de detección de señales puede incluir un sistema de excitación para hacer que una sonda, como un tinte fluorescente, emita una señal. El sistema de detección puede incluir una fuente de iluminación, tal como una lámpara de arco, un láser, un diodo emisor de luz (LED) o similares. En realizaciones particulares, la unidad de detección de señales incluye ópticas para la transmisión de luz desde una fuente de iluminación a la muestra o desde la muestra al sensor de detección o de imagen. Como alternativa, la unidad de detección de señales puede no incluir una fuente de iluminación, tal como, por ejemplo, cuando una señal se produce espontáneamente como resultado de una reacción de secuenciación. Por ejemplo, puede producirse una
30 señal mediante la interacción de un resto liberado, tal como un ion liberado que interacciona con una capa sensible a los iones, o un pirofosfato que reacciona con una enzima u otro catalizador para producir una señal quimioluminiscente. En otro ejemplo, los cambios en una corriente eléctrica, voltaje o resistencia se detectan sin la necesidad de una fuente de iluminación.

En algunas realizaciones, una unidad de análisis y control de adquisición de datos supervisa varios parámetros del sistema. Los parámetros del sistema pueden incluir la temperatura de varias partes del instrumento, tal como una
40 unidad de procesamiento de muestras o depósitos de reactivos, volúmenes de varios reactivos, el estado de varios subcomponentes del sistema, como un manipulador, un motor paso a paso, una bomba o similar o cualquier combinación de los mismos.

Un experto en la materia apreciará que se utilizan diversas realizaciones de los instrumentos y sistemas para practicar métodos de secuenciación tales como secuenciación por síntesis, métodos de molécula única y otras técnicas de secuenciación. La secuenciación por síntesis puede incluir la incorporación de nucleótidos marcados con colorante, la terminación de la cadena, la secuenciación de iones/protones, la secuenciación de pirofosfatos o similares. Las técnicas de una sola molécula pueden incluir una secuenciación gradual, donde las reacciones de
45 secuenciación se detiene para determinar la identidad del nucleótido incorporado.

50 En algunas realizaciones, el instrumento de secuenciación determina la secuencia de un ácido nucleico, tal como un polinucleótido o un oligonucleótido. El ácido nucleico puede incluir ADN o ARN, y puede ser monocatenario, tal como ADNss y ARN, o bicatenario, como ADNds o un par de ARNA/ADNc. En algunas realizaciones, el ácido nucleico puede incluir o derivar de una biblioteca de fragmentos, una biblioteca de pares de parejas, un fragmento de ChIP, o similar. En realizaciones particulares, el instrumento de secuenciación puede obtener la información de secuencia de una única molécula de ácido nucleico o de un grupo de moléculas de ácido nucleico sustancialmente idénticas.

En algunas realizaciones, el instrumento de secuenciación puede generar datos de lectura de secuenciación de ácidos nucleicos en diversos tipos/formatos diferentes de archivos de datos de salida, que incluyen, entre otros: *.txt,
60 *.fasta, *.csfasta, *.seq.txt, *.qseq.txt, *.fastq, *.sff, *.prb.txt, *.sms, *.srs, y/o *.qv.

Análisis de ácidos nucleicos

En algunas realizaciones, se utiliza un programa de análisis informático para traducir los datos sin procesar generados por el ensayo de detección (por ejemplo, lecturas de secuenciación) en datos de valor predictivo para un
65 usuario final (por ejemplo, personal médico). El usuario puede acceder a los datos predictivos utilizando cualquier

medio adecuado. Por lo tanto, en algunas realizaciones preferidas, la presente tecnología proporciona el beneficio adicional de que el usuario, que probablemente no esté capacitado en genética o biología molecular, no requiere comprender los datos en bruto. Los datos se presentan directamente al usuario final en su forma más útil. El usuario puede utilizar inmediatamente la información para determinar información útil (por ejemplo, en diagnósticos médicos, investigación o detección).

Algunas realizaciones proporcionan un sistema para reconstruir una secuencia de ácido nucleico. El sistema puede incluir un secuenciador de ácido nucleico, un almacenamiento de datos de la secuencia de muestra, un almacenamiento de datos de la secuencia de referencia y un dispositivo/servidor/nodo de computación analítica. En algunas realizaciones, el dispositivo/servidor/nodo de computación analítica puede ser una estación de trabajo, ordenador principal, ordenador personal, dispositivo móvil, etc. El secuenciador de ácido nucleico puede configurarse para analizar (por ejemplo, interrogar) un fragmento de ácido nucleico (por ejemplo, fragmento individual, fragmento de pareja apareada, fragmento de extremo pareado, etc.) utilizando todas las variedades disponibles de técnicas, plataformas o tecnologías para obtener información de la secuencia de ácido nucleico, en particular los métodos descritos en el presente documento utilizando las composiciones proporcionadas en el presente documento. En algunas realizaciones, el secuenciador de ácido nucleico está en comunicación con el almacenamiento de datos de la secuencia de muestra, ya sea directamente a través de un cable de datos (por ejemplo, cable en serie, conexión directa por cable, etc.) o enlace de bus o, alternativamente, a través de una conexión de red (por ejemplo, Internet, LAN, WAN, VPN, etc.). En algunas realizaciones, la conexión de red puede ser una conexión física "cableada". Por ejemplo, el secuenciador de ácido nucleico se puede conectar comunicativamente (a través de la Categoría 5 (CAT5), fibra óptica o cableado equivalente) a un servidor de datos que está conectado comunicativamente (a través de CAT5, fibra óptica o cableado equivalente) a través de Internet y secuencia de muestras de almacenamiento de datos. En algunas realizaciones, la conexión de red es una conexión de red inalámbrica (por ejemplo, Wi-Fi, WLAN, etc.), por ejemplo, utilizando un formato de transmisión 802.11 a/b/g/n o equivalente. En la práctica, la conexión de red utilizada depende de los requisitos particulares del sistema. En algunas realizaciones, el almacenamiento de datos de la secuencia de muestra es una parte integrada del secuenciador de ácido nucleico.

En algunas realizaciones, el almacenamiento de datos de la secuencia de muestra es cualquier dispositivo de almacenamiento, sistema o implementación de la base de datos (por ejemplo, partición de almacenamiento de datos, etc.) que está configurado para organizar y almacenar los datos de lectura de la secuencia de ácido nucleico generados por el secuenciador de ácido nucleico de manera tal que los datos se pueden buscar y recuperar manualmente (por ejemplo, por un administrador de base de datos u operador del cliente) o automáticamente mediante un programa informático, una aplicación o un script de software. En algunas realizaciones, el almacenamiento de datos de referencia puede ser cualquier dispositivo de base de datos, sistema de almacenamiento o implementación (por ejemplo, partición de almacenamiento de datos, etc.) que esté configurado para organizar y almacenar secuencias de referencia (por ejemplo, genoma total o parcial, exoma total o parcial, SNP, gen, etc.) de modo que los datos puedan buscarse y recuperarse manualmente (por ejemplo, un administrador de la base de datos u operador del cliente) o automáticamente mediante un programa de computadora, una aplicación y/o un script de software. En algunas realizaciones, los datos de lectura de la secuencia de ácidos nucleicos de muestra pueden almacenarse en el almacenamiento de datos de secuencia de muestra y/o en el almacenamiento de datos de referencia en una variedad de diferentes tipos/formatos de archivos de datos, incluidos, entre otros: *.txt, *.fasta, *.csfasta, *.seq.txt, *.qseq.txt, *.fastq, *.sff, *.prb.txt, *.sms, *.srs y/o *.qv.

En algunas realizaciones, el almacenamiento de datos de la secuencia de muestra y el almacenamiento de datos de referencia son dispositivos/sistemas autónomos independientes o implementados en diferentes dispositivos. En algunas realizaciones, el almacenamiento de datos de la secuencia de muestra y el almacenamiento de datos de referencia se implementan en el mismo dispositivo/sistema. En algunas realizaciones, el almacenamiento de datos de la secuencia de muestra y/o el almacenamiento de datos de referencia pueden implementarse en el dispositivo/servidor/nodo de cálculo informático. El dispositivo/servidor/nodo informático analítico puede estar en comunicación con el almacenamiento de datos de secuencia de muestra y el almacenamiento de datos de referencia directamente a través de un cable de datos (por ejemplo, cable en serie, conexión directa por cable, etc.) o enlace de bus o, alternativamente, a través de una conexión de red (por ejemplo, Internet, LAN, WAN, VPN, etc.). En algunas realizaciones, el dispositivo informático analítico/servidor/nodo puede alojar un motor de mapeo de referencia, un módulo de mapeo de novo y/o un motor de análisis terciario. En algunas realizaciones, el motor de mapeo de referencia puede configurarse para obtener lecturas de secuencia de ácido nucleico de muestra del almacenamiento de datos de muestra y mapearlas contra una o más secuencias de referencia obtenidas del almacenamiento de datos de referencia para ensamblar las lecturas en una secuencia que es similar pero no necesariamente idéntica a la secuencia de referencia que utiliza todas las variedades de técnicas y métodos de mapeo/alineación de referencia. La secuencia reensamblada puede analizarse después mediante uno o más motores de análisis terciarios opcionales para identificar diferencias en la composición genética (genotipo), expresión génica o estado epigenético de individuos que pueden resultar en grandes diferencias en las características físicas (fenotipo). Por ejemplo, en algunas realizaciones, el motor de análisis terciario puede configurarse para identificar varias variantes genómicas (en la secuencia ensamblada) debido a mutaciones, recombinación/cruce o deriva genética. Los ejemplos de tipos de variantes genómicas incluyen, pero no se limitan a: polimorfismos de un solo nucleótido (SNP), variaciones en el número de copias (CNV), inserciones/deleciones

(Indels), inversiones, etc. El módulo opcional de mapeo de novo se puede configurar para ensamblar la secuencia de ácido nucleico de muestra se lee del almacenamiento de datos de la muestra en secuencias nuevas y previamente desconocidas. Sin embargo, debe entenderse que los diversos motores y módulos alojados en el dispositivo/servidor/nodo de cómputo analítico se pueden combinar o colapsar en un solo motor o módulo, según los requisitos de la aplicación particular o la arquitectura del sistema. Además, en algunas realizaciones, el dispositivo/servidor/nodo informático analítico puede alojar motores o módulos adicionales según lo requiera la aplicación particular o la arquitectura del sistema.

En algunas realizaciones, los motores de mapeo y/o análisis terciario están configurados para procesar el ácido nucleico y/o las secuencias de referencia en el espacio de color. En algunas realizaciones, los motores de mapeo y/o análisis terciario están configurados para procesar el ácido nucleico y/o las secuencias de referencia en el espacio base. Sin embargo, debe entenderse que los motores de mapeo y/o análisis terciario descritos en el presente documento pueden procesar o analizar los datos de la secuencia de ácido nucleico en cualquier esquema o formato, siempre que el esquema o formato pueda transmitir la identidad y posición de la base de la secuencia de ácido nucleico.

En algunas realizaciones, la lectura de la secuenciación de ácidos nucleicos de muestra y los datos de la secuencia de referencia pueden suministrarse al dispositivo/servidor/nodo informáticos analíticos en una variedad de diferentes tipos/formatos de archivos de datos de entrada, que incluyen, entre otros: *.txt, *.fasta, *.csfasta, *.seq.txt, *.qseq.txt, *.fastq, *.sff, *.prb.txt, *.sms, *.srs y/o *.qv.

Además, un terminal de cliente puede ser un dispositivo informático de cliente ligero o cliente pesado. En algunas realizaciones, el terminal del cliente puede tener un navegador web que se puede usar para controlar el funcionamiento del motor de mapeo de referencia, el módulo de mapeo de novo y/o el motor de análisis terciario. Es decir, el terminal del cliente puede acceder al motor de mapeo de referencia, el módulo de mapeo de novo y/o el motor de análisis terciario usando un navegador para controlar su función. Por ejemplo, el terminal del cliente puede usarse para configurar los parámetros operativos (por ejemplo, restricción de falta de coincidencia, umbrales de valor de calidad, etc.) de los diversos motores, según los requisitos de la aplicación en particular. De manera similar, el terminal del cliente también puede mostrar los resultados del análisis realizado por el motor de mapeo de referencia, el módulo de mapeo de novo y/o el motor de análisis terciario.

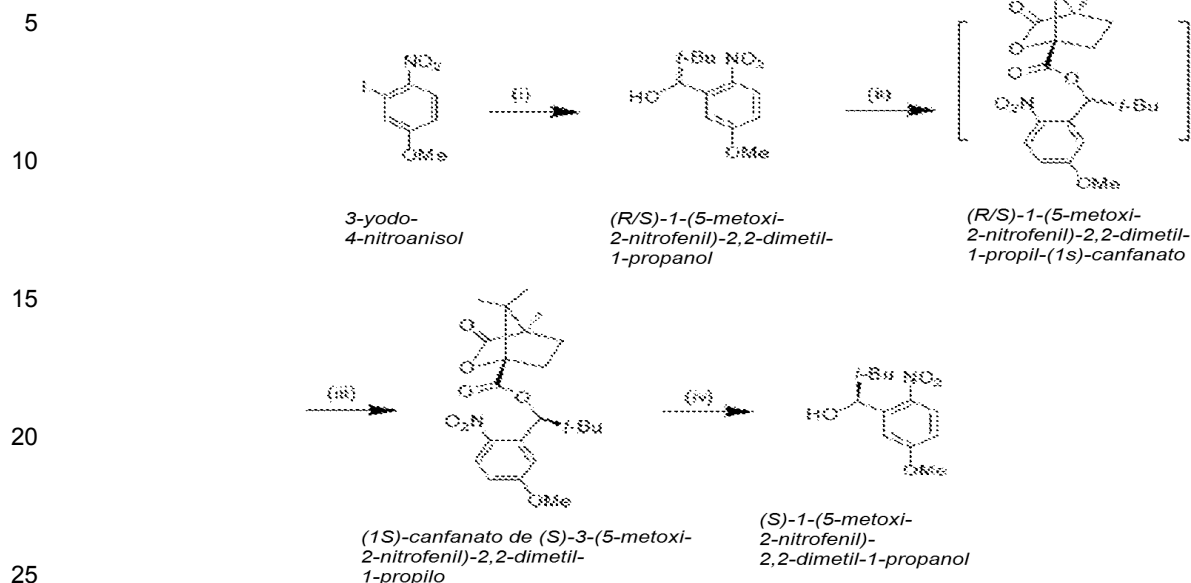
La presente tecnología también abarca cualquier método capaz de recibir, procesar y transmitir la información hacia y desde los laboratorios que realizan los ensayos, la información, el personal médico y los sujetos.

Aunque la divulgación en el presente documento se refiere a ciertas realizaciones ilustradas, debe entenderse que estas realizaciones se presentan a modo de ejemplo y no a modo de limitación.

EJEMPLOS

1. Síntesis del intermedio de alcohol (S)-t-butil-5-metoxi-2-nitrobencilico

(R/S)-2-(5-metoxi-2-nitrofenil)-2,2-dimetil-1-propanol y (S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-1-propanol



Esquema 54. Síntesis de (R/S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-1-propanol y (S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-1-propanol. Reactivos y condiciones: (i) PhMgCl, THF anhidro, menos 40°C; (CH₃)₂CHO, menos 40°C a temperatura ambiente, 88 %. (ii) cloruro de ácido (1S)-canfánico, DMAP, CH₂Cl₂, Temperatura ambiente; (iii) cristalización fraccionada en acetato de etilo/hexano, 43 %; (iv) 2CO₂, MeOH, reflujo, 99 %.

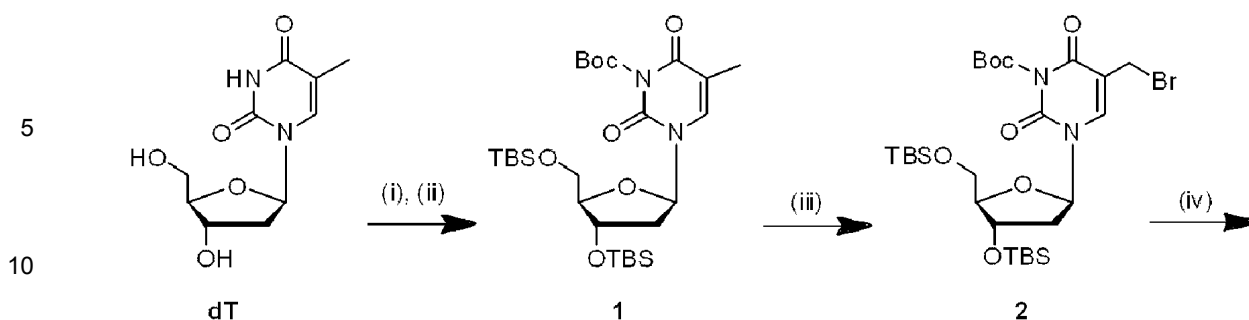
A una solución de 3-yodo-4-nitroanisol (2,79 g, 10,0 mmol) en THF anhidro (10 ml) a menos 40 °C bajo una atmósfera de nitrógeno, se utilizó cloruro de fenilmagnesio (2 M en THF, 4,2 ml, 8,3 mmol) se añade gota a gota a una velocidad tal que la temperatura no exceda menos 35 1C. Una vez completada la adición, la mezcla se agitó a menos 40 °C durante dos horas, seguido de la adición de trimetilacetaldérido (1,1 ml, 10 mmol). La mezcla se agitó a menos 40 °C durante dos horas y luego a temperatura ambiente durante una hora más. A continuación, la reacción se detuvo con salmuera (100 ml) y la mezcla se extrajo con CH₂Cl₂ (40 ml) tres veces. La fase orgánica combinada se secó sobre Na₂SO₄, se concentró al vacío, y el residuo se purificó por cromatografía en columna de gel de sílice para producir (R/S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-1-propanol racémico (1,76 g, 88 %).

RMN ¹H (400 MHz, CDCl₃): δ 7,89 (d, 1 H, J = 9,2 Hz, Ph-H), 7,27 (d, 1 H, J = 2,8 Hz, Ph-H), 6,84 (dd, 1 H, J = 8,8 and 2,8 Hz, Ph-H), 5,62 (d, 1 H, J = 4,0 Hz, PhCH), 3,89 (s, 3 H, OCH₃), 2,08 (d, 1 H, J = 4,0 Hz, OH), 0,89 (s, 9 H, C(CH₃)₃).

A una solución de (R/S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-1-propanol racémico (1,75 g, 7,3 mmol) y DMAP (2,92 g, 23,9 mmol) en CH₂Cl₂ anhidro (10 ml), se añadió cloruro (1S)-canfánico (2,6 g, 12 mmol) y la mezcla se agitó durante la noche a temperatura ambiente bajo una atmósfera de nitrógeno. La mezcla de reacción se diluyó con CH₂Cl₂ (50 ml) y se lavó con una solución de NaHCO₃ saturada (50 ml). La fase orgánica se secó sobre Na₂SO₄, se concentró al vacío, y el residuo se purificó por cromatografía en columna de gel de sílice para producir (1S)-canfanato de (R/S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-1-propilo (2,5 g, 85 %, mezcla 1:1 de diastereoisómeros). El canfanato se disolvió en acetato de etilo (30 ml) seguido de la adición lenta de hexano (120 ml) con agitación. Los cristales de aguja se formaron gradualmente a partir de la solución durante un período de dos horas. Los cristales se recogieron por filtración para producir diastereómero único (1S)-canfanato de (S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-1-propilo puro. El filtrado se concentró al vacío y el proceso de cristalización se repitió dos veces para proporcionar (1S)-canfanato de (S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-1-propilo adicional (total 1,08 g, 43 %).

RMN ¹H (400 MHz, CDCl₃): δ 8,04 (d, 1 H, J = 9,2 Hz, Ph-H), 7,27 (d, 1 H, J = 2,8 Hz, Ph-H), 6,88 (dd, 1 H, J = 2,8 and 8,8 Hz, Ph-H), 6,81 (3, 1 H, Ph-CH), 3,87 (s, 3 H, OCH₃), 2,36 (m, 1 H, CH), 1,92 (m, 2 H, CH₂), 1,66 (m, 1 H, CH), 1,12 (s, 3 H, CH₃), 1,06 (s, 3 H, CH₃), 1,02 (s, 3 H, CH₃), 0,95 (s, 9 H, C(CH₃)₃).

2. Síntesis de análogo de 5-bromometil-2'-desoxiuridina



- Esquema S5. Síntesis de los análogos de 5-(2-nitrobenzoyl) metil-dUTP. Reactivos y condiciones: (i) TBSCl, imidazol, DMF anhidra, temperatura ambiente, 16 h, 95 %; (ii) Boc₂O, DMAP, DMF anhidro, temperatura ambiente, 16 h, 78 %; (iii) N-bromosuccinimida, peróxido de benzoilo, CCl₄, reflujo, una hora, 43 %;

3',5'-O-bis-terc-butildimetilsilil-N₃-terc-butiloxycarbonil-timidina (1). A una solución de timidina (5,00 g, 20,64 mmol) e imidazol (9,0 g, 132,1 mmol) en DMF anhidra (11 ml), se añadió gota a gota una solución de TBSCl (9,96 g, 66,05 mmol) en DMF anhidra (11 ml) y la mezcla se agitó a temperatura ambiente durante 16 h bajo una atmósfera de nitrógeno. El disolvente se eliminó al vacío y el residuo se purificó mediante cromatografía en columna de gel de sílice para producir 3',5'-O-bis-terc-butildimetilsilil-timidina (9,23 g, 95 %) como un sólido blanco.

RMN ¹H (400 MHz, CDCl₃): δ 8,51 (s, 1 H, NH), 7,48 (d, 1 H, J = 1,2 Hz, H-6), 6,34 (dd, 1 H, J = 5,8 and 8,0 Hz, H-1'), 4,41 (m, 1 H, H-3'), 3,93 (m, 1 H, H-4'), 3,87 (dd, 1 H, J = 2,6 and 11,4 Hz, H-5'a), 3,76 (dd, 1 H, J = 2,6 and 11,4 Hz, H-5'b), 2,17 (m, 1 H, H-2'a), 2,01 (m, 1 H, H-2'b), 1,92 (d, 3 H, J = 1,2 Hz, CH₃), 0,93 (s, 9 H, (CH₃)₃CSi), 0,88 (s, 9 H, (CH₃)₃CSi), 0,11 (s, 6 H, (CH₃)₂Si), 0,08 (s, 6 H, (CH₃)₂Si).

A una solución del compuesto 3',5'-O-bis-terc-butildimetilsilil-timidina (2,43 g, 5,15 mmol) y DMAP (1,39 g, 11,34 mmol) en DMF anhidra (45 ml), se añadió gota a gota una solución de di-terc-butildicarbonato (2,47 g, 11,34 mmol) en DMF anhidra (9 ml). La mezcla se agitó a temperatura ambiente durante 16 horas bajo una atmósfera de nitrógeno. La mezcla se concentró después al vacío, y el residuo se disolvió en CH₂Cl₂ (80 ml), se lavó con una solución de NH₄Cl saturado (10 ml), se secó sobre Na₂SO₄, y se concentró al vacío. El residuo se purificó por cromatografía en columna de gel de sílice para dar 3', 5'-O-bis-terc-butildimetilsilil-N₃-terc-butiloxycarbonil-timidina 1 (2,30 g, 78 %) como un sólido blanco.

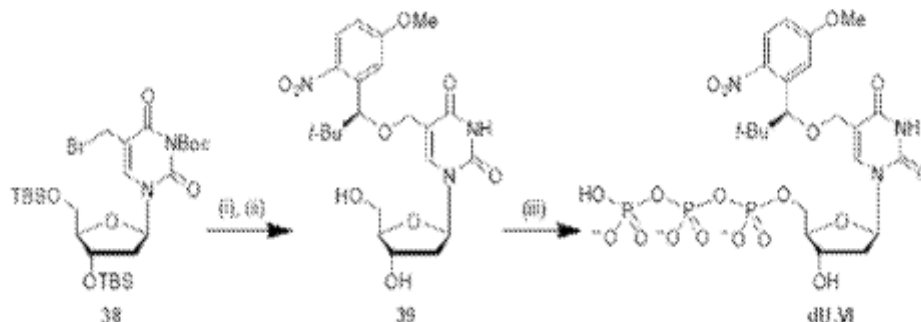
RMN ¹H (400 MHz, CDCl₃): δ 7,50 (d, 1 H, J = 1,2 Hz, H-6), 6,34 (dd, 1 H, J = 5,8 and 8,0 Hz, H-1'), 4,42 (m, 1 H, H-3'), 3,95 (m, 1 H, H-4'), 3,87 (dd, 1 H, J = 2,5 and 11,4 Hz, H-5'a), 3,76 (dd, 1 H, J = 2,5 and 11,4 Hz, H-5'b), 2,17 (m, 1 H, H-2'a), 2,01 (m, 1 H, H-2'b), 1,92 (d, 3 H, J = 1,2 Hz, CH₃), 1,60 (s, 9 H, (CH₃)₃), 0,93 (s, 9 H, (CH₃)₃CSi), 0,88 (s, 9 H, (CH₃)₃CSi), 0,11 (s, 6 H, (CH₃)₂Si), 0,08 (s, 6 H, (CH₃)₂Si).

3',5'-Bis-O-terc-butildimetilsilil-N₃-terc-butiloxycarbonil-5-bromometil-2'-desoxiuridina (2). Una solución del compuesto 1 (570 mg, 1,00 mmol), N-bromosuccinimida (0,37 g, 2,10 mmol) y peróxido de benzoilo (10 mg, solución acuosa al 75 %) en CCl₄ (10 ml) se calentó a reflujo durante una hora (7). La mezcla se filtró y el filtrado se concentró al vacío y se purificó por cromatografía en columna de gel de sílice para dar 3',5'-bis-O-terc-butildimetilsilil-N₃-terc-butiloxycarbonil-5-bromometil-2'-desoxiuridina 2 (281 mg, 43 %) como un sólido amarillo.

RMN ¹H (400 MHz, CDCl₃): δ 7,89 (s, 1 H, H-6), 6,27 (dd, 1 H, J = 5,8 and 8,0 Hz, H-1'), 4,38 (m, 1 H, H-3'), 4,27 (d, 1 H, J = 10,6 Hz, CH₂Br), 4,20 (d, 1 H, J = 10,6 Hz, CH₂Br), 3,98 (m, 1 H, H-4'), 3,88 (dd, 1 H, J = 2,5 and 11,4 Hz, H-5'b), 3,77 (dd, 1 H, J = 2,6 and 11,4 Hz, H-5'a), 2,33 (m, 1 H, H-2'a), 2,01 (m, 1 H, H-2'b), 1,61 (s, 9 H, (CH₃)₃), 0,95 (s, 9 H, (CH₃)₃CSi), 0,89 (s, 9 H, (CH₃)₃CSi), 0,14 (s, 6 H, (CH₃)₂Si), 0,07 (s, 6 H, (CH₃)₂Si); ¹³C NMR (100 MHz, CDCl₃): δ 159,21 (C), 147,99 (C), 147,32 (C), 138,46 (CH), 111,30 (C), 88,34 (CH), 87,22 (C), 86,00 (CH), 72,28 (CH), 63,04 (CH₂), 41,93 (CH₂), 27,42 (CH₃), 25,99 (CH₃), 25,71 (CH₃), 25,65 (CH₃), 24,91 (CH₂), 18,47 (C), 17,97 (C), -3,58 (CH₃), -4,65 (CH₃), -4,86 (CH₃), -5,32 (CH₃).

3. Síntesis del análogo de trifosfato de 5-HOMe-2'-desoxiuridina

S-((S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-propiloxi)metil-2'-desoxiuridin-5'-trifosfato



Esquema S19. Síntesis de S-((S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-propiloxi)metil-2'-desoxiuridin-5'-trifosfato.

Reactivos y condiciones: (i) (S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-1-propanol, 110 °C

; (ii) NH₄F, MeOH, 50 °C. 56 % para dos etapas; (iii)

POCl₃, esponja de protones, (MeO)P(O)O⁻, 0°C; (n-Bu₃NH)2H₂P3O₃, n-Bu₃N, DMF; 1M HNEt₃HCO₂

El compuesto 38 (315 mg, 0,49 mmol) y (S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-1-propanol (490 mg, 2,1 mmol) se calentaron a 110 °C durante 45 min bajo una atmósfera de nitrógeno. La mezcla se enfrió a temperatura ambiente, se disolvió en MeOH (10 ml) y se siguió con la adición de NH₄F (400 mg, 11 mmol). La mezcla se agitó a 50 °C durante 12 horas, se concentró al vacío, se disolvió en CH₂Cl₂ (50 ml) y se lavó con salmuera (50 ml). La fase orgánica se secó sobre Na₂SO₄, se concentró al vacío, y el residuo se purificó por cromatografía en gel de sílice para producir 5-[(S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-propiloxi] metil-2'-desoxiuridina 39 (130 mg, 56 %).

RMN ¹H (400 MHz, CDCl₃): δ 9,14 (s a, 1 H, NH), 7,90 (d, 1 H, J = 9,2 Hz, Ph-H), 7,67 (s, 1 H, H-6), 7,17 (d, 1 H, J = 2,8 Hz, Ph-H), 6,84 (dd, 1 H, J = 9,2 and 2,8 Hz, Ph-H), 6,18 (t, 1 H, J = 6,4 Hz, H-1'), 5,22 (s, 1 H, Ph-CH), 4,56 (m, 1 H, H-3'), 4,24 (d, 1 H, J = 12,4 Hz, 5-CH₂a), 4,15 (d, 1 H, J = 12,4 Hz, 5-CH₂b), 4,00 (m, 1 H, H-4'), 3,90 (m, 1 H, H-5'a), 3,88 (s, 3 H, OCH₃), 3,81 (m, 1 H, H-5'b), 2,35 (m, 2 H, H-2), 0,83 (s, 9 H, C(CH₃)₃).

El compuesto 39 (30 mg, 0,065 mmol) se fosforiló con POCl₃ (9 µl, 0,097 mmol) y esponja de protones (28 mg, 0,13 mmol) en fosfato de trimetilo (0,35 ml) a 0 °C durante una hora bajo una atmósfera de nitrógeno. Se añadió una solución de pirofosfato de tri-n-butilamonio (147 mg, 0,32 mmol) y tri-n-butilamina (64 µl) en DMF anhidra (0,64 ml). Después de 10 minutos de agitación, se añadió tampón bicarbonato de trietilamonio (0,1 M, pH 7,5; 10 ml). La reacción se agitó a temperatura ambiente durante una hora y luego se concentró al vacío. El residuo se disolvió en acetonitrilo acuoso al 20 % (10 ml), se filtró y se purificó por cromatografía de intercambio aniónico. Las fracciones que contenían trifosfato se combinaron y se liofilizaron para producir 5-[(S)-1-(5-metoxi-2-nitrofenil)-2,2-dimetil-propiloxi] metil-2'-desoxiuridina-5'-trifosfato dU. VI, que se purificó adicionalmente usando RP-HPLC.

REIVINDICACIONES

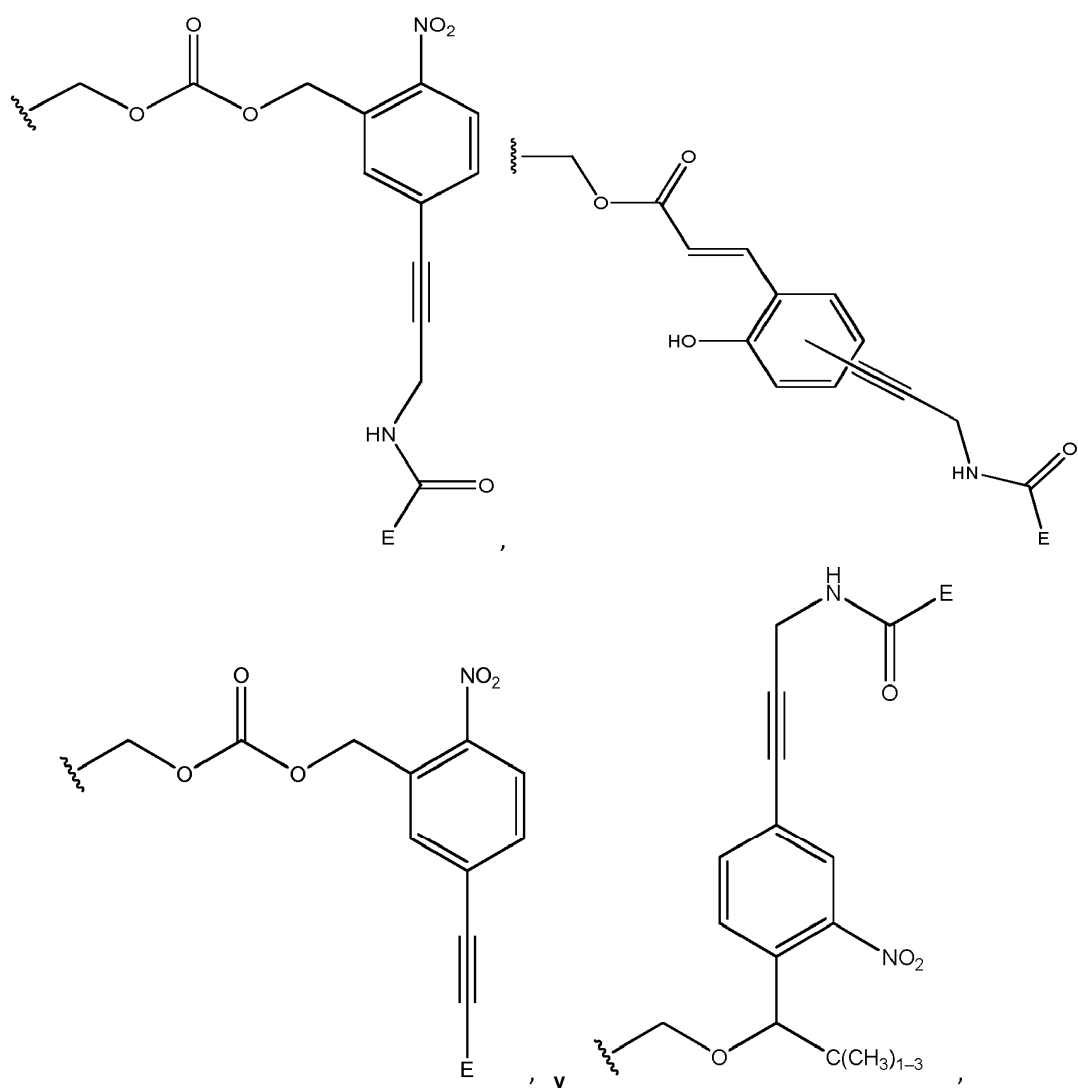
1. Un análogo nucleotídico que comprende:

- a) un nucleótido que comprende un resto fosfato, una base y un azúcar;
- b) un resto de terminación fotoescindible, detectable electroquímicamente, unido al nucleótido; y
- c) un segundo resto de terminación fotoescindible unido al nucleótido.

2. El análogo nucleotídico de la reivindicación 1, en el que el resto de terminación fotoescindible y detectable electroquímicamente está unido al resto de fosfato, la base o el azúcar del nucleótido.

3. El análogo nucleotídico de una cualquiera de las reivindicaciones 1 a 2, en el que el resto de terminación fotoescindible, detectable electroquímicamente, está unido al nucleótido mediante un enlazador fotoescindible.

4. El análogo nucleotídico de la reivindicación 3, en el que el resto de terminación fotoescindible, detectable electroquímicamente, se selecciona del grupo que consiste en



en el que E representa un resto detectable electroquímicamente.

5. El análogo nucleotídico de la reivindicación 3, en el que el resto de terminación fotoescindible, detectable electroquímicamente, está unido al nucleótido mediante un enlazador fotoescindible seleccionado del grupo que

consiste en éter, éster, diéster, 2-nitrobencilo, 2-nitrobencilo alfa-sustituido, 1- (nitrofenil)etilo, 3,5-dimetoxibencilo, ácido tiohidroxámico, 7-nitroindolina, 9-fenilxantilo, benzoína, hidroxifenacilo, nitroveratril y ácido N-hidroxi-succinimidil-4-azidosalicílico.

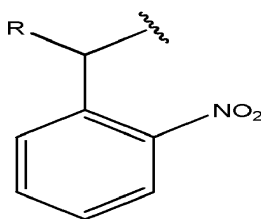
6. El análogo nucleotídico de la reivindicación 1, en el que el resto de terminación fotoescindible electroquímicamente detectable comprende un resto detectable electroquímicamente seleccionado del grupo que consiste en un grupo redox orgánico, un grupo organometálico, una nanopartícula metálica y un punto cuántico.

7. El análogo nucleotídico de la reivindicación 6, en el que:

- a) el grupo redox orgánico se selecciona del grupo que consiste en azul de metileno, antraquinona y tionina;
- b) el grupo organometálico se selecciona del grupo que consiste en ferroceno, derivados de ferroceno, un complejo de bipyrideno de rutenio y un complejo de bipyrideno de osmio;
- c) la nanopartícula metálica se selecciona del grupo que consiste en oro y plata; o
- d) el punto cuántico se selecciona del grupo que consiste en CdS, CdSe y ZnS.

8. El análogo nucleotídico de la reivindicación 1, en el que el segundo resto de terminación fotoescindible está unido al azúcar del nucleótido, en el que opcionalmente el segundo resto de terminación fotoescindible está unido a una posición 3' del azúcar del nucleótido.

9. El análogo nucleotídico de la reivindicación 1, en el que el segundo resto de terminación fotoescindible se selecciona del grupo que consiste en



en la que R es cualquier grupo alifático que consiste en 1-3 carbonos o un hidrógeno.

10. Una composición que comprende un análogo nucleotídico de una cualquiera de las reivindicaciones 1 a 9 y una polimerasa.

11. Uso de un análogo nucleotídico de una cualquiera de las reivindicaciones 1 a 9 en una reacción de secuenciación de ácidos nucleicos.

12. Un método para secuenciar un ácido nucleico, comprendiendo el método:

- a) hibridar un cebador con un ácido nucleico para formar un complejo de cebador/ácido nucleico hibridado;
- b) proporcionar una pluralidad de análogos nucleotídicos, comprendiendo cada análogo nucleotídico un nucleótido, un resto de terminación fotoescindible y detectable electroquímicamente unido al nucleótido y un segundo resto de terminación fotoescindible unido al nucleótido;
- c) hacer reaccionar el complejo de cebador/ácido nucleico hibridado y el análogo nucleotídico con una polimerasa para agregar el análogo nucleotídico al cebador mediante una reacción de polimerasa para formar un producto extendido que comprende un análogo nucleotídico incorporado;
- d) consultar el producto extendido para identificar el análogo nucleotídico incorporado;
- e) exponer el producto extendido a una fuente de luz que proporciona una longitud de onda de la luz para eliminar un resto de terminación fotoeléctrico, detectable electroquímicamente, del análogo nucleotídico incorporado; y
- f) exponer el producto extendido a una fuente de luz que proporciona una segunda longitud de onda de luz para eliminar un segundo resto de terminación fotoescindible del análogo nucleotídico incorporado.

13. Un kit o mezcla de reacción para secuenciar un ácido nucleico que comprende cuatro análogos nucleotídicos según una cualquiera de las reivindicaciones 1 a 9, en el que un primer análogo nucleotídico comprende un nucleótido adenina, un segundo análogo nucleotídico comprende un nucleótido citosina, un tercer análogo nucleotídico comprende un nucleótido guanina y un cuarto análogo nucleotídico comprende un nucleótido timina o un nucleótido uracilo.

14. Un sistema para secuenciar un ácido nucleico, comprendiendo el sistema:

- a) un análogo nucleotídico de acuerdo con una cualquiera de las reivindicaciones 1 a 9,

- b) una o más fuentes de luz coherentes; y
- c) un elemento de detección de fluorescencia o electroquímica.

15. El sistema según la reivindicación 14, que comprende además un sustrato que comprende una guía de onda de modo cero.

5

10

15

20

25

30

35

40

45

50

55

60

65

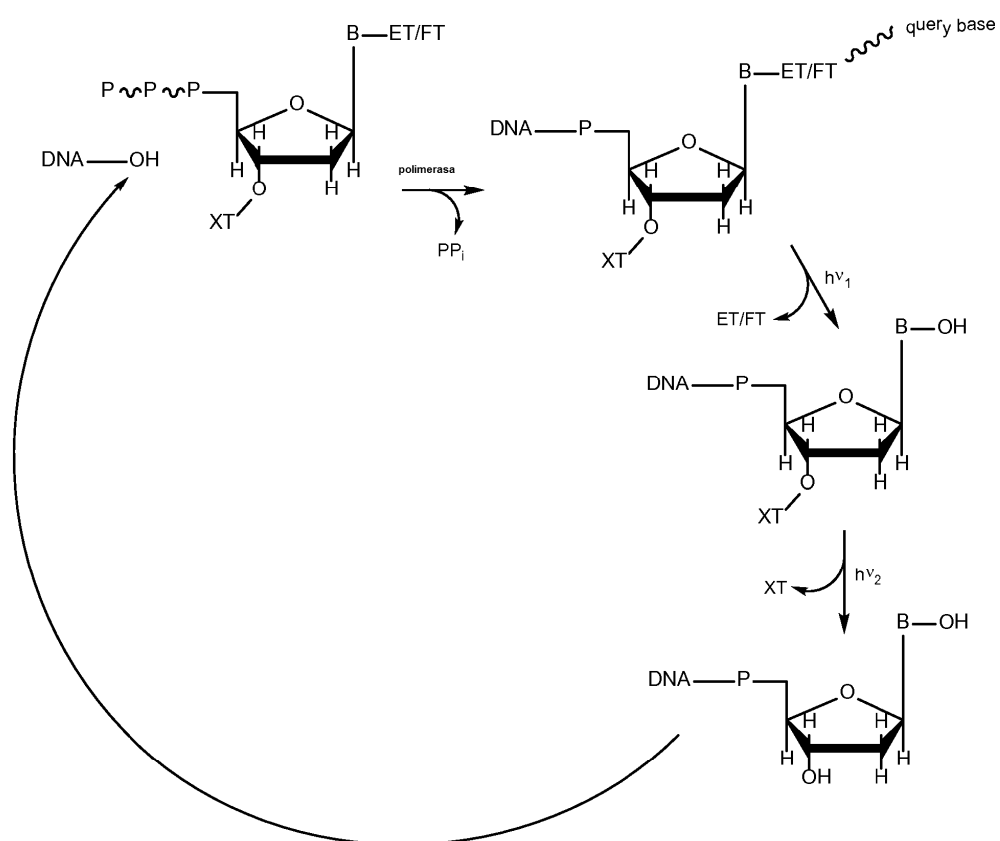


FIG. 1