

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 982 551**

51 Int. Cl.:

C12P 19/34 (2006.01)

C12N 9/12 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **26.05.2020 PCT/EP2020/064524**

87 Fecha y número de publicación internacional: **03.12.2020 WO20239737**

96 Fecha de presentación y número de la solicitud europea: **26.05.2020 E 20727647 (8)**

97 Fecha y número de publicación de la concesión europea: **05.06.2024 EP 3976809**

54 Título: **Variantes de desoxinucleotidil transferasa terminal y usos de las mismas**

30 Prioridad:

28.05.2019 EP 19177018
28.05.2019 US 201916423972

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
16.10.2024

73 Titular/es:

DNA SCRIPT (33.3%)
Immeube Okabé, 67 avenue de Fontainebleau
94270 Le Kremlin-Bicêtre, FR;
INSTITUT PASTEUR (33.3%) y
CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (33.3%)

72 Inventor/es:

CHAMPION, ELISE;
SOSKINE, MIKHAEL;
YBERT, THOMAS y
DELARUE, MARC

74 Agente/Representante:

VEIGA SERRANO, Mikel

Observaciones:

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

ES 2 982 551 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

Variantes de desoxinucleotidil transferasa terminal y usos de las mismas

5 **Sector de la técnica**

La invención se refiere a variantes de desoxinucleotidil transferasa terminal (TdT) y a usos de las mismas para la síntesis enzimática de secuencias de ácido nucleico sin una cadena molde. Más particularmente, la presente invención se refiere a dichas variantes adecuadas para incorporar nucleótidos modificados, para la síntesis de moléculas de

10 ácido nucleico con secuencias determinadas o controladas.

Estado de la técnica

15 Durante los últimos 40 años se han empleado ampliamente métodos para la síntesis química *de novo* de ácidos nucleicos basándose en la química de fosforamidita en fase sólida. La técnica consiste en un ciclo de elongación de la cadena en cuatro etapas que añade una base por ciclo a una cadena de oligonucleótido en crecimiento unida a una matriz de soporte sólido. Aunque ha sido el método de elección para sintetizar ácidos nucleicos durante las últimas décadas, esta tecnología tiene algunas limitaciones importantes: Requiere usar múltiples disolventes y reactivos y debido a las limitaciones en la eficiencia de la reacción química, la longitud de los oligonucleótidos sintéticos

20 normalmente no supera las 150–200 bases. Además, es necesario ensamblar adicionalmente estos fragmentos cortos para obtener la secuencia de ADN deseada.

Una alternativa a la síntesis química consiste en usar ADN polimerasas independientes de la cadena molde que añadirán nucleótidos modificados de terminador reversible a una cadena de ácidos nucleicos monocatenaria en crecimiento. Esto permite la adición de un tipo de nucleótido por ciclo de modo controlado.

25

Algunas enzimas nativas tienen la capacidad de actuar sobre los nucleótidos naturales en ausencia de cadena molde y de este modo, pueden catalizar la síntesis de ácidos nucleicos de una manera no controlada. Sin embargo, son particularmente ineficaces a la hora de incorporar nucleótidos modificados y más concretamente, nucleótidos modificados de terminador reversible. Se han practicado esfuerzos para desarrollar nuevas ADN polimerasas capaces de actuar sobre los nucleótidos modificados, pero las enzimas obtenidas no son plenamente satisfactorias en cuanto al rendimiento para la síntesis de cualquier tipo de ácido nucleico.

30

Hasta ahora, se han identificado algunas ADN polimerasas que pueden actuar eficazmente en el ADN monocatenario (sin emplear una cadena molde). La polimerasa más caracterizada que tiene dicha actividad independiente de la cadena molde es la desoxinucleotidil transferasa terminal (TdT). Las enzimas TdT se han utilizado ampliamente para modificar ADN monocatenario para diversos tipos de aplicaciones, entre las que se incluyen biotecnología, investigación biomédica y biología sintética. Sin embargo, la TdT nativa apenas es capaz de utilizar nucleótidos modificados.

35

Se han realizado varios intentos de desarrollar TdT modificada con un rendimiento aceptable para la incorporación de nucleótidos modificados. Sin embargo, el rendimiento de incorporación de dichos nucleótidos modificados sigue siendo un factor limitante. La eficiencia de incorporación es el principal parámetro que determina la pureza y el rendimiento generales de la síntesis. Estas dos características del proceso de síntesis tienen un impacto significativo en la calidad, el tiempo de obtención y el coste de los productos de ácido nucleico.

45

Por tanto, existe la necesidad de desarrollar TdT mejoradas capaces de utilizar nucleótidos modificados en ausencia de cadena molde, para desarrollar métodos eficientes y rentables para la síntesis de ácidos nucleicos.

50 **Objeto de la invención**

Al trabajar en TdT para la síntesis *de novo* de polinucleótidos con una secuencia controlada y sin utilizar una cadena molde, los inventores han descubierto que pueden modificarse algunos restos de aminoácido específicos del dominio catalítico de la TdT con el fin de mejorar la capacidad de dicha TdT modificada para sintetizar polinucleótidos. Más particularmente, los inventores han desarrollado TdT modificadas con sustituciones de aminoácidos específicas que mejoran la síntesis enzimática de polinucleótidos y reducen el coste general de la síntesis de polinucleótidos. Los inventores desarrollaron previamente variantes de TdT murina (SEQ ID NO: 1 o SEQ ID NO: 2). Al seguir estudiando la TdT, los inventores han desarrollado nuevas variantes de TdT.

55

En algunas realizaciones, cada una de las TdT modificadas presenta una o más sustituciones de aminoácidos específicas en comparación con las TdT de tipo silvestre (tales como las SEQ ID NO: 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 o 34) y versiones truncadas N-terminales de las mismas que comprenden un dominio catalítico de TdT (tales como las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35). En algunas realizaciones, cada una de las TdT modificadas de la invención posee una secuencia de aminoácidos que tiene un porcentaje de identidad de secuencia específico con respecto a un dominio catalítico de una TdT (tal como las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35) y que tiene una o más sustituciones de aminoácidos específicas. Las polimerasas

60

65

independientes de la cadena molde de la invención permiten las síntesis enzimática de polinucleótidos a mayor velocidad, con un coste menor y mayor calidad.

5 Por tanto, un objeto de la invención es proporcionar variantes de desoxinucleotidil transferasas terminales (TdT) que comprenden una secuencia de aminoácidos de un dominio catalítico de TdT o de un porcentaje de identidad de secuencia de un dominio catalítico de una TdT, como se expone en las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35, con al menos una sustitución de aminoácidos en la posición correspondiente al resto C173 (con respecto a la numeración de aminoácidos de la SEQ ID NO: 11), o un resto funcionalmente equivalente, es capaz de sintetizar un fragmento de ácido nucleico sin cadena molde y es capaz de incorporar un nucleótido modificado, tal como un nucleótido modificado en 3'-O, en un fragmento de ácido nucleico. Concretamente, el nucleótido modificado se incorpora en un 3'-hidroxilo libre del fragmento de ácido nucleico.

Más particularmente, un objeto de la presente invención es proporcionar variantes de desoxinucleotidil transferasa terminal (TdT) que comprende una secuencia de aminoácidos al menos un 90 % idéntica a la SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35 con una sustitución en la posición correspondiente al resto C173 con respecto a las SEQ ID NO: 11, 13, 17, 19, 21, 29 o 31, o en la posición correspondiente al resto C172 con respecto a la SEQ ID NO: 15, o en la posición correspondiente al resto C178 con respecto a la SEQ ID NO: 23, o en la posición correspondiente al resto C174 con respecto a la SEQ ID NO: 25, o en la posición correspondiente al resto C171 con respecto a la SEQ ID NO: 27, o en la posición correspondiente al resto C182 con respecto a la SEQ ID NO: 33, o en la posición correspondiente al resto C176 con respecto a la SEQ ID NO: 35, y en donde la secuencia de aminoácidos comprende una sustitución adicional de un resto de arginina en la posición 207 con respecto a las SEQ ID NO: 11, 13, 17, 19, 23, 29 o 31, o de un resto de arginina funcionalmente equivalente seleccionado entre una arginina en la posición 206 con respecto a la SEQ ID NO: 15, o una arginina en la posición 208 con respecto a las SEQ ID NO: 21 o 25, o una arginina en la posición 205 con respecto a la SEQ ID NO: 27, o una arginina en la posición 216 con respecto a la SEQ ID NO: 33, o una arginina en la posición 210 con respecto a la SEQ ID NO: 35, en donde la variante de TdT (i) es capaz de sintetizar un fragmento de ácido nucleico sin una cadena molde y (ii) es capaz de incorporar un nucleótido modificado en un fragmento de ácido nucleico.

30 En algunas realizaciones, el valor de porcentaje de identidad anterior es al menos un 95 % de identidad con respecto a las SEQ ID NO indicadas; en algunas realizaciones, el valor de porcentaje de identidad anterior es al menos un 97 % de identidad; en algunas realizaciones, el valor de porcentaje de identidad anterior es al menos un 98 % de identidad; en algunas realizaciones, el valor de porcentaje de identidad anterior es al menos un 99 % de identidad.

35 Ventajosamente, con respecto a (ii), dicho nucleótido modificado es un nucleótido modificado en 3'-O que puede incorporarse en un 3'-hidroxilo libre del fragmento de ácido nucleico. El nucleótido modificado en 3'-O puede comprender un trifosfato de 3'-O-NH₂-nucleósido, un trifosfato de 3'-O-azidometil-nucleósido, un trifosfato de 3'-O-allyl-nucleósido, un trifosfato de 3'-O-(2-nitrobenzil)-nucleósido o un trifosfato de 3'-O-propargil-nucleósido.

40 Una parte de la invención son variantes que tienen sustituciones únicamente en al menos las posiciones 173 con respecto a las SEQ ID NO: 11, 13, 17, 19, 21, 29 o 31 y 207 con respecto a las SEQ ID NO: 11, 13, 17, 19, 23, 29 o 31 y posiciones funcionalmente equivalentes. Cualquier otra divulgación es una divulgación que no forma parte de la invención. La sustitución se selecciona entre

45 C313G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 10; o C173G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 11; o C302G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 12; o C173G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 13; o C302G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 14; o C172G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 15; o C304G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 16; o C173G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 17; o C304G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 18; o C173G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 19; o C293G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 20; o C174G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 21; o C282G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 22; o C173G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 23; o C304G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 24; o C174G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 25; o C300G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 26; o C171G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 27; o C305G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 28; o C173G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 29; o C302G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 30; o C173G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 31; o C313G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 32; o C182G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 33; o C271G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 34; o C176G/R/P/A/V/S/N/Q/D con respecto a la SEQ ID NO: 35.

60 En una realización adicional, la sustitución se selecciona entre C302G/R con respecto a la SEQ ID NO: 4; o C302G/R con respecto a la SEQ ID NO: 9; o C313G/R con respecto a la SEQ ID NO: 10; o C173G/R con respecto a la SEQ ID NO: 11; o C302G/R con respecto a la SEQ ID NO: 12; o C173G/R con respecto a la SEQ ID NO: 13; o C302G/R con respecto a la SEQ ID NO: 14; o C172G/R con respecto a la SEQ ID NO: 15; o C304G/R con respecto a la SEQ ID NO: 16; o C173G/R con respecto a la SEQ ID NO: 17; o C304G/R con respecto a la SEQ ID NO: 18; o C173G/R con respecto a la SEQ ID NO: 19; o C293G/R con respecto a la SEQ ID NO: 20; o C173G/R con respecto a la SEQ ID NO: 21; o C282G/R con respecto a la SEQ ID NO: 22; o C173G/R con respecto a la SEQ ID NO: 23; o C304G/R con respecto a la SEQ ID NO: 24; o C174G/R con respecto a

la SEQ ID NO: 25; o C300G/R con respecto a la SEQ ID NO: 26; o C171G/R con respecto a la SEQ ID NO: 27; o C305G/R con respecto a la SEQ ID NO: 28; o C173G/R con respecto a la SEQ ID NO: 29; o C302G/R con respecto a la SEQ ID NO: 30; o C173G/R con respecto a la SEQ ID NO: 31; o C313G/R con respecto a la SEQ ID NO: 32; o C182G/R con respecto a la SEQ ID NO: 33; o C271G/R con respecto a la SEQ ID NO: 34; o C176G/R con respecto a la SEQ ID NO: 35.

La divulgación se refiere a composiciones que comprenden variantes de TdT que comprenden una secuencia de aminoácidos que tiene al menos un 90 por ciento de identidad, o en algunas realizaciones, al menos un 95 por ciento de identidad, al menos un 97 por ciento de identidad o, en algunas realizaciones, al menos un 98 por ciento de identidad o al menos un 99 % de identidad con respecto a una secuencia de TdT de referencia o de tipo silvestre seleccionada entre el grupo que consiste en las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 y 35, en donde dichas variantes de TdT tienen una mutación seleccionada entre C173G/R/P/A/V/S/N/Q/D, preferentemente C173G/R (en donde el resto de aminoácido es con respecto a la SEQ ID NO: 11, o un número de resto equivalente de las SEQ ID NO: 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35) (i) es capaz de sintetizar un fragmento de ácido nucleico sin una cadena molde y (ii) dichas variantes de TdT incorporan trifosfatos de nucleósidos modificados en 3'-O con una mayor eficiencia o a una mayor velocidad, que la TdT de referencia o de tipo silvestre.

La divulgación proporciona variantes truncadas de desoxinucleotidil transferasas terminales (TdT), comprendiendo cada una de ellas una secuencia de aminoácidos con al menos un 90 por ciento de identidad respecto de cualquiera de las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35 con al menos dos sustituciones de aminoácidos, preferentemente al menos tres sustituciones de aminoácidos, seleccionadas entre M63R/Q, L131P, C173G/R, R207L/N, D250V, R325P/N y E328N/L/T/S, (en donde los números de resto son con respecto a la SEQ ID NO: 11 o con respecto a sus números de resto funcionalmente equivalentes en las SEQ ID NO: 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35), (i) es capaz de sintetizar un fragmento de ácido nucleico sin una cadena molde y (ii) es capaz de incorporar un nucleótido modificado en el fragmento de ácido nucleico, por ejemplo, un trifosfato de desoxinucleósido bloqueado reversiblemente en 3'-O en un 3'-hidroxilo libre de un fragmento de ácido nucleico.

El valor de porcentaje de identidad de secuencia anterior es al menos un 95, 96, 97, 98 o 99 por ciento de identidad respecto de las secuencias especificadas.

Otro objeto de la invención es proporcionar una molécula de ácido nucleico que codifica una variante de una TdT como se ha definido anteriormente y/o un vector de expresión que comprende dicha molécula de ácido nucleico y/o una célula hospedadora que comprende dicha molécula de ácido nucleico o dicho vector de expresión.

Un objeto adicional de la invención es proporcionar un proceso para producir una variante de TdT de acuerdo con la invención, en donde se cultiva una célula hospedadora como se ha definido anteriormente en condiciones de cultivo que permiten la expresión del ácido nucleico que codifica dicha variante, y en donde la variante se recoge opcionalmente.

La invención se refiere además al uso de una variante de TdT, para sintetizar una molécula de ácido nucleico sin cadena molde, mediante la adición sucesiva de uno o más nucleótidos modificados en 3'-O a un fragmento de ácido nucleico. En algunas realizaciones, dichos métodos comprenden las etapas de (a) proporcionar un iniciador que comprende un oligonucleótido que tiene un 3'-hidroxilo libre; (b) hacer reaccionar en condiciones de extensión enzimática una variante de TdT de la invención con el iniciador o un iniciador extendido en presencia de un nucleósido bloqueado reversiblemente en 3'-O. En algunas realizaciones, dicho método incluye además las etapas de (c) desbloquear los iniciadores extendidos para formar iniciadores extendidos con 3'-hidroxilos libres y (d) repetir las etapas (b) y (c) hasta que se haya sintetizado una molécula de ácido nucleico con una secuencia predeterminada.

También es un objeto de la invención proporcionar un proceso para sintetizar una molécula de ácido nucleico sin cadena molde, que comprende una etapa de poner en contacto un cebador de ácido nucleico tanto con al menos un nucleótido, preferentemente al menos un nucleótido modificado en 3'-O y una variante de TdT de acuerdo con la invención.

La presente invención proporciona además un kit para realizar una reacción de incorporación de nucleótidos que comprende una variante de TdT de acuerdo con la invención y uno o más nucleótidos, preferentemente uno o más nucleótidos modificados en 3'-O y opcionalmente, al menos un cebador de ácido nucleico.

Descripción detallada de la invención

Las familias de ADN polimerasas se dividen en siete familias en función de su homología de secuencia y estructura cristalina. Entre ellas, las polimerasas de la familia PolX representan una gran variedad de polimerasas que abarcan desde polimerasas replicativas hasta enzimas transferasas terminales. Las polimerasas de la familia PolX están presentes en una gran variedad de organismos eucariotas. Las polimerasas de la familia PolX están implicadas en una gran variedad de procesos biológicos y concretamente, en los mecanismos de reparación del daño en el ADN o los mecanismos de corrección de errores. La familia PolX abarca la polimerasa β (Pol β), μ (Pol μ), λ (Pol λ), IV de levadura (Pol IV) y la desoxinucleotidil transferasa terminal (TdT). La TdT está implicada naturalmente en los

mecanismos de reparación y mantenimiento del ADN. Concretamente, la TdT tiene la capacidad única de conservar una actividad de polimerización de nucleótidos incluso en ausencia de una cadena molde. En condiciones específicas y con nucleótidos naturales, la TdT es capaz de alargar fragmentos de ADN con varios cientos de nucleótidos, en ausencia de cualquier hebra complementaria. Sin embargo, la TdT de tipo silvestre es completamente incapaz de incorporar eficazmente nucleótidos con el azúcar modificado.

Por tanto, el objetivo de la presente invención es proporcionar variantes de TdT con mutaciones específicas que les permiten incorporar nucleótidos modificados en un fragmento de ácido nucleico durante la síntesis de dicho fragmento de nucleótido. Más particularmente, los inventores han identificado restos de aminoácidos específicos que pueden sustituirse ventajosamente, solos o en combinación, para mejorar la capacidad de la enzima para sintetizar fragmentos de ácido nucleico de diferentes longitudes y con secuencias predeterminadas, lo que incluye el uso de nucleótidos modificados. La invención se define solo mediante las reivindicaciones.

Definiciones

Como se usan en el presente documento, los términos "*mutante*" y "*variante*" pueden utilizarse indistintamente para hacer referencia a polipéptidos relacionados con o procedentes de la SEQ ID NO: 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 o 35 y que comprenden una modificación o una alteración, es decir, una sustitución, inserción y/o eliminación, en una o más (p. ej., varias) posiciones y que tienen tanto actividad polimerasa sin cadena molde como la capacidad de incorporar trifosfatos de nucleótidos modificados en 3'-O en una cadena de ácido nucleico. Las variantes pueden obtenerse mediante diversas técnicas conocidas en la técnica. Concretamente, algunos ejemplos de técnicas para alterar la secuencia del ADN que codifica la proteína de tipo silvestre, incluyen, pero sin limitación, mutagénesis dirigida, mutagénesis aleatoria y construcción de oligonucleótidos sintéticos. Las actividades de la mutagénesis consisten en eliminar, insertar o sustituir uno o varios aminoácidos en la secuencia de una proteína o en el caso de la invención, de una polimerasa. Los aminoácidos diana pueden ser consecutivos o estar distribuidos a lo largo de toda la secuencia de la polimerasa. Por ejemplo, pueden usarse como diana motivos específicos o características estructurales.

Los términos "*modificación*" o "*alteración*", como se usan en el presente documento en relación con una posición o un aminoácido, significan que el aminoácido en la posición específica se ha modificado en comparación con el aminoácido de la proteína de tipo silvestre.

Una "*sustitución*" significa que un resto de aminoácido se reemplaza por otro resto de aminoácido. Preferentemente, el término "*sustitución*" se refiere al reemplazo de un resto de aminoácido por otro seleccionado entre los 20 restos de aminoácidos convencionales de origen natural, restos de aminoácidos raros de origen natural (p. ej., hidroxiprolina, hidroxilisina, alohidroxilisina, 6-N-metil-lisina, N-etilglicina, N-metilglicina, N-etilasparagina, aloisoleucina, N-metilisoleucina, N-metilvalina, piroglutamina, ácido aminobutírico, ornitina, norleucina, norvalina), y restos de aminoácidos de origen no natural, a menudo producidos sintéticamente, (p. ej., ciclohexil-alanina). Preferentemente, el término "*sustitución*" se refiere al reemplazo de un resto de aminoácido por otro seleccionado entre los 20 restos de aminoácidos convencionales de origen natural. El signo "+" indica una combinación de sustituciones.

Los aminoácidos se representan, en el presente documento, por su código de una letra o de tres letras de acuerdo con la siguiente nomenclatura: A: alanina (Ala); C: cisteína (Cys); D: ácido aspártico (Asp); E: ácido glutámico (Glu); F: fenilalanina (Phe); G: glicina (Gly); H: histidina (His); I: isoleucina (Ile); K: lisina (Lys); L: leucina (Leu); M: metionina (Met); N: asparagina (Asn); P: prolina (Pro); Q: glutamina (Gln); R: arginina (Arg); S: serina (Ser); T: treonina (Thr); V: valina (Val); W: triptófano (Trp) e Y: tirosina (Tyr).

En el presente documento, la siguiente terminología se usa para designar una sustitución: L238A indica que el resto de aminoácido (leucina, L) en la posición 238 de la secuencia precursora se cambia a una alanina (A). A132V/I/M significa que un resto de aminoácido (alanina, A) en la posición 132 de la secuencia precursora se cambia a uno de los siguientes aminoácidos: valina (V), isoleucina (I) o metionina (M). La sustitución puede ser una sustitución conservadora o no conservadora. Algunos ejemplos de sustituciones conservadoras se encuentran en los grupos de aminoácidos básicos (arginina, lisina e histidina), aminoácidos ácidos (ácido glutámico y ácido aspártico), aminoácidos polares (glutamina, asparagina y treonina), aminoácidos hidrófobos (metionina, leucina, isoleucina, cisteína y valina), aminoácidos aromáticos (fenilalanina, triptófano y tirosina) y aminoácidos pequeños (glicina, alanina y serina).

Como se usan en el presente documento, las expresiones "*identidad de secuencia*" o "*identidad*" se refieren al número (o fracción expresada como porcentaje, [%]) de coincidencias (restos de aminoácidos idénticos) entre dos secuencias de polipéptido. La identidad de secuencia se determina comparando las secuencias cuando están alineadas para maximizar la superposición y la identidad a la vez que se minimizan los huecos de secuencia. Concretamente, la identidad de secuencia se puede determinar utilizando cualquiera de una serie de algoritmos matemáticos de alineación globales o locales, dependiendo de la longitud de las dos secuencias. Las secuencias de longitudes similares se alinean preferentemente utilizando un algoritmo de alineación global (por ejemplo, el algoritmo de Needleman y Wunsch; Needleman y Wunsch, 1970) que alinea las secuencias de manera óptima en toda la longitud, mientras que las secuencias de longitudes sustancialmente diferentes se alinean preferentemente utilizando un algoritmo de alineación local (por ejemplo, el algoritmo de Smith y Waterman (Smith y Waterman, 1981) o el algoritmo

de Altschul (Altschul *et al.*, 1997; Altschul *et al.*, 2005)). La alineación para determinar el porcentaje de identidad de secuencia de aminoácidos puede lograrse de varias maneras que se encuentran dentro de las capacidades del experto en la materia, por ejemplo, utilizando el software informático disponible públicamente en sitios web de Internet tales como <http://blast.ncbi.nlm.nih.gov/> o <http://www.ebi.ac.uk/Tools/emboss/>. Los expertos en la materia pueden determinar los parámetros adecuados para medir la alineación, incluido cualquier algoritmo necesario para lograr un alineamiento máximo sobre la longitud completa de las secuencias que se están comparando. A efectos del presente documento, los valores de % de identidad de secuencia de aminoácido se refieren a valores generando el programa de alineamiento de secuencias por pares EMBOSS Needle, que crea un alineamiento global óptimo de dos secuencias utilizando el algoritmo de Needleman-Wunsch, en donde todos los parámetros de búsqueda se establecen en los valores predeterminados, es decir, matriz de puntuación = BLOSUM62, Hueco abierto = 10, Hueco extendido = 0,5, Penalización por hueco final = falso, Hueco final abierto = 10 y Hueco final extendido = 0,5.

En el presente documento, los términos "*péptido*", "*polipéptido*", "*proteína*", "*enzima*", se refieren a una cadena de aminoácidos unidos mediante enlaces peptídicos, independientemente del número de aminoácidos que forman dicha cadena.

Salvo que se indique lo contrario, las posiciones divulgadas en la presente solicitud se numeran por referencia a la secuencia de aminoácidos expuesta en una SEQ ID NO específica.

Variantes de TdT

La presente invención proporciona variantes de la enzima TdT que pueden utilizarse para sintetizar polinucleótidos de secuencias predeterminadas, tales como ADN o ARN, sin utilizar una cadena molde. Las variantes de TdT de la invención permiten utilizar nucleótidos modificados, y más concretamente, nucleótidos modificados en 3'-O, en un método enzimático de síntesis de polinucleótidos, tal como el descrito por Hiatt *et al.*, patente de EE. UU. 5763594.

En algunas realizaciones de la invención, "desoxirribonucleotidil transferasa terminal modificada", "TdT modificada", "variantes de desoxirribonucleotidil transferasa terminal" y "variantes de TdT" se refieren a enzimas que comparten al menos un 90 % de identidad con las SEQ ID NO: 2, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, exceptuando al menos las sustituciones de restos de aminoácidos en la posición 173 y 207. En otras realizaciones más, la variante de TdT comprende una secuencia de aminoácidos que comparte al menos un 95 % de identidad con las SEQ ID NO: 2, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, exceptuando al menos las sustituciones de aminoácidos en las posiciones 173 y 207. En otras realizaciones más, la variante de TdT comprende una secuencia de aminoácidos que comparte al menos un 97 %, 98 % o 99 % de identidad con las SEQ ID NO: 2, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, exceptuando al menos las sustituciones de aminoácidos en las posiciones 173 y 207.

En algunos casos, las variantes de la presente invención pueden describirse de acuerdo con sus mutaciones en restos específicos, cuyas posiciones se determinan mediante alineamiento con o en referencia a una SEQ ID NO.

Por "resto funcionalmente equivalente" se entiende un resto en una secuencia dada de una TdT que tiene un papel funcional idéntico en comparación con un resto correspondiente en una secuencia funcionalmente equivalente. En el contexto de la invención, cada secuencia de TdT seleccionada entre la SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35 puede considerarse una "secuencia funcionalmente equivalente" de una cualquiera de las otras secuencias.

En algunas realizaciones, la invención comprende una desoxinucleotidil transferasa terminal (TdT) que comprende una secuencia de aminoácidos que tiene al menos un 90 %, preferentemente

al menos un 95 %, 97 %, 98 % o 99 % de identidad con una secuencia de aminoácidos seleccionada entre las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35, con al menos una sustitución de aminoácidos en una posición correspondiente a un resto funcionalmente equivalente del resto C173 con respecto a la SEQ ID NO: 11, (i) es capaz de sintetizar un fragmento de ácido nucleico sin cadena molde y (ii) es capaz de incorporar un trifosfato de nucleósido modificado en 3'-O, tal como un trifosfato de nucleósido bloqueado en 3'-O, en el fragmento de ácido nucleico.

De hecho, los inventores han descubierto que dicha sustitución tiene un impacto significativo en las propiedades tanto superficiales como de interacción de la enzima con los nucleótidos, lo que puede permitir la incorporación de nucleótidos modificados en 3'-O en una secuencia de ácido nucleico.

Las realizaciones adicionales de variantes de TdT de la invención se enumeran como elementos en las tablas 1A a 1C (sustituciones únicas), las tablas 2A a 2C (dos sustituciones), las tablas 3A a 3C (tres sustituciones) y las tablas 4A a 4F (cuatro sustituciones), en donde cada una de dichas TdT variantes se define por la SEQ ID NO indicada en la columna de la derecha, modificada con las sustituciones indicadas en la columna de la izquierda de la misma fila que la SEQ ID NO. Una sustitución "no de tipo silvestre" significa que la sustitución puede ser cualquier aminoácido, excepto por el ácido nucleico en la posición indicada en la secuencia de tipo silvestre, o de modo equivalente, la secuencia de la SEQ ID NO indicada.

ES 2 982 551 T3

Tabla 1A: Variantes de TdT en la posición C173 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Sustitución no de tipo silvestre en	SEQ ID NO
C313	10
C173	11
C302	12
C173	13
C302	14
C172	15
C304	16
C173	17
C304	18
C173	19
C293	20
C173	21
C282	22
C178	23
C304	24
C174	25
C300	26
C171	27
C305	28
C173	29
C302	30
C173	31
C313	32
C182	33
C271	34
C176	35

Tabla 1B: Variantes de TdT adicionales en la posición C173 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

5

Sustitución	SEQ ID NO
C313/G/R/P/A/V/S/N/Q/D	10
C173/G/R/P/A/V/S/N/Q/D	11
C302/G/R/P/A/V/S/N/Q/D	12
C173/G/R/P/A/V/S/N/Q/D	13
C302/G/R/P/A/V/S/N/Q/D	14
C172/G/R/P/A/V/S/N/Q/D	15
C304/G/R/P/A/V/S/N/Q/D	16
C173/G/R/P/A/V/S/N/Q/D	17
C304/G/R/P/A/V/S/N/Q/D	18
C173/G/R/P/A/V/S/N/Q/D	19
C293/G/R/P/A/V/S/N/Q/D	20
C173/G/R/P/A/V/S/N/Q/D	21
C282/G/R/P/A/V/S/N/Q/D	22
C178/G/R/P/A/V/S/N/Q/D	23
C304/G/R/P/A/V/S/N/Q/D	24
C174/G/R/P/A/V/S/N/Q/D	25
C300/G/R/P/A/V/S/N/Q/D	26
C171/G/R/P/A/V/S/N/Q/D	27
C305/G/R/P/A/V/S/N/Q/D	28
C173/G/R/P/A/V/S/N/Q/D	29
C302/G/R/P/A/V/S/N/Q/D	30
C173/G/R/P/A/V/S/N/Q/D	31
C313/G/R/P/A/V/S/N/Q/D	32
C182/G/R/P/A/V/S/N/Q/D	33
C271/G/R/P/A/V/S/N/Q/D	34
C176/G/R/P/A/V/S/N/Q/D	35

ES 2 982 551 T3

Tabla 1C: Variantes de TdT adicionales en la posición C173 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Sustituciones	SEQ ID NO
C313/G/R	10
C173/G/R	11
C302/G/R	12
C173/G/R	13
C302/G/R	14
C172/G/R	15
C304/G/R	16
C173/G/R	17
C304/G/R	18
C173/G/R	19
C293/G/R	20
C173/G/R	21
C282/G/R	22
C178/G/R	23
C304/G/R	24
C174/G/R	25
C300/G/R	26
C171/G/R	27
C305/G/R	28
C173/G/R	29
C302/G/R	30
C173/G/R	31
C313/G/R	32
C182/G/R	33
C271/G/R	34
C176/G/R	35

5 Tabla 2A: Variantes de TdT adicionales en la posición C173 (SEQ ID NO: 11) y la posición M63 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Sustituciones no de tipo silvestre en las ubicaciones	SEQ ID NO
M63 + C173	11
M63 + C173	13
L62 + C172	15
M63 + C173	17
M63 + C173	19
R64 + C173	21
M73 + C178	23
M64 + C174	25
M61 + C171	27
M63 + C173	29
L63 + C173	31
M63 + C182	33
M66 + C176	35

Tabla 2B: Variantes de TdT adicionales en la posición C173 (SEQ ID NO: 11) y la posición M63 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Sustituciones y posiciones de las sustituciones	SEQ ID NO
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D	11
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D	13
L62R/Q/G/A/V/D/N/H/E + C172G/R/P/A/V/S/N/Q/D	15
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D	17
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D	19
R64R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D	21
M73R/Q/G/A/V/D/N/H/E + C178G/R/P/A/V/S/N/Q/D	23
M64R/Q/G/A/V/D/N/H/E + C174G/R/P/A/V/S/N/Q/D	25
M61R/Q/G/A/V/D/N/H/E + C171G/R/P/A/V/S/N/Q/D	27
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D	29
L63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D	31
M63R/Q/G/A/V/D/N/H/E + C182G/R/P/A/V/S/N/Q/D	33
M66R/Q/G/A/V/D/N/H/E + C176G/R/P/A/V/S/N/Q/D	35

Tabla 2C: Variantes de TdT adicionales en la posición C173 (SEQ ID NO: 11) y la posición M63 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Sustituciones y posiciones de las sustituciones	SEQ ID NO
M63R/Q + C173G/R	11
M63R/Q + C173G/R	13
L62R/Q + C172G/R	15
M63R/Q + C173G/R	17
M63R/Q + C173G/R	19
R64R/Q + C173G/R	21
M73R/Q + C178G/R	23
M64R/Q + C174G/R	25
M61R/Q + C171G/R	27
M63R/Q + C173G/R	29
L63R/Q + C173G/R	31
M63R/Q + C182G/R	33
M66R/Q + C176G/R	35

5 Tabla 3A: Variantes de TdT adicionales en las posiciones C173 (SEQ ID NO: 11), M63 (SEQ ID NO: 11) y R207 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Mutaciones	SEQ ID NO
M63 + C173 + R207	11
M63 + C173 + R207	13
L62 + C172 + R206	15
M63 + C173 + R207	17
M63 + C173 + R207	19
R64 + C173 + R208	21
M73 + C178 + R207	23
M64 + C174 + R208	25
M61 + C171 + R205	27
M63 + C173 + R207	29
L63 + C173 + R207	31
M63 + C182 + R216	33
M66 + C176 + R210	35

Tabla 3B: Variantes de TdT adicionales en las posiciones C173 (SEQ ID NO: 11), M63 (SEQ ID NO: 11) y R207 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Mutaciones	SEQ ID NO
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P	11
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P	13
L62R/Q/G/A/V/D/N/H/E + C172G/R/P/A/V/S/N/Q/D + R206 N/L/K/H/G/D/A/P	15
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P	17
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P	19
R64Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R208 N/L/K/H/G/D/A/P	21
M73R/Q/G/A/V/D/N/H/E + C178G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P	23
M64R/Q/G/A/V/D/N/H/E + C174G/R/P/A/V/S/N/Q/D + R208 N/L/K/H/G/D/A/P	25
M61R/Q/G/A/V/D/N/H/E + C171G/R/P/A/V/S/N/Q/D + R205 N/L/K/H/G/D/A/P	27
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P	29
L63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207N/L/K/H/G/D/A/P	31
M63R/Q/G/A/V/D/N/H/E + C182G/R/P/A/V/S/N/Q/D + R216N/L/K/H/G/D/A/P	33
M66R/Q/G/A/V/D/N/H/E + C176G/R/P/A/V/S/N/Q/D + R210N/L/K/H/G/D/A/P	35

10 Tabla 3C: Variantes de TdT adicionales en las posiciones C173 (SEQ ID NO: 11), M63 (SEQ ID NO: 11) y R207 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Mutaciones	SEQ ID NO
M63R/Q + C173G/R + R207L/N	11
M63R/Q + C173G/R + R207L/N	13
M62R/Q + C172G/R + R206L/N	15
M63R/Q + C173G/R + R207L/N	17
M63R/Q + C173G/R + R207L/N	19
R64Q + C173G/R + R208L/N	21
M73R/Q + C178G/R + R207N/L	23

(continuación)

Mutaciones	SEQ ID NO
M64R/Q + C 174G/R + R208 N/L	25
M61R/Q + C171G/R + R205N/L	27
M63R/Q + C173G/R + R207L/N	29
L63R/Q + C173G/R + R207L/N	31
M63R/Q + C182G/R + R216N/L	33
M66R/Q + C 176G/R + R210N/L	35

Tabla 4A: Variantes de TdT adicionales en las posiciones C173 (SEQ ID NO: 11), M63 (SEQ ID NO: 11), R207 (SEQ ID NO: 11) y R324 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Mutaciones	SEQ ID NO
M63 + C173 + R207 + R324	11
M63 + C173 + R207 + R324	13
L62 + C172 + R206 + R320	15
M63 + C173 + R207 + R331	17
M63 + C173 + R207 + P325	19
R64 + C173 + R208 + T331	21
M73 + C178 + R207 + R325	23
M64 + C174 + R208 + P326	25
M61 + C171 + R205 + R323	27
M63 + C173 + R207 + R328	29
L63 + C173 + R207 + R325	31
M63 + C182 + R216 + R338	33
M66 + C176 + R210 + R328	35

5

Tabla 4B: Variantes de TdT adicionales en las posiciones C173 (SEQ ID NO: 11), M63 (SEQ ID NO: 11), R207 (SEQ ID NO: 11) y R324 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Mutaciones	SEQ ID NO
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P + R324P/N/A/L/K/H/G/D	11
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P + R324P/N/A/L/K/H/G/D	13
L62R/Q/G/A/V/D/N/H/E + C172G/R/P/A/V/S/N/Q/D + R206 N/L/K/H/G/D/A/P + R320P/N/A/L/K/H/G/D	15
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P + R331P/N/A/L/K/H/G/D	17
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P + P325N/A/L/K/H/G/D	19
R64Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R208 N/L/K/H/G/D/A/P + T331P/N/A/L/K/H/G/D	21
M73R/Q/G/A/V/D/N/H/E + C178G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P + R325P/N/A/L/K/H/G/D	23
M64R/Q/G/A/V/D/N/H/E + C174G/R/P/A/V/S/N/Q/D + R208 N/L/K/H/G/D/A/P + P326N/A/L/K/H/G/D	25
M61R/Q/G/A/V/D/N/H/E + C171G/R/P/A/V/S/N/Q/D + R205 N/L/K/H/G/D/A/P + R323P/N/A/L/K/H/G/D	27
M63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207 N/L/K/H/G/D/A/P + R328P/N/A/L/K/H/G/D	29
L63R/Q/G/A/V/D/N/H/E + C173G/R/P/A/V/S/N/Q/D + R207N/L/K/H/G/D/A/P + R325P/N/A/L/K/H/G/D	31
M63R/Q/G/A/V/D/N/H/E + C182G/R/P/A/V/S/N/Q/D + R216N/L/K/H/G/D/A/P + R338P/N/A/L/K/H/G/D	33
M66R/Q/G/A/V/D/N/H/E + C176G/R/P/A/V/S/N/Q/D + R210N/L/K/H/G/D/A/P + R328P/N/A/L/K/H/G/D	35

10

Tabla 4C: Variantes de TdT adicionales en las posiciones C173 (SEQ ID NO: 11), M63 (SEQ ID NO: 11), R207 (SEQ ID NO: 11) y R324 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Mutaciones	SEQ ID NO
M63R/Q + C173G/R + R207N/L + R324P/N	11
M63R/Q + C173G/R + R207N/L + R324P/N	13
L62R/Q + C172G/R + R206N/L + R320P/N	15
M63R/Q + C173G/R + R207N/L + R331P/N	17
M63R/Q + C173G/R + R207N/L + P325N	19
R64Q/G + C173G/R + R208N/L + T331P/N	21

(continuación)

Mutaciones	SEQ ID NO
M73R/Q/G + C178G/R + R207N/L + R325P/N	23
M64R/Q + C174G/R + R208N/L + P326N	25
M61R/Q + C171G/R + R205N/L + R323P/N	27
M63R/Q + C173G/R + R207N/L + R328P/N	29
L63R/Q + C173G/R + R207N/L + R325P/N	31
M63R/Q + C182G/R + R216N/L + R338P/N	33
M66R/Q + C176G/R + R210N/L + R328P/N	35

Tabla 4D: Variantes de TdT adicionales en las posiciones C173 (SEQ ID NO: 11), M63 (SEQ ID NO: 11), R207 (SEQ ID NO: 11) y E327 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Mutaciones	SEQ ID NO
M63 + C173 + R207 + E327	11
M63 + C173 + R207 + E327	13
L62 + C172 + R206 + G323	15
M63 + C173 + R207 + E334	17
M63 + C173 + R207 + E327	19
R64 + C173 + R208 + E334	21
M73 + C178 + R207 + E328	23
M64 + C174 + R208 + E329	25
M61 + C171 + R205 + E326	27
M63 + C173 + R207 + E331	29
L63 + C173 + R207 + E328	31
M63 + C182 + R216 + E341	33
M66 + C176 + R210 + E331	35

5

Tabla 4E: Variantes de TdT adicionales en las posiciones C173 (SEQ ID NO: 11), M63 (SEQ ID NO: 11), R207 (SEQ ID NO: 11) y E327 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Mutaciones	SEQ ID NO
M63R/Q/G/AN/D/N/H/E + C173G/R/P/AV/S/N/Q/D + R207 N/L/K/H/G/D/A/P + E327N/L/T/S	11
M63R/Q/G/AN/D/N/H/E + C173G/R/P/AV/S/N/Q/D + R207 N/L/K/H/G/D/A/P + E327N/L/T/S	13
L62R/Q/G/AV/D/N/H/E + C172G/R/P/AV/S/N/Q/D + R206 N/L/K/H/G/D/A/P + G323N/L/T/S	15
M63R/Q/G/AN/D/N/H/E + C173G/R/P/AV/S/N/Q/D + R207 N/L/K/H/G/D/A/P + E334N/L/T/S	17
M63R/Q/G/AN/D/N/H/E + C173G/R/P/AV/S/N/Q/D + R207 N/L/K/H/G/D/A/P + E327N/L/T/S	19
R64Q/G/AV/D/N/H/E + C173G/R/P/AV/S/N/Q/D + R208 N/L/K/H/G/D/A/P + E334N/L/T/S	21
M73R/Q/G/AN/D/N/H/E + C178G/R/P/AV/S/N/Q/D + R207 N/L/K/H/G/D/A/P + E328N/L/T/S	23
M64R/Q/G/AV/D/N/H/E + C174G/R/P/AV/S/N/Q/D + R208 N/L/K/H/G/D/A/P + E329N/L/T/S	25
M61R/Q/G/AV/D/N/H/E + C171G/R/P/AV/S/N/Q/D + R205 N/L/K/H/G/D/A/P + E326N/L/T/S	27
M63R/Q/G/AV/D/N/H/E + C173G/R/P/AV/S/N/Q/D + R207 N/L/K/H/G/D/A/P + E331N/L/T/S	29
L63R/Q/G/AV/D/N/H/E + C173G/R/P/AV/S/N/Q/D + R207N/L/K/H/G/D/A/P + E328N/L/T/S	31
M63R/Q/G/AV/D/N/H/E + C182G/R/P/AV/S/N/Q/D + R216N/L/K/H/G/D/A/P + E341N/L/T/S	33
M66R/Q/G/AV/D/N/H/E + C176G/R/P/AV/S/N/Q/D + R210N/L/K/H/G/D/A/P + E331N/L/T/S	35

10

Tabla 4F: Variantes de TdT adicionales en las posiciones C173 (SEQ ID NO: 11), M63 (SEQ ID NO: 11), R207 (SEQ ID NO: 11) y E327 (SEQ ID NO: 11) o posiciones funcionalmente equivalentes de la SEQ ID NO indicada

Mutaciones	SEQ ID NO
M63R/Q + C173G/R + R207 N/L + E327N/L/T/S	11
M63R/Q + C173G/R + R207N/L + E327N/L/T/S	13
L62R/Q + C172G/R + R206N/L + G323N/L/T/S	15
M63R/Q + C173G/R + R207N/L + E334N/L/T/S	17
M63R/Q + C173G/R + R207N/L + E327N/L/T/S	19
R64Q/G + C173G/R + R208N/L + E334N/L/T/S	21
M73R/Q + C178G/R + R207N/L + E328N/L/T/S	23
M64R/Q + C174G/R + R208N/L + E329N/L/T/S	25
M61R/Q + C171G/R + R205N/L + E326N/L/T/S	27
M63R/Q/G + C173G/R + R207N/L + E331N/L/T/S	29
L63R/Q + C173G/R + R207N/L + E328N/L/T/S	31
M63R/Q + C182G/R + R216N/L + E341N/L/T/S	33
M66R/Q + C176G/R + R210N/L + E331N/L/T/S	35

La sustitución se selecciona entre CzzzG/R/P/AV/S/N/Q/D, donde Czzz representa un número de resto de aminoácido funcionalmente equivalente a C173 de la SEQ ID NO: 11 en las SEQ ID NO: 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33

y 35, respectivamente, y preferentemente de CzzzG/R, donde Czzz representa un número de resto de aminoácido funcionalmente equivalente a C173 de la SEQ ID NO: 11 en las SEQ ID NO: 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 y 35, respectivamente.

- 5 La variante comprende además al menos una sustitución de aminoácidos en una posición correspondiente a restos funcionalmente equivalentes a los restos seleccionados entre M63, R207, R324 y E327, de la SEQ ID NO: 11.

De acuerdo con la invención, todas las variantes de TdT como se han divulgado anteriormente son capaces tanto de sintetizar un fragmento de ácido nucleico sin una cadena molde como de incorporar un nucleótido modificado en el fragmento de ácido nucleico. Ventajosamente, dichas variantes tienen una mayor capacidad de incorporar un nucleótido modificado, preferiblemente un nucleótido modificado en 3'O, en un fragmento de ácido nucleico en comparación con una TdT de las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 y 35. Cabe destacar que el alcance de la invención se define únicamente por las reivindicaciones.

- 15 En algunas de las realizaciones descritas anteriormente, la eficiencia de una TdT variante para incorporar un trifosfato de nucleósido modificado en 3'O es al menos un 110 por ciento la de una TdT de tipo silvestre con la secuencia de la SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 y 35; en otras realizaciones, la eficiencia de una TdT variante para incorporar un trifosfato de nucleósido modificado en 3'O es al menos un 150 por ciento la de una TdT de tipo silvestre con la secuencia de la SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 y 35; en otras realizaciones, la eficiencia de una TdT variante para incorporar un trifosfato de nucleósido modificado en 3'O es al menos un 200 por ciento la de una TdT de tipo silvestre con la secuencia de la SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 y 35.

Las variantes de la invención comprenden la secuencia de aminoácidos de las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35 y opcionalmente fragmentos de aminoácidos en los extremos carboxílico o amínico. En otra realización, las variantes de la invención consisten únicamente en la secuencia de aminoácidos de las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35.

La variante de desoxinucleotidil transferasa terminal (TdT) comprende una secuencia de aminoácidos como se expone en las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35 o una secuencia de aminoácidos que tiene un 90 % de identidad de secuencia, preferentemente un 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia con cualquiera de las secuencias anteriores, con al menos tres sustituciones de aminoácidos seleccionadas entre M63R/Q, L131P, C173G/R, R207L/N, D250V, R324P/N y E327N/L/T/S, o un resto funcionalmente equivalente, en donde las posiciones se numeran por referencia a la secuencia de aminoácidos expuesta en la SEQ ID NO: 11 o como se expone directamente en otras partes del presente documento con respecto a sus SEQ ID NO individuales, (i) es capaz de sintetizar un fragmento de ácido nucleico sin cadena molde y (ii) es capaz de incorporar un nucleótido modificado en 3'-O en un fragmento de ácido nucleico.

Por ejemplo, la variante de TdT comprende una secuencia de aminoácidos con un porcentaje especificado de identidad de secuencia con la SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 y 35 y una combinación de sustituciones seleccionadas entre M63R + L131P + R207L, M63R + L131P + R207N, M63R + L131P + D250V, M63R + L131P + R324P, M63R + L131P + R324A, M63R + L131P + E327L, M63R + L131P + E327N, M63R + R207L + D250V, M63R + R207L + R324P, M63R + R207L + R324A, M63R + R207L + E327L, M63R + R207L + E327N, M63R + R207N + D250V, M63R + R207N + R324P, M63R + R207N + R324A, M63R + R207N + E327L, M63R + R207N + E327N, M63R + D250V + R324P, M63R + D250V + R324A, M63R + R324P + E327L, M63R + R324P + E327N, M63R + R324A + E327L, M63R + R324A + E327N, M63Q + L131P + R207L, M63Q + L131P + R207N, M63Q + L131P + D250V, M63Q + L131P + R324P, M63Q + L131P + R324A, M63Q + L131P + E327L, M63Q + L131P + E327N, M63Q + R207L + D250V, M63Q + R207L + R324P, M63Q + R207L + R324A, M63Q + R207L + E327L, M63Q + R207L + E327N, M63Q + D250V + R324P, M63Q + D250V + R324A, M63Q + D250V + E327L, M63Q + D250V + E327N, M63Q + R324P + E327L, M63Q + R324P + E327N, L131P + R207L + D250V, L131P + R207L + R324A, L131P + R207L + E327L, L131P + R207L + E327N, L131P + R207N + D250V, L131P + R207N + R324P, L131P + R207N + R324A, L131P + R207N + E327L, L131P + R207N + E327N, L131P + D250V + R324P, L131P + D250V + R324A, L131P + D250V + E327L, L131P + D250V + E327N, L131P + R324P + E327L, L131P + R324P + E327N, L131P + R324A + E327L, L131P + R324A + E327N, R207L + D250V + R324P, R207L + D250V + R324A, R207L + D250V + E327L, R207L + D250V + E327N, R207L + R324P + E327L, R207L + R324P + E327N, R207L + R324A + E327L, R207L + R324A + E327N, R207N + D250V + R324P, R207N + D250V + R324A, R207N + D250V + E327L, R207N + E327N + R324P + E327L, R207N + R324P + E327N, R207N + R324A + E327L, R207N + R324A + E327N, D250V + R324P + E327L, D250V + R324P + E327N, D250V + R324A + E327L, D250V + R324A + E327N y R207L + D250V + R324P, o uno o más restos funcionalmente equivalentes, en donde los números de posición anteriores son con respecto a la SEQ ID NO: 11.

La variante de TdT comprende una secuencia de aminoácidos con un porcentaje de identidad de secuencia respecto de la SEQ ID NO: 11, o una secuencia funcionalmente equivalente, con la combinación de sustituciones R207L + R324P + E327L, o restos funcionalmente equivalentes.

La variante de TdT comprende una secuencia de aminoácidos con un porcentaje de identidad de secuencia respecto

de la SEQ ID NO: 11, o una secuencia funcionalmente equivalente, con la combinación de sustituciones R207N + R324A + E327N, o restos funcionalmente equivalentes. Dicha variante puede comprender además al menos una sustitución en la posición correspondiente a los restos seleccionados entre L52, A108, L131, T340, G284, H287, E289, W450, R354 y A510, con respecto a la SEQ ID NO: 11, o uno o más restos funcionalmente equivalentes en una secuencia funcionalmente equivalente. Como se ha expuesto anteriormente, dicha variante también puede comprender la combinación de mutaciones constantes L52F+A108V+R354K y/o G284L/S+H287D+E289A, con respecto a la SEQ ID NO:11, o uno o más restos funcionalmente equivalentes en una secuencia funcionalmente equivalente.

La divulgación proporciona una variante de didesoxinucleotidil transferasa terminal (TdT) que comprende una secuencia de aminoácidos con un porcentaje de identidad de secuencia especificado respecto de la SEQ ID NO: 11 o una secuencia funcionalmente equivalente, con al menos una sustitución de aminoácidos seleccionada entre M63R, M63Q, L131P, R207L, R207N, D250V, R324P, R324A, E327L, E327N, o uno o más restos funcionalmente equivalentes, (i) es capaz de sintetizar un fragmento de ácido nucleico sin una cadena molde y (ii) es capaz de incorporar un nucleótido modificado en 3'-O en un fragmento de ácido nucleico.

La divulgación proporciona una variante de didesoxinucleotidil transferasa terminal (TdT) que comprende una secuencia de aminoácidos con un porcentaje de identidad de secuencia especificado respecto de la SEQ ID NO: 11 o una secuencia funcionalmente equivalente, con al menos la combinación de sustituciones seleccionadas entre M63R + L131P, M63R + R207L, M63R + R207N, M63R + D250V, M63R + R324P, M63R + R324A, M63R + E327L, M63R + E327N, M63Q + L131P, M63Q + R207L, M63Q + R207N, M63Q + D250V, M63Q + R324P, M63Q + R324A, M63Q + E327L, M63Q + E327N, L131P + R207L, L131P + R207N, L131P + D250V, L131P + R324P, L131P + R324A, L131P + E327L, L131P + E327N, R207L + D250V, R207L + R324P, R207L + R324A, R207L + E327L, R207L + E327N, R207N + D250V, R207N + R324P, R207N + R324A, R207N + E327L, R207N + E327N, D250V + R324P, D250V + R324A, D250V + E327L, D250V + E327N, R324P + E327L, R324P + E327N, R324A + E327L and R324A + E327N, o uno o más restos funcionalmente equivalentes, en donde las posiciones se numeran por referencia a la secuencia de aminoácidos expuesta en la SEQ ID NO: 11, (i) es capaz de sintetizar un fragmento de ácido nucleico sin una cadena molde y (ii) es capaz de incorporar un nucleótido modificado en 3'-O en un fragmento de ácido nucleico.

Una variante de TdT tiene una sustitución o combinación de sustituciones descritas anteriormente y tiene una secuencia de aminoácidos con al menos un 80 % de identidad respecto de las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35; en algunas realizaciones, dicha secuencia de aminoácidos tiene al menos un 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad respecto de las SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35.

Modificaciones adicionales

En una realización, la variante de TdT incluye además cualquier tipo de péptido marcador en su extremo amínico, en su extremo carboxílico o en ambos extremos, tal como la secuencia de marcador His. Dicho péptido marcador puede usarse para purificación, identificación, aumentar la expresión, la capacidad de secreción o aumentar la actividad catalítica. Se entenderá que dichos diferentes marcadores se describen ampliamente en la bibliografía y por tanto, la presente invención abarca todos los marcadores conocidos por un experto en la materia.

Las variantes de la invención también pueden incluir una o más características exógenas o heterólogas en las regiones del extremo amínico y/o carboxílico de la proteína para su uso, p. ej., en la purificación de la polimerasa recombinante.

La variante de la invención puede comprender además una sustitución de restos entre las posiciones C378 a L406, en donde las posiciones se numeran por referencia a la secuencia de aminoácidos expuesta en la SEQ ID NO: 1 o restos funcionalmente equivalentes, por los restos H363 a C390 de la polimerasa Pol μ con la secuencia de SEQ ID NO: 3, en donde las posiciones se numeran por referencia a la secuencia de aminoácidos expuesta en la SEQ ID NO: 3 o restos funcionalmente equivalentes.

Ventajosamente, la variante de TdT comprende al menos la secuencia de aminoácidos de SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35, con las sustituciones y los valores de porcentaje de identidad de secuencia divulgados.

Ácidos nucleicos, casete de expresión, vector

También es el objetivo de la invención proporcionar una molécula de ácido nucleico que codifica una variante de la invención. Como se usan en el presente documento, las expresiones "ácido nucleico", "secuencia de ácido nucleico", "polinucleótido", "oligonucleótido" y "secuencia de nucleótidos" se usan indistintamente y se refieren a una secuencia de desoxirribonucleótidos y/o ribonucleótidos. En una realización, el ácido nucleico es un ADN. En una realización alternativa, el ácido nucleico es RNA. En una realización alternativa, el ácido nucleico es XNA.

Los ácidos nucleicos pueden estar en forma monocatenaria o en forma dúplex o una mezcla de ambas. Puede ser de origen recombinante, artificial y/o sintético y puede comprender nucleótidos modificados. Dichas modificaciones

pueden ser modificaciones naturales, tales como modificaciones epigenéticas, o modificaciones no naturales, tales como marcadores, enlaces modificados, una base de purina o pirimidina modificada o un azúcar modificado. En una realización, las moléculas de ácido nucleico son ADN, ARN o XNA que portan modificaciones epigenéticas de origen natural, tales como metilación, hidroximetilación, formilación o 5-carboxilación. En una realización, las moléculas de ácido nucleico son ADN, ARN o XNA que portan modificaciones de origen no natural, tales como una etiqueta fluorescente, marcador fluorescente, grupos de interacción.

Los ácidos nucleicos de la invención pueden estar en forma aislada o purificada, y producirse, aislarse y/o manipularse mediante técnicas en sí conocidas en la técnica, p. ej., clonación y expresión de bibliotecas de ADNc, amplificación, síntesis enzimática o tecnología recombinante. Los ácidos nucleicos también pueden sintetizarse *in vitro* mediante técnicas de síntesis química bien conocidas, como se describe en, p. ej., Belousov (1997) *Nucleic Acids Res.* 25:3440-3444.

La invención también abarca ácidos nucleicos que hibridan, en condiciones rigurosas, con un ácido nucleico que codifica una variante de TdT como se ha definido anteriormente. Preferentemente, dichas condiciones rigurosas incluyen incubaciones de filtros de hibridación a aproximadamente 42 °C durante aproximadamente 2,5 horas en 2 X SSC/SDS al 0,1 %, seguida del lavado de los filtros cuatro veces de 15 minutos en 1 X SSC/SDS al 0,1 % a 65 °C. Los protocolos utilizados se describen en referencias, como en Sambrook *et al.* (*Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor N.Y. (1988)) y Ausubel (*Current Protocols in Molecular Biology* (1989)).

La invención también abarca ácidos nucleicos que codifican una variante de TdT de la invención, en donde la secuencia de dichos ácidos nucleicos, o al menos una porción de dicha secuencia, se ha modificado utilizando un uso de codones optimizado.

Como alternativa, los ácidos nucleicos de acuerdo con la invención pueden deducirse de la secuencia de la variante de TdT de acuerdo con la invención y el uso de codones puede adaptarse en función de la célula hospedadora en la que se transcribirán los ácidos nucleicos. Estas etapas pueden llevarse a cabo de acuerdo con métodos bien conocidos por un experto en la materia y algunos se describen en el manual de referencia Sambrook *et al.* (Sambrook *et al.*, 2001).

En una realización, las moléculas de ácido nucleico son moléculas poliméricas que tienen una longitud de más de 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 15000, 20000, 30000, 40000, 50000 o 100000 nucleótidos.

Los ácidos nucleicos de la invención pueden comprender además secuencias de nucleótidos adicionales, tales como regiones reguladoras, es decir, promotores, potenciadores, silenciadores, terminadores, péptidos de señal y similares que pueden utilizarse para provocar o regular la expresión del polipéptido en una célula hospedadora o sistema seleccionado.

La presente invención se refiere además a un casete de expresión que comprende un ácido nucleico de acuerdo con la invención unido operativamente a una o más secuencias de control que dirigen la expresión de dicho ácido nucleico en una célula hospedadora adecuada. Típicamente, el casete de expresión comprende, o consiste en, un ácido nucleico de acuerdo con la invención unido operativamente a una secuencia de control, tal como un promotor transcripcional y/o un terminador de la transcripción. La secuencia de control puede incluir un promotor que es reconocido por una célula hospedadora o un sistema de expresión *in vitro* para la expresión de un ácido nucleico que codifica una variante de TdT de la presente invención. El promotor contiene secuencias de control de la transcripción que median en la expresión de la enzima. El promotor puede ser cualquier polinucleótido que muestra actividad transcripcional en la célula hospedadora, incluyendo promotores mutantes, truncados e híbridos, y puede obtenerse a partir de genes que codifican polipéptidos extracelulares o intracelulares, bien homólogos o heterólogos de la célula hospedadora. La secuencia de control también puede ser un terminador de la transcripción, que es reconocido por una célula hospedadora para terminar la transcripción. El terminador está unido operativamente al extremo 3' del ácido nucleico que codifica la esterasa. En la presente invención puede usarse cualquier terminador que sea funcional en la célula hospedadora. Típicamente, el casete de expresión comprende, o consiste en, un ácido nucleico de acuerdo con la invención unido operativamente a un promotor transcripcional y un terminador de la transcripción.

La invención también se refiere a un vector que comprende un ácido nucleico o un casete de expresión como se ha definido anteriormente.

El término "*vector*" se refiere a una molécula de ADN usada como vehículo para transferir material genético recombinante en una célula hospedadora. Los principales tipos de vectores son plásmidos, bacteriófagos, virus, cósmidos y cromosomas artificiales. El vector en sí es generalmente una secuencia de ADN que consiste en un fragmento (una secuencia de ácido nucleico heteróloga, transgén) y una secuencia de mayor tamaño que sirve de "armazón" del vector. El propósito de un vector que transfiere información genética al hospedador es, por lo general, aislar, multiplicar o expresar el fragmento en la célula diana. Los vectores denominados vectores de expresión (construcciones de expresión) se adaptan específicamente para la expresión de las secuencias heterólogas en la

célula diana y, en general, tienen una secuencia promotora que dirige la expresión de las secuencias heterólogas que codifican un polipéptido. En general, los elementos reguladores que están presentes en un vector de expresión incluyen un promotor transcripcional, un sitio de unión al ribosoma, un terminador y un operador opcionalmente presente. Preferentemente, un vector de expresión también contiene un origen de replicación para la replicación autónoma en una célula hospedadora, un marcador de selección, un número limitado de sitios útiles de enzimas de restricción y un potencial para un alto número de copias. Algunos ejemplos de vectores de expresión son vectores de clonación, vectores de clonación modificados, plásmidos y virus diseñados específicamente. Los vectores de expresión que proporcionan niveles adecuados de expresión de polipéptidos en diferentes hospedadores son bien conocidos en la materia. La elección del vector normalmente dependerá de la compatibilidad del vector con la célula hospedadora en la que ha de introducirse el vector.

Otro objeto de la invención es proporcionar una célula hospedadora que comprende un ácido nucleico, un casete de expresión o un vector como se ha descrito anteriormente. La presente invención se refiere, por lo tanto, al uso de un ácido nucleico, casete de expresión o vector de acuerdo con la invención para transformar, transfectar o transducir una célula hospedadora. La elección del vector normalmente dependerá de la compatibilidad del vector con la célula hospedadora en la que se debe introducir.

De acuerdo con la invención, la célula hospedadora puede transformarse, transfectarse o transducirse de una manera transitoria o estable. El casete o vector de expresión de la invención se introduce en una célula hospedadora de modo que el casete o vector se mantiene como un integrante cromosómico o como un vector extracromosómico autorreplicante. La expresión "célula hospedadora" también abarca cualquier descendencia de una célula hospedadora precursora que no es idéntica a la célula hospedadora precursora debido a mutaciones que se producen durante la replicación. La célula hospedadora puede ser cualquier célula útil en la producción de una variante de la presente invención, p. ej., un procarionta o un eucariota. La célula hospedadora procarionta puede ser cualquier bacteria grampositiva o gramnegativa. La célula hospedadora también puede ser una célula eucariota, tales como células de levadura, fúngicas, de mamífero, de insecto o vegetales.

El ácido nucleico, el casete de expresión o el vector de expresión de acuerdo con la invención pueden introducirse en la célula hospedadora mediante cualquier método conocido por el experto en la materia, tal como electroporación, conjugación, transducción, transformación en células competentes, transformación de protoplastos, fusión de protoplastos, transformación biolística con "pistola génica", transformación mediada por PEG, transformación o transfección asistida por lípidos, transfección mediada químicamente, transformación mediada por acetato de litio, transformación mediada por liposomas.

Opcionalmente, más de una copia de un ácido nucleico, casete o vector de la presente invención puede insertarse en una célula hospedadora para aumentar la producción de la variante.

Nucleótidos modificados

De acuerdo con la invención, las variantes de TdT pueden incorporar nucleótidos modificados, preferentemente nucleótidos modificados en 3'O y más preferentemente, nucleótidos bloqueados en 3'-O.

En el contexto de la invención, la expresión "nucleótido modificado" se refiere a una molécula que contiene un nucleósido (es decir, una base unida a una molécula de azúcar de desoxirribosa o ribosa) unido a tres grupos fosfato que tiene al menos un grupo adicional en uno de su extremo: 2', 3', 5' o base. Dicho grupo adicional bloquea el seguir añadiendo nucleótidos al impedir la formación de cualquier enlace fosfodiéster (modificación en 3'O, en 2' o en 2'O) o impidiendo estéricamente que la polimerasa se una a cualquier fragmento de ácido nucleico que comprenda en su extremo 3' dicho nucleósido modificado (modificación 5' o base). Además, dicho grupo adicional tiene ventajosamente una naturaleza reversible, lo que permite que dicho grupo se elimine mediante una reacción de escisión específica.

Los nucleósidos o trifosfatos de nucleótidos incluyen trifosfato de desoxiadenosina (dATP), trifosfato de desoxiguanosina (dGTP), trifosfato de desoxicitidina (dCTP) o trifosfato de desoxitimidina (dTTP), como ejemplos de nucleótidos que contienen desoxirribosa. El trifosfato de adenosina (ATP), el trifosfato de guanosina (GTP), el trifosfato de citidina (CTP) o el trifosfato de uridina (UTP) son ejemplos adicionales de trifosfatos de nucleótidos que contienen ribosa. Pueden unirse otros tipos de nucleósidos a tres fosfatos, formándose trifosfatos de nucleótidos, tales como nucleósidos modificados de origen natural y nucleósidos artificiales.

En una realización particular, el nucleótido modificado es un nucleótido bloqueado en 3'O, que comprende un grupo unido reversiblemente al extremo 3' del trifosfato de nucleótido, para impedir que se sigan añadiendo nucleótidos. Dicho grupo puede tener diversas naturalezas químicas, tales como azidometilo, aminoxi y alilo.

Ventajosamente, el nucleótido modificado se selecciona entre un trifosfato de 3'-O-NH₂-nucleósido, un trifosfato de 3'-O-azidometil-nucleósido, un trifosfato de 3'-O-alil-nucleósido, un trifosfato de 3'-O-(2-nitrobenzil)-nucleósido o un trifosfato de 3'-O-propargil-nucleósido.

En algunas realizaciones, los nucleótidos modificados comprenden un nucleótido modificado o una molécula de

nucleósido que comprende una base de purina o pirimidina y un resto de azúcar de ribosa o desoxirribosa que tiene un grupo de bloqueo 3'-OH eliminable unido covalentemente a la misma, de tal forma que el átomo de carbono 3' tiene unido un grupo de la estructura:



10 en donde -Z es cualquiera de -C(R')₂-O-R", -C(R')₂-N(R'')₂, -C(R')₂-N(H)R", -C(R')₂-S-R" y -C(R')₂-F, en donde cada R" es o forma parte de un grupo protector retirable; cada R' es independientemente un átomo de hidrógeno, un grupo alquilo, alquilo sustituido, arilalquilo, alquenilo, alquinilo, arilo, heteroarilo, heterocíclico, acilo, ciano, alcoxi, ariloxi, heteroariloxi o amido, o un marcador detectable unido a través de un grupo de enlace; a condición de que en algunas realizaciones, dichos sustituyentes tienen hasta 10 átomos de carbono y/o hasta 5 heteroátomos de oxígeno o nitrógeno; o (R')₂ representa un grupo alquilideno de fórmula =C(R''')₂ en donde R''' puede ser igual o diferente y se selecciona del grupo que comprende átomos de hidrógeno y halógeno y grupos alquilo, a condición de que en algunas realizaciones, el alquilo de cada R''' tiene de 1 a 3 átomos de carbono; y en donde la molécula se puede hacer reaccionar, obteniéndose un intermedio en el que cada R" se intercambia por H o, donde Z es -(R')₂-F, el F se intercambia por OH, SH o NH₂, preferentemente OH, disociándose dicho intermedio en condiciones acuosas, obteniéndose una molécula con un 3'-OH libre; a condición de que cuando Z es -C(R')₂-S-R", ambos grupos R' no son H. En ciertas realizaciones, R' del nucleótido o nucleósido modificado es un alquilo o alquilo sustituido, a condición de que dicho alquilo o alquilo sustituido tiene de 1 a 10 átomos de carbono y de 0 a 4 heteroátomos de oxígeno o nitrógeno. En ciertas realizaciones, -Z del nucleótido o nucleósido modificado es de fórmula -C(R')₂-N₃. En ciertas realizaciones, Z es un grupo azidometilo.

25 En algunas realizaciones, Z es un resto orgánico escindible con o sin heteroátomos que tiene un peso molecular de 200 o menos. En otras realizaciones, Z es un resto orgánico escindible con o sin heteroátomos que tiene un peso molecular de 100 o menos. En otras realizaciones, Z es un resto orgánico escindible con o sin heteroátomos que tiene un peso molecular de 50 o menos.

30 En una realización particular adicional, "nucleótido modificado en 3'O" se refiere a un trifosfato de nucleótido que porta en el extremo 3' un grupo 3'-O-metilo, 3'-azido, 3'-O-azidometilo, 3'-O-amino, 3'-aminoxi o 3'-O-alilo. En una realización adicional, el trifosfato de nucleótido bloqueado en 3' se bloquea con un grupo 3'-O-azidometilo, 3'-aminoxi o 3'-O-alilo. En otras realizaciones, "nucleótido modificado en 3'O" se refiere a un trifosfato de nucleótido que porta en el extremo 3' ésteres, éteres, carbonitrilos, fosfatos, carbonatos, carbamatos, hidroxilamina, boratos, nitratos, azúcares, fosforamida, fosforoamidatos, fenilsulfenatos, sulfatos, sulfonas o aminoácidos. En algunas realizaciones, los grupos de bloqueo en 3'-O anteriores tienen un peso molecular de 100 o menos.

35 En otras realizaciones, los grupos de bloqueo en 3'-O de la invención incluyen metilo, 3'-O-(2-nitrobencilo), alilo, amina, azidometilo, *terc*-butoxi, etoxi o propargilo.

40 En una realización particular adicional, "nucleótido modificado en 3'O" se refiere a un trifosfato de nucleótido que tiene un grupo modificador efector de terminación, como los descritos en el documento WO2016034807.

45 De manera interesante, las variantes de la invención muestran una afinidad aumentada por los nucleótidos modificados, en comparación con la TdT de tipo silvestre y por tanto, una capacidad aumentada de incorporar dicho nucleótido modificado en una secuencia de ácido nucleico durante la síntesis de ácidos nucleicos. Más particularmente, las variantes de la invención tienen la capacidad de usar e incorporar nucleótidos modificados en 3'O (y más particularmente, nucleótidos bloqueados en 3'O) en una secuencia de ácido nucleico, lo que no es posible con la TdT de tipo silvestre (véase Knapp *et al.* Chem. Eur. J., 2011, 17:2903).

50 De acuerdo con un aspecto particular, la invención se refiere a variantes de TdT que pueden funcionar con nucleótidos modificados en un proceso de síntesis enzimática de ácidos nucleicos, especialmente con nucleótidos modificados en 3'O (p. ej., nucleótidos bloqueados en 3'O), y que tienen la capacidad de producir moléculas de ácido nucleico o derivados de moléculas de ácido nucleico de gran longitud.

55 Síntesis enzimática de ácidos nucleicos

60 El objetivo de la presente invención es proporcionar variantes de TdT que pueden usarse para la síntesis de ácidos nucleicos, como se describe en Ybert *et al.*, documento WO2015/159023; Jensen *et al.*, Biochemistry, 57: 1821-1832 (2018); Hiatt *et al.*, patente de EE. UU. 5808045. Más particularmente, el objetivo de la presente invención es proporcionar variantes de TdT adecuadas para añadir nucleótidos modificados a una hebra de ácidos nucleicos en crecimiento. A continuación, puede eliminarse el grupo de bloqueo para permitir la adición de un nuevo nucleótido modificado.

65 De acuerdo con la invención, mediante el uso de una variante de la invención, es posible implementar ciclos sucesivos que comprenden adiciones y desprotecciones. Por tanto, este proceso permitirá múltiples ciclos de adición de un nucleótido modificado reversible y la eliminación adicional del grupo de bloqueo, lo que permitirá la extensión controlada de una hebra de ácido nucleico en crecimiento en una secuencia definida.

La presente invención contempla el uso de TdT modificada de acuerdo con la presente invención en un proceso de síntesis enzimática de ácidos nucleicos.

5 Por tanto, es un objetivo de la presente invención proporcionar un método para sintetizar un polinucleótido que tiene una secuencia predeterminada, que comprende las etapas de:

- a) proporcionar un iniciador que tiene un nucleótido terminal 3' que tiene un 3'-hidroxilo libre;
- b) repetir ciclos de (i) poner en contacto en condiciones de elongación el iniciador o los fragmentos elongados que tienen 3'-O-hidroxilos libres con un trifosfato de nucleósido bloqueado en 3'-O y una variante de TdT de la presente invención, de tal forma que el iniciador o los fragmentos elongados se elongan mediante la incorporación de un trifosfato de nucleósido bloqueado en 3'-O, formándose fragmentos elongados bloqueados en 3'-O y (ii) desbloquear los fragmentos elongados, formándose fragmentos elongados que tienen 3'-hidroxilos libres, hasta que se forma el polinucleótido.

15 También es el objetivo de la presente invención proporcionar un proceso para sintetizar una molécula de ácido nucleico sin una cadena molde, que comprende una etapa de poner en contacto un cebador de ácido nucleico tanto con al menos un nucleótido, preferentemente al menos un nucleótido modificado en 3'O y una variante de la invención.

20 La presente invención contempla el concepto de un proceso de síntesis de ácidos nucleicos enzimática. En dicho proceso, las moléculas de ácido nucleico se sintetizan *de novo* en ausencia de una cadena molde. Por consiguiente, se acopla una secuencia ordenada de nucleótidos a un fragmento de ácido nucleico iniciador con la ayuda de la variante de la invención. Se entenderá que cobra gran importancia el acoplamiento cuantitativo y de manera más general, una alta eficacia de acoplamiento de cada nucleótido a la cadena de ácido nucleico en crecimiento. También se entenderá que los nucleótidos no terminadores, tales como los nucleótidos naturales o los nucleótidos marcados permanentemente, no permitirán control alguno sobre la secuencia sintetizada y darán como resultado, por ejemplo, poliadiciones no controladas y no deseadas.

30 En algunas realizaciones, el método para sintetizar un polinucleótido comprende las etapas de (a) proporcionar un iniciador que tiene un 3'-hidroxilo libre; (b) hacer reaccionar en condiciones de extensión el iniciador o un intermedio de extensión que tiene un 3'-hidroxilo libre con una TdT variante de la invención, en presencia de un trifosfato de nucleósido bloqueado en 3'-O, produciéndose un intermedio de extensión bloqueado en 3'-O; (c) desbloquear el intermedio de extensión, produciéndose un intermedio de extensión con un 3'-hidroxilo libre; y (d) repetir las etapas (b) y (c) hasta que se sintetiza el polinucleótido.

35 En algunas realizaciones, el método para sintetizar un polinucleótido comprende las etapas de (a) proporcionar un iniciador unido a un soporte sólido, siendo el iniciador un oligonucleótido que tiene un 3'-hidroxilo libre; (b) hacer reaccionar en condiciones de extensión el iniciador o un intermedio de extensión que tiene un 3'-hidroxilo libre con una TdT variante de la invención, en presencia de un trifosfato de nucleósido bloqueado en 3'-O, produciéndose un intermedio de extensión bloqueado en 3'-O; (c) lavar el soporte sólido para eliminar el trifosfato de nucleósido bloqueado en 3'-O no incorporado; (d) desbloquear el intermedio de extensión exponiendo el soporte sólido a un agente de desbloqueo, produciéndose un intermedio de extensión que tiene un 3'-hidroxilo libre; y (e) repetir las etapas (b) y (d) hasta que se sintetiza el polinucleótido. El método puede incluir una etapa adicional de escindir el polinucleótido completado del soporte sólido.

40 En algunas realizaciones, para las reacciones de adición catalizadas por TdT, las condiciones enzimáticas pueden contener de aproximadamente 0,20 a aproximadamente 200 μ M del nucleótido que tiene el resto de bloqueo eliminable que protege el 3'-hidroxilo y de aproximadamente 0,20 a 200 μ M de 3'-hidroxilos libres y no modificados procedentes del sustrato de inicio. En algunas realizaciones, el tampón de reacción contiene de aproximadamente 10 a aproximadamente 500 mM de tampón de cacodilato potásico (pH entre 6,5 y 7,5), y de aproximadamente 0,01 a aproximadamente 10 mM de un catión divalente (p. ej., CoCl_2 o MnCl_2). Otras composiciones y componentes del tampón pueden ser adecuados para realizaciones particulares deseadas de la presente invención.

55 En el contexto de la invención, la expresión "reacción de escisión" se refiere a cualquier acción del sustrato o condiciones físicas, que tienen la capacidad de escindir el grupo adicional anteriormente descrito en los nucleótidos modificados reversibles. Un experto en la materia es capaz de determinar una reacción de escisión para cualquier grupo enumerado anteriormente.

60 En una realización, el agente de escisión es un agente de escisión química. En una realización alternativa, el agente de escisión es un agente de escisión enzimática.

65 Un experto en la materia entenderá que la selección del agente de escisión depende del tipo de grupo de bloqueo de nucleótido 3' empleado. Por ejemplo, puede utilizarse tris(2-carboxietil)fosfina (TCEP) para escindir los grupos 3'-O-azidometilo, pueden usarse complejos de paladio para escindir los grupos 3'-O-alilo o puede usarse nitrito de sodio para escindir un grupo 3'-O-amino. En una realización particular, la reacción de escisión implica: TCEP, un complejo de paladio o nitrito de sodio.

En una realización particular, la reacción de escisión se realiza en presencia de componentes adicionales, tales como un desnaturante (urea, cloruro de guanidinio, formamida o betaína, por ejemplo). En una realización adicional, la reacción de escisión se realiza con uno o más tampones. Un experto en la materia entenderá que la elección del
5 tampón depende del mecanismo de reacción exacto.

La presente invención se refiere a variantes de TdT con la capacidad de incorporar, de manera cuantitativa, nucleótidos modificados. Por "cuantitativamente" o "reacción cuantitativa", se entiende una reacción que alcanza la compleción, es decir, en la que los reactivos se convierten completamente en el producto. La polimerasa que incorpora de manera
10 cuantitativa un nucleótido modificado reversible es una polimerasa capaz de alargar cada fragmento de ácido nucleico con todos los nucleótidos disponibles, lo que da lugar a la conversión de todos los fragmentos de iniciación de longitud n , a fragmentos de longitud $n+1$.

Como se usan en el presente documento, "fragmento de iniciación" se refiere a una secuencia de oligonucleótido sintética con un extremo 3' libre, que puede elongarse adicionalmente. En una realización, el fragmento de iniciación es un fragmento de iniciación de ADN. En una realización alternativa, el fragmento de iniciación es un fragmento de
15 iniciación de ARN.

En una realización, el fragmento de iniciación posee entre 3 y 100 nucleótidos, concretamente entre 3 y 20 nucleótidos.
20

En una realización, el fragmento de iniciación es monocatenario. En una realización alternativa, el fragmento de iniciación es bicatenario.

En una realización, el fragmento de iniciación se inmoviliza sobre un soporte sólido. El fragmento de iniciación puede unirse mediante diversos métodos a un soporte sólido, lo que da como resultado un compuesto estable en las diversas condiciones de reacción enzimáticas o de síntesis a las que se someterá el fragmento.
25

En una realización, el fragmento de iniciación se inmoviliza sobre un soporte sólido mediante un resto de interacción reversible, tal como un enlazador escindible químicamente, un epítipo de anticuerpo/inmunogénico, una biotina/proteína de unión a biotina o un marcador de glutatión-GST. En una realización adicional, el fragmento de iniciación se inmoviliza sobre un soporte sólido mediante un enlazador escindible químicamente, tal como un enlazador de éter de hemiaminal enmascarado con disulfuro, alilo o azida.
30

En un fragmento de iniciación, la parte inmovilizada contiene al menos un sitio de restricción. Se ha descrito en la bibliografía el uso de enzimas de restricción y sitios de restricción para hidrolizar cadenas de ácido nucleico de manera selectiva en un sitio específico. Cualquier experto en la materia será capaz de elegir la enzimas de restricción adecuada para la secuencia del sitio de escisión del fragmento de iniciación.
35

En una realización alternativa, el fragmento de iniciación contiene al menos una uridina. El tratamiento con uracilo-ADN glicosilasa (UDG) genera un sitio abásico. El tratamiento sobre un sustrato adecuado con una endonucleasa de sitioapurínico/apirimidínico (AP) extraerá la hebra de ácido nucleico.
40

Aplicaciones

En el presente documento se describe el uso de variantes de TdT para su uso en la síntesis de ácidos nucleicos, la síntesis de oligonucleótidos, la síntesis de sondas, marcaje, la amplificación de ácidos nucleicos, aptámeros, moléculas de ácido nucleico terapéuticas, el descubrimiento y la validación de dianas farmacológicas, el diagnóstico de enfermedades, la ingeniería metabólica, el almacenamiento de datos, la mejora de cultivos, el diseño de bibliotecas, mezclas de secuenciación, el marcaje o la unión de ácidos nucleicos o cualquier otra aplicación que implica moléculas de ácido nucleico.
45
50

Producción de TdT variantes

Las variantes de la invención pueden producirse mutando polinucleótidos codificantes de TdT de referencia o de tipo silvestre conocidos, y después expresarlos utilizando técnicas de biología molecular convencionales. Por ejemplo, el gen de TdT de ratón (SEQ ID NO: 1) puede ensamblarse a partir de fragmentos sintéticos utilizando técnicas de biología molecular convencionales, p. ej., utilizando protocolos descritos por Stemmer et al., Gene, 164: 49-53 (1995); Kodumal et al, Proc. Natl. Acad. Sci., 101: 15573-15578 (2004); o similares, o puede clonarse directamente a partir de células de ratón utilizando protocolos descritos por Boule et al., Mol. Biotechnology, 10: 199-208 (1998) o Bentolila et al, EMBO J., 14: 4221-4229 (1995); o similares.
55
60

Por ejemplo, puede insertarse un fragmento de TdT aislado en un vector de expresión, tal como pET32 (Novagen) obteniéndose un vector pCTdT que después puede utilizarse para producir y expresar proteínas TdT variantes utilizando protocolos convencionales. Los vectores con la secuencia correcta pueden transformarse en cepas de *E. coli* productoras.
65

Las cepas transformadas se cultivan empleando técnicas convencionales hasta obtener gránulos de los que se extrae la proteína TdT. Por ejemplo, los gránulos previamente preparados se descongelan en un baño de agua a una temperatura de 30 a 37 °C. Una vez completamente descongelados, los gránulos se resuspenden en tampón de lisis formado por Tris-HCl 50 mm (Sigma) pH 7,5, NaCl 150 mm (Sigma), mercaptoetanol 0,5 mm (Sigma), glicerol al 5 % (Sigma), imidazol 20 mm (Sigma) y 1 comprimido para 100 ml de cóctel inhibidor de proteasas (ThermoFisher). La resuspensión cuidadosa se realiza para evitar la lisis prematura y el resto de agregados. Las células resuspendidas se lisan mediante varios ciclos de prensa francesa, hasta que se obtiene una plena homogeneidad del color. La presión habitual utilizada es 96,5 mPa (14.000 psi). A continuación, se centrifuga el lisado durante 1 h a 1 h 30 min a 10.000 rpm. El centrifugado se hace pasar a través de un filtro de 0,2 µm para eliminar los restos antes de la purificación en columna.

La proteína TdT puede purificarse del centrifugado en un procedimiento por afinidad en una etapa. Por ejemplo, se usa una columna de afinidad de Ni-NTA (GE Healthcare) para unirse a las polimerasas. Inicialmente, la columna se ha lavado y equilibrado con 15 volúmenes de columna de Tris-HCl 50 mm (Sigma) a pH 7,5, NaCl 150 mm (Sigma) e imidazol 20 mm (Sigma). Las polimerasas se unen a la columna después del equilibrado. A continuación, se aplican a la columna 15 volúmenes de columna de un tampón de lavado, compuesto por Tris-HCl 50 mm (Sigma) a pH 7,5, NaCl 500 mm (Sigma) e imidazol 20 mm (Sigma). Después del lavado, se eluyen las polimerasas con Tris-HCl 50 mm (Sigma) a pH 7,5, NaCl 500 mm (Sigma) e imidazol 0,5 m (Sigma). Las fracciones correspondientes a la máxima concentración de polimerasa de interés se recogen y agrupan en una sola muestra. Las fracciones agrupadas se dializan frente a tampón de diálisis (Tris-HCl 20 mm, pH 6,8, NaCl 200 mm, MgOAc 50 mm, [NH₄]₂SO₄ 100 mm). El dializado se concentra posteriormente con la ayuda de filtros de concentración (Amicon Ultra-30, Merk Millipore). La enzima concentrada se distribuye en pequeñas alícuotas, se añade una concentración final de glicerol al 50 % y a continuación, se congelan dichas alícuotas a -20 °C y se almacenan a largo plazo. Se analizan 5 µl de diversas fracciones de las enzimas purificadas en geles de SDS-PAGE.

25

Kits, composición de enzima y nucleótidos

Un aspecto particular de la invención se refiere a la composición y el uso de kits que comprenden una variante de TdT de acuerdo con la invención, o a cualquier aspecto particular de la presente invención, con, opcionalmente, cualquier combinación de uno o más componentes seleccionados entre: un fragmento de iniciación, uno o más nucleótidos terminadores reversibles, enzima adicional y reactivos utilizados en una reacción de escisión. Dichos kits pueden usarse en un método de síntesis enzimática de ácidos nucleicos.

La presente invención abarca la composición de materia que comprende variantes de TdT de acuerdo con la invención, o con cualquier aspecto particular de la presente invención, con un nucleótido modificado reversible en una mezcla con un tampón adecuado a una concentración adecuada.

Ejemplos

40 **Ejemplo 1: Generación, expresión y purificación de variantes de TdT de acuerdo con la invención**

Generación de la cepa de expresión

El gen de TdT de ratón se genera a partir del plásmido pET28 descrito en [Boulé *et al.*, 1998, Mol. Biotechnol. 10, 199-208]. La secuencia de SEQ ID NO: 4 (Marcador TdT) se amplifica utilizando los siguientes cebadores:

- T7-pro: TAATACGACTCACTATAGGG (SEQ ID NO: 5)
- T7-ter: GCTAGTTATTGCTCAGCGG (SEQ ID NO: 6)

50

mediante técnicas de biología molecular convencionales. A continuación, la secuencia se clona en el armazón del plásmido pET32, obteniéndose el nuevo plásmido pCTdT.

Después de la secuenciación, pCTdT se transforma en células de *E. coli* comerciales, BL21 (DE3, de Novagen). Las colonias en proliferación sobre la placa con kanamicina se aíslan y denominan Ec-CTdT.

55

Generación de variantes de polimerasa

Como vector de partida, se utiliza el vector pCTdT. Los cebadores específicos que comprenden una o varias mutaciones puntuales se generan mediante el software online de Agilent (<http://www.genomics.agilent.com:80/primerDesignProgram.jsp>). Se ha usado el kit comercial QuickChange II (Agilent) para generar la polimerasa modificada deseada que comprende las mutaciones diana. El procedimiento experimental ha seguido el protocolo del proveedor. Después de generar los diferentes vectores, se secuenció cada uno de ellos. Los vectores con la secuencia correcta se han transformado en cepas de *E. coli* productoras, como se ha descrito anteriormente. Los clones con capacidad de proliferar en placas de LB-agar con kanamicina se aíslan.

65

Expresión

Se utilizan las cepas Ec-CTdT y Ec-clon para inocular matraces Erlenmeyer de 250 ml con 50 ml de medio LB suplementado con la cantidad adecuada de kanamicina. Después del crecimiento durante una noche a 37 °C, se utilizan volúmenes adecuados de estos precultivos para inocular matraces Erlenmeyer de 5 l con 2 l de medio LB con kanamicina. La DO inicial de los cultivos de 5 l se selecciona para que sea de 0,01. Los matraces Erlenmeyer se ponen a 37 °C con agitación vigorosa y se comprueba regularmente la DO de los diferentes cultivos. Después de alcanzar una DO comprendida entre 0,6 y 0,9, se suplementa cada matraz Erlenmeyer mediante la adición de 1 ml de IPTG 1 m (β -D-1-tiogalactopiranosido de isopropilo, Sigma). Se vuelven a poner los matraces Erlenmeyer en agitación a una temperatura controlada de 37 °C. Después de la expresión durante una noche, se recolectan las células en varios gránulos. Los gránulos que expresan las mismas variantes se agrupan y almacenan a -20 °C, en algunos casos durante varios meses.

Extracción

Los gránulos previamente preparados se descongelan en un baño de agua a una temperatura de 30 a 37 °C. Una vez completamente descongelados, los gránulos se resuspenden en tampón de lisis formado por Tris-HCl 50 mm (Sigma) pH 7,5, NaCl 150 mm (Sigma), mercaptoetanol 0,5 mm (Sigma), glicerol al 5 % (Sigma), imidazol 20 mm (Sigma) y 1 comprimido para 100 ml de cóctel inhibidor de proteasas (ThermoFisher). La resuspensión cuidadosa se realiza para evitar la lisis prematura y el resto de agregados. Las células resuspendidas se lisan mediante varios ciclos de prensa francesa, hasta que se obtiene una plena homogeneidad del color. La presión habitual utilizada es 96,5 mPa (14.000 psi). A continuación, se centrifuga el lisado durante 1 h a 1 h 30 min a 10.000 rpm. El centrifugado se hace pasar a través de un filtro de 0,2 μ m para eliminar los restos antes de la purificación en columna.

Purificación

Se emplea un procedimiento de purificación en una etapa para purificar las enzimas polimerasas producidas y extraídas. Se usa una columna de afinidad de Ni-NTA (GE Healthcare) para unirse a las polimerasas. Inicialmente, la columna se ha lavado y equilibrado con 15 volúmenes de columna de Tris-HCl 50 mm (Sigma) a pH 7,5, NaCl 150 mm (Sigma) e imidazol 20 mm (Sigma). Las polimerasas se unen a la columna después del equilibrado. A continuación, se aplican a la columna 15 volúmenes de columna de un tampón de lavado, compuesto por Tris-HCl 50 mm (Sigma) a pH 7,5, NaCl 500 mm (Sigma) e imidazol 20 mm (Sigma). Después del lavado, se eluyen las polimerasas con Tris-HCl 50 mm (Sigma) a pH 7,5, NaCl 500 mm (Sigma) e imidazol 0,5 m (Sigma). Las fracciones correspondientes a la máxima concentración de polimerasa de interés se recogen y agrupan en una sola muestra. Las fracciones agrupadas se dializan frente a tampón de diálisis (Tris-HCl 20 mm, pH 6,8, NaCl 200 mm, MgOAc 50 mm, $[\text{NH}_4]_2\text{SO}_4$ 100 mm). El dializado se concentra posteriormente con la ayuda de filtros de concentración (Amicon Ultra-30, Merk Millipore). La enzima concentrada se distribuye en pequeñas alícuotas, se añade una concentración final de glicerol al 50 % y a continuación, se congelan dichas alícuotas a -20 °C y se almacenan a largo plazo. Se analizan 5 μ l de diversas fracciones de las enzimas purificadas en geles de SDS-PAGE.

Ejemplo 2: Evaluación de la actividad de las variantes de TdT con cebadores fluorescentes***Ensayo de actividad***

Se evalúa el rendimiento de elongación de las variantes de TdT generadas, expresadas y purificadas de acuerdo con el ejemplo 1 mediante el siguiente ensayo. Todos los resultados se comparan entre sí y con la enzima TdT de tipo silvestre y con un tubo de control que no tenía enzima polimerasa.

Tabla 5: Ensayo de actividad

Reactivo	Concentración	Volumen
H ₂ O	-	12 μ l
Tampón de actividad	10x	2 μ l
dNTP	250 μ M	2 μ l
Enzima purificada	20 μ M	2 μ l
ADN de cebador fluorescente	500 nM	2 μ l

El tampón de actividad comprende, por ejemplo, tampón de reacción de TdT (disponible de New England BioLabs) suplementado con CoCl₂. El cebador usado es el siguiente:
5'-AAAAAAAAAAAAAAAAAGGGG-3' (SEQ ID NO: 7)

El cebador también tiene un colorante fluorescente ATTO en el extremo 5'.

Los nucleótidos usados (indicados como dNTP en la tabla 5) son 3'-O-amino-2',3'-didesoxinucleótidos-5'-trifosfato (ONH₂, Firebird Biosciences) tales como 3'-O-amino-2',3'-didesoxiadenosina-5'-trifosfato, por ejemplo.

Por cada diferente variante analizada, se utiliza un tubo para la reacción. Los reactivos se añaden al tubo, comenzando con agua y después en el orden de la tabla 5. Después de 30 min a 37 °C, se detiene la reacción mediante la adición de formamida (Sigma).

5 *Análisis*

El análisis implica análisis en gel de poliacrilamida. Las muestras del ensayo de actividad se analizan con un gel de poliacrilamida al 16 % (Biorad). Los geles se preparan justo antes del análisis vertiendo la poliacrilamida dentro de placas de vidrio y se dejan polimerizar. El gel dentro de las placas de vidrio se monta sobre un tanque adaptado relleno con tampón TBE (Sigma) para la etapa de electroforesis. Las muestras que se van a analizar se cargan sobre el gel. Se aplica una tensión de 500 a 2000 V entre la parte superior y la inferior del gel durante 3 a 6 h a temperatura ambiente. Una vez que han migrado según el tamaño objetivo de la muestra, se desmonta el gel y se lee la fluorescencia mediante el uso del instrumento Typhoon (GE Life Sciences). Después de la adquisición de imágenes, se utiliza el software ImageJ (imagej.nih.gov/ij/) para analizar el porcentaje de incorporación de los nucleótidos modificados.

Ejemplo 3: Evaluación de la actividad de variantes de TdT con cebador no marcado

Ensayo de actividad

Se evalúa el rendimiento de elongación de las variantes 5 generadas, expresadas y purificadas de acuerdo con el ejemplo 1 mediante el siguiente ensayo. Todos los resultados se comparan con una variante de referencia (SEQ ID NO: 9) obtenida de la investigación anterior y con un tubo de control que no contiene enzima polimerasa.

Tabla 6: Ensayo de actividad

Reactivo	Concentración	Volumen
H ₂ O	-	12 µl
Tampón de actividad	10x	2 µl
dNTP	250 µM	2 µl
Enzima purificada	20 µM	2 µl
ADN de cebador fluorescente	500 nM	2 µl

El cebador usado es el siguiente:

5'- TTTTTTTTTTTTAAATAAGG-3' (SEQ ID NO: 8)

Los nucleótidos utilizados (indicados como dNTP en la tabla 6) fueron 3'-O-amino-2',3'-didesoxinucleótidos-5'-trifosfato (ONH₂, Firebird Biosciences) tales como 3'-O-amino-2',3'-didesoxiadenosina-5'-trifosfato, por ejemplo.

Por cada variante analizada, se utilizó un tubo para la reacción. Los reactivos se añaden en el tubo comenzando con el agua y después en el orden indicado en la tabla 6. Después de 30min a 37 °C, se detuvo la reacción mediante la adición de formamida (Sigma).

Análisis

El análisis utilizó cromatografía de líquidos y detección y cuantificación por espectrometría de masas (LC/MS). Las muestras del ensayo de actividad se analizan mediante LC/MS. Las muestras se cargan en el instrumento de LC/MS y se realiza un método de separación de oligonucleótidos convencional. La adquisición de datos fue seguida de su análisis y cálculo de espectros.

REIVINDICACIONES

1. Una variante de didesoxinucleotidil transferasa terminal (TdT) que comprende una secuencia de aminoácidos al menos un 90 % idéntica a la secuencia de aminoácidos como se expone en la secuencia de aminoácidos seleccionada de la SEQ ID NO: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 o 35,
- 5
- en donde la secuencia de aminoácidos comprende al menos una sustitución de aminoácidos de un resto de cisteína en la posición 173 con respecto a las SEQ ID NO: 11, 13, 17, 19, 21, 29 o 31, o de un resto de cisteína funcionalmente equivalente seleccionado entre una cisteína en la posición 172 con respecto a la SEQ ID NO: 15, o una cisteína en la posición 178 con respecto a la SEQ ID NO: 23, o una cisteína en la posición 174 con respecto a la SEQ ID NO: 25, o una cisteína en la posición 171 con respecto a la SEQ ID NO: 27, o una cisteína en la posición 182 con respecto a la SEQ ID NO: 33, o una cisteína en la posición 176 con respecto a la SEQ ID NO: 35,
 - 10
 - en donde la secuencia de aminoácidos comprende una sustitución adicional de un resto de arginina en la posición 207 con respecto a las SEQ ID NO: 11, 13, 17, 19, 23, 29 o 31, o de un resto de arginina funcionalmente equivalente seleccionado entre una arginina en la posición 206 con respecto a la SEQ ID NO: 15, o una arginina en la posición 208 con respecto a las SEQ ID NO: 21 o 25, o una arginina en la posición 205 con respecto a la SEQ ID NO: 27, o una arginina en la posición 216 con respecto a la SEQ ID NO: 33, o una arginina en la posición 210 con respecto a la SEQ ID NO: 35,
 - 15
 - 20
- en donde la variante de TdT (i) es capaz de sintetizar un fragmento de ácido nucleico sin una cadena molde y (ii) es capaz de incorporar un nucleótido modificado en un fragmento de ácido nucleico.
2. La variante de TdT de la reivindicación 1 que tiene una sustitución adicional de un resto de metionina en la posición 63 con respecto a las SEQ ID NO: 11, 13, 15, 17, 19, 29 o 33 o de un resto de metionina funcionalmente equivalente seleccionado entre una metionina en la posición 73 con respecto a la SEQ ID NO: 23, o una metionina en la posición 73 con respecto a la SEQ ID NO: 23, o una metionina en la posición 64 con respecto a la SEQ ID NO: 25, metionina en la posición 61 con respecto a la SEQ ID NO: 27, o una metionina en la posición 66 con respecto a la SEQ ID NO: 35.
- 25
3. La variante de TdT de la reivindicación 1 o 2, que tiene una sustitución adicional de un resto de arginina en la posición 324 con respecto a las SEQ ID NO: 11 o 13, o de un resto de arginina funcionalmente equivalente seleccionado entre una arginina en la posición 320 con respecto a la SEQ ID NO: 15, o una arginina en la posición 331 con respecto a la SEQ ID NO: 17, o una arginina en la posición 325 con respecto a las SEQ ID NO: 19, 23 o 31, o una arginina en la posición 326 con respecto a la SEQ ID NO: 25, o una arginina en la posición 323 con respecto a la SEQ ID NO: 27, o una arginina en la posición 328 con respecto a la SEQ ID NO: 29, o una arginina en la posición 338 con respecto a la SEQ ID NO: 33, o una arginina en la posición 328 con respecto a la SEQ ID NO: 35.
- 30
4. La variante de TdT de una cualquiera de las reivindicaciones anteriores, que tiene una sustitución adicional de un resto de ácido glutámico en la posición 327 con respecto a las SEQ ID NO: 11 o 13, o de un resto de ácido glutámico funcionalmente equivalente seleccionado entre un ácido glutámico en la posición 334 con respecto a la SEQ ID NO: 17 o 21, o un ácido glutámico en la posición 331 con respecto a las SEQ ID NO: 29 o 35, o un ácido glutámico en la posición 328 con respecto a las SEQ ID NO: 19, 23 o 31, o un ácido glutámico en la posición 329 con respecto a la SEQ ID NO: 25, o un ácido glutámico en la posición 326 con respecto a la SEQ ID NO: 27, o un ácido glutámico en la posición 341 con respecto a la SEQ ID NO: 33.
- 35
- 40
5. La variante de TdT de una cualquiera de las reivindicaciones anteriores, en donde dicha sustitución de dicha cisteína es G, R, P, A, V, S, N, Q o D, preferentemente G o R.
- 45
6. La variante de TdT de una cualquiera de las reivindicaciones 2 a 5, en donde dicha sustitución de dicha metionina es R, Q, G, A, V, D, N, H o E, preferentemente R o Q.
- 50
7. La variante de TdT de una cualquiera de las reivindicaciones anteriores, en donde dicha sustitución de dicha arginina es N, L, K, H, G, D, A o P, preferentemente N o L.
- 55
8. La variante de TdT de una cualquiera de las reivindicaciones 4 a 7, en donde dicha sustitución de dicho ácido glutámico es N, L, T o S.
- 60
9. La variante de TdT de una cualquiera de las reivindicaciones 1 a 8, en donde dicha variante de TdT es capaz de incorporar en un fragmento de ácido nucleico, preferentemente en un 3'-hidroxilo libre de un fragmento de ácido nucleico, un nucleótido modificado en 3'-O, preferentemente seleccionado entre un trifosfato de 3'-O-NH₂-nucleósido, un trifosfato de 3'-O-azidometil-nucleósido, un trifosfato de 3'-O-alil-nucleósido, un trifosfato de 3'-O-(2-nitrobencil)-nucleósido o un trifosfato de 3'-O-propargil-nucleósido, más preferentemente un trifosfato de 3'-O-NH₂-nucleósido.
- 65
10. La variante de TdT de una cualquiera de las reivindicaciones 1 a 9, en donde dicha variante de TdT incorpora dicho nucleótido modificado en 3'-O en un intervalo mayor que el de una TdT de tipo silvestre.

11. Un ácido nucleico que codifica una variante de TdT como se define en una cualquiera de las reivindicaciones 1 a 10.
- 5 12. Un método para producir una variante de TdT que comprende:
- (a) cultivar una célula hospedadora que comprende un ácido nucleico de acuerdo con la reivindicación 11, en condiciones adecuadas para expresar el ácido nucleico que codifica la variante de TdT; y opcionalmente
- 10 (b) recuperar dicha variante de TdT del cultivo celular.
13. Un método para sintetizar un polinucleótido que tiene una secuencia predeterminada, comprendiendo el método las etapas de:
- 15 a) proporcionar un iniciador que tiene un nucleótido terminal 3' que tiene un 3'-hidroxilo libre;
- b) repetir ciclos de (i) poner en contacto en condiciones de elongación el iniciador o los fragmentos elongados que tienen 3'-O-hidroxilos libres con un trifosfato de nucleósido bloqueado en 3'-O y una variante de TdT de acuerdo con una cualquiera de las reivindicaciones 1 a 10, de tal forma que el iniciador o los fragmentos elongados se elongan mediante la incorporación de un trifosfato de nucleósido bloqueado en 3'-O, formándose fragmentos elongados bloqueados en 3'-O y (ii) desbloquear los fragmentos elongados, formándose fragmentos elongados
- 20 que tienen 3'-hidroxilos libres, hasta que se forma el polinucleótido.
14. El método de la reivindicación 13, en donde dicho nucleósido bloqueado en 3'-O es un trifosfato de 3'-O-NH₂-nucleósido, un trifosfato de 3'-O-azidometil-nucleósido, un trifosfato de 3'-O-alil-nucleósido, o un trifosfato de 3'-O-(2-nitrobencil)-nucleósido.
- 25 15. Un kit para realizar una reacción de incorporación de nucleótidos que comprende
- una variante de TdT de acuerdo con una cualquiera de las reivindicaciones 1 a 10,
- uno o más nucleótidos, preferentemente uno o más nucleótidos modificados en 3'O, y
- 30 - opcionalmente al menos un cebador de ácido nucleico.