

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 987 024**

51 Int. Cl.:

C12Q 1/68 (2008.01)

C12N 9/12 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **03.05.2018** **PCT/US2018/030972**

87 Fecha y número de publicación internacional: **08.11.2018** **WO18204707**

96 Fecha de presentación y número de la solicitud europea: **03.05.2018** **E 18793821 (2)**

97 Fecha y número de publicación de la concesión europea: **17.07.2024** **EP 3619321**

54 Título: **Variantes de polimerasa DP04**

30 Prioridad:

04.05.2017 US 201762501526 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
13.11.2024

73 Titular/es:

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

72 Inventor/es:

KOKORIS, MARK STAMATIOS;
PRINDLE, MARC;
CHASE, JACK;
BUSAM, ROBERT;
KOVARIK, MICHAEL;
KELLER, SALK;
MURT, MEGAN y
THIESSEN, GREG

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

Observaciones:

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

ES 2 987 024 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 polimerasa DP04

5 Campo de la invención

La divulgación se refiere, en general, a composiciones y procedimientos con polimerasa. Más en particular, la divulgación se refiere a polimerasas DPO4 modificadas y su uso en aplicaciones biológicas incluyendo, por ejemplo, incorporación de análogos de nucleótidos, extensión del cebador y reacciones de secuenciación de molécula única.

Antecedentes de la invención

Las ADN polimerasas replican los genomas de los organismos vivos. Además de este papel central en biología, las ADN polimerasas también son herramientas ubicuas de la biotecnología. Se usan ampliamente, por ejemplo, para retrotranscripción, amplificación, marcaje y secuenciación, todas tecnologías centrales para una variedad de aplicaciones, tales como secuenciación de ácidos nucleicos, amplificación de ácidos nucleicos, clonación, manipulación de proteínas, diagnóstico, medicina molecular y muchas otras tecnologías.

Debido a su importancia, las ADN polimerasas se han estudiado extensamente, con el foco puesto, por ejemplo, en las relaciones filogenéticas entre polimerasas, estructura de las polimerasas, rasgos característicos de estructura-función de las polimerasas y el papel de las polimerasas en la replicación de ADN y otros procesos biológicos básicos, así como modos de usar ADN polimerasas en biotecnología. Para una revisión de las polimerasas, véanse, por ejemplo, Hubscher *et al.* (2002) "Eukaryotic DNA Polymerases" Annual Review of Biochemistry, vol. 71: 133-163, Alba (2001) "Protein Family Review: Replicative DNA Polymerases" Genome Biology 2(1): revisiones 3002.1-3002.4, Steitz (1999) "DNA polymerases: structural diversity and common mechanisms" J Biol Chem 274:17395-17398, y Burgers *et al.* (2001) "Eukaryotic DNA polymerases: proposal for a revised nomenclature" J Biol. Chem. 276(47): 43487-90. Se han resuelto estructuras cristalinas para muchas polimerasas, que, a menudo, comparten una arquitectura similar. Se han determinado los mecanismos básicos de acción de muchas polimerasas.

Una aplicación fundamental de las ADN polimerasas es en las tecnologías de secuenciación de ADN. Desde el clásico procedimiento de secuenciación de Sanger hasta las recientes tecnologías de secuenciación de "nueva generación" (SNG), los sustratos de nucleótidos usados para la secuenciación han cambiado necesariamente con el tiempo. La serie de modificaciones nucleotídicas requeridas por estas tecnologías rápidamente cambiantes ha introducido tareas ingentes para los investigadores de las ADN polimerasas a la hora de buscar, diseñar o desarrollar enzimas compatibles para las químicas de secuenciación de ADN siempre cambiantes. Se han identificado mutantes de ADN polimerasa que tienen una variedad de propiedades útiles, incluyendo capacidades de incorporación de análogos de nucleótidos alteradas en relación con enzimas homólogas naturales. Por ejemplo, la ADN polimerasa Vent^{A488L} puede incorporar determinados nucleótidos no estándar con una mayor eficacia que la ADN polimerasa Vent natural. Véanse Gardner *et al.* (2004) "Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase" J. Biol. Chem. 279(12): 11834-11842 y Gardner y Jack (1999) "Determinants of nucleotide sugar recognition in an archaeon DNA polymerase" Nucleic Acids Research 27(12):2545-2553. Se predice que el residuo alterado en este mutante, A488, estará de espaldas al sitio de unión a nucleótido de la enzima. El patrón de especificidad relajada en esta posición se correlaciona aproximadamente con el tamaño de la cadena lateral del aminoácido sustituido y afecta a la incorporación por la enzima de una variedad de glúcidos de nucleótidos modificados.

Más recientemente, las tecnologías de SNG han introducido la necesidad de adaptar las enzimas ADN polimerasas para aceptar sustratos de nucleótidos modificados con finalizadores reversibles en el -OH en 3', tal como -ONH₂. Para este fin, Chen y sus colaboradores combinaron análisis estructurales con un análisis de "vías adaptativas evolutivas reconstruidas" para generar una variante de TAQ^{L616A} que pueda incorporar eficazmente finalizadores tanto reversibles como irreversibles. Véase Chen *et al.* (2010) "Reconstructed Evolutionary Adaptive Paths Give Polymerases Accepting Reversible Terminators for Sequencing and SNP Detection" Proc. Nat. Acad. Sci. 107(5):1948-1953. Los estudios de modelado sugirieron que esta variante podría abrir espacio detrás de Phe-667, permitiéndole acomodar los sustituyentes en 3' más grandes. La patente de EE. UU. n.º 8.999.676 de Emig *et al.* divulga polimerasas modificadas adicionales que presentan propiedades mejoradas útiles para tecnologías de secuenciación de molécula única en base a detección fluorescente. En particular, se descubrió que la sustitución de la ADN polimerasa φ29 en las posiciones E375 y K512 potencia la capacidad de la polimerasa para utilizar análogos de nucleótidos marcados con fosfato no naturales que incorporan diferentes tintes fluorescentes.

Recientemente, Kokoris *et al.* han descrito un procedimiento, denominado "secuenciación por expansión" (SBX), que usa una ADN polimerasa para transcribir la secuencia de ADN en un polímero mensurable llamado Xpandomer (véase, por ejemplo, la patente de EE. UU. n.º 8.324.360 de Kokoris *et al.*). La secuencia transcrita se codifica a lo largo de la cadena principal de Xpandomer en indicadores de alta señal con respecto a ruido que están separados por ~10 nm y están diseñados para respuestas de alta señal con respecto a ruido bien diferenciadas al leerse por

sistemas de secuenciación basada en nanoporos. Los Xpandomer se generan a partir de análogos de nucleótidos no naturales, denominados XNTP, caracterizados por sustituyentes voluminosos que posibilitan que la cadena principal de Xpandomer se expanda tras la síntesis. Estos análogos XNTP introducen novedosos desafíos como sustratos para las ADN polimerasas disponibles actualmente.

El documento WO 2017/087281 A1 se refiere a variantes de ADN polimerasa de tipo DPO4 recombinante con sustituciones aminoácidas que confieren propiedades modificadas a la polimerasa para aplicaciones de secuenciación de molécula única mejoradas.

Los documentos WO 2013/170963 A2 y WO 2015/074756 A1 se refieren a un procedimiento para incorporar ácidos L-nucleicos al ADN que se va a sintetizar usando la ADN polimerasa DPO4 o variantes de la misma.

Por tanto, nuevas polimerasas modificadas, por ejemplo, polimerasas modificadas que presenten propiedades mejoradas útiles para secuenciación basada en nanoporos y otras aplicaciones de polimerasa (por ejemplo, la amplificación, secuenciación, marcaje, detección, clonación de ADN, etc., se divulgan en el presente documento), encontrarían valor en la técnica. En el presente documento se divulgan nuevas ADN polimerasas recombinantes con propiedades deseables, incluyendo la capacidad de incorporar análogos de nucleótidos con sustituciones voluminosas con eficacia mejorada. También se proporcionan procedimientos de preparación y uso de dichas polimerasas y muchos otros rasgos característicos que serán evidentes tras una revisión completa de lo que sigue.

Sumario

La presente invención se define por las reivindicaciones adjuntas. En particular, la presente invención proporciona:

[1] Una ADN polimerasa recombinante aislada, comprendiendo dicha ADN polimerasa recombinante una secuencia de aminoácidos que es al menos un 90 % idéntica a los aminoácidos 1-340 de SEQ ID NO: 1, comprendiendo dicha ADN polimerasa recombinante mutaciones en las posiciones de aminoácido 42, 56, 76, 78, 79, 82, 83, 86, 184, 189, 248, 289, 290, 291, 292 y 293, y al menos una posición seleccionada del grupo que consiste en 7, 24, 26, 77, 87, 97, 116, 157, 164, 170, 173, 174, 176, 183, 191, 192, 196, 211, 212, 219, 221, 223, 242, 253, 256, 258, 259, 260, 262, 267, 270, 277, 278, 279, 282, 283, 294, 296, 302, 303, 304, 308, 309, 314, 317, 320, 321, 322, 324, 325, 327 y 328, en la que la identificación de posiciones es en relación con la polimerasa DPO4 natural (SEQ ID NO: 1), en la que los aminoácidos 341-352 en relación con la polimerasa DPO4 natural (SEQ ID NO: 1) están delecionados; y presentando dicha ADN polimerasa recombinante actividad polimerasa.

[2] La ADN polimerasa recombinante de [1], en la que la mutación en la posición 42 es A42V, la mutación en la posición 56 es K56Y, la mutación en la posición 76 es M76W, la mutación en la posición 78 es K78N, la mutación en la posición 79 es E79L, la mutación en la posición 82 es Q82W, la mutación en la posición 83 es Q83G, la mutación en la posición 86 es S86E, la mutación en la posición 184 es P184L, la mutación en la posición 189 es I189W, la mutación en la posición 248 es I248T, la mutación en la posición 289 es V289W, la mutación en la posición 290 es T290K o T290R, la mutación en la posición 291 es E291S, la mutación en la posición 292 es D292Y y la mutación en la posición 293 es L293W.

[3] La ADN polimerasa recombinante de [1], en la que la mutación se selecciona del grupo que consiste en:

- (a) la mutación en la posición 7 es D7V;
- (b) la mutación en la posición 24 es K24S;
- (c) la mutación en la posición 26 es K26L;
- (d) la mutación en la posición 57 es A57P;
- (e) la mutación en la posición 77 es R77S;
- (f) la mutación en la posición 87 es R87Q;
- (g) la mutación en la posición 97 es E97S;
- (h) la mutación en la posición 116 es R116N;
- (i) la mutación en la posición 152 se selecciona del grupo que consiste en K152E, K152G, K152L, K152Q, K152A, K152Y, K152V, K152M y K152L;
- (j) la mutación en la posición 153 se selecciona del grupo que consiste en I153M, I153T, I153R, I153H, I153S, I153Y, I153Q, I153K e I153V;

ES 2 987 024 T3

- (k) la mutación en la posición 155 se selecciona del grupo que consiste en A155G, A155M y A155C;
- (l) la mutación en la posición 156 se selecciona del grupo que consiste en D156R, D156L, D156W, D156A, D156G, D156Y, D156S, D156I, D156M y D156T;
- (m) la mutación en la posición 157 es M157R;
- (n) la mutación en la posición 164 es K164T;
- (o) la mutación en la posición 170 es E170M;
- (p) la mutación en la posición 173 es R173N o R173G;
- (q) la mutación en la posición 176 es R176N;
- (r) la mutación en la posición 183 es V183L;
- (s) la mutación en la posición 187 es G187P o G183E;
- (t) la mutación en la posición 188 es N188Y o N188S;
- (u) la mutación en la posición en 190 se selecciona del grupo que consiste en T190Y, T190V y T190F;
- (v) la mutación en la posición 191 es A191E;
- (w) la mutación en la posición 192 es E192R;
- (x) la mutación en la posición 196 es K196S o K196L;
- (y) la mutación en la posición 211 se selecciona del grupo que consiste en D211P, D211S y D211T;
- (z) la mutación en la posición 212 es K212D o K212F;
- (aa) la mutación en la posición 214 se selecciona del grupo que consiste en K214V, K214S, K214A, K214R y K214Q;
- (bb) la mutación en la posición 215 se selecciona del grupo que consiste en G215R, G215S, G215K, G215T, G215E y G215A;
- (cc) la mutación en la posición 219 se selecciona del grupo que consiste en E219W, E219G, E219A, E219R, E219V y E219L;
- (dd) la mutación en la posición 221 es K221R o K221G;
- (ee) la mutación en la posición 223 se selecciona del grupo que consiste en K223C, K223T, K223R, K223V y K223S;
- (ff) la mutación en la posición 242 es R242W;
- (gg) la mutación en la posición 253 es R253N;
- (hh) la mutación en la posición 256 es R256T o R256N;
- (ii) la mutación en la posición 258 es L258P;
- (jj) la mutación en la posición 259 es R259S;
- (kk) la mutación en la posición 260 es E260Q;
- (ll) la mutación en la posición 262 es K262N;
- (mm) la mutación en la posición 267 es R267Q;
- (nn) la mutación en la posición 270 es E270Q;
- (oo) la mutación en la posición 277 es D277G;

- (pp) la mutación en la posición 278 es K278N;
- (qq) la mutación en la posición 279 es R279Q;
- 5 (rr) la mutación en la posición 282 es K282G o K282S;
- (ss) la mutación en la posición 294 se selecciona del grupo que consiste en D294Q, D294A, D294C, D294S, D294N, D294L, D294P y D294H;
- 10 (tt) la mutación en la posición 295 se selecciona del grupo que consiste en I295Y, I295C, I295P, I295T, I295S, I295W, I295Y, I295F, I295M e I295A;
- (uu) la mutación en la posición 296 se selecciona del grupo que consiste en V296A, V296R, V296S, V296Q y V296Y;
- 15 (vv) la mutación en la posición 297 se selecciona del grupo que consiste en S297H, S297T, S297P, S297W, S297L, S297Q, S297Y, S297C, S297R y S297F;
- (ww) la mutación en la posición 298 se selecciona del grupo que consiste en R298L, R298K, R298T, R298E y R298G;
- 20 (xx) la mutación en la posición 299 se selecciona del grupo que consiste en G299L, G299F, G299W, G299I, G299A, G299H, G299S y G299Y;
- 25 (yy) la mutación en la posición 300 se selecciona del grupo que consiste en R300C, R300S, R300V y R300G;
- (zz) la mutación en la posición 301 se selecciona del grupo que consiste en T301R, T301H, T301F, T301V, T301W, T301E, T301Y, T301M, T301K, T301S y T301Q;
- 30 (aaa) la mutación en la posición 302 es F302E;
- (bbb) la mutación en la posición 303 es P303R;
- (ccc) la mutación en la posición 304 es H304G;
- 35 (ddd) la mutación en la posición 308 es K308Q;
- (eee) la mutación en la posición 309 se selecciona del grupo que consiste en E309N, E309P y E309D;
- 40 (fff) la mutación en la posición 314 es E314Q;
- (ggg) la mutación en la posición 317 es K317Q;
- (hhh) la mutación en la posición 320 se selecciona del grupo que consiste en Q320K, Q320R, Q320M, Q320I, Q320T, Q320N y Q320S;
- 45 (iii) la mutación en la posición 321 se selecciona del grupo que consiste en K321Q, K321N, K321R, K321S, K321G y K321M;
- 50 (jjj) la mutación en la posición 322 se selecciona del grupo que consiste en I322A, I322L, I322C, I322V e I322T;
- (kkk) la mutación en la posición 324 se selecciona del grupo que consiste en E324K, E324N, E324R, E324A, E324G, E324S, E324R, E324W, E3234D y E324K;
- 55 (lll) la mutación en la posición 325 se selecciona del grupo que consiste en E325K, E325Q, E325R, E325G, E325V, E325S y E325A;
- (mmm) la mutación en la posición 326 es D326K;
- 60 (nnn) la mutación en la posición 327 se selecciona del grupo que consiste en E327K, E337Q y E327R; y
- (ooo) la mutación en la posición 328 es R328N o R328S.
- 65 [4] La ADN polimerasa recombinante de [1], que comprende la secuencia de aminoácidos como se expone en una cualquiera de SEQ ID NO: 2-106.

[5] Una composición que comprende una ADN polimerasa recombinante como se expone en uno cualquiera de [1]-[4].

5 [6] La composición de [5], en la que la composición está presente en un sistema de secuenciación de ADN que comprende al menos un sustrato análogo de nucleótido no natural.

[7] Un ácido nucleico modificado que codifica una ADN polimerasa de tipo DPO4 modificada como se expone en uno cualquiera de [1]-[4].

10 Además, en el presente documento se divulgan ADN polimerasas recombinantes y ADN polimerasas modificadas, por ejemplo, DPO4 modificadas, que pueden encontrar uso en aplicaciones tales como, por ejemplo, secuenciación de molécula única por expansión (SBX). Además en el presente documento se divulgan ADN polimerasas recombinantes y variantes de ADN polimerasa modificada que comprenden mutaciones que confieren
15 propiedades, que pueden ser, en particular, deseables para estas aplicaciones. Estas propiedades, por ejemplo, pueden mejorar la capacidad de la polimerasa para utilizar análogos de nucleótidos voluminosos como sustratos durante la polimerización dependiente de molde de una hebra hija. En el presente documento también se divulgan composiciones que componen dichas ADN polimerasas y polimerasas de tipo DPO4 modificadas, ácidos nucleicos que codifican dichas polimerasas modificadas, procedimientos de generación de dichas polimerasas modificadas y procedimientos en los que dichas polimerasas se pueden usar, por ejemplo, para secuenciar un molde de ADN.

En el presente documento también se divulga una ADN polimerasa recombinante aislada, que tiene una secuencia de aminoácidos que es al menos un 80 % idéntica a los aminoácidos 1-340 de SEQ ID NO: 1, y tiene mutaciones en las posiciones de aminoácido 42, 56, 76, 78, 79, 82, 83, 86, 184, 189, 248, 289, 290, 291, 292, 293, y al menos
25 una posición seleccionada del grupo que consiste en 7, 24, 26, 57, 77, 87, 97, 116, 152, 153, 155, 156, 157, 164, 170, 173, 174, 176, 183, 187, 188, 190, 191, 192, 196, 211, 212, 214, 215, 219, 221, 223, 242, 253, 256, 258, 259, 260, 263, 267, 270, 277, 278, 279, 282, 283, 290, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 308, 309, 314, 317, 320, 321, 322, 324, 325, 326, 327 y 328, en la que la identificación de posiciones es en relación con la polimerasa DPO4 natural (SEQ ID NO: 1), y presentando dicha ADN polimerasa recombinante actividad
30 polimerasa. Las mutaciones ejemplares en las posiciones 42, 56, 76, 78, 79, 82, 83, 86, 184, 189, 248, 289, 290, 291, 292 y 293 incluyen A42V, K56Y, M76W, K78N, E79L, Q82W, Q83G, S86E, P184L, I189W, I248T, V289W, T290K o T290R, E291S, D292Y y L293. Las mutaciones ejemplares en las posiciones 7, 24, 26, 57, 77, 87, 97, 116, 152, 153, 155, 156, 157, 164, 170, 173, 174, 176, 183, 187, 188, 190, 191, 192, 196, 211, 212, 214, 215, 219, 221, 223, 242, 253, 256, 258, 259, 260, 263, 267, 270, 277, 278, 279, 282, 283, 290, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 308, 309, 314, 317, 320, 321, 322, 324, 325, 326, 327 y 328, incluyen D7V, K24S, K26L, A57P, R77S, R87Q, E97S, R116N, K152E, K152G, K152L, K152W, K152A, K152Y, K152V, K152M, o K152L, I153M, I153T, I153R, I153H, I153S, I153Y, I153Q, I153K, o I153V, A155G, A155M o A155C D156R, D156L, D156W, D156A, D156G, D156Y, D156S, D156I, D156M, o D156T, K164T, E170M, R173N o R173G, R176N, V183L, G187P o G183E, N188Y o N188S, T190Y, T190V o T190F, A191E, E192R, K196S o K196L, D211P, D211S o D211T, K212D o K212F, K214V, K214S, K214A, K214R o K214Q, G215R, G215S, G215K, G215T, G215E o G215A, E219W, E219G, E219A, E219R, E219V o E219L, K221R o K221G, K223C, K223T, K223R, K223V o K223S, R242W, R253N, R256T o R256N, L258P, R259S, E260Q, K262N, R267Q, E270Q, D277G, K278N, R279Q, K282G o K282S, R290K, D294Q, D294A, D294C, D294S, D294N, D294L, D294P o D294H, I295Y, I295C, I295P, I295T, I295S, I295W, I295Y, I295F, I295M o I295A, V296A, V296R, V296S, V296Q o V296Y, S297H, S297P, S297W, S297L, S297Q, S297Y, S297C, S297R, S297F o S297T, R298L, R298K, R298T, R298E o R298G, G299L, G299W, G299I, G299A, G299H, G299S, G299Y o G299F, R300C, R300S, R300V o R300G, T301R, T301F, T301V, T301W, T301E, T301Y, T301M, T310K, T301S, T301Q, o T301H, F302E, P303R, H304G, K308Q, E309N, E309P o E309D, E314Q, K317Q, Q320K, Q320R, Q320M, Q320I, Q320T, Q320N o Q320S, K321Q, K321N, K321R, K321S, K321G o K321M, I322A, I322L, I322C, I322V o I322T, E324K, E324N, E324A, E324G, E324S, E324R, E324W, E324D o E324K, o E324R, E325K, E325R, E325G, E325V, E325S, E325A o E325Q, D326K, E327K, E337Q, o E327R, y R328N o R328S. La ADN polimerasa de tipo DPO4 recombinante se puede
50 representar por la secuencia de aminoácidos como se expone en una cualquiera de SEQ ID NO: 2-106.

En el presente documento se divulgan composiciones que contienen cualquiera de las ADN polimerasas de tipo DPO4 recombinantes expuestas anteriormente. Las composiciones también pueden contener al menos un sustrato análogo de nucleótido no natural.

Además en el presente documento se divulgan ácidos nucleicos modificados que codifican cualquiera de las ADN polimerasas de tipo DPO4 modificadas expuestas anteriormente.

Breve descripción de las figuras

Las FIGS. 1A y 1B muestran la secuencia de aminoácidos de la proteína polimerasa DPO4 con las regiones de Mut_1 a Mut_13 señaladas y los aminoácidos variables subrayados.

Definiciones

Como se usa en la presente memoria descriptiva y en las reivindicaciones adjuntas, las formas en singular "un", "una" y "el/la" incluyen referentes al plural a menos que el contexto lo indique claramente de otro modo. Por tanto, por ejemplo, la referencia a "una proteína" incluye una pluralidad de proteínas; la referencia a "una célula" incluye mezclas de células, y similares.

El término "aproximadamente" como se usa en el presente documento indica que el valor de una cantidad dada varía en $\pm 10\%$ del valor, u opcionalmente $\pm 5\%$ del valor, o en $\pm 1\%$ del valor así descrito.

"Nucleobase" es una base heterocíclica, tal como adenina, guanina, citosina, timina, uracilo, inosina, xantina, hipoxantina o un derivado heterocíclico, análogo o tautómero de la misma. Una nucleobase puede ser natural o sintética. Los ejemplos no limitantes de nucleobases son adenina, guanina, timina, citosina, uracilo, xantina, hipoxantina, 8-azapurina, purinas sustituidas en la posición 8 con metilo o bromo, 9-oxo-N6-metiladenina, 2-aminoadenina, 7-desazaxantina, 7-desazaguanina, 7-desaza-adenina, N4-etanocitosina, 2,6-diaminopurina, N6-etano-2,6-diaminopurina, 5-metilcitosina, 5-(C3-C6)-alquilcitosina, 5-fluorouracilo, 5-bromouracilo, tiouracilo, pseudoisocitosina, 2-hidroxi-5-metil-4-triazolopiridina, isocitosina, isoguanina, inosina, 7,8-dimetilaloxazina, 6-dihidrotimina, 5,6-dihidrouracilo, 4-metilindol, etenoadenina y las nucleobases no naturales descritas en la pat. de EE. UU. n.ºs 5.432.272 y 6.150.510 y la publicación PCT n.ºs WO 92/002258, WO 93/10820, WO 94/22892 y WO 94/24144, y Fasman ("Practical Handbook of Biochemistry and Molecular Biology", pp. 385-394, 1989, CRC Press, Boca Raton, La.).

"Residuo de nucleobase" incluye nucleótidos, nucleósidos, fragmentos de los mismos y moléculas relacionadas que tienen la propiedad de unirse a un nucleótido complementario. Los desoxinucleótidos y ribonucleótidos, y sus diversos análogos, se contemplan dentro del alcance de esta definición. Los residuos de nucleobase pueden ser miembros de oligómeros y sondas. "Nucleobase" y "residuo de nucleobase" se pueden usar de manera intercambiable en el presente documento y, en general, son sinónimos a menos que el contexto lo indique claramente de otro modo.

Los "polinucleótidos", también llamados ácidos nucleicos, son series de nucleótidos enlazados covalentemente en los que la posición 3' de la pentosa de un nucleótido está unida por un grupo fosfodiéster a la posición 5' del siguiente. El ADN (ácido desoxirribonucleico) y el ARN (ácido ribonucleico) son polinucleótidos biológicos en los que los residuos de nucleótido están enlazados en una secuencia específica por enlaces fosfodiéster. Como se usa en el presente documento, los términos "polinucleótido" u "oligonucleótido" engloban cualquier compuesto polimérico que tenga una cadena principal lineal de nucleótidos. Los oligonucleótidos, también denominados oligómeros, son, en general, polinucleótidos de cadena más corta.

"Ácido nucleico" es un polinucleótido o un oligonucleótido. Una molécula de ácido nucleico puede ser ácido desoxirribonucleico (ADN), ácido ribonucleico (ARN) o una combinación de ambos. Los ácidos nucleicos se denominan, en general, "ácidos nucleicos diana" o "secuencia diana" si se seleccionan como diana para secuenciación. Los ácidos nucleicos pueden ser mezclas o conjuntos de moléculas seleccionados como diana para secuenciación.

Una "secuencia de polinucleótido" o "secuencia de nucleótidos" es un polímero de nucleótidos (un oligonucleótido, un ADN, un ácido nucleico, etc.) o una cadena de caracteres que representa un polímero de nucleótidos, dependiendo del contexto. A partir de cualquier secuencia de polinucleótido especificada, se puede determinar el ácido nucleico dado o bien la secuencia de polinucleótido complementaria (por ejemplo, el ácido nucleico complementario).

Un "polipéptido" es un polímero que comprende dos o más residuos de aminoácido (por ejemplo, un péptido o una proteína). El polímero puede comprender adicionalmente elementos no aminoácidos, tales como marcadores, extintores, grupos bloqueantes o similares, y opcionalmente puede comprender modificaciones, tales como glucosilación o similares. Los residuos de aminoácido del polipéptido pueden ser naturales o no naturales y pueden estar no sustituidos, no modificados, sustituidos o modificados.

Una "secuencia de aminoácidos" es un polímero de residuos de aminoácido (una proteína, polipéptido, etc.) o una cadena de caracteres que representa un polímero de aminoácidos, dependiendo del contexto.

La numeración de un polímero de aminoácidos o nucleótidos dado "corresponde a la numeración de" o es "en relación con" un polímero de aminoácidos o ácido nucleico seleccionado cuando se designa la posición de cualquier componente polimérico dado (residuo de aminoácido, nucleótido incorporado, etc.) por referencia a la misma posición del residuo en el polímero de aminoácidos o nucleótidos seleccionado, en lugar de por la posición real del componente en el polímero dado. De forma similar, la identificación de una posición dada dentro de un polímero de aminoácidos o nucleótidos dado es "en relación con" un polímero de aminoácidos o nucleótidos seleccionado cuando se designa la posición de cualquier componente polimérico dado (residuo de aminoácido, nucleótido incorporado, etc.) por referencia al nombre y posición del residuo en el polímero de aminoácidos o nucleótidos seleccionado, en lugar de por el nombre y posición reales del componente en el polímero dado. La

correspondencia de posiciones típicamente se determina alineando las secuencias de aminoácidos o polinucleótido pertinentes.

El término "recombinante" indica que el material (por ejemplo, un ácido nucleico o una proteína) se ha alterado artificial o sintéticamente (no naturalmente) por intervención humana. La alteración se puede realizar en el material dentro de, o retirado de, su entorno o estado natural. Por ejemplo, un "ácido nucleico recombinante" es uno que se prepara recombinando ácidos nucleicos, por ejemplo, durante la clonación, barajado de ADN u otros procedimientos, o por mutagénesis química u otra; un "polipéptido recombinante" o "proteína recombinante", por ejemplo, es un polipéptido o proteína que se produce por expresión de un ácido nucleico recombinante.

Una "ADN polimerasa de tipo DPO4" es una ADN polimerasa expresada naturalmente por las arqueas, *Sulfolobus solfataricus*, o una ADN polimerasa de la familia Y relacionada que, en general, funciona en la replicación del ADN dañado por un proceso conocido como síntesis por translesión (TLS). Las ADN polimerasas de la familia Y son homólogas a la polimerasa DPO4 (por ejemplo, como se enumera en SEQ ID NO: 1); los ejemplos incluyen las enzimas procarióticas, PolII, PolIV, PolV, la enzima de arquea, Dbh, y las enzimas eucarióticas, ADN polimerasas Rev3p, Rev1p, Pol η, REV3, REV1, Pol ι y Pol κ, así como quimeras de las mismas. Una ADN polimerasa de tipo DPO4 recombinante modificada incluye una o más mutaciones en relación con ADN polimerasas de tipo DPO4 naturales, por ejemplo, una o más mutaciones que incrementan la capacidad de utilizar análogos de nucleótidos voluminosos como sustratos u otra propiedad de la polimerasa, y puede incluir alteraciones o modificaciones adicionales en la ADN polimerasa de tipo DPO4 natural, tales como una o más delecciones, inserciones y/o fusiones de secuencias peptídicas o proteicas adicionales (por ejemplo, para inmovilizar la polimerasa en una superficie o de otro modo marcar la enzima polimerasa).

"Síntesis dirigida por molde", "ensamblaje dirigido por molde", "hibridación dirigida por molde", "unión dirigida por molde" y cualquier otro proceso dirigido por molde, por ejemplo, extensión del cebador, se refiere a un proceso con el que los residuos de nucleótidos o análogos de nucleótidos se unen selectivamente a un ácido nucleico diana complementario y se incorporan a una hebra hija naciente. Una hebra hija producida por una síntesis dirigida por molde es complementaria a la diana monocatenaria a partir de la que se sintetiza. Cabe destacar que la secuencia correspondiente de una hebra diana se puede inferir a partir de la secuencia de su hebra hija, si es conocida. La "polimerización dirigida por molde" es un caso especial de síntesis dirigida por molde, con lo que se polimeriza la hebra hija resultante.

"XNTP" es un sustrato de nucleótido modificado en 5' trifosfato expandible compatible con polimerización enzimática dependiente de molde. Un XNTP tiene dos componentes funcionales distintos; a saber, una nucleobase 5'-trifosfato y una conexión o precursor de conexión que se fija dentro de cada nucleótido en posiciones que permiten la expansión controlada de RT por escisión intranucleotídica.

El "intermedio de Xpandomer" es un producto intermedio (también denominado en el presente documento "hebra hija") ensamblado a partir de XNTP y está formado por un ensamblaje dirigido por molde de XNTP usando un molde de ácido nucleico diana. El intermedio de Xpandomer contiene dos estructuras; a saber, el Xpandomer restringido y la cadena principal primaria. El Xpandomer restringido comprende todas las conexiones en la hebra hija, pero puede comprender todas, una porción o ninguna de las nucleobases 5'-trifosfato según se requiere por el procedimiento. La cadena principal primaria comprende todas las nucleobases 5'-trifosfato contiguas. Durante la etapa del procedimiento en la que la cadena principal primaria se fragmenta o disocia, el Xpandomer restringido ya no está restringido y es el producto de Xpandomer que se extiende a medida que se estiran las conexiones. "Hebra hija híbrida" se refiere a un intermedio de Xpandomer que se hibrida al molde diana.

"Xpandomer" o "producto de Xpandomer" es una construcción molecular sintética producida por expansión de un Xpandomer restringido, que por sí misma se sintetiza por ensamblaje dirigido por molde de XNTP. El Xpandomer se alarga en relación con el molde diana a partir del que se produjo. Se compone de una concatenación de XNTP, incluyendo cada XNTP una conexión que comprende uno o más indicadores que codifican información de secuencia. El Xpandomer está diseñado para expandirse para ser más largo que el molde diana, reduciendo, de este modo, la densidad lineal de la información de secuencia del molde diana a lo largo de su longitud. Además, el Xpandomer proporciona opcionalmente una plataforma para incrementar el tamaño y abundancia de los indicadores, lo que, a su vez, mejora la señal con respecto a ruido para la detección. Una densidad de información lineal menor y señales más fuertes incrementan la resolución y reducen los requisitos de sensibilidad para detectar y descodificar la secuencia de la hebra molde.

"Conexión" o "miembro de conexión" se refiere a un polímero o construcción molecular que tiene una dimensión, en general, lineal y con un resto de extremo en cada uno de dos extremos opuestos. Se fija una conexión a una nucleobase 5'-trifosfato con un enlace en al menos un resto de extremo para formar un XNTP. Los restos de extremo de la conexión se pueden conectar a enlaces escindibles con respecto a la nucleobase 5'-trifosfato que sirven para restringir la conexión en una "configuración restringida". Después de que se sintetice la hebra hija, cada resto de extremo tiene un enlace de extremo que se acopla directa o indirectamente a otras conexiones. Las conexiones acopladas comprenden el Xpandomer restringido que comprende además la hebra hija. Las conexiones tienen una "configuración restringida" y una "configuración expandida". La configuración restringida se

encuentra en XNTP y en la hebra hija. La configuración restringida de la conexión es la precursora de la configuración expandida, como se encuentra en los productos de Xpandomer. La transición de la configuración restringida a la configuración expandida da como resultado la escisión de enlaces selectivamente escindibles que pueden estar dentro de la cadena principal primaria de la hebra hija o enlaces de intraconexión. También se usa una conexión en una configuración restringida donde se añade una conexión para formar la hebra hija después del ensamblaje de la "cadena principal primaria". Las conexiones pueden comprender opcionalmente uno o más indicadores o construcciones de indicador a lo largo de su longitud que pueden codificar información de secuencia de sustratos. La conexión proporciona un medio para expandir la longitud del Xpandomer y, de este modo, reducir la densidad lineal de información de secuencia.

El "elemento de conexión" o "segmento de conexión" es un polímero que tiene una dimensión, en general, lineal con dos extremos terminales, donde los extremos forman enlaces de extremo para concatenar los elementos de conexión. Los elementos de conexión pueden ser segmentos de construcciones de conexión. Dichos polímeros pueden incluir, pero no se limitan a: polietilenglicoles, poliglicoles, polipiridinas, poliisocianatos, poliisocianatos, poli(metacrilatos de triarilmétilo), polialdehídos, polipirrolinonas, poliureas, fosfodiésteres de poliglicol, poliacrilatos, polimetacrilatos, poli(acrilamidas, poli(ésteres de vinilo), poliestirenos, poliamidas, poliuretanos, policarbonatos, polibutiratos, polibutadienos, polibutirilactonas, polipirrolidinonas, poli(fosfonatos de vinilo), poli(acetamidas, polisacáridos, polihialuronatos, poliamidas, polimidas, poliésteres, polietilenos, polipropilenos, poliestirenos, policarbonatos, politereftalatos, polisilanos, poliuretanos, poliéteres, poliaminoácidos, poliglicinas, poliprolinas, polilisina N-sustituida, polipéptidos, péptidos N-sustituidos de cadena lateral, glicina poli-N-sustituida, peptoides, péptidos sustituidos con carboxilo de cadena lateral, homopéptidos, oligonucleótidos, oligonucleótidos de ácido ribonucleico, oligonucleótidos de ácido desoxinucleico, oligonucleótidos modificados para evitar el emparejamiento de bases de Watson-Crick, análogos de oligonucleótidos, ácido policitidílico, ácido poliadenílico, ácido poliuridílico, politimidina, polifosfato, polinucleótidos, polirribonucleótidos, fosfodiésteres de polietilenglicol, análogos de polinucleótidos peptídicos, análogos de treosil-polinucleótidos, análogos de glicol-polinucleótidos, análogos de morfolino-polinucleótidos, análogos de oligómeros de nucleótidos bloqueados, análogos de polipéptidos, polímeros ramificados, polímeros de tipo peine, polímeros de tipo estrella, polímeros dendríticos, copolímeros aleatorios, de gradiente y de bloque, polímeros aniónicos, polímeros catiónicos, polímeros formadores de tallos-bucles, segmentos rígidos y segmentos flexibles.

Una variedad de términos adicionales se definen o de otro modo caracterizan en el presente documento.

Descripción detallada

La divulgación engloba composiciones que comprenden una polimerasa recombinante, por ejemplo, una ADN polimerasa de tipo DPO4 recombinante que incluye una o más mutaciones en comparación con una polimerasa de referencia, por ejemplo, una polimerasa de tipo DPO4 natural. Dependiendo de la mutación o combinación de mutaciones particular, la polimerasa presenta una o más propiedades que encuentran uso, por ejemplo, en aplicaciones de secuenciación de molécula única. Las propiedades ejemplares que presentan diversas polimerasas divulgadas en el presente documento incluyen la capacidad de incorporar análogos de nucleótidos "voluminosos" a una hebra hija en crecimiento durante la replicación de ADN. Las polimerasas pueden incluir uno o más rasgos característicos exógenos o heterólogos en las regiones N y/o C terminales de la proteína para su uso, por ejemplo, en la purificación de la polimerasa recombinante. Las polimerasas también pueden incluir una o más delecciones que faciliten la purificación de la proteína, por ejemplo, incrementando la solubilidad de la proteína producida de forma recombinante.

Estas nuevas polimerasas son, en particular, muy adecuadas para aplicaciones de replicación y/o secuenciación de ADN, en particular, protocolos de secuenciación que incluyan la incorporación de análogos de nucleótidos voluminosos a una hebra hija de ácido nucleico replicada, tal como en el protocolo de secuenciación por expansión (SBX), como se describe además a continuación.

Las polimerasas divulgadas en el presente documento incluyen, por ejemplo, una ADN polimerasa de tipo DPO4 recombinante que comprende mutaciones en las posiciones 42, 56, 76, 78, 79, 82, 83, 86, 184, 189, 248, 289, 290, 291, 292, 293, y al menos una posición seleccionada del grupo que consiste en 7, 24, 26, 57, 77, 87, 97, 116, 152, 153, 155, 156, 157, 164, 170, 173, 174, 176, 183, 187, 188, 190, 191, 192, 196, 211, 212, 214, 215, 219, 221, 223, 242, 253, 256, 258, 259, 260, 263, 267, 270, 277, 278, 279, 282, 283, 290, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 308, 309, 314, 317, 320, 321, 322, 324, 325, 326, 327 y 328, en las que la identificación de posiciones es en relación con la polimerasa DPO4 natural (SEQ ID NO: 1). Opcionalmente, la polimerasa comprende mutaciones en 16 o más, hasta 20 o más, hasta 30 o más, hasta 40 o más, o de 40 a 70 o más de estas posiciones. Las polimerasas divulgadas en el presente documento también pueden poseer una delección de los aminoácidos 341-352 de la proteína natural, correspondiente al "recuadro PIP". Las polimerasas divulgadas en el presente documento pueden incluir mutaciones en residuos adicionales no citados en el presente documento, siempre que dichas mutaciones proporcionen las ventajas funcionales como se analiza además en el presente documento. Las polimerasas divulgadas en el presente documento pueden ser al menos un 80 % idénticas a la SEQ ID NO: 1 (aminoácidos 1-340 de la ADN polimerasa DPO4 natural). Las polimerasas divulgadas en el presente documento pueden ser menos de un 80 % idénticas a la SEQ ID NO: 1, siempre que dichas polimerasas

demuestren capacidades potenciadas para utilizar XNTP como sustratos de polimerización. En el presente documento se describe una serie de sustituciones ejemplares en estas (y otras) posiciones.

ADN polimerasas

En general, están disponibles ADN polimerasas que se pueden modificar para incrementar la capacidad de incorporar sustratos análogos de nucleótidos voluminosos a una hebra de ácido nucleico hija en crecimiento y/u otras propiedades deseables como se describe en el presente documento. Las ADN polimerasas a veces se clasifican en seis grupos o familias, principales, en base a diversas relaciones filogenéticas, por ejemplo, con Pol I de *E. coli* (clase A), Pol II de *E. coli* (clase B), Pol III de *E. coli* (clase C), Pol II de *Euryarchaeotic* (clase D), Pol beta humana (clase X) y UmuC/DinB de *E. coli* y la variante de xerodermia pigmentosa/RAD30 eucariótica (clase Y). Para una revisión de la nomenclatura reciente, véase, por ejemplo, Burgers *et al.* (2001) "Eukaryotic DNA polymerases: proposal for a revised nomenclature" *J Biol. Chem.* 276(47):43487-90. Para una revisión de las polimerasas, véanse, por ejemplo, Hubscher *et al.* (2002) "Eukaryotic DNA Polymerases" *Annual Review of Biochemistry*, vol. 71: 133-163; Alba (2001) "Protein Family Review: Replicative DNA Polymerases", *Genome Biology* 2(1): revisiones 3002.1-3002.4; y Steitz (1999) "DNA polymerases: structural diversity and common mechanisms" *J Biol Chem* 274:17395-17398. La ADN polimerasa se ha estudiado extensamente y se han determinado los mecanismos básicos de acción de muchas. Además, las secuencias de literalmente cientos de polimerasas están disponibles públicamente, y las estructuras cristalinas de muchas de estas se han determinado o se pueden inferir en base a similitud con las estructuras cristalinas resueltas para polimerasas homólogas. Por ejemplo, está disponible la estructura cristalina de DPO4, un tipo preferente de enzima original que se va a modificar de acuerdo con la presente divulgación, véase, por ejemplo, Ling *et al.* (2001) "Crystal Structure of a Y-Family DNA Polymerase in Action: A Mechanism for Error-Prone and Lesion-Bypass Replication" *Cell* 107:91-102.

Las ADN polimerasas que son sustratos preferentes para la mutación para incrementar el uso de análogos de nucleótidos voluminosos como sustratos para la incorporación a hebras hijas de ácido nucleico en crecimiento, y/o para alterar una o más de otras propiedades descritas en el presente documento, incluyen polimerasas DPO4 y otros miembros de la familia Y de ADN polimerasas translesionales, tal como Dbh, y derivados de dichas polimerasas.

La polimerasa que se puede modificar es una ADN polimerasa de tipo DPO4. Por ejemplo, la ADN polimerasa recombinante modificada puede ser homóloga a una ADN polimerasa DPO4 natural. De forma alternativa, la ADN polimerasa recombinante modificada puede ser homóloga a otras ADN polimerasas de clase Y, también conocidas como ADN polimerasas de "translesión", tales como polimerasa Dbh de *Sulfolobus acidocaldarius*. Para una revisión, véase Goodwin y Woodgate (2013) "Translesion DNA Polymerases" *Cold Spring Harb Perspect in Biol* doi:10.1101/cshperspect.a010363. Véase, por ejemplo, la SEQ ID NO: 1 para la secuencia de aminoácidos de la polimerasa DPO4 natural.

La polimerasa que se puede modificar es una polimerasa de tipo ADN Pol Kappa, una polimerasa de tipo ADN Pol Eta, una polimerasa de tipo PrimPol o una polimerasa de tipo Terminator Gamma derivada de cualquier especie adecuada, tal como levadura, ser humano, *S. islandicus* o *T. thermophilus*. La polimerasa que se modifica puede ser versiones de longitud completa o truncadas que incluyan o carezcan de diversos rasgos característicos de la proteína. Determinados rasgos característicos deseados de estas polimerasas también se pueden combinar con cualquiera de las variantes de DPO4 divulgadas en el presente documento.

Muchas polimerasas que son adecuadas para la modificación, por ejemplo, para su uso en tecnologías de secuenciación, están disponibles comercialmente. Por ejemplo, la polimerasa DPO4 está disponible de TREVEGAN® y New England Biolabs®.

Además de las polimerasas naturales, se pueden usar polimerasas quiméricas preparadas a partir de un mosaico de diferentes fuentes. Por ejemplo, las polimerasas de tipo DPO4 preparadas teniendo en cuenta secuencias de más de una polimerasa original se pueden usar como punto de partida para la mutación para producir las polimerasas divulgadas. Se pueden producir quimeras, por ejemplo, usando la consideración de regiones de similitud entre las polimerasas para definir secuencias consenso que se usan en la quimera, o usando tecnologías de barajado de genes en las que múltiples polimerasas relacionadas con DPO4 se barajan aleatoria o semialeatoriamente por medio de técnicas de barajado de genes disponibles (por ejemplo, por medio de "barajado de genes familiares"; véanse Crameri *et al.* (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291; Clackson *et al.* (1991) "Making antibody fragments using phage display libraries" *Nature* 352:624-628; Gibbs *et al.* (2001) "Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling" *Gene* 271:13-20; y Hiraga y Arnold (2003) "General method for sequence-independent site-directed chimeragenesis" *J. Mol. Biol.* 330:287-296). En estos procedimientos, los puntos de recombinación se pueden predeterminar de modo que los fragmentos génicos se ensamblen en el orden correcto. Sin embargo, las combinaciones, por ejemplo, quimeras, se pueden formar aleatoriamente. Se pueden introducir en las quimeras mutaciones apropiadas para mejorar la incorporación de sustratos análogos de nucleótidos voluminosos u otra propiedad deseable.

Análogos de nucleótidos

Como se analiza, diversas polimerasas divulgadas en el presente documento pueden incorporar uno o más análogos de nucleótidos a una cadena oligonucleotídica en crecimiento. Tras la incorporación, el análogo puede dejar un residuo que sea el mismo que o diferente a un nucleótido natural en el oligonucleótido en crecimiento (la polimerasa puede incorporar cualquier resto no estándar del análogo o puede escindirlo durante la incorporación al oligonucleótido). Un "análogo de nucleótido" en el presente documento es un compuesto que, en una aplicación particular, funciona de una manera similar o análoga a un nucleósido trifosfato natural (un "nucleótido") y de otro modo no indica ninguna estructura particular. Un análogo de nucleótido es un análogo distinto de un nucleótido natural estándar, es decir, distinto de A, G, C, T o U, aunque, tras la incorporación al oligonucleótido, el residuo resultante en el oligonucleótido puede ser el mismo que (o diferente de) un residuo de A, G, C, T o U.

Están disponibles muchos análogos de nucleótidos y se pueden incorporar por las polimerasas. Estos incluyen estructuras de análogos con similitud central con nucleótidos naturales, tales como las que comprenden uno o más sustituyentes en un resto fosfato, glucídico o de base del nucleósido o nucleótido en relación con un nucleósido o nucleótido natural.

Los análogos de nucleótidos también se pueden modificar para lograr cualquiera de las propiedades mejoradas deseadas. Por ejemplo, se pueden incorporar diversas conexiones, conectores u otros sustituyentes a análogos para crear un análogo de nucleótido "voluminoso", en el que se entiende que el término "voluminoso" significa que el tamaño del análogo es sustancialmente más grande que el de un nucleótido natural, no indicando al mismo tiempo ninguna dimensión particular. Por ejemplo, el análogo puede incluir un compuesto sustituido (es decir, un "XNTP", como se divulga en la patente de EE. UU. n.º 7.939.259 y en la publicación PCT n.º WO 2016/081871 de Kokoris *et al.*) de fórmula:



Como se muestra en la fórmula anterior, la construcción de XNTP monomérica tiene un residuo de nucleobase, N, que tiene dos restos separados por un enlace selectivamente escindible (V), fijándose cada resto a un extremo de una conexión (T). Los extremos de conexión se pueden fijar a las modificaciones del grupo conector en el heterociclo, el grupo ribosa o la cadena principal de fosfato. El sustrato monomérico también tiene un sitio de escisión intrasustrato situado dentro de la cadena principal de fosforribosilo de modo que la escisión proporcione la expansión de la conexión restringida. Por ejemplo, para sintetizar un monómero de XATP, se puede usar el conector de amino en 8-[(6-amino)hexil]-amino-ATP o N6-(6-amino)hexil-ATP como primer punto de fijación de conexión y, se puede usar un conector de cadena principal mixta, tal como la modificación no puente fosforamidato de (N-1-aminoalquilo) o fosfonato de (2-aminoetil), como segundo punto de fijación de conexión. Además, se puede usar una modificación de cadena principal puente, tal como un fosforamidato (3' O--P--N 5') o un fosforotiolato (3' O--P--S 5'), por ejemplo, para la escisión química selectiva de la cadena principal primaria. R¹ y R² son grupos de extremo configurados como es apropiado para el protocolo de síntesis en el que se usa la construcción de sustrato. Por ejemplo, R¹ = 5'-trifosfato y R² = 3'-OH para un protocolo de polimerasa. El R¹ 5' trifosfato puede incluir modificaciones de cadena principal mixta, tales como un fosfonato de aminoetil o un fosforotiolato 3'-O--P--S-5', para posibilitar el enlace de conexión y la escisión de la cadena principal, respectivamente. Opcionalmente, R² se puede configurar con un grupo de bloqueo reversible para la adición de único sustrato cíclica. De forma alternativa, R¹ y R² se pueden configurar con grupos de extremo conectores para el acoplamiento químico. R¹ y R² pueden ser del tipo general XR, en el que X es un grupo de enlace y el R divulgado es un grupo funcional. Se pueden encontrar estructuras atómicas detalladas de sustratos adecuados para las variantes de polimerasa divulgadas, por ejemplo, en Vaghefi, M. (2005) "Nucleoside Triphosphates and their Analogs" CRC Press Taylor & Francis Group.

Aplicaciones para capacidades incrementadas para incorporar sustratos análogos de nucleótidos voluminosos

Las polimerasas divulgadas en el presente documento, por ejemplo, polimerasas recombinantes modificadas o variantes, se pueden usar en combinación con nucleótidos y/o análogos de nucleótidos y moldes de ácido nucleico (ADN o ARN) para copiar el ácido nucleico molde. Es decir, se hace reaccionar una mezcla de la polimerasa, nucleótidos/análogos y opcionalmente otros reactivos apropiados, el molde y un resto de iniciación de la replicación (por ejemplo, cebador) de modo que la polimerasa sintetice una hebra de ácido nucleico hija (por ejemplo, extienda el cebador) de una manera dependiente de molde. El resto de iniciación de la replicación puede ser un cebador oligonucleotídico estándar o, de forma alternativa, un componente del molde, por ejemplo, el molde puede ser un ADN monocatenario autocebante, un ADN bicatenario mellado o similares. De forma similar, una proteína terminal puede servir como resto de iniciación. Se puede incorporar al menos un análogo de nucleótido al ADN. El ADN

molde puede ser un ADN lineal o circular y, en determinadas aplicaciones, es deseable un molde circular (por ejemplo, para replicación por círculo rodante o para secuenciación de moldes circulares). Opcionalmente, la composición puede estar presente en un sistema de replicación y/o secuenciación de ADN automatizado.

La hebra de ácido nucleico hija puede ser un intermedio de Xpandomer compuesto por XNTP, como se divulga en la patente de EE. UU. n.º 7.939.259 y en la publicación PCT n.º WO 2016/081871 de Kokoris *et al.* y cedida a Stratos Genomics. Stratos Genomics ha desarrollado un procedimiento llamado secuenciación por expansión ("SBX") que usa una ADN polimerasa para transcribir la secuencia de ADN en un polímero mensurable llamado "Xpandomer". En términos generales, un Xpandomer codifica (analiza) los datos de secuencia de nucleótidos del ácido nucleico diana en un formato linealmente expandido, mejorando, de este modo, la resolución espacial, opcionalmente con amplificación de la intensidad de la señal. La secuencia transcrita se codifica a lo largo de la cadena principal de Xpandomer en indicadores de alta señal con respecto a ruido que están separados por ~10 nm y están diseñados para respuestas de alta señal con respecto a ruido bien diferenciadas. Estas diferencias proporcionan significativas potenciaciones del rendimiento en la eficacia de lectura de secuencia y exactitud de los Xpandomer en relación con el ADN natural. Los Xpandomer pueden posibilitar varias tecnologías de secuenciación de ADN de nueva generación y son muy adecuados para la secuenciación por nanoporos. Como se analiza anteriormente, un procedimiento de síntesis de Xpandomer usa XNTP como análogos de ácidos nucleicos para extender la síntesis dependiente de molde y usa una variante de ADN polimerasa como catalizador.

Polimerasas mutantes

En la presente divulgación se usan opcionalmente diversos tipos de mutagénesis, por ejemplo, para modificar polimerasas para producir variantes, por ejemplo, de acuerdo con modelos de polimerasa y predicciones de modelo como se analiza anteriormente, o usando enfoques mutacionales aleatorios o semialeatorios. En general, se puede usar cualquier procedimiento de mutagénesis disponible para preparar mutantes de polimerasa. Dichos procedimientos de mutagénesis incluyen opcionalmente la selección de ácidos nucleicos y polipéptidos mutantes para una o más actividades de interés (por ejemplo, la capacidad de incorporar análogos de nucleótidos voluminosos en una hebra de ácido nucleico hija). Los procedimientos que se pueden usar incluyen, pero no se limitan a: mutagénesis puntual dirigida a sitio, mutagénesis puntual aleatoria, recombinación homóloga *in vitro* o *in vivo* (barajado de ADN y PCR por superposición combinatoria), mutagénesis que usa moldes que contienen uracilo, mutagénesis dirigida a oligonucleótido, mutagénesis de ADN modificado con fosforotioato, mutagénesis usando ADN bicatenario con huecos, reparación de emparejamientos erróneos puntuales, mutagénesis usando cepas huésped con alteración de la vía reparadora, mutagénesis por delección por restricción-selección y restricción-purificación, mutagénesis por síntesis génica total, PCR degenerada, reparación de roturas bicatenarias, y muchos otros conocidos por expertos en la técnica. La polimerasa de partida para la mutación puede ser cualquiera de las destacadas en el presente documento, incluyendo polimerasa DPO4 natural.

Opcionalmente, la mutagénesis se puede guiar por información conocida (por ejemplo, diseño "racional" o "semirracional") de una molécula de polimerasa natural, o de una polimerasa alterada o mutada conocida (por ejemplo, usando una polimerasa mutante existente como se destaca en las referencias precedentes), por ejemplo, secuencia, comparaciones de secuencias, propiedades físicas, estructura cristalina y/o similares como se analiza anteriormente. Sin embargo, la modificación puede ser esencialmente aleatoria (por ejemplo, como en el barajado de ADN clásico o "familiar", véase, por ejemplo, Cramer *et al.* (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291).

Se encuentra información adicional sobre los formatos de mutación en: Sambrook *et al.*, *Molecular Cloning--A Laboratory Manual* (3.ª ed.), vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 2000 ("Sambrook"); *Current Protocols in Molecular Biology*, F. M. Ausubel *et al.*, eds., Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (complementados hasta 2011) ("Ausubel") y *PCR Protocols A Guide to Methods and Applications* (Innis *et al.* eds. Academic Press Inc. San Diego, Calif. (1990) ("Innis"). Las siguientes publicaciones y referencias citadas proporcionan detalles adicionales sobre los formatos de mutación: Arnold, Protein engineering for unusual environments, *Current Opinion in Biotechnology* 4:450-455 (1993); Bass *et al.*, Mutant Trp repressors with new DNA-binding specificities, *Science* 242:240-245 (1988); Bordo y Argos (1991) Suggestions for "Safe" Residue Substitutions in Site-directed Mutagenesis 217:721-729; Botstein y Shortle, Strategies and applications of in vitro mutagenesis, *Science* 229:1193-1201 (1985); Carter *et al.*, Improved oligonucleotide site-directed mutagenesis using M13 vectors, *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, Site-directed mutagenesis, *Biochem. J.* 237:1-7 (1986); Carter, Improved oligonucleotide-directed mutagenesis using M13 vectors, *Methods in Enzymol.* 154: 382-403 (1987); Dale *et al.*, Oligonucleotide-directed random mutagenesis using the phosphorothioate method, *Methods Mol. Biol.* 57:369-374 (1996); Eghtedarzadeh y Henikoff, Use of oligonucleotides to generate large deletions, *Nucl. Acids Res.* 14: 5115 (1986); Fritz *et al.*, Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro, *Nucl. Acids Res.* 16: 6987-6999 (1988); Grundstrom *et al.*, Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis, *Nucl. Acids Res.* 13: 3305-3316 (1985); Hayes (2002) Combining Computational and Experimental Screening for rapid Optimization of Protein Properties *PNAS* 99(25) 15926-15931; Kunkel, The efficiency of oligonucleotide directed mutagenesis, en *Nucleic Acids & Molecular Biology* (Eckstein, F. y Lilley, D. M. J. eds., Springer Verlag, Berlín) (1987); Kunkel, Rapid and efficient site-specific mutagenesis without phenotypic selection,

Proc. Natl. Acad. Sci. USA 82:488-492 (1985); Kunkel *et al.*, Rapid and efficient site-specific mutagenesis without phenotypic selection, *Methods in Enzymol.* 154, 367-382 (1987); Kramer *et al.*, The gapped duplex DNA approach to oligonucleotide-directed mutation construction, *Nucl. Acids Res.* 12: 9441-9456 (1984); Kramer y Fritz Oligonucleotide-directed construction of mutations via gapped duplex DNA, *Methods in Enzymol.* 154:350-367 (1987); Kramer *et al.*, Point Mismatch Repair, *Cell* 38:879-887 (1984); Kramer *et al.*, Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations, *Nucl. Acids Res.* 16: 7207 (1988); Ling *et al.*, Approaches to DNA mutagenesis: an overview, *Anal Biochem.* 254(2): 157-178 (1997); Lorimer y Pastan *Nucleic Acids Res.* 23, 3067-8 (1995); Mandecki, Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis, *Proc. Natl. Acad. Sci. USA*, 83:7177-7181(1986); Nakamaye y Eckstein, Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 14: 9679-9698 (1986); Nambiar *et al.*, Total synthesis and cloning of a gene coding for the ribonuclease S protein, *Science* 223: 1299-1301(1984); Sakamar y Khorana, Total synthesis and expression of a gene for the α -subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin), *Nucl. Acids Res.* 14: 6361-6372 (1988); Sayers *et al.*, Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 16:791-802 (1988); Sayers *et al.* Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide, (1988) *Nucl. Acids Res.* 16: 803-814; Sieber, *et al.*, *Nature Biotechnology*, 19:456-460 (2001); Smith, In vitro mutagenesis, *Ann. Rev. Genet.* 19:423-462 (1985); *Methods in Enzymol.* 100: 468-500 (1983); *Methods in Enzymol.* 154: 329-350 (1987); Stemmer, *Nature* 370, 389-91(1994); Taylor *et al.*, The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA, *Nucl. Acids Res.* 13: 8749-8764 (1985); Taylor *et al.*, The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA, *Nucl. Acids Res.* 13: 8765-8787 (1985); Wells *et al.*, Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin, *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986); Wells *et al.*, Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites, *Gene* 34:315-323 (1985); Zoller y Smith, Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment, *Nucleic Acids Res.* 10:6487-6500 (1982); Zoller y Smith, Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, *Methods in Enzymol.* 100:468-500 (1983); Zoller y Smith, Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template, *Methods in Enzymol.* 154:329-350 (1987); Clackson *et al.* (1991) "Making antibody fragments using phage display libraries" *Nature* 352:624-628; Gibbs *et al.* (2001) "Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling" *Gene* 271:13-20; y Hiraga y Arnold (2003) "General method for sequence-independent site-directed chimeragenesis: *J. Mol. Biol.* 330:287-296. Se pueden encontrar detalles adicionales sobre muchos de los procedimientos anteriores en *Methods in Enzymology*, volumen 154, que también describe controles útiles para solucionar problemas con diversos procedimientos de mutagénesis.

Polimerasas de cribado

Se pueden usar cribado u otros protocolos para determinar si una polimerasa presenta una actividad modificada, por ejemplo, para un análogo de nucleótido, en comparación con una ADN polimerasa original. Por ejemplo, la capacidad de unirse e incorporar análogos de nucleótidos voluminosos en una hebra hija durante la síntesis de ADN dependiente de molde. En el presente documento se describen ensayos para dichas propiedades y similares. El rendimiento de una polimerasa recombinante en una reacción de extensión del cebador se puede examinar para someter a ensayo propiedades tales como incorporaciones de análogos de nucleótidos, etc., como se describe en el presente documento.

Se puede preparar una colección de ADN polimerasas recombinantes y cribar para determinar estas propiedades. Por ejemplo, se puede hacer que una pluralidad de miembros de la colección incluya una o más mutaciones que alteren las incorporaciones y/o mutaciones generadas aleatoriamente (por ejemplo, donde diferentes miembros incluyen diferentes mutaciones o diferentes combinaciones de mutaciones) y, a continuación, la colección se puede cribar para determinar las propiedades de interés (por ejemplo, incorporaciones, etc.). En general, la colección se puede cribar para identificar al menos un miembro que comprenda una actividad modificada de interés.

Las colecciones de polimerasas pueden ser de naturaleza física o bien lógica. Además, se puede usar cualquiera de una amplia variedad de formatos de colecciones. Por ejemplo, las polimerasas se pueden fijar a superficies sólidas en matrices de proteínas. De forma similar, se pueden construir matrices de polimerasas en fase líquida (por ejemplo, en placas de micropocillos) para manipulaciones de fluidos de alto rendimiento convenientes de soluciones que comprenden polimerasas. También se pueden construir colecciones líquidas, en emulsión o en fase de gel de células que expresen polimerasas recombinantes, por ejemplo, en placas de micropocillos o en placas de agar. Se pueden producir colecciones de presentación en fagos de polimerasas o dominios de polimerasa (por ejemplo, incluyendo la región del sitio activo o regiones de estabilidad entre dominios). Asimismo, se pueden usar colecciones de presentación en levaduras. Las instrucciones para preparar y usar colecciones se pueden encontrar, por ejemplo, en Sambrook, Ausubel y Berger, a los que se hace referencia en el presente documento.

Para la generación de colecciones que implican la transferencia de fluidos hacia o desde placas de microvaloración,

se usa opcionalmente una estación de manipulación de fluidos. Están disponibles comercialmente varias estaciones de manipulación de fluidos existentes para realizar dichas transferencias, incluyendo, por ejemplo, los sistemas Zymate de Caliper Life Sciences (Hopkinton, Mass.) y otras estaciones que utilizan pipetas automáticas, por ejemplo, junto con la robótica para el movimiento de placas (por ejemplo, el robot ORCA®, que se usa en una variedad de sistemas de laboratorio disponibles, por ejemplo, de Beckman Coulter, Inc. (Fullerton, Calif.).

La manipulación de fluidos se puede realizar en microchips, por ejemplo, que impliquen la transferencia de materiales desde placas de micropocillos u otros pocillos a través de microcanales en los chips a sitios de destino (regiones de microcanales, pocillos, cámaras o similares). Los sistemas de microfluidos disponibles comercialmente incluyen los de Hewlett-Packard/Agilent Technologies (por ejemplo, el bioanalizador HP2100) y el sistema de cribado ultrarrápido Caliper. El sistema de cribado ultrarrápido Caliper proporciona una interfaz de ejemplo entre los formatos de colecciones de micropocillos estándar y las tecnologías Labchip. La plataforma de nanogotículas de RainDance Technologies proporciona otro procedimiento para manipular una gran serie de reacciones espacialmente separadas. Además, la literatura técnica y de patente incluye muchos ejemplos de sistemas de microfluidos que se pueden interconectar directamente con placas de micropocillos para la manipulación de fluidos.

Marcas y otros rasgos característicos de polimerasa opcionales

La ADN polimerasa recombinante incluye opcionalmente rasgos característicos exógenos o heterólogos adicionales a la polimerasa. Por ejemplo, la polimerasa recombinante incluye opcionalmente una o más marcas, por ejemplo, purificación, unión a sustrato, u otras marcas, tales como una marca de polihistidina, una marca His10, una marca His6, una marca de alanina, una marca Ala16, una marca de biotina, una secuencia de reconocimiento de biotina-ligasa u otro sitio de fijación a biotina (por ejemplo, una BiTag o una Btag o variante de la misma, por ejemplo, BtagV1-11), una marca GST, una marca S, una marca SNAP, una marca HA, una marca DSB (Sso7D), una marca de lisina, una NanoTag, una marca Cmyc, una marca o conector que comprende los aminoácidos glicina y serina, una marca o conector que comprende los aminoácidos glicina, serina, alanina e histidina, una marca o conector que comprende los aminoácidos glicina, arginina, lisina, glutamina y prolina, una pluralidad de marcas de polihistidina, una pluralidad de marcas His10, una pluralidad de marcas His6, una pluralidad de marcas de alanina, una pluralidad de marcas Ala16, una pluralidad de marcas de biotina, una pluralidad de marcas GST, una pluralidad de BiTag, una pluralidad de marcas S, una pluralidad de marcas SNAP, una pluralidad de marcas HA, una pluralidad de marcas DSB (Sso7D), una pluralidad de marcas de lisina, una pluralidad de NanoTag, una pluralidad de marcas Cmyc, una pluralidad de marcas o conectores que comprenden los aminoácidos glicina y serina, una pluralidad de marcas o conectores que comprenden los aminoácidos glicina, serina, alanina e histidina, una pluralidad de marcas o conectores que comprenden los aminoácidos glicina, arginina, lisina, glutamina y prolina, biotina, avidina, un anticuerpo o dominio de anticuerpo, fragmento de anticuerpo, antígeno, receptor, dominio de receptor, fragmento de receptor, o ligando, uno o más sitios de proteasa (por ejemplo, sitio de factor Xa, enterocinasa o trombina), un tinte, un aceptor, un extintor, un dominio de unión a ADN (por ejemplo, un dominio hélice-horquilla-hélice de topoisomerasa V), o una combinación de los mismos. Los uno o más rasgos característicos exógenos o heterólogos en las regiones N y/o C terminales de la polimerasa pueden encontrar uso no solo con propósitos de purificación, inmovilización de la polimerasa en un sustrato y similares, sino que también pueden ser útiles para alterar una o más propiedades de la polimerasa.

Los uno o más rasgos característicos exógenos o heterólogos se pueden incluir internos a la polimerasa, en la región N terminal de la polimerasa, en la región C terminal de la polimerasa, o en ambas regiones N terminal y C terminal de la polimerasa. Cuando la polimerasa incluye un rasgo característico exógeno o heterólogo en las regiones N terminal y C terminal, los rasgos característicos exógenos o heterólogos pueden ser los mismos (por ejemplo, una marca de polihistidina, por ejemplo, una marca His10, tanto en las regiones N como C terminales) o diferentes (por ejemplo, una secuencia de reconocimiento de biotina-ligasa en la región N terminal y una marca de polihistidina, por ejemplo, marca His10, en la región C terminal). Opcionalmente, una región terminal (por ejemplo, la región N o C terminal) de una polimerasa divulgada en el presente documento puede comprender dos o más rasgos característicos exógenos o heterólogos que pueden ser los mismos o diferentes (por ejemplo, una secuencia de reconocimiento de biotina-ligasa y una marca de polihistidina en la región N terminal, una secuencia de reconocimiento de biotina-ligasa, una marca de polihistidina y un sitio de reconocimiento del factor Xa en la región N terminal, y similares). Como algunos ejemplos, la polimerasa puede incluir una marca de polihistidina en la región C terminal, una secuencia de reconocimiento de biotina-ligasa y una marca de polihistidina en la región N terminal, una secuencia de reconocimiento de biotina-ligasa y una marca de polihistidina en la región N terminal y una marca de polihistidina en la región C terminal, o una marca de polihistidina y una secuencia de reconocimiento de biotina-ligasa en la región C terminal.

Preparación y aislamiento de polimerasas recombinantes

En general, los ácidos nucleicos que codifican una polimerasa divulgada en el presente documento se pueden preparar por clonación, recombinación, síntesis *in vitro*, amplificación *in vitro* y/u otros procedimientos disponibles. Se puede usar una variedad de procedimientos recombinantes para expresar un vector de expresión que codifique

una polimerasa divulgada en el presente documento. Los procedimientos para preparar ácidos nucleicos recombinantes, la expresión y el aislamiento de productos expresados son bien conocidos y se describen en la técnica. En el presente documento se describe una serie de mutaciones ejemplares y combinaciones de mutaciones, así como estrategias para el diseño de mutaciones deseables. Los procedimientos para preparar y seleccionar mutaciones en el sitio activo de las polimerasas, incluyendo la modificación de rasgos característicos estéricos en o cerca del sitio activo para permitir un acceso mejorado por análogos de nucleótidos, se encuentran anteriormente en el presente documento y, por ejemplo, en las publicaciones PCT n.ºs WO 2007/076057 y WO 2008/051530.

Las referencias útiles adicionales para procedimientos de mutación, recombinación y manipulación de ácidos nucleicos *in vitro* (incluyendo clonación, expresión, PCR y similares) incluyen Berger y Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, volumen 152, Academic Press, Inc., San Diego, Calif. (Berger); Kaufman *et al.* (2003) *Handbook of Molecular and Cellular Methods in Biology and Medicine*, segunda edición, Ceske (ed) CRC Press (Kaufman); y *The Nucleic Acid Protocols Handbook* Ralph Rapley (ed) (2000) Cold Spring Harbor, Humana Press Inc (Rapley); Chen *et al.* (ed) *PCR Cloning Protocols*, segunda edición (Methods in Molecular Biology, volumen 192) Humana Press; y en Viljoen *et al.* (2005) *Molecular Diagnostic PCR Handbook* Springer, ISBN 1402034032.

Además, está disponible comercialmente una plétora de kits para la purificación de plásmidos u otros ácidos nucleicos pertinentes de células (véase, por ejemplo, EasyPrep™ FlexiPrep™, ambos de Pharmacia Biotech; StrataClean™, de Stratagene; y, QIAprep™ de Qiagen). Cualquier ácido nucleico aislado y/o purificado se puede manipular además para producir otros ácidos nucleicos, usar para transfectar células, incorporar en vectores relacionados para infectar organismos para la expresión y/o similares. Los vectores de clonación típicos contienen finalizadores de transcripción y traducción, secuencias de iniciación de transcripción y traducción y promotores útiles para la regulación de la expresión del ácido nucleico diana particular. Los vectores comprenden opcionalmente casetes de expresión genéricos que contienen al menos una secuencia de finalizador independiente, secuencias que permiten la replicación del casete en eucariotas o procariotas, o ambos (por ejemplo, vectores transportadores) y marcadores de selección para sistemas tanto procarióticos como eucarióticos. Los vectores son adecuados para la replicación e integración en procariotas, eucariotas o ambos.

Otras referencias útiles, por ejemplo, para el aislamiento y cultivo celulares (por ejemplo, para posterior aislamiento de ácidos nucleicos) incluyen Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, tercera edición, Wiley-Liss, New York y las referencias citadas en la misma; Payne *et al.* (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. New York, N.Y.; Gamborg y Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture; Fundamental Methods* Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) y Atlas y Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, Fla.

La divulgación también engloba ácidos nucleicos que codifican las polimerasas recombinantes divulgadas. Un aminoácido particular se puede codificar por múltiples codones, y determinados sistemas de traducción (por ejemplo, células procarióticas o eucarióticas) a menudo presentan un sesgo de codones; por ejemplo, diferentes organismos a menudo prefieren uno de los varios sinónimos que pueden ser codones que codifiquen el mismo aminoácido. Como tales, los ácidos nucleicos pueden estar opcionalmente "optimizados a nivel de codón", lo que significa que los ácidos nucleicos se sintetizan para que incluyan codones que sean preferentes por el sistema de traducción particular que se emplea para expresar la polimerasa. Por ejemplo, cuando es deseable expresar la polimerasa en una célula bacteriana (o incluso en una cepa particular de bacterias), el ácido nucleico se puede sintetizar para que incluya codones encontrados con mayor frecuencia en el genoma de esa célula bacteriana, para una expresión eficaz de la polimerasa. Se puede emplear una estrategia similar cuando es deseable expresar la polimerasa en una célula eucariótica, por ejemplo, el ácido nucleico puede incluir codones preferentes por esa célula eucariótica.

Son conocidos una variedad de procedimientos de aislamiento y detección de proteínas que se pueden usar para aislar polimerasas, por ejemplo, de cultivos recombinantes de células que expresan las polimerasas recombinantes divulgadas. Son bien conocidos en la técnica una variedad de procedimientos de aislamiento y detección de proteínas, incluyendo, por ejemplo, los expuestos en R. Scopes, *Protein Purification*, Springer-Verlag, N.Y. (1982); Deutscher, *Methods in Enzymology*, vol. 182: *Guide to Protein Purification*, Academic Press, Inc. N.Y. (1990); Sandana (1997) *Bioseparation of Proteins*, Academic Press, Inc.; Bollag *et al.* (1996) *Protein Methods*, 2.ª edición, Wiley-Liss, NY; Walker (1996) *The Protein Protocols Handbook* Humana Press, NJ, Harris y Angal (1990) *Protein Purification Applications: A Practical Approach* IRL Press at Oxford, Oxford, Inglaterra; Harris y Angal *Protein Purification Methods: A Practical Approach* IRL Press at Oxford, Oxford, Inglaterra; Scopes (1993) *Protein Purification: Principles and Practice*, 3.ª edición, Springer Verlag, NY; Janson y Ryden (1998) *Protein Purification: Principles, High Resolution Methods and Applications*, segunda edición, Wiley-VCH, NY; y Walker (1998) *Protein Protocols on CD-ROM* Humana Press, NJ; y las referencias citadas en los mismos. Se pueden encontrar detalles adicionales con respecto a los procedimientos de detección y purificación de proteínas en Satinder Ahuja ed., *Handbook of Bioseparations*, Academic Press (2000).

Variantes y secuencias de ácido nucleico y polipeptídicas

En el presente documento también se divulgan secuencias de polinucleótido que codifican, por ejemplo, una polimerasa como se describe en el presente documento. Se proporcionan ejemplos de secuencias de polimerasa que incluyen rasgos característicos encontrados en el presente documento, por ejemplo, como en la tabla 2. Las combinaciones de cualquiera de las mutaciones destacadas en el presente documento también son rasgos característicos de la divulgación.

En consecuencia, en el presente documento se divulgan una variedad de polipéptidos (polimerasas) y polinucleótidos (ácidos nucleicos que codifican polimerasas). Los polinucleótidos ejemplares pueden incluir, por ejemplo, cualquier polinucleótido que codifique una polimerasa de la tabla 2 o de otro modo descrita en el presente documento. Debido a la redundancia del código genético, muchos polinucleótidos codifican de forma equivalente una secuencia de polimerasa dada. De forma similar, un ácido nucleico artificial o recombinante que se hibrida a un polinucleótido indicado anteriormente en condiciones altamente rigurosas en sustancialmente toda la longitud del ácido nucleico (y que sea distinto de un polinucleótido natural) es un polinucleótido divulgado en el presente documento. Una composición puede incluir un polipéptido divulgado en el presente documento y un excipiente (por ejemplo, tampón, agua, excipiente farmacéuticamente aceptable, etc.). En el presente documento también se divulga un anticuerpo o antisuero específicamente inmunorreactivo con un polipéptido divulgado en el presente documento (por ejemplo, que reconoce específicamente un rasgo característico de la polimerasa que confiere ramificación disminuida o estabilidad de complejo incrementada).

Un vector (por ejemplo, un plásmido, un cósmido, un fago, un virus, etc.) puede comprender un polinucleótido divulgado en el presente documento. El vector puede ser un vector de expresión. El vector de expresión puede incluir un promotor enlazado de forma funcional a uno o más de los polinucleótidos divulgados en el presente documento. Una célula puede comprender un vector que incluya un polinucleótido divulgado en el presente documento.

Variaciones conservadoras

Debido a la redundancia del código genético, las "sustituciones sinónimas" (es decir, sustituciones en una secuencia de ácido nucleico que no dan como resultado una alteración en un polipéptido codificado) son un rasgo característico implícito de cada secuencia de ácido nucleico que codifica una secuencia de aminoácidos. De forma similar, las "sustituciones aminoácidas conservadoras", donde uno o un número limitado de aminoácidos en una secuencia de aminoácidos se sustituyen con diferentes aminoácidos con propiedades altamente similares, también se identifican fácilmente como altamente similares a una construcción divulgada. Dichas variaciones conservadoras de cada secuencia divulgada son un rasgo característico de la presente divulgación.

"Variaciones conservadoras" de una secuencia de ácido nucleico particular se refiere a aquellos ácidos nucleicos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas o, cuando el ácido nucleico no codifica una secuencia de aminoácidos, a secuencias esencialmente idénticas. Un experto reconocerá que las sustituciones, deleciones o adiciones individuales que alteran, añaden o delecionan un único aminoácido o un pequeño porcentaje de aminoácidos (típicamente menos de un 5 %, más típicamente menos de un 4 %, 2 % o 1 %) en una secuencia codificada son "variaciones modificadas de forma conservadora" donde las alteraciones dan como resultado la deleción de un aminoácido, la adición de un aminoácido o la sustitución de un aminoácido con un aminoácido químicamente similar, reteniendo al mismo tiempo el rasgo característico mutacional pertinente (por ejemplo, la sustitución conservadora puede ser de un residuo distal a la región del sitio activo, o distal a una región de estabilidad entre dominios). Por tanto, las "variaciones conservadoras" de la divulgación de una secuencia polipeptídica enumerada de la presente divulgación incluyen sustituciones de un pequeño porcentaje, típicamente menos de un 5 %, más típicamente menos de un 2 % o un 1 %, de los aminoácidos de la secuencia polipeptídica, con un aminoácido del mismo grupo de sustitución conservadora. Finalmente, la adición de secuencias que no alteren la actividad codificada de una molécula de ácido nucleico, tal como la adición de una secuencia no funcional o marcadora (intrones en el ácido nucleico, poli His o secuencias similares en el polipéptido codificado, etc.), es una variación conservadora del ácido nucleico o polipéptido básico.

Las tablas de sustituciones conservadoras que proporcionan aminoácidos funcionalmente similares son bien conocidas en la técnica, donde un residuo de aminoácido se sustituye por otro residuo de aminoácido que tiene propiedades químicas similares (por ejemplo, cadenas laterales aromáticas o cadenas laterales positivamente cargadas) y, por lo tanto, no cambia sustancialmente las propiedades funcionales de la molécula polipeptídica. Lo que sigue expone grupos de ejemplo que contienen aminoácidos naturales de propiedades químicas similares, donde las sustituciones dentro de un grupo son una "sustitución conservadora".

Tabla 1

Sustituciones aminoácidas conservadoras

Cadenas laterales no polares y/o alifáticas	Cadenas laterales polares no cargadas	Cadenas laterales aromáticas	Cadenas laterales positivamente cargadas	Cadenas laterales negativamente cargadas
Glicina	Serina	Fenilalanina	Lisina	Aspartato
Alanina	Treonina	Tirosina	Arginina	Glutamato
Valina	Cisteína	Triptófano	Histidina	
Leucina	Metionina			
Isoleucina	Asparagina			
Prolina	Glutamina			

Hibridación de ácidos nucleicos

Se puede usar hibridación comparativa para identificar los ácidos nucleicos divulgados en el presente documento, incluyendo variaciones conservadoras de los ácidos nucleicos divulgados. Además, los ácidos nucleicos diana que se hibridan a un ácido nucleico divulgado en el presente documento en condiciones de alta, ultraalta y ultraultraalta restricción, donde los ácidos nucleicos codifican mutantes correspondientes a los destacados en las tablas 2 y 3 u otras polimerasas enumeradas, son un rasgo característico de la divulgación. Los ejemplos de dichos ácidos nucleicos incluyen aquellos con una o unas pocas sustituciones de ácido nucleico sinónimas o conservadoras en comparación con una secuencia de ácido nucleico dada que codifica una polimerasa de la tabla 2 (u otra polimerasa ejemplificada), donde cualquier sustitución conservadora es para residuos distintos de los destacados en la tabla 2 o en otra parte como que es pertinente para un rasgo característico de interés (incorporaciones de análogos de nucleótidos mejoradas, etc.).

Se dice que un ácido nucleico de prueba se hibrida específicamente a un ácido nucleico de sonda cuando se hibrida al menos en un 50 % tanto a la sonda como a la diana complementaria perfectamente emparejada, es decir, con una proporción señal con respecto a ruido de al menos la mitad de alta que la hibridación de la sonda a la diana en condiciones en las que la sonda perfectamente emparejada se une a la diana complementaria perfectamente emparejada con una proporción señal con respecto a ruido que sea al menos aproximadamente 5x-10x tan alta como la observada para la hibridación a cualquiera de los ácidos nucleicos diana no emparejados.

Los ácidos nucleicos se "hibridan" cuando se asocian, típicamente en solución. Los ácidos nucleicos se hibridan debido a una variedad de fuerzas fisicoquímicas bien caracterizadas, tales como enlaces de hidrógeno, exclusión de disolvente, apilamiento de bases y similares. Se encuentra una guía extensa para la hibridación de ácidos nucleicos en Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes*, parte I, capítulo 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays" (Elsevier, N.Y.), así como en *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley y Sons, Inc., (complementados hasta 2011); Hames y Higgins (1995) *Gene Probes 1* IRL Press at Oxford University Press, Oxford, Inglaterra, (Hames y Higgins 1) y Hames y Higgins (1995) *Gene Probes 2* IRL Press at Oxford University Press, Oxford, Inglaterra (Hames y Higgins 2) proporcionan detalles sobre la síntesis, marcaje, detección y cuantificación de ADN y ARN, incluyendo oligonucleótidos.

Un ejemplo de condiciones de hibridación rigurosas para la hibridación de ácidos nucleicos complementarios que tienen más de 100 residuos complementarios en un filtro en una transferencia de Southern o Northern es formol al 50 % con 1 mg de heparina a 42 °C, llevándose a cabo la hibridación durante la noche. Un ejemplo de condiciones de lavado rigurosas es un lavado con 0,2x SSC a 65 °C durante 15 minutos (véase Sambrook, *supra* para una descripción del tampón SSC). A menudo, el lavado de alta restricción está precedido por un lavado de baja restricción para retirar la señal de sonda de fondo. Un lavado de baja restricción de ejemplo es 2x SSC a 40 °C durante 15 minutos. En general, una proporción señal con respecto a ruido de 5x (o mayor) que la observada para una sonda no relacionada en el ensayo de hibridación particular indica la detección de una hibridación específica.

Las "condiciones de hibridación-lavado rigurosas" en el contexto de experimentos de hibridación de ácidos nucleicos, tales como hibridaciones de Southern y Northern, son dependientes de la secuencia y son diferentes bajo diferentes parámetros ambientales. Una guía extensa para la hibridación de ácidos nucleicos se encuentra en Tijssen (1993), *supra*, y en Hames y Higgins, 1 y 2. Las condiciones de hibridación y lavado rigurosas fácilmente se pueden determinar empíricamente para cualquier ácido nucleico de prueba. Por ejemplo, al determinar las condiciones de hibridación y lavado rigurosas, las condiciones de hibridación y lavado se incrementan gradualmente (por ejemplo, incrementado la temperatura, disminuyendo la concentración de sal, incrementado la concentración de detergente y/o incrementado la concentración de disolventes orgánicos, tales como formol, en la hibridación o lavado), hasta que se cumpla un conjunto seleccionado de criterios. Por ejemplo, en condiciones de

hibridación y lavado altamente rigurosas, las condiciones de hibridación y lavado se incrementan gradualmente hasta que una sonda se une a una diana complementaria perfectamente emparejada con una proporción señal con respecto a ruido que sea al menos 5x tan alta como la observada para la hibridación de la sonda a una diana no emparejada.

Se seleccionan condiciones "muy rigurosas" para que sean iguales al punto de fusión térmica (T_f) para una sonda particular. La T_f es la temperatura (bajo fuerza iónica y pH definidos) a la que un 50 % de la secuencia de prueba se hibrida a una sonda perfectamente emparejada. En general, las condiciones de hibridación y lavado "altamente rigurosas" se seleccionan para que estén aproximadamente 5° C por debajo de la T_f para la secuencia específica a una fuerza iónica y pH definidos.

Las condiciones de hibridación y lavado de "ultraalta restricción" son aquellas en las que la restricción de las condiciones de hibridación y lavado se incrementa hasta que la proporción señal con respecto a ruido para la unión de la sonda al ácido nucleico diana complementario perfectamente emparejado sea al menos 10x tan alta como la observada para la hibridación a cualquiera de los ácidos nucleicos diana no emparejados. Se dice que un ácido nucleico diana que se hibrida a una sonda en dichas condiciones, con una proporción señal con respecto a ruido de al menos 1/2 de la del ácido nucleico diana complementario perfectamente emparejado se une a la sonda en condiciones de ultraalta restricción.

De forma similar, se pueden determinar niveles de restricción incluso mayores incrementando gradualmente las condiciones de hibridación y/o lavado del ensayo de hibridación pertinente. Por ejemplo, aquellos en los que la restricción de las condiciones de hibridación y lavado se incrementan hasta que la proporción señal con respecto a ruido para la unión de la sonda al ácido nucleico diana complementario perfectamente emparejado sea al menos 10x, 20x, 50x, 100x o 500x o más tan alta como la observada para la hibridación con cualquiera de los ácidos nucleicos diana no emparejados. Se dice que un ácido nucleico diana que se hibrida a una sonda en dichas condiciones, con una proporción señal con respecto a ruido de al menos 1/2 de la del ácido nucleico diana complementario perfectamente emparejado se une a la sonda en condiciones de ultraalta restricción.

Los ácidos nucleicos que no se hibridan entre sí en condiciones rigurosas todavía son sustancialmente idénticos si los polipéptidos que codifican son sustancialmente idénticos. Esto se produce, por ejemplo, cuando se crea una copia de un ácido nucleico usando la máxima redundancia de codones permitida por el código genético.

Comparación, identidad y homología de secuencias

Los términos "idéntico" o "identidad en porcentaje", en el contexto de dos o más secuencias de ácido nucleico o polipeptídicas, se refieren a dos o más secuencias o subsecuencias que son las mismas o tienen un porcentaje especificado de residuos de aminoácido o nucleótidos que son los mismos, al compararse y alinearse en cuanto a una correspondencia máxima, como se mide usando uno de los algoritmos de comparación de secuencias descritos a continuación (u otros algoritmos disponibles para expertos en la técnica) o por inspección visual.

La frase "sustancialmente idénticas", en el contexto de dos ácidos nucleicos o polipéptidos (por ejemplo, ADN que codifican una polimerasa, o la secuencia de aminoácidos de una polimerasa) se refiere a dos o más secuencias o subsecuencias que tienen identidad de residuos de aminoácido o nucleótidos en al menos aproximadamente un 60 %, aproximadamente un 80 %, aproximadamente un 90-95 %, aproximadamente un 98 %, aproximadamente un 99 % o más al compararse y alinearse en cuanto a una correspondencia máxima, como se mide usando un algoritmo de comparación de secuencias o por inspección visual. Dichas secuencias "sustancialmente idénticas" típicamente se consideran "homólogas", sin referencia a la ascendencia real. Preferentemente, la "identidad sustancial" existe en una región de las secuencias que tiene al menos aproximadamente 50 residuos de longitud, más preferentemente en una región de al menos aproximadamente 100 residuos, y lo más preferentemente, las secuencias son sustancialmente idénticas en al menos aproximadamente 150 residuos, o en la longitud completa de las dos secuencias que se van a comparar.

Las proteínas y/o secuencias proteicas son "homólogas" cuando se derivan, natural o artificialmente, de una proteína o secuencia proteica ancestral común. De forma similar, los ácidos nucleicos y/o las secuencias de ácido nucleico son homólogos cuando se derivan, natural o artificialmente, de un ácido nucleico o secuencia de ácido nucleico ancestral común. La homología, en general, se infiere a partir de la similitud de secuencias entre dos o más ácidos nucleicos o proteínas (o secuencias de los mismos). El porcentaje preciso de similitud entre secuencias que es útil para establecer la homología varía con el ácido nucleico y la proteína en cuestión, pero para establecer la homología se usa de forma rutinaria tan solo una similitud de secuencias de un 25 % en 50, 100, 150 o más residuos. También se pueden usar mayores niveles de similitud de secuencias, por ejemplo, identidad en un 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 % o 99 % o más, para establecer la homología. Los procedimientos para determinar los porcentajes de similitud de secuencias (por ejemplo, BLASTP y BLASTN usando parámetros por defecto) se describen en el presente documento y, en general, están disponibles.

Para la comparación de secuencias y la determinación de homología, típicamente una secuencia actúa como secuencia de referencia con la que se comparan las secuencias de prueba. Cuando se usa un algoritmo de

comparación de secuencias, las secuencias de prueba y de referencia se introducen en un ordenador, se designan las coordenadas de subsecuencia, si es necesario, y se designan los parámetros del programa de algoritmo de secuencias. A continuación, el algoritmo de comparación de secuencias calcula la identidad de secuencias en porcentaje para la(s) secuencia(s) de prueba en relación con la secuencia de referencia, en base a los parámetros del programa designados.

Se puede llevar a cabo la alineación óptima de secuencias para su comparación, por ejemplo, por el algoritmo de homología local de Smith y Waterman, Adv. Appl. Math. 2:482 (1981), por el algoritmo de alineación de homología de Needleman y Wunsch, J. Mol. Biol. 48:443 (1970), por la búsqueda del procedimiento de similitud de Pearson y Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), por implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis) o por inspección visual (véanse, en general, Current Protocols in Molecular Biology, Ausubel *et al.*, eds., Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., complementados hasta 2011).

Un ejemplo de un algoritmo que es adecuado para determinar la identidad de secuencias y la similitud de secuencias en porcentaje es el algoritmo BLAST, que se describe en Altschul *et al.*, J. Mol. Biol. 215:403-410 (1990). El programa informático para realizar los análisis por BLAST está disponible públicamente a través del National Center for Biotechnology Information. Este algoritmo implica en primer lugar identificar pares de secuencias de puntuación alta (HSP) identificando palabras cortas de longitud W en la secuencia de consulta, que coincidan o bien satisfagan alguna puntuación umbral de valor positivo T cuando están alineadas con una palabra de la misma longitud en una secuencia de base de datos. T se refiere al umbral de puntuación de palabras vecinas (Altschul *et al.*, *supra*). Estos resultados positivos de palabras vecinas iniciales actúan como semillas para iniciar búsquedas para encontrar HSP más largas que los contengan. A continuación, los resultados positivos de palabras se extienden en ambas direcciones a lo largo de cada secuencia hasta donde se pueda incrementar la puntuación de alineación acumulada. Las puntuaciones acumuladas se calculan usando, para secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de residuos emparejados, siempre >0) y N (puntuación de penalización para los residuos de emparejamiento erróneo, siempre <0). Para secuencias de aminoácidos, se usa una matriz de puntuación para calcular la puntuación acumulada. La extensión de los resultados positivos de palabras en cada dirección se detiene cuando: la puntuación de alineación acumulada cae en la cantidad X desde su valor máximo logrado; la puntuación acumulada pasa a ser cero o inferior debido a la acumulación de una o más alineaciones de residuos de puntuación negativa; o se alcanza el final de cualquiera de las secuencias. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y velocidad de la alineación. El programa BLASTN (para secuencias de nucleótidos) usa por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, un valor de corte de 100, M=5, N=-4 y una comparación de ambas hebras. Para secuencias de aminoácidos, el programa BLASTP usa por defecto una longitud de palabra (W) de 3, una expectativa (E) de 10 y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

Además de calcular la identidad de secuencias en porcentaje, el algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul (1993) Proc. Nat'l. Acad. Sci. USA 90:5873-5787). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña (P(N)), que proporciona una indicación de la probabilidad de que se produzca por casualidad un emparejamiento entre dos secuencias de aminoácidos o nucleótidos. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de suma más pequeña en una comparación del ácido nucleico de prueba con el ácido nucleico de referencia es menor de aproximadamente 0,1, más preferentemente menor de aproximadamente 0,01 y lo más preferentemente menor de aproximadamente 0,001.

Como referencia, la secuencia de aminoácidos de una polimerasa DPO4 natural se presenta en la tabla 2.

Combinaciones de mutaciones ejemplares

En la tabla 2 se proporciona una lista de combinaciones de mutaciones de polimerasas ejemplares y las secuencias de aminoácidos de polimerasas DPO4 recombinantes que albergan las combinaciones de mutaciones ejemplares. Las posiciones de las sustituciones aminoacídicas se identifican en relación con una ADN polimerasa DPO4 natural (SEQ ID NO: 1). Las polimerasas de la divulgación (incluyendo las proporcionadas en la tabla 2) pueden incluir cualquier rasgo característico exógeno o heterólogo (o combinación de dichos rasgos característicos) en la región N y/o C terminal. Por ejemplo, se entenderá que los mutantes de polimerasa en la tabla 2 que no incluyen, por ejemplo, una marca de polihistidina C terminal se pueden modificar para que incluyan una marca de polihistidina en la región C terminal, en solitario o en combinación con cualquiera de los rasgos característicos exógenos o heterólogos descritos en el presente documento. Las variantes expuestas en el presente documento incluyen una delección de los últimos 12 aminoácidos de la proteína (es decir, los aminoácidos 341-352) para, por ejemplo, incrementar la solubilidad de proteínas en sistemas de expresión bacterianos.

Tabla 2

Variantes de DPO4 identificadas a través de diseño semirracional

SEQ ID NO	Secuencia de aminoácidos
1 ADN polimerasa DPO4 wt	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGAVA TANYEARKFGVKAGIPIVEAKKILPNAVYLPMPRKEVYQQVSSRI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAKIAADMAKPNGIKVIDDEEVKRLIR ELDIADVPGIGNITA EK LKKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRIVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVAVTEDLDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKFIEAIGLDKFFDT
2 CO1065 A42V_K56Y_M76W_K78N_E79L_Q82W_ Q83G_S86E_K152E_I153X1_A155G_D1 56R_P184L_G187P_N188Y_I189W_T190 Y_I248T_V289W_T290K_E291S_D292Y_ L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVF AEMAGRM A K P N G I K V I D D E E V K R L I R ELDIADVLGIPYWYAEK L K L G I N K L V D T L S I E F D K L G M I G E A KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKS Y W D I V S R G R T F P H G I S K ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
3 CO1066 A42V_K56Y_M76W_K78N_E79L_Q82W_ Q83G_S86E_K152G_I153T_A155G_D15 6L_P184L_G187P_N188Y_I189W_T190Y _I248T_V289W_T290K_E291S_D292Y_L 293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEK L K L G I N K L V D T L S I E F D K L G M I G E A KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKS Y W D I V S R G R T F P H G I S K ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
4 CO1067 A42V_K56Y_M76W_K78N_E79L_Q82W_ Q83G_S86E_K152L_I153T_A155G_D156 R_P184L_G187P_N188Y_I189W_T190Y _I248T_V289W_T290K_E291S_D292Y_L 293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEK L K L G I N K L V D T L S I E F D K L G M I G E A KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKS Y W D I V S R G R T F P H G I S K ETAYSESVKLLQKILEEDERKIRRIGVRFSKF

5 C1222 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_R87Q_K152G_A155_M_D156W_P184L_I189W_K196S_I248T_R253N_R256T_K278N_R279Q_V289W_T290R_E291S_D292Y_L293W_I295Y_S297H_G299L_T301RΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSQI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKSLGINKLVDTLSIEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKNNSTNLBEEKPY LFRAIEESYYKLDNQIPKAIHVVAWRSYWDYVHRLRRFPFGISK ETAYSESVKLLQKILEEDERKIRRIGVRFBSKF
6 C1224 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_R116N_K152G_A155M_D156W_R173N_R176N_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_I295Y_S297H_G299L_T301R_K317Q_Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVNDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKNLIN ELDIADVLGIGNWTAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEEKPY LFRAIEESYYKLDKRIKAIHVVAWRSYWDYVHRLRRFPFGISK ETAYSESVQLLQKILEEDERKIRRIGVRFBSKF
7 C1226 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_I295Y_S297H_G299L_T301R_E324K_E327KΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEEKPY LFRAIEESYYKLDKRIKAIHVVAWRSYWDYVHRLRRFPFGISK ETAYSESVKLLQKILKEDKRKIRRIGVRFBSKF
8 C1337 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_R116N_K152G_A155M_D156W_R173N_R176N_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_I295Y_S297H_G299L_T301R_E324K_E327K_Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVNDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKNLIN ELDIADVLGIGNWTAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEEKPY LFRAIEESYYKLDKRIKAIHVVAWRSYWDYVHRLRRFPFGISK ETAYSESVKLLQKILKEDKRKIRRIGVRFBSKF
9 C1363 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_K164T_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_I295Y_S297H_G299L_T301RΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGITVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEEKPY LFRAIEESYYKLDKRIKAIHVVAWRSYWDYVHRLRRFPFGISK ETAYSESVKLLQKILEEDERKIRRIGVRFBSKF

10 C1368 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_E192R_K196L_R242W_I248T_K282G_V289W_T290R_E291S_D292Y_L293W_I295Y_S297H_G299L_T301RΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTARKLKLKGINKLVDTLSEFDKLGKMIGEAL KAKYLISLARDEYNEPIRTRVWKSIGRTVTMKNRSRNLLEEIKPY LFRAIEESYYKLDKRIPGAIHVVAWRSYWDYVHRLRRFPFGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
11 C1417 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_R116N_K152G_A155M_D156W_R173N_R176N_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_I295Y_S297H_G299L_T301R_K317Q_K321Q_E324K_E325K_E327K_Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVNDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKNLIN ELDIADVLGIGNWTAEKLLKLKGINKLVDTLSEFDKLGKMIGEAL KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWDYVHRLRRFPFGISK ETAYSESVQLLQQLKKDKRKIRRIGVRFSKF
12 C1418 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_R116N_K152G_A155M_D156W_R173N_R176N_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_I295Y_S297H_G299L_T301R_K317Q_K321Q_E324N_E325Q_E327Q_Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVNDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKNLIN ELDIADVLGIGNWTAEKLLKLKGINKLVDTLSEFDKLGKMIGEAL KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWDYVHRLRRFPFGISK ETAYSESVQLLQQLNQQDKRKIRRIGVRFSKF
13 C1428 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_I189W_I248T_V289W_T290K_E291S_D292Y_L293W_E324K_E327K Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLKGINKLVDTLSEFDKLGKMIGEAL KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYSYWDIVSRGRTFPFGISK ETAYSESVKLLQKILKEDDKRKIRRIGVRFSKF
14 C1429 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_E324K_E327KΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLKGINKLVDTLSEFDKLGKMIGEAL KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYSYWDIVSRGRTFPFGISK ETAYSESVKLLQKILKEDDKRKIRRIGVRFSKF

15 C1452 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_E192R_K196L_R242W_I248T_K282G_V289W_T290R_E291S_D292Y_L293W_I295Y_S297H_G299L_T301R_E324R_E327RΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTARKLKLKGINKLVDTLSEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVWKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPGAIVHVAWRSYWDYVHRLRRFPFGISK ETAYSESVKLLQKILREDRRKIRRIGVRFSKF
16 C1454 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_I189W_I248T_V289W_T290K_E291S_D292Y_L293W_K317Q_K321Q_E324K_E325K_E327K Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRIVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYWDIVSRGRTFPHGISK ETAYSESVQLLQQLKKDKRKIRRIGVRFSKF
17 C1456 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_K317Q_K321Q_E324K_E325K_E327KΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYWDIVSRGRTFPHGISK ETAYSESVQLLQQLKKDKRKIRRIGVRFSKF
18 C1457 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_K317Q_K321Q_E324N_E325Q_E327Q Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYWDIVSRGRTFPHGISK ETAYSESVQLLQQLNQDQRKIRRIGVRFSKF
19 C1506 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_I189W_I248T_V289W_T290K_E291S_D292Y_L293W_E325K_D326K Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRIVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEKKERKIRRIGVRFSKF

20 C1395 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152Q_I153R_A155G_D156A_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293WΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFQAQRAGAMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEIKPY LFRAIEESYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
21 C1187 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_D294Q_I295M_V296A_S297T_R298L_G299F_T301HΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEIKPY LFRAIEESYKLDKRIPKAIHVVAWRSYWMATLFRHFFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
22 CO672 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_I295A_S297P_G299W_T301FΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEIKPY LFRAIEESYKLDKRIPKAIHVVAWRSYWDVAVPRWRFFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
23 CO677 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293_I295Y_S297W_G299W_T301VΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEIKPY LFRAIEESYKLDKRIPKAIHVVAWRSYWDYVWRWRVFFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
24 CO982 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293_I295C_S297W_G299I_T301WΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEIKPY LFRAIEESYKLDKRIPKAIHVVAWRSYWDVWRIRWFFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF

25 C1152 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_K282G_A283L_V289W_T290K_E291S_D292Y_L293W_F302E_P303R_H304GΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEIKPY LFRAIEESYYKLDKRIPGLIHVVAWKSZYWDIVSRGRTERGGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
26 C1159 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153H_A155G_D156R_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293WΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGHAGRMKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
27 C1185 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_D294A_I295P_V296R_S297L_R298K_G299A_R300C_T301EΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWPRLKACEFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
28 C1190 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_D294C_V296S_S297Q_R298T_T301LΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWCISQTGRLFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
29 C1191 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_D294S_I295T_V296R_R298T_G299H_R300S_T301YΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMTKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWSTRSTHSYFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF

30 C1234 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_D294N_I295S_V296Q_S297Y_G299W_R300S_T301W_Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEAE KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWSQYRWSWFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
31 C1237 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P1841L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_D294L_I295T_V296R_S297C_T301RΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEAE KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWLTRCRGRFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
32 C1238 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_D294A_I295W_V296R_S297R_R298E_G299S_R300S_T301MΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEAE KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYAWRRESSMFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
33 C1240 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152A_I153R_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293WΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEAE KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
34 C1278 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_D294P_I295P_V296R_G299A_R300C_T301KΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEAE KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWPSPRSRACKFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF

35 C1287 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152Y_I153S_A155_C_D156G_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293WΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAYSACGMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
36 C1288 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152V_I153Y_A155C_D156G_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293WΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAYYACGMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
37 C1289 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152M_I153Q_A155G_D156Y_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293WΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAMQAGYMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
38 C1292 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153K_A155G_D156S_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293WΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALKAGSMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
39 C1328 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_I295Y_S297F_G299Y_T301SΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSIEFDKLGKMIGEAA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSZYWDYVFRYRSFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF

40 C1384 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_D294H_I295W_V296Y_S297QΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEAE KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWHWYQGRGTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEKF
41 C1392 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152E_I153V_A155G_D156R_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_I295Y_S297H_G299L_T301RΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEAE KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWDYVHRLRRFPFHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEKF
42 C1404 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_D294S_I295F_V296A_S297T_R298G_T301QΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEAE KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWSFATGGRQFPFHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEKF
43 C1428 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_E324K_E327K Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGIMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEAE KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILKEDKRKIRRIGVRFSEKF
44 C1442 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_A155M_D156W_P184L_I189W_I248T_V289W_T290R_E291S_D292Y_L293W_D294P_I295P_S297Y_R298L_R300V_T301RΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEAE KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWPVYLGVRFPPFHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEKF

45 C1462 A42V_K56Y_M76W_R77S_K78N_E79L_Q82W_Q83G_S86E_E97S_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWSNLVYWGVSERI MNLLREYSSKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLI SLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
46 C1464 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_R298E_K321N_R328N Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLI SLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSEGRRTFPHGISK ETAYSESVKLLQNILEEDENKIRRIGVRF SKF
47 C1516 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_P184L_G187E_N188S_I189W_T190V_I248T_V289W_T290K_E291S_D292Y_L293W_R300I Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIESWVAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLI SLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGITFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
48 C1517 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152A_I153K_D156I_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNNKVFAAKAAIMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLI SLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
49 C1520 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153Q_A155G_D156M_P184L_G187E_N188S_I189W_T190V_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNNKVFALQAGMMAKPNGIKVIDDEEVKRLIR ELDIADVLGIESWVAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLI SLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

50 C1554 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y-I189W_T190Y_I248T_V289W_D211P_K214V_E219W_K223C_T290R_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFPKLVGMIGWA KACYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
51 C1555 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_D211S_K214S_E219G_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFSLSCMIGGA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
52 C1556 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_D211P_K214V_E219A_K223T_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFPKLVGMIGAA KATYLLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
53 C1557 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_D211P_K214A_E219A_K223R_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFPKLAGMIGAA KARYLLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF

54 C1558 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189Y_T190F_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYFYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
55 C1560 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_Q320K_K321R_I322A_E324A_E325R Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLKRALARDEKIRRIGVRF SKF
56 C1561 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_Q320R_K321S_I322L_E324G_E325G Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLRSLGGDERKIRRIGVRF SKF
57 C1562 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_K321R_I322C_E324S_E325G Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQRCLSGDERKIRRIGVRF SKF

58 C1569 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_K214S_G215R_E219A_K221R_K223R_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLSRMIGAA RARYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEK
59 C1570 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_K214V_G215S_E219R_K221R_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLVSMIGRA RAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEK
60 C1575 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_E170M_R173G_L174F_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKGFIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEK
61 C1577 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153Y_A155G_D15_6T_M157R_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293WΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGYAGTRAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEK

62 C1581 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_K212D_K214R_G215K_K221G_K223V_I248T_V289W_T290K_E291S_D292Y_L2_93W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDDLKRMIGEA GAVYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
63 C1584 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_K212F_K214S_G215T_K221R_K223R_I248T_V289W_T290K_E291S_D292Y_L2_93W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDFLSTMIGEA RARYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
64 C1589 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_K214S_G215E_E219G_K221G_K223R_I248T_V289W_T290K_E291S_D292Y_L2_93W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDFKLSEMIGGA GARYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
65 C1592 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_Q320M_K321G_I322V_E324S_E325V Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDFKLKGMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLMGVLSVDERKIRRIGVRFSKF

66 C1593 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_K321R_E324M_E325SΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYKLDKRIKPAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQIRILMSDERKIRRIGVRFSEK
67 C1594 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_Q320K_K321Q_I322V_E324W_E325G Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYKLDKRIKPAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQVWLWGDERKIRRIGVRFSEK
68 C1599 K24S_K26L_A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLSGLPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYKLDKRIKPAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEK
69 C1614 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_R253N_R256N_E259S_E260Q_K262N_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEAA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNNSNNLSQINPY LFRAIEESYKLDKRIKPAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEK

70 C1615 D7V_A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_R267Q_E270Q_D277G_K278N_R279Q_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLEYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFQAIQESYYKLGNIQIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
71 C1616 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_K282S_V289W_T290K_E291S_D292Y_L293W_R300S_K308Q_E309N_Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLEYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPSAIHVVAWKSZYWDIVSRGSTFPHGISQ NTAYSESVKLLQKILEEDERKIRRIGVRFSKF
72 C1617 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_E314Q_K317Q_E324K_E327K_R328SΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLEYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSQSVQLLQKILKEDKSKIRRIGVRFSKF
73 C1618 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_D277G_K278N_R279Q_V289W_T290K_E291S_D292Y_L293W_E309D_E324K_E327K Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLEYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLGNIQIPKAIHVVAWKSZYWDIVSRGRTFPHGISK DTAYSESVKLLQKILKEDERKIRRIGVRFSKF
74 C1621 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_K214R_G215R_E219G_K221R_K223T_I248T_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLEYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLRMIGGA RATYLLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF

75 C1622 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_K214A_G215E_E219V_K221G_K223S_I248T_V289W_T290K_E291S_D292Y_L2_93W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLAEMIGVA GASYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEF
76 C1623 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_K214Q_G215A_E219L_K221G_K223R_I248T_V289W_T290K_E291S_D292Y_L2_93W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLAEMIGLA GARYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEF
77 C1626 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L_293W_Q320I_K321Q_I322T_E324D_E32_5V Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLIQTLVDVDERKIRRIGVRFSEF
78 C1627 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L_293W_Q320T_K321R_I322T_E324A_E32_5G Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSZYWDIVSRGRTFPHGISK ETAYSESVKLLTRTLAGDERKIRRIGVRFSEF

79 C1628 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_Q320N_K321R_E324R_E325R Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKFPVVVCFVSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMGIEA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPDIVSRGRTFFPHGISK ETAYSESVKLLNRLRRDERKIRRIIGVRFSKF
80 C1630 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L29W_Q320N_K321S_I322L_E324R_E325S Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKFPVVVCFVSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMGIEA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPDIVSRGRTFFPHGISK ETAYSESVKLLNSLLRSDEKIRRIIGVRFSKF
81 C1666 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290R_E291S_D292Y_L293W_D294S_I295P_S297T_R300G_T3_01Q Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKFPVVVCFVSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMGIEA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPVTRGGQFPFHGISK ETAYSESVKLLQKILEEDERKIRRIIGVRFSKF
82 C1788 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_D211T_K214A_E219V_I248T_V289W_T290K_E291S_D292Y_L293W_Q320T_K321R_I322C_E324G_E325G Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKFPVVVCFVSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFTLAGMGIVA KAKYLIISLARDEYNEPIRTPVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPDIVSRGRTFFPHGISK ETAYSESVELLTRLGVDERKIRRIIGVRFSKF
83 C1789 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D15_6L_P184L_G187P_N188Y_I189W_T190Y_D211T_K214A_E219V_I248T_V289W_T290K_E291S_D292Y_L293W_Q320S_K321M_I322C_E324S_E325A Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKFPVVVCFVSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFTLAGMGIVA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPDIVSRGRTFFPHGISK ETAYSESVKLLSMCLSDERKIRRIIGVRFSKF

84 C1800 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_M157R_E170M_R173G_L174F_P184L_G187P_N188Y_I189W_T190Y_D211P_K214A_E219A_K223R_I248T_D277G_K278N_R279Q_V289W_T290K_E291S_D292Y_L293W_D294Q_I295M_V296A_S297T_R298L_G299F_T301H_E309D_K317Q_K321Q_E324K_E325K_E327K Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKEFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISEKNKVFAGTAGLRAKPNGIKVIDDEMVKGFIR ELDIADVLGIPYWYAEXLKKLGINKLVDTLSIEFPKLAGMIGAA KARYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLESIKPY LFRAIEESYKLGNIQIPKAIHVVAWKSYSQMATLFRHFFPHGISK ETAYSESQVLLQQLKKDKRKIRIRIGVRFSEKF
85 C1801 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_I295A_S297P_G299W_T301FΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKEFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISEKNKVFALTAGERMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEXLKKLGINKLVDTLSIEFDKLGKMIGBA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLESIKPY LFRAIEESYKLDKRIIPKAIHVVAWKSYSWDAPVRWRFFPHGISK ETAYSESQVLLQKILEEDERKIRIRIGVRFSEKF
86 C1802 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_I295A_S297P_G299W_T301F_K317Q_K321Q_E324K_E325K_E327KΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKEFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISEKNKVFALTAGERMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEXLKKLGINKLVDTLSIEFDKLGKMIGBA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMERNSENLESIKPY LFRAIEESYKLDKRIIPKAIHVVAWKSYSWDAPVRWRFFPHGISK ETAYSESQVLLQQLKKDKRKIRIRIGVRFSEKF
87 C1803 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_I295C_S297W_G299I_T301 W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKEFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISEKNKVFALTAGERMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEXLKKLGINKLVDTLSIEFDKLGKMIGBA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLESIKPY LFRAIEESYKLDKRIIPKAIHVVAWKSYSWDCVWRIRWFFPHGISK ETAYSESQVLLQKILEEDERKIRIRIGVRFSEKF

88 C1804 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_I295C_S297W_G299I_T301W_K317Q_K321Q_E324K_E325K_E327KΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPWDCVWRIRWFPHGISK ETAYSESVQLLQQLKKDKRKIRRIGVRFSEF
89 C1805 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_I295Y_S297W_G299W_T301VΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPWYVWRVWFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEF
90 C1807 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_I295Y_S297H_G299L_T301RΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPWYVHRLRRFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEF
91 C1808 A42V_K56Y_A57P_M76WK78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_I295Y_S297H_G299L_T301R_K317Q_K321Q_E324K_E325K_E327KΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPWYVHRLRRFPHGISK ETAYSESVQLLQQLKKDKRKIRRIGVRFSEF

92 C1809 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83GS86E_K152L_I153T_A155_G_D156R_P184L_G187P_N188Y_I189W_T190Y_E192R_K196L_R242W_I248T_K282G_V289W_T290K_E291S_D292Y_L293W_I295Y_S297H_G299L_T301R_E324K_E327K_Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYARKLKLLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVWKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPGAIHVVAWKSYSWDYVHRLRRFPHGISK ETAYSESVKLLQKILKEDKRKIRRIGVRFSKF
93 C1810 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155_G_D156R_P184L_G187P_N188Y_I189W_T190Y_E192R_K196L_R242W_I248T_K282G_V289W_T290K_E291S_D292Y_L293W_I295Y_S297H_G299L_T301R_E324K_E327K_Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYARKLKLLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVWKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPGAIHVVAWKSYSWQMATLFRHFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRFSKF
94 C1812 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155_G_D156R_P184L_G187P_N188Y_I189W_T190Y_E192R_K196L_R242W_I248T_K282G_V289W_T290K_E291S_D292Y_L293W_D294Q_I295M_V296A_S297T_R298L_G299F_T301H_K317Q_K321Q_E324K_E325K_E327K_Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYARKLKLLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVWKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPGAIHVVAWKSYSWQMATLFRHFPHGISK ETAYSESVKLLQKILKEDKRKIRRIGVRFSKF
95 C1816 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_D294N_I295S_V296Q_S297Y_G299W_R300S_T301W_K317Q_K321Q_E324K_E325K_E327K_Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYSWNSQYRWSWFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRFSKF

96 C1817 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152A_I153K_D156I_V183L_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_D294Q_I295M_V296A_S297T_R298L_G299F_T301HΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEBIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAAKAAIMAKPNGIKVIDDEEVKRLIR ELDIADLLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFBSKF
97 C1818 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152Q_I153R_A155G_D156A_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_D294Q_I295M_V296A_S297T_R298L_G299F_T301HΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEBIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAQRAGAMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFBSKF
98 C1819 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_D156G_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_D294Q_I295M_V296A_S297T_R298L_G299F_T301HΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEBIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAVYACGMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFBSKF
99 C1820 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_E170M_R173G_L174F_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_D294Q_I295M_V296A_S297T_R298L_G299F_T301HΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEBIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFALTAGMAKPNGIKVIDDEMVKGFIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYPQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFBSKF

100 C1821 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_P184L_G187P_N188Y_I189W_T190Y_D211P_K214V_E219A_K223T_I248T_V289W_T290K_E291S_D292Y_L293W_D294Q_I295M_V296A_S297T_R298L_G299F_T301H Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFFPKLVGMIGAA KATYLIISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSQYQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
101 C1823 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152A_I153K_D156I_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_K317Q_K321Q_E324K_E325K_E327K Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKFVFAAKAAIMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSQYWDIVSRGRTFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRF SKF
102 C1824 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152Q_I153R_A155G_D156A_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_K317Q_K321Q_E324K_E325K_E327K Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKFVFAQRAGAMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSQYWDIVSRGRTFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRF SKF
103 C1826 A42V_K56Y_A57P_M76W_K78N_E79L_Q82W_Q83G_S86E_K152Q_I153R_A155G_D156A_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_I295Y_S297H_G299L_T301R_K317Q_K321Q_E324K_E325K_E327K Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKFVFAQRAGAMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSQYWDYVHRLRRFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRF SKF

104 C1827 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152Q_1153R_A155G_D156A_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_T290K_E291S_D292Y_L293W_D294Q_I295M_V296A_S297T_R298L_G299F_T301H_K317Q_K321Q_E324K_E325K_E327K Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLEYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFQAQRAGAMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYSYQMATLFRHFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRFSKF
105 C1828 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_P184L_G187P_N188Y_I189W_T190Y_I248T_V289W_D211P_K214A_E219A_K223R_T290K_E291S_D292Y_L293W_K317Q_K321Q_E324K_E325K_E327K Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLEYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFPKLAGMIGAA KARYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYSYWDIVSRGRTFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRFSKF
106 C1895 A42V_K56Y_M76W_K78N_E79I_Q82W_Q83G_S86E_K152L_I153T_A155G_D156R_E170M_R173G_L174F_P184L_G187P_N188Y_I189W_T190Y_I248T_L258P_V289W_T290K_E291S_D292Y_L293W_D294Q_I295M_V296A_S297T_R298L_G299F_T301H Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLEYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFALTAGRMAKPNGIKVIDDEMVKGFIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNPEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYSYQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF

Los ejemplos y variantes de polimerasa proporcionados a continuación ilustran y ejemplifican además las composiciones de la presente divulgación y los procedimientos de preparación y uso de dichas composiciones.

5 Ejemplos

Ejemplo 1

10 Identificación de DPO4 como ADN polimerasa de translesión candidata para la incorporación de análogos de nucleótidos voluminosos durante la síntesis de ADN mediada por molde

15 Para identificar una ADN polimerasa con la capacidad de sintetizar hebras hijas usando sustratos "voluminosos" (es decir, que se puedan unir e incorporar análogos de nucleótidos fuertemente sustituidos a una hebra de ácido nucleico en crecimiento), se llevó a cabo un cribado de varias polimerasas disponibles comercialmente. Se evaluaron las polimerasas candidatas para determinar la capacidad de extender un cebador unido a oligonucleótido usando un conjunto de análogos de dNTP sustituidos con conectores de alquino tanto en el α-fosfato de cadena principal como en los restos de nucleobase (sustratos voluminosos modelo, denominados en el presente documento "dNTP-2c"). Las polimerasas cribadas para determinar su actividad incluyeron las siguientes: Ventra (Exo-), Deep Ventra® (Exo-), Therminator, Therminator II, Therminator III, Therminator Y, 9°Nm, PWO, PWO

SuperYield, PyroPhage 3173 (Exo-), Bst, fragmento grande, Exo- Pfu, Platinum Genotype TSP, Hemo Klen Taq, Taq, MasterAMP Taq, Phi29, Bsu, fragmento grande, Exo-Minus Klenow (D355A, E357A), Sequenase, versión 2.0, Transcriptor, Maxima, Thermoscript, M-MuLV (RNasa H-), AMV, M-MuLV, Monsterscript y DPO4. De las polimerasas sometidas a prueba, DPO4 (expresada naturalmente por la arqueas, *Sulfolobus solfataricus*) fue la que más pudo extender eficazmente un cebador unido a molde con análogos de nucleótidos dNTP-2c. Sin ceñirse a la teoría, se especuló que DPO4, y posiblemente otros miembros de la familia de ADN polimerasas de translesión (es decir, ADN polimerasas de clase Y), podrían utilizar eficazmente análogos de nucleótidos voluminosos debido a sus sitios de unión a sustrato relativamente grandes, que se han desarrollado para acomodar lesiones de ADN voluminosas naturales.

Ejemplo 2

Identificación de "puntos calientes" para la mutagénesis dirigida en la proteína DPO4 y cribado de colecciones de mutantes de DPO4 para identificar motivos de secuencia optimizados

Como primera etapa en la generación de variantes de DPO4 con actividad polimerasa mejorada al exponerse a sustratos voluminosos, se usó la herramienta web "HotSpot Wizard" para identificar aminoácidos en la proteína DPO4 para seleccionarse como diana para la mutagénesis. Esta herramienta implementa un protocolo de genomanipulación de proteínas que selecciona como diana posiciones de aminoácido evolutivamente variables localizadas, por ejemplo, en el sitio activo de enzima. Los "puntos calientes" para la mutación se seleccionan a través de la integración de información estructural, funcional y evolutiva (véase, por ejemplo, Pavelka *et al.*, "HotSpot Wizard: a Web Server for Identification of Hot Spots in Protein Engineering" (2009) *Nuc Acids Res* 37 doi:10.1093/nar/gkp410). Al aplicar esta herramienta a la proteína DPO4, se observó que los residuos de puntos calientes identificados tendían a agruparse en determinadas zonas o regiones, repartidas por la secuencia de aminoácidos completa. Se establecieron límites arbitrarios para distinguir 13 de dichas regiones, designadas "Mut1" - "Mut13", en las que se concentran los puntos calientes de mutagénesis. Estas 13 regiones "Mut" se ilustran en la FIG. 1 con residuos de puntos calientes identificados por subrayado.

Para cribar variantes de DPO4 con actividad polimerasa mejorada en base a la cartografía de puntos calientes, se creó una colección de mutagénesis por saturación para cada una de las 13 regiones Mut, en la que se cambiaron los aminoácidos de puntos calientes, mientras que los aminoácidos conservados se dejaron inalterados. El cribado se llevó a cabo usando una plataforma de placa de 96 pocillos y la actividad polimerasa se evaluó con un ensayo de extensión del cebador usando análogos de nucleótidos "dNTP-OAc" como sustratos. Estos sustratos voluminosos modelo se sustituyen con restos acetato de triazol conjugados con sustituyentes de alquino tanto en el α -fosfato como en los restos de nucleobase. Los resultados del cribado identificaron dos regiones Mut en particular que producían consecuentemente mutantes de DPO4 con actividad potenciada. Estas regiones, "Mut_4" y "Mut_11", corresponden a los aminoácidos 76-86 y los aminoácidos 289-304, respectivamente, de la proteína DPO4. Otro análisis de las variantes Mut_4 y Mut_11 de alto rendimiento da lugar a la identificación de una secuencia de motivo variante optimizado para cada región. El motivo Mut_4 optimizado identificado en el presente documento es como sigue: M76W, K78N, E79L, Q82W, Q83G y S86E, mientras que el de la región Mut_11 es como sigue: V289W, T290K, E291S, D292Y y L293W.

Ejemplo 3

Enfoques semirracionales para diseñar variantes de DPO4 con actividad de polimerización potenciada

Para continuar desarrollando variantes de DPO4 con una utilización mejorada de sustratos voluminosos, se adoptó un enfoque de diseño "semirracional" siguiendo una serie de estrategias diferentes. En una estrategia, se creó una colección que dirigió la mutagénesis aleatoria en las posiciones 152, 153, 155 y 156 (correspondientes a la región Mut_6) en la variante de alto rendimiento identificada previamente, CO534 (divulgada en la solicitud de patente PCT pendiente de tramitación de los solicitantes n.º PCT/US2016/061661, divulgada en el presente documento en su totalidad por referencia). El cribado de esta colección se llevó a cabo como se describe anteriormente y los resultados identificaron tres variantes en particular que demostraron actividad de extensión del cebador superior usando sustratos análogos de nucleótidos voluminosos. Estas variantes se exponen en la tabla 2 como C1065 (SEQ ID NO: 2), C1066 (SEQ ID NO: 3) y CO1067 (SEQ ID NO: 4).

En otra estrategia, se seleccionaron como diana en cuanto a la mutación los aminoácidos cargados expuestos en la superficie de la proteína DPO4 para reducir, eliminar o invertir la carga. Se planteó la hipótesis de que las alteraciones en el perfil de carga superficial global de la proteína podrían potenciar la interacción entre la polimerasa y los sustratos análogos de nucleótidos que portan grandes sustituyentes positivamente cargados, tales como los sustratos XNTP divulgados en el presente documento. Este enfoque identificó varios mutantes que demostraban una actividad potencialmente potenciada en ensayos de extensión del cebador usando sustratos análogos de nucleótidos voluminosos. Estas variantes se exponen como SEQ ID NO: 5-21 en la tabla 2.

En aún otras estrategias, las variantes de alto rendimiento, incluyendo CO416 y CO534 (divulgadas en la solicitud PCT pendiente de tramitación de los solicitantes n.º PCT/US2016/061661) y C1066 (SEQ ID NO: 3), CO1067 (SEQ

ID NO: 4), C1454 (SEQ ID NO: 16) y C1187 (SEQ ID NO: 21) se sometieron a mutagénesis aleatoria o dirigida como se divulga en el presente documento para crear colecciones para su cribado adicional. El cribado de estas colecciones se llevó a cabo como se describe anteriormente y los resultados identificaron varias variantes que funcionaron al menos tan bien como las variantes originales en ensayos de extensión del cebador usando sustratos

5 XNTP voluminosos. Estas variantes se exponen en la tabla 2 como SEQ ID NO: 22-106.

REIVINDICACIONES

1. Una ADN polimerasa recombinante aislada, comprendiendo dicha ADN polimerasa recombinante una secuencia de aminoácidos que es al menos un 90 % idéntica a los aminoácidos 1-340 de SEQ ID NO: 1, comprendiendo dicha ADN polimerasa recombinante mutaciones en las posiciones de aminoácido 42, 56, 76, 78, 79, 82, 83, 86, 184, 189, 248, 289, 290, 291, 292 y 293, y al menos una posición seleccionada del grupo que consiste en 7, 24, 26, 77, 87, 97, 116, 157, 164, 170, 173, 174, 176, 183, 191, 192, 196, 211, 212, 219, 221, 223, 242, 253, 256, 258, 259, 260, 262, 267, 270, 277, 278, 279, 282, 283, 294, 296, 302, 303, 304, 308, 309, 314, 317, 320, 321, 322, 324, 325, 327 y 328, en la que la identificación de posiciones es en relación con la polimerasa DPO4 natural (SEQ ID NO: 1), en la que los aminoácidos 341-352 en relación con la polimerasa DPO4 natural (SEQ ID NO: 1) están deletados; y presentando dicha ADN polimerasa recombinante actividad polimerasa.
2. La ADN polimerasa recombinante de la reivindicación 1, en la que la mutación en la posición 42 es A42V, la mutación en la posición 56 es K56Y, la mutación en la posición 76 es M76W, la mutación en la posición 78 es K78N, la mutación en la posición 79 es E79L, la mutación en la posición 82 es Q82W, la mutación en la posición 83 es Q83G, la mutación en la posición 86 es S86E, la mutación en la posición 184 es P184L, la mutación en la posición 189 es I189W, la mutación en la posición 248 es I248T, la mutación en la posición 289 es V289W, la mutación en la posición 290 es T290K o T290R, la mutación en la posición 291 es E291S, la mutación en la posición 292 es D292Y y la mutación en la posición 293 es L293W.
3. La ADN polimerasa recombinante de la reivindicación 1, en la que la mutación se selecciona del grupo que consiste en:
 - (a) la mutación en la posición 7 es D7V;
 - (b) la mutación en la posición 24 es K24S;
 - (c) la mutación en la posición 26 es K26L;
 - (d) la mutación en la posición 57 es A57P;
 - (e) la mutación en la posición 77 es R77S;
 - (f) la mutación en la posición 87 es R87Q;
 - (g) la mutación en la posición 97 es E97S;
 - (h) la mutación en la posición 116 es R116N;
 - (i) la mutación en la posición 152 se selecciona del grupo que consiste en K152E, K152G, K152L, K152Q, K152A, K152Y, K152V, K152M y K152L;
 - (j) la mutación en la posición 153 se selecciona del grupo que consiste en I153M, I153T, I153R, I153H, I153S, I153Y, I153Q, I153K e I153V;
 - (k) la mutación en la posición 155 se selecciona del grupo que consiste en A155G, A155M y A155C;
 - (l) la mutación en la posición 156 se selecciona del grupo que consiste en D156R, D156L, D156W, D156A, D156G, D156Y, D156S, D156I, D156M y D156T;
 - (m) la mutación en la posición 157 es M157R;
 - (n) la mutación en la posición 164 es K164T;
 - (o) la mutación en la posición 170 es E170M;
 - (p) la mutación en la posición 173 es R173N o R173G;
 - (q) la mutación en la posición 176 es R176N;
 - (r) la mutación en la posición 183 es V183L;
 - (s) la mutación en la posición 187 es G187P o G183E;
 - (t) la mutación en la posición 188 es N188Y o N188S;

- (u) la mutación en la posición 190 se selecciona del grupo que consiste en T190Y, T190V y T190F;
- (v) la mutación en la posición 191 es A191E;
- 5 (w) la mutación en la posición 192 es E192R;
- (x) la mutación en la posición 196 es K196S o K196L;
- (y) la mutación en la posición 211 se selecciona del grupo que consiste en D211P, D211S y D211T,
- 10 (z) la mutación en la posición 212 es K212D o K212F;
- (aa) la mutación en la posición 214 se selecciona del grupo que consiste en K214V, K214S, K214A, K214R y K214Q;
- 15 (bb) la mutación en la posición 215 se selecciona del grupo que consiste en G215R, G215S, G215K, G215T, G215E y G215A;
- (cc) la mutación en la posición 219 se selecciona del grupo que consiste en E219W, E219G, E219A, E219R, E219V y E219L;
- 20 (dd) la mutación en la posición 221 es K221R o K221G;
- (ee) la mutación en la posición 223 se selecciona del grupo que consiste en K223C, K223T, K223R, K223V y K223S;
- 25 (ff) la mutación en la posición 242 es R242W;
- (gg) la mutación en la posición 253 es R253N;
- 30 (hh) la mutación en la posición 256 es R256T o R256N;
- (ii) la mutación en la posición 258 es L258P;
- 35 (jj) la mutación en la posición 259 es R259S;
- (kk) la mutación en la posición 260 es E260Q;
- (ll) la mutación en la posición 262 es K262N;
- 40 (mm) la mutación en la posición 267 es R267Q;
- (nn) la mutación en la posición 270 es E270Q;
- 45 (oo) la mutación en la posición 277 es D277G;
- (pp) la mutación en la posición 278 es K278N;
- (qq) la mutación en la posición 279 es R279Q;
- 50 (rr) la mutación en la posición 282 es K282G o K282S;
- (ss) la mutación en la posición 294 se selecciona del grupo que consiste en D294Q, D294A, D294C, D294S, D294N, D294L, D294P y D294H;
- 55 (tt) la mutación en la posición 295 se selecciona del grupo que consiste en I295Y, I295C, I295P, I295T, I295S, I295W, I295Y, I295F, I295M e I295A;
- (uu) la mutación en la posición 296 se selecciona del grupo que consiste en V296A, V296R, V296S, V296Q y V296Y;
- 60 (vv) la mutación en la posición 297 se selecciona del grupo que consiste en S297H, S297T, S297P, S297W, S297L, S297Q, S297Y, S297C, S297R y S297F;
- 65 (ww) la mutación en la posición 298 se selecciona del grupo que consiste en R298L, R298K, R298T, R298E y R298G;

(xx) la mutación en la posición 299 se selecciona del grupo que consiste en G299L, G299F, G299W, G299I, G299A, G299H, G299S y G299Y;

5 (yy) la mutación en la posición 300 se selecciona del grupo que consiste en R300C, R300S, R300V y R300G;

(zz) la mutación en la posición 301 se selecciona del grupo que consiste en T301R, T301H, T301F, T301V, T301W, T301E, T301Y, T301M, T301K, T301S y T301Q;

10 (aaa) la mutación en la posición 302 es F302E;

(bbb) la mutación en la posición 303 es P303R;

(ccc) la mutación en la posición 304 es H304G;

15 (ddd) la mutación en la posición 308 es K308Q;

(eee) la mutación en la posición 309 se selecciona del grupo que consiste en E309N, E309P y E309D;

20 (fff) la mutación en la posición 314 es E314Q;

(ggg) la mutación en la posición 317 es K317Q;

25 (hhh) la mutación en la posición 320 se selecciona del grupo que consiste en Q320K, Q320R, Q320M, Q320I, Q320T, Q320N y Q320S;

(iii) la mutación en la posición 321 se selecciona del grupo que consiste en K321Q, K321N, K321R, K321S, K321G y K321M;

30 (jjj) la mutación en la posición 322 se selecciona del grupo que consiste en I322A, I322L, I322C, I322V e I322T;

(kkk) la mutación en la posición 324 se selecciona del grupo que consiste en E324K, E324N, E324R, E324A, E324G, E324S, E324R, E324W, E324D y E324K;

35 (lll) la mutación en la posición 325 se selecciona del grupo que consiste en E325K, E325Q, E325R, E325G, E325V, E325S y E325A;

(mmm) la mutación en la posición 326 es D326K;

40 (nnn) la mutación en la posición 327 se selecciona del grupo que consiste en E327K, E327Q y E327R; y

(ooo) la mutación en la posición 328 es R328N o R328S.

45 4. La ADN polimerasa recombinante de la reivindicación 1, que comprende la secuencia de aminoácidos como se expone en una cualquiera de SEQ ID NO: 2-106.

5. Una composición que comprende una ADN polimerasa recombinante como se expone en una cualquiera de las reivindicaciones 1-4.

50 6. La composición de la reivindicación 5, en la que la composición está presente en un sistema de secuenciación de ADN que comprende al menos un sustrato análogo de nucleótido no natural.

7. Un ácido nucleico modificado que codifica una ADN polimerasa de tipo DPO4 modificada como se expone en una cualquiera de las reivindicaciones 1-4.

55

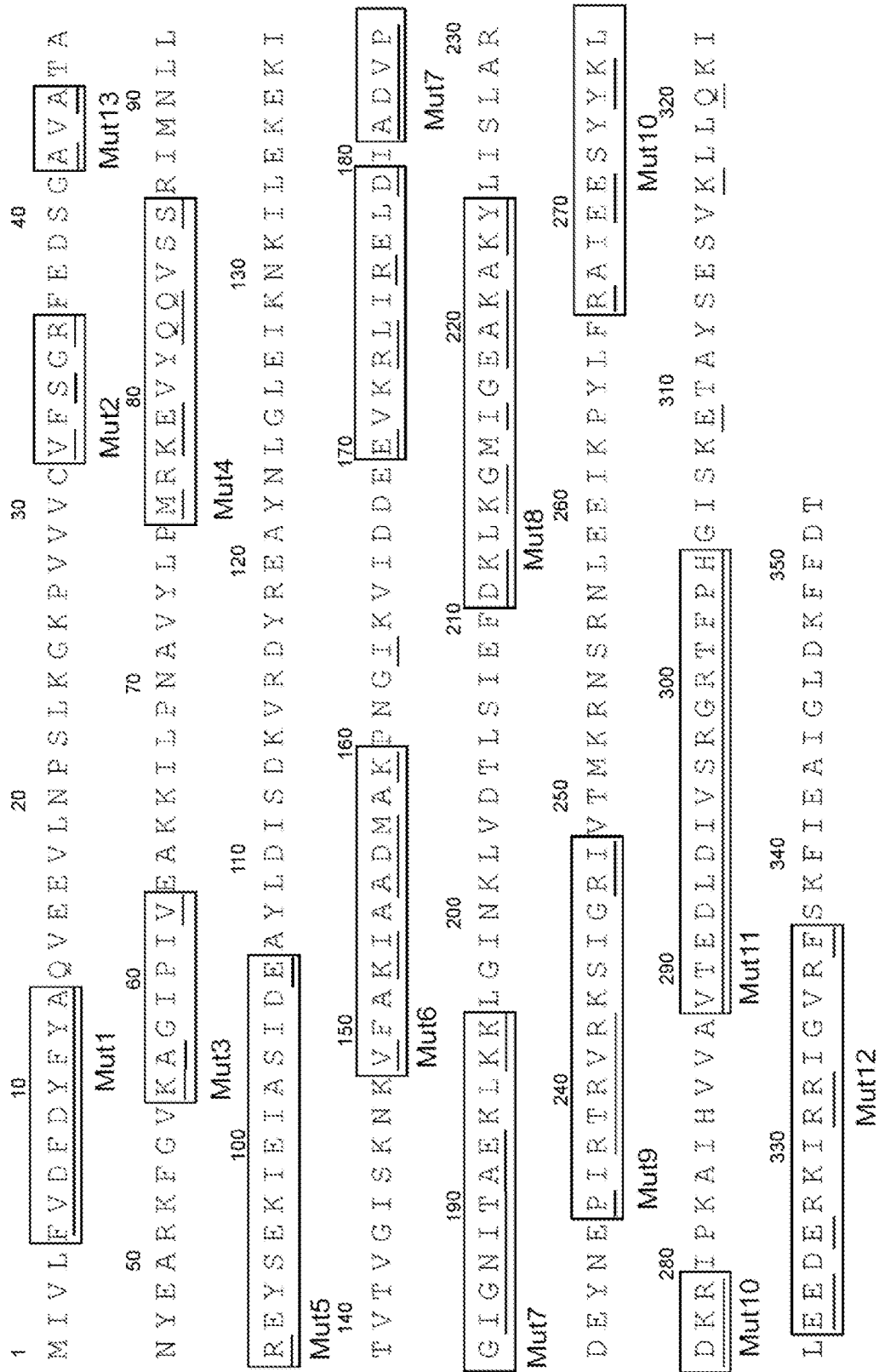


FIG. 1A

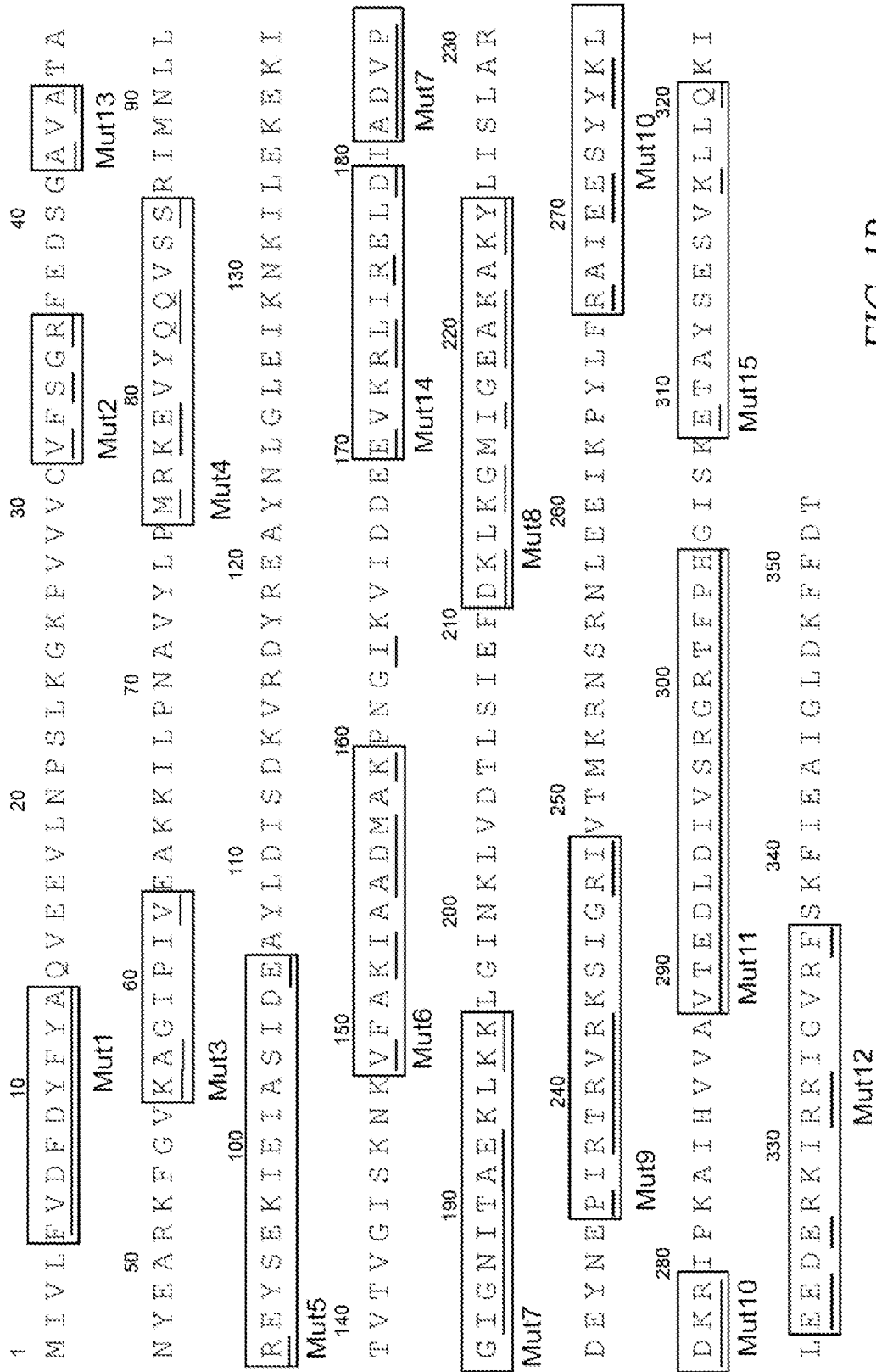


FIG. 1B