(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 May 2020 (22.05.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/099451 A1

(51) International Patent Classification:
*C12N 9/12* (2006.01)

(21) International Application Number:
PCT/EP2019/081099

(22) International Filing Date:
13 November 2019 (13.11.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|---|---|---|
| 18206298.4 | 14 November 2018 (14.11.2018) | EP |
| 19305677.7 | 28 May 2019 (28.05.2019) | EP |
| 19195662.2 | 05 September 2019 (05.09.2019) | EP |


(71) Applicant: **DNA SCRIPT** [FR/FR]; Immeube Okabé, 67 avenue de Fontainebleau, 94270 LE KREMLIN-BICÊTRE (FR).

(72) Inventors: **CHAMPION, Elise**; 198 bis rue de Tolbiac, 75013 PARIS (FR). **LO'CH, Jérôme**; c.o. DNA SCRIPT, Immeuble Okabé, 67 avenue de Fontainebleau, 94270 LE KREMLIN-BICÊTRE (FR). **SOSKINE, Mikhael**; 139 rue du Plessis Bouchard, 95130 FRANCONVILLE (FR). **SUNE, Elodie**; c.o. DNA SCRIPT, Immeuble Okabé, 67 avenue de Fontainebleau, 94270 LE KREMLIN-BICÊTRE (FR).

(74) Agent: **CABINET BECKER ET ASSOCIES**; 25, rue Louis le Grand, 75002 PARIS (FR).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

WO 2020/099451 A1

(54) Title: TERMINAL DEOXYNUCLEOTIDYL TRANSFERASE VARIANTS AND USES THEREOF

(57) Abstract: The present invention is directed to terminal deoxynucleotidyltransferase (TdT) variants that (i) comprise an amino acid sequence that is at least a specified percent identical to an indicated SEQ ID NOs and have at least one substitution at Q455 or at least Q455 plus at least one further substitution at G186, S248, T331, Q390, K394 or H466 (where positions are with respect to SEQ ID NO 1 and functionally equivalent positions in indicated SEQ ID NOs), (ii) are capable of template-free extension of a polynucleotide, and (iii) exhibit enhanced stability or enhanced efficiency in incorporating 3'-0-blocked nucleoside triphosphates into a polynucleotide. The invention is also directed to the use of these TdT variants for synthesizing polynucleotides of any predetermined sequence.

# TERMINAL DEOXYNUCLEOTIDYL TRANSFERASE VARIANTS AND USES THEREOF

## BACKGROUND

**[0001]** The use of highly purified inexpensive polynucleotides of predetermined sequences in a wide range of lengths has become central to a host of technologies, including genomic and diagnostic sequencing, multiplex nucleic acid amplification, therapeutic antibody development, synthetic biology, nucleic acid-based therapeutics, DNA origami, DNA-based data storage, and the like. Recently, interest has arisen in supplementing or replacing chemically-based synthesis methods by enzymatically-based methods using template-free polymerases, such as, terminal deoxynucleotidyl transferase (TdT), because of the proven efficiency of such enzymes and the benefit of mild non-toxic reaction conditions, e.g. Ybert et al, International patent publication WO2015/159023; Hiatt et al, U.S. patent 5763594; Jensen et al, Biochemistry, 57: 1821-1832 (2018); and the like. Most approaches in enzyme-based synthesis require the use of reversibly blocked nucleoside triphosphates in order to obtain a desired sequence in the polynucleotide product. Unfortunately, natural TdTs incorporate such modified nucleoside triphosphates with greatly reduced efficiency as compared to unmodified nucleoside triphosphates.

**[0002]** In view of the above, the field of template-free enzymatically-based polynucleotide synthesis would be advanced if new template-free polymerases, such as variant TdTs, were available that could incorporate reversibly blocked nucleoside triphosphates with greater efficiency.

## SUMMARY OF THE INVENTION

**[0003]** The present invention is directed to terminal deoxynucleotidyl transferase (TdT) variants that display enhanced efficiency in incorporating reversibly blocked nucleoside triphosphates into a polynucleotide, and to their use in synthesizing polynucleotides of any predetermined sequence. Additionally, in some embodiments, TdT variants of the invention exhibit enhanced stability with respect to wildtype enzymes. In part the invention is based on the discovery that the efficiency of TdT-based nucleotide incorporation depends in part on the nucleotide sequence of the 3' end of the polynucleotide being extended; thus, the invention is in part an appreciation and recognition that TdT variants of the invention are capable of efficiently extending a polynucleotide independent of the nucleotide sequence of its 3' end.

**[0004]** In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence at least sixty percent identical to an amino acid sequence selected from SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or 16 with a substitution of glutamine at position 326 with respect to SEQ ID NOs 3, 8, 10 and 14, or glutamine at position 325 with respect to SEQ ID NO: 4 and 5, or glutamine at position 332 with respect to SEQ ID NO: 7 and 9, or glutamine at position 329 with respect to SEQ ID NO: 13 and 16, or glutamine at position 321 with respect to SEQ ID NO: 6, or glutamine at position 327 with respect to SEQ ID NO: 11, or glutamine at position 324 with respect to SEQ ID NO: 12, or glutamine at position 339 with respect to SEQ ID NO: 15, or glutamine at position 309 with respect to SEQ ID NO: 24 and 26, or glutamine at position 307 with respect to SEQ ID NO: 25, wherein the TdT variant (i) is capable of synthesizing a nucleic acid fragment without a template and (ii) is capable of incorporating a 3'-O-modified nucleotide onto a free 3'-hydroxyl of a nucleic acid fragment.

**[0005]** In some embodiments, the above percent identity value is at least 80 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 90 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 95 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 97 percent identity; in some embodiments, the above percent identity value is at least 98 percent identity; in some embodiments, the above percent identity value is at least 99 percent identity. In some embodiments, the variant consists in a sequence selected from the group consisting in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 24, 25 or 26, with one or more of the disclosed amino acid substitutions. As used herein, the percent identity values used to compare a reference sequence to a variant sequence do not include the expressly specified amino acid positions containing substitutions of the variant sequence; that is, the percent identity relationship is between sequences of a reference protein and sequences of a variant protein outside of the expressly specified positions containing substitutions in the variant. Thus, for example, if the reference sequence and the variant sequence each comprised 100 amino acids and the variant sequence had mutations at positions 25 and 81, then the percent identity would be in regard to sequences 1-24, 26-80 and 82-100.

**[0006]** In regard to (ii), such 3'-O-modified nucleotide may comprise a 3'-O-NH2-nucleoside triphosphate, a 3'-O-azidomethyl-nucleoside triphosphate, a 3'-O-allyl-nucleoside triphosphate, a 3'O—(2-nitrobenzyl)-nucleoside triphosphate, or a 3'-O-propargyl-nucleoside triphosphate.

**[0007]** In some embodiments, the above substitution for glutamine is selected from the group consisting of T, F, L and M; in other embodiments, the above substitution for glutamine is selected from the group consisting of T, F, L, M, I, V and Y. In some embodiments, said substitution is F. In some embodiments, the substitution of glutamine may be in combination with other mutations described herein, such as those at the lysine, histidine, alanine, tryptophan, glycine or glutamine below.

**[0008]** In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence at least sixty percent 60% identical to an amino acid sequence selected from SEQ ID NO: 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or 16 and comprising a substitution of lysine at position 265 with respect to SEQ ID NOs 3, 7, 8, 9, 10, 13 and 14, or lysine at position 263 with respect to SEQ ID NOs 6 and 12, or lysine at position 264 with respect to SEQ ID NO 5, or lysine at position 266 with respect to SEQ ID NO 11, or lysine at position 268 with respect to SEQ ID NO 16, or lysine at position 272 with respect to SEQ ID NO 15, wherein the TdT variant (i) is capable of synthesizing a nucleic acid fragment without a template and (ii) is capable of incorporating a 3'-O-modified nucleotide onto a free 3'-hydroxyl of a nucleic acid fragment. In some embodiments, the above percent identity value is at least 80 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 90 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 95 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 97 percent identity; in some embodiments, the above percent identity value is at least 98 percent identity; in some embodiments, the above percent identity value is at least 99 percent identity. In regard to (ii), such 3'-O-modified nucleotide may comprise a 3'-O-NH2-nucleoside triphosphate, a 3'-O-azidomethyl-nucleoside triphosphate, a 3'-O-allyl-nucleoside triphosphate, a 3'O—(2-nitrobenzyl)-nucleoside triphosphate, or a 3'-O-propargyl-nucleoside triphosphate. In some embodiments, the amino acid substitution of lysine is selected from the group consisting of E, T, A and R. In some embodiments, said substitution

is T. In some embodiments, the substitution of lysine may be in combination with other mutations described herein, such as those at the glutamine above, or the histidine, alanine, tryptophan, glycine or glutamine below.

**[0009]** In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence at least sixty percent 60% identical to an amino acid sequence selected from SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or 16 and comprising a substitution of histidine at position 337 with respect to SEQ ID NOs 3, 8, 10 and 14, or histidine at position 336 with respect to SEQ ID NOs 4 and 5, or histidine at position 343 with respect to SEQ ID NOs 7 and 9, or histidine at position 340 with respect to SEQ ID NOs 13 and 16, or histidine at position 332 with respect to SEQ ID NO 6, or histidine at position 338 with respect to SEQ ID NO 11, or histidine at position 335 with respect to SEQ ID NO 12, or histidine at position 350 with respect to SEQ ID NO 15, wherein the TdT variant (i) is capable of synthesizing a nucleic acid fragment without a template and (ii) is capable of incorporating a 3'-O-modified nucleotide onto a free 3'-hydroxyl of a nucleic acid fragment. In some embodiments, the above percent identity value is at least 80 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 90 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 95 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 97 percent identity; in some embodiments, the above percent identity value is at least 98 percent identity; in some embodiments, the above percent identity value is at least 99 percent identity. In regard to (ii), such 3'-O-modified nucleotide may comprise a 3'-O-NH2-nucleoside triphosphate, a 3'-O-azidomethyl-nucleoside triphosphate, a 3'-O-allyl-nucleoside triphosphate, a 3'O—(2-nitrobenzyl)-nucleoside triphosphate, or a 3'-O-propargyl-nucleoside triphosphate. In some embodiments, the amino acid substitution of histidine is selected from the group consisting of Y, F, N and D. In some embodiments, the substitution of histidine may be in combination with other mutations described herein, such as those at the glutamine or lysine above, or the alanine, tryptophan, glycine or glutamine below.

**[0010]** In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence at least sixty percent 60% identical to an amino acid sequence selected from SEQ ID NO: 3, 4, 5, 6, 8, 9, 10, 11, 12, 13,

14 or 15 and comprising a substitution of tryptophan at position 377 with respect to SEQ ID NOs 3, 8, 10 and 14, or tryptophan at position 376 with respect to SEQ ID NOs 4 and 5, or tryptophan at position 372 with respect to SEQ ID NO 6, or tryptophan at position 380 with respect to SEQ ID NO 13, or tryptophan at position 383 with respect to SEQ ID NO 9, or tryptophan at position 378 with respect to SEQ ID NO 11, or tryptophan at position 375 with respect to SEQ ID NO 12, or tryptophan at position 380 with respect to SEQ ID NO 13, or tryptophan at position 390 with respect to SEQ ID NO 15, wherein the TdT variant (i) is capable of synthesizing a nucleic acid fragment without a template and (ii) is capable of incorporating a 3'-O-modified nucleotide onto a free 3'-hydroxyl of a nucleic acid fragment. In some embodiments, the above percent identity value is at least 80 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 90 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 95 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 97 percent identity; in some embodiments, the above percent identity value is at least 98 percent identity; in some embodiments, the above percent identity value is at least 99 percent identity. In regard to (ii), such 3'-O-modified nucleotide may comprise a 3'-O-NH2-nucleoside triphosphate, a 3'-O-azidomethyl-nucleoside triphosphate, a 3'-O-allyl-nucleoside triphosphate, a 3'O—(2-nitrobenzyl)-nucleoside triphosphate, or a 3'-O-propargyl-nucleoside triphosphate. In some embodiments, tryptophan is substituted with R or K. In some embodiments, tryptophan is substituted with R. In some embodiments, the substitution of tryptophan may be in combination with other mutations described herein, such as those at the glutamine, lysine or histidine above, or the alanine, or glycine, or glutamine below.

**[0011]** In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence at least sixty percent 60% identical to an amino acid sequence selected from SEQ ID NO: 3, 4, 5, 6, 8, 10, 11, 13, 14, 15 or 16 with a substitution of alanine at position 17 with respect to SEQ ID NOs 3, 4, 5, 6, 8, 10, 13, 14 and 15, or alanine at position 18 with respect to SEQ ID NO: 11 and 16, wherein the TdT variant (i) is capable of synthesizing a nucleic acid fragment without a template and (ii) is capable of incorporating a 3'-O-modified nucleotide onto a free 3'-hydroxyl of a nucleic acid fragment. In some embodiments, the above percent identity value is at least 80 percent

identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 90 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 95 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 97 percent identity; in some embodiments, the above percent identity value is at least 98 percent identity; in some embodiments, the above percent identity value is at least 99 percent identity. In regard to (iii), such 3'-O-modified nucleotide may comprise a 3'-O-NH2-nucleoside triphosphate, a 3'-O-azidomethyl-nucleoside triphosphate, a 3'-O-allyl-nucleoside triphosphate, a 3'O—(2-nitrobenzyl)-nucleoside triphosphate, or a 3'-O-propargyl-nucleoside triphosphate. In some embodiments, the above substitution for alanine is V, I or L. In some embodiments, the above substitution for alanine is V. In some embodiments, the substitution of alanine may be in combination with other mutations described herein, such as those at the glutamine, lysine, histidine or tryptophan above, or the glycine or glutamine below.

**[0012]** In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence at least sixty percent 60% identical to an amino acid sequence selected from SEQ ID NO: 3, 9, 11 or 15 with a substitution of glycine at position 57 with respect to SEQ ID NOs 3 and 15, or glycine at position 58 with respect to SEQ ID NO: 9 and 11, wherein the TdT variant (i) is capable of synthesizing a nucleic acid fragment without a template and (ii) is capable of incorporating a 3'-O-modified nucleotide onto a free 3'-hydroxyl of a nucleic acid fragment. In some embodiments, the above percent identity value is at least 80 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 90 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 95 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 97 percent identity; in some embodiments, the above percent identity value is at least 98 percent identity; in some embodiments, the above percent identity value is at least 99 percent identity. In regard to (ii), such 3'-O-modified nucleotide may comprise a 3'-O-NH2-nucleoside triphosphate, a 3'-O-azidomethyl-nucleoside triphosphate, a 3'-O-allyl-nucleoside triphosphate, a 3'O—(2-nitrobenzyl)-nucleoside triphosphate, or a 3'-O-propargyl-nucleoside triphosphate. In some embodiments, the above substitution for glycine is E. In some embodiments, the substitution of glycine may be in

combination with other mutations described herein, such as those at the glutamine, lysine, histidine, tryptophan or alanine above and/or the glutamine described below.

**[0013]** In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence at least sixty percent 60% identical to an amino acid sequence selected from SEQ ID NO: 3, 4, or 6and comprising a substitution of glutamine at position 261 with respect to SEQ ID NO: 3, or glutamine at position 262 with respect to SEQ ID NO: 6, or glutamine at position 264 with respect to SEQ ID NO 4, wherein the TdT variant (i) is capable of synthesizing a nucleic acid fragment without a template and (ii) is capable of incorporating a 3'-O-modified nucleotide onto a free 3'-hydroxyl of a nucleic acid fragment. In some embodiments, the above percent identity value is at least 80 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 90 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 95 percent identity with the indicated SEQ ID NOs; in some embodiments, the above percent identity value is at least 97 percent identity; in some embodiments, the above percent identity value is at least 98 percent identity; in some embodiments, the above percent identity value is at least 99 percent identity. In regard to (ii), such 3'-O-modified nucleotide may comprise a 3'-O-NH2-nucleoside triphosphate, a 3'-O-azidomethyl-nucleoside triphosphate, a 3'-O-allyl-nucleoside triphosphate, a 3'O—(2-nitrobenzyl)-nucleoside triphosphate, or a 3'-O-propargyl-nucleoside triphosphate. In some embodiments, the amino acid substitution of glutamine is R. In some embodiments, the substitution of glutamine at this position may be in combination with other mutations described herein, such as those at the glutamine, histidine, lysine, alanine, tryptophan or glycine above.

**[0014]** In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence at least ninety percent identical to the amino acid sequence as set forth in SEQ ID NO: 4 with substitutions at position M63, R207, R324 and E327. In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising or consisting in an amino acid sequence at least ninety percent identical to the amino acid sequence as set forth in SEQ ID NO: 24 with substitutions at positions M47, R190, R308 and E311. In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising or

consisting in an amino acid sequence at least ninety percent identical to the amino acid sequence as set forth in SEQ ID NO: 26 with substitutions at positions M46, R190 and E311. In a particular embodiment, at least one of the amino acid substitutions is selected from the group consisting in M46R, M47R, M63R, R190L, R207L, E227N and E311N. In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence at least ninety percent identical to the amino acid sequence as set forth in SEQ ID NO: 25 with substitutions at positions R184, R306 and E309. Particularly, the substitutions are selected from the group consisting in R184L, R306A and E309N. In an embodiment, the variant consists in an amino acid sequence having at least 90% identity with the amino acid sequence set forth in SEQ ID NO: 25 with the substitutions R184L, R306A and E309N.

**[0015]** In some embodiments, a TdT variant of the invention comprises substitutions of glutamine (first occurrence above), lysine, histidine, tryptophan, alanine, glycine and glutamine (second occurrence above) at the indicated positions and specified SEQ ID NOs described above.

**[0016]** In some embodiments, a TdT variant of the invention comprises an isolated protein.

**[0017]** In addition to the above TdT variants which comprise one or more substitutions of glutamine, lysine, histidine, tryptophan, alanine and/or glycine at the indicated positions of the specified SEQ ID NOs, in some embodiments, each such TdT variant may further comprise from 1 to 11 additional substitutions selected from the positions listed in the rows of Table 1 (set forth in Fig. 2) for the specified SEQ ID NO. Such additional substitutions are sometimes referred to herein as "supplemental substitutions" and contribute to an increased efficiency or rate of incorporation of 3'-O-modified nucleoside triphosphates onto a polynucleotide. In some embodiments, the supplemental substitutions are selected from Table 1. For example, a TdT variant with arginine substituted for glutamine at position 326 of SEQ ID NO 3 (as described above) may further comprise the following substitutions: L52F, M63R, A108V and L131P (that is, substitutions for the first four positions listed in row 2 of Table 1). Thus, in this example, the TdT variant may be characterized as follows using the nomenclature described below: Q326R + L52F + M63R + A108V + L131P with respect to SEQ ID NO: 3, and otherwise comprising at least 90 percent sequence identity with SEQ ID NO 3.

**[0018]** In another example, a TdT variant with arginine substituted for glutamine at position 325 of SEQ ID NO 5, aspartic acid substituted for histidine at position 336 of SEQ ID NO 5, and arginine substituted for tryptophan at position 376 of SEQ ID NO 5, may further comprise the substitutions M63R, A110V, L131P, C173G, R207N, E294A, R325P, E327N and R353K (selected from row 4 of Table 1, set forth in Fig. 2). Thus, in this example, the TdT variant may be characterized as Q325R + H336D + W376D + M63R + A110V + L131P + C173G + R207N + E294A + R325P + E327N + R353K with respect to SEQ ID NO: 5, and otherwise comprise a specified percent sequence identity, such as at least 90 percent sequence identity, with SEQ ID NO: 5.

**[0019]** In addition to the above TdT variants which comprise one or more substitutions of glutamine, lysine, histidine, tryptophan, alanine and/or glycine at the indicated positions of the specified SEQ ID NOs, and in addition to the Supplemental Substitutions described above, in some embodiments, each such TdT variant may further comprise from 1 to 12 additional substitutions selected from the positions listed in the rows of Table 2 (set forth in Fig. 3) for the specified SEQ ID NO. Such additional substitutions are sometimes referred to herein as "Stability Substitutions" and contribute to an increased stability of a TdT variant, particularly with respect to reaction conditions of template-free enzymatic polynucleotide incorporation, and with respect to elevated temperature. For example, a TdT variant may comprise substitutions A17V, K265E and W377R from Table 3B, supplemental substitutions M63R, L131P, C173G, R207L, G284L and R325P from Table 1, and stability substitutions S119A and S146E from Table 2, wherein the position numbers are with respect to SEQ ID NO: 3. Such TdT variant may be designated as A17V + M63R + S119A + L131P + S146E + C173G + R207L + K265E + G284L + R325P + W377R and it is understood that in addition to these specific substitution the TdT variant comprises a specified percent sequence identity, such as at least 90 percent sequence identity, with SEQ ID NO: 3.

**[0020]** In some embodiments, a TdT variant of the invention comprises all or part of a BRCT-like segment attached to its N-terminus, e.g. see Delarue et al, EMBO J., 21(3): 427-439 (2002).

**[0021]** In some embodiments, the invention is directed to TdT variants comprising an amino acid sequence that has at least 90 percent identity with SEQ ID NO: 1, wherein each such TdT variant (i) comprises at least one mutation at one or more positions selected from the

group consisting of G186, S248, T331, Q390, K394, Q455 or H466 with respect to SEQ ID NO: 1, or a functional equivalent thereof, and (ii) is capable of extending a polynucleotide without a template, and (iii) is capable of incorporating 3'-O-modified nucleoside triphosphates with greater efficiency than a wild type TdT. In some embodiments, TdT variants of the invention comprise G186E. In some embodiments, TdT variants of the invention comprise K394 E/T/A/R. In some embodiments, the invention is directed to TdT variants comprising an amino acid sequence that has at least 90 percent identity with SEQ ID NO: 3, wherein each such TdT variant (i) comprises at least one mutation at one or more positions selected from the group consisting of G57, S119, T202, Q261, K265, Q326 or H337 with respect to SEQ ID NO: 3, or a functional equivalent thereof, and (ii) is capable of extending a polynucleotide without a template, and (iii) is capable of incorporating 3'-O-modified nucleoside triphosphates with greater efficiency than a wild type TdT.

**[0022]** In the embodiments of this paragraph, the amino acid position numbers are with respect to SEQ ID NO: 3. In some embodiments, TdT variants of the invention comprise G57E. In some embodiments, TdT variants of the invention comprise K265E/T/A/R. In some embodiments, TdT variants of the invention comprise T202A. In some embodiments, TdT variants of the invention comprise Q326T/F/L/M or Q326T/F/L/M/I/V/Y. In some embodiments, TdT variants of the invention comprise S119A. In some embodiments, TdT variants of the invention comprise Q261R. In some embodiments, TdT variants of the invention comprise H337Y/F/D. In some embodiments, TdT variants of the invention comprise one or more amino acid changes selected from the group consisting of T202A, Q326T/F/L/M or Q326T/F/L/M/I/V/Y/W, S119A, Q261R, H337Y/F/D, G57E and K265E/T/A/R. In some embodiments, TdT variants of the invention comprise K265E/T/A; and in other embodiments comprise both G57E and K265E/T/A. In some embodiments, variant TdTs of the invention have at least 95 percent identity with the reference or wild type TdT sequence SEQ ID NO: 3. In some embodiments, variant TdTs of the invention have at least 98 percent identity with SEQ ID NO: 3. In some embodiments, TdT variants of the invention displaying increase efficiency of incorporation comprise one or more amino acid changes selected from the group consisting of T202A, Q326T/F/L/M or Q326T/F/L/M/I/V/Y, Q261R, H337Y/F/D, G57E and K265E/T/A/R. In some embodiments, TdT variants of the invention displaying enhanced stability comprise the amino acid change S119A. In some embodiments,

TdT variants of the invention displaying increased efficiency of incorporation comprise, either individually or in combination, G57E, K265T/E/R/A, Q326T/F/L/M, or H337Y/F/D.

**[0023]** The invention further relates to the use of a TdT variant of the invention for synthesizing a nucleic acid molecule without template by the successive addition of one or more 3'-O-modified nucleotides to a nucleic acid fragment. In some embodiments, such methods comprise the steps of (a) providing an initiator comprising an oligonucleotide having a free 3'-hydroxyl; (b) reacting under enzymatic extension conditions a TdT variant of the invention with the initiator or an extended initiator in the presence of a 3'-O-reversibly blocked nucleoside triphosphate. In some embodiments, such method further includes steps of (c) deblocking the extended initiators to form extended initiators with free 3'-hydroxyls and (d) repeating steps (b) and (c) until a nucleic acid molecule of a predetermined sequence is synthesized.

**[0024]** In further embodiments, the invention includes nucleic acid molecules encoding a variant TdTs described above, expression vectors comprising such nucleic acid molecules, and host cells comprising the aforementioned nucleic acid molecules or the aforementioned expression vectors. In still further embodiments, the invention includes processes for producing a variant TdT of the invention, wherein a host cell is cultivated under culture conditions allowing the expression of the nucleic acid encoding said variant TdT, and wherein the variant TdT is optionally retrieved. The invention also includes kits for performing template-free polynucleotide elongations of any predetermine sequence, wherein the kits include a TdT variant of the invention. Such kits may further comprise 3'-O-blocked deoxyribonucleoside triphosphates (dNTPs) for A, C, G and T for DNA elongation, or 3'-O-blocked ribonucleoside triphosphates (rNTPs) for rA, rC, rG and U for RNA elongation.

**[0025]** The present invention advantageously overcomes problems in the field of template-free enzymatic nucleic acid synthesis related to the efficient incorporation of 3'-O-modified nucleoside triphosphates by providing new TdT variants with a capability of incorporating 3'-O-modified nucleotides with greater efficiency or at a higher rate than wild type TdTs or previously available TdT variants, particularly with respect to incorporation of nucleotides onto polynucleotides that comprise certain 3' nucleotide sequences described herein. In some embodiments, the present invention also advantageously overcomes problems in the above

field by providing new TdT variants with increased stability in comparison with wild type TdTs.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 illustrates diagrammatically the steps of a method of template-free enzymatic nucleic acid synthesis using TdT variants of the invention.

Fig. 2 contains Table 1 listing supplementary substitutions for the various SEQ ID NOs.

Fig. 3 contains Table 2 listing stability substitutions for the various SEQ ID NOs.

DETAILED DESCRIPTION OF THE INVENTION

**[0026]** While the invention is amenable to various modifications and alternative forms, specifics thereof have been shown by way of example in the drawings and will be described in detail. It should be understood that the intention is not to limit the invention to the particular embodiments described. It is the intention to cover all modifications, equivalents, and alternatives falling within the spirit and scope of the invention. Guidance for aspects of the invention is found in many available references and treatises well known to those with ordinary skill in the art, including, for example, Sambrook et al. (1989), Molecular cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, and the like.

**[0027]** The present invention provides variants of the TdT polymerase that can be used for synthesizing polynucleotides, such as DNA or RNA, of predetermined sequences without the use of template strand. The TdT variants of the invention allow modified nucleotides, and more particularly 3'O-reversibly blocked nucleoside triphosphates, to be used in an enzyme-based method of polynucleotide synthesis. The variants of the present invention are described according to their mutations or substitutions at specific residues, whose positions are designated with respect to a specified SEQ ID NO.

**[0028]** In some embodiments, a TdT variant may be operably linked to a linker moiety including a covalent or non-covalent bond; amino acid tag (e.g., poly-amino acid tag, poly-His tag, 6His-tag); chemical compound (e.g., polyethylene glycol); protein-protein binding pair (e.g., biotin-avidin); affinity coupling; capture probes; or any combination of these. The linker moiety can be separate from or part of a TdT variant (e.g., recombinant His-tagged polymerase, such as exemplified by the following pairs of SEQ ID NOs: 19 and 20, 21 and 22, 23 and 24, and 25 and 26). Typically, the linker moiety does not interfere with the nucleotide binding activity, or catalytic activity of the mutant TdT.

**[0029]** In some of the embodiments described above, the efficiency of a variant TdT in incorporating a 3'O-modified nucleoside triphosphate is at least 105 percent that of a previous available TdT wildtype or variant; in other embodiments, the efficiency of a variant TdT in incorporating a 3'O-modified nucleoside triphosphate is at least 110 percent that of a previous available TdT wildtype or variant; in other embodiments, the efficiency of a variant TdT in incorporating a 3'O-modified nucleoside triphosphate is at least 150 percent that of a previous available TdT wildtype or variant.

**[0030]** In some embodiments, a TdT variant of the invention comprises an amino acid sequences at least 60 percent identical to the SEQ ID NOs specified in Table 3A and comprises at least a substitution of glutamine at the position indicated in column 1 for the specified SEQ ID NO. In some embodiments, a TdT variant of the invention comprises an amino acid sequences at least 60 percent identical to the SEQ ID NOs specified in Table 3A and comprises from 1 to 7 substitutions of glutamine, lysine, histidine, tryptophan, alanine or glycine at the positions indicated in columns 1 to 7, respectively, for the specified SEQ ID NOs. In some embodiments, a TdT variant of the invention comprises an amino acid sequences at least 60 percent identical to the SEQ ID NOs specified in Table 3B and comprises at least a substitution of glutamine with one of T, F, L or M at the indicated position for the specified SEQ ID NO. In some embodiments, a TdT variant of the invention comprises an amino acid sequences at least 60 percent identical to the SEQ ID NOs specified in Table 3B and comprises from 1 to 7 substitutions of glutamine with one of T, F, L or M, lysine with one of E, T, A or R, histidine with one of Y, F or D, tryptophan with R, alanine with V or glycine with E at the positions indicated in columns 1 to 7, respectively, for the specified SEQ ID NOs. Where a cell of the Table is blank at a column and specified SEQ ID NO, the amino acid associated with the column is not present in the specified SEQ ID NO so that there is no substitution at that position of the TdT variant.

**[0031]** Each of the TdT variants described in the previous paragraph further (i) is capable of synthesizing a nucleic acid fragment without a template and (ii) is capable of incorporating a 3'-O-modified nucleotide onto a free 3'-hydroxyl of a nucleic acid fragment. In some embodiments, the above percent identity value is at least 80 percent identity with the indicated SEQ ID NOs. In some embodiments, the percent identity value of the previous paragraph is at least 90 percent identity; in some embodiments, such percent identity value is at least 95

percent identity; in some embodiments, such percent identity value is at least 98 percent identity; in some embodiments, such percent identity value is at least 99 percent identity. In some embodiments, the above-mentioned 3'-O-modified nucleotide may comprise a 3'-O-NH2-nucleoside triphosphate, a 3'-O-azidomethyl-nucleoside triphosphate, a 3'-O-allyl-nucleoside triphosphate, a 3'O—(2-nitrobenzyl)-nucleoside triphosphate, or a 3'-O-propargyl-nucleoside triphosphate.

Table 3A: TdT variants with substitutions at positions Q455, K394, H466, W506, A146, G186 and/or Q390 (with respect to SEQ ID NO: 1) or functionally equivalent positions of the specified SEQ ID NO

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | SEQ ID NO |
|---|---|---|---|---|---|---|---|
| Q455 | K394 | H466 | W506 | A146 | G186 | Q390 | 1 |
| Q326 | K265 | H337 | W377 | A17 | G57 | Q261 | 3 |
| Q325 | --- | H336 | W376 | A17 | --- | Q264 | 4 |
| Q325 | K264 | H336 | W376 | A17 | --- | --- | 5 |
| Q321 | K263 | H332 | W372 | A17 | --- | Q262 | 6 |
| Q332 | K265 | H343 | --- | --- | --- | --- | 7 |
| Q326 | K265 | H337 | W377 | A17 | --- | --- | 8 |
| Q332 | K265 | H343 | W383 | --- | G58 | --- | 9 |
| Q326 | K265 | H337 | W377 | A17 | --- | --- | 10 |
| Q327 | K266 | H338 | W378 | A18 | G58 | --- | 11 |
| Q324 | K263 | H335 | W375 | --- | --- | Q259 | 12 |
| Q329 | K265 | H340 | W380 | A17 | --- | --- | 13 |
| Q326 | K265 | H337 | W377 | A17 | --- | --- | 14 |
| Q339 | K272 | H350 | W390 | A17 | G57 | --- | 15 |
| Q329 | K268 | H340 | --- | A18 | --- | --- | 16 |

Table 3B: TdT variants with indicated substitutions at positions Q455, K394, H466, W506, A146, G186 and/or Q390 (with respect to SEQ ID NO: 1) or functionally equivalent positions of the specified SEQ ID NO

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | SEQ ID NO |
|---|---|---|---|---|---|---|---|
| Q455T/F/L/M | K394E/T/A/R | H466Y/F/D | W506R/K | A146V/I/L | G186E | Q390R | 1 |
| Q326T/F/L/M | K265E/T/A/R | H337Y/F/D | W377R/K | A17V/I/L | G57E | Q261R | 3 |
| Q325T/F/L/M | --- | H336Y/F/D | W376R/K | A17V/I/L | --- | Q264R | 4 |
| Q325T/F/L/M | K264E/T/A/R | H336Y/F/D | W376R/K | A17V/I/L | --- | --- | 5 |
| Q321T/F/L/M | K263E/T/A/R | H332Y/F/D | W372R/K | A17V/I/L | --- | Q262R | 6 |
| Q332T/F/L/M | K265E/T/A/R | H343Y/F/D | --- | --- | --- | --- | 7 |
| Q326T/F/L/M | K265E/T/A/R | H337Y/F/D | W377R/K | A17V/I/L | --- | --- | 8 |
| Q332T/F/L/M | K265E/T/A/R | H343Y/F/D | W383R/K | --- | G58E | --- | 9 |
| Q326T/F/L/M | K265E/T/A/R | H337Y/F/D | W377R/K | A17V/I/L | --- | --- | 10 |
| Q327T/F/L/M | K266E/T/A/R | H338Y/F/D | W378R/K | A18V/I/L | G58E | --- | 11 |
| Q324T/F/L/M | K263E/T/A/R | H335Y/F/D | W375R/K | --- | --- | Q259R | 12 |
| Q329T/F/L/M | K265E/T/A/R | H340Y/F/D | W380R/K | A17V/I/L | --- | --- | 13 |
| Q326T/F/L/M | K265E/T/A/R | H337Y/F/D | W377R/K | A17V/I/L | --- | --- | 14 |
| Q339T/F/L/M | K272E/T/A/R | H350Y/F/D | W390R/K | A17V/I/L | G57E | --- | 15 |
| Q329T/F/L/M | K268E/T/A/R | H340Y/F/D | --- | A18V/I/L | --- | --- | 16 |

As noted above, in some embodiments, TdT variants of the invention may comprise, in addition to the substitutions set forth in Tables 3A and 3B, one or more Supplementary Substitutions at the positions listed in Table 1 (set forth in Fig. 2) with respect to the specified SEQ ID NOs, and/or one or more stability enhancing substitutions at the positions listed in Table 2 with respect to the specified SEQ ID NOs (set forth in Fig. 3).

In regard to the stability enhancing mutations of Table 2, several adjacent substitutions are believed to exert a stabilizing effect by forming salt bridges between side chains. Thus, the equivalent stabilizing effect of substitutions Q166E and D170R may be obtained by switching the positions of E and R. Accordingly, in addition to the substitutions shown in Table 2, stabilizing substitutions also include the pairs, Q166R with D170E, C188D with L189R, and S275R with Q278E.

Particular TdT variants of the invention, DS1001 to DS1018, are set forth in Table 4. Each of the TdT variants DS1001 through DS1018 comprises an amino acid sequence at least 60 percent identical to SEQ ID NO 3 and comprises the substitutions at the indicated positions. In some embodiments, TdT variants DS1001 through DS1018 comprises an amino acid sequence at least 80 percent identical to SEQ ID NO 3 and comprises the substitutions at the indicated positions; in some embodiments, TdT variants DS1001 through DS1018 comprises an amino acid sequence at least 90 percent identical to SEQ ID NO 3 and comprises the substitutions at the indicated positions; in some embodiments, TdT variants DS1001 through DS1018 comprises an amino acid sequence at least 95 percent identical to SEQ ID NO 3 and comprises the substitutions at the indicated positions; in some embodiments, TdT variants DS1001 through DS1018 comprises an amino acid sequence at least 97 percent identical to SEQ ID NO 3 and comprises the substitutions at the indicated positions; in some embodiments, TdT variants DS1001 through DS1018 comprises an amino acid sequence at least 98 percent identical to SEQ ID NO 3 and comprises the substitutions at the indicated positions; in some embodiments, TdT variants DS1001 through DS1018 comprises an amino acid sequence at least 99 percent identical to SEQ ID NO 3 and comprises the substitutions at the indicated positions.

Table 4: Specific TdT Variants of the Invention

| | |
|---|---|
| DS1001 (TH M27)<br>SEQ ID NO: 17 | A17V + L52F + M63R + A108V + C173G + R207L + K265T + G284P + E289V + R325P + E328N + R351K |
| DS1002<br>(M44) | A17V + Q37E + D41R + L52F + G57E + M63R + S94R + G98E + A108V + S119A + L131R + S146E + Q149R + C173G + R207L + K265T + G284P + E289V + R325P + Q326F + E328N + H337D + R351K + W377R |
| DS1003 | A17V + Q37E + D41R + L52F + G57E + M63R + S94R + G98E + A108V + S146E + Q149R + C173G + F193Y + V199M + M201V + R207L + K265T + G284P + E289V + Q326F + E328N + R351K |
| DS1004<br>(M45) | A17V + Q37E + D41R + L52F + G57E + M63R + S94R + G98E + A108V + S146E + Q149R + C173G + F193Y + V199M + M201V + R207L + K265T + G284P + E289V + R325A + Q326F + E328N + R351K |
| DS1005 | A17V + Q37E + D41R + L52F + G57E + M63R + S94R + G98E + A108V + S146E + Q149R + C173G + F193Y + V199M + M201V + R207L + K265T + G284P + E289V + Q326F + E328N + R351K |
| DS1006<br>(M46) | L52F + A108V + R351K + A17V + Q37E + D41R + G57E+ C59R + L60D + M63R + S94R + G98E + S119A + L131R + S146E + Q149R + C173G + R207L + K265T + G284P + E289V + R325A + Q326F + E328N |
| DS1007<br>(M47) | L52F + A108V + R351K + A17V + Q37E + D41R + G57E + C59R + L60D + M63R + S94R + G98E + K118Q + S119A + L131R + S146E + Q149R + C173G + R207L + K265T + G284P + E289V + R325A + Q326F + E328N + W377R |
| DS1008 | A17V + Q37E + D41R + L52F + G57E + C59R + L60D + M63R + S94R + G98E + A108V + S119A + L131R + S146E + Q149R + C173G + R207L + F259S + Q261L + G284P + E289V + R325A + Q326F + E328N + R351K + W377R |
| DS1009<br>(MS 13-34)<br>SEQ ID NO: 18 | A17V + D41R + L53F + G57E + C59R + L60D + M63R + S94R + G98E + K118Q + S119A + L131R + S146E + Q149R + C173G + R207L + K265T + G284P + E289V + R325A + Q326F + R351K + W377R |
| DS1010<br>(MS 34-1)<br>SEQ ID NO: 19 | A17V + D41R + L52F + G57E + C59R + L60D + M63R + S94R + G98E + A108V + S119A + L131R + S146E + Q149R + R207L + K265T + G284P + E289V + R325A + Q326F + R351K |
| DS1011 | A17V + D41R + L53F + G57E + C59R + L60D + M63R + S94R + G98E + K118Q + S119A + L131R + S146E + Q149R + C173G + R207L + K265T + G284P + E289V + Q326F + R351K + W377R |
| DS1012<br>(M48) | A17V + Q37E + D41R + L52F + G57E + C59R + L60D + M63R + S94R + G98E + A108V + S119A + L131R + S146E + Q149R + C173G + R207L + F259S + Q261L, G284P + E289V + R325A + Q326F + E328N + R351K + W377R |
| DS1013 | A17V + Q37E + D41R + L52F + G57E + M63R + S94R + G98E + A108V + S146E + Q149R + C173G + R207L + K265T + G284P + E289V + R325A + Q326F + E328N + R351K |

| | |
|---|---|
| DS1014 (M49) | A17V + Q37E + D41R + L52F + G57E + C59R + L60D + M63R + S94R + G98E + A108V + S119A + L131R + S146E + Q149R + C173G + R207L + E257D + F259S + K260R + Q261L + G284P + E289V + R325A + Q326F + E328N + R351K + W377R |
| DS1015 | A17V + Q37E + D41R + L52F + G57E + C59R + L60D + M63R + S94R + G98E + A108V + S119A + L131R + S146E + Q149R + C173G + F193Y + V199M + M201V + R207L + E257D + F259S + K260R + Q261L + G284P + E289V + R325A + Q326F + E328N + R351K + W377R |
| DS1016 (TH c2_5) SEQ ID NO: 20 | A17V + D41R + L52F + G57E + M63R + S94R + G98E + A108V + S146E + Q149R + C173G + M184T + R207L + K209H + G284L + E289A + R325V + E328K + R351K |
| DS1017 (M27) SEQ ID NO: 32 | A17V + L52F + G57E + M63R + A108V + C173G + R207L + K265T + G284P + E289V + R325P + E328N + R351K |
| DS1018 (M60) | A17V + L32T + Q37R + D41R + L52F + G57E + C59R + L60D + M63R + S67A + S94R + G98E + A108V + S119A +L131R + S146E + Q149R + V171A + S172E + C173R + V182I + S183E + R207L + K209H + M210K + T211I + E223G + A224P + E228D + Q261L + G284P + E289V + R325A + Q326F + E328N + R351K + D372E |

TdT variants of the invention as described above each comprise an amino acid sequence having a percent sequence identity with a specified SEQ ID NO, subject to the presence of indicated substitutions. In some embodiments, the number and type of sequence differences between a TdT variant of the invention described in this manner and the specified SEQ ID NO may be due to substitutions, deletion and/or insertions, and the amino acids substituted, deleted and/or inserted may comprise any amino acid. In some embodiments, such deletions, substitutions and/or insertions comprise only naturally occurring amino acids. In some embodiments, substitutions comprise only conservative, or synonymous, amino acid changes, as described in Grantham, Science, 185: 862-864 (1974). That is, a substitution of an amino acid can occur only among members of its set of synonymous amino acids. In some embodiments, sets of synonymous amino acids that may be employed are set forth in Table 5A.

Table 5A: Synonymous Sets of Amino Acids I

| Amino Acid | Synonymous Set |
|---|---|
| Ser | Ser, Thr, Gly, Asn |
| Arg | Arg, Gln, Lys, Glu, His |
| Leu | Ile, Phe, Tyr, Met, Val, Leu |
| Pro | Gly, Ala, Thr, Pro |
| Thr | Pro, Ser, Ala, Gly, His, Gln, Thr |
| Ala | Gly, Thr, Pro, Ala |
| Val | Met, Tyr, Phe, Ile, Leu, Val |
| Gly | Gly, Ala, Thr, Pro, Ser |
| Ile | Met, Tyr, Phe, Val, Leu, Ile |
| Phe | Trp, Met, Tyr, Ile, Val, Leu, Phe |
| Tyr | Trp, Met, Phe, Ile, Val, Leu, Tyr |
| Cys | Cys, Ser, Thr |
| His | His, Glu, Lys, Gln, Thr, Arg |
| Gln | Gln, Glu, Lys, Asn, His, Thr, Arg |
| Asn | Asn, Gln, Asp, Ser |
| Lys | Lys, Glu, Gln, His, Arg |
| Asp | Asp, Glu, Asn |
| Glu | Glu, Asp, Lys, Asn, Gln, His, Arg |
| Met | Met, Phe, Ile, Val, Leu |
| Trp | Trp |

In some embodiments, sets of synonymous amino acids that may be employed are set forth in Table 5B.

Table 5B: Synonymous Sets of Amino Acids II

| Amino Acid | Synonymous Set |
|---|---|
| Ser | Ser |
| Arg | Arg, Lys, His |
| Leu | Ile, Phe, Met, Leu |
| Pro | Ala, Pro |
| Thr | Thr |
| Ala | Pro, Ala |
| Val | Met, Ile Val |
| Gly | Gly |
| Ile | Met, Phe, Val, Leu, Ile |
| Phe | Met, Tyr, Ile, Leu, Phe |
| Tyr | Trp, Met |
| Cys | Cys, Ser |
| His | His, Gln, Arg |
| Gln | Gln, Glu, His |
| Asn | Asn, Asp |
| Lys | Lys, Arg |

| | |
|---|---|
| Asp | Asp, Asn |
| Glu | Glu, Gln |
| Met | Met, Phe, Ile, Val, Leu |
| Trp | Trp |

Measurement of Nucleotide Incorporation Activity

**[0032]** The efficiency of nucleotide incorporation by variants of the invention may be measured by an extension, or elongation, assay, e.g. as described in Boule et al (cited below); Bentolila et al (cited below); and Hiatt et al, U.S. patent 5808045, the latter of which is incorporated herein by reference. Briefly, in one form of such an assay, a fluorescently labeled oligonucleotide having a free 3'-hydroxyl is reacted under TdT extension conditions with a variant TdT to be tested for a predetermined duration in the presence of a reversibly blocked nucleoside triphosphate, after which the extension reaction is stopped and the amounts of extension products and unextended initiator oligonucleotide are quantified after separation by gel electrophoresis. By such assays, the incorporation efficiency of a variant TdT may be readily compared to the efficiencies of other variants or to that of wild type or reference TdTs, or other polymerases. In some embodiments, a measure of variant TdT efficiency may be a ratio (given as a percentage) of amount of extended product using the variant TdT over the amount of extended product using wild type TdT in an equivalent assay.

**[0033]** In some embodiments, the following particular extension assay may be used to measure incorporation efficiencies of TdTs: Primer used is the following:

5'-AAAAAAAAAAAAAAGGGG-3' (SEQ ID NO: 2)

The primer has also an ATTO fluorescent dye on the 5' extremity. Representative modified nucleotides used (noted as dNTP in Table 5) include 3'-O-amino-2',3'-dideoxynucleotides-5'-triphosphates (ONH2, Firebird Biosciences), such as 3'-O-amino-2',3'-dideoxyadenosine-5'-triphosphate. For each different variant tested, one tube is used for the reaction. The reagents are added in the tube, starting from water, and then in the order of Table 6. After 30 min at 37°C the reaction is stopped by addition of formamide (Sigma).

Table 6: Extension Activity Assay Reagents

| Reagent | Concentration | Volume |
|---|---|---|
| $H_2O$ | - | 12 μL |
| Activity buffer | 10x | 2 μL |
| dNTP | 250 μM | 2 μL |
| Purified enzyme | 20 μM | 2 μL |
| Fluorescent primer | 500 nM | 2 μL |

The Activity buffer comprises, for example, TdT reaction buffer (available from New England Biolabs) supplemented with $CoCl_2$.

**[0034]** The product of the assay is analyzed by conventional polyacrylamide gel electrophoresis. For example, products of the above assay may be analyzed in a 16 percent polyacrylamide denaturing gel (Bio-Rad). Gels are made just before the analysis by pouring polyacrylamide inside glass plates and let it polymerize. The gel inside the glass plates is mounted on an adapted tank filed with TBE buffer (Sigma) for the electrophoresis step. The samples to be analyzed are loaded on the top of the gel. A voltage of 500 to 2,000V is applied between the top and bottom of the gel for 3 to 6h at room temperature. After separation, gel fluorescence is scanned using, for example, a Typhoon scanner (GE Life Sciences). The gel image is analyzed using ImageJ software (imagej.nih.gov/ij/), or its equivalent, to calculate the percentage of incorporation of the modified nucleotides.

**[0035]** Hairpin completion assay. In one aspect, the invention includes methods of measuring the capability of a polymerase, such as a TdT variant, to incorporate a dNTP onto a 3' end of a polynucleotide (i.e. a "test polynucleotide"). One such method comprises providing a test polynucleotide with a free 3' hydroxyl under reaction conditions in which it is substantially only single stranded, but that upon extension with a polymerase, such as a TdT variant, forms a stable hairpin structure comprising a single stranded loop and a double stranded stem, thereby allowing detection of an extension of the 3' end by the presence of the double stranded polynucleotide. The double stranded structure may be detected in a variety of ways including, but not limited to, fluorescent dyes that preferentially fluoresce upon intercalation into the double stranded structure, fluorescent resonance energy transfer (FRET) between an acceptor (or donor) on the extended polynucleotide and a donor (or acceptor) on an oligonucleotide that forms a triplex with the newly formed hairpin stem, FRET acceptors and donors that are both attached to the test polynucleotide and that are brought into FRET

proximity upon formation of a hairpin, or the like. In some embodiments, a stem portion of a test polynucleotide after extension by a single nucleotide is in the range of 4 to 6 basepairs in length; in other embodiments, such stem portion is 4 to 5 basepairs in length; and in still other embodiments, such stem portion is 4 basepairs in length. In some embodiments, a test polynucleotide has a length in the range of from 10 to 20 nucleotides; in other embodiments, a test polynucleotide has a length in the range of from 12 to 15 nucleotides. In some embodiments, it is advantageous or convenient to extend the test polynucleotide with a nucleotide that maximizes the difference between the melting temperatures of the stem without extension and the stem with extension; thus, in some embodiments, a test polynucleotide is extended with a dC or dG (and accordingly the test polynucleotide is selected to have an appropriate complementary nucleotide for stem formation).

**[0036]** Exemplary test polynucleotides for hairpin completion assays include p875 (5'-CAGTTAAAAACT) (SEQ ID NO: 21) which is completed by extending with a dGTP; p876 (5'- GAGTTAAAACT) (SEQ ID NO: 22) which is completed by extending with a dCTP; and p877 (5'- CAGCAAGGCT) (SEQ ID NO: 23) which is completed by extending with a dGTP. Exemplary reaction conditions for such test polynucleotides may comprise: 2.5 - 5 μM of test polynucleotide, 1:4000 dilution of GelRed® (intercalating dye from Biotium, Inc., Fremont, CA), 200mM Cacodylate KOH pH 6.8, 1mM $CoCl_2$, 0-20% of DMSO and 3' $ONH_2$ dGTP and TdT at desired concentrations. Completion of the hairpin may be monitored by an increase in fluorescence of GelRed® dye using a conventional fluorimeter, such as a TECAN reader at a reaction temperature of 28-38°C, using an excitation filter set to 360nm and anemission filter set to 635nm.

**[0037]** In some embodiments of this aspect of the invention, TdT variants may be tested for their capacity for template-free incorporate of nucleoside triphosphates by the following steps: (a) combining a test polynucleotide having a free 3'-hydroxyl, a TdT variant and a nucleoside triphosphate under conditions wherein the test polynucleotide is single stranded but upon incorporation of the nucleoside triphosphate forms a hairpin having a double stranded stem region, and (b) detecting the amount of double stranded stem regions formed as a measure of the capacity of the TdT variant to incorporate the nucleoside triphosphate. In some embodiments, the nucleoside triphosphate is a 3'-O-blocked nucleoside triphosphate.

Measurement of Enzyme Stability

**[0038]** In some embodiments, enzyme stability means a capability of an enzyme (or variant thereof) to retain a particular activity after it has been subjected to destabilizing conditions for a period of time, such as, elevated temperature, lowered temperature, low pH, high pH, exposure to a chaotropic agent, or the like. In some embodiments, enzyme stability may be measured by exposing the enzyme to elevated temperatures, e.g. in the range of 50-70° C. for a period of time, e.g. in the range of 15-30 minutes, after which the activity of template-free elongation of an initiator stranded using a 3'-modified NTP is tested. In other embodiments, enzyme stability may be measured by exposing the enzyme to low pH, e.g. pH in the range of 1-4, for a period of time, e.g. in the range of 15-30 minutes. In some embodiments, TdT variants of the invention having enhanced stability with respect to elevated temperature display template-free initiator elongation activity using 3'-O-modified dNTPs equal to or greater than that of wild type TdT. In some embodiments, TdT variants of the invention having enhanced stability with respect to pH display template-free initiator elongation activity using 3'-O-modified dNTPs equal to or greater than that of wild type TdT. In some embodiments, such elevated temperature or pH stability is with respect to template-free initiator elongation activity using 3'-O-amine protected dNTPs.

Template-Free Enzymatic Synthesis

**[0039]** Template-free enzymatic synthesis of polynucleotides may be carried out by a variety of known protocols using template-free polymerases, such as terminal deoxynucleotidyl transferase (TdT), including variants thereof engineered to have improved characteristics, such as greater temperatue stability or greater efficiency in the incorporation of 3'-O-blocked deoxynucleoside triphosphates (3'-O-blocked dNTPs), e.g. Ybert et al, International patent publication WO/2015/159023; Ybert et al, International patent publication WO/2017/216472; Hyman, U.S. patent 5436143; Hiatt et al, U.S. patent 5763594; Jensen et al, Biochemistry, 57: 1821-1832 (2018); Mathews et al, Organic & Biomolecular Chemistry, DOI: 0.1039/c6ob01371f (2016); Schmitz et al, Organic Lett., 1(11): 1729-1731 (1999).

**[0040]** In some embodiments, the method of enzymatic DNA synthesis comprises repeated cycles of steps, such as are illustrated in Fig. 1, in which a predetermined nucleotide is added in each cycle. Initiator polynucleotides (100) are provided, for example, attached to solid

support (102), which have free 3'-hydroxyl groups (103). To the initiator polynucleotides (100) (or elongated initiator polynucleotides in subsequent cycles) are added a 3'-O-protected-dNTP and a TdT variant under conditions (104) effective for the enzymatic incorporation of the 3'-O-protected-dNTP onto the 3' end of the initiator polynucleotides (100) (or elongated initiator polynucleotides). This reaction produces elongated initiator polynucleotides whose 3'-hydroxyls are protected (106). If the elongated initiator polynucleotide contains a competed sequence, then the 3'-O-protection group is removed, or deprotected, and the desired sequence is cleaved from the original initiator polynucleotide. Such cleavage may be carried out using any of a variety of single strand cleavage techniques, for example, by inserting a cleavable nucleotide or cleavable linker at a predetermined location within the original initiator polynucleotide. Exemplary cleavable nucleotides or linkers include, but are not limited to, (i) a uracil nucleotide which is cleaved by uracil DNA glycosylase; (ii) a photocleavable group, such as a nitrobenzyl linker, as described in U.S. patent 5,739,386; or an inosine which is cleaved by endonuclease V. In some embodiments, a cleaved polynucleotide may have a free 5'-hydroxyl; in other embodiments, a cleaved polynucleotide may have a 5'-phosphorylated end. If the elongated initiator polynucleotide does not contain a completed sequence, then the 3'-O-protection groups are removed to expose free 3'-hydroxyls (103) and the elongated initiator polynucleotides are subjected to another cycle of nucleotide addition and deprotection.

**[0041]** In some embodiments, an ordered sequence of nucleotides is coupled to an initiator nucleic acid using a TdT in the presence of 3'-O-reversibly blocked dNTPs in each synthesis step. In some embodiments, the method of synthesizing an oligonucleotide comprises the steps of (a) providing an initiator having a free 3'-hydroxyl; (b) reacting under extension conditions the initiator or an extension intermediate having a free 3'-hydroxyl with a TdT in the presence of a 3'-O-blocked nucleoside triphosphate to produce a 3'-O-blocked extension intermediate; (c) deblocking the extension intermediate to produce an extension intermediate with a free 3'-hydroxyl; and (d) repeating steps (b) and (c) until the polynucleotide is synthesized. (Sometime "an extension intermediate" is also referred to as an "elongation fragment."). In some embodiments, an initiator is provided as an oligonucleotide attached to a solid support, e.g. by its 5' end. The above method may also include washing steps after the reaction, or extension, step, as well as after the de-blocking step. For example, the step of reacting may include a sub-step of removing unincorporated nucleoside triphosphates, e.g. by

washing, after a predetermined incubation period, or reaction time. Such predetermined incubation periods or reaction times may be a few seconds, e.g. 30 sec, to several minutes, e.g. 30 min.

**[0042]** The above method may also include capping step(s) as well as washing steps after the reacting, or extending, step, as well as after the deblocking step. As mentioned above, in some embodiments, capping steps may be included in which non-extended free 3'-hydroxyls are reacted with compounds that prevents any further extensions of the capped strand. In some embodiments, such compound may be a dideoxynucleoside triphosphate. In other embodiments, non-extended strands with free 3'-hydroxyls may be degraded by treating them with a 3'-exonuclease activity, e.g. Exo I. For example, see Hyman, U.S. patent 5436143. Likewise, in some embodiments, strands that fail to be deblocked may be treated to either remove the strand or render it inert to further extensions.

**[0043]** In some embodiments that comprise serial synthesis of oligonucleotides, capping steps may be undesirable as capping may prevent the production of equal molar amounts of a plurality of oligonucleotides. Without capping, sequences will have a uniform distribution of deletion errors, but each of a plurality of oligonucleotides will be present in equal molar amounts. This would not be the case where non-extended fragments are capped.

**[0044]** In some embodiments, reaction conditions for an extension or elongation step may comprising the following: 2.0 μM purified TdT; 125-600 μM 3'-O-blocked dNTP (e.g. 3'-O-$NH_2$-blocked dNTP); about 10 to about 500 mM potassium cacodylate buffer (pH between 6.5 and 7.5) and from about 0.01 to about 10 mM of a divalent cation (e.g. $CoCl_2$ or $MnCl_2$), where the elongation reaction may be carried out in a 50 μL reaction volume, at a temperature within the range RT to 45°C, for 3 minutes. In embodiments, in which the 3'-O-blocked dNTPs are 3'-O-$NH_2$-blocked dNTPs, reaction conditions for a deblocking step may comprise the following: 700 mM $NaNO_2$; 1 M sodium acetate (adjusted with acetic acid to pH in the range of 4.8-6.5), where the deblocking reaction may be carried out in a 50 μL volume, at a temperature within the range of RT to 45°C for 30 seconds to several minutes.

**[0045]** Depending on particular applications, the steps of deblocking and/or cleaving may include a variety of chemical or physical conditions, e.g. light, heat, pH, presence of specific reagents, such as enzymes, which are able to cleave a specified chemical bond. Guidance in selecting 3'-O-blocking groups and corresponding de-blocking conditions may be found in the

following references, which are incorporated by reference: U.S. patent 5808045; U.S. patent 8808988; International patent publication WO91/06678; and references cited below. In some embodiments, the cleaving agent (also sometimes referred to as a de-blocking reagent or agent) is a chemical cleaving agent, such as, for example, dithiothreitol (DTT). In alternative embodiments, a cleaving agent may be an enzymatic cleaving agent, such as, for example, a phosphatase, which may cleave a 3'-phosphate blocking group. It will be understood by the person skilled in the art that the selection of deblocking agent depends on the type of 3'-nucleotide blocking group used, whether one or multiple blocking groups are being used, whether initiators are attached to living cells or organisms or to solid supports, and the like, that necessitate mild treatment. For example, a phosphine, such as tris(2-carboxyethyl)phosphine (TCEP) can be used to cleave a 3'O-azidomethyl groups, palladium complexes can be used to cleave a 3'O-allyl groups, or sodium nitrite can be used to cleave a 3'O-amino group. In particular embodiments, the cleaving reaction involves TCEP, a palladium complex or sodium nitrite.

**[0046]** As noted above, in some embodiments it is desirable to employ two or more blocking groups that may be removed using orthogonal de-blocking conditions. The following exemplary pairs of blocking groups may be used in parallel synthesis embodiments, such as those described above. It is understood that other blocking group pairs, or groups containing more than two, may be available for use in these embodiments of the invention.

| | |
|---|---|
| 3'-O-NH2 | 3'-O-azidomethyl |
| 3'-O-NH2 | 3'-O-allyl |
| 3'-O-NH2 | 3'-O-phosphate |
| 3'-O-azidomethyl | 3'-O-allyl |
| 3'-O-azidomethyl | 3'-O-phosphate |
| 3'-O-allyl | 3'-O-phosphate |

**[0047]** Synthesizing oligonucleotides on living cells requires mild deblocking, or deprotection, conditions, that is, conditions that do not disrupt cellular membranes, denature proteins, interfere with key cellular functions, or the like. In some embodiments, deprotection

conditions are within a range of physiological conditions compatible with cell survival. In such embodiments, enzymatic deprotection is desirable because it may be carried out under physiological conditions. In some embodiments specific enzymatically removable blocking groups are associated with specific enzymes for their removal. For example, ester- or acyl-based blocking groups may be removed with an esterase, such as acetylesterase, or like enzyme, and a phosphate blocking group may be removed with a 3' phosphatase, such as T4 polynucleotide kinase. By way of example, 3'-O-phosphates may be removed by treatment with as solution of 100 mM Tris-HCl (pH 6.5) 10 mM $MgCl_2$ , 5 mM 2-mercaptoethanol, and one Unit T4 polynucleotide kinase. The reaction proceeds for one minute at a temperature of 37°C.

**[0048]** A "3'-phosphate-blocked" or "3'-phosphate-protected" nucleotide refers to nucleotides in which the hydroxyl group at the 3'-position is blocked by the presence of a phosphate containing moiety. Examples of 3'-phosphate-blocked nucleotides in accordance with the invention arc nucleotidyl-3'-phosphate monoester/nucleotidyl-2',3'-cyclic phosphate, nuclcotidyl-2'-phosphate monoester and nucleotidyl-2' or 3'-alkylphosphate diester, and nucleotidyl-2' or 3'-pyrophosphate. Thiophosphate or other analogs of such compounds can also be used, provided that the substitution does not prevent dephosphorylation resulting in a free 3'-OH by a phosphatase.

**[0049]** Further examples of synthesis and enzymatic deprotection of 3'-O-ester-protected dNTPs or 3'-O-phosphate-protected dNTPs are described in the following references: Canard et al, Proc. Natl. Acad. Sci., 92:10859-10863 (1995); Canard et al, Gene, 148: 1-6 (1994); Cameron et al, Biochemistry, 16(23): 5120-5126 (1977); Rasolonjatovo et al, Nucleosides & Nucleotides, 18(4&5): 1021-1022 (1999); Ferrero et al, Monatshefte fur Chemie, 131: 585-616 (2000); Taunton-Rigby et al, J. Org. Chem., 38(5): 977-985 (1973); Uemura et al, Tetrahedron Lett., 30(29): 3819-3820 (1989); Becker et al, J. Biol. Chem., 242(5): 936-950 (1967); Tsien, International patent publication WO1991/006678.

**[0050]** As used herein, an "initiator" (or equivalent terms, such as, "initiating fragment," "initiator nucleic acid," "initiator oligonucleotide," or the like) refers to a short oligonucleotide sequence with a free 3'-end, which can be further elongated by a template-free polymerase, such as TdT. In one embodiment, the initiating fragment is a DNA initiating fragment. In an alternative embodiment, the initiating fragment is an RNA initiating fragment. In one

embodiment, the initiating fragment possesses between 3 and 100 nucleotides, in particular between 3 and 20 nucleotides. In one embodiment, the initiating fragment is single-stranded. In an alternative embodiment, the initiating fragment is double-stranded. In a particular embodiment, an initiator oligonucleotide synthesized with a 5'-primary amine may be covalently linked to magnetic beads using the manufacturer's protocol. Likewise, an initiator oligonucleotide synthesized with a 3'-primary amine may be covalently linked to magnetic beads using the manufacturer's protocol. A variety of other attachment chemistries amenable for use with embodiments of the invention are well-known in the art, e.g. Integrated DNA Technologies brochure, "Strategies for Attaching Oligonucleotides to Solid Supports," v.6 (2014); Hermanson, Bioconjugate Techniques, Second Edition (Academic Press, 2008); and like references.

**[0051]** Many of the 3'-O-blocked dNTPs employed in the invention may be purchased from commercial vendors or synthesized using published techniques, e.g. U.S. patent 7057026; International patent publications WO2004/005667, WO91/06678; Canard et al, Gene (cited above); Metzker et al, Nucleic Acids Research, 22: 4259-4267 (1994); Meng et al, J. Org. Chem., 14: 3248-3252 (3006); U.S. patent publication 2005/037991. In some embodiments, the modified nucleotides comprise a modified nucleotide or nucleoside molecule comprising a purine or pyrimidine base and a ribose or deoxyribose sugar moiety having a removable 3'-OH blocking group covalently attached thereto, such that the 3' carbon atom has attached a group of the structure:

-O-Z

wherein –Z is any of $–C(R')_2-O-R"$, $-C(R')_2-N(R")_2$, $-C(R')_2-N(H)R"$, $-C(R')_2-S-R"$ and $–C(R')_2-F$, wherein each R" is or is part of a removable protecting group; each R' is independently a hydrogen atom, an alkyl, substituted alkyl, arylalkyl, alkenyl, alkynyl, aryl, heteroaryl, heterocyclic, acyl, cyano, alkoxy, aryloxy, heteroaryloxy or amido group, or a detectable label attached through a linking group; with the proviso that in some embodiments such substituents have up to 10 carbon atoms and/or up to 5 oxygen or nitrogen heteroatoms; or $(R')_2$ represents a group of formula $=C(R"')_2$ wherein each R"' may be the same or different and is selected from the group comprising hydrogen and halogen atoms and alkyl groups, with the proviso that in some embodiments the alkyl of each R"' has from 1 to 3 carbon atoms; and wherein the molecule may be reacted to yield an intermediate in which each R" is exchanged

for H or, where Z is $–(R')_2$-F, the F is exchanged for OH, SH or $NH_2$, preferably OH, which intermediate dissociates under aqueous conditions to afford a molecule with a free 3'-OH; with the proviso that where Z is $–C(R')_2$-S-R", both R' groups are not H. In certain embodiments, R' of the modified nucleotide or nucleoside is an alkyl or substituted alkyl, with the proviso that such alkyl or substituted alkyl has from 1 to 10 carbon atoms and from 0 to 4 oxygen or nitrogen heteroatoms. In certain embodiments, -Z of the modified nucleotide or nucleoside is of formula $–C(R')_2$-N3. In certain embodiments, Z is an azidomethyl group.

**[0052]** In some embodiments, Z is a cleavable organic moiety with or without heteroatoms having a molecular weight of 200 or less. In other embodiments, Z is a cleavable organic moiety with or without heteroatoms having a molecular weight of 100 or less. In other embodiments, Z is a cleavable organic moiety with or without heteroatoms having a molecular weight of 50 or less. In some embodiments, Z is an enzymatically cleavable organic moiety with or without heteroatoms having a molecular weight of 200 or less. In other embodiments, Z is an enzymatically cleavable organic moiety with or without heteroatoms having a molecular weight of 100 or less. In other embodiments, Z is an enzymatically cleavable organic moiety with or without heteroatoms having a molecular weight of 50 or less. In other embodiments, Z is an enzymatically cleavable ester group having a molecular weight of 200 or less. In other embodiments, Z is a phosphate group removable by a 3'-phosphatase. In some embodiments, one or more of the following 3'-phosphatases may be used with the manufacturer's recommended protocols: T4 polynucleotide kinase, calf intestinal alkaline phosphatase, recombinant shrimp alkaline phosphatase (e.g. available from New England Biolabs, Beverly, MA)

**[0053]** In a further particular embodiment, the 3'-blocked nucleotide triphosphate is blocked by either a 3'-O-azidomethyl, 3'-O-$NH_2$ or 3'-O-allyl group.

**[0054]** In still other embodiments, 3'-O-blocking groups of the invention include 3'-O-methyl, 3'-O-(2-nitrobenzyl), 3'-O-allyl, 3'-O-amine, 3'-O-azidomethyl, 3'-O-tert-butoxy ethoxy, 3'-O-(2-cyanoethyl), and 3'-O-propargyl.

Production of Variant TdTs

**[0055]** Variants of the invention may be produced by mutating known reference or wild type TdT-coding polynucleotides, then expressing it using conventional molecular biology techniques. For example, the mouse TdT gene (SEQ ID NO: 1) may be assembled from synthetic fragments using conventional molecular biology techniques, e.g. using protocols described by Stemmer et al, Gene, 164: 49-53 (1995); Kodumal et al, Proc. Natl. Acad. Sci., 101: 15573-15578 (2004); or the like, or it may be directly cloned from mouse cells using protocols described by Boule et al, Mol. Biotechnology, 10: 199-208 (1998), or Bentolila et al, EMBO J., 14: 4221-4229 (1995); or the like.

**[0056]** For example, an isolated TdT gene may be inserted into an expression vector, such as pET32 (Novagen) to give a vector pCTdT which then may be used to make and express variant TdT proteins using conventional protocols. Vectors with the correct sequence may be transformed in E. coli producer strains.

**[0057]** Transformed strains are cultured using conventional techniques to pellets from which TdT protein is extracted. For example, previously prepared pellets are thawed in 30 to 37°C water bath. Once fully thawed, pellets are resuspended in lysis buffer composed of 50mM tris-HCL (Sigma) pH 7.5, 150mM NaCl (Sigma), 0.5mM mercaptoethanol (Sigma), 5% glycerol (Sigma), 20mM imidazole (Sigma) and 1 tab for 100mL of protease cocktail inhibitor (Thermofisher). Careful resuspension is carried out in order to avoid premature lysis and remaining of aggregates. Resuspended cells are lysed through several cycles of French press, until full color homogeneity is obtained. Usual pressure used is 14,000psi. Lysate is then centrifuged for 1h to 1h30 at 10,000 rpm. Centrifugate is pass through a 0.2µm filter to remove any debris before column purification.

**[0058]** TdT protein may be purified from the centrifugate in a one-step affinity procedure. For example, Ni-NTA affinity column (GE Healthcare) is used to bind the polymerases. Initially the column has been washed and equilibrated with 15 column volumes of 50mM tris-HCL (Sigma) pH 7.5, 150mM NaCl (Sigma) and 20mM imidazole (Sigma). Polymerases are bound to the column after equilibration. Then a washing buffer, composed of 50mM tris-HCL (Sigma) pH 7.5, 500mM NaCl (Sigma) and 20mM imidazole (Sigma), is applied to the column for 15 column volumes. After wash the polymerases are eluted with 50mM tris-HCL (Sigma)

pH 7.5, 500mM NaCl (Sigma) and 0.5M imidazole (Sigma). Fractions corresponding to the highest concentration of polymerases of interest are collected and pooled in a single sample. The pooled fractions are dialyzed against the dialysis buffer (20 mM Tris-HCl, pH 6.8, 200mM Na Cl, 50mM MgOAc, 100mM [NH4]2SO4). The dialysate is subsequently concentrated with the help of concentration filters (Amicon Ultra-30, Merk Millipore). Concentrated enzyme is distributed in small aliquots, 50% glycerol final is added, and those aliquots are then frozen at -20°C and stored for long term. 5µL of various fraction of the purified enzymes are analyzed in SDSPAGE gels.

Kits for Practicing Methods of the Invention

**[0059]** The invention includes a variety of kits for practicing methods of the invention. In one aspect, kits of the invention comprise a TdT variant of the invention in a formulation suitable for carrying out template-free enzymatic polynucleotide synthesis as described herein. Such kits may also include synthesis buffers that provide reaction conditions for optimizing the template-free addition or incorporation of a 3'-O-protected dNTP to a growing strand. In some embodiments, kits of the invention further include 3'-O-reversibly protected dNTPs. In such embodiments, the 3'-O-reversibly protected dNTPs may comprise 3'-O-amino-dNTPs or 3'-O-azidomethyl-dNTPs. In further embodiments, kits may include one or more of the following items, either separately or together with the above-mentioned items: (i) deprotection reagents for carrying out a deprotecting step as described herein, (ii) solid supports with initiators attached thereto, (iii) cleavage reagents for releasing completed polynucleotides from solid supports, (iv) wash reagents or buffers for removing unreacted 3'-O-protected dNTPs at the end of an enzymatic addition or coupling step, and (v) post-synthesis processing reagents, such as purification columns, desalting reagents, eluting reagents, and the like.

**[0060]** In regard to items (ii) and (iii) above, certain initiators and cleavage reagents go together. For example, an initiator comprising an inosine cleavable nucleotide may come with an endonuclease V cleavage reagent; an initiator comprising a nitrobenzyl photocleavable linker may come with a suitable light source for cleaving the photocleavable linker; an initiator comprising a uracil may come with a uracil DNA glycosylase cleavage reagent; and the like.

Example 1: Generation, Expression and Purification of TdT Variants

**[0061]** Expression strain generation. The TdT mouse gene may be generated from the pET28 plasmid described in [Boulé et al., 1998, Mol. Biotechnol. 10, 199-208]. For example, the gene may be amplified by using the following primers:

- T7-pro: TAATACGACTCACTATAGGG (SEQ ID N°33)
- T7-ter: GCTAGTTATTGCTCAGCGG (SEQ ID N°34)

through standard molecular biology techniques. The sequence is then cloned into plasmid pET32 backbone to give the new pCTdT plasmid. After sequencing pCTdT is transformed into commercial E. coli cells, BL21 (DE3, from Novagen). Growing colonies on plate with kanamycin are isolated and named Ec-CTdT.Polymerase variants generation. The pCTdT vector is used as starting vector. Specific primers comprising one or several point mutations have been generated from Agilent online software (http://www.genomics.agilent.com:80/primerDesignProgram.jsp). The commercially available kit QuickChange II (Agilent) may be used to generate the desired modified polymerase comprising the targeted mutations. Experimental procedure follows the supplier's protocol. After generation of the different vectors, each of them is sequenced. Vectors with the correct sequence are transformed in E. coli producer strains. Clones able to grow on kanamycin LB-agar plates are isolated.

**[0062]** Expression. The Ec-CTdT and Ec-DSi or Ec-DSi' strains may be used for inoculating 250mL erlens with 50mL of LB media supplemented with appropriate amount of kanamycin. After overnight growth at 37°C, appropriate volumes of these pre-cultures are used to inoculate 5L erlens with 2L LB media with kanamycin. The initial OD for the 5L cultures is chosen to be 0.01. The erlens are put at 37°C under strong agitation and the OD of the different cultures are regularly checked. After reaching an OD comprised between 0.6 and 0.9 each erlen is supplemented by the addition of 1mL of 1M IPTG (Isopropyl β-D-1-thiogalactopyranoside, Sigma). The erlens are put back to agitation under a controlled temperature of 37°C. After overnight expression, the cells are harvested in several pellets. Pellets expressing the same variants are pooled and stored at -20°C, eventually for several months.

**[0063]** Extraction. Previously prepared pellets are thawed in 30 to 37°C water bath. Once fully thawed, pellets are resuspended in lysis buffer composed of 50mM tris-HCL (Sigma) pH

7.5, 150mM NaCl (Sigma), 0.5mM mercaptoethanol (Sigma), 5% glycerol (Sigma), 20mM imidazole (Sigma) and 1 tab for 100mL of protease cocktail inhibitor (Thermofisher). Careful resuspension is carried out in order to avoid premature lysis and remaining of aggregates. Resuspended cells are lysed through several cycles of French press, until full color homogeneity is obtained. Usual pressure used is 14,000psi. Lysate is then centrifuged for 1h to 1h30 at 10,000 rpm. Centrifugate is pass through a 0.2µm filter to remove any debris before column purification.

**[0064]** Purification. A one-step affinity procedure is used to purify the produced and extracted polymerase enzymes. A Ni-NTA affinity column (GE Healthcare) is used to bind the polymerases. Initially the column has been washed and equilibrated with 15 column volumes of 50mM tris-HCL (Sigma) pH 7.5, 150mM NaCl (Sigma) and 20mM imidazole (Sigma). Polymerases are bound to the column after equilibration. Then a washing buffer, composed of 50mM tris-HCL (Sigma) pH 7.5, 500mM NaCl (Sigma) and 20mM imidazole (Sigma), is applied to the column for 15 column volumes. After wash the polymerases are eluted with 50mM tris-HCL (Sigma) pH 7.5, 500mM NaCl (Sigma) and 0.5M imidazole (Sigma). Fractions corresponding to the highest concentration of polymerases of interest are collected and pooled in a single sample. The pooled fractions are dialyzed against the dialysis buffer (20 mM Tris-HCl, pH 6.8, 200mM Na Cl, 50mM MgOAc, 100mM $[NH_4]_2SO_4$). The dialysate is subsequently concentrated with the help of concentration filters (Amicon Ultra-30, Merk Millipore). Concentrated enzyme is distributed in small aliquots, 50% glycerol final is added, and those aliquots are then frozen at -20°C and stored for long term. 5µL of various fraction of the purified enzymes are analyzed in SDS-PAGE gels.

Example 2: Efficiency of TdT Variants For Synthesizing Difficult Sequences

**[0065]** As noted above, the invention is based in part on a recognition and appreciation by the inventors that certain nucleotide sequences are difficult for TdTs to extend. Thus, an object of this experiment was to discover new TdT variants that exhibit enhanced capability to synthesize such difficult sequences based on comparison to an earlier TdT variant, designated M27 (SEQ ID NO: 32). In this example, a mutation library of mouse TdT (SEQ ID NO: 3) was produced based on structural information from mouse TdT, the activities of TdT variants from prior libraries, and from conventional protein engineering techniques. TdT variants from

the library are expressed and purified as described above and were screened for their capability to synthesize certain difficult-to-synthesize sequences (shown in Table 7) at a higher rate than that of M27. Synthesis of the short sequences was performed by repeating 5 cycles of synthesis wherein each cycle comprises two steps: (i) an extension step (typical reaction volume 200µL) wherein a primer ( or initiator) is incubated for 3 min at 37°C with 1 mM of a 3'-O-NH2-dXTP and a defined concentration of TdT variant in the activity buffer (X can be A, T, C or G depending on the sequence to synthesized; for example, to synthesize AACTA, X=A for the first cycle, X=A for the second cycle, X=C for the third cycle, and so on); and (ii) a deprotection step (typical reaction volume 200µL) to deblock the extended primer to remove the 3'-protection group and allow a second extension step. The deblocking reaction is performed in 700 mM NaNO2, 1 M Sodium Acetate, pH=5.5 for 3 min. Volume of reaction is typically 200µL (e.g. see Benner, U.S. patent 7544794). The extended primer is linked to a solid support that allows removal of the reaction buffer between each step. As noted above, the reactions were run for 3 min at 37°C after which the extended primers were (i) cleaved from their supports, (ii) separated by electrophoresis and (iii) the fluorescent intensity of the bands was measured to evaluate the efficiency of synthesis for each variant. The results are shown in Table 7 where the entries are relative values that reflect the relative extension rates among the variants.

Table 7: Synthesis Efficiency of Various TdT Variants

| | 3' Sequences of Test Initiators | | | | |
|---|---|---|---|---|---|
| Variant | -AACTA | -AAGCT | -CCCCA | -GGCAT | -GGCTG |
| DS1017 (M27) SEQ ID NO: 32 | 75 | 96 | 26 | 96 | 56 |
| Lib34-20 | 72 | 96 | 18 | 96 | 57 |
| DS1002 (M44-1) | 74 | 97 | 50 | 97 | 36 |
| DS1002 (M44-2) | 78 | 95 | 50 | 96 | 35 |
| DS1004 (M45) | 84 | 97 | 59 | 95 | 58 |
| DS1006 (M46) | 93 | 96 | 28 | 96 | 71 |
| DS1007 (M47) | 92 | 97 | 30 | 96 | 78 |
| Lib34-16 | 91 | 94 | 7 | 0 | 75 |

An inspection of the sequences of the variants that exceed extension activity of M27 for one or more of the test initiators showed that each one at least possessed a substitution at Q326 and

usually shared one or more additional substitutions from the set K265, H337, W377, A17 and G57, with respect to SEQ ID NO: 3.

**[0066]** Elongation efficiency of variants M27, M44 and M45 were also measured under different reaction temperatures. In these experiments only a single elongation step was performed. The Reaction conditions were as follows: primer (i.e. initiator) concentration of 0.5 μM, a single 3'-ONH2-dTTP at 125 μM, and TdT reaction buffer (for example, available from New England Biolabs) supplemented with CoCl2. Separate reactions were run using different primers (not attached to supports) having different compositions and lengths in the range of from 20 to 25 nucleotides. Primers were labeled with an ATTO fluorescent dye on the 5' extremity. The reactions were run for 10 min separately at 37°C, 45°C, 50°C and 55°C. The results are shown in Table 8 where the values in the Table are relative magnitudes.

Table 8: Optimal Extension Reaction Temperatures For Various TdT Variants

| | Incorporation on Mixed Nucleotide Primer | | | | | Incorporation on PolyC Primer | | | |
|---|---|---|---|---|---|---|---|---|---|
| Temperature | 37°C | 45°C | 50°C | 55°C | | 37°C | 45°C | 50°C | 55°C |
| DS1017 (M27) SEQ ID NO: 32 | 63 | 100 | 77 | 24 | | 71 | 97 | 100 | 44 |
| DS1002 (M44) | 95 | 131 | 137 | 120 | | 78 | 104 | 127 | 128 |
| DS1004 (M45) | 62 | 101 | 105 | 82 | | 72 | 101 | 122 | 116 |

The data show that M44 and M45 exhibit equivalently higher yields than M27 at temperatures of 50°C and above, which is evidence of greater temperature stability.

Example 3: TdT Variants of Various Species

In this example, non-mouse TdT variants were constructed from publicly available genes and were tested to determine their ability to incorporate 3'-O-amino-dNTPs into a test polynucleotide (p877) in hairpin completion assays as described above. The TdT variants are identified in Table 9 along with their incorporation capacity as compared to mouse TdT variant, M27.

Table 9: Characteristics of Non-Mouse TdT Variants

| Species | Accession Number Of sources species | Percent Identity to M27 | Percent Activity Relative to M27* | SEQ ID NO |
|---|---|---|---|---|
| Bovine | NP_803461.1 | 82 | 484 | 27 |
| Related to Latmeria | XP_005999893.1 | 80 | 31 | 28 |
| Puma | XP_026918530.1 | 81 | 54 | 29 |
| N139 reptilian | XP_016851390.1 | 68 | 131 | 30 |
| Shrew | XP_006880141.1 | 82 | 57 | 31 |
| Mouse M27 | | 100 | -- | 32 |

* Hairpin completion assay

Definitions

**[0067]** Amino acids are represented by either their one-letter or three-letters code according to the following nomenclature: A: alanine (Ala); C: cysteine (Cys); D: aspartic acid (Asp); E: glutamic acid (Glu); F: phenylalanine (Phe); G: glycine (Gly); H: histidine (His); I: isoleucine (Ile); K: lysine (Lys); L: leucine (Leu); M: methionine (Met); N: asparagine (Asn); P: proline (Pro); Q: glutamine (Gln); R: arginine (Arg); S: serine (Ser); T: threonine (Thr); V: valine (Val); W: tryptophan (Trp ) and Y: tyrosine (Tyr).

**[0068]** "Functionally equivalent" in reference to a substituted residue means the substituted residue of a variant TdT has an identical functional role as a residue in a sequence of another TdT having a sequence homologous to SEQ ID NO: 1. Functionally equivalent residues may be identified by using sequence alignments, for example, using the Mutalin line alignment software (http://multalin.toulouse.inra.fr/multalin/multalin.html; 1988, Nucl. Acids Res., 16 (22), 25 10881-10890). After alignment, the functionally equivalent residues are at homologous positions on the different sequences considered. Sequence alignments and identification of functionally equivalent residues may be determined between any TdT and their natural variants, including inter-species.

**[0069]** "Isolated" in reference to protein means such a compound which has been identified and separated and/or recovered from a component of its natural environment or from a heterogeneous reaction mixture. Contaminant components of a natural environment or reaction mixture are materials which would interfere with a protein's function, and may

include enzymes, hormones, and other proteinaceous or nonproteinaceous solutes. In some embodiments, a protein of the invention is purified (1) to greater than 95% by weight of protein as determined by the Lowry method, and most preferably more than 99% by weight, (2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (3) to homogeneity by SDS-PAGE under reducing or nonreducing conditions using Coomassie blue or, preferably, silver stain. When manufactured by recombinant methodologies, an isolated protein of the invention may include the protein of the invention in situ within recombinant cells since at least one component of the protein's natural environment will not be present. Ordinarily, an isolated protein of the invention is prepared by at least one purification step.

**[0070]** "Kit" refers to any delivery system for delivering materials or reagents for carrying out a method of the invention. In the context of reaction assays, such delivery systems include systems and/or compounds (such as dilutants, surfactants, carriers, or the like) that allow for the storage, transport, or delivery of reaction reagents (e.g., one or more TdT variants, reaction buffers, 3'-O-protected-dNTPs, deprotection reagents, solid supports with initiators attached , etc. in the appropriate containers) and/or supporting materials (e.g., buffers, written instructions for performing the assay etc.) from one location to another. For example, kits include one or more enclosures (e.g., boxes) containing the relevant reaction reagents and/or supporting materials. Such contents may be delivered to the intended recipient together or separately. For example, a first container may contain one or more TdT variants for use in a synthesis method, while a second or additional containers may contain deprotection agents, solid supports with initiators, 3'-O-protected dNTPs, or the like.

**[0071]** "Mutant" or "variant," which are used interchangeably, refer to polypeptides derived from a natural or reference TdT polypeptide described herein, and comprising a modification or an alteration, i.e., a substitution, insertion, and/or deletion, at one or more positions. Variants may be obtained by various techniques well known in the art. In particular, examples of techniques for altering the DNA sequence encoding the wild-type protein, include, but are not limited to, site-directed mutagenesis, random mutagenesis, sequence shuffling and synthetic oligonucleotide construction. Mutagenesis activities consist in deleting, inserting or substituting one or several amino-acids in the sequence of a protein or in the case of the invention of a polymerase. The following terminology is used to designate a

substitution: L238A denotes that amino acid residue (Leucine, L) at position 238 of a reference, or wild type, sequence is changed to an Alanine (A). A132V/I/M denotes that amino acid residue (Alanine, A) at position 132 of the parent sequence is substituted by one of the following amino acids: Valine (V), Isoleucine (I), or Methionine (M). The substitution can be a conservative or non-conservative substitution. Examples of conservative substitutions are within the groups of basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine, asparagine and threonine), hydrophobic amino acids (methionine, leucine, isoleucine, cysteine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine and serine).

**[0072]** "Sequence identity" refers to the number (or fraction, usually expressed as a percentage) of matches (e.g., identical amino acid residues) between two sequences, such as two polypeptide sequences or two polynucleotide sequences. The sequence identity is determined by comparing the sequences when aligned so as to maximize overlap and identity while minimizing sequence gaps. In particular, sequence identity may be determined using any of a number of mathematical global or local alignment algorithms, depending on the length of the two sequences. Sequences of similar lengths are preferably aligned using a global alignment algorithm (e.g. Needleman and Wunsch algorithm; Needleman and Wunsch, 1970) which aligns the sequences optimally over the entire length, while sequences of substantially different lengths are preferably aligned using a local alignment algorithm (e.g. Smith and Waterman algorithm (Smith and Waterman, 1981) or Altschul algorithm (Altschul et al., 1997; Altschul et al., 2005)). Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software available on internet web sites such as http://blast.ncbi.nlm.nih.gov/ or ttp://www.ebi.ac.uk/Tools/emboss/. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithm needed to achieve maximal alignment over the full length of the sequences being compared. For purposes herein, % amino acid sequence identity values refer to values generated using the pair wise sequence alignment program EMBOSS Needle, that creates an optimal global alignment of two sequences using the Needleman-Wunsch algorithm, wherein all search

parameters are set to default values, i.e. Scoring matrix = BLOSUM62, Gap open = 10, Gap extend = 0.5, End gap penalty = false, End gap open = 10 and End gap extend = 0.5.

**[0073]** "Polynucleotide" or "oligonucleotide" are used interchangeably and each mean a linear polymer of nucleotide monomers or analogs thereof. Monomers making up polynucleotides and oligonucleotides are capable of specifically binding to a natural polynucleotide by way of a regular pattern of monomer-to-monomer interactions, such as Watson-Crick type of base pairing, base stacking, Hoogsteen or reverse Hoogsteen types of base pairing, or the like. Such monomers and their internucleosidic linkages may be naturally occurring or may be analogs thereof, e.g. naturally occurring or non-naturally occurring analogs. Non-naturally occurring analogs may include PNAs, phosphorothioate internucleosidic linkages, bases containing linking groups permitting the attachment of labels, such as fluorophores, or haptens, and the like. Whenever the use of an oligonucleotide or polynucleotide requires enzymatic processing, such as extension by a polymerase, ligation by a ligase, or the like, one of ordinary skill would understand that oligonucleotides or polynucleotides in those instances would not contain certain analogs of internucleosidic linkages, sugar moieties, or bases at any or some positions. Polynucleotides typically range in size from a few monomeric units, e.g. 5-40, when they are usually referred to as "oligonucleotides," to several thousand monomeric units. Whenever a polynucleotide or oligonucleotide is represented by a sequence of letters (upper or lower case), such as "ATGCCTG," it will be understood that the nucleotides are in 5'→3' order from left to right and that "A" denotes deoxyadenosine, "C" denotes deoxycytidine, "G" denotes deoxyguanosine, and "T" denotes thymidine, "I" denotes deoxyinosine, "U" denotes uridine, unless otherwise indicated or obvious from context. Unless otherwise noted the terminology and atom numbering conventions will follow those disclosed in Strachan and Read, Human Molecular Genetics 2 (Wiley-Liss, New York, 1999). Usually polynucleotides comprise the four natural nucleosides (e.g. deoxyadenosine, deoxycytidine, deoxyguanosine, deoxythymidine for DNA or their ribose counterparts for RNA) linked by phosphodiester linkages; however, they may also comprise non-natural nucleotide analogs, e.g. including modified bases, sugars, or internucleosidic linkages. It is clear to those skilled in the art that where an enzyme has specific oligonucleotide or polynucleotide substrate requirements for activity, e.g. single stranded DNA, RNA/DNA duplex, or the like, then selection of

appropriate composition for the oligonucleotide or polynucleotide substrates is well within the knowledge of one of ordinary skill, especially with guidance from treatises, such as Sambrook et al, Molecular Cloning, Second Edition (Cold Spring Harbor Laboratory, New York, 1989), and like references. Likewise, the oligonucleotide and polynucleotide may refer to either a single stranded form or a double stranded form (i.e. duplexes of an oligonucleotide or polynucleotide and its respective complement). It will be clear to one of ordinary skill which form or whether both forms are intended from the context of the terms usage.

**[0074]** "Primer" means an oligonucleotide, either natural or synthetic that is capable, upon forming a duplex with a polynucleotide template, of acting as a point of initiation of nucleic acid synthesis and being extended from its 3' end along the template so that an extended duplex is formed. Extension of a primer is usually carried out with a nucleic acid polymerase, such as a DNA or RNA polymerase. The sequence of nucleotides added in the extension process is determined by the sequence of the template polynucleotide. Usually primers are extended by a DNA polymerase. Primers usually have a length in the range of from 14 to 40 nucleotides, or in the range of from 18 to 36 nucleotides. Primers are employed in a variety of nucleic amplification reactions, for example, linear amplification reactions using a single primer, or polymerase chain reactions, employing two or more primers. Guidance for selecting the lengths and sequences of primers for particular applications is well known to those of ordinary skill in the art, as evidenced by the following references that are incorporated by reference: Dieffenbach, editor, PCR Primer: A Laboratory Manual, 2nd Edition (Cold Spring Harbor Press, New York, 2003).

**[0075]** A "substitution" means that an amino acid residue is replaced by another amino acid residue. Preferably, the term "*substitution*" refers to the replacement of an amino acid residue by another selected from the naturally-occurring standard 20 amino acid residues, rare naturally occurring amino acid residues (e.g. hydroxyproline, hydroxylysine, allohydroxylysine, 6-N-methylysine, N-ethylglycine, N-methylglycine, N-ethylasparagine, allo-isoleucine, N-methylisoleucine, N-methylvaline, pyroglutamine, aminobutyric acid, ornithine, norleucine, norvaline), and non-naturally occurring amino acid residue, often made synthetically, (e.g. cyclohexyl-alanine). Preferably, the term "*substitution*" refers to the replacement of an amino acid residue by another selected from the naturally-occurring standard 20 amino acid residues. The sign "+" indicates a combination of substitutions.

**[0076]** The amino acids are herein represented by their one-letter or three-letters code according to the following nomenclature: A: alanine (Ala); C: cysteine (Cys); D: aspartic acid (Asp); E: glutamic acid (Glu); F: phenylalanine (Phe); G: glycine (Gly); H: histidine (His); I: isoleucine (Ile); K: lysine (Lys); L: leucine (Leu); M: methionine (Met); N: asparagine (Asn); P: proline (Pro); Q: glutamine (Gln); R: arginine (Arg); S: serine (Ser); T: threonine (Thr); V: valine (Val); W: tryptophan (Trp ) and Y: tyrosine (Tyr).

**[0077]** In the present document, the following terminology is used to designate a substitution: L238A denotes that amino acid residue (Leucine, L) at position 238 of the parent sequence is changed to an Alanine (A). A132V/I/M denotes that amino acid residue (Alanine, A) at position 132 of the parent sequence is substituted by one of the following amino acids: Valine (V), Isoleucine (I), or Methionine (M). The substitution can be a conservative or non-conservative substitution. Examples of conservative substitutions are within the groups of basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine, asparagine and threonine), hydrophobic amino acids (methionine, leucine, isoleucine, cysteine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine and serine).

**[0078]** This disclosure is not intended to be limited to the scope of the particular forms set forth, but is intended to cover alternatives, modifications, and equivalents of the variations described herein. Further, the scope of the disclosure fully encompasses other variations that may become obvious to those skilled in the art in view of this disclosure. The scope of the present invention is limited only by the appended claims.

# CLAIMS

1. A terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence at least ninety percent identical to an amino acid sequence selected from SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 24, 25 or 26, with a substitution at position Q326 with respect to SEQ ID NOs 3, 8, 10 and 14, or at position Q325 with respect to SEQ ID NO: 4 and 5, or at position Q332 with respect to SEQ ID NO: 7 and 9, or at position Q329 with respect to SEQ ID NO: 13 and 16, or at position Q321 with respect to SEQ ID NO: 6, or at position Q327 with respect to SEQ ID NO: 11, or at position Q324 with respect to SEQ ID NO: 12, or at position Q339 with respect to SEQ ID NO: 15, or at position Q309 with respect to SEQ ID NO: 24 and 26, or at position Q307 with respect to SEQ ID NO: 25, wherein the TdT variant (i) is capable of synthesizing a nucleic acid fragment without a template and (ii) is capable of incorporating a 3'-O-modified nucleotide onto a nucleic acid fragment.

2. The TdT variant of claim 1 further comprising one or more of:
(i) a substitution at position K265 with respect to SEQ ID NOs 3, 7, 8, 9, 10, 13 and 14, or at position K263 with respect to SEQ ID NOs 6 and 12, or at position K264 with respect to SEQ ID NO 5, or at position K266 with respect to SEQ ID NO 11, or at position K268 with respect to SEQ ID NO 16, or at position K272 with respect to SEQ ID NO 15;
(ii) a substitution at position H337 with respect to SEQ ID NOs 3, 8, 10 and 14, or at position H336 with respect to SEQ ID NOs 4 and 5, or at position H343 with respect to SEQ ID NOs 7 and 9, or at position H340 with respect to SEQ ID NOs 13 and 16, or at position H332 with respect to SEQ ID NO 6, or at position H338 with respect to SEQ ID NO 11, or at position H335 with respect to SEQ ID NO 12, or at position H350 with respect to SEQ ID NO 15;
(iii) a substitution at position W377 with respect to SEQ ID NOs 3, 8, 10 and 14, or at position W376 with respect to SEQ ID NOs 4 and 5, or at position W372 with respect to SEQ ID NO 6, or at position W380 with respect to SEQ ID NO 13, or at position W372 with respect to SEQ ID NO 6, or at position W383 with respect to SEQ ID NO 9, or at position W378 with respect to SEQ ID NO 11, or at position W375 with respect to SEQ ID NO 12, or at position W380 with respect to SEQ ID NO 13, or at position W390 with respect to SEQ ID NO 15;

(iv) a substitution at position A17 with respect to SEQ ID NOs 3, 4, 5, 6, 8, 10, 13, 14 and 15, or at position A18 with respect to SEQ ID NO: 11 and 16;

(v) a substitution at position G57 with respect to SEQ ID NOs 3 and 15; or

(vi) a substitution at position Q261 with respect to SEQ ID NO: 3, or at position Q262 with respect to SEQ ID NO: 6, or at position Q264 with respect to SEQ ID NO 4.

3. The TdT variant of claims 1 or 2 wherein:

(i) said substitution at position Q326 with respect to SEQ ID NOs 3, 8, 10 and 14, or at position Q325 with respect to SEQ ID NO: 4 and 5, or at position Q332 with respect to SEQ ID NO: 7 and 9, or at position Q329 with respect to SEQ ID NO: 13 and 16, or at position Q321 with respect to SEQ ID NO: 6, or at position Q327 with respect to SEQ ID NO: 11, or at position Q324 with respect to SEQ ID NO: 12, or at position Q339 with respect to SEQ ID NO: 15 is selected from the group consisting of T, F, L, M, I, V and Y;

(ii) said substitution at position K265 with respect to SEQ ID NOs 3, 7, 8, 9, 10, 13 and 14, or at position K263 with respect to SEQ ID NOs 6 and 12, or at position K264 with respect to SEQ ID NO 5, or at position K266 with respect to SEQ ID NO 11, or at position K268 with respect to SEQ ID NO 16, or at position K272 with respect to SEQ ID NO 15 is selected from the group consisting of E, T, A and R;

(iii) said substitution of at position H337 with respect to SEQ ID NOs 3, 8, 10 and 14, or at position H336 with respect to SEQ ID NOs 4 and 5, or at position H343 with respect to SEQ ID NOs 7 and 9, or at position H340 with respect to SEQ ID NOs 13 and 16, or at position H332 with respect to SEQ ID NO 6, or at position H338 with respect to SEQ ID NO 11, or at position H335 with respect to SEQ ID NO 12, or at position H350 with respect to SEQ ID NO 15 is selected from the group consisting of Y, F, N and D;

(iv) said substitution at position W377 with respect to SEQ ID NOs 3, 8, 10 and 14, or at position W376 with respect to SEQ ID NOs 4 and 5, or at position W372 with respect to SEQ ID NO 6, or at position W380 with respect to SEQ ID NO 13, or at position W383 with respect to SEQ ID NO 9, or at position W378 with respect to SEQ ID NO 11, or at position W375 with respect to SEQ ID NO 12, or at position W380 with respect to SEQ ID NO 13, or at position W390 with respect to SEQ ID NO 15 is R;

(v) said substitution at position A17 with respect to SEQ ID NOs 3, 4, 5, 6, 8, 10, 13, 14 and 15, or at position A18 with respect to SEQ ID NO: 11 and 16 is V, I or L;

(vi) said substitution at position G57 with respect to SEQ ID NOs 3 and 15, or at position G58 with respect to SEQ ID NO: 9 and 11 is E; and

(vii) said substitution at position Q261 with respect to SEQ ID NO: 3, or at position Q262 with respect to SEQ ID NO: 6, or at position Q264 with respect to SEQ ID NO 4 is R.

4. The TdT variant of claims 1, 2 or 3 further comprising one or more of:

(i) a substitution at position L52 with respect to SEQ ID NO: 3;

(ii) a substitution at position M63 with respect to SEQ ID NOs: 3, 4, 5, 7, 8, 13 and 15; or at position M73 with respect to SEQ ID NO: 10; or at position M64 with respect to SEQ ID NO: 11; or at position M61 with respect to SEQ ID NO: 12; or at position M66 with respect to SEQ ID NO: 16;

(iii) a substitution at position A108 with respect to SEQ ID NOs: 3 and 12; or at position A110 with respect to SEQ ID NOs: 4, 5, 7, 8, 10, 13, 14 and 15; or at position A107 with respect to SEQ ID NO: 6; or at position A111 with respect to SEQ ID NO: 9; or at position A113 with respect to SEQ ID NO: 16;

(iv) a substitution at position L131 with respect to SEQ ID NOs: 3, 4, 5, 7, 8, 10, 13, 14 and 15; or at position L132 with respect to SEQ ID NOs: 9 and 11; or at position L129 with respect to SEQ ID NO: 12; or at position L113 with respect to SEQ ID NO: 16;

(v) a substitution at position C173 with respect to SEQ ID NOs: 3, 4, 5, 7, 8, 10, 13 and 14; or at position C172 with respect to SEQ ID NO: 6; or at position C174 with respect to SEQ ID NOs: 9 and 11; or at position C171 with respect to SEQ ID NO: 12; or at position C176 with respect to SEQ ID NO: 16; or at position C182 with respect to SEQ ID NO: 15;

(vi) a substitution at position R207 with respect to SEQ ID NOs: 3, 4, 5, 7, 8, 10, 13 and 14; or at position R206 with respect to SEQ ID NO: 6; or at position R208 with respect to SEQ ID NOs: 9 or 11; or at position R205 with respect to SEQ ID NO: 12; or at position R216 with respect to SEQ ID NO: 216; or at position R210 with respect to SEQ ID NO: 16;

(vii) a substitution at position G284 with respect to SEQ ID NO: 3 or 12;

(viii) a substitution at position E289 with respect to SEQ ID NO: 3; or at position E294 with respect to SEQ ID NO: 5, or at position E290 with respect to SEQ ID NO: 6; or at position

E292 with respect to SEQ ID NO: 7; or at position E295 with respect to SEQ ID NOs: 10 and 14; or at position E298 with respect to SEQ ID NOs: 13 and 16;

(ix) a substitution at position R325 with respect to SEQ ID NO: 3, 10 and 14; or at position R324 with respect to SEQ ID NOs 4 and 5; or at position R320 with respect to SEQ ID NO: 6; or at position R331 with respect to SEQ ID NOs: 7 and 9; or at position R323 with respect to SEQ ID NO: 12; or at position R328 with respect to SEQ ID NOs: 13 and 16; or at position R338 with respect to SEQ ID NO: 15;

(x) a substitution at position E328 with respect to SEQ ID NOs: 3, 8, 10 and 14; or at position E327 with respect to SEQ ID NOs: 4 and 5; or at position E334 with respect to SEQ ID NOs: 7 and 9; or at position E329 with respect to SEQ ID NO: 11; or at position E326 with respect to SEQ ID NO: 12; or at position E331 with respect to SEQ ID NOs: 13 and 16; or

(xi) a substitution at position R351 with respect to SEQ ID NO: 3; or at position R353 with respect to SEQ ID NOs: 4 and 5; or at position R354 with respect to SEQ ID NO: 8 and 14; or at position R355 with respect to SEQ ID NO: 11; or at position R352 with respect to position 13.

5. The TdT variant of any one of the previous claims further comprising one or more of:

(i) a substitution Q37E with respect to SEQ ID NOs: 3, 5, 7, 9 and 15;

(ii) a substitution D41R with respect to SEQ ID NOs: 3, 4, 5, 7, 8, 10, 13, 14 and 15, or D40R with respect to SEQ ID NO: 6, or D42R with respect to SEQ ID NO: 9 and 11, or D44R with respect to SEQ ID NO: 16, or D39R with respect to SEQ ID NO:12;

(iii) a substitution C59R with respect to SEQ ID NOs: 3, 5, 7, 13 and 15, or C58R with respect to SEQ ID NO: 6, or C60R with respect to SEQ ID NO: 10, or C62R with respect to SEQ ID NO:16;

(iv) a substitution L60D with respect to SEQ ID NOs: 3, 7, 10 and 15, or L59D with respect to SEQ ID NO:6;

(v) a substitution S94R with respect to SEQ ID NOs: 3, 12 and 15;

(vi) a substitution G98E with respect to SEQ ID NO: 3;

(vii) a substitution S119A with respect to SEQ ID NOs: 3, 4, 5, 7, 8, 10, 13, 14 and 14, or S118A with respect to SEQ ID NO: 6, or S120A with respect to SEQ ID NOs 9 and 11, or S122A with respect to SEQ ID NO: 16;

(viii) a substitution S146E with respect to SEQ ID NOs: 3, 4, 5, 8, 10 and 14, or S147E with respect to SEQ ID NO: 11, or S144E with respect to SEQ ID NO: 12;
(ix) a substitution Q149R with respect to SEQ ID NO: 3;
(x) a substitution F193Y with respect to SEQ ID NOs: 3, 4, 5, 7, 8, 10, 13 and 14, or F192Y with respect to SEQ ID NO: 6, or F194Y with respect to SEQ ID NO: 11, or F191Y with respect to SEQ ID NO: 12, or F196Y with respect to SEQ ID NO: 16;
(xi) a substitution V199M with respect to SEQ ID NO: 3, 4, 5, 7, 8, 10, 13, 14 and 15, or V198M with respect to SEQ ID NO: 6, or V200M with respect to SEQ ID NOs: 9 and 11; or
(xii) a substitution M201V with respect to SEQ ID NOs: 3, 4, 5, 8, 10 and 13, or M202V with respect to SEQ ID NO: 11, or M199V with respect to SEQ ID NO: 12.

6. The TdT variant of any one of the previous claims comprising the following combination of substitutions with respect to SEQ ID NO: 3:
(i) A17V + Q37E + D41R + L52F + G57E + M63R + S94R + G98E + A108V + S119A + L131R + S146E + Q149R + C173G + R207L + K265T + G284P + E289V + R325P + Q326F + E328N + H337D + R351K + W377R;
(ii) A17V + Q37E + D41R + L52F + G57E + M63R + S94R + G98E + A108V + S146E + Q149R + C173G + F193Y + V199M + M201V + R207L + K265T + G284P + E289V + R325A + Q326F + E328N + R351K;
(iii) L52F + A108V + R351K + A17V + Q37E + D41R + G57E+ C59R + L60D + M63R + S94R + G98E + S119A + L131R + S146E + Q149R + C173G + R207L + K265T + G284P + E289V + R325A + Q326F + E328N;
(iv) L52F + A108V + R351K + A17V + Q37E + D41R + G57E + C59R + L60D + M63R + S94R + G98E + K118Q + S119A + L131R + S146E + Q149R + C173G + R207L + K265T + G284P + E289V + R325A + Q326F + E328N + W377R;
(v) A17V + Q37E + D41R + L52F + G57E + C59R + L60D + M63R + S94R + G98E + A108V + S119A + L131R + S146E + Q149R + C173G + R207L + F259S + Q261L, G284P + E289V + R325A + Q326F + E328N + R351K + W377R;
(vi) A17V + Q37E + D41R + L52F + G57E + C59R + L60D + M63R + S94R + G98E + A108V + S119A + L131R + S146E + Q149R + C173G + R207L + E257D + F259S + K260R + Q261L + G284P + E289V + R325A + Q326F + E328N + R351K + W377R;

(vii) A17V + D41R + L53F + G57E + C59R + L60D + M63R + S94R + G98E + K118Q + S119A + L131R + S146E + Q149R + C173G + R207L + K265T + G284P + E289V + R325A + Q326F + R351K + W377R;

(viii) A17V + D41R + L52F + G57E + C59R + L60D + M63R + S94R + G98E + A108V + S119A + L131R + S146E + Q149R + R207L + K265T + G284P + E289V + R325A + Q326F + R351K;

(ix) A17V + L52F + M63R + A108V + C173G + R207L + K265T + G284P + E289V + R325P + E328N + R351K;

(x) A17V + D41R + L52F + G57E + M63R + S94R + G98E + A108V + S146E + Q149R + C173G + M184T + R207L + K209H + G284L + E289A + R325V + E328K + R351K;

(xi) A17V + L52F + G57E + M63R + A108V + C173G + R207L + K265T + G284P + E289V + R325P + E328N + R351K;

(xii) A17V + L32T + Q37R + D41R + L52F + G57E + C59R + L60D + M63R + S67A + S94R + G98E + A108V + S119A +L131R + S146E + Q149R + V171A + S172E + C173R + V182I + S183E + R207L + K209H + M210K + T211I + E223G + A224P + E228D + Q261L + G284P + E289V + R325A + Q326F + E328N + R351K + D372E; or

(xiii) comprising SEQ ID NOs: 27, 28, 29, 30 or 31.

7. The TdT variant of any one of claims 1 to 6having amino acid sequence at least ninety percent identical to SEQ ID NO: 4 and further having substitutions at positions M63, R207, R324 and E327.

8. The TdT variant of claim 7, comprising the combination of substitutions M63R+ R207L+E227N.

9 The TdT variant of any one of claims 1 to 6, having amino acid sequence at least ninety percent identical to SEQ ID NO: 24 and further having substitutions at positions M47, R190, R308 and E311.

10. The TdT variant of claim 9, comprising the combination of substitutions M47R+ R190L+E311N.

11. The TdT variant of any one of claims 1 to 6 having amino acid sequence at least ninety percent identical to SEQ ID NO: 26 and further having substitutions at positions M46, R190 and E311.

12. The TdT variant of claim 11, comprising the combination of substitutions M46R+ R190 L+ E311N.

13. The TdT variant of any one of claims 1 to 6, having amino acid sequence at least ninety percent identical to SEQ ID NO: 25 and further having substitutions at positions R184, R306 and E309.

14. The TdT variant of claim 13, comprising the combination of substitutions R184L+R306A+E309N.

15. A method of synthesizing a polynucleotide having a predetermined sequence, the method comprising the steps of:

a) providing an initiator having a 3'-terminal nucleotide having a free 3'-hydroxyl;

b) repeating cycles of (i) contacting under elongation conditions the initiator or elongated fragments having free 3'-O-hydroxyls with a 3'-O-blocked nucleoside triphosphate and a TdT variant according to any one of claims 1 to 14, so that the initiator or elongated fragments are elongated by incorporation of a 3'-O-blocked nucleoside triphosphate to form 3'-O-blocked elongated fragments, and (ii) deblocking the elongated fragments to form elongated fragments having free 3'-hydroxyls, until the polynucleotide is formed.

16. The method of claim 15, wherein said 3'-O-blocked nucleoside triphosphate is a 3'-O-NH2-nucleoside triphosphate, a 3'-O-azidomethyl-nucleoside triphosphate, a 3'-O-allyl-nucleoside triphosphate, or a 3'-O-(2-nitrobenzyl)-nucleoside triphosphate.

17. A kit for performing a nucleotide incorporation reaction comprising:

a) a variant of TdT according to any one of claims 1 to 14,

b) one or more nucleotides, preferably one or more 3'O-modified nucleotides, and

c) optionally at least one nucleic acid primer.

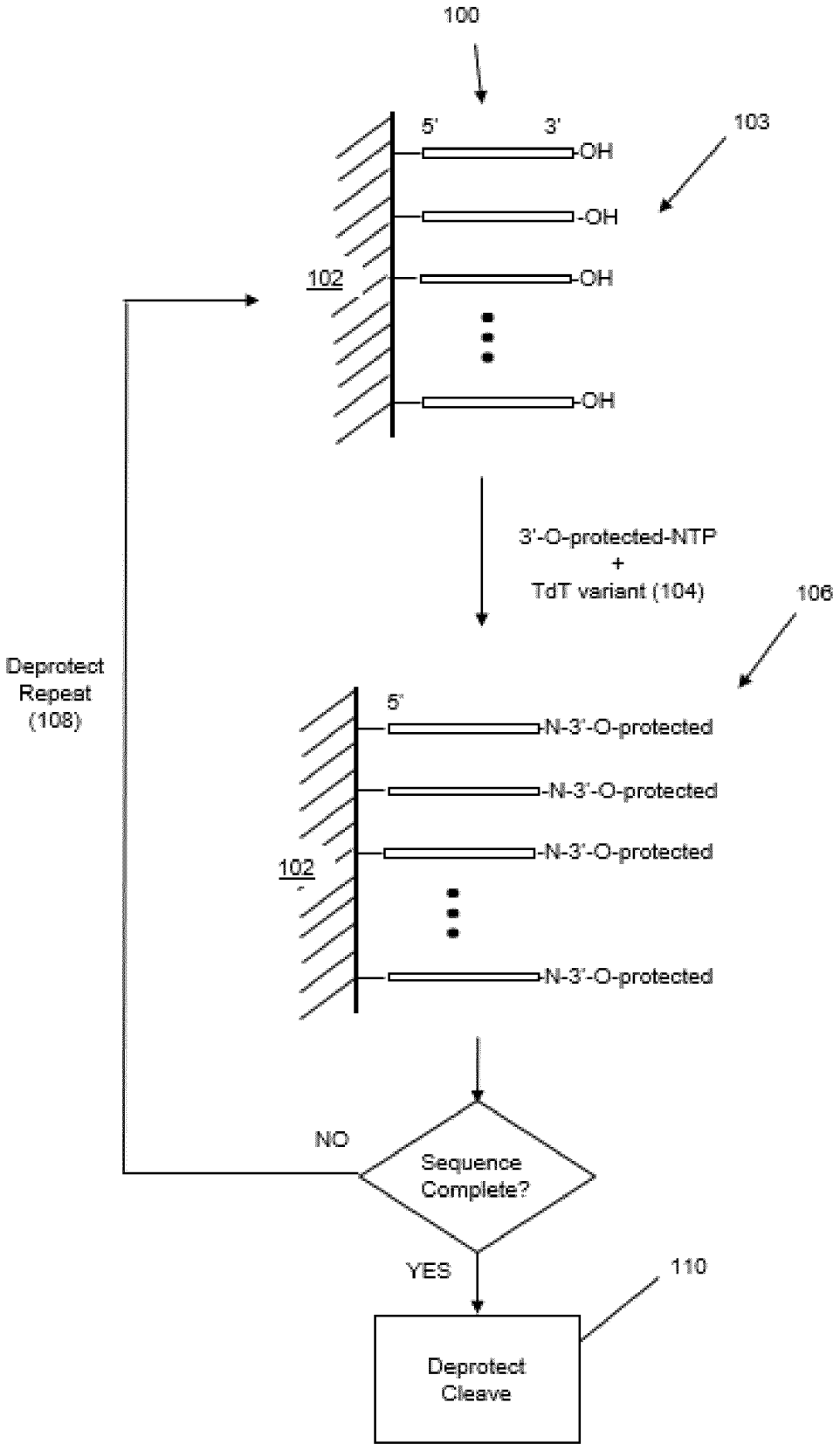
100
5'
3'
-OH
-OH
-OH
-OH
103
102
3'-O-protected-NTP
+
TdT variant (104)
106
Deprotect
Repeat
(108)
5'
-N-3'-O-protected
-N-3'-O-protected
-N-3'-O-protected
-N-3'-O-protected
102
NO
Sequence
Complete?
YES
110
Deprotect
Cleave

FIGURE 1

Table 1

| SEQ ID NO | Supplemental Substitutions | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | L181F/A/M/Y | M192R/Q | A237V | L260P/R/K | C302G/R | R336L/N | G413L/S | E418A/V | R454P/N/A/V | E457N/L/T/S/K | R480K/E/D |
| 3 | L52F/A/M/Y | M63R/Q | A108V | L131P/R/K | C173G/R | R207L/N | G284L/S | E289A/V | R325P/N/A/V | E328N/L/T/S/K | R351K/E/D |
| 4 | --- | M63R/Q | A110V | L131P/R/K | C173G/R | R207L/N | --- | --- | R324P/N/A/V | E327N/L/T/S/K | R353K/E/D |
| 5 | --- | M63R/Q | A110V | L131P/R/K | C173G/R | R207L/N | --- | E294A/V | R324P/N/A/V | E327N/L/T/S/K | R353K/E/D |
| 6 | --- | --- | A107V | --- | C172G/R | R206L/N | --- | E290A/V | R320P/N/A/V | --- | --- |
| 7 | --- | M63R/Q | A110V | L131P/R/K | C173G/R | R207L/N | --- | E292A/V | R331P/N/A/V | E334N/L/T/S/K | --- |
| 8 | --- | M63R/Q | A110V | L131P/R/K | C173G/R | R207L/N | --- | --- | --- | E328N/L/T/S/K | R354K/E/D |
| 9 | --- | --- | A111V | L132P/R/K | C174G/R | R208L/N | --- | --- | R331P/N/A/V | E334N/L/T/S/K | --- |
| 10 | --- | M73R/Q | A110V | L131P/R/K | C173G/R | R207L/N | --- | E295A/V | R325P/N/A/V | E328N/L/T/S/K | --- |
| 11 | --- | M64R/Q | A111V | L132P/R/K | C174G/R | R208L/N | --- | --- | --- | E329N/L/T/S/K | R355K/E/D |
| 12 | --- | M61R/Q | A108V | L129P/R/K | C171G/R | R205L/N | G284L/S | E293A/V | R323P/N/A/V | E326N/L/T/S/K | R352K/E/D |
| 13 | --- | M63R/Q | A110V | L131P/R/K | C173G/R | R207L/N | --- | E298A/V | R328P/N/A/V | E331N/L/T/S/K | --- |
| 14 | --- | --- | A110V | L131P/R/K | C173G/R | R207L/N | --- | E295A/V | R325P/N/A/V | E328N/L/T/S/K | R354K/E/D |
| 15 | --- | M63R/Q | A110V | L131P/R/K | C182G/R | R216L/N | --- | --- | R338P/N/A/V | E341N/L/T/S/K | --- |
| 16 | --- | M66R/Q | A113V | L134P/R/K | C176G/R | R210L/N | --- | E298A/V | R328P/N/A/V | E331N/L/T/S/K | --- |

FIGURE 2

Table 2

| SEQ ID NO | Stability Substitutions | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Q166E | D170R | C188R | L189D | S223R | G227E | S248A | S275E | Q278R | F322Y | V328M | M330V |
| 3 | Q37E | D41R | C59R | L60D | S94R | G98E | S119A | S146E | Q149R | F193Y | V199M | M201V |
| 4 | -- | D41R | -- | -- | -- | -- | S119A | S146E | -- | F193Y | V199M | M201V |
| 5 | Q37E | D41R | C59R | -- | -- | -- | S119A | S146E | -- | F193Y | V199M | M201V |
| 6 | -- | D40R | C58R | L59D | -- | -- | S118A | -- | -- | F192Y | V198M | -- |
| 7 | Q37E | D41R | C59R | L60D | -- | -- | S119A | -- | -- | F193Y | V199M | -- |
| 8 | -- | D41R | -- | -- | -- | -- | S119A | S146E | -- | F193Y | V199M | M201V |
| 9 | Q37E | D42R | C60R | -- | -- | -- | S120A | -- | -- | -- | V200M | -- |
| 10 | -- | D41R | -- | L60D | -- | -- | S119A | S146E | -- | F193Y | V199M | M201V |
| 11 | -- | D42R | -- | -- | -- | -- | S120A | S147E | -- | F194Y | V200M | M202V |
| 12 | -- | D39R | -- | -- | S94R | -- | -- | S144E | -- | F191Y | -- | M199V |
| 13 | -- | D41R | C59R | -- | -- | -- | S119A | -- | -- | F193Y | V199M | M201V |
| 14 | -- | D41R | -- | -- | -- | -- | S119A | S146E | -- | F193Y | V199M | -- |
| 15 | Q37E | D41R | C59R | L60D | S94R | -- | S119A | -- | -- | -- | V199M | -- |
| 16 | -- | D44R | C62R | -- | -- | -- | S122A | -- | -- | F196Y | -- | -- |

FIGURE 3

# INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/081099

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/12
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, CHEM ABS Data, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | CA 3 024 184 A1 (DNA SCRIPT [FR]; PASTEUR INSTITUT [FR]; CENTRE NAT RECH SCIENT [FR]) 21 December 2017 (2017-12-21)<br>page 1, paragraphs 1, 2<br>page 3, paragraph 2 - page 6, paragraph 3<br>page 12, paragraph 3<br>examples 5, 6<br>claims 1-3, 21-24<br>-----<br>-/-- | 1-17 |

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 20 December 2019 | 21/01/2020 |

| Name and mailing address of the ISA/ | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016 | Surdej, Patrick |

1

# INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/081099

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | DELARUE M ET AL: "Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase", THE EMBO JOURNAL,, vol. 21, no. 3, 1 February 2002 (2002-02-01), pages 427-439, XP002767368, ISSN: 1460-2075, DOI: 10.1093/EMBOJ/21.3.427 the whole document<br>----- | 1-17 |

1

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2019/081099

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| CA 3024184 A1 | 21-12-2017 | AU 2017286477 A1 | 06-12-2018 |
| | | CA 3024184 A1 | 21-12-2017 |
| | | CN 109477080 A | 15-03-2019 |
| | | EP 3469075 A2 | 17-04-2019 |
| | | FR 3052462 A1 | 15-12-2017 |
| | | FR 3053699 A1 | 12-01-2018 |
| | | JP 2019517799 A | 27-06-2019 |
| | | KR 20190024883 A | 08-03-2019 |
| | | SG 11201809961T A | 28-12-2018 |
| | | WO 2017216472 A2 | 21-12-2017 |