

TERMINAL DEOXYNUCLEOTIDYL TRANSFERASE VARIANTS
AND USES THEREOF

5

BACKGROUND

[0001] Synthetic polynucleotides of predetermined sequences are central to a host of technologies, including molecular diagnostics, genomic and diagnostic sequencing, 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, however, 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 from a variety of species which display enhanced efficiency in incorporating reversibly blocked nucleoside triphosphates into a polynucleotide, and to their use in synthesizing polynucleotides of any predetermined sequence.

[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: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44 or 47, wherein

5 with respect to SEQ ID NO: 2 leucine at position 61 is substituted, cysteine at position 170 is substituted, arginine at position 204 is substituted, arginine at position 326 is substituted and glycine at position 329 is substituted;

 with respect to SEQ ID NO: 5 leucine at position 48 is substituted, cysteine at position 158 is substituted, arginine at position 192 is substituted, arginine at position 314 is substituted
10 and glycine at position 317 is substituted;

 with respect to SEQ ID NO: 8 leucine at position 61 is substituted, tyrosine at position 171 is substituted, arginine at position 205 is substituted, arginine at position 327 is substituted and glutamic acid at position 329 is substituted;

 with respect to SEQ ID NO: 11 leucine at position 61 is substituted, tyrosine at position
15 171 is substituted, arginine at position 205 is substituted, arginine at position 324 is substituted and glutamic acid at position 327 is substituted;

 with respect to SEQ ID NO: 14 glycine at position 61 is substituted, arginine at position 205 is substituted, arginine at position 327 is substituted and glutamic acid at position 330 is substituted;

20 with respect to SEQ ID NO: 17 arginine at position 61 may or may not be substituted, alanine at position 158 is substituted, arginine at position 192 is substituted, arginine at position 311 is substituted and glutamic acid at position 314 is substituted;

 with respect to SEQ ID NO: 20 arginine at position 61 may or may not be substituted, proline at position 171 may or may not be substituted, arginine at position 205 is substituted,
25 threonine at position 321 is substituted and glutamic acid at position 324 is substituted;

 with respect to SEQ ID NO: 23 arginine at position 61 may or may not be substituted, alanine at position 171 may or may not be substituted, arginine at position 205 is substituted, alanine at position 322 may or may not be substituted and glutamic acid at position 325 is substituted;

with respect to SEQ ID NO: 26 arginine at position 61 may or may not be substituted, alanine at position 171 may or may not be substituted, arginine at position 205 is substituted, arginine at position 322 is substituted and glutamic acid at position 325 is substituted;

5 with respect to SEQ ID NO: 29 methionine at position 48 is substituted, cysteine at position 158 is substituted, arginine at position 192 is substituted, arginine at position 310 is substituted and glutamic acid at position 313 is substituted;

with respect to SEQ ID NO: 32 methionine at position 61 is substituted, cysteine at position 171 is substituted, arginine at position 205 is substituted, arginine at position 323 is substituted and glutamic acid at position 326 is substituted;

10 with respect to SEQ ID NO: 35 methionine at position 61 is substituted, cysteine at position 171 is substituted, arginine at position 205 is substituted, arginine at position 323 is substituted and glutamic acid at position 326 is substituted;

with respect to SEQ ID NO: 38 methionine at position 61 is substituted, cysteine at position 171 is substituted, arginine at position 205 is substituted, arginine at position 323 is substituted and glutamic acid at position 326 is substituted;

15 with respect to SEQ ID NO: 41 methionine at position 61 is substituted, cysteine at position 171 is substituted, arginine at position 205 is substituted, arginine at position 323 is substituted and glutamic acid at position 326 is substituted;

20 with respect to SEQ ID NO: 44 methionine at position 48 is substituted, cysteine at position 158 is substituted, arginine at position 192 is substituted, arginine at position 309 is substituted and glutamic acid at position 312 is substituted; and

with respect to SEQ ID NO: 47 methionine at position 61 is substituted, cysteine at position 171 is substituted, arginine at position 205 is substituted, arginine at position 323 is substituted and glutamic acid at position 326 is substituted; and

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 substitution for leucine at position 61 in SEQ ID NO: 2, 8 or 11 or position 48 in SEQ ID NO: 5 is selected from the group consisting of R or Q. In some embodiments, the above substitution for cysteine at position 170 in SEQ ID NO: 2 or position 158 in SEQ ID NO: 5, 29 or 44 or position 171 in SEQ ID NO: 32, 35, 38, 41 or 47 is selected from the group consisting of G, R, P, A, V, S, N, Q or D; and in other embodiments, the substitution for cysteine at the foregoing positions is selected from the group consisting of G or R. In some embodiments, the above substitution for tyrosine at position 171 of SEQ ID NO: 8 or 11 is selected from the group consisting of G, R, P, A, V, S, N, Q or D; and in other embodiments, the substitution for tyrosine at the foregoing positions is selected from the group consisting of G or R. In some embodiments, the above substitution for alanine at position 158 of SEQ ID NO: 17 or position 171 of SEQ ID NO: 23 or 26 is selected from the group consisting of G, R, P, V, S, N, Q or D; and in other embodiments, the substitution for alanine at the foregoing positions is selected from the group consisting of G or R. In some embodiments, the above substitution for proline at position 171 of SEQ ID NO: 20 is selected from the group consisting of G, R, A, V, S, N, Q or D; and in other embodiments, the substitution for tyrosine at the foregoing positions is selected from the group consisting of G or R. In some embodiments, the above substitution for arginine at position 204 of SEQ ID NO: 2 or position 192 of SEQ ID NO: 5, 17, 29 or 44 or position 205 of SEQ ID NO: 8, 11, 14, 20, 23, 26, 32, 35, 38, 41 or 47 is selected from the group consisting of L or N. In some embodiments, the above substitution for arginine at position 326 of SEQ ID NO: 2 or position 314 of SEQ ID NO: 5 or position 327 of SEQ ID NO: 8 or 14 or position 324 of SEQ ID NO: 11 or position 311 of SEQ ID NO: 17 or position 321 of SEQ ID NO: 20 or position 322 of SEQ ID NO: 23 or 26 or position 310 of SEQ ID NO: 29 or position 323 of SEQ ID NO: 32, 35, 38, 41 or 47 or position 309 of SEQ ID NO: 44 is selected from the group consisting of P, N or A. In some embodiments, the above substitution for threonine at position 321 of SEQ ID NO: 20 is selected from the group consisting of P, N or A. In some embodiments, the above substitution for glycine at position 329 of SEQ ID NO: 2 or 5 is selected from the group consisting of N, L, T or S. In some embodiments, the above substitution for glutamic acid at position 330 of SEQ ID NO: 8 or 14 or position 327 of SEQ ID NO: 11 or position 311 of SEQ ID NO: 17 or position 324 of SEQ ID NO: 20 or position 325 of SEQ ID NO: 23 or 26 or

position 313 of SEQ ID NO: 29 position 326 of SEQ ID NO: 32, 35, 38, 41 or 47 or position 312 of SEQ ID NO: 44 is selected from the group consisting of N, L, T or S.

[0006] In further 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: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44 or 47, substituted as described above and further substituted as follows:

with respect to SEQ ID NO: 2 glutamine at position 327 is substituted;

with respect to SEQ ID NO: 5 glutamic acid at position 315 is substituted;

10 with respect to SEQ ID NO: 8 glutamine at position 328 is substituted;

with respect to SEQ ID NO: 11 glutamine at position 325 is substituted;

with respect to SEQ ID NO: 14 glutamine at position 328 is substituted;

with respect to SEQ ID NO: 17 glutamine at position 312 is substituted;

with respect to SEQ ID NO: 20 glutamine at position 322 is substituted;

15 with respect to SEQ ID NO: 23 glutamine at position 323 is substituted;

with respect to SEQ ID NO: 26 methionine at position 323 is substituted;

with respect to SEQ ID NO: 29 glutamine at position 311 is substituted;

with respect to SEQ ID NO: 32 glutamine at position 324 is substituted;

with respect to SEQ ID NO: 35 glutamine at position 324 is substituted;

20 with respect to SEQ ID NO: 38 glutamine at position 324 is substituted;

with respect to SEQ ID NO: 41 glutamine at position 324 is substituted;

with respect to SEQ ID NO: 44 glutamine at position 310 is substituted; and

with respect to SEQ ID NO: 47 glutamine at position 324 is substituted;

and wherein the TdT variant (i) is capable of synthesizing a nucleic acid fragment without a

25 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 foregoing substitutions of glutamine and glutamic acid are T, F, L or M.

[0007] In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence at least ninety percent identical

30

to an amino acid sequence selected from SEQ ID NO: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45 or 48, wherein:

with respect to SEQ ID NO: 3 amino acid position 61 is R, amino acid position 171 is R, amino acid position 205 is L, amino acid position 327 is A, and amino acid position 330 is N;

with respect to SEQ ID NO: 6 amino acid position 48 is R, amino acid position 158 is R, amino acid position 192 is L, amino acid position 314 is P, and amino acid position 317 is N;

with respect to SEQ ID NO: 9 amino acid position 61 is R, amino acid position 171 is R, amino acid position 205 is L, amino acid position 327 is A, and amino acid position 330 is N;

with respect to SEQ ID NO: 12 amino acid position 61 is R, amino acid position 171 is A, amino acid position 205 is L, amino acid position 324 is P, and amino acid position 327 is N;

with respect to SEQ ID NO: 15 amino acid position 61 is R, amino acid position 171 is R, amino acid position 205 is L, amino acid position 328 is P, and amino acid position 331 is N;

with respect to SEQ ID NO: 18 amino acid position 48 is R, amino acid position 158 is R, amino acid position 192 is L, amino acid position 311 is P, and amino acid position 314 is N;

with respect to SEQ ID NO: 21 amino acid position 61 is R, amino acid position 171 is P, amino acid position 205 is L, amino acid position 321 is A, and amino acid position 324 is N;

with respect to SEQ ID NO: 24 amino acid position 61 is R, amino acid position 171 is A, amino acid position 205 is L, amino acid position 322 is A, and amino acid position 325 is N;

with respect to SEQ ID NO: 27 amino acid position 61 is R, amino acid position 171 is R, amino acid position 205 is L, amino acid position 327 is A, and amino acid position 330 is N;

with respect to SEQ ID NO: 30 amino acid position 48 is R, amino acid position 158 is R, amino acid position 192 is L, amino acid position 310 is P, and amino acid position 313 is N;

5 with respect to SEQ ID NO: 33 amino acid position 61 is R, amino acid position 171 is R, amino acid position 205 is L, amino acid position 323 is P, and amino acid position 326 is N;

with respect to SEQ ID NO: 36 amino acid position 61 is R, amino acid position 171 is R, amino acid position 205 is L, amino acid position 323 is A, and amino acid position 326 is N;

10 with respect to SEQ ID NO: 39 amino acid position 61 is R, amino acid position 171 is R, amino acid position 205 is L, amino acid position 323 is P, and amino acid position 326 is N;

with respect to SEQ ID NO: 42 amino acid position 61 is R, amino acid position 171 is R, amino acid position 205 is L, amino acid position 323 is A, and amino acid position 326 is N;

15 with respect to SEQ ID NO: 45 amino acid position 48 is R, amino acid position 158 is R, amino acid position 192 is L, amino acid position 310 is A, and amino acid position 313 is N;

20 with respect to SEQ ID NO: 48 amino acid position 61 is R, amino acid position 171 is R, amino acid position 205 is L, amino acid position 323 is P, and amino acid position 326 is N;

and 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.

25 In some embodiments, in the foregoing TdT variants, with respect to SEQ ID NO: 3 amino acid position 328 is Q or is selected from the group consisting of T, F, L or M;

with respect to SEQ ID NO: 6 amino acid position 315 is Q or is selected from the group consisting of T, F, L or M;

30 with respect to SEQ ID NO: 9 amino acid position 328 is Q or is selected from the group consisting of T, F, L or M;

with respect to SEQ ID NO: 12 amino acid position 325 is Q or is selected from the group consisting of T, F, L or M;

with respect to SEQ ID NO: 15 amino acid position 329 is Q or is selected from the group consisting of T, F, L or M;

5 with respect to SEQ ID NO: 18 amino acid position 312 is Q or is selected from the group consisting of T, F, L or M;

with respect to SEQ ID NO: 21 amino acid position 322 is Q or is selected from the group consisting of T, F, L or M;

10 with respect to SEQ ID NO: 24 amino acid position 323 is Q or is selected from the group consisting of T, F, L or M;

with respect to SEQ ID NO: 27 amino acid position 328 is Q or is selected from the group consisting of T, F, L or M;

with respect to SEQ ID NO: 30 amino acid position 311 is Q or is selected from the group consisting of T, F, L or M;

15 with respect to SEQ ID NO: 33 amino acid position 324 is Q or is selected from the group consisting of T, F, L or M;

with respect to SEQ ID NO: 36 amino acid position 324 is Q or is selected from the group consisting of T, F, L or M;

20 with respect to SEQ ID NO: 39 amino acid position 324 is Q or is selected from the group consisting of T, F, L or M;

with respect to SEQ ID NO: 42 amino acid position 324 is Q or is selected from the group consisting of T, F, L or M;

with respect to SEQ ID NO: 45 amino acid position 311 is Q or is selected from the group consisting of T, F, L or M; and

25 with respect to SEQ ID NO: 48 amino acid position 324 is Q or is selected from the group consisting of T, F, L or M.

[0008] In some embodiments, the invention is directed to a terminal deoxynucleotidyl transferase (TdT) variant comprising an amino acid sequence selected from SEQ ID NO: 3, 6,
30 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45 or 48.

[0009] In some embodiments, the percent identity value with respect to each of the foregoing TdT variants 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. 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 homology would be in regard to sequences 1-24, 26-80 and 82-100. In regard to (ii) above, in some embodiments, 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, a 3'-O-nitro-nucleoside triphosphate or a 3'-O-propargyl-nucleoside triphosphate. In other embodiments, such 3'-O-modified nucleotide may comprise a 3'-O-NH₂-nucleoside triphosphate or a 3'-O-azidomethyl-nucleoside triphosphate.

[0010] 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.

[0011] 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.

[0012] 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. 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.

20

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] Fig. 1 illustrates diagrammatically the steps of a method of template-free enzymatic nucleic acid synthesis using TdT variants of the invention.

25

DETAILED DESCRIPTION OF THE INVENTION

[0014] 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.

[0015] 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.

[0016] TdT variants of the invention are derived from natural TdTs without their BRCT-like N-terminal amino acid sequences. Such truncated sequences are sometimes referred to herein as "truncated wild type" TdTs. More particularly, TdT variants of the invention are derived from the following natural TdTs:

Table 1

Designation	Species	NCBI Accession No.	SEQ ID NO
B10 trunc wt	Serinus canaria	XP_030097130.1	2
B10 variant	Serinus canaria	XP_030097130.1	3
B10 tagged variant	Serinus canaria	XP_030097130.1	4
B19 trunc wt	Neopelma chrysocephalum	XP_027527888.1	5
B19 variant	Neopelma chrysocephalum	XP_027527888.1	6
B19 tagged variant	Neopelma chrysocephalum	XP_027527888.1	7
B57 trunc wt	Alligator sinensis	XP_025069068.1	8
B57 variant	Alligator sinensis	XP_025069068.1	9
B57 tagged variant	Alligator sinensis	XP_025069068.1	10
B59 trunc wt	Xenopus laevis	AAI29712.1	11
B59 variant	Xenopus laevis	AAI29712.1	12
B59 tagged variant	Xenopus laevis	AAI29712.1	13
B63 trunc wt	Notechis scutatus	XP_026526500.1	14
B63 variant	Notechis scutatus	XP_026526500.1	15
B63 tagged variant	Notechis scutatus	XP_026526500.1	16
F15 trunc wt	Salmo truttal	XP_029608991.1	17
F15 variant	Salmo truttal	XP_029608991.1	18
F15 tagged variant	Salmo truttal	XP_029608991.1	19
F24 trunc wt	Electrophorus electricus	XP_026869907.1	20
F24 variant	Electrophorus electricus	XP_026869907.1	21

F24 tagged variant	Electrophorus electricus	XP_026869907.1	22
F32 trunc wt	Anabas testudineus	XP_026209749.1	23
F32 variant	Anabas testudineus	XP_026209749.1	24
F32 tagged variant	Anabas testudineus	XP_026209749.1	25
F57 trunc wt	Poecilia reticulata	XP_008428203.1	26
F57 variant	Poecilia reticulata	XP_008428203.1	27
F57 tagged variant	Poecilia reticulata	XP_008428203.1	28
N100 trunc wt	Rattus norvegicus	XP_001012479.1	29
N100 variant	Rattus norvegicus	XP_001012479.1	30
N100 tagged variant	Rattus norvegicus	XP_001012479.1	31
N100or trunc wt	Rattus norvegicus	XP_001012479.1	32
N100or variant	Rattus norvegicus	XP_001012479.1	33
N100or tagged variant	Rattus norvegicus	XP_001012479.1	34
N14 trunc wt	Piliocolobus tephrosceles	XP_023062037.1	35
N14 variant	Piliocolobus tephrosceles	XP_023062037.1	36
N14 tagged variant	Piliocolobus tephrosceles	XP_023062037.1	37
N27 trunc wt	Sus scrofa	XP_003133204.1	38
N27 variant	Sus scrofa	XP_003133204.1	39
N27 tagged variant	Sus scrofa	XP_003133204.1	40
N35 trunc wt	Panthera tigris altaica	XP_007080320.1	41
N35 variant	Panthera tigris altaica	XP_007080320.1	42
N35 tagged variant	Panthera tigris altaica	XP_007080320.1	43
N72 trunc wt	Bubalus bubalis	XP_006061247.1	44
N72 variant	Bubalus bubalis	XP_006061247.1	45
N72 tagged variant	Bubalus bubalis	XP_006061247.1	46
N93or trunc wt	Marmota flaviventris	XP_027786017.1	47
N93or variant	Marmota flaviventris	XP_027786017.1	48
N93or tagged variant	Marmota flaviventris	XP_027786017.1	49

[0017] 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: (3,4), (6,7), (9,10), (12,13), (15,16), (18,19), (21,22), (24,25), (27,28), (30,31), (33,34), (36,37), (39,40), (42,43), (45,46) and (48,49)). Typically, the linker moiety does not interfere with the nucleotide binding activity, or catalytic activity of the TdT variant.

[0018] 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.

[0019] 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 2A.

Table 2A: 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

[0020] In some embodiments, sets of synonymous amino acids that may be employed are set forth in Table 2B.

Table 2B: 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

5

Measurement of Nucleotide Incorporation Activity

[0021] 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

10

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.

[0022] In some embodiments, the following particular extension assay may be used to measure incorporation efficiencies of TdTs: Primer used is the following:

5'-AAAAAAAAAAAAAGGGG-3' (SEQ ID NO: 50)

The primer has also an ATTO fluorescent dye on the 5' extremity. Representative modified nucleotides used (noted as dNTP in Table 3) 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 3. After 30 min at 37°C the reaction is stopped by addition of formamide (Sigma).

Table 3: Extension Activity Assay Reagents

Reagent	Concentration	Volume
H ₂ O	-	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 CoCl₂.

[0023] 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.

[0024] 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).

[0025] Exemplary test polynucleotides for hairpin completion assays include p875 (5'-CAGTTAAAAACT) (SEQ ID NO: 51) which is completed by extending with a dGTP; p876 (5'-GAGTTAAAACT) (SEQ ID NO: 52) which is completed by extending with a dCTP; and p877 (5'-CAGCAAGGCT) (SEQ ID NO: 53) which is completed by extending with a dGTP.

5 Exemplary reaction conditions for such test polynucleotides may comprise: 2.5 - 5 μ M of test polynucleotide, 1:4000 dilution of GelRed[®] (intercalating dye from Biotium, Inc., Fremont, CA), 200mM Cacodylate KOH pH 6.8, 1mM CoCl₂, 0-20% of DMSO and 3' ONH₂ 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
10 a reaction temperature of 28-38°C, using an excitation filter set to 360nm and an emission filter set to 635nm. In some embodiments of this aspect of the invention, TdT variants may be tested for their capacity for template-free incorporation 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
15 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.

20

Template-Free Enzymatic Synthesis of Oligonucleotides

[0026] Generally, methods of template-free (or equivalently, "template-independent") enzymatic DNA synthesis comprise repeated cycles of steps, such as are illustrated in Fig. 1, in which a predetermined nucleotide is coupled to an initiator or growing chain in each cycle.

25 The general elements of template-free enzymatic synthesis is described in the following references: 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
30 Lett., 1(11): 1729-1731 (1999).

[0027] 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 (104) a 3'-O-protected-dNTP and a template-free polymerase, such as a TdT or variant thereof (e.g. Ybert et al, WO/2017/216472; Champion et al, WO2019/135007) under conditions 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 sequence is not complete, then another cycle of addition is implemented (108). If the elongated initiator polynucleotide contains a completed sequence, then the 3'-O-protection group may be removed, or deprotected, and the desired sequence may be cleaved from the original initiator polynucleotide (110). Such cleavage may be carried out using any of a variety of single strand cleavage techniques, for example, by inserting a cleavable nucleotide at a predetermined location within the original initiator polynucleotide. An exemplary cleavable nucleotide may be a uracil nucleotide which is cleaved by uracil DNA glycosylase. 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.

[0028] As used herein, the terms "protected" and "blocked" in reference to specified groups, such as, a 3'-hydroxyls of a nucleotide or a nucleoside, are used interchangeably and are intended to mean a moiety is attached covalently to the specified group that prevents a chemical change to the group during a chemical or enzymatic process. Whenever the specified group is a 3'-hydroxyl of a nucleoside triphosphate, or an extended fragment (or "extension intermediate") in which a 3'-protected (or blocked)-nucleoside triphosphate has been incorporated, the prevented chemical change is a further, or subsequent, extension of the extended fragment (or "extension intermediate") by an enzymatic coupling reaction.

[0029] As used herein, an "initiator" (or equivalent terms, such as, "initiating fragment," "initiator nucleic acid," "initiator oligonucleotide," or the like) usually refers to a short oligonucleotide sequence with a free 3'-hydroxyl at its 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 some embodiments, an initiating fragment possesses between 3 and 100 nucleotides, in particular between 3 and 20 nucleotides. In some embodiments, the initiating fragment is single-stranded. In alternative embodiments, the initiating fragment may be double-stranded. In some embodiments, an initiator oligonucleotide may be attached to a synthesis support by its 5' end; and in other embodiments, an initiator oligonucleotide may be attached indirectly to a synthesis support by forming a duplex with a complementary oligonucleotide that is directly attached to the synthesis support, e.g. through a covalent bond. In some embodiments a synthesis support is a solid support which may be a discrete region of a solid planar solid, or may be a bead.

[0030] In some embodiments, an initiator may comprise a non-nucleic acid compound having a free hydroxyl to which a TdT may couple a 3'-O-protected dNTP, e.g. Baiga, U.S. patent publications US2019/0078065 and US2019/0078126.

[0031] After synthesis is completed polynucleotides with the desired nucleotide sequence may be released from initiators and the solid supports by cleavage. A wide variety of cleavable linkages or cleavable nucleotides may be used for this purpose. In some embodiments, cleaving the desired polynucleotide leaves a natural free 5'-hydroxyl on a cleaved strand; however, in alternative embodiments, a cleaving step may leave a moiety, e.g. a 5'-phosphate, that may be removed in a subsequent step, e.g. by phosphatase treatment. Cleaving steps may be carried out chemically, thermally, enzymatically or by photochemical methods. In some embodiments, cleavable nucleotides may be nucleotide analogs such as deoxyuridine or 8-oxo-deoxyguanosine that are recognized by specific glycosylases (e.g. uracil deoxyglycosylase followed by endonuclease VIII, and 8-oxoguanine DNA glycosylase, respectively). In some embodiments, cleavage may be accomplished by providing initiators with a deoxyinosine as the penultimate 3' nucleotide, which may be cleaved by endonuclease V at the 3' end of the initiator leaving a 5'-phosphate on the released polynucleotide. Further methods for cleaving single stranded polynucleotides are disclosed in the following references, which are incorporated by reference: U.S. Pat. Nos. 5,739,386, 5,700,642 and 5,830,655; and U.S. Patent Publication Nos. 2003/0186226 and 2004/0106728; and in Urdea and Horn, U.S. patent 5367066.

[0032] In some embodiments, cleavage by glycosylases and/or endonucleases may require a double stranded DNA substrate.

[0033] Returning to Fig. 1, in some embodiments, an ordered sequence of nucleotides are coupled to an initiator nucleic acid using a template-free polymerase, such as TdT, in the presence of 3'-O-protected dNTPs in each synthesis step. In some embodiments, the method of synthesizing an oligonucleotide comprises the steps of (a) providing an initiator (e.g. (100)) having a free 3'-hydroxyl (103); (b) reacting (104) under extension conditions the initiator or an extension intermediate having a free 3'-hydroxyl with a template-free polymerase in the presence of a 3'-O-protected nucleoside triphosphate to produce a 3'-O-protected extension intermediate (106); (c) deprotecting the extension intermediate to produce an extension intermediate with a free 3'-hydroxyl; and (d) repeating steps (b) and (c) (108) until the polynucleotide is synthesized (110). (Sometimes the terms "extension intermediate" and "elongation fragment" are used interchangeably). 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-protecting 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.

[0034] When the sequence of polynucleotides on a synthesis support includes reverse complementary subsequences, secondary intra-molecular or cross-molecular structures may be created by the formation of hydrogen bonds between the reverse complementary regions. In some embodiments, base protecting moieties for exocyclic amines are selected so that hydrogens of the protected nitrogen cannot participate in hydrogen bonding, thereby preventing the formation of such secondary structures. That is, base protecting moieties may be employed to prevent the formation of hydrogen bonds, such as are formed in normal base pairing, for example, between nucleosides A and T and between G and C. At the end of a synthesis, the base protecting moieties may be removed and the polynucleotide product may be cleaved from the solid support, for example, by cleaving it from its initiator.

[0035] In addition to providing 3'-O-blocked dNTP monomers with base protection groups, elongation reactions may be performed at higher temperatures using thermal stable template-

free polymerases. For example, a thermal stable template-free polymerase having activity above 40°C may be employed; or, in some embodiments, a thermal stable template-free polymerase having activity in the range of from 40-85°C may be employed; or, in some embodiments, a thermal stable template-free polymerase having activity in the range of from 40-65°C may be employed.

[0036] In some embodiments, elongation conditions may include adding solvents to an elongation reaction mixture that inhibit hydrogen bonding or base stacking. Such solvents include water miscible solvents with low dielectric constants, such as dimethyl sulfoxide (DMSO), methanol, and the like. Likewise, in some embodiments, elongation conditions may include the provision of chaotropic agents that include, but are not limited to, n-butanol, ethanol, guanidinium chloride, lithium perchlorate, lithium acetate, magnesium chloride, phenol, 2-propanol, sodium dodecyl sulfate, thiourea, urea, and the like. In some embodiments, elongation conditions include the presence of a secondary-structure-suppressing amount of DMSO. In some embodiments, elongation conditions may include the provision of DNA binding proteins that inhibit the formation of secondary structures, wherein such proteins include, but are not limited to, single-stranded binding proteins, helicases, DNA glycolases, and the like.

[0037] 3'-O-blocked dNTPs without base protection may be purchased from commercial vendors or synthesized using published techniques, e.g. U.S. patent 7057026; Guo et al, Proc. Natl. Acad. Sci., 105(27): 9145-9150 (2008); Benner, U.S. patents 7544794 and 8212020; International patent publications WO2004/005667, WO91/06678; Canard et al, Gene (cited herein); 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. 3'-O-blocked dNTPs with base protection may be synthesized as described below.

[0038] When base-protected dNTPs are employed the above method of Fig. 1 may further include a step (e) removing base protecting moieties, which in the case of acyl or amidine protection groups may (for example) include treating with concentrated ammonia.

[0039] 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 deprotecting 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
5 remove the strand or render it inert to further extensions.

[0040] In some embodiments, reaction conditions for an elongation step (also sometimes referred to as an extension step or a coupling step) may comprising the following: 2.0 μ M purified TdT; 125-600 μ M 3'-O-blocked dNTP (e.g. 3'-O-NH₂-blocked dNTP); about 10 to about 500 mM potassium cacodylate buffer (pH between 6.5 and 7.5) and from about 0.01 to
10 about 10 mM of a divalent cation (e.g. CoCl₂ or MnCl₂), where the elongation reaction may be carried out in a 50 μ L 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-NH₂-blocked dNTPs, reaction conditions for a deblocking step may comprise the following: 700 mM NaNO₂; 1 M sodium acetate (adjusted with acetic acid to pH in the range of 4.8-6.5), where the deblocking
15 reaction may be carried out in a 50 μ L volume, at a temperature within the range of RT to 45°C for 30 seconds to several minutes. Washes may be performed with the cacodylate buffer without the components of the coupling reaction (e.g. enzyme, monomer, divalent cations).

[0041] 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
20 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: Benner, U.S. patents 7544794 and 8212020; U.S. patent 5808045; U.S. patent 8808988; International patent publication WO91/06678; and references cited below. In some embodiments, the cleaving agent (also
25 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,
30 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.

- 5 [0042] 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 (Table 4) may be used in parallel synthesis embodiments. 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.

10 Table 4: Exemplary pairs of blocking groups

3'-O-NH ₂	3'-O-azidomethyl
3'-O-NH ₂	3'-O-allyl
3'-O-NH ₂	3'-O-phosphate
3'-O-azidomethyl	3'-O-allyl
3'-O-azidomethyl	3'-O-phosphate
3'-O-allyl	3'-O-phosphate

- [0043] Synthesizing oligonucleotides on living cells requires mild deblocking, or deprotection, conditions, that is, conditions that do not disrupt cellular membranes, denature
 15 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
 20 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 acetylcholinesterase, 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 MgCl₂, 5 mM 2-mercaptoethanol, and one Unit T4 polynucleotide kinase. The reaction proceeds for one minute at a temperature of
 25 37°C.

[0044] 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 are nucleotidyl-3'-phosphate monoester/nucleotidyl-2',3'-cyclic phosphate, 5 nucleotidyl-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.

[0045] Further examples of synthesis and enzymatic deprotection of 3'-O-ester-protected 10 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 15 Lett., 30(29): 3819-3820 (1989); Becker et al, J. Biol. Chem., 242(5): 936-950 (1967); Tsien, International patent publication WO1991/006678.

[0046] 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' 20 carbon atom has attached a group of the structure:



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 25 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; 30 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
- 5 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 $-C(R')_2-N_3$. In certain embodiments, Z is an azidomethyl group.
- 10 **[0047]** 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
- 15 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
- 20 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)
- 25 **[0048]** In a further embodiments, the 3'-blocked nucleotide triphosphate is blocked by either a 3'-O-azidomethyl, 3'-O- NH_2 or 3'-O-allyl group. In 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), 3'-O-nitro, and 3'-O-propargyl. In other embodiments, the 3'-blocked nucleotide triphosphate is blocked by either
- 30 a 3'-O-azidomethyl or a 3'-O- NH_2 . Synthesis and use of such 3'-blocked nucleoside triphosphates are disclosed in the following references: U.S. patents 9410197; 8808988;

6664097; 5744595; 7544794; 8034923; 8212020; 10472383; Guo et al, Proc. Natl. Acad. Sci., 105(27): 9145-9150 (2008); and like references.

[0049] In some embodiments, 3'-O- protection groups are electrochemically labile groups. That is, deprotection or cleavage of the protection group is accomplished by changing the electrochemical conditions in the vicinity of the protection group which result in cleavage. Such changes in electrochemical conditions may be brought about by changing or applying a physical quantity, such as a voltage difference or light to activate auxiliary species which, in turn, cause changes in the electrochemical conditions at the site of the protection group, such as an increase or decrease in pH. In some embodiments, electrochemically labile groups include, for example, pH-sensitive protection groups that are cleaved whenever the pH is changed to a predetermined value. In other embodiments, electrochemically labile groups include protecting groups which are cleaved directly whenever reducing or oxidizing conditions are changed, for example, by increasing or decreasing a voltage difference at the site of the protection group.

Production of Variant TdTs

[0050] 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, a desired gene or DNA fragment encoding a polypeptide of desired sequence 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 such gene or DNA fragment may be directly cloned from cells of a selected species using conventional protocols, e.g. described by Boule et al, Mol. Biotechnology, 10: 199-208 (1998), or Bentolila et al, EMBO J., 14: 4221-4229 (1995); or the like.

[0051] An isolated gene encoding a desired TdT variant may be inserted into an expression vector, such as pET32 (Novagen) to give an expression vector 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.

[0052] 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.

[0053] TdT protein may be purified from the centrifugate in a one-step affinity procedure.

For example, Ni-NTA affinity column (GE Healthcare) may be used to bind the TdT polymerases. Initially the column is washed and equilibrated with 15 column volumes of 50mM tris-HCL (Sigma) pH 7.5, 150mM NaCl (Sigma) and 20mM imidazole (Sigma). TdT polymerases are bound to the column after equilibration; then, a washing buffer, for example, composed of 50mM tris-HCL (Sigma) pH 7.5, 500mM NaCl (Sigma) and 20mM imidazole (Sigma), may be applied to the column for 15 column volumes. After such washing, the TdT 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 TdT 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.

[0054] 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, or the like); 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. An exemplary His-tag for use with TdT variants of the invention is MASSHHHHHHSSGSENLYFQTGSSG- (SEQ ID NO: 54)). The tag-linker moiety does not interfere with the nucleotide binding activity, or catalytic activity of the TdT variant.

[0055] The above processes, or equivalent processes, result in an isolated TdT variant that may be mixed with a variety of reagents, such as, salts, pH buffers, carrier compounds, and the like, that are necessary or useful for activity and/or preservation.

5

Kits for Practicing Methods of the Invention

[0056] 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. In some embodiments, kits of the invention may include arrays of reaction wells for carrying out multiple synthesis reactions in a single operation. In some embodiments, such arrays may be conventional filter plates comprising 24-, 48-, 96-, 384- or 1536-wells.

[0057] 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.

30

Example 1

Testing Incorporation Activity of TdT Variants

[0058] TdT variants listed in Table 5 were prepared using conventional techniques, purified and tested in two separate hairpin assays as described above using 3'-O-amino-2'-deoxynucleoside triphosphate monomers. Their performance in the assays was compared to that of a mouse-bovine chimeric TdT (designated M57, SEQ ID NO: 55) known to have a high incorporation rate.

Table 5

Designation	SEQ ID NO	mg/ml expression	Hairpin +G (%M57)	Hairpin +C (%M57)	3'-OH (no addition) (%M57)
B10	4	2.3	34	36	25
B19	7	1.0	83	107	42
B57	10	0.8	34	91	62
B59	13	2.8	9	-73	-1
B63	16	2.8	21	162	41
F15	19	7.3	54	625	85
F24	22	5.0	25	324	85
F32	25	5.8	35	536	91
F57	28	0.4	15	172	24
N100	31	8.1	31	90	10
N100or	34	6.2	53	362	28
N14	37	2.8	69	510	96
N27	40	4.3	111	363	85
N35	43	2.8	70	480	101
N72	46	1.2	70	222	87
N93or	49	0.4	37	267	61
M57	55	8.9	100	100	100

As can be seen from the data many of the TdT variants have coupling efficiencies much greater than the standard, M57. Reactions were carried out in presence of low concentration of 3' terminated dNTPs (20 μ M), so the increases are mostly attributed to better affinity towards particular dNTP. Expression levels correspond to protein concentration after elution from Ni-NTA (300 μ L total, derived from 50 ml culture).

Definitions

[0059] 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).

[0060] "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 Multalin 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.

[0061] "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.

[0062] "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.

[0063] "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).

[0064] "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 <http://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.

[0065] “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
5 "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,
10 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
15 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
20 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
25 form or whether both forms are intended from the context of the terms usage.

[0066] "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,
30 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
5 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).

[0067] A “substitution” means that an amino acid residue is replaced by another amino acid
10 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-methyllysine, N-ethylglycine, N-methylglycine, N-ethylasparagine, allo-isoleucine, N-methylisoleucine, N-methylvaline, pyroglutamine, aminobutyric acid,
15 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.

20
[0068] 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
25 present invention is limited only by the appended claims.