



(51) International Patent Classification:

C12N 9/12 (2006.01)

(21) International Application Number:

PCT/GB2020/050247

(22) International Filing Date:

04 February 2020 (04.02.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1901501.5 04 February 2019 (04.02.2019) GB

(71) Applicant: **NUCLERA NUCLEICS LTD** [GB/GB]; 137 Cambridge Science Park, Milton Road, Cambridge CB4 0GD (GB).

(72) Inventors: **CHEN, Michael Chun Hao**; Nuclera Nucleics Ltd, 137 Cambridge Science Park, Milton Road, Cambridge CB4 0GD (GB). **MCINROY, Gordon Ross**; Nuclera Nucleics Ltd, 137 Cambridge Science Park, Milton Road, Cambridge CB4 0GD (GB).

(74) Agent: **STRATAGEM INTELLECTUAL PROPERTY MANAGEMENT LIMITED**; Meridian Court, Comberton Road, Toft, Cambridge CB23 2RY (GB).

(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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: MODIFIED TERMINAL DEOXYNUCLEOTIDYL TRANSFERASE (TdT) ENZYMES

(57) Abstract: The invention relates to engineered terminal deoxynucleotidyl transferase (TdT) enzymes or the homologous amino acid sequence of Pol μ , Pol β , Pol λ , and Pol θ of any species or the homologous amino acid sequence of X family polymerases of any species and their use in a method of nucleic acid synthesis, to methods of synthesizing nucleic acids, and to the use of kits comprising said enzymes in a method of nucleic acid synthesis. The invention also relates to the use of new terminal deoxynucleotidyl transferases and 3'-blocked nucleoside triphosphates in a method of template independent nucleic acid synthesis.



Modified terminal deoxynucleotidyl transferase (TdT) enzymes**FIELD OF THE INVENTION**

The invention relates to the use of specific terminal deoxynucleotidyl transferase (TdT) enzymes or the homologous amino acid sequence of Pol μ , Pol β , Pol λ , and Pol θ of any species or the homologous amino acid sequence of X family polymerases of any species in a method of nucleic acid synthesis, to methods of synthesizing nucleic acids, and to the use of kits comprising said enzymes in a method of nucleic acid synthesis. The invention also relates to the use of terminal deoxynucleotidyl transferases or homologous enzymes and 3'-blocked nucleoside triphosphates in a method of template independent nucleic acid synthesis.

BACKGROUND OF THE INVENTION

Nucleic acid synthesis is vital to modern biotechnology. The rapid pace of development in the biotechnology arena has been made possible by the scientific community's ability to artificially synthesise DNA, RNA and proteins.

Artificial DNA synthesis allows biotechnology and pharmaceutical companies to develop a range of peptide therapeutics, such as insulin for the treatment of diabetes. It allows researchers to characterise cellular proteins to develop new small molecule therapies for the treatment of diseases our aging population faces today, such as heart disease and cancer. It even paves the way forward to creating life, as the Venter Institute demonstrated in 2010 when they placed an artificially synthesised genome into a bacterial cell.

However, current DNA synthesis technology does not meet the demands of the biotechnology industry. Despite being a mature technology, it is practically impossible to synthesise a DNA strand greater than 200 nucleotides in length, and most DNA synthesis companies only offer up to 120 nucleotides. In comparison, an average protein-coding gene is of the order of 2000- 3000 contiguous nucleotides, a chromosome is at least a million contiguous nucleotides in length and an average eukaryotic genome numbers in the billions of nucleotides. In order to prepare nucleic acid strands thousands of base pairs in length, all major gene synthesis companies today rely on variations of a 'synthesise and stitch' technique, where overlapping 40-60-mer fragments are synthesised and stitched together by enzymatic copying and extension. Current methods generally allow up to 3 kb in length for routine production.

The reason DNA cannot be synthesised beyond 120-200 nucleotides at a time is due to the current methodology for generating DNA, which uses synthetic chemistry (i.e., phosphoramidite technology) to couple a nucleotide one at a time to make DNA. Even if the efficiency of each nucleotide-coupling step is 99% efficient, it is mathematically impossible to synthesise DNA longer than 200 nucleotides in acceptable yields. The Venter Institute illustrated this laborious process by spending 4 years and 20 million USD to synthesise the relatively small genome of a bacterium.

Known methods of DNA sequencing use template-dependent DNA polymerases to add 3'-reversibly terminated nucleotides to a growing double-stranded substrate. In the 'sequencing-by-synthesis' process, each added nucleotide contains a dye, allowing the user to identify the exact sequence of the template strand. Albeit on double-stranded DNA, this technology is able to produce strands of between 500-1000 bps long. However, this technology is not suitable for de novo nucleic acid synthesis because of the requirement for an existing nucleic acid strand to act as a template.

Various attempts have been made to use a terminal deoxynucleotidyl transferase for de novo single-stranded DNA synthesis. Uncontrolled de novo single-stranded DNA synthesis, as opposed to controlled, takes advantage of TdT's deoxynucleoside 5'-triphosphate (dNTP) 3'-tailing properties on single-stranded DNA to create, for example, homopolymeric adaptor sequences for next-generation sequencing library preparation. In controlled extensions, reversible deoxynucleoside 5'-triphosphate termination technology needs to be employed to prevent uncontrolled addition of dNTPs to the 3'-end of a growing DNA strand. The development of a controlled single-stranded DNA synthesis process through TdT would be invaluable to in situ DNA synthesis for gene assembly or hybridization microarrays as it removes the need for an anhydrous environment and allows the use of various polymers incompatible with organic solvents.

However, TdT has not been shown to efficiently add nucleoside triphosphates containing 3'-O-reversibly terminating moieties for building up a nascent single-stranded DNA chain necessary for a de novo synthesis cycle. A 3'-O-reversible terminating moiety would prevent a terminal transferase like TdT from catalysing the nucleotide transferase reaction between the 3'-end of a growing DNA strand and the 5'-triphosphate of an incoming nucleoside triphosphate.

There is therefore a need to identify modified terminal deoxynucleotidyl transferases that readily incorporate 3'-O-reversibly terminated nucleotides. Said modified terminal deoxynucleotidyl transferases can be used to incorporate 3'-O-reversibly terminated nucleotides in a fashion useful

for biotechnology and single-stranded DNA synthesis processes in order to provide an improved method of nucleic acid synthesis that is able to overcome the problems associated with currently available methods.

5 BRIEF DESCRIPTION OF THE FIGURES

Figure 1. Incorporation of 3'-O-CH₂-N₃ nucleoside 5'-triphosphates by terminal deoxynucleotidyl transferase (TdT) SEQ ID NOs 1-173. Incorporation rate is defined by quantity (pmol) of reversibly terminated nucleoside 5'-triphosphate incorporation to the 3'-end of a single-stranded DNA molecule per nanogram of TdT per minute. Wild-type bovine TdT activity is represented by the dotted line near y-axis = 0. The dynamic range of the assay saturates at 2.0 pmol incorporate per nanogram of TdT per minute.

Figure 2. Sequence alignment of selected orthologs of wild-type terminal deoxynucleotidyl transferases using the Clustal Omega multiple sequence alignment program provided by the European Molecular Biology Laboratory (EMBL) multiple sequence alignment site.

Figure 3. Top panel: Base-modified substrates A*, C*, G*, and T* were supplied to engineered variants in a single nucleotide incorporation assay. A* describes 3'-aminooxy 6-azido 2'-deoxyadenosine 5'-triphosphate; C* describes 3'-aminooxy 4-azido 5-methyl 2'-deoxycytidine 5'-triphosphate; G* describes 3'-aminooxy 2-azido 2'-deoxyguanosine 5'-triphosphate; T* describes 3'-aminooxy 5-(-hydroxy-1-butynyl) 2'-deoxyuridine 5'-triphosphate. A pool of DNA initiators with degenerate ends was immobilised to a solid support and exposed to nucleotide addition mixture (NAM) solutions containing (1) TdT variant, (2) neutral pH buffer, (3) monovalent salt, (4) cobalt chloride, and (5) 3'-ONH₂-dXTP, where X is a modified nucleobase resulting in A*, C*, G*, or T*. The incubation temperature was 37°C and the reaction time was 2.5 minutes. The solid support was then washed with a high salt solution at neutral pH followed by a low salt solution at neutral pH. The initiator pools were transformed into sequencing libraries and analysed by next-generation sequencing (NGS) on an Illumina NextSeq500 with PE30 reads. Bcl files were converted to fastq files with Illumina's bcl2fastq conversion software and analysed in R. The incorporation efficiency ($\frac{[\text{reads containing N+1 product}]}{[\text{reads containing N initiator}] + [\text{reads containing N+1 product}]}$) was calculated for the addition of each modified base against to all possible initiator contexts. The average incorporation efficiency across all contexts is presented in Figure 3 (top panel). The single nucleotide incorporations by wild-type bovine and wild-type spotted gar TdT of the base-modified reversibly terminated nucleotides A*, C*, G*, and T* are indicated by the dotted line. The dotted

line occurs at $y=0$ due to the inability of wild-type bovine or spotted gar TdTs to incorporate these base-modified reversibly terminated nucleotides in the assay. All of the presented mutations in the TdT variants resulted in improvements in TdT modified-base incorporation relative to wild-type bovine and wild-type spotted gar TdTs.

5 *Bottom panel:* Solid support synthesis of an 8-nt nucleic acid sequence using 3'-ONH₂ nucleoside 5'-triphosphates with terminal deoxynucleotidyl transferase (TdT) SEQ ID NOs 1 & 344-727. The DNA sequence 5'-ATCGATCG-3' was synthesized by repeated exposure of a solid support bound DNA initiator to nucleotide addition mixture (NAM) solutions containing (1) TdT, (2) neutral pH buffer, (3) monovalent salt, (4) cobalt chloride, and (5) 3'-ONH₂-dNTP, where N is selected from adenine, 10 thymine, cytosine, or guanine. A cycle consists of the following: (A) NAM solution with a specified A, T, C, or G reversibly terminated nucleotide was incubated for 5 minutes on the solid support at 37 °C; (B) the solid support was then washed with a high salt solution at neutral pH; (C) the solid support was then exposed to acidic aqueous sodium nitrite; and (D) the solid support was then washed with the same high salt solution at neutral pH from (B). (A)-(D) were then repeated 7 more times to 15 synthesize the desired 8-nt sequence. Synthesized DNA were analysed by running reactions on a denaturing polyacrylamide gel and quantified by virtue of a fluorophore covalently attached to the DNA initiator. The fraction full length (8-nt species) was determined by taking fluorescence intensity of the 8-nt band and dividing by total lane intensity. Wild-type bovine and spotted gar TdT activities as it relates to fraction full-length are indicated by the dotted line. All mutations included in the TdT 20 variants resulted in improvements in TdT relative to wild-type bovine and spotted gar TdTs.

Figure 4. Spotted gar TdT homology structure based on the co-crystal structure of a preferred engineered Nuclera TdT variant bound to a DNA initiator and a reversibly terminated dNTP. An engineered TdT variant with a BRCT-domain truncation was expressed in *Escherichia coli* and 25 purified through immobilized metal affinity chromatography followed by size exclusion chromatography. The aforementioned TdT (60 micromolar) was then co-crystallized by sitting drop vapor diffusion at 19 °C with a DNA oligonucleotide (5' – TTTT[ddC] – 3'; 120 micromolar), cobalt chloride (1 mM), in the absence (top) or presence (bottom) of dATP-ONH₂ (1 mM) in the following reservoir solutions: Bis-Tris HCl (100 mM), NaCl (27 mM), 22% w/v PEG 3350 or Bis-Tris HCl (10 mM) 30 and 27.5% w/v PEG MME 500, respectively. Plate-like crystals appeared and grew to maximum dimensions of 400 x 50 x 5 μm^3 over 1 week and diffracted to 2.6 and 2.5 angstroms, respectively, with a space group of P2₁2₁2₁ and unit cell dimensions of $a = 56.5 \text{ \AA}$, $b = 80.0 \text{ \AA}$, $c = 88.2 \text{ \AA}$. and $a = 55.6 \text{ \AA}$, $b = 79.1 \text{ \AA}$, $c = 88.6 \text{ \AA}$, respectively. Data were indexed, integrated, and scaled using the automatic software pipeline at the Diamond Light Source. Molecular replacement with a homology

model of *Lepisosteus oculatus* (spotted gar) followed by cycles of manual model building interspersed with rigid-body, simulated annealing, energy minimization and individual isotropic D-factor produced complete structural models ($R_{\text{free}} = 27.8\%$ and 28.6% , respectively). Cobalt ions were identified by virtue of their anomalous scattering. The spotted gar TdT wild-type sequence (>95% sequence identity) with a BRCT-domain truncation was modelled onto the two structures.

Figure 5. Surface representation of spotted gar TdT homology structure modelled onto the crystal structure of an engineered TdT variant as described in Figure 4 with MENHNQI motif, which includes the ENHNQ motif, highlighted in black. TdT contains two entrances to the nucleotidyl transferase active site in which the DNA initiator and reversibly terminated dNTP can enter and bind. Determined by modelling and visual inspection of the active site, the MENHNQI and ENHNQ motifs controls access to the nucleotidyl transferase active site, and the mutations shown in this patent demonstrate that mutating this motif results in increased TdT incorporation activity of reversibly terminated nucleotides.

Figure 6. Surface representations (right is 180 degree rotation of left view) of spotted gar TdT homology structure modelled onto the crystal structure of an engineered TdT variant as described in Figure 4 with SEGPLCLAFMRA, which includes the FMRA motif, highlighted in black. TdT contains two entrances to the nucleotidyl transferase active site in which the DNA initiator and reversibly terminated dNTP can enter and bind. Determined by modelling and visual inspection of the active site, the SEGPLCLAFMRA and FMRA motifs (1) control access of the reversibly terminated nucleotide to the nucleotidyl transferase active site through steric effects and electrostatic interactions and (2) directly packs against the critical TGSR motif, which directly contacts the reversibly terminated nucleotide. Specifically in relation to the TGSR motif, the SEGPLCLAFMRA and FMRA motifs directly contact and modulate the positioning of R438, which directly contacts the reversibly terminated nucleotide. Mutations shown in this patent demonstrate that mutating these motifs result in increased TdT incorporation activity of reversibly terminated nucleotides.

Figure 7. Surface representation of spotted gar TdT homology structure modelled onto the crystal structure of an engineered TdT variant as described in Figure 4 with HFTKMQK motif, which includes the MQK motif, highlighted in black. TdT contains two entrances to the nucleotidyl transferase active site in which the DNA initiator and reversibly terminated dNTP can enter and bind. Determined by modelling and visual inspection of the active site, the HFTKMQK and MQK motifs bind to the nucleic acid initiator and assist in positioning the 3'-end of the initiator for nucleophilic attack against the

incoming nucleotide. Mutations shown in this patent demonstrate that mutating this motif results in increased TdT incorporation activity of reversibly terminated nucleotides.

Figure 8. Surface representation of spotted gar TdT homology structure modelled onto the crystal structure of an engineered TdT variant as described in Figure 4 with SAAVCK motif highlighted in black. TdT contains two entrances to the nucleotidyl transferase active site in which the DNA initiator and reversibly terminated dNTP can enter and bind. Determined by modelling and visual inspection of the active site, the SAAVCK motif controls reversibly terminated nucleotide access to the nucleotidyl transferase active site. Mutations shown in this patent demonstrate that mutating this motif results in increased TdT incorporation activity of reversibly terminated nucleotides.

Figure 9. Surface (left) and cartoon/sphere (right) representation of spotted gar TdT homology structure modelled onto the crystal structure of an engineered TdT variant as described in Figure 4 with the GKEC motif, which includes the KEC motif, highlighted in black. TdT contains two entrances to the nucleotidyl transferase active site in which the DNA initiator and reversibly terminated dNTP can enter and bind. Determined by modelling and visual inspection of the active site, the GKEC and KEC motifs control reversibly terminated nucleotide access to and direct binding to the nucleotidyl transferase active site. Mutations shown in this patent demonstrate that mutating this motif results in increased TdT incorporation activity of reversibly terminated nucleotides.

Figure 10. Surface representation of spotted gar TdT homology structure modelled onto the crystal structure of an engineered TdT variant as described in Figure 4 with the DHFQK motif, which includes the DHFQ motif, highlighted in black. TdT contains two entrances to the nucleotidyl transferase active site in which the DNA initiator and reversibly terminated dNTP can enter and bind. Determined by modelling and visual inspection of the active site, the DHFQK and DHFQ motifs control nucleic acid initiator access to the nucleotidyl transferase active site by sterically impinging on the nucleic acid initiator. Mutations shown in this patent demonstrate that mutating this motif results in increased TdT incorporation activity of reversibly terminated nucleotides.

Figure 11. Cartoon representation of an engineered TdT variant crystal structure as described in Figure 4 bound to a DNA initiator and dATP-ONH₂ with the FARHERKMLLDNHA motif highlighted in black spheres. TdT contains two entrances to the nucleotidyl transferase active site in which the DNA initiator and reversibly terminated dNTP can enter and bind. Determined by modelling and visual inspection of the active site, the FARHERKMLLDNHA motif controls nucleotide and nucleic acid

initiator binding in the nucleotidyl transferase active site by sterically impinging on the nucleotide and initiator. Additionally, motif FARHERKMLLDNHA is mutated to FARHERKMLLDRHA. The Asn to Arg mutation results in the Arg residue directly binding one out of two of the catalytic cobalt ions, which is required for activity. Mutations shown in this patent demonstrate that mutating residues within this motif adjust the positioning of critical amino acids in contact with either a Co^{2+} , the reversibly terminated nucleotide, and/or the nucleic acid initiator to result in increased TdT incorporation activity of reversibly terminated nucleotides.

Figure 12. Cartoon representation of an engineered TdT variant crystal structure as described in Figure 4 bound to a DNA initiator and dATP- ONH_2 with the FARHERKMLLDNHALYDKTKK motif highlighted in black spheres. TdT contains two entrances to the nucleotidyl transferase active site in which the DNA initiator and reversibly terminated dNTP can enter and bind. Determined by modelling and visual inspection of the active site, the FARHERKMLLDNHALYDKTKK motif controls nucleotide and nucleic acid initiator binding in the nucleotidyl transferase active site by sterically impinging on the nucleotide and initiator. Additionally, motif FARHERKMLLDNHALYDKTKK is mutated to FARHERKMLLDRHALYDKTKK. The Asn to Arg mutation results in the Arg residue directly binding one out of two of the catalytic cobalt ions, which is required for activity. Mutations shown in this patent demonstrate that mutating residues within this motif adjust the positioning of critical amino acids in contact with either a Co^{2+} , the reversibly terminated nucleotide, and/or the nucleic acid initiator to result in increased TdT incorporation activity of reversibly terminated nucleotides.

Figure 13. Cartoon representation of an engineered TdT variant crystal structure as described in Figure 4 bound to a DNA initiator and dATP- ONH_2 with the DYIDP motif, which contains the YIDP motif, highlighted in black spheres. TdT contains two entrances to the nucleotidyl transferase active site in which the DNA initiator and reversibly terminated dNTP can enter and bind. Determined by modelling and visual inspection of the active site, the DYIDP and YIDP motifs control nucleotide binding in the nucleotidyl transferase active site by modulating the positioning of A494 within the active site. This most C-terminal residue (Ala494) sterically impinges upon the incoming nucleotide. Mutating the YIDP motif, modulates the positioning of the most C-terminal Ala residue resulting in increased TdT incorporation activity of reversibly terminated nucleotides.

Figure 14. Cartoon representation of an engineered TdT variant crystal structure as described in Figure 4 bound to a DNA initiator and dATP- ONH_2 with the YYDIV motif highlighted in black spheres. TdT contains two entrances to the nucleotidyl transferase active site in which the DNA initiator and

reversibly terminated dNTP can enter and bind. Determined by modelling and visual inspection of the active site, the YYDIV motif controls positioning of the nucleic acid initiator by sterically impinging upon the antepenultimate (III), penultimate (II), and ultimate (I) 3'-nucleotide in the nucleic acid initiator. In particular, this motif impinges upon pyrimidines (left, as crystallized, I & II shown in sphere representation), but especially purines (right, II modelled into the structure as a purine). Mutating the YYDIV motif, modulates the positioning and access of the nucleic acid initiator into the nucleotidyl transferase active site resulting in increased TdT incorporation activity of reversibly terminated nucleotides.

Figure 15. Cartoon representation of an engineered TdT variant crystal structure as described in Figure 4 bound to a DNA initiator and dATP-ONH₂ with the APPVDNF, which contains the APPVDN motif, motif highlighted in black spheres. In the structure, the APPVDNF motif is shown with an Ala to Met mutation to MPPVDNF. Through modelling, it was determined that the APPVDNF and APPVDN motifs control positioning of the following motifs by directly packing against the YYDIV (upper right) and FARHERKMLLDNHALYDKTKK (lower left) motifs, both of which are shown in light grey spheres. Additionally determined through modelling, the APPVDNF and APPVDN motifs control positioning of the following motifs by determining the loop conformation of the TGSR motif (upper left) and the catalytic triad of aspartates (lower right) critical to the nucleotidyl transferase mechanism, both of which are shown in light grey spheres. Mutations in the APPVDNF and APPVDN motifs directly influence how the aforementioned critical protein motifs to modulate their contacts with the reversibly terminated nucleotide and nucleic acid initiator. Mutations in the APPVDNF and APPVDN motifs have been shown in this patent to result in increased TdT incorporation activity of reversibly terminated nucleotides.

SUMMARY OF THE INVENTION

Described herein are modified terminal deoxynucleotidyl transferase (TdT) enzymes or the homologous amino acid sequence of Polμ, Polβ, Polλ, and Polθ of any species or the homologous amino acid sequence of X family polymerases of any species. Terminal transferase enzymes are ubiquitous in nature and are present in many species. Many known TdT sequences have been reported in the NCBI database <http://www.ncbi.nlm.nih.gov/>.

GI Number	Species http://www.ncbi.nlm.nih.gov/
gi 768	Bos taurus
gi 460163	Gallus gallus

	gi 494987	Xenopus laevis
	gi 1354475	Oncorhynchus mykiss
	gi 2149634	Monodelphis domestica
	gi 12802441	Mus musculus
5	gi 28852989	Ambystoma mexicanum
	gi 38603668	Takifugu rubripes
	gi 40037389	Raja eglanteria
	gi 40218593	Ginglymostoma cirratum
	gi 46369889	Danio rerio
10	gi 73998101	Canis lupus familiaris
	gi 139001476	Lemur catta
	gi 139001490	Microcebus murinus
	gi 139001511	Otolemur garnettii
	gi 148708614	Mus musculus
15	gi 149040157	Rattus norvegicus
	gi 149704611	Equus caballus
	gi 164451472	Bos taurus
	gi 169642654	Xenopus (Silurana) tropicalis
	gi 291394899	Oryctolagus cuniculus
20	gi 291404551	Oryctolagus cuniculus
	gi 301763246	Ailuropoda melanoleuca
	gi 311271684	Sus scrofa
	gi 327280070	Anolis carolinensis
	gi 334313404	Monodelphis domestica
25	gi 344274915	Loxodonta africana
	gi 345330196	Ornithorhynchus anatinus
	gi 348588114	Cavia porcellus
	gi 351697151	Heterocephalus glaber
	gi 355562663	Macaca mulatta
30	gi 395501816	Sarcophilus harrisii
	gi 395508711	Sarcophilus harrisii
	gi 395850042	Otolemur garnettii
	gi 397467153	Pan paniscus
	gi 403278452	Saimiri boliviensis boliviensis

	gi 410903980	Takifugu rubripes
	gi 410975770	Felis catus
	gi 432092624	Myotis davidii
	gi 432113117	Myotis davidii
5	gi 444708211	Tupaia chinensis
	gi 460417122	Pleurodeles waltl
	gi 466001476	Orcinus orca
	gi 471358897	Trichechus manatus latirostris
	gi 478507321	Ceratotherium simum simum
10	gi 478528402	Ceratotherium simum simum
	gi 488530524	Dasypus novemcinctus
	gi 499037612	Maylandia zebra
	gi 504135178	Ochotona princeps
	gi 505844004	Sorex araneus
15	gi 505845913	Sorex araneus
	gi 507537868	Jaculus jaculus
	gi 507572662	Jaculus jaculus
	gi 507622751	Octodon degus
	gi 507640406	Echinops telfairi
20	gi 507669049	Echinops telfairi
	gi 507930719	Condylura cristata
	gi 507940587	Condylura cristata
	gi 511850623	Mustela putorius furo
	gi 512856623	Xenopus (Silurana) tropicalis
25	gi 512952456	Heterocephalus glaber
	gi 524918754	Mesocricetus auratus
	gi 527251632	Melopsittacus undulatus
	gi 528493137	Danio rerio
	gi 528493139	Danio rerio
30	gi 529438486	Falco peregrinus
	gi 530565557	Chrysemys picta bellii
	gi 532017142	Microtus ochrogaster
	gi 532099471	Ictidomys tridecemlineatus
	gi 533166077	Chinchilla lanigera

- gi|533189443 Chinchilla lanigera
- gi|537205041 Cricetulus griseus
- gi|537263119 Cricetulus griseus
- gi|543247043 Geospiza fortis
- 5 gi|543351492 Pseudopodoces humilis
- gi|543731985 Columba livia
- gi|544420267 Macaca fascicularis
- gi|545193630 Equus caballus
- gi|548384565 Pundamilia nyererei
- 10 gi|551487466 Xiphophorus maculatus
- gi|551523268 Xiphophorus maculatus
- gi|554582962 Myotis brandtii
- gi|554588252 Myotis brandtii
- gi|556778822 Pantholops hodgsonii
- 15 gi|556990133 Latimeria chalumnae
- gi|557297894 Alligator sinensis
- gi|558116760 Pelodiscus sinensis
- gi|558207237 Myotis lucifugus
- gi|560895997 Camelus ferus
- 20 gi|560897502 Camelus ferus
- gi|562857949 Tupaia chinensis
- gi|562876575 Tupaia chinensis
- gi|564229057 Alligator mississippiensis
- gi|564236372 Alligator mississippiensis
- 25 gi|564384286 Rattus norvegicus
- gi|573884994 Lepisosteus oculatus

The sequences of the various described terminal transferases show some regions of highly conserved sequence, and some regions which are highly diverse between different species. A
 30 sequence alignment for sequences from a selection of species is shown in Figure 2.

The inventors have modified the terminal transferase from *Lepisosteus oculatus* TdT (spotted gar) (shown as SED ID 1). However the corresponding modifications can be introduced into the analagous

terminal transferase sequences from any other species, including the sequences listed above in the various NCBI entries, including those shown in Figure 2 or truncated versions thereof.

The amino acid sequence of the spotted gar (*Lepisosteus oculatus*) is shown below (SEQ ID 1)

5
MLHIPIFPPIKKRQKLPEsrNSCKYEVKFSEVAIFLVERKMGSSRRKFLTNLARSkgFRIEDVLSDAV
THVVAEDNSADELWQWLQNSSLGDLskIEVLDisWFTECMGAGKPVQVEARHCLVKSCPVIDQYLEPS
TVETVSQYACQRRTTMENHNQIFtdaFaiLAENAEFNESEGpCLAFMRAASLLKSLPHAISSSKDLEG
LPCLGDQTKAVIEDILEYgQCSKVQDVLCDdRYQTIKLFTSVFGVGLKTAekWYRKGFHSLeeVQADN
10 AIHFTKMqKAGFLYYDDISAAVCKAEaQAIGQIVEETVRLIAPDAIVTLTGGFRRGKECGHDVDFLIT
TPeMGKEVWLLNRLINRLQNQGILLYDIVESTFDKTRLPCRkFEAMdHFQKCFaiIKLKKELAAGRV
QKDWKAIRVDFVAPPVDNFAFALLGWTGSRQFERDLRRFARHERKMLLDNHALYDKTKKI FLPAKTEE
DIFaHLGLDYIDPWQRNA

15 An engineered variant of this sequence was previously identified as SEQ ID NO 8 in publication WO2016/128731. Further engineered improvements to this published sequence are described herein. The modified sequences disclosed herein are different to SEQ ID NO 8 disclosed in the prior art. WO2016/128731 SEQ ID NO 2 is a "mis-annotated" wild-type gar sequence.

20 SEQ ID NO 8 in publication WO2016/128731 is shown below with the engineered mutations identified:

MLHIPIFPPIKKRQKLPEsrNSCKYEVKFSEVAIFLVERKMGSSRRKFLTNLARSkgFRIEDVLSDAV
THVVAE~~EN~~NSADEL~~EQ~~WQWLQNSSLGDLskIEVLDisWFTECMGAGKPVQVEARHCLVKSCPVIDQYLEPS
25 TVETVSQYACQRRTTMENHNQIFtdaFaiLAENAEFNESEGpCLAFMRAASLLKSLPHAISSSKDLEG
LPCLGDQTKAVIEDILEYgQCSKVQDVLCDdRYQTIKLFTSVFGVGL~~ET~~TAekWYRKGFHSLeeVQADN
AIHFTKMqKAGFLYYDDISAAVCKAEaQAIGQIVEETVRLIAPDAIVTLTGGFRRGKECGHDVDFLIT
TPeMGKEVWLLNRLINRLQNQGILLYDIVESTFDKTRLPCRkFEAMdHFQKCFaiIKLKKELAAGRV
QKDWKAIRVDFVAPPVDNFAFALLGWTGSRQFERDLRRFARHERKMLLDNHALYDKTKKI FLPAKTEE
30 DIFaHLGLDYIDPWQRNA

The inventors have identified various amino acids modifications in the amino acid sequence having improved properties. Certain regions improve the solubility and handling of the enzyme. Certain other regions improve the ability to incorporate nucleotides with modifications; these modifications
35 include modifications at the 3'-position of the sugar and modifications to the base.

Described herein are modified terminal deoxynucleotidyl transferase (TdT) enzymes comprising amino acid modifications when compared to a wild type sequence SEQ ID NO 1 or a truncated version thereof or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species or the homologous amino acid sequence of Pol μ , Pol β , Pol λ , and Pol θ of any species or the homologous amino acid sequence of X family polymerases of any species, wherein the amino acid is modified at one or more of the amino acids:

V32, A33, I34, F35, A53, V68, V71, E97, I101, M108, G109, A110, Q115, V116, S125, T137, Q143, M152, E153, N154, H155, N156, Q157, I158, I165, N169, N173, S175, E176, G177, P178, C179, L180, A181, F182, M183, R184, A185, L188, H194, A195, I196, S197, S198, S199, K200, E203, G204, D210, Q211, T212, K213, A214, I216, E217, D218, L220, Y222, V228, D230, Q238, T239, L242, L251, K260, G261, F262, H263, S264, L265, E267, Q269, A270, D271, N272, A273, H275, F276, T277, K278, M279, Q280, K281, S291, A292, A293, V294, C295, K296, E298, A299, Q300, A301, Q304, I305, T309, V310, R311, L312, I313, A314, I318, V319, T320, G328, K329, E330, C331, L338, T341, P342, E343, M344, G345, K346, W349, L350, L351, N352, R353, L354, I355, N356, R357, L358, Q359, N360, Q361, G362, I363, L364, L365, Y366, Y367, D368, I369, V370, K376, T377, C381, K383, D388, H389, F390, Q391, K392, F394, I397, K398, K400, K401, E402, L403, A404, A405, G406, R407, D411, A421, P422, P423, V424, D425, N426, F427, A430, R438, F447, A448, R449, H450, E451, R452, K453, M454, L455, L456, D457, N458, H459, A460, L461, Y462, D463, K464, T465, K466, K467, T474, D477, D485, Y486, I487, D488, P489.

Modifications which improve the solubility include a modification within the amino acid region WLLNRLINRLQNQGILLYDIV shown highlighted in the sequence below.

MLHIPIFPPIKKRQKLPEsrNSCKYEVKFSEVAIFLVERKMGSSRRKFLTNLARSkgFRIEDVLSDAV
THVVAEDNSADELWQWLQNSSLGDLskIEVLDisWFTECMGAGKPVQVEARHCLVKSCPVIDQYLEPS
TVETVSQYACQRRTTMENHNQIFTDaFaiLAENAEfNESEGpCLAFmRAASLLKSLPHaISSSKDLEG
LPCLGDQTKAVIEDILEYgQCSKVQDVLCDdRYQTIKLFTSVFGVGLKTAekWYRKGFHsLEEVQADN
AIHFtKMqKAGFLYYDDISAAVCKAEaQAIGQIVEETVRLIAPDAIVTLTGGFRGKECGHDVDFLIT
TPeMGKEVWLLNRLINRLQNQGILLYDIVeSTFDKTRLPCRKFAMdHFQKCFAIIKLKkELAAgrV
QKDWKAIRVDFVAPPVDNFAFaLLGWTGSrQFERDLRRFaRHERKMLLDNHAlYDKTKKIFLPaKTEE
DIFaHLGLDYIDPWQRNA

Modifications which improve the incorporation of modified nucleotides can be at one or more of selected regions shown below. Regions were selected according to mutation data (Figures 1 and 3),

sequence alignment (Figure 2), and structural data obtained from spotted gar TdT co-crystallized with DNA and a 3'-modified dNTP (Figures 4-14). The second modification can be selected from one or more of the amino acid regions VAIF, MGA, MENHNQI, SEGPCLAFMRA, HAISSS, DQTKA, KGFHS, QADNA, HFTKMQK, SAAVCK, EAQA, TVRLI, GKEC, TPENMGK, YYDIV, DHFQK, LAAG, APPVDNF, FARHERKMLLDNHALYDKTKK, and DYIDP shown highlighted in the sequence below.

MLHIPIFPPIKKRQKLPEsrNSCKYEVKFSEVAIFLVERKMGSSRRKFLTNLARSkgFRIEDVLSDAV
 THVVAEDNSADELWQWLQNSSLGDLskIEVLDisWFTECMGAGKPVQVEARHCLVKSCPVIDQYLEPS
 TVETVSQYACQRRRTTMENHNQIFTDAFAILAENAEFNESEGPCLAFMRAASLLKSLPHAISSSKDLEG
 LPCLGDOQKAVIEDILEYGQCSKVQDVLCDdryQTIKLFTSVFGVGLKTAekWYRKGFHSLeeVQADN
 AIHFTKMQRAGFLYYDDISAAVCKABAQAIQIveETVRLIAPDAIVTLTGgFRRGKECGHDVDFLIT
 TPENMGKEVWLLNRLINRLQNQGILLYDIVESTFDKTRLPCRKFeamDHFQKCFaIIKLKKEIAAGRV
 QKDWKAIRVDFVAPPVDNFAFALLGWTGSRQFERDLRRFARHERKMLLDNHALYDKTKKIFLPakTEE
 DIFaHLGLDYIDPWQRNA

Particular modifications which improve the incorporation of modified nucleotides can be at one or more of selected regions shown below. The second modification can be selected from one or more of the amino acid regions VAIF, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA, TVR, KEC, TPENMGK, DHFQ, LAAG, APPVDN, FARHERKMLLDNHA, and YIDP shown highlighted in the sequence below.

MLHIPIFPPIKKRQKLPEsrNSCKYEVKFSEVAIFLVERKMGSSRRKFLTNLARSkgFRIEDVLSDAV
 THVVAEDNSADELWQWLQNSSLGDLskIEVLDisWFTECMGAGKPVQVEARHCLVKSCPVIDQYLEPS
 TVETVSQYACQRRRTTMENHNQIFTDAFAILAENAEFNESEGPCLAFMRAASLLKSLPHAISSSKDLEG
 LPCLGDOQKAVIEDILEYGQCSKVQDVLCDdryQTIKLFTSVFGVGLKTAekWYRKGFHSLeeVQADN
 AIHFTKMQRAGFLYYDDISAAVCKABAQAIQIveETVRLIAPDAIVTLTGgFRRGKECGHDVDFLIT
 TPENMGKEVWLLNRLINRLQNQGILLYDIVESTFDKTRLPCRKFeamDHFQKCFaIIKLKKEIAAGRV
 QKDWKAIRVDFVAPPVDNFAFALLGWTGSRQFERDLRRFARHERKMLLDNHALYDKTKKIFLPakTEE
 DIFaHLGLDYIDPWQRNA

Described herein is a modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising at least one amino acid modification when compared to a wild type sequence SEQ ID NO 1 or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species, wherein the modification is selected from one or more of the amino acid regions WLLNRLINRLQNQGILLYDIV, VAIF, MGA, MENHNQI, SEGPCLAFMRA, HAISSS, DQTKA, KGFHS, QADNA, HFTKMQK, SAAVCK, EAQA, TVRLI, GKEC, TPENMGK, DHFQK, LAAG, APPVDNF,

FARHERKMLLDNHALYDKTKK, and DYIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species.

References to particular sequences include truncations thereof. Included herein are modified
 5 terminal deoxynucleotidyl transferase (TdT) enzyme comprising at least one amino acid modification
 when compared to a wild type sequence SEQ ID NO 1 or a truncated version thereof, or the
 homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other
 species, wherein the modification is selected from one or more of the amino acid regions
 WLLNRLINRLQNQGILLYDIV, VAIF, MGA, MENHNQI, SEGPCLAFMRA, HAIS, DQTKA, KGFHS,
 10 QADNA, HFTKMQK, SAAVCK, EAQA, TVRLI, GKEC, TPEMGK, DHFQK, LAAG, APPVDNF,
 FARHERKMLLDNHALYDKTKK, and DYIDP of the sequence of SEQ ID NO 1 or the homologous regions
 in other species.

Truncated proteins may include at least the region shown below

15
 TVSQYACQRRRTMENHNQIFTDFAFILAENAEFNESEGPCLAFMRAASLLKSLPHAISSSKDLEGLPC
 LGDQTKAVIEDILEYGQCSKVQDVLCDTRYQTIKFTSVFVGLKTAEKWYRKGFSLEEVQADNAIH
 FTKMQKAGFLYYDDISAAVCKAEQAIGQIVEETVRLIAPDAIVTLTGGFRRGKECGHDVDFLITPE
 MGKEVWLLNRLINRLQNQGILLYDVESTFDKTRLPCRKFAMDFQKCFAIKLLKELAAAGRVQKD
 20 WKAIRVDFVAPPVDNFALLGWTGSRQFERDLRRFARHERKMLLDNHALYDKTKKIFLPAKTEEDIF
 AHLGLDYIDPWQRNA

Described herein is a modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising at
 least the sequence:

25
 TVSQYACQRRRTMENHNQIFTDFAFILAENAEFNESEGPCLAFMRAASLLKSLPHAISSSKDLEGLPC
 LGDQTKAVIEDILEYGQCSKVQDVLCDTRYQTIKFTSVFVGLKTAEKWYRKGFSLEEVQADNAIH
 FTKMQKAGFLYYDDISAAVCKAEQAIGQIVEETVRLIAPDAIVTLTGGFRRGKECGHDVDFLITPE
 MGKEVWLLNRLINRLQNQGILLYDVESTFDKTRLPCRKFAMDFQKCFAIKLLKELAAAGRVQKD
 30 WKAIRVDFVAPPVDNFALLGWTGSRQFERDLRRFARHERKMLLDNHALYDKTKKIFLPAKTEEDIF
 AHLGLDYIDPWQRNA

or the homologous regions in other species, wherein the sequence has one or more amino acid
 modifications in one or more of the amino acid regions WLLNRLINRLQNQGILLYDI, MENHNQI,

SEGPCLAFMRA, HAISSS, DQTKA, KGFHS, QADNA, HFTKMQK, SAAVCK, EAQA, TVRLI, GKEC, TPMEGK, DHFQK, LAAG, APPVDNF, FARHERKMMLLDNHALYDKTKK, and DYIDP of the sequence.

TVSQYACQRRRTT MENHNQI FTDAFAILAENAEFNE SEGPCLAFMRAASLLKSLPHATSSSKDLEGLPC
 5 LGDQTEAVIEDILEYGQCSKVQDVLCDTRYQTIKLFTSVFVGLKTAEKWYRKGHSLSEEVQADNAIH
 FTKMKAGFLYYDDISAAVCKABQAIGQIVEETVRLIAPDAIVTLTGGRRGKGGHGDVDFLITPE
 MGKEVWLENRIINFLQNGTLLYYDI VESTFDKTRLPCRKFEMDHFQKCFAI IKLKKELAAGRVQKD
 WKAIRVDFVAPPVDNEAFALLGWTGSRQFERDLRRFARHERKMMLLDNHALYDKTKKIFLPAKTEEDIF
 AHLGLDIDFWQRNA

10

Disclosed herein is a modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising at least one amino acid modification when compared to a wild type sequence, wherein the modification is selected from one or more of the amino acid regions WLLNRLINRLQNQGILLYDI, VAIF, MGA, MENHNQI, SEGPCLAFMRA, HAISSS, DQTKA, KGFHS, QADNA, HFTKMQK, SAAVCK, EAQA,
 15 TVRLI, GKEC, TPMEGK, DHFQK, LAAG, APPVDNF, FARHERKMMLLDNHALYDKTKK, and DYIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species.

20

Disclosed herein is a modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising at least one amino acid modification when compared to a wild type sequence, wherein the modification is selected from one or more of the amino acid regions WLLNRLINRLQNQGILLYDI, VAIF, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA, TVR, KEC, TPMEGK, DHFQ, LAAG, APPVDN, FARHERKMMLLDNHA, and YIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species.

25

Described herein is a modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising at least one amino acid modification when compared to a wild type sequence SEQ ID NO 1 or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species, wherein the modification is selected from one or more of the amino acid regions WLLNRLINRLQNQGILLYDI, VAIF, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA,
 30 TVR, KEC, TPMEGK, DHFQ, LAAG, APPVDN, FARHERKMMLLDNHA, and YIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species.

35

Homologous refers to protein sequences between two or more proteins that possess a common evolutionary origin, including proteins from superfamilies in the same species of organism as well as homologous proteins from different species. Such proteins (and their encoding nucleic acids) have

sequence homology, as reflected by their sequence similarity, whether in terms of percent identity or by the presence of specific residues or motifs and conserved positions. A variety of protein (and their encoding nucleic acid) sequence alignment tools may be used to determine sequence homology. For example, the Clustal Omega multiple sequence alignment program provided by the European Molecular Biology Laboratory (EMBL) can be used to determine sequence homology or homologous regions.

Improved sequences as described herein can contain both modifications, namely

- a. a first modification is within the amino acid region WLLNRLINRLQNQGILLYDIV of the sequence of SEQ ID NO 1 or the homologous region in other species; and
- b. a second modification is selected from one or more of the amino acid regions VAIF, MGA, MENHNQI, SEGPCLAFMRA, HAISSS, DQTKA, KGFHS, QADNA, HFTKMQK, SAAVCK, EAQA, TVRLI, GKEC, TPENMGK, DHFQK, LAAG, APPVDNF, FARHERKMILLDNHALYDKTKK, and DYIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species.

The sequence can be truncated

Improved sequences as described herein can contain both modifications, namely

- a. a first modification is within the amino acid region WLLNRLINRLQNQGILLYDI of the sequence of SEQ ID NO 1 or the homologous region in other species; and
- b. a second modification is selected from one or more of the amino acid regions VAIF, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA, TVR, KEC, TPENMGK, DHFQ, LAAG, APPVDN, FARHERKMILLDNHA, and YIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species.

As a comparison with other species, the sequence of *Bos taurus* (cow) TdT is shown below:

```
MDPLCTASSGPRKKRPRQVGASMASPPHDIKFQNLVLFLEKKMGTRRNFLMELARRKGFVENELS
DSVTHIVAENNSGSEVLEWLQVQNIRASSQLELLDVSWLIESMGAGKPVEITGKHQLVVRTDY SATPN
PGFQKTPPLAVKKISQYACQRKTTLNYNHIFTDAFEILAENSEFKENEVS YVTFMRAASVLKSLPFT
IISMKDTEGIPCLGDKVKCIIEEIIEDGESSEVKAVLNDERYQSFKLFTSVFGVGLKTSEKWFRMGFR
SLSKIMSDKTLKFTKMQKAGFLYYEDLVSCVTRAEAEAVGVLVKEAVWAFLPDAFVTMTGGFRRGKKI
GHDVDFLITSPGSAEDEEQLLPKVINLWEKKGLLLYYDLVESTFEKFKLPSRQVDTLDHFQKCFILIK
LHHQRVDSSKSNQQEGKTKAIRVDLMCPYENRAFALLGWTGSRQFERDIRRYATHERKMMLDNHAL
YDKTKRVFLKAESEEEIFAHGLDYIEPWERN
```

The homologous regions in the sequences are highlighted below.

Modifications which improve the solubility include a modification within the amino acid region QLLPKVINLWEKKGLLLYYDLV shown highlighted in the sequence below.

5

MDPLCTASSGPRKKRPRQVGASMASPPHDIKFQNLVLFIEKKMGTTTRRNFLMELARRKGFRVENELS
 DSVTHIVAENNSGSEVLEWLQVQNI RASSQLELLDVSWLIESMGAGKPVEITGKHQLVVRTDY SATPN
 PGFQKTPPLAVKKISQYACQRKTTLN NNYNHI FTDAFEILAENSEFKENEVS YVTFMRAASVLKSLPFT
 IISMKDTEGIPCLGDKVKCIIEEIIEDGESSEVKAVLNDERYQSFKLFTSVFGVGLKTSEKWFRMGFR
 10 SLSKIMSDKTLKFTKMQKAGFLYYEDLVSCVTRAEAEAVGVLVKEAVWAFLPDAFVTMTGGFRRGKKI
 GHDVDFLITSPGSAEDEEQLLPKVINLWEKKGLLLYYDLVESTFEKFKLPSRQVDTLDHFQKCFILIK
 LHHQRVDSSKSNQQEGKTKWKAIRVDLMCPYENRAFALLGWTGSRQFERDIRRYATHERKMMLDNHAL
 YDKTKRVFLKAESEEEIFAHLGLDYIEPWERNA

15

Modifications which improve the incorporation of modified nucleotides can be at one or more of selected regions shown below. The second modification can be selected from one or more of the amino acid regions LVLF, MGA, LNNYNHI, NEVS YVTFMRA, FTIISM, DKVKC, MGFRS, MSDKT, KFTKMQK, VSCVTR, EAEA, AVWAFL, GKKI, SPGSAE, YYDLV, DHFQK, MCPYENR, YATHERKMMLDNHALYDKTKR, and DYIEP shown highlighted in the sequence below.

20

MDPLCTASSGPRKKRPRQVGASMASPPHDIKFQNLVLFIEKKMGTTTRRNFLMELARRKGFRVENELS
 DSVTHIVAENNSGSEVLEWLQVQNI RASSQLELLDVSWLIESMGAGKPVEITGKHQLVVRTDY SATPN
 PGFQKTPPLAVKKISQYACQRKTTLN NNYNHI FTDAFEILAENSEFKENEVS YVTFMRAASVLKSLPFT
 IISMKDTEGIPCLGDKVKCIIEEIIEDGESSEVKAVLNDERYQSFKLFTSVFGVGLKTSEKWFRMGFR
 25 SLSKIMSDKTLKFTKMQKAGFLYYEDLVSCVTRAEAEAVGVLVKEAVWAFLPDAFVTMTGGFRRGKKI
 GHDVDFLITSPGSAEDEEQLLPKVINLWEKKGLLLYYDLVESTFEKFKLPSRQVDTLDHFQKCFILIK
 LHHQRVDSSKSNQQEGKTKWKAIRVDLMCPYENRAFALLGWTGSRQFERDIRRYATHERKMMLDNHAL
 YDKTKRVFLKAESEEEIFAHLGLDYIEPWERNA

30

Modifications which improve the incorporation of modified nucleotides can be at one or more of selected regions shown below. The second modification can be selected from one or more of the amino acid regions LVLF, MGA, NNNYNH, FMRA, FTI, VKC, FRS, MSDKT, MQK, EAEA, AVW, KKI, SPGSAE, DHFQ, MCPYEN, YATHERKMMLDNHA, and YIEP shown highlighted in the sequence below.

35

MDPLCTASSGPRKKRPRQVGASMASPPHDIKFQNLVLFIEKKMGTTTRRNFLMELARRKGFRVENELS
 DSVTHIVAENNSGSEVLEWLQVQNI RASSQLELLDVSWLIESMGAGKPVEITGKHQLVVRTDY SATPN

PGFQKTPPLAVKKISQYACQRKTTLLNNYNI FTDAFEILAENSEFKENEVSIVTMMRAASVLKSLPPT
 ITSMKDTEGIPCLGDKVKSIIEEIIEDGESSEVKAVLNDERYQSFKLFTSVFGVGLKTSEKWFRMGFR
 SLSKIMSDEKSLRFTQMOKAGFLYYEDLVSCVTRAEAEAVGVLVKEAVTFLPDADFVTMTGGFRRGKMT
 GHDVDFLITSPEATEDEEQQLLPKVINLWEKKGLLLYDLVESTFEKFKLPSRQVDTHFQKCFILK
 5 LHHQRVDSSKSNQQEGKTKAIRVDLMCPYDRRAFALLGWTGSRQFERDIRRYATHERKMMLDNHAL
 YDKTKRVFLKAESEEEI FAHLGLDYIEPWERNA

As a comparison with other species, the sequence of *Mus musculus* (mouse) TdT is shown below:

10 MDPLQAVHLGPRKKRPRQLGTPVASTPYDIRFRDLVLFIEKKMGTTTTRAFMELARRKGFVRVENELS
 DSVTHIVAENNSGSDVLEWLQLQNIKASSELELLDISWLIIECMGAGKPVEMMGRHQLVVRNNSPSPV
 PGSQNVPAVAKKISQYACQRRTTLNNYNQLFTDALDILAENDELRENEGSCLAFLMRASSVLKSLPFP
 ITSMKDTEGIPCLGDKVKSIIIEGIIEDGESSEAKAVLNDERYKSFKLFTSVFGVGLKTAEKWFRMGFR
 TLSKIQSDKSLRFTQMOKAGFLYYEDLVSCVNRPEAEAVSMLVKEAVVTFLPDALVTMTGGFRRGKMT
 15 GHDVDFLITSPEATEDEEQQLLHKVTDVFWKQQGLLLYCDILESTFEKFKQPSRKVDALDHFQKCFILK
 KLDHGRVHSEKSGQEGKGWKAIRVDLMCPYDRRAFALLGWTGSRQFERDLRRYATHERKMMLDNHA
 LYDRTKGKTVTISPLDGKVS KLQKALRVFLEAESEEEI FAHLGLDYIEPWERNA

Modifications which improve the solubility include a modification within the amino acid region

20 QLLHKVTDVFWKQQGLLLYCDIL shown highlighted in the sequence below.

MDPLQAVHLGPRKKRPRQLGTPVASTPYDIRFRDLVLFIEKKMGTTTTRAFMELARRKGFVRVENELS
 DSVTHIVAENNSGSDVLEWLQLQNIKASSELELLDISWLIIECMGAGKPVEMMGRHQLVVRNNSPSPV
 PGSQNVPAVAKKISQYACQRRTTLNNYNQLFTDALDILAENDELRENEGSCLAFLMRASSVLKSLPFP
 25 ITSMKDTEGIPCLGDKVKSIIIEGIIEDGESSEAKAVLNDERYKSFKLFTSVFGVGLKTAEKWFRMGFR
 TLSKIQSDKSLRFTQMOKAGFLYYEDLVSCVNRPEAEAVSMLVKEAVVTFLPDALVTMTGGFRRGKMT
 GHDVDFLITSPEATEDEEQQLLHKVTDVFWKQQGLLLYCDILESTFEKFKQPSRKVDALDHFQKCFILK
 KLDHGRVHSEKSGQEGKGWKAIRVDLMCPYDRRAFALLGWTGSRQFERDLRRYATHERKMMLDNHA
 LYDRTKGKTVTISPLDGKVS KLQKALRVFLEAESEEEI FAHLGLDYIEPWERNA

30

Modifications which improve the incorporation of modified nucleotides can be at one or more of
 selected regions shown below. The second modification can be selected from one or more of the
 amino acid regions LVL, MGA, LNNYNQL, NEGSCLAFLMRA, FPITSM, DKVKS, MGFR, QSDKS,
 RFTQMOK, VSCVNR, EAEA, AVVTFL, GKMT, SPEATE, DHFQK, SGQ, MCPYDRR,
 35 YATHERKMMLDNHALYDRT, R, and DYIEP shown highlighted in the sequence below.

MDPLQAVHLGPRKKRPRQLGTPVASTPYDIRFRDILVLEKMKGTTRRAFLMELARRKGFRVENELS
 DSVTHIVAENNSGSDVLEWLQLQNIKASSELELLDISWLEIECKGAGKPVEMMGRHQLVVRNRNSSPSPV
 PGSQNVPAFAVKKISQYACQRRTTNNYNQLFTDALDILAENDELRENEGSCCLAFMRASSVLKSLPFF
 ITSMKDTEGIPCLGDKVKSIIIEGIIEDGESSEAKAVLNDERYKSFKLFTSVFGVGLKTAEKWFRMGFR
 5 TLSKIQSDKSLRFTQMOKAGFLYYEDLVSCVNRPEAEAVSMLVKEAVVTFLPDALVTMTGGFRRGKMT
 GHDVDFLITSPEATEDEEQQLLHKVTD FWKQQGLLLYCDILESTFEKFKQPSRKVDALDHFQKCFLLIL
 KLDHGRVHSEKSGQEGKGWKAIRVDLVMCOPYDRAFALLGWTGSRQFERDLRRYATHERKMMLDNHA
 LYDRTKGKTVTISPLDGKVS KLQKALRVFLEAESEEEIFAHLGLDYIEPWERNA

- 10 Modifications which improve the incorporation of modified nucleotides can be at one or more of selected regions shown below. The second modification can be selected from one or more of the amino acid regions LVLF, MGA, LNNYNQ, NEGSCCLAFMRA, FPI, VKS, FRT, SKIQSDKS, MQK, VSCVNR, EAEA, AVV, KMT, SPEATE, DHFQK, MCPYDR, YATHERKMMLDNHA, and YIEP shown highlighted in the sequence below.

15

MDPLQAVHLGPRKKRPRQLGTPVASTPYDIRFRDILVLEKMKGTTRRAFLMELARRKGFRVENELS
 DSVTHIVAENNSGSDVLEWLQLQNIKASSELELLDISWLEIECKGAGKPVEMMGRHQLVVRNRNSSPSPV
 PGSQNVPAFAVKKISQYACQRRTTNNYNQLFTDALDILAENDELRENEGSCCLAFMRASSVLKSLPFF
 ITSMKDTEGIPCLGDKVKSIIIEGIIEDGESSEAKAVLNDERYKSFKLFTSVFGVGLKTAEKWFRMGFR
 20 TLSKIQSDKSLRFTQMOKAGFLYYEDLVSCVNRPEAEAVSMLVKEAVVTFLPDALVTMTGGFRRGKMT
 GHDVDFLITSPEATEDEEQQLLHKVTD FWKQQGLLLYCDILESTFEKFKQPSRKVDALDHFQKCFLLIL
 KLDHGRVHSEKSGQEGKGWKAIRVDLVMCOPYDRAFALLGWTGSRQFERDLRRYATHERKMMLDNHA
 LYDRTKGKTVTISPLDGKVS KLQKALRVFLEAESEEEIFAHLGLDYIEPWERNA

- 25 Modifications which improve the incorporation of modified nucleotides can be at one or more of selected regions shown below. The second modification can be selected from one or more of the amino acid regions LVLF, MGA, NNYNQ, FMRA, FPI, VKS, FRT, QSDKS, MQK, VSCVNR, EAEA, AVV, KMT, SPEATE, DHFQ, MCPYDR, YATHERKMMLDNHA, and YIEP shown highlighted in the sequence below.

30

MDPLQAVHLGPRKKRPRQLGTPVASTPYDIRFRDILVLEKMKGTTRRAFLMELARRKGFRVENELS
 DSVTHIVAENNSGSDVLEWLQLQNIKASSELELLDISWLEIECKGAGKPVEMMGRHQLVVRNRNSSPSPV
 PGSQNVPAFAVKKISQYACQRRTTNNYNQLFTDALDILAENDELRENEGSCCLAFMRASSVLKSLPFF
 ITSMKDTEGIPCLGDKVKSIIIEGIIEDGESSEAKAVLNDERYKSFKLFTSVFGVGLKTAEKWFRMGFR
 35 TLSKIQSDKSLRFTQMOKAGFLYYEDLVSCVNRPEAEAVSMLVKEAVVTFLPDALVTMTGGFRRGKMT
 GHDVDFLITSPEATEDEEQQLLHKVTD FWKQQGLLLYCDILESTFEKFKQPSRKVDALDHFQKCFLLIL

KLDHGRVHSEKSGQEGKGWKAIRVDLVMPYDRRAFALLGWTGSRQFERDLRRYATHERKMMIDNHA
LYDRTKGKTVTISPLDGKVS KLQKALRVFLEAESEEEI FAHLGLDYIEPWERNA

Thus by a process of aligning sequences, it is immediately apparent which regions in the sequences
of terminal transferases from other species correspond to the sequences described herein with
respect to the spotted gar sequence shown in SEQ ID NO 1.

Sequence homology extends to all modified or wild-type members of family X polymerases, such as
DNA Pol μ (also known as DNA polymerase mu or POLM), DNA Pol β (also known as DNA polymerase
beta or POLB), and DNA Pol λ (also known known as DNA polymerase lambda or POLL). It is well
known in the art that all family X member polymerases, of which TdT is a member, either have
terminal transferase activity or can be engineered to gain terminal transferase activity akin to
terminal deoxynucleotidyl transferase (Biochim Biophys Acta. 2010 May; 1804(5): 1136–1150). For
example, when the following human TdT loop1 amino acid sequence

...ESTFEKLR LPSRKVDALDHF...

was engineered to replace the following human Pol μ amino acid residues

...HSCCESPTRLAQQSHMDAF...,

the chimeric human Pol μ containing human TdT loop1 gained robust terminal transferase activity
(Nucleic Acids Res. 2006 Sep; 34(16): 4572–4582).

Furthermore, it was generally demonstrated in US patent application no. 2019/0078065 that family
X polymerases when engineered to contain TdT loop1 chimeras could gain robust terminal
transferase activity. Additionally, it was demonstrated that TdT could be converted into a template-
dependent polymerase through specific mutations in the loop1 motif (Nucleic Acids Research, Jun
2009, 37(14):4642-4656). As it has been shown in the art, family X polymerases can be trivially modified
to either display template-dependent or template-independent nucleotidyl transferase activities.
Therefore, all motifs, regions, and mutations demonstrated in this patent can be trivially extended to
modified X family polymerases to enable modified X family polymerases to incorporate 3'-modified
nucleotides, reversibly terminated nucleotides, and modified nucleotides in general to effect
methods of nucleic acid synthesis.

As a comparison with other family X polymerases, the human Pol μ sequence is shown below:

MLPKRRRRARVGSPSGDAASSTPPSTRFPGVAIYLVEPRMGRSRRRAFLTGLARSKGFRVLDACSSEATH
 VVMEETSAAEEAVSWQERRMAAAPPGCTPPALLDISWLTESLGAGQVPVVECRHRLEVAGPRKGPLSPA
 5 WMPAYACQRPTPLTHHNTGLSEALEILAEAAAGFEGSEGRLLTFCRAASVLKALPSPVTTLSQLQGLPH
 FGEHSSRVVQELLEHGVCCEEVERVRRSERYQTMKLFQTQIFGVGVKTADRWYREGLRTLDDDLREQPQKL
 TQQQKAGLQHHQDLSTPVLRSVDVDALQQVVEEAVGQALPGATVTTLTGGFRRGKLQGHVDVDFLITHPKE
 GQEAGLLPRVMCRLQDQGLILYHQHQHSCCESPTRLAQQSHMDAFERSFCIFRLPQPPGAAVGGSTRP
 CPSWKAVRVDLVVAPVSQFPFALLGWIGSKLQRELRRFSRKEKGLWLNHGLFDPEQKTFQAASEE
 10 DIFRHLGLEYLPPPEQRNA

Modifications which improve the solubility include a modification within the amino acid region
 GLLPRVMCRLQDQGLILYHQHQ shown highlighted in the sequence below.

MLPKRRRRARVGSPSGDAASSTPPSTRFPGVAIYLVEPRMGRSRRRAFLTGLARSKGFRVLDACSSEATH
 VVMEETSAAEEAVSWQERRMAAAPPGCTPPALLDISWLTESLGAGQVPVVECRHRLEVAGPRKGPLSPA
 WMPAYACQRPTPLTHHNTGLSEALEILAEAAAGFEGSEGRLLTFCRAASVLKALPSPVTTLSQLQGLPH
 FGEHSSRVVQELLEHGVCCEEVERVRRSERYQTMKLFQTQIFGVGVKTADRWYREGLRTLDDDLREQPQKL
 TQQQKAGLQHHQDLSTPVLRSVDVDALQQVVEEAVGQALPGATVTTLTGGFRRGKLQGHVDVDFLITHPKE
 20 GQEAGLLPRVMCRLQDQGLILYHQHQHSCCESPTRLAQQSHMDAFERSFCIFRLPQPPGAAVGGSTRP
 CPSWKAVRVDLVVAPVSQFPFALLGWIGSKLQRELRRFSRKEKGLWLNHGLFDPEQKTFQAASEE
 DIFRHLGLEYLPPPEQRNA

Modifications which improve the incorporation of modified nucleotides can be at one or more of
 25 selected regions shown below. The second modification can be selected from one or more of the
 amino acid regions VAIY, LGA, LTHHNTG, SEGRLLTFCRAA, SPVTTL, EHSSR, EGLRT, REQ, KLTQQQKA,
 STPVL, DVDA, AVGQA, GKQ, HPKEGQ, YHQHQ, DAFER, VAPVSQ, FSRKEKGLWLNHGLFDPEQK, AND
 EYLPP shown highlighted in the sequence below.

MLPKRRRRARVGSPSGDAASSTPPSTRFPGVAIYLVEPRMGRSRRRAFLTGLARSKGFRVLDACSSEATH
 VVMEETSAAEEAVSWQERRMAAAPPGCTPPALLDISWLTESLGAGQVPVVECRHRLEVAGPRKGPLSPA
 WMPAYACQRPTPLTHHNTGLSEALEILAEAAAGFEGSEGRLLTFCRAASVLKALPSPVTTLSQLQGLPH
 FGEHSSRVVQELLEHGVCCEEVERVRRSERYQTMKLFQTQIFGVGVKTADRWYREGLRTLDDDLREQPQKL
 TQQQKAGLQHHQDLSTPVLRSVDVDALQQVVEEAVGQALPGATVTTLTGGFRRGKLQGHVDVDFLITHPKE
 35 GQEAGLLPRVMCRLQDQGLILYHQHQHSCCESPTRLAQQSHMDAFERSFCIFRLPQPPGAAVGGSTRP
 CPSWKAVRVDLVVAPVSQFPFALLGWIGSKLQRELRRFSRKEKGLWLNHGLFDPEQKTFQAASEE
 DIFRHLGLEYLPPPEQRNA

Thus by a process of aligning sequences, it is immediately apparent which regions in the sequences of all family X polymerases from any species correspond to the sequences described herein with respect to the spotted gar sequence shown in SEQ ID NO 1.

Furthermore, the A family polymerase, DNA Pol θ (also known as DNA polymerase theta or POLQ) was demonstrated to display robust terminal transferase capability ([eLife](#). 2016; 5: e13740). DNA Pol θ was also demonstrated to be useful in methods of nucleic acid synthesis (GB patent application no. 2553274). In US patent application no. 2019/0078065, it was demonstrated that chimeras of DNA Pol θ and family X polymerases could be engineered to gain robust terminal transferase activity and become competent for methods of nucleic acid synthesis. Therefore, all motifs, regions, and mutations demonstrated in this patent can be trivially extended to modified A family polymerases, especially DNA Pol θ , to enable modified A family polymerases to incorporate 3'-modified nucleotides, reversibly terminated nucleotides, and modified nucleotides in general to effect methods of nucleic acid synthesis.

DETAILED DESCRIPTION OF THE INVENTION

Described herein are modified terminal deoxynucleotidyl transferase (TdT) enzymes. Terminal transferase enzymes are ubiquitous in nature and are present in many species. Many known TdT sequences have been reported in the NCBI database. The sequences described herein are modified from the sequence of the Spotted Gar, but the corresponding changes can be introduced into the homologous sequences from other species. Homologous amino acid sequences of Pol μ , Pol β , Pol λ , and Pol θ or the homologous amino acid sequence of X family polymerases also possess terminal transferase activity. References to terminal transferase also include homologous amino acid sequences of Pol μ , Pol β , Pol λ , and Pol θ or the homologous amino acid sequence of X family polymerases where such sequences possess terminal transferase activity.

Disclosed herein is a modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising at least one amino acid modification when compared to a wild type sequence, wherein the modification is selected from one or more of the amino acid regions WLLNRLINRLQNQGILLYDIV, VAIF, MGA, MENHNQI, SEGPCLA FMRA, HAISSS, DQTKA, KGFHS, QADNA, HFTKMQK, SAAVCK, EAQA, TVRLI, GKEC, TPENMGK, DHFQK, LAAG, APPVDNF, FARHERKMMLLDNHALYDKTKK, and DYIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species.

Described herein is a modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising at least the sequence ID 729:

TVSQYACQRRRTMENHNQIFTDAFAILAENAEFNESEGPCLAFMRAASLLKSLPHAISSSKDLEGLPC
 5 LGDQTKAVIEDILEYGQCSKVQDVLCDTRYQTIKLFTSVFVGLKTAEKWYRKGFHSLEEVQADNAIH
 FTKMQKAGFLYYDDISAAVCKAEAQAIGQIVEETVRLIAPDAIVTLTGGFRRGKECGHDVDFLITTP
 MGKEVWLLNRLINRLQNQGILLYDIVESTFDKTRLPCRKFEMDHFQKCFAI IKLKKELAAGRVQKD
 WKAIRVDFVAPPVDNFALFALLGWTGSRQFERDLRRFARHERKMMLLDNHALYDKTKKIFLPAKTEEDIF
 AHLGLDYIDPWQRNA

10

or the equivalent homologous region in other species, wherein the sequence has one or more amino acid modifications in one or more of the amino acid regions WLLNRLINRLQNQGILLYDIV, MENHNQI, SEGPCLAFMRA, HAIS, DQTKA, KGFHS, QADNA, HFTKMQK, SAAVCK, EAQA, TVRLI, GKEC, TPMEGK, DHFQK, LAAG, APPVDNF, FARHERKMMLLDNHALYDKTKK, and DYIDP of the sequence. The sequence
 15 above of 355 amino acids can be attached to other amino acids without affecting the function of the enzyme. For example there can be a further N-terminal sequence that is incorporated simply as a protease cleavage site, for example the sequence MENLYFQG.

20

Disclosed is a modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising at least one amino acid modification when compared to a wild type sequence SEQ ID NO 1 or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species, wherein the modification is selected from one or more of the amino acid regions WLLNRLINRLQNQGILLYDI, VAIF, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA, TVR, KEC, TPMEGK, DHFQ, LAAG, APPVDN, FARHERKMMLLDNHA, and YIDP of the sequence of SEQ ID
 25 NO 1 or the homologous regions in other species.

30

Further disclosed is a modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising at least two amino acid modifications when compared to a wild type sequence SEQ ID NO 1 or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species, wherein;

- a. a first modification is within the amino acid region WLLNRLINRLQNQGILLYDIV of the sequence of SEQ ID NO 1 or the homologous region in other species; and
- b. a second modification is selected from one or more of the amino acid regions VAIF, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA, TVR, KEC, TPMEGK, DHFQ, LAAG,

APPVDN, FARHERKMMLDNHA, and YIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species.

When compared to the sequence of *Bos taurus* (cow) TdT; SEQ ID NO 2,

- 5 a. a first modification is within the amino acid region QLLPKVINLWEKKGLLYYDLV of the sequence of SEQ ID NO 2 or the homologous region in other species; and
- b. a second modification is selected from one or more of the amino acid regions LVLF, MGA, NNYNH, FMRA, FTI, VKC, FRS, MSDKT, MQK, EAEA, AVW, KKI, SPGSAE, MCP, YATHERKMMLDNHA, and YIEP of the sequence of SEQ ID NO 2 or the homologous regions in other species.

10

When compared to the sequence of *Mus musculus* (mouse) TdT; SEQ ID NO 3,

- a. a first modification is within the amino acid region QLLHKVTDFWKQGLLYCDIL of the sequence of SEQ ID NO 3 or the homologous region in other species; and
- 15 b. a second modification is selected from one or more of the amino acid regions LVLF, MGA, NNYNQ, FMRA, FPI, VKS, FRT, QSDKS, MQK, VSCVNR, EAEA, AVV, KMT, SPEATE, DHFQ, MCPYDR, YATHERKMMLDNHA, and YIEP of the sequence of SEQ ID NO 3 or the homologous regions in other species.

20 The modifications can be chosen from any amino acid that differs from the wild type sequence. The amino acid can be a naturally occurring amino acid. The modified amino acid can be selected from ala, arg, asn, asp, cys, gln, glu, gly, his, ile, leu lys, met, phe, pro, ser, thr, trp, val, and sec.

For the purposes of brevity, the modifications are further described in relation to SEQ ID NO 1, but the modifications are applicable to the sequences from other species, for example those sequences
25 listed above having sequences in the NCBI database. The sequence modifications also apply to truncated versions of SEQ ID NO 1.

The sequences can be modified at positions in addition to those regions described. Embodiments on the invention may include for example sequences having modifications to amino acids outside the
30 defined positions, providing those sequences retain terminal transferase activity. Embodiments of the invention may include for example sequences having truncations of amino acids outside the defined positions, providing those sequences retain terminal transferase activity. For example the sequences may be BRCT truncated as described in application WO2018215803 where amino acids are removed from the N-terminus whilst retaining or improving activity. Alterations, additions,

insertions or deletions or truncations to amino acid positions outside the claimed regions are therefore within the scope of the invention, providing that the claimed regions as defined are modified as claimed. The sequences described herein refer to TdT enzymes, which are typically at least 300 amino acids in length. All sequences described herein can be seen as having at least 300 amino acids. The claims do not cover peptide fragments or sequences which do not function as terminal transferase enzymes.

The modification within the region WLLNRLINRLQNQGILLYYDIV or the corresponding region from other species help improve the solubility of the enzyme. The modification within the amino acid region WLLNRLINRLQNQGILLYYDIV can be at one or more of the underlined amino acids.

Particular changes can be selected from W-Q, N-P, R-K, L-V, R-L, L-W, Q-E, N-K, Q-K or I-L.

The sequence WLLNRLINRLQNQGILLYYDIV can be altered to QLLPKVINLWEKKGLLYYDLV.

The second modification improves incorporation of nucleotides having a modification at the 3' position in comparison to the wild type sequence. The second modification can be selected from one or more of the amino acid regions VAIF, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA, TVR, KEC, TPENMGK, DHFQ, LAAG, APPVDN, FARHERKMLLDNHA, and YIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species. The second modification can be selected from two or more of the amino acid regions VAIF, EDN, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA, TVR, KEC, TPENMGK, DHFQ, LAAG, APPVDN, FARHERKMLLDNHA, and YIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species shown highlighted in the sequence below.

MLHIPIFPPIKKRQKLPEsrNSCKYEVKFSEVAIFLVERKMGSSRRKFLTNLARSKGFRIEDVLSDAV
 THVVAEDNSADELWQWLQNSSLGDLskIEVLDisWFTECHGAGKPVQVEARHCLVKSCPVIDQYLEPS
 TVETVSQYACQRRtTMEENHNQIFtdaFaiLAENAEfNESEGpCLAFMRAASLLKSLPHAISSSKDLEG
 LPCLGDQTHAVIEDILEYGQCSKVQDVLCDdRYQTIKLFTSVFGVGLKTAekWYRKGFHSLeeVQADN
 AIHFTKMGAGFLYYDDISAAVCKAEQAIGQIVEETVRLIAPDAIVTLTGgFRRGKEGHdVDFLIT
 TPENMGKEVWLLNRLINRLQNQGILLYYDIVESTFDKTRLPCRKFeamDHFQKCFaiIKLkKELAAgRV
 QKDWKAIRVDFVAPPVDNIFAfALLGWTGSRQFERDLRRFARHERKMLLDNHAlYDKTKKIFLPakTEE
 DIFaHLGLDYIDFWQRNA

The identified positions commence at positions V32, M108, F182, T212, D271, M279, E298, A421, L456, Y486. Modifications disclosed herein contain at least one modification at the defined positions.

5 In the sequence below the modified regions are numbered thus

MLHIPIFPPIKKRQKLPESRNSCKYEVKFSEVAIFLVERKMGSSRRKFLTNLARSKGFRIEDVLSDAV
 THVVAEDNSADELWQWLQNSSLGDLISKIEVLDISWFTECHGAGKPVQVEARHCLVKSCPVIDQYLEPS
 TVETVSQYACQRRTMENHNQIFTDFAFAILAENAEFNESEGPCLAFMRAASLLKSLPHAISSSKDLEG
 10 LPCLGDQTKAVIEDILEYGQCSKVQDVLCDTRYQTIKLFTSVFGVGLKTAEKWYRKGFSHLEEVQADN
 HFTKMQKAGFLYYDDISAAVCKAEQAIGQIVEETVRLIAPDAIVTLTGGRGKECGHGDVDFLIT
 TPENMGKEVWLLNRLINRLQNGILLYYDIVESTFDKTRLPCRKFEMDHFQKCFAIKLLKKEIAGGRV
 QKDWKAIRVDFVAPPVDNFAFALLGWTGSRQFERDLRRFARHERKMLLDNHALYDKTKKIFLPAKTEE
 DIFAHGLDYIDPWQRNA

15

WLLNRLINRLQNGILLYYDIV, 349 to 370

VAIF, 32 to 35

MGA, 108 to 110

MENHNQI, 152 to 158

20 SEGPCLAFMRA, 175 to 185

HAISSS, 194 to 199

DQTKA, 210 to 214

KGFSH, 260 to 264

QADNA, 269 to 273

25 HFTKMQK, 275 to 281

SAAVCK, 291 to 296

EAQA, 298 to 301

TVRLI, 309 to 313

GKEC, 328 to 331

30 TPENMGK, 341 to 346

DHFQK, 388 to 392

LAAG, 403 to 406

APPVDNF, 421 to 427

FARHERKMLLDNHALYDKTKK, 447 to 467

35 DYIDP, 485 to 489

The modified amino acid can be in the region FMRA. The modified amino acid can be in the region QADNA. The modified amino acid can be in the region EAQA. The modified amino acid can be in the region APP. The modified amino acid can be in the region LDNHA. The modified amino acid can be in the region YIDP. The region FARHERKMILLDNHA is advantageous for removing substrate biases in modifications. The FARHERKMILLDNHA region appears highly conserved across species.

The modification selected from one or more of the amino acid regions FMRA, QADNA, EAQA, APP, FARHERKMILLDNHA, and YIDP can be at the underlined amino acid(s).

Rather than the invention being described by modifications in defined domains, the sequences can be defined by modifications at certain defined amino acids. Described herein are modified terminal deoxynucleotidyl transferase (TdT) enzymes comprising amino acid modifications when compared to a wild type sequence SEQ ID NO 1 or a truncated version thereof or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species, wherein the amino acid is modified at one or more of the amino acids A53, V68, V71, E97, I101, G109, Q115, V116, S125, T137, Q143, M152, N154, H155, Q157, I158, I165, N169, S175, G177, C179, L180, A181, M183, A195, S197, S198, S199, K200, D210, Q211, T212, K213, A214, E217, T239, K260, F262, S264, Q269, D271, N272, A273, H275, T277, K281, S291, K296, Q300, T309, R311, L312, I313, G328, E330, C331, T341, P342, E343, M344, G345, K346, N352, R353, L354, I355, N356, L358, N360, Q361, G362, I363, Y366, Y367, D368, V370, H389, K392, L403, A405, G406, D411, A421, P422, V424, N426, F427, R438, F447, R452, K453, L455, Y462, K464, T465, K467, D485, I487 and/or D488.

Described herein are modified terminal deoxynucleotidyl transferase (TdT) enzymes comprising amino acid modifications when compared to a wild type sequence SEQ ID NO 1 or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species, wherein the amino acid is modified at one or more of the amino acids

V32, A33, I34, F35, A53, V68, V71, E97, I101, M108, G109, A110, Q115, V116, S125, T137, Q143, M152, E153, N154, H155, N156, Q157, I158, I165, N169, N173, S175, E176, G177, P178, C179, L180, A181, F182, M183, R184, A185, L188, H194, A195, I196, S197, S198, S199, K200, E203, G204, D210, Q211, T212, K213, A214, I216, E217, D218, L220, Y222, V228, D230, Q238, T239, L242, L251, K260, G261, F262, H263, S264, L265, E267, Q269, A270, D271, N272, A273, H275, F276, T277, K278, M279, Q280, K281, S291, A292, A293, V294, C295, K296, E298, A299, Q300, A301, Q304, I305, T309, V310, R311, L312, I313, A314, I318, V319, T320, G328, K329, E330, C331, L338, T341, P342, E343, M344,

G345, K346, W349, L350, L351, N352, R353, L354, I355, N356, R357, L358, Q359, N360, Q361, G362, I363, L364, L365, Y366, Y367, D368, I369, V370, K376, T377, C381, K383, D388, H389, F390, Q391, K392, F394, I397, K398, K400, K401, E402, L403, A404, A405, G406, R407, D411, A421, P422, P423, V424, D425, N426, F427, A430, R438, F447, A448, R449, H450, E451, R452, K453, M454, L455, L456, D457, N458, H459, A460, L461, Y462, D463, K464, T465, K466, K467, T474, D477, D485, Y486, I487, D488, P489.

Specific amino acid changes can include any one of A53G, V68I, V71I, E97A, I101V, G109E, G109R, Q115E, V116I, V116S, S125R, T137A, Q143P, M152L, M152T, N154H, H155C, H155N, H155R, Q157K, Q157R, I158V, I158L, I158M, I165V, N169R, N173R, S175N, G177S, G177V, G177D, C179A, C179S, L180V, A181E, M183R, L188V, A195T, A195P, S197I, S197L, S198R, S198C, S199L, K200R, E203Q, G204D, D210E, Q211R, T212S, K213S, A214R, A214C, A214G, I216M, E217Q, D218E, L220I, L220F, L220Y, Y222C, V228A, D230E, Q238K, T239S, L242Q, L251P, K260M, F262L, L265F, S264T, E267D, Q269K, D271N, D271E, N272K, A273S, A273T, H275Q, T277S, K281R, S291N, S291T, K296R, Q300D, Q304R, Q304C, Q304H, I305V, T309A, R311W, R311H, L312G, L312K, L312A, I313M, A314T, I318L, V319L, T320A, G328A, E330N, E330S, C331I, L338I, T341S, P342A, E343G, E343Q, M344A, M344Q, M344K, G345R, K346R, N352Q, K353R, K353D, V354I, V354L, I355V, I355M, N356R, N356D, W358L, N360K, Q361K, G362E, I363L, Y366F, Y367C, D368H, D368E, V370I, K376I, T377S, T377L, T377A, C381S, K383R, H389N, H389A, K392Q, F394W, I397L, K398R, K400E, K400N, K401A, K401Q, E402Q, L403Q, L403R, A405D, G406R, R407C, D411N, A421L, A421M, A421V, P422A, P422C, V424Y, V424I, V424A, N426R, N426C, F427Y, A430T, R438K, F447W, R452K, L455I, K453R, Y462F, K464R, T465R, K467R, T474S, D477E, D485E, I487V, and/or D488P.

Amino acid changes include any two or more of A53G, V68I, V71I, E97A, I101V, G109E, G109R, Q115E, V116I, V116S, S125R, T137A, Q143P, M152L, M152T, N154H, H155C, H155N, H155R, Q157K, Q157R, I158V, I158L, I158M, I165V, N169R, N173R, S175N, G177S, G177V, G177D, C179A, C179S, L180V, A181E, M183R, L188V, A195T, A195P, S197I, S197L, S198R, S198C, S199L, K200R, E203Q, G204D, D210E, Q211R, T212S, K213S, A214R, A214C, A214G, I216M, E217Q, D218E, L220I, L220F, L220Y, Y222C, V228A, D230E, Q238K, T239S, L242Q, L251P, K260M, F262L, L265F, S264T, E267D, Q269K, D271N, D271E, N272K, A273S, A273T, H275Q, T277S, K281R, S291N, S291T, K296R, Q300D, Q304R, Q304C, Q304H, I305V, T309A, R311W, R311H, L312G, L312K, L312A, I313M, A314T, I318L, V319L, T320A, G328A, E330N, E330S, C331I, L338I, T341S, P342A, E343G, E343Q, M344A, M344Q, M344K, G345R, K346R, N352Q, K353R, K353D, V354I, V354L, I355V, I355M, N356R, N356D, W358L, N360K, Q361K, G362E, I363L, Y366F, Y367C, D368H, D368E, V370I, K376I, T377S, T377L, T377A,

C381S, K383R, H389N, H389A, K392Q, F394W, I397L, K398R, K400E, K400N, K401A, K401Q, E402Q, L403Q, L403R, A405D, G406R, R407C, D411N, A421L, A421M, A421V, P422A, P422C, V424Y, V424I, V424A, N426R, N426C, F427Y, A430T, R438K, F447W, R452K, L455I, K453R, Y462F, K464R, T465R, K467R, T474S, D477E, D485E, I487V, and/or D488P.

5

The modification of QADNA to KADKA, QADKA, KADNA, QADNS, KADNT, or QADNT is advantageous for the incorporation of 3'-O-modified nucleoside triphosphates to the 3'-end of nucleic acids and removing substrate biases during the incorporation of modified nucleoside triphosphates. The modification of APPVDN to MCPVDN, MPPVDN, ACPVDR, VPPVDN, LPPVDR, ACPYDN, LCPVDN, or
 10 MAPVDN is advantageous for the incorporation of 3'-O-modified nucleoside triphosphates to the 3'-end of nucleic acids and removing substrate biases during the incorporation of modified nucleoside triphosphates. The modification of FARHERKMILLDRHA to WARHERKMILDNHA, FARHERKMILDNHA, WARHERKMILLDNHA, FARHERKMILLDRHA, or FARHEKKMILLDNHA is also advantageous for the incorporation of 3'-O-modified nucleoside triphosphates to the 3'-end of nucleic acids and removing
 15 substrate biases during the incorporation of modified nucleoside triphosphates.

The modification can be selected from one or more of the following sequences FRRA, QADKA, EADA, MPP, FARHERKMILLDRHA, and YIPP. Included is a modified terminal deoxynucleotidyl transferase (TdT) enzyme wherein the second modification is selected from two or more of the following
 20 sequences FRRA, QADKA, EADA, MPP, FARHERKMILLDRHA, and YIPP. Included is a modified terminal deoxynucleotidyl transferase (TdT) enzyme wherein the second modification contains each of the following sequences FRRA, QADKA, EADA, MPP, FARHERKMILLDRHA, and YIPP.

Crystal structures shown herein show the following domains, which may be preferred as domains to
 25 modify:

Figure 5. MENHNQI motif (152 to 158), which includes the ENHNQ motif

Figure 6. SEGPCLAFMRA (175 to 185), which includes the FMRA motif, highlighted in black.

30

Figure 7. HFTKMQK motif (275 to 281), which includes the MQK motif, highlighted in black.

Figure 8. SAAVCK motif (291 to 296) highlighted in black.

Figure 9. GKEC motif (328 to 331), which includes the KEC motif, highlighted in black.

Figure 10. DHFQK motif (388 to 392), which includes the DHFQ motif, highlighted in black.

5 **Figure 11.** FARHERKMLLDNHA motif (447 to 460) highlighted in black spheres. Additionally, motif FARHERKMLLDNHA is mutated to FARHERKMLLDRHA.

Figure 12. FARHERKMLLDNHALYDKTKK motif (447 to 467) highlighted in black spheres. Additionally, motif FARHERKMLLDNHALYDKTKK is mutated to FARHERKMLLDRHALYDKTKK.

10

Figure 13. DYIDP motif (485 to 489), which contains the YIDP motif, highlighted in black spheres.

Figure 14. YYDIV motif (366 to 370) highlighted in black spheres.

15 **Figure 15.** APPVDNF, which contains the APPVDN motif (421 to 427), motif highlighted in black spheres. In the structure, the APPVDNF motif is shown with an Ala to Met mutation to MPPVDNF.

In order to aid purification of the expressed sequence, the amino acid can be further modified. For example the amino acid sequence can contain one or more further histidine residues at the terminus. Included is a modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising any one of SEQ ID NOs 4 to 173 or a truncated version thereof. Sequences 4-173 are the full length sequences derived from the spotted gar. Included is a modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising any one of SEQ ID NOs 174 to 343. Sequences 174 to 343 are N-truncated sequences as spotted gar/bovine chimeras. Sequences 344 to 727 are spotted Gar sequences in truncated form. Additionally, for these sequences, there is an N-terminal sequence that is incorporated simply as a protease cleavage site (MENLYFQG...).

25

Also disclosed is a method of nucleic acid synthesis, which comprises the steps of:

- (a) providing an initiator oligonucleotide;
- 30 (b) adding a 3'-blocked nucleoside triphosphate to said initiator oligonucleotide in the presence of a terminal deoxynucleotidyl transferase (TdT) as defined herein;
- (c) removal of all reagents from the initiator oligonucleotide;
- (d) cleaving the blocking group in the presence of a cleaving agent; and
- (e) removal of the cleaving agent.

The method can add greater than 1 nucleotide by repeating steps (b) to (e).

References herein to 'nucleoside triphosphates' refer to a molecule containing a nucleoside (i.e. a base attached to a deoxyribose or ribose sugar molecule) bound to three phosphate groups.

5 Examples of nucleoside triphosphates that contain deoxyribose are: deoxyadenosine triphosphate (dATP), deoxyguanosine triphosphate (dGTP), deoxycytidine triphosphate (dCTP) or deoxythymidine triphosphate (dTTP). Examples of nucleoside triphosphates that contain ribose are: adenosine triphosphate (ATP), guanosine triphosphate (GTP), cytidine triphosphate (CTP) or uridine triphosphate (UTP). Other types of nucleosides may be bound to three phosphates to form
10 nucleoside triphosphates, such as naturally occurring modified nucleosides and artificial nucleosides.

Therefore, references herein to '3'-blocked nucleoside triphosphates' refer to nucleoside triphosphates (e.g., dATP, dGTP, dCTP or dTTP) which have an additional group on the 3' end which prevents further addition of nucleotides, i.e., by replacing the 3'-OH group with a protecting group.

15

It will be understood that references herein to '3'-block', '3'-blocking group' or '3'-protecting group' refer to the group attached to the 3' end of the nucleoside triphosphate which prevents further nucleotide addition. The present method uses reversible 3'-blocking groups which can be removed by cleavage to allow the addition of further nucleotides. By contrast, irreversible 3'-blocking groups
20 refer to dNTPs where the 3'-OH group can neither be exposed nor uncovered by cleavage.

The 3'-blocked nucleoside 5'-triphosphate can be blocked by any chemical group that can be unmasked to reveal a 3'-OH. The 3'-blocked nucleoside triphosphate can be blocked by a 3'-O-azidomethyl, 3'-aminooxy, 3'-O-(N-oxime) (3'-O-N=CR₁R₂, where R₁ and R₂ are each a C1-C3 alkyl
25 group, for example CH₃, such that the oxime can be O-N=C(CH₃)₂ (N-acetoneoxime)), 3'-O-allyl group, 3'-O-cyanoethyl, 3'-O-acetyl, 3'-O-nitrate, 3'-phosphate, 3'-O-acetyl levulinic ester, 3'-O-tert butyl dimethyl silane, 3'-O-trimethyl(silyl)ethoxymethyl, 3'-O-ortho-nitrobenzyl, and 3'-O-para-nitrobenzyl.

30 The 3'-blocked nucleoside 5'-triphosphate can also be blocked by any chemical group that can be directly utilized in chemical ligations, such as copper-catalyzed or copper-free azide-alkyne click reactions and tetrazine-alkene click reactions. The 3'-blocked nucleoside triphosphate can include chemical moieties containing an azide, alkyne, alkene, and tetrazine.

References herein to 'cleaving agent' refer to a substance which is able to cleave the 3'-blocking group from the 3'-blocked nucleoside triphosphate. In one embodiment, the cleaving agent is a chemical cleaving agent. In an alternative embodiment, the cleaving agent is an enzymatic cleaving agent. The cleaving can be done in a single step, or can be a multi-step process, for example to transform an oxime (such as for example 3'-O-(N-oxime), 3'-O-N=C(CH₃)₂, into aminooxy (O-NH₂), followed by cleaving the aminooxy to OH.

It will be understood by the person skilled in the art that the selection of cleaving agent is dependent on the type of 3'-nucleotide blocking group used. For instance, tris(2-carboxyethyl)phosphine (TCEP) or tris(hydroxypropyl)phosphine (THPP) can be used to cleave a 3'-O-azidomethyl group, palladium complexes can be used to cleave a 3'-O-allyl group, or sodium nitrite can be used to cleave a 3'-aminooxy group. Therefore, in one embodiment, the cleaving agent is selected from: tris(2-carboxyethyl)phosphine (TCEP), a palladium complex or sodium nitrite.

In one embodiment, the cleaving agent is added in the presence of a cleavage solution comprising a denaturant, such as urea, guanidinium chloride, formamide or betaine. The addition of a denaturant has the advantage of being able to disrupt any undesirable secondary structures in the DNA. In a further embodiment, the cleavage solution comprises one or more buffers. It will be understood by the person skilled in the art that the choice of buffer is dependent on the exact cleavage chemistry and cleaving agent required.

References herein to an 'initiator oligonucleotide' or 'initiator sequence' refer to a short oligonucleotide with a free 3'-end which the 3'-blocked nucleoside triphosphate can be attached to. In one embodiment, the initiator sequence is a DNA initiator sequence. In an alternative embodiment, the initiator sequence is an RNA initiator sequence.

References herein to a 'DNA initiator sequence' refer to a small sequence of DNA which the 3'-blocked nucleoside triphosphate can be attached to, i.e., DNA will be synthesised from the end of the DNA initiator sequence.

In one embodiment, the initiator sequence is between 5 and 50 nucleotides long, such as between 5 and 30 nucleotides long (i.e. between 10 and 30), in particular between 5 and 20 nucleotides long (i.e., approximately 20 nucleotides long), more particularly 5 to 15 nucleotides long, for example 10 to 15 nucleotides long, especially 12 nucleotides long.

In one embodiment, the initiator sequence is single-stranded. In an alternative embodiment, the initiator sequence is double-stranded. It will be understood by persons skilled in the art that a 3'-overhang (i.e., a free 3'-end) allows for efficient addition.

- 5 In one embodiment, the initiator sequence is immobilised on a solid support. This allows TdT and the cleaving agent to be removed (in steps (c) and (e), respectively) without washing away the synthesised nucleic acid. The initiator sequence may be attached to a solid support stable under aqueous conditions so that the method can be easily performed via a flow setup.
- 10 In one embodiment, the initiator sequence is immobilised on a solid support via a reversible interacting moiety, such as a chemically-cleavable linker, an antibody/immunogenic epitope, a biotin/biotin binding protein (such as avidin or streptavidin), or glutathione-GST tag. Therefore, in a further embodiment, the method additionally comprises extracting the resultant nucleic acid by removing the reversible interacting moiety in the initiator sequence, such as by incubating with
- 15 proteinase K.

- In one embodiment, the initiator sequence contains a base or base sequence recognisable by an enzyme. A base recognised by an enzyme, such as a glycosylase, may be removed to generate an abasic site which may be cleaved by chemical or enzymatic means. A base sequence may be
- 20 recognised and cleaved by a restriction enzyme.

- In a further embodiment, the initiator sequence is immobilised on a solid support via a chemically-cleavable linker, such as a disulfide, allyl, or azide-masked hemiaminal ether linker. Therefore, in one embodiment, the method additionally comprises extracting the resultant nucleic acid by cleaving the
- 25 chemical linker through the addition of tris(2-carboxyethyl)phosphine (TCEP) or dithiothreitol (DTT) for a disulfide linker; palladium complexes or an allyl linker; or TCEP for an azide-masked hemiaminal ether linker.

- In one embodiment, the resultant nucleic acid is extracted and amplified by polymerase chain reaction using the nucleic acid bound to the solid support as a template. The initiator sequence
- 30 could therefore contain an appropriate forward primer sequence and an appropriate reverse primer could be synthesised.

In one embodiment, the terminal deoxynucleotidyl transferase (TdT) of the invention is added in the presence of an extension solution comprising one or more buffers (e.g., Tris or cacodylate), one or more salts (e.g., Na⁺, K⁺, Mg²⁺, Mn²⁺, Cu²⁺, Zn²⁺, Co²⁺, etc. all with appropriate counterions, such as Cl) and inorganic pyrophosphatase (e.g., the *Saccharomyces cerevisiae* homolog). It will be understood that the choice of buffers and salts depends on the optimal enzyme activity and stability. The use of an inorganic pyrophosphatase helps to reduce the build-up of pyrophosphate due to nucleoside triphosphate hydrolysis by TdT. Therefore, the use of an inorganic pyrophosphatase has the advantage of reducing the rate of (1) backwards reaction and (2) TdT strand dismutation.

10 In one embodiment, step (b) is performed at a pH range between 5 and 10. Therefore, it will be understood that any buffer with a buffering range of pH 5-10 could be used, for example cacodylate, Tris, HEPES or Tricine, in particular cacodylate or Tris.

15 In one embodiment, step (d) is performed at a temperature less than 99 °C, such as less than 95 °C, 90 °C, 85 °C, 80 °C, 75 °C, 70 °C, 65 °C, 60 °C, 55 °C, 50 °C, 45 °C, 40 °C, 35 °C, or 30 °C. It will be understood that the optimal temperature will depend on the cleavage agent utilised. The temperature used helps to assist cleavage and disrupt any secondary structures formed during nucleotide addition.

20 In one embodiment, steps (c) and (e) are performed by applying a wash solution. In one embodiment, the wash solution comprises the same buffers and salts as used in the extension solution described herein. This has the advantage of allowing the wash solution to be collected after step (c) and recycled as extension solution in step (b) when the method steps are repeated.

25 Also disclosed is a kit comprising a terminal deoxynucleotidyl transferase (TdT) as defined herein in combination with an initiator sequence and one or more 3'-blocked nucleoside triphosphates.

The invention includes the nucleic acid sequence used to express the modified terminal transferase. Included within the invention are the codon-optimized cDNA sequences which express the modified terminal transferase. Included are the codon-optimized cDNA sequences for each of the protein variants (SEQ ID NOs 4-727).

The nucleic acid sequence may be the sequence below (ID 728):

ATGCTGCATATTCGGATTTTCCCGCCGATTAAGAAGCGCCAGAAGCTGCCGGAGAGCAGAAA
 TTCCTGCAAATATGAAGTCAAGTTCTCTGAAGTTGCGATTTTCTGGTTGAGCGCAAGATGG
 GTAGCAGCCGTCGCAAGTTTCTGACCAACCTGGCGCGTAGCAAGGGCTTTCGTATCGAAGAT
 GTTTTGAGCGATGCGGTGACCCACGTTGTGCGCCGAGGACAACAGCGCGGACGAGCTGTGGCA
 5 GTGGCTCCAAAATTCCAGCCTGGGTGATCTGTCCAAAATTGAAGTCCTGGACATTTTCATGGT
 TCACCGAGTGTATGGGTGCCGGTAAACCTGTGCAAGTCGAGGCCCGTCACTGCCTGGTCAAG
 AGCTGTCCGGTGATCGATCAGTACCTTGAGCCATCCACGGTTGAAACGGTCAGCCAATATGC
 GTGCCAACGTCGTACGACCATGGAGAACCATAACCAGATTTTTCACGGACGCATTTGCCATCC
 TGGCGGAGAATGCTGAGTTTAAACGAGAGCGAGGGTCCGTGCCTGGCGTTTATGCGTGACGCT
 10 AGCCTGCTGAAATCGTTGCCGCACGCAATTTCTAGCAGCAAAGACCTGGAAGGTCTGCCGTG
 CCTGGGCGACCAGACGAAAGCCGTAATCGAGGACATCTTAGAATACGGTCAGTGCAGCAAAG
 TTCAAGATGTGTTGTGTGATGACCGTTACCAGACCATCAAACGTTCACGAGCGTTTTTGGC
 GTGGGCTGAAAACCTGCAGAGAAGTGGTACCGCAAGGGTTTCCACTCTCTGGAAGAAGTGCA
 AGCGGATAATGCGATCCATTTACGAAAATGCAAAAAGCGGGCTTCCTGTACTATGATGACA
 15 TTAGCGCAGCAGTTTGTAAAGGCTGAGGCCCAAGCGATCGGTCAGATCGTCGAAGAGACTGTG
 CGTCTGATCGCACCGGATGCGATTGTCACCGTGACGGGCGGTTTTCGCCGTGGTAAAGAGTG
 CGGTCACGACGTTGACTTTCTGATCACCACCCCGGAAATGGGCAAAGAAGTGTGGCTGCTGA
 ATCGTCTGATTAACCGTCTGCAGAATCAGGGTATTCTGTTGTATTATGACATTGTTGAGAGC
 ACCTTCGACAAAACCCGTCTGCCGTGTGCAAGTTCGAGGCAATGGACCATTTTCAGAAGTG
 20 CTTTGCGATCATTAAGTTGAAGAAAGAACTGGCGGCTGGCCGTGTGCAAAAAGACTGGAAAG
 CGATCCGTGTTGATTTCTGTTGGCTCCGCCGGTCGACAATTTTCGCCTTTGCATTGCTGGGTTGG
 ACCGGTAGCCGTGAGTTCGAGCGTGATCTGCGCCGCTTTGCGCGCCACGAACGTAAGATGCT
 GCTGGATAACCATGCACTGTACGATAAGACCAAAAAGATTTTCCTGCCGGCCAAAACCGAAG
 AAGATATCTTTGCGCACCTGGGCTTGGACTACATCGATCCATGGCAACGCAACGCGTAATGA
 25

The invention includes a cell line producing the modified terminal transferase.

Examples

30 Expression of TdT variants

Briefly, plasmids containing genes encoding terminal transferase enzymes were transformed into
 BL21 *E. coli* competent cells. Starter Luria broth (LB) cultures were grown overnight at 37 °C and
 inoculated into LB expression cultures. Expression cultures were grown to an optical density of 0.6 at
 600 nm and induced by the addition of IPTG to 1 mM. Cultures were induced and grown overnight at
 35 25°C. The following morning, cultures were lysed in a detergent lysis buffer and purified to
 homogeneity by immobilized metal affinity chromatography (IMAC).

Assaying the incorporation of reversible terminators by TdT variants

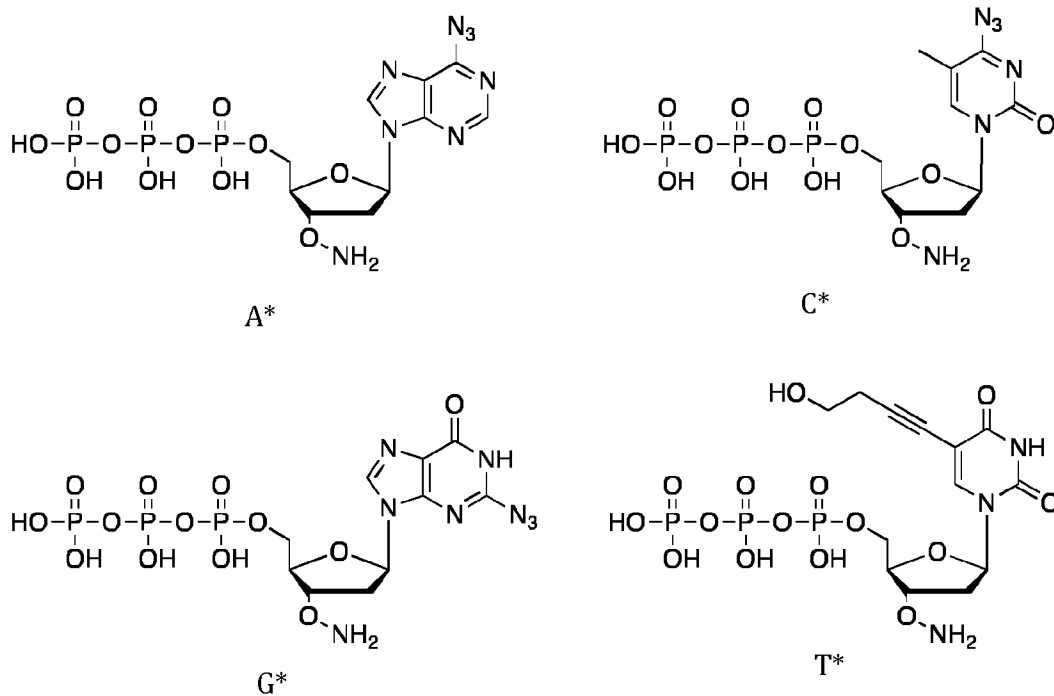
173 terminal transferase enzymes were expressed, purified, and compared against wild-type bovine
 40 TdT (SEQ ID NO 2). Purified engineered TdTs were then used in the following assays: a fluorescently
 labeled 15-nt ssDNA primer was incubated with 1 x TdT buffer (Thermo Fisher Scientific), yeast
 inorganic pyrophosphatase (Sigma-Aldrich, 0.1 mU/μl), 3'-azidomethyl dTTP or 3'-aminooxy dATP,

and engineered TdT (24 µg/µl) for 10 min at 37 °C. Formamide (Fisher Scientific) was then used to quench the reaction, and samples were directly loaded onto and analysed by denaturing polyacrylamide gel electrophoresis. Gels were imaged and resulting gel bands were quantified with a Typhoon scanner (GE).

5

The results from the 173 TdT enzymes (SEQ ID NOs 1-173) are shown in Figure 1.

Assaying the incorporation of base-modified reversible terminators by TdT variants



10

15

A* describes 3'-aminooxy 6-azido 2'-deoxyadenosine 5'-triphosphate; C* describes 3'-aminooxy 4-azido 5-methyl 2'-deoxycytidine 5'-triphosphate; G* describes 3'-aminooxy 2-azido 2'-deoxyguanosine 5'-triphosphate; T* describes 3'-aminooxy 5-(-hydroxy-1-butynyl) 2'-deoxyuridine 5'-triphosphate.

20

192 terminal deoxynucleotidyl transferases (TdT) variants were expressed and purified as described above. The expressed variants had SEQ ID NOs:

25

345, 347, 352, 357, 359, 360, 361, 362, 364, 365, 366, 367, 368, 370, 371, 372, 375, 376, 377, 378, 380, 382, 383, 384, 385, 387, 388, 392, 393, 394, 395, 397, 398, 399, 401, 405, 406, 410, 411, 416, 418, 422, 426, 427, 430, 433, 436, 439, 440, 442, 444, 445, 446, 447, 450, 453, 454, 455, 457, 460, 461, 462, 463, 464, 467, 472, 473, 475, 476, 477, 478, 479, 480, 485, 486, 487, 489, 492, 494, 495, 497, 499, 500, 503, 505, 506, 507, 509, 510, 514, 516, 517, 519, 524, 525, 526, 527, 528, 529, 531,

532, 533, 535, 543, 544, 546, 550, 553, 555, 557, 559, 560, 561, 562, 564, 565, 567, 568, 570, 572, 573, 575, 580, 582, 584, 589, 593, 595, 598, 599, 600, 601, 604, 605, 606, 609, 611, 612, 614, 618, 619, 620, 623, 629, 637, 638, 639, 641, 643, 644, 646, 648, 649, 651, 652, 654, 657, 658, 660, 661, 662, 664, 665, 666, 667, 670, 673, 678, 679, 681, 684, 685, 687, 690, 692, 698, 699, 700, 703, 706, 707, 708, 711, 712, 715, 716, 717, 718, 720, 721, 722, 725.

The base-modified reversibly terminators A*, C*, G*, and T* were supplied as substrates to engineered variants in a single nucleotide incorporation assay. A pool of DNA initiators with degenerate ends (...NNN, where N = A, C, G, T) was immobilised to a solid support and exposed to nucleotide addition mixture (NAM) solutions containing (1) a TdT variant, (2) neutral pH buffer, (3) monovalent salt, (4) cobalt chloride, and (5) 3'-ONH₂-dXTP, where X is the modified nucleobase resulting in A*, C*, G*, or T*. The incubation temperature was 37°C and the reaction time was 2.5 minutes. The solid support was then washed with a high salt solution at neutral pH followed by a low salt solution at neutral pH. The initiator pools transformed into sequencing libraries and analysed by next-generation sequencing (NGS) on an Illumina NextSeq500 with paired-end 30 cycle reads (PE30 reads). Bcl files were converted to fastq files with Illumina's bcl2fastq conversion software and analysed in R. The incorporation efficiency ($\frac{[\text{reads containing N+1 product}]}{[\text{reads containing N initiator}] + [\text{reads containing N+1 product}]}$) was calculated for the addition of each modified base against to all possible initiator contexts. The average incorporation efficiency across all contexts is presented in Figure 3 (top panel). Wild-type bovine and spotted gar TdT activities as it relates to the single nucleotide incorporation of the base-modified reversibly terminated nucleotides A*, C*, G*, and T* are indicated by the dotted line. The dotted line occurs at $\gamma=0$ due to the inability of wild-type bovine or spotted gar TdTs to incorporate these base-modified reversibly terminated nucleotides in our assay. All of the presented mutations in the TdT variants resulted in improvements in TdT modified-base incorporation relative to wild-type bovine and wild-type spotted gar TdTs.

Assaying the multicycling capabilities of TdT variants

Solid support synthesis of an 8-nt nucleic acid sequence using 3'-ONH₂ nucleoside 5'-triphosphates with terminal deoxynucleotidyl transferase (TdT) SEQ ID NOs 1 & 344-727. The TdT variants were expressed and purified as described above. The DNA sequence 5'-ATCGATCG-3' was synthesized by repeated exposure of a solid support bound DNA initiator to nucleotide addition mixture (NAM) solutions containing (1) TdT, (2) neutral pH buffer, (3) monovalent salt, (4) cobalt chloride, and (5) 3'-ONH₂-dNTP, where N is selected from adenine, thymine, cytosine, or guanine. A cycle consists of the

following: (A) NAM solution with a specified A, T, C, or G reversibly terminated nucleotide were incubated for 5 minutes on the solid support at 37°C; (B) the solid support was then washed with a high salt solution at neutral pH; (C) the solid support was then exposed to acidic aqueous sodium nitrite; and (D) the solid support was then washed with the same high salt solution at neutral pH from (B). (A)-(D) were then repeated 7 more times to synthesize the desired 8-nt sequence. Synthesized DNA were analysed by running reactions on a denaturing polyacrylamide gel and quantified by virtue of a fluorophore covalently attached to the DNA initiator. The fraction full length (8-nt species) was determined by taking fluorescence intensity of the 8-nt band and dividing by total lane intensity. Wild-type bovine and spotted gar TdT activities as it relates to fraction full-length are indicated by the dotted line at $y = 0$ indicating that they are incapable of synthesizing any 8-nt product. All mutations included in the TdT variants resulted in improvements in TdT relative to wild-type bovine and spotted gar TdTs. Results are shown in Figure 3 (bottom panel).

Claims

1. A modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising amino acid modifications when compared to a wild type sequence SEQ ID NO 1 or a truncated version thereof or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species or the homologous amino acid sequence of Pol μ , Pol β , Pol λ , and Pol θ of any species or the homologous amino acid sequence of X family polymerases of any species, wherein the amino acid is modified at one or more of the amino acids:

V32, A33, I34, F35, A53, V68, V71, E97, I101, M108, G109, A110, Q115, V116, S125, T137, Q143, M152, E153, N154, H155, N156, Q157, I158, I165, N169, N173, S175, E176, G177, P178, C179, L180, A181, F182, M183, R184, A185, L188, H194, A195, I196, S197, S198, S199, K200, E203, G204, D210, Q211, T212, K213, A214, I216, E217, D218, L220, Y222, V228, D230, Q238, T239, L242, L251, K260, G261, F262, H263, S264, L265, E267, Q269, A270, D271, N272, A273, H275, F276, T277, K278, M279, Q280, K281, S291, A292, A293, V294, C295, K296, E298, A299, Q300, A301, Q304, I305, T309, V310, R311, L312, I313, A314, I318, V319, T320, G328, K329, E330, C331, L338, T341, P342, E343, M344, G345, K346, W349, L350, L351, N352, R353, L354, I355, N356, R357, L358, Q359, N360, Q361, G362, I363, L364, L365, Y366, Y367, D368, I369, V370, K376, T377, C381, K383, D388, H389, F390, Q391, K392, F394, I397, K398, K400, K401, E402, L403, A404, A405, G406, R407, D411, A421, P422, P423, V424, D425, N426, F427, A430, R438, F447, A448, R449, H450, E451, R452, K453, M454, L455, L456, D457, N458, H459, A460, L461, Y462, D463, K464, T465, K466, K467, T474, D477, D485, Y486, I487, D488, P489.

2. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to claim 1 comprising at least one amino acid modification when compared to a wild type sequence or a truncated version thereof, wherein the modification is selected from one or more of the amino acid regions WLLNRLINRLQNQGILLYDIV, VAIF, MGA, MENHNQI, SEGPCLAFMRA, HAISSS, DQTKA, KGFHS, QADNA, HFTKMQK, SAAVCK, EAQA, TVRLI, GKEC, TPENMGK, DHFQK, LAAG, APPVDNF, FARHERKMLLDNHALYDKTKK, and DYIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species or the homologous regions of Pol μ , Pol β , Pol λ , and Pol θ of any species or the homologous regions of X family polymerases of any species.

3. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to claim 1 comprising at least one amino acid modification when compared to a wild type sequence SEQ ID NO 1 or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species, wherein the modification is selected from one or more of the amino acid regions WLLNRLINRLQNQGILLYDI, VAIF, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA, TVR, KEC, TPENMGK, DHFQ, LAAG, APPVDN,

FARHERKMMLLDNHA, and YIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species.

4. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to any one of claims 1 to 3 comprising at least the sequence:

TVSQYACQRRTTMENHNQIFTDFAFAILAENAEFNESEGPCCLAFMRAASLLKSLPHAISSSKD
LEGLPCLGDQTKAVIEDILEYGQCSKVQDVLCDTRYQTIKLFSTVFGVGLKTAEKWYRKGFH
SLEEVQADNAIHFTKMQKAGFLYYDDISAAVCKAEQAIGQIVEETVRLLIAPDAIVTLTGGF
RRGKECGHDVDFLITTPEMGKEVWLLNRLINRLQNQGILLYYDIVESTFDKTRLPCKRFEAM
DHFQKCFAI IKLKKELAAGRVQKDWKAIRVDFVAPPVDNF AFALLGWTGSRQFERDLRRFAR
HERKMMLLDNHALYDKTKKIFLPAKTEEDI FAHLGLDYIDPWQRNA

or the homologous regions in other species or the homologous regions of Pol μ , Pol β , Pol λ , and Pol θ of any species or the homologous regions of X family polymerases of any species, wherein the sequence has one or more amino acid modifications in one or more of the amino acid regions WLLNRLINRLQNQGILLYYDIV, MENHNQI, SEGPCCLAFMRA, HAISSS, DQTKA, KGFHS, QADNA, HFTKMQK, SAAVCK, EAQA, TVRLI, GKEC, TPENMGK, DHFQK, LAAG, APPVDNF, FARHERKMMLLDNHALYDKTKK, and DYIDP of the sequence.

5. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to any one of claims 1 to 4 comprising at least two amino acid modifications when compared to a wild type sequence SEQ ID NO 1 or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species, wherein;

- a. a first modification is within the amino acid region WLLNRLINRLQNQGILLYYDIV of the sequence of SEQ ID NO 1 or the homologous region in other species; and
- b. a second modification is selected from one or more of the amino acid regions VAIF, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA, TVR, KEC, TPENMGK, DHFQ, LAAG, APPVDN, FARHERKMMLLDNHA, and YIDP of the sequence of SEQ ID NO 1 or the homologous regions in other species.

6. The modified terminal deoxynucleotidyl transferase (TdT) enzyme according to claim 5 wherein the modification in amino acid region WLLNRLINRLQNQGILLYYDIV improves the enzyme solubility and the modification(s) in VAIF, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA, TVR, KEC, TPENMGK, DHFQ, LAAG, APPVDN, FARHERKMMLLDNHA, and YIDP improves the incorporation of nucleotides having a modification at the 3' position in comparison to the wild type sequence.

7. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to any one of claims 1 to 6 wherein the wild type sequence is selected from

	gi 768	Bos taurus
	gi 460163	Gallus gallus
5	gi 494987	Xenopus laevis
	gi 1354475	Oncorhynchus mykiss
	gi 2149634	Monodelphis domestica
	gi 12802441	Mus musculus
	gi 28852989	Ambystoma mexicanum
10	gi 38603668	Takifugu rubripes
	gi 40037389	Raja eglanteria
	gi 40218593	Ginglymostoma cirratum
	gi 46369889	Danio rerio
	gi 73998101	Canis lupus familiaris
15	gi 139001476	Lemur catta
	gi 139001490	Microcebus murinus
	gi 139001511	Otolemur garnettii
	gi 148708614	Mus musculus
	gi 149040157	Rattus norvegicus
20	gi 149704611	Equus caballus
	gi 164451472	Bos taurus
	gi 169642654	Xenopus (Silurana) tropicalis
	gi 291394899	Oryctolagus cuniculus
	gi 291404551	Oryctolagus cuniculus
25	gi 301763246	Ailuropoda melanoleuca
	gi 311271684	Sus scrofa
	gi 327280070	Anolis carolinensis
	gi 334313404	Monodelphis domestica
	gi 344274915	Loxodonta africana
30	gi 345330196	Ornithorhynchus anatinus
	gi 348588114	Cavia porcellus
	gi 351697151	Heterocephalus glaber
	gi 355562663	Macaca mulatta
	gi 395501816	Sarcophilus harrisii

	gi 395508711	Sarcophilus harrisii
	gi 395850042	Otolemur garnettii
	gi 397467153	Pan paniscus
	gi 403278452	Saimiri boliviensis boliviensis
5	gi 410903980	Takifugu rubripes
	gi 410975770	Felis catus
	gi 432092624	Myotis davidii
	gi 432113117	Myotis davidii
	gi 444708211	Tupaia chinensis
10	gi 460417122	Pleurodeles waltl
	gi 466001476	Orcinus orca
	gi 471358897	Trichechus manatus latirostris
	gi 478507321	Ceratotherium simum simum
	gi 478528402	Ceratotherium simum simum
15	gi 488530524	Dasyus novemcinctus
	gi 499037612	Maylandia zebra
	gi 504135178	Ochotona princeps
	gi 505844004	Sorex araneus
	gi 505845913	Sorex araneus
20	gi 507537868	Jaculus jaculus
	gi 507572662	Jaculus jaculus
	gi 507622751	Octodon degus
	gi 507640406	Echinops telfairi
	gi 507669049	Echinops telfairi
25	gi 507930719	Condylura cristata
	gi 507940587	Condylura cristata
	gi 511850623	Mustela putorius furo
	gi 512856623	Xenopus (Silurana) tropicalis
	gi 512952456	Heterocephalus glaber
30	gi 524918754	Mesocricetus auratus
	gi 527251632	Melopsittacus undulatus
	gi 528493137	Danio rerio
	gi 528493139	Danio rerio
	gi 529438486	Falco peregrinus

	gi 530565557	Chrysemys picta bellii
	gi 532017142	Microtus ochrogaster
	gi 532099471	Ictidomys tridecemlineatus
	gi 533166077	Chinchilla lanigera
5	gi 533189443	Chinchilla lanigera
	gi 537205041	Cricetulus griseus
	gi 537263119	Cricetulus griseus
	gi 543247043	Geospiza fortis
	gi 543351492	Pseudopodoces humilis
10	gi 543731985	Columba livia
	gi 544420267	Macaca fascicularis
	gi 545193630	Equus caballus
	gi 548384565	Pundamilia nyererei
	gi 551487466	Xiphophorus maculatus
15	gi 551523268	Xiphophorus maculatus
	gi 554582962	Myotis brandtii
	gi 554588252	Myotis brandtii
	gi 556778822	Pantholops hodgsonii
	gi 556990133	Latimeria chalumnae
20	gi 557297894	Alligator sinensis
	gi 558116760	Pelodiscus sinensis
	gi 558207237	Myotis lucifugus
	gi 560895997	Camelus ferus
	gi 560897502	Camelus ferus
25	gi 562857949	Tupaia chinensis
	gi 562876575	Tupaia chinensis
	gi 564229057	Alligator mississippiensis
	gi 564236372	Alligator mississippiensis
	gi 564384286	Rattus norvegicus
30	8. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to any one of claims 1 to 7 wherein the modification within the amino acid region <u>WLLNRLINRLQNQGILLYD</u> IIV is at one or more of the underlined amino acids.	

9. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to claim 8 wherein the modification within the amino acid region WLLNRLINRLQNQGILLYDIV is to QLLPKVINLWEKKGLLLYYDLV.
10. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to any one of
 5 claims 1 to 9 wherein the modification is selected from one or more of the amino acid regions VAIF, MGA, ENHNQ, FMRA, HAI, TKA, FHS, QADNA, MQK, SAAVCK, EAQA, TVR, KEC, TPMEGK, DHFQ, LAAG, APPVDN, FARHERKMLLDNHA, and YIDP.
11. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to any one of
 10 claims 1 to 9 wherein the modification is selected from one or more of the amino acids A53, V68, V71, E97, I101, G109, Q115, V116, S125, T137, Q143, N154, H155, Q157, I158, I165, G177, L180, A181, M183, A195, K200, T212, K213, A214, E217, T239, F262, S264, Q269, N272, A273, K281, S291, K296, Q300, T309, R311, E330, T341, E343, G345, N352, N360, Q361, I363, Y367, H389, L403, G406, D411, A421, P422, V424, N426, R438, F447, R452, L455, and/or D488.
12. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to any one of
 15 claims 1 to 9 wherein the modification is selected from one or more of the amino acid changes A53G, V68I, V71I, E97A, I101V, G109E, G109R, Q115E, V116I, V116S, S125R, T137A, Q143P, M152L, M152T, N154H, H155C, H155N, H155R, Q157K, Q157R, I158V, I158L, I158M, I165V, N169R, N173R, S175N, G177S, G177V, G177D, C179A, C179S, L180V, A181E, M183R,
 20 L188V, A195T, A195P, S197I, S197L, S198R, S198C, S199L, K200R, E203Q, G204D, D210E, Q211R, T212S, K213S, A214R, A214C, A214G, I216M, E217Q, D218E, L220I, L220F, L220Y, Y222C, V228A, D230E, Q238K, T239S, L242Q, L251P, K260M, F262L, L265F, S264T, E267D, Q269K, D271N, D271E, N272K, A273S, A273T, H275Q, T277S, K281R, S291N, S291T, K296R, Q300D, Q304R, Q304C, Q304H, I305V, T309A, R311W, R311H, L312G, L312K, L312A, I313M,
 25 A314T, I318L, V319L, T320A, G328A, E330N, E330S, C331I, L338I, T341S, P342A, E343G, E343Q, M344A, M344Q, M344K, G345R, K346R, N352Q, K353R, K353D, V354I, V354L, I355V, I355M, N356R, N356D, W358L, N360K, Q361K, G362E, I363L, Y366F, Y367C, D368H, D368E, V370I, K376I, T377S, T377L, T377A, C381S, K383R, H389N, H389A, K392Q, F394W, I397L, K398R, K400E, K400N, K401A, K401Q, E402Q, L403Q, L403R, A405D, G406R, R407C, D411N,
 30 A421L, A421M, A421V, P422A, P422C, V424Y, V424I, V424A, N426R, N426C, F427Y, A430T, R438K, F447W, R452K, L455I, K453R, Y462F, K464R, T465R, K467R, T474S, D477E, D485E, I487V, and/or D488P.

13. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to claim 10 wherein the modification is selected from one or more of the amino acid regions FMRA, QADNA, EAQA, APPVDN, FARHERKMLLDNHA, and YIDP is at the underlined amino acid(s).
14. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to claim 13 wherein the modification is selected from one or more of the following sequences FRRA, QADKA, EADA, MPPVDN, FARHERKMLLDRHA, and YIPP.
15. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to claim 13 wherein the modification is selected from two or more of the following sequences FRRA, QADKA, EADA, MPPVDN, FARHERKMLLDRHA, and YIPP.
16. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to claim 15 wherein the modification contain each of the following sequences FRRA, QADKA, EADA, MPPVDN, FARHERKMLLDRHA, and YIPP.
17. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to any one of claims 1 to 16 comprising a sequence selected from SEQ ID NO 4 to SEQ ID NO 173 or a truncated version thereof.
18. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to any one of claims 1 to 16 comprising a sequence selected from SEQ ID NO 174 to SEQ ID NO 343.
19. A modified terminal deoxynucleotidyl transferase (TdT) enzyme according to any one of claims 1 to 16 comprising a sequence selected from SEQ ID NO 344 to 727.
20. A method of nucleic acid synthesis, which comprises the steps of:
 - (a) providing an initiator oligonucleotide;
 - (b) adding a 3'-blocked nucleoside triphosphate to said initiator oligonucleotide in the presence of a terminal deoxynucleotidyl transferase (TdT) as defined in any one of claims 1 to 19;
 - (c) removal of all reagents from the initiator oligonucleotide;
 - (d) cleaving the blocking group in the presence of a cleaving agent; and
 - (e) removal of the cleaving agent.
21. The method as defined in claim 20, wherein greater than 1 nucleotide is added by repeating steps (b) to (e).
22. The method as defined in claim 20 or claim 21 wherein the 3'-blocked nucleoside triphosphate is blocked a group selected from 3'-O-azidomethyl, 3'-aminooxy, 3'-O-(N-oxime), 3'-O-allyl 3'-O-cyanoethyl, 3'-O-acetyl, 3'-O-nitrate, 3'-phosphate, 3'-O-acetyl levulinic ester, 3'-O-tert butyl dimethyl silane, 3'-O-trimethyl(silyl)ethoxymethyl, 3'-O-ortho-nitrobenzyl, or 3'-O-para-nitrobenzyl.

23. The method as defined in claim 22 wherein the 3'-blocked nucleoside triphosphate is blocked by either a 3'-O-azidomethyl, 3'-aminooxy or 3'-O-allyl group.
24. A kit comprising a terminal deoxynucleotidyl transferase (TdT) as defined in any one of claims 1 to 19 in combination with an initiator oligonucleotide and one or more 3'-blocked nucleoside triphosphates.

5

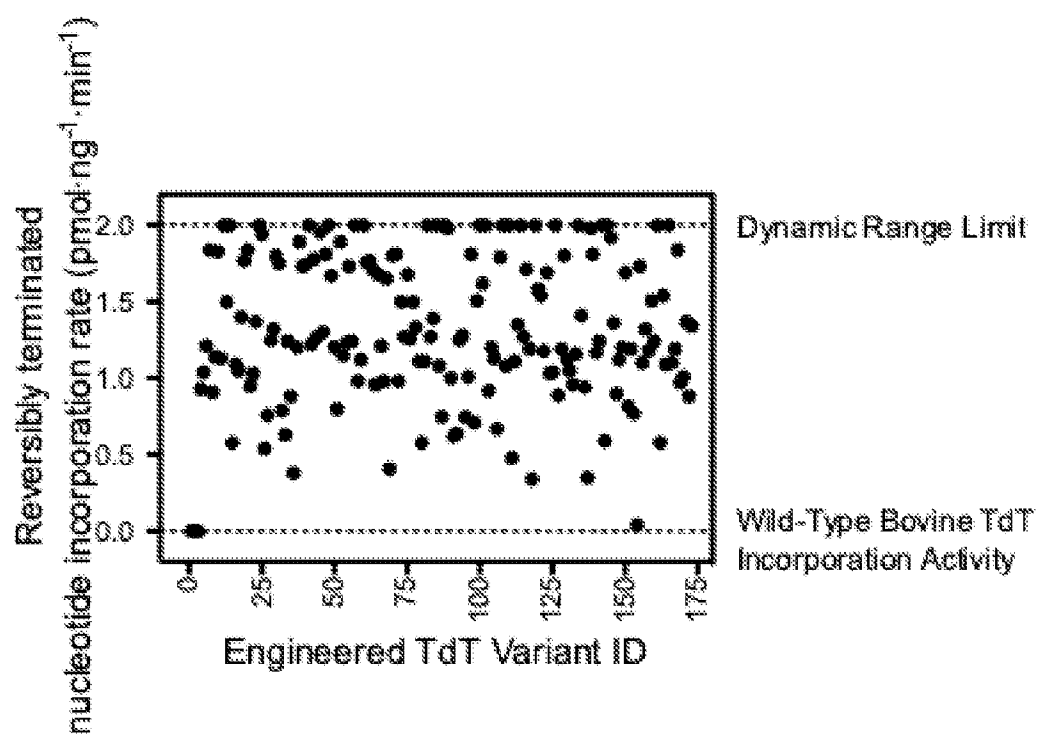


Figure 1

[illegible]

Figure 2

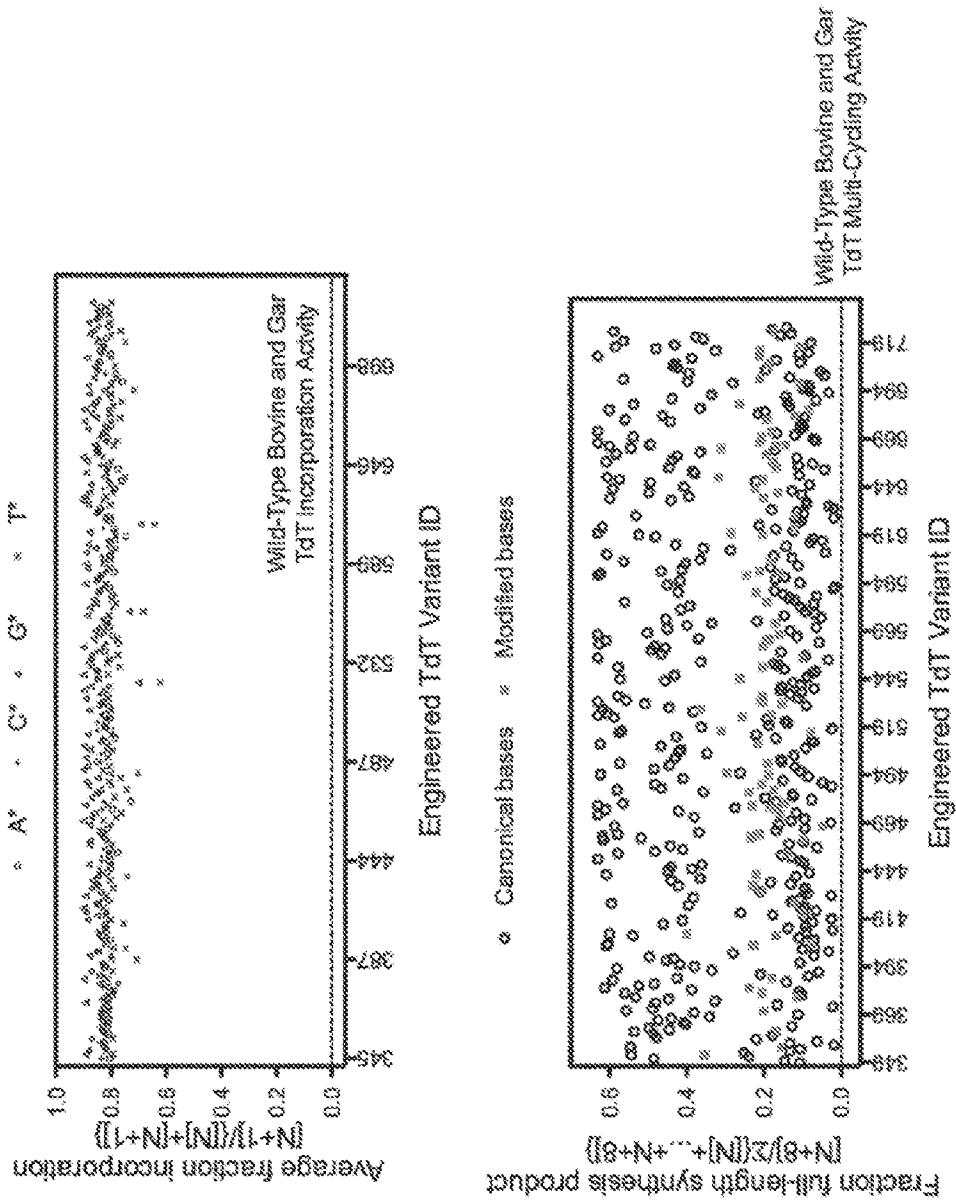


Figure 3

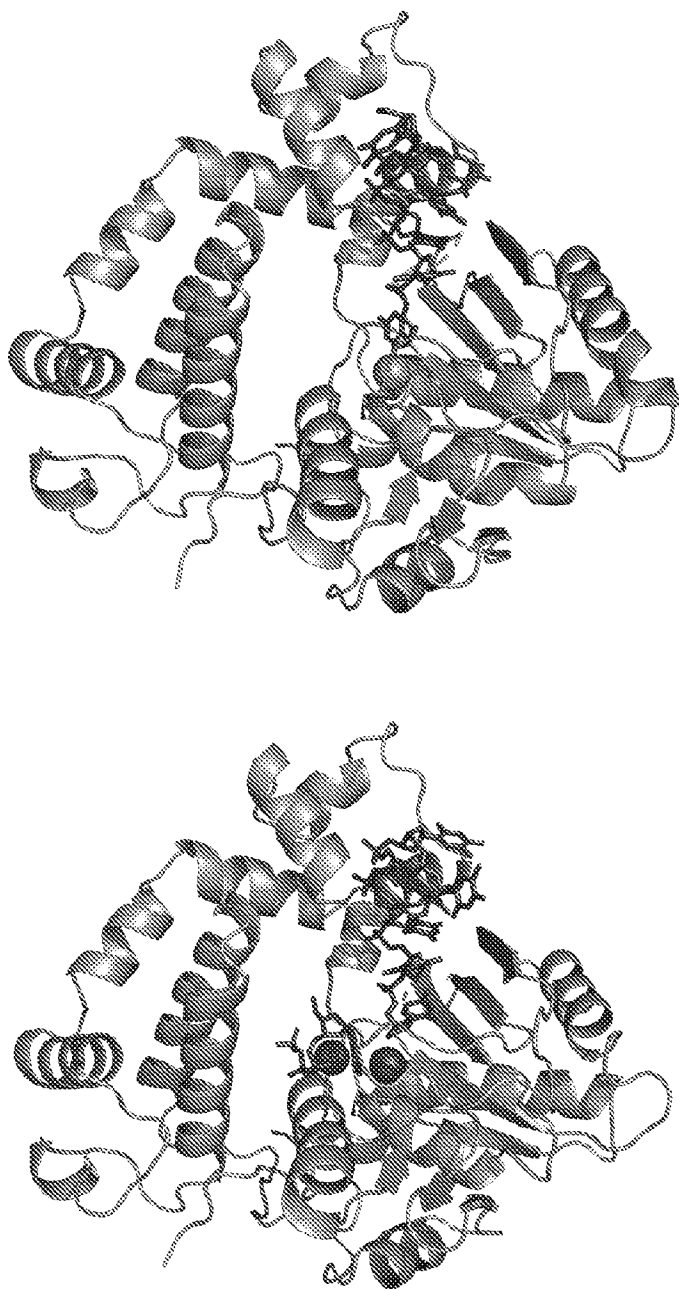


Figure 4

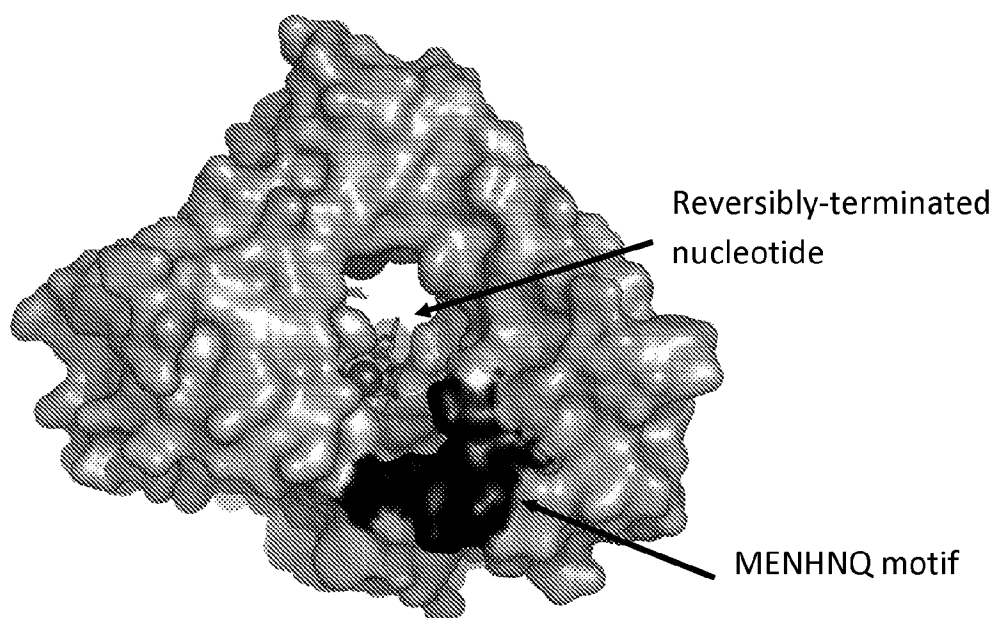


Figure 5

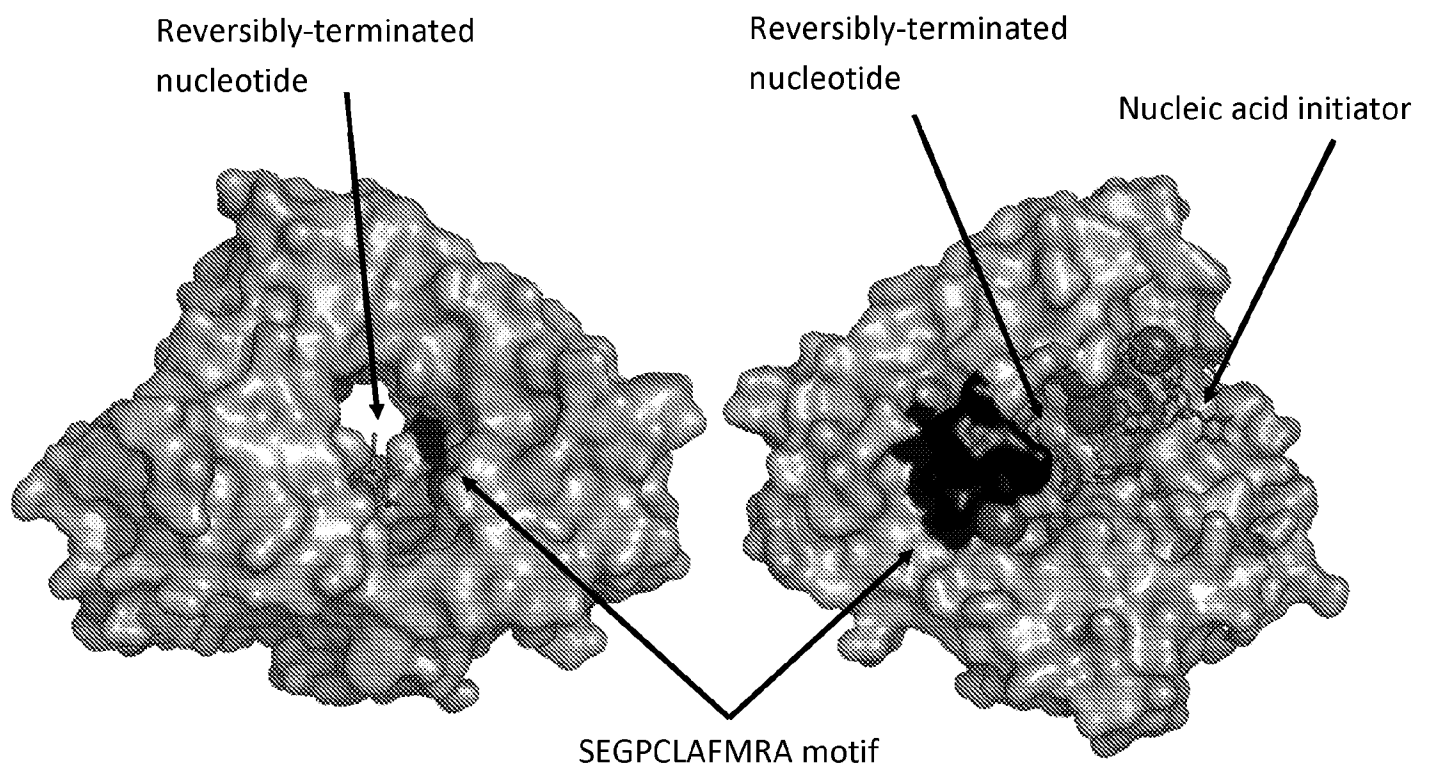


Figure 6

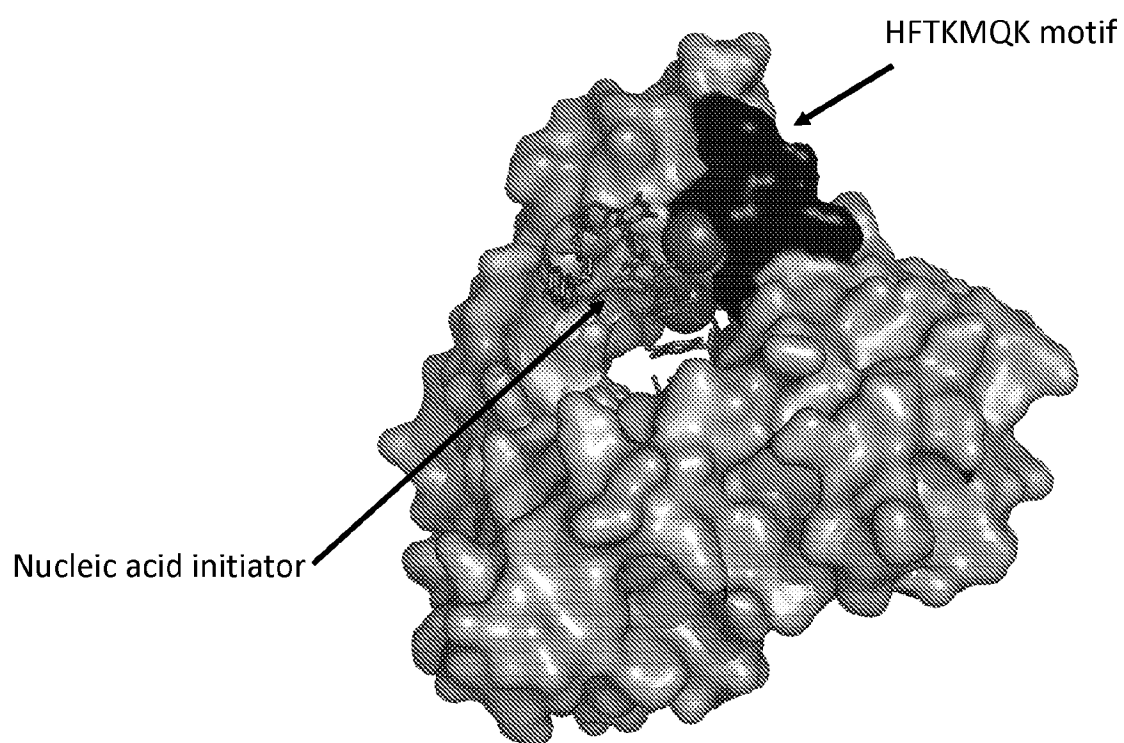


Figure 7

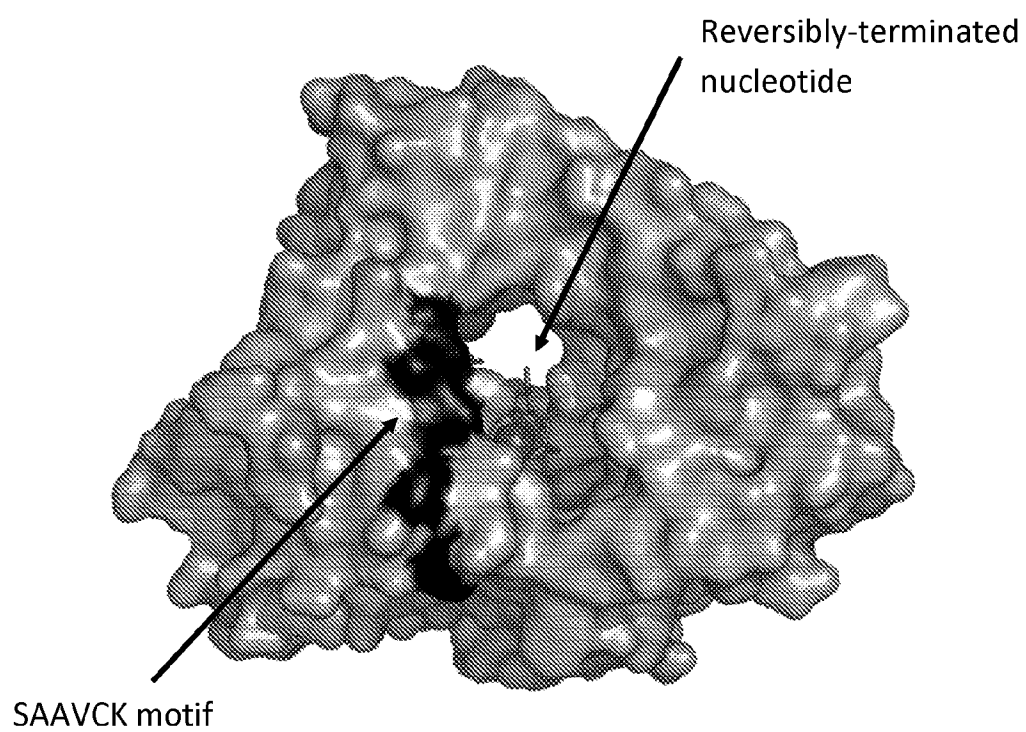


Figure 8

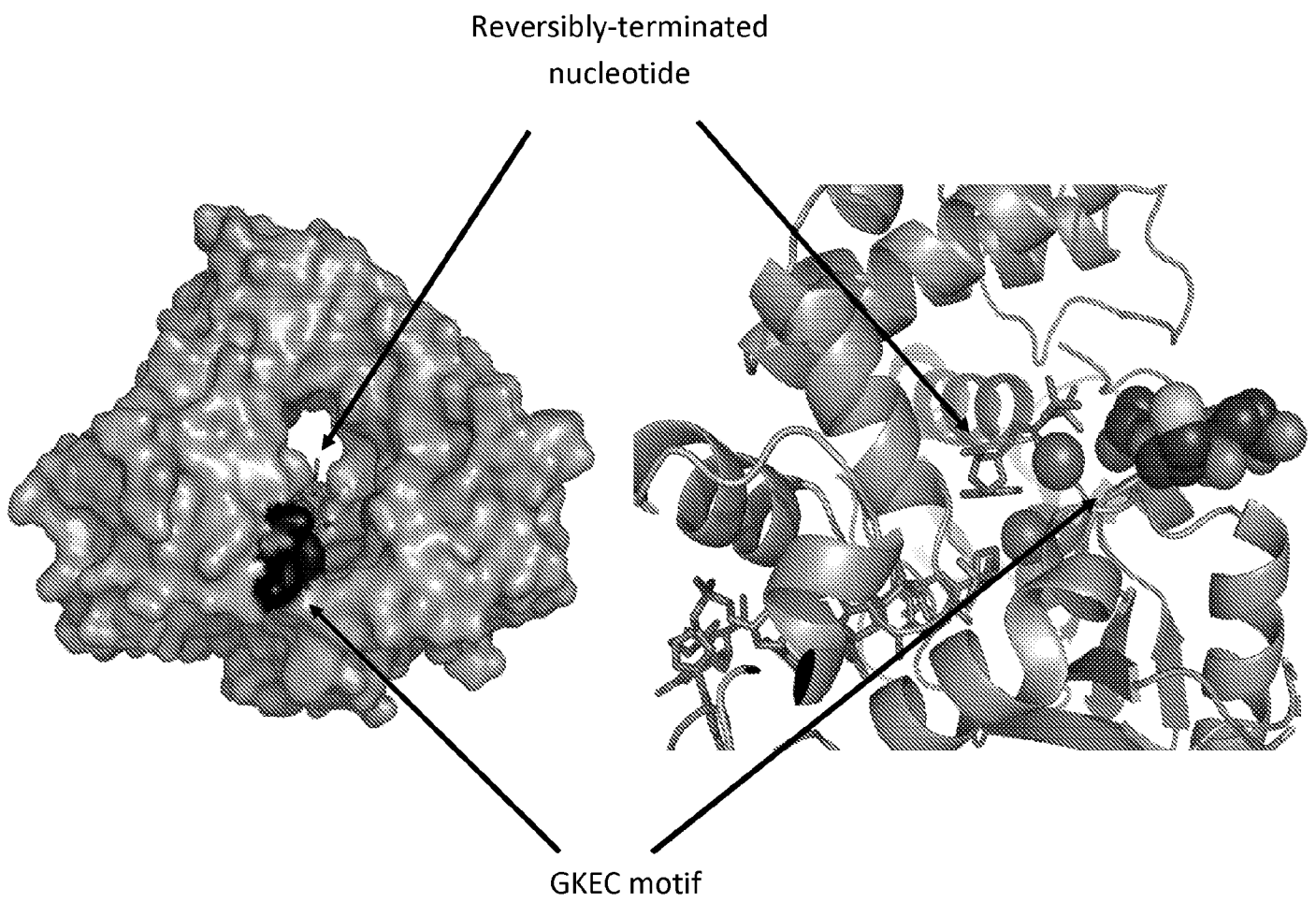


Figure 9

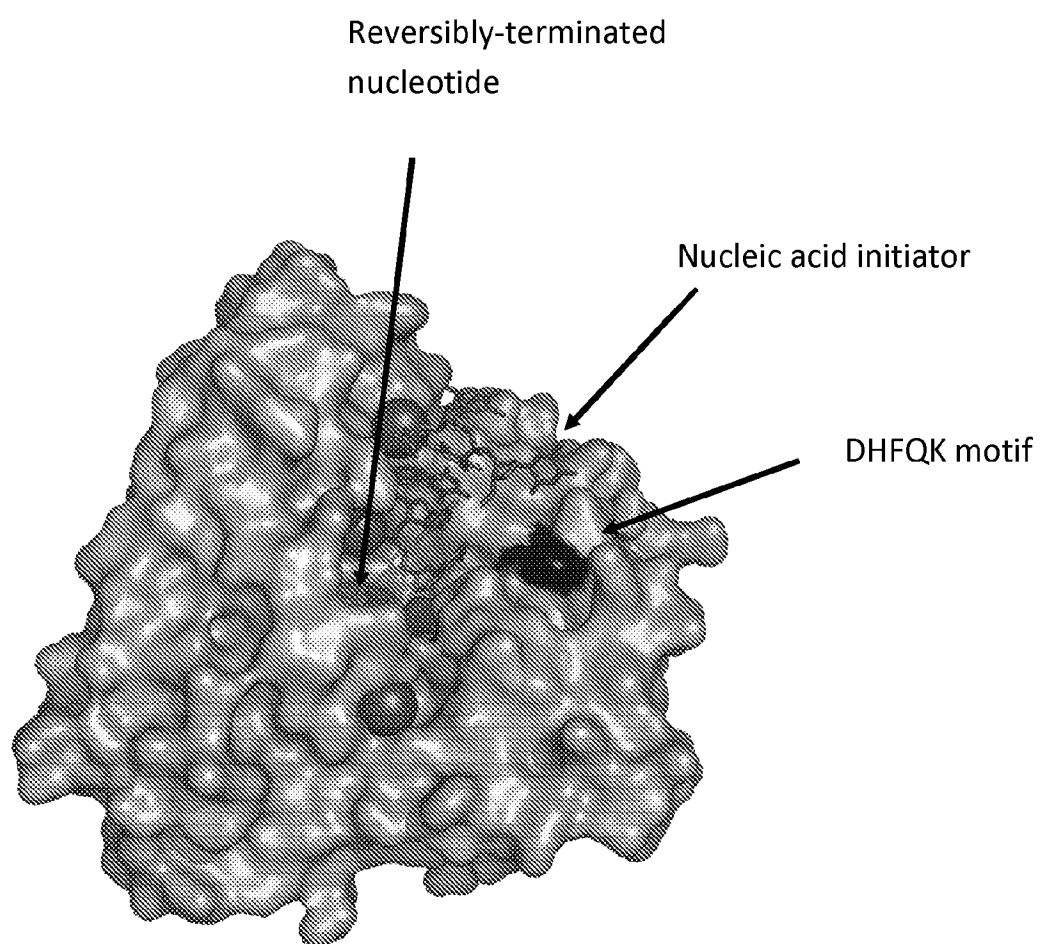


Figure 10

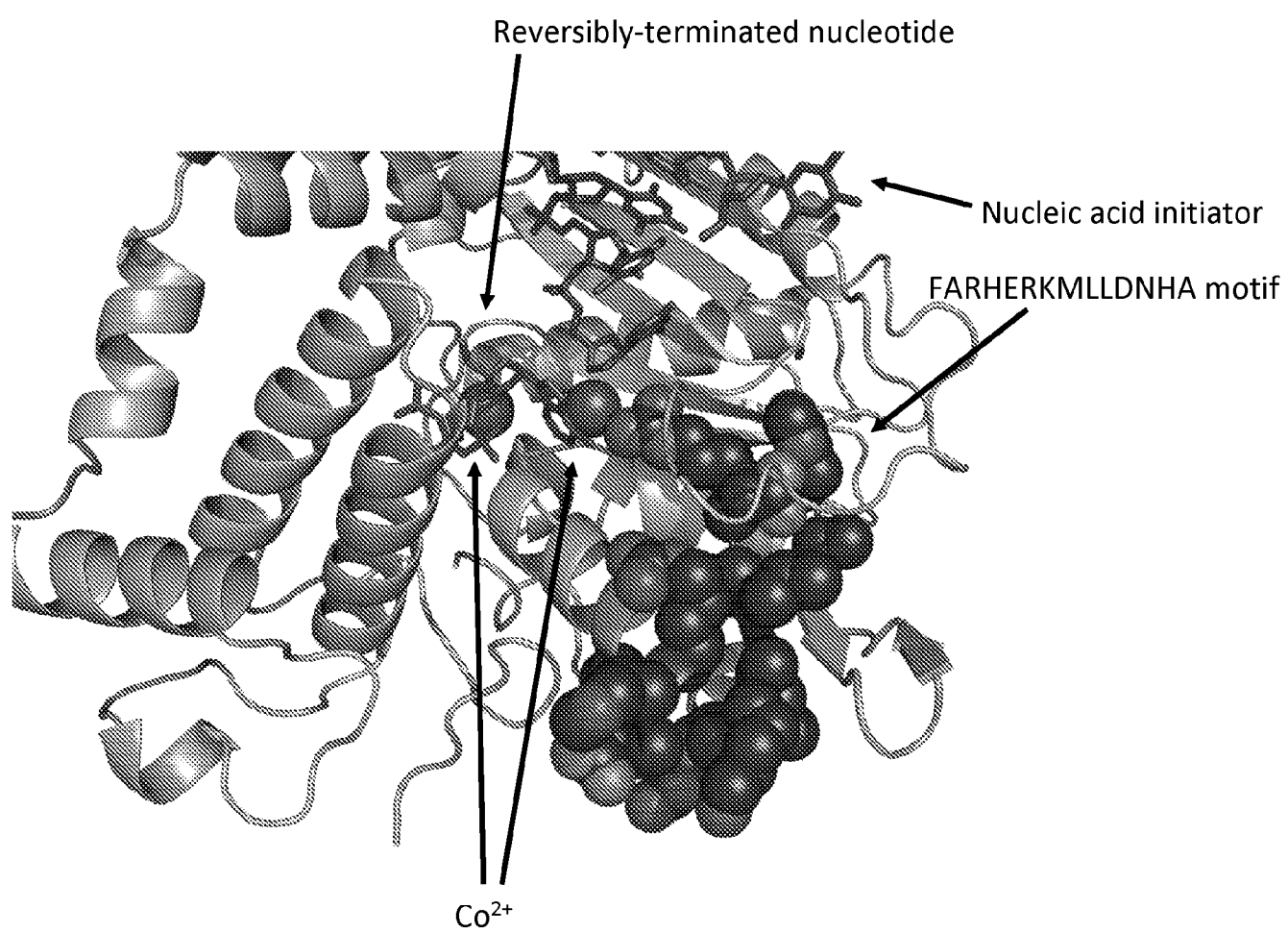


Figure 11

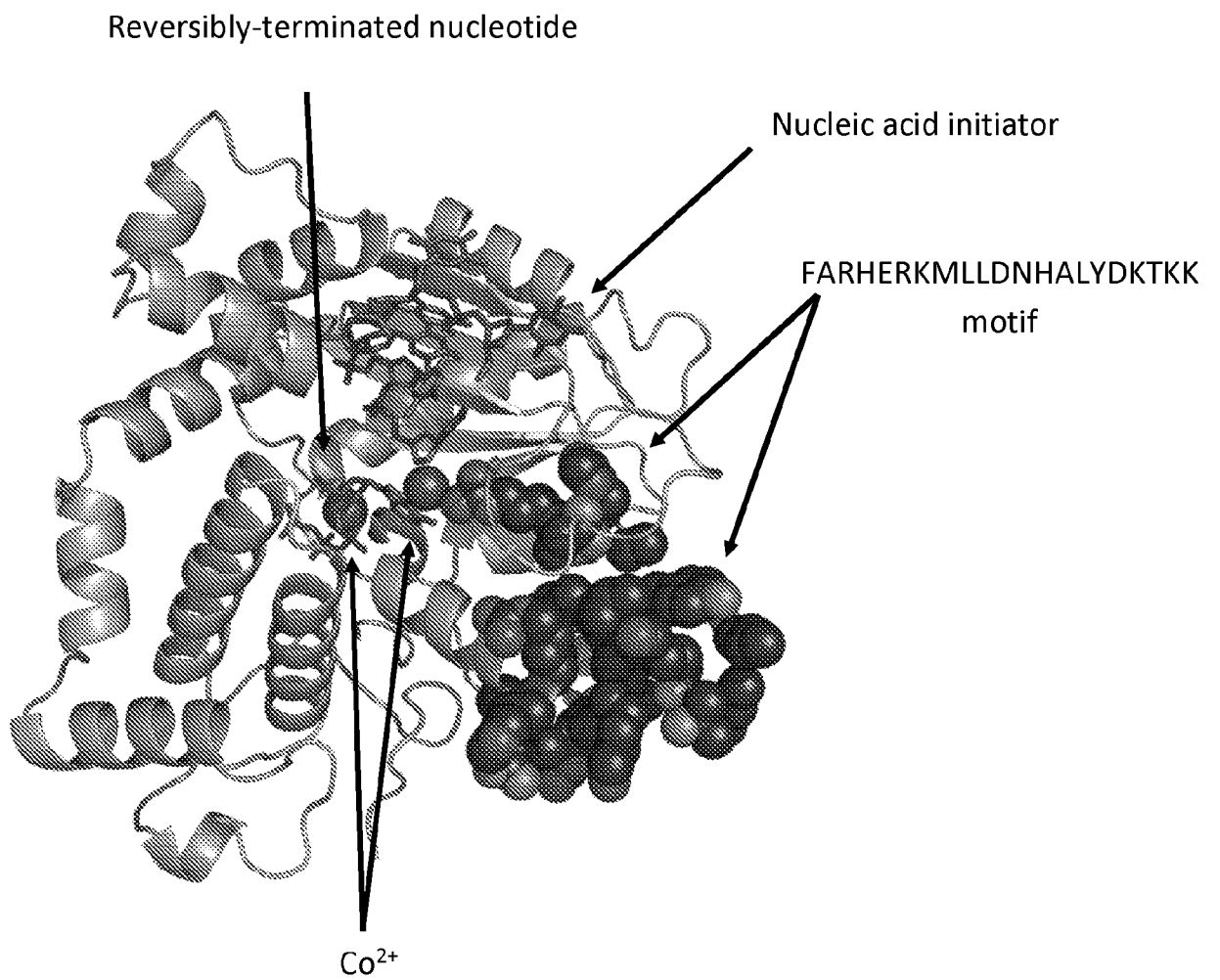


Figure 12

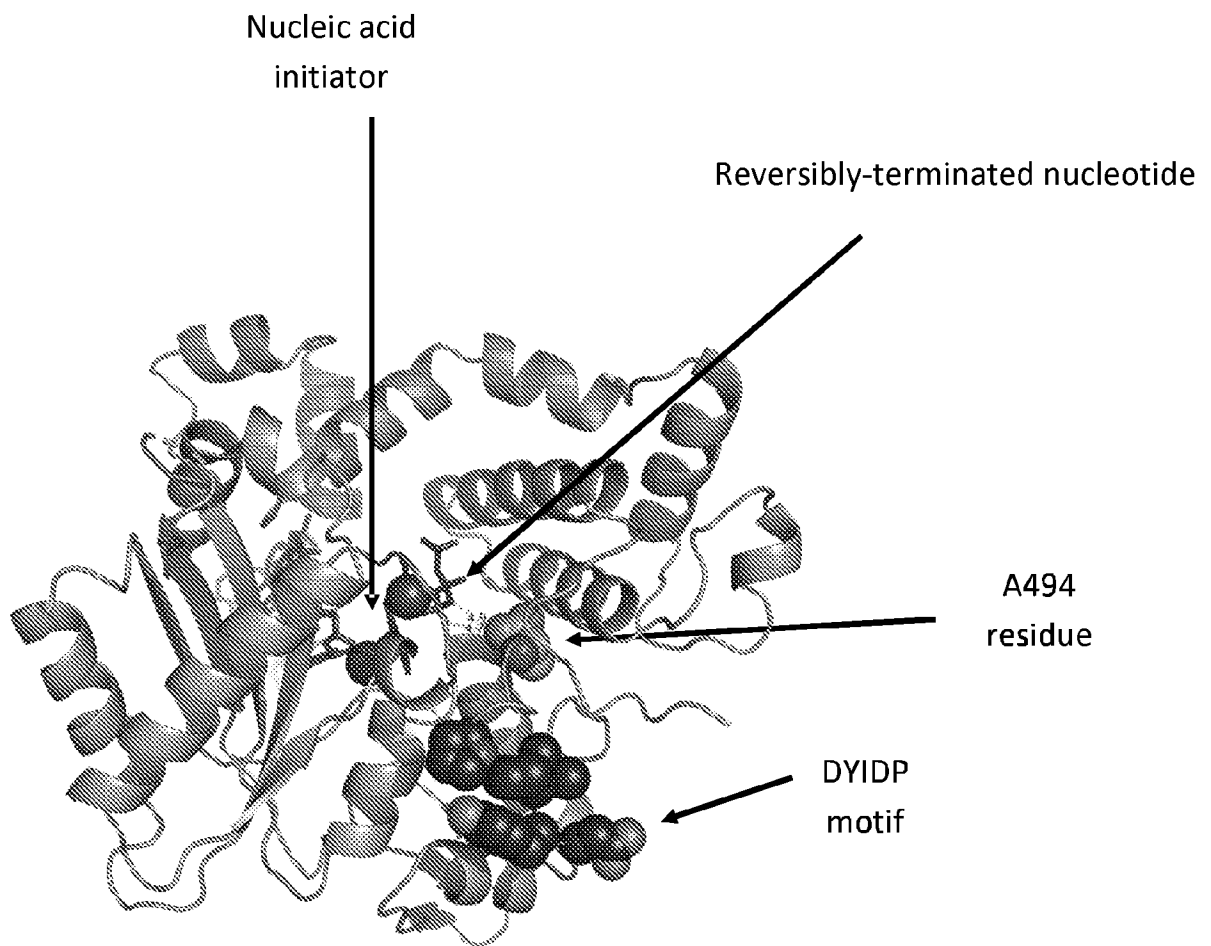


Figure 13

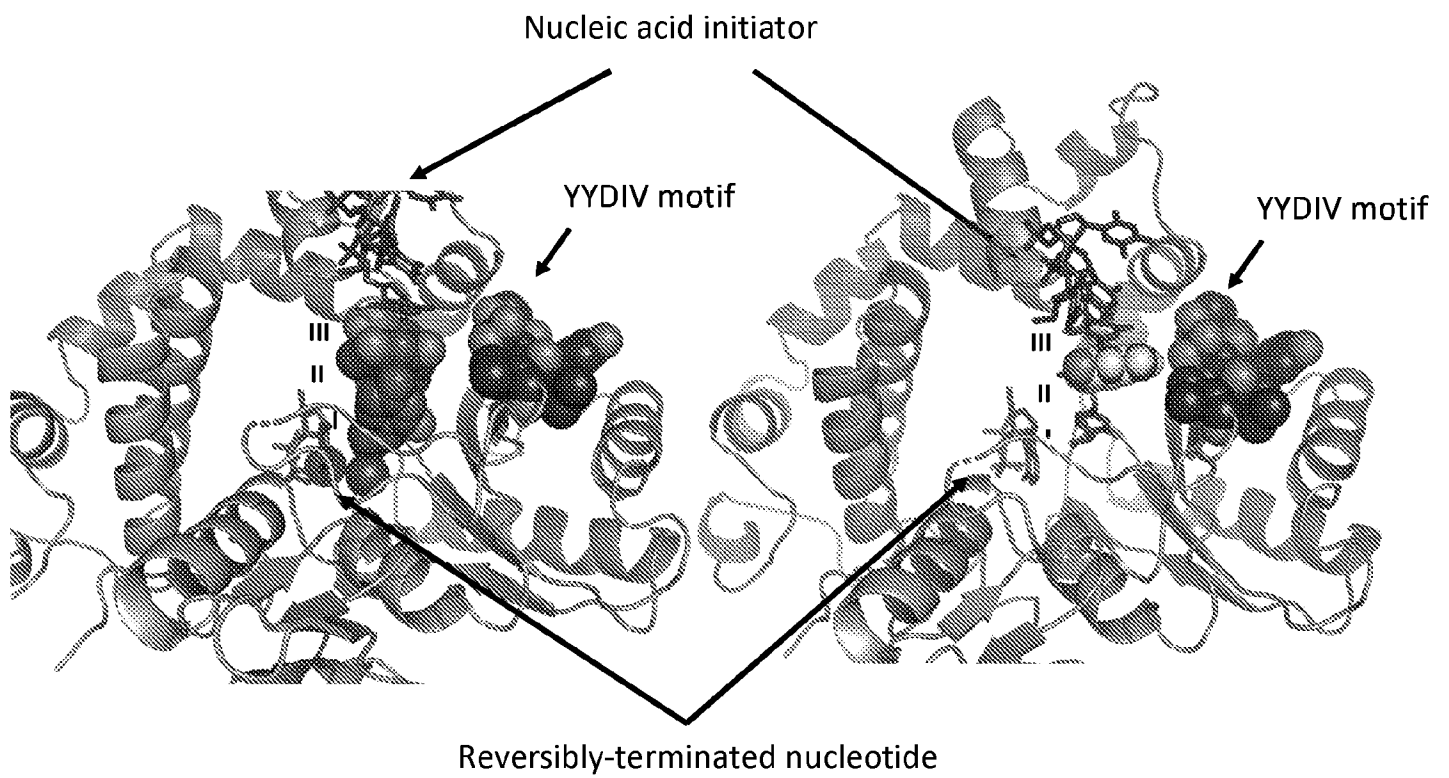


Figure 14

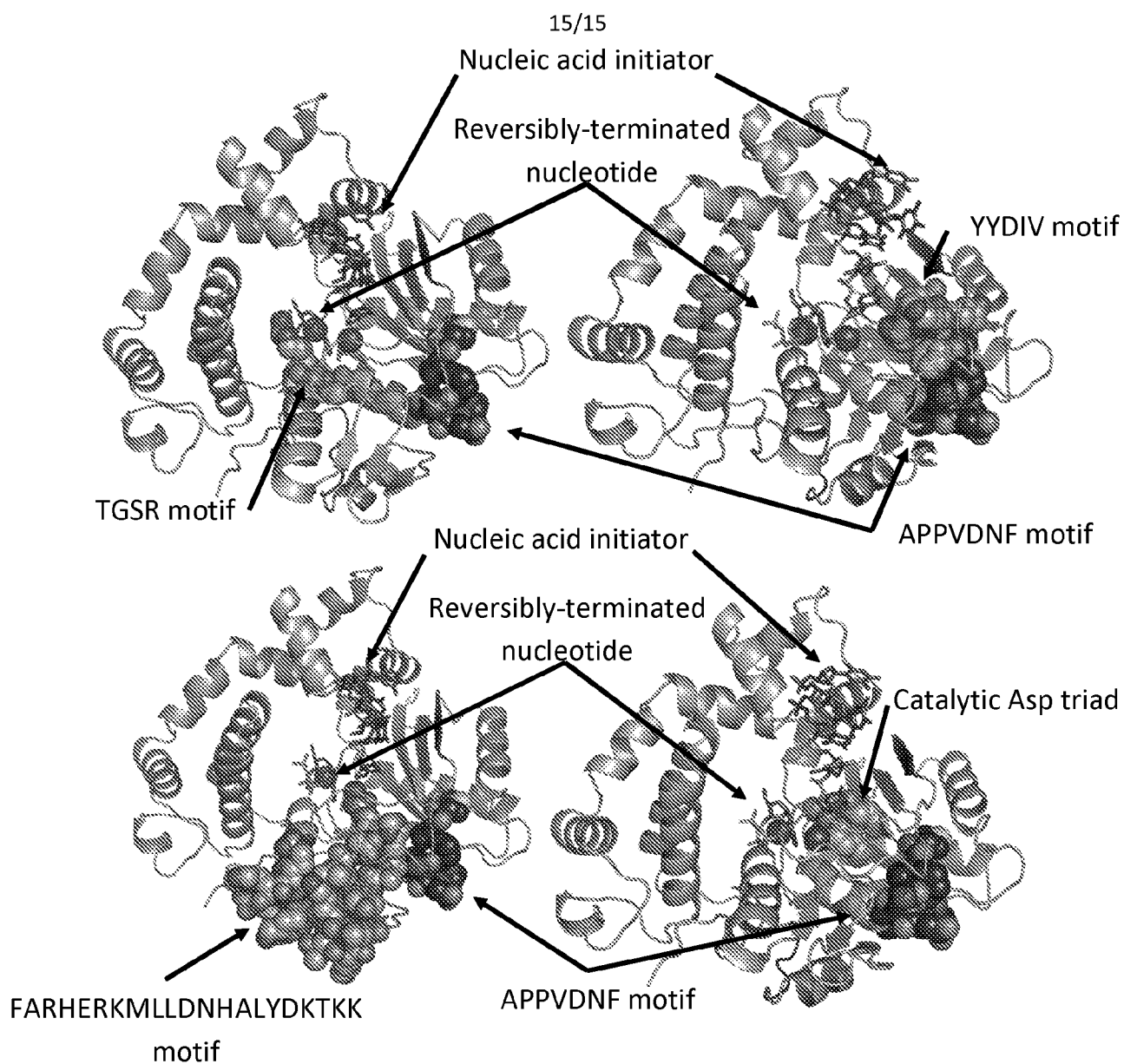


Figure 15

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2020/050247

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	US 2018/023108 A1 (CHEN MICHAEL C [GB] ET AL) 25 January 2018 (2018-01-25) paragraphs [0001], [0010], [0011] - [0026], [0029] - [0031], [0044], [0049], [0152] - [0155] claims 1-5, 38 examples 3, 4 ----- -/--	1-7, 10-24



Further documents are listed in the continuation of Box C.



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

7 July 2020

Date of mailing of the international search report

13/07/2020

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Surdej, Patrick

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2020/050247

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
1-7, 10-24(all partially)
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☒ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2020/050247

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2017/216472 A2 (DNA SCRIPT [FR]; PASTEUR INSTITUT [FR]; CENTRE NAT RECH SCIENT [FR]) 21 December 2017 (2017-12-21) page 1, line 3 - line 7 page 1, paragraphs 1, 2 page 2, line 18 - line 26 page 3, line 16 - line 23 page 3, paragraph 2 - page 6, paragraph 3 page 12, paragraph 3 page 13, line 10 - line 21 page 22, line 23 - page 23, line 6 table 1 examples 1, 3 examples 5, 6 claims 1-5, 21-24 -----	1-7, 10-24
X	US 2016/108382 A1 (EFCAVITCH J WILLIAM [US] ET AL) 21 April 2016 (2016-04-21) the whole document -----	1-7, 10-24
X	WO 2018/215803 A1 (NUCLERA NUCLEICS LTD [GB]) 29 November 2018 (2018-11-29) the whole document -----	1-7, 10-24
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 -----	1-7, 10-24

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2020/050247

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2018023108 A1	25-01-2018	AU 2016217663 A1 CA 2975789 A1 CN 107567501 A EP 3256596 A1 JP 2018509177 A US 2018023108 A1 WO 2016128731 A1	17-08-2017 18-08-2016 09-01-2018 20-12-2017 05-04-2018 25-01-2018 18-08-2016
WO 2017216472 A2	21-12-2017	AU 2017286477 A1 CA 3024184 A1 CN 109477080 A EP 3469075 A2 FR 3052462 A1 FR 3053699 A1 JP 2019517799 A KR 20190024883 A SG 11201809961T A US 2020002690 A1 WO 2017216472 A2	06-12-2018 21-12-2017 15-03-2019 17-04-2019 15-12-2017 12-01-2018 27-06-2019 08-03-2019 28-12-2018 02-01-2020 21-12-2017
US 2016108382 A1	21-04-2016	EP 3209781 A1 US 2016108382 A1 US 2020190490 A1 WO 2016064880 A1	30-08-2017 21-04-2016 18-06-2020 28-04-2016
WO 2018215803 A1	29-11-2018	AU 2018272302 A1 CA 3064820 A1 CN 110997907 A EP 3630965 A1 GB 2577020 A WO 2018215803 A1	19-12-2019 29-11-2018 10-04-2020 08-04-2020 11-03-2020 29-11-2018

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-7, 10-24(all partially)

A modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising amino acid modifications when compared to a wild type sequence SEQ ID NO 1 or a truncated version thereof or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species or the homologous amino acid sequence of Polmu, Polbeta, Pollambda, and Poltheta of any species or the homologous amino acid sequence of X family polymerases of any species, wherein the amino acid is modified at one or more of the amino acids: V32, A33, 134, F35

2-69. claims: 1-24(partially)

A modified terminal deoxynucleotidyl transferase (TdT) enzyme comprising amino acid modifications when compared to a wild type sequence SEQ ID NO 1 or a truncated version thereof or the homologous amino acid sequence of a terminal deoxynucleotidyl transferase (TdT) enzyme in other species or the homologous amino acid sequence of Polmu, Polbeta, Pollambda, and Poltheta of any species or the homologous amino acid sequence of X family polymerases of any species, wherein the amino acid is modified, respectively (groups are separated by a semicolon), at one or more of the amino acids: A53; V68; V71; E97; I101; M108, G109, A110; Q115; V116; S125; T137; Q143; M152, E153, N154, H155, N156, Q157, I158; L188; H194, A195, I196, S197, S198, S199; K200; E203; G204; D210, Q211, T212, K213, A214; I216; E217; D218; L220; Y222; V228; D230; Q238; T239; L242; L251; K260, G261, F262, H263, S264; L265; E267; Q269, A270, D271, N272, A273; H275, F276, T277, K278, M279, Q280, K281; S291, A292, A293, V294, C295, K296; E298, A299, Q300, A301; Q304; I305; T309, V310, R311, L312, I313; A314; I318; V319; T320; G328, K329, E330, C331; L338; T341, P342, E343, M344, G345, K346; W349, L350, L351, N352, R353, L354, I355, N356, R357, L358, Q359, N360, Q361, G362, I363, L364, L365, Y366, Y367, D368, I369, V370; K376; T377; C381; K383; D388, H389, F390, Q391, K392; F394; I397; K398; K400; K401; E402; L403, A404, A405, G406; R407; D411; A421, P422, P423, V424, D425, N426, F427; A430; R438; F447, A448, R449, H450, E451, R452, K453, M454, L455, L456, D457, N458, H459, A460, L461, Y462, D463, K464, T465, K466, K467; T474; D477; D485, Y486, I487, D488, P489
