



- (51) International Patent Classification:
C12Q 1/68 (2018.01) C12N 9/12 (2006.01)
- (21) International Application Number:
PCT/US2018/030972
- (22) International Filing Date:
03 May 2018 (03.05.2018)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
62/501,526 04 May 2017 (04.05.2017) US
- (71) Applicant: STRATOS GENOMICS INC. [US/US]; 2211 Elliott Avenue, Suite 210, Seattle, Washington 98121 (US).
- (72) Inventors: KOKORIS, Mark Stamatios; 10535 North-east 196th Street, Bothell, Washington 98011 (US). PRINDLE, Marc; 7506 Winona Ave N, Seattle, Washington 98103 (US). CHASE, Jack; 708 6th Ave N, Apt. 421, Seattle, Washington 98109 (US). BUSAM, Robert; 2539 Piedmont Pl W, Seattle, Washington 98119 (US). KO-

VARIK, Michael; 500 3rd Avenue West, Apt. 409, Seattle, Washington 98119 (US). KELLER, Salka; 16737 6th Ave. NE, Shoreline, Washington 98155 (US). MURT, Megan; 4027A California Ave SW, Seattle, Washington 98116 (US). THIESEN, Greg; 4703 South Juneau St., Seattle, Washington 98118 (US).

(74) Agent: HERMANN, Karl R.; Seed Intellectual Property Law Group, Suite 5400, 701 Fifth Avenue, Seattle, Washington 98104-7064 (US).

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

(54) Title: DP04 POLYMERASE VARIANTS

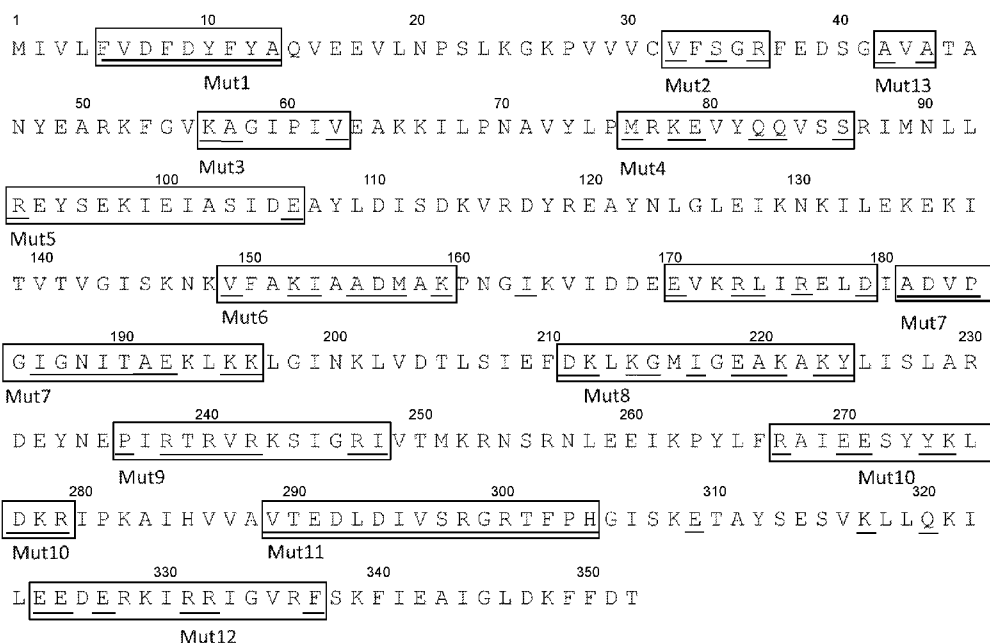


FIG. 1A

(57) Abstract: Recombinant DP04-type DNA polymerase variants with amino acid substitutions that confer modified properties upon the polymerase for improved single molecule sequencing applications are provided. Such properties may include enhanced binding and incorporation of bulky nucleotide analog substrates into daughter strands and the like. Also provided are compositions comprising such DP04 variants and nucleotide analogs, as well as nucleic acids which encode the polymerases with the aforementioned phenotypes.



(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).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

DP04 POLYMERASE VARIANTS

FIELD OF THE INVENTION

The disclosure relates generally to polymerase compositions and methods. More particularly, the disclosure relates to modified DPO4 polymerases and
5 their use in biological applications including, for example, nucleotide analogue incorporation, primer extension and single molecule sequencing reactions.

BACKGROUND OF THE INVENTION

DNA polymerases replicate the genomes of living organisms. In addition to this central role in biology, DNA polymerases are also ubiquitous tools of
10 biotechnology. They are widely used, e.g., for reverse transcription, amplification, labeling, and sequencing, all central technologies for a variety of applications, such as nucleic acid sequencing, nucleic acid amplification, cloning, protein engineering, diagnostics, molecular medicine, and many other technologies.

Because of their significance, DNA polymerases have been extensively
15 studied, with a focus, e.g., on phylogenetic relationships among polymerases, structure of polymerases, structure-function features of polymerases, and the role of polymerases in DNA replication and other basic biological processes, as well as ways of using DNA polymerases in biotechnology. For a review of polymerases, see, e.g., Hubscher et al. (2002) "Eukaryotic DNA Polymerases" Annual Review of Biochemistry Vol. 71: 133-
20 163, Alba (2001) "Protein Family Review: Replicative DNA Polymerases" Genome Biology 2(1): reviews 3002.1-3002.4, Steitz (1999) "DNA polymerases: structural diversity and common mechanisms" J Biol Chem 274:17395-17398, and Burgers et al. (2001) "Eukaryotic DNA polymerases: proposal for a revised nomenclature" J Biol. Chem. 276(47): 43487-90. Crystal structures have been solved for many polymerases,
25 which often share a similar architecture. The basic mechanisms of action for many polymerases have been determined.

A fundamental application of DNA polymerases is in DNA sequencing technologies. From the classical Sanger sequencing method to recent "next-generation"

sequencing (NGS) technologies, the nucleotide substrates used for sequencing have necessarily changed over time. The series of nucleotide modifications required by these rapidly changing technologies has introduced daunting tasks for DNA polymerase researchers to look for, design, or evolve compatible enzymes for ever-changing DNA sequencing chemistries. DNA polymerase mutants have been identified that have a variety of useful properties, including altered nucleotide analog incorporation abilities relative to wild-type counterpart enzymes. For example, Vent^{A488L} DNA polymerase can incorporate certain non-standard nucleotides with a higher efficiency than native Vent DNA polymerase. See Gardner et al. (2004) "Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase" J. Biol. Chem. 279(12):11834-11842 and Gardner and Jack (1999) "Determinants of nucleotide sugar recognition in an archaeon DNA polymerase" Nucleic Acids Research 27(12):2545-2553. The altered residue in this mutant, A488, is predicted to be facing away from the nucleotide binding site of the enzyme. The pattern of relaxed specificity at this position roughly correlates with the size of the substituted amino acid side chain and affects incorporation by the enzyme of a variety of modified nucleotide sugars.

More recently, NGS technologies have introduced the need to adapt DNA polymerase enzymes to accept nucleotide substrates modified with reversible terminators on the 3' -OH, such as -ONH₂. To this end, Chen and colleagues combined structural analyses with a "reconstructed evolutionary adaptive path" analysis to generate a TAQ^{L616A} variant that is able to efficiently incorporate both reversible and irreversible terminators. See Chen et al. (2010) "Reconstructed Evolutionary Adaptive Paths Give Polymerases Accepting Reversible Terminators for Sequencing and SNP Detection" Proc. Nat. Acad. Sci. 107(5):1948-1953. Modeling studies suggested that this variant might open space behind Phe-667, allowing it to accommodate the larger 3' substituents. U.S. Patent No. 8,999,676 to Emig et al. discloses additional modified polymerases that display improved properties useful for single molecule sequencing technologies based on fluorescent detection. In particular, substitution of φ29 DNA polymerase at positions E375 and K512 was found to enhance the ability of the

polymerase to utilize non-natural, phosphate-labeled nucleotide analogs incorporating different fluorescent dyes.

Recently, Kokoris et al. have described a method, termed “sequencing by expansion” (SBX), that uses a DNA polymerase to transcribe the sequence of DNA onto a measurable polymer called an Xpandomer (see, e.g., U.S. Patent No. 8,324,360 to Kokoris et al.). The transcribed sequence is encoded along the Xpandomer backbone in high signal-to-noise reporters that are separated by ~10nm and are designed for high signal-to-noise, well differentiated responses when read by nanopore-based sequencing systems. Xpandomers are generated from non-natural nucleotide analogs, termed XNTPs, characterized by bulky substituents that enable the Xpandomer backbone to be expanded following synthesis. Such XNTP analogs introduce novel challenges as substrates for currently available DNA polymerases.

Thus, new modified polymerases, e.g., modified polymerases that display improved properties useful for nanopore-based sequencing and other polymerase applications (e.g., DNA amplification, sequencing, labeling, detection, cloning, etc.), would find value in the art. The present invention provides new recombinant DNA polymerases with desirable properties, including the ability to incorporate nucleotide analogs with bulky substitutions with improved efficiency. Also provided are methods of making and using such polymerases, and many other features that will become apparent upon a complete review of the following.

SUMMARY

Recombinant DNA polymerases and modified DNA polymerases, e.g. modified DPO4, can find use in such applications as, e.g., single-molecule sequencing by expansion (SBX). Among other aspects, the invention provides recombinant DNA polymerases and modified DNA polymerase variants comprising mutations that confer properties, which can be particularly desirable for these applications. These properties can, e.g., improve the ability of the polymerase to utilize bulky nucleotide analogs as substrates during template-dependent polymerization of a daughter strand. Also provided are compositions comprising such DNA polymerases and modified DPO4-

type polymerases, nucleic acids encoding such modified polymerases, methods of generating such modified polymerases and methods in which such polymerases can be used, e.g., to sequence a DNA template.

One general class of embodiments provides an isolated recombinant
5 DNA polymerase, that has an amino acid sequence that is at least 80% identical to amino acids 1-340 of SEQ ID NO:1, and has mutations at amino acid positions 42, 56, 76, 78, 79, 82, 83, 86, 184, 189, 248, 289, 290, 291, 292, 293, and at least one position selected from the group consisting of 7, 24, 26, 57, 77, 87, 97, 116, 152, 153, 155, 156, 157, 164, 170, 173, 174, 176, 183, 187, 188, 190, 191, 192, 196, 211, 212, 214, 215,
10 219, 221, 223, 242, 253, 256, 258, 259, 260, 263, 267, 270, 277, 278, 279, 282, 283, 290, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 308, 309, 314, 317, 320, 321, 322, 324, 325, 326, 327 and 328, wherein identification of positions is relative to wildtype DPO4 polymerase (SEQ ID NO:1), and which recombinant DNA polymerase exhibits polymerase activity. Exemplary mutations at positions 42, 56, 76, 78, 79, 82,
15 83, 86, 184, 189, 248, 289, 290, 291, 292, and 293 include A42V, K56Y, M76W, K78N, E79L, Q82W, Q83G, S86E, P184L, I189W, I248T, V289W, T290K or T290R, E291S, D292Y, and L293. Exemplary mutations at positions 7, 24, 26, 57, 77, 87, 97, 116, 152, 153, 155, 156, 157, 164, 170, 173, 174, 176, 183, 187, 188, 190, 191, 192, 196, 211, 212, 214, 215, 219, 221, 223, 242, 253, 256, 258, 259, 260, 263, 267, 270,
20 277, 278, 279, 282, 283, 290, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 308, 309, 314, 317, 320, 321, 322, 324, 325, 326, 327 and 328, include D7V, K24S, K26L, A57P, R77S, R87Q, E97S, R116N, K152E, K152G, K152L, K152W, K152A, K152Y, K152V, K152M, orK152L, I153M, I153T, I153R, I153H, I153S, I153Y, I153Q, I153K, or I153V, A155G, A155M or A155C D156R, D156L, D156W, D156A,
25 D156G, D156Y, D156S, D156I, D156M, or D156T, K164T, E170M, R173N OR R173G, R176N, V183L, G187P OR G183E, N188Y or N188S, T190Y, T190V or T190F, A191E, E192R, K196S or K196L, D211P, D211S or D211T, K212D or K212F, K214V, K214S, K214A, K214R or K214Q, G215R, G215S, G215K, G215T, G215E or G215A, E219W, E219G, E219A, E219R, E219V or E219L, K221R or K221G, K223C,
30 K223T, K223R, K223V or K223S, R242W, R253N, R256T or R256N, L258P, R259S,

E260Q, K262N, R267Q, E270Q, D277G, K278N, R279Q, K282G or K282S, R290K, D294Q, D294A, D294C, D294S, D294N, D294L, D294P or D294H, I295Y, I295C, I295P, I295T, I295S, I295W, I295Y, I295F, I295M or I295A, V296A, V296R, V296S, V296Q or V296Y, S297H, S297P, S297W, S297L, S297Q, S297Y, S297C, S297R, 5 S297F or S297T, R298L, R298K, R298T, R298E or R298G, G299L, G299W, G299I, G299A, G299H, G299S, G299Y or G299F, R300C, R300S, R300V or R300G, T301R, T301F, T301V, T301W, T301E, T301Y, T301M, T310K, T301S, T301Q, or T301H, F302E, P303R, H304G, K308Q, E309N, E309P or E309D, E314Q, K317Q, Q320K, Q320R, Q320M, Q320I, Q320T, Q320N or Q320S, K321Q, K321N, K321R, K321S, 10 K321G or K321M, I322A, I322L, I322C, I322V or I322T, E324K, E324N, E324A, E324G, E324S, E324R, E324W, E324D or E324K, or E324R, E325K, E325R, E325G, E325V, E325S, E325A or E325Q, D326K, E327K, E337Q, or E327R, and R328N or R328S. In other embodiments, the recombinant DPO4-type DNA polymerase is represented by the amino acid sequence as set forth in any one of SEQ ID NOs: 2-106.

15 In a related aspect, the invention provides compositions containing any of the recombinant DPO4-type DNA polymerase set forth above. In certain embodiments, the compositions may also contain at least one non-natural nucleotide analog substrate.

In another related aspect, the invention provides modified nucleic acids 20 encoding any of the modified DPO4-type DNA polymerase set forth above.

BRIEF DESCRIPTION OF THE FIGURES

FIGS. 1A and 1B show the amino acid sequence of the DPO4 polymerase protein with the Mut_1 through Mut_13 regions outlined and variable amino acids underscored.

25 DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. The following definitions supplement those in the art and

are directed to the current application and are not to be imputed to any related or unrelated case, e.g., to any commonly owned patent or application. Although any methods and materials similar or equivalent to those described herein can be used in the practice for testing of the present invention, the preferred materials and methods are described herein. Accordingly, the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a protein" includes a plurality of proteins; reference to "a cell" includes mixtures of cells, and the like.

The term "about" as used herein indicates the value of a given quantity varies by +/-10% of the value, or optionally +/-5% of the value, or in some embodiments, by +/-1% of the value so described.

"Nucleobase" is a heterocyclic base such as adenine, guanine, cytosine, thymine, uracil, inosine, xanthine, hypoxanthine, or a heterocyclic derivative, analog, or tautomer thereof. A nucleobase can be naturally occurring or synthetic. Non-limiting examples of nucleobases are adenine, guanine, thymine, cytosine, uracil, xanthine, hypoxanthine, 8-azapurine, purines substituted at the 8 position with methyl or bromine, 9-oxo-N6-methyladenine, 2-aminoadenine, 7-deazaxanthine, 7-deazaguanine, 7-deazaadenine, N4-ethanocytosine, 2,6-diaminopurine, N6-ethano-2,6-diaminopurine, 5-methylcytosine, 5-(C3-C6)-alkynylcytosine, 5-fluorouracil, 5-bromouracil, thiouracil, pseudoisocytosine, 2-hydroxy-5-methyl-4-triazolopyridine, isocytosine, isoguanine, inosine, 7,8-dimethylalloxazine, 6-dihydrothymine, 5,6-dihydrouracil, 4-methyl-indole, ethenoadenine and the non-naturally occurring nucleobases described in U.S. Pat. Nos. 5,432,272 and 6,150,510 and PCT Publication Nos. WO 92/002258, WO 93/10820, WO 94/22892, and WO 94/24144, and Fasman ("Practical Handbook of Biochemistry and Molecular Biology", pp. 385-394, 1989, CRC Press, Boca Raton, La.), all herein incorporated by reference in their entireties.

"Nucleobase residue" includes nucleotides, nucleosides, fragments thereof, and related molecules having the property of binding to a complementary

nucleotide. Deoxynucleotides and ribonucleotides, and their various analogs, are contemplated within the scope of this definition. Nucleobase residues may be members of oligomers and probes. "Nucleobase" and "nucleobase residue" may be used interchangeably herein and are generally synonymous unless context dictates otherwise.

5 "Polynucleotides", also called nucleic acids, are covalently linked series of nucleotides in which the 3' position of the pentose of one nucleotide is joined by a phosphodiester group to the 5' position of the next. DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are biologically occurring polynucleotides in which the nucleotide residues are linked in a specific sequence by phosphodiester linkages. As
10 used herein, the terms "polynucleotide" or "oligonucleotide" encompass any polymer compound having a linear backbone of nucleotides. Oligonucleotides, also termed oligomers, are generally shorter chained polynucleotides.

"Nucleic acid" is a polynucleotide or an oligonucleotide. A nucleic acid molecule can be deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or a
15 combination of both. Nucleic acids are generally referred to as "target nucleic acids" or "target sequence" if targeted for sequencing. Nucleic acids can be mixtures or pools of molecules targeted for sequencing.

A "polynucleotide sequence" or "nucleotide sequence" is a polymer of nucleotides (an oligonucleotide, a DNA, a nucleic acid, etc.) or a character string
20 representing a nucleotide polymer, depending on context. From any specified polynucleotide sequence, either the given nucleic acid or the complementary polynucleotide sequence (e.g., the complementary nucleic acid) can be determined.

A "polypeptide" is a polymer comprising two or more amino acid residues (e.g., a peptide or a protein). The polymer can additionally comprise non-
25 amino acid elements such as labels, quenchers, blocking groups, or the like and can optionally comprise modifications such as glycosylation or the like. The amino acid residues of the polypeptide can be natural or non-natural and can be unsubstituted, unmodified, substituted or modified.

An "amino acid sequence" is a polymer of amino acid residues (a protein, polypeptide, etc.) or a character string representing an amino acid polymer, depending on context.

Numbering of a given amino acid or nucleotide polymer "corresponds to
5 numbering of" or is "relative to" a selected amino acid polymer or nucleic acid when the position of any given polymer component (amino acid residue, incorporated nucleotide, etc.) is designated by reference to the same residue position in the selected amino acid or nucleotide polymer, rather than by the actual position of the component in the given polymer. Similarly, identification of a given position within a given amino
10 acid or nucleotide polymer is "relative to" a selected amino acid or nucleotide polymer when the position of any given polymer component (amino acid residue, incorporated nucleotide, etc.) is designated by reference to the residue name and position in the selected amino acid or nucleotide polymer, rather than by the actual name and position of the component in the given polymer. Correspondence of positions is typically
15 determined by aligning the relevant amino acid or polynucleotide sequences.

The term "recombinant" indicates that the material (e.g., a nucleic acid or a protein) has been artificially or synthetically (non-naturally) altered by human intervention. The alteration can be performed on the material within, or removed from, its natural environment or state. For example, a "recombinant nucleic acid" is one that is
20 made by recombining nucleic acids, e.g., during cloning, DNA shuffling or other procedures, or by chemical or other mutagenesis; a "recombinant polypeptide" or "recombinant protein" is, e.g., a polypeptide or protein which is produced by expression of a recombinant nucleic acid.

A "DPO4-type DNA polymerase" is a DNA polymerase naturally
25 expressed by the archaea, *Sulfolobus solfataricus*, or a related Y-family DNA polymerase, which generally function in the replication of damaged DNA by a process known as translesion synthesis (TLS). Y-family DNA polymerases are homologous to the DPO4 polymerase (e.g., as listed in SEQ ID NO:1); examples include the prokaryotic enzymes, PolIII, PolIV, PolV, the archaeal enzyme, Dbh, and the eukaryotic
30 enzymes, Rev3p, Rev1p, Pol η , REV3, REV1, Pol I, and Pol κ DNA polymerases, as

well as chimeras thereof. A modified recombinant DPO4-type DNA polymerase includes one or more mutations relative to naturally-occurring wild-type DPO4-type DNA polymerases, for example, one or more mutations that increase the ability to utilize bulky nucleotide analogs as substrates or another polymerase property, and may include additional alterations or modifications over the wild-type DPO4-type DNA polymerase, such as one or more deletions, insertions, and/or fusions of additional peptide or protein sequences (e.g., for immobilizing the polymerase on a surface or otherwise tagging the polymerase enzyme).

"Template-directed synthesis", "template-directed assembly", "template-directed hybridization", "template-directed binding" and any other template-directed processes, e.g., primer extension, refers to a process whereby nucleotide residues or nucleotide analogs bind selectively to a complementary target nucleic acid, and are incorporated into a nascent daughter strand. A daughter strand produced by a template-directed synthesis is complementary to the single-stranded target from which it is synthesized. It should be noted that the corresponding sequence of a target strand can be inferred from the sequence of its daughter strand, if that is known. "Template-directed polymerization" is a special case of template-directed synthesis whereby the resulting daughter strand is polymerized.

"XNTP" is an expandable, 5' triphosphate modified nucleotide substrate compatible with template dependent enzymatic polymerization. An XNTP has two distinct functional components; namely, a nucleobase 5'-triphosphate and a tether or tether precursor that is attached within each nucleotide at positions that allow for controlled RT expansion by intra-nucleotide cleavage.

"Xpandomer intermediate" is an intermediate product (also referred to herein as a "daughter strand") assembled from XNTPs, and is formed by a template-directed assembly of XNTPs using a target nucleic acid template. The Xpandomer intermediate contains two structures; namely, the constrained Xpandomer and the primary backbone. The constrained Xpandomer comprises all of the tethers in the daughter strand but may comprise all, a portion or none of the nucleobase 5'-triphosphates as required by the method. The primary backbone comprises all of the

abutted nucleobase 5'-triphosphates. Under the process step in which the primary backbone is fragmented or dissociated, the constrained Xpandomer is no longer constrained and is the Xpandomer product which is extended as the tethers are stretched out. "Duplex daughter strand" refers to an Xpandomer intermediate that is hybridized or
5 duplexed to the target template.

"Xpandomer" or "Xpandomer product" is a synthetic molecular construct produced by expansion of a constrained Xpandomer, which is itself synthesized by template-directed assembly of XNTPs. The Xpandomer is elongated relative to the target template it was produced from. It is composed of a concatenation
10 of XNTPs, each XNTP including a tether comprising one or more reporters encoding sequence information. The Xpandomer is designed to expand to be longer than the target template thereby lowering the linear density of the sequence information of the target template along its length. In addition, the Xpandomer optionally provides a platform for increasing the size and abundance of reporters which in turn improves
15 signal to noise for detection. Lower linear information density and stronger signals increase the resolution and reduce sensitivity requirements to detect and decode the sequence of the template strand.

"Tether" or "tether member" refers to a polymer or molecular construct having a generally linear dimension and with an end moiety at each of two opposing
20 ends. A tether is attached to a nucleobase 5'-triphosphate with a linkage in at least one end moiety to form an XNTP. The end moieties of the tether may be connected to cleavable linkages to the nucleobase 5'-triphosphate that serve to constrain the tether in a "constrained configuration". After the daughter strand is synthesized, each end moiety has an end linkage that couples directly or indirectly to other tethers. The coupled
25 tethers comprise the constrained Xpandomer that further comprises the daughter strand. Tethers have a "constrained configuration" and an "expanded configuration". The constrained configuration is found in XNTPs and in the daughter strand. The constrained configuration of the tether is the precursor to the expanded configuration, as found in Xpandomer products. The transition from the constrained configuration to the
30 expanded configuration results cleaving of selectively cleavable bonds that may be

within the primary backbone of the daughter strand or intra-tether linkages. A tether in a constrained configuration is also used where a tether is added to form the daughter strand after assembly of the "primary backbone". Tethers can optionally comprise one or more reporters or reporter constructs along its length that can encode sequence information of substrates. The tether provides a means to expand the length of the Xpandomer and thereby lower the sequence information linear density.

"Tether element" or "tether segment" is a polymer having a generally linear dimension with two terminal ends, where the ends form end-linkages for concatenating the tether elements. Tether elements may be segments of tether constructs. Such polymers can include, but are not limited to: polyethylene glycols, polyglycols, polypyridines, polyisocyanides, polyisocyanates, poly(triarylmethyl)methacrylates, polyaldehydes, polypyrrolinones, polyureas, polyglycol phosphodiester, polyacrylates, polymethacrylates, polyacrylamides, polyvinyl esters, polystyrenes, polyamides, polyurethanes, polycarbonates, polybutyrates, polybutadienes, polybutyrolactones, polypyrrolidinones, polyvinylphosphonates, polyacetamides, polysaccharides, polyhyaluronates, polyamides, polyimides, polyesters, polyethylenes, polypropylenes, polystyrenes, polycarbonates, polyterephthalates, polysilanes, polyurethanes, polyethers, polyamino acids, polyglycines, polyprolines, N-substituted polylysine, polypeptides, side-chain N-substituted peptides, poly-N-substituted glycine, peptoids, side-chain carboxyl-substituted peptides, homopeptides, oligonucleotides, ribonucleic acid oligonucleotides, deoxynucleic acid oligonucleotides, oligonucleotides modified to prevent Watson-Crick base pairing, oligonucleotide analogs, polycytidylic acid, polyadenylic acid, polyuridylic acid, polythymidine, polyphosphate, polynucleotides, polyribonucleotides, polyethylene glycol-phosphodiester, peptide polynucleotide analogues, threosyl-polynucleotide analogues, glycol-polynucleotide analogues, morpholino-polynucleotide analogues, locked nucleotide oligomer analogues, polypeptide analogues, branched polymers, comb polymers, star polymers, dendritic polymers, random, gradient and block copolymers, anionic polymers, cationic polymers, polymers forming stem-loops, rigid segments and flexible segments.

A variety of additional terms are defined or otherwise characterized herein.

DETAILED DESCRIPTION

One aspect of the invention is generally directed to compositions comprising a recombinant polymerase, e.g., a recombinant DPO4-type DNA polymerase that includes one or more mutations as compared to a reference polymerase, e.g., a wildtype DPO4-type polymerase. Depending on the particular mutation or combination of mutations, the polymerase exhibits one or more properties that find use, e.g., in single molecule sequencing applications. Exemplary properties exhibited by various polymerases of the invention include the ability to incorporate “bulky” nucleotide analogs into a growing daughter strand during DNA replication. The polymerases can include one or more exogenous or heterologous features at the N- and/or C-terminal regions of the protein for use, e.g., in the purification of the recombinant polymerase. The polymerases can also include one or more deletions that facilitate purification of the protein, e.g., by increasing the solubility of recombinantly produced protein.

These new polymerases are particularly well suited to DNA replication and/or sequencing applications, particularly sequencing protocols that include incorporation of bulky nucleotide analogs into a replicated nucleic acid daughter strand, such as in the sequencing by expansion (SBX) protocol, as further described below.

Polymerases of the invention include, for example, a recombinant DPO4-type DNA polymerase that comprises mutations at positions 42, 56, 76, 78, 79, 82, 83, 86, 184, 189, 248, 289, 290, 291, 292, 293, and at least one position selected from the group consisting of 7, 24, 26, 57, 77, 87, 97, 116, 152, 153, 155, 156, 157, 164, 170, 173, 174, 176, 183, 187, 188, 190, 191, 192, 196, 211, 212, 214, 215, 219, 221, 223, 242, 253, 256, 258, 259, 260, 263, 267, 270, 277, 278, 279, 282, 283, 290, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 308, 309, 314, 317, 320, 321, 322, 324, 325, 326, 327 and 328 wherein identification of positions is relative to wild-type DPO4 polymerase (SEQ ID NO:1). Optionally, the polymerase comprises

mutations at 16 or more, up to 20 or more, up to 30 or more, up to 40 or more, or from 40 to 70 or more or more of these positions. The polymerases of the invention may also possess a deletion of amino acids 341-352 of the wildtype protein, corresponding to the "PIP box". In some embodiments, the polymerases of the invention may include
5 mutations at additional residues not cited herein, provided that such mutations provide functional advantages as discussed further herein. In certain embodiments the polymerases of the invention are at least 80% identical to SEQ ID NO:1 (amino acids 1-340 of wildtype DPO4 DNA polymerase. In other embodiments, the polymerases of the invention may be less than 80% identical to SEQ ID NO:1, provided that such
10 polymerases demonstrate enhanced abilities to utilize XNTPs as polymerization substrates. A number of exemplary substitutions at these (and other) positions are described herein.

DNA Polymerases

DNA polymerases that can be modified to increase the ability to
15 incorporate bulky nucleotide analog substrates into a growing daughter nucleic acid strand and/or other desirable properties as described herein are generally available. DNA polymerases are sometimes classified into six main groups, or families, based upon various phylogenetic relationships, e.g., with *E. coli* Pol I (class A), *E. coli* Pol II (class B), *E. coli* Pol III (class C), Euryarchaeotic Pol II (class D), human Pol beta (class X), and *E. coli* UmuC/DinB and eukaryotic RAD30/xeroderma pigmentosum
20 variant (class Y). For a review of recent nomenclature, see, e.g., Burgers et al. (2001) "Eukaryotic DNA polymerases: proposal for a revised nomenclature" *J Biol. Chem.* 276(47):43487-90. For a review of polymerases, see, e.g., Hubscher et al. (2002) "Eukaryotic DNA Polymerases" *Annual Review of Biochemistry* Vol. 71: 133-163;
25 Alba (2001) "Protein Family Review: Replicative DNA Polymerases" *Genome Biology* 2(1): reviews 3002.1-3002.4; and Steitz (1999) "DNA polymerases: structural diversity and common mechanisms" *J Biol Chem* 274:17395-17398. DNA polymerase have been extensively studied and the basic mechanisms of action for many have been determined. In addition, the sequences of literally hundreds of polymerases are publicly

available, and the crystal structures for many of these have been determined or can be inferred based upon similarity to solved crystal structures for homologous polymerases. For example, the crystal structure of DPO4, a preferred type of parental enzyme to be modified according to the present invention, is available see, e.g., Ling et al. (2001)
5 “Crystal Structure of a Y-Family DNA Polymerase in Action: A Mechanism for Error-Prone and Lesion-Bypass Replication” Cell 107:91-102.

DNA polymerases that are preferred substrates for mutation to increase the use of bulky nucleotide analog as substrates for incorporation into growing nucleic acid daughter strands, and/or to alter one or more other property described herein
10 include DPO4 polymerases and other members of the Y family of translesional DNA polymerases, such as Dbh, and derivatives of such polymerases.

In one aspect, the polymerase that is modified is a DPO4-type DNA polymerase. For example, the modified recombinant DNA polymerase can be homologous to a wildtype DPO4 DNA polymerase. Alternately, the modified
15 recombinant DNA polymerase can be homologous to other Class Y DNA polymerases, also known as “translesion” DNA polymerases, such as *Sulfolobus acidocaldarius* Dbh polymerase. For a review, see Goodwin and Woodgate (2013) “Translesion DNA Polymerases” Cold Spring Harb Perspect in Biol doi:10.1101/cshperspect.a010363. See, e.g., SEQ ID NO:1 for the amino acid sequence of wildtype DPO4 polymerase.

20 In other aspects, the polymerase that is modified is a DNA Pol Kappa-type polymerase, a DNA Pol Eta-type polymerase, a PrimPol-type polymerase, or a Terminator Gamma-type polymerase derived from any suitable species, such as yeast, human, *S. islandicus*, or *T. thermophilus*. The polymerase that is modified may be full length or truncated versions that include or lack various features of the protein. Certain
25 desired features of these polymerases may also be combined with any of the DPO4 variants disclosed herein.

Many polymerases that are suitable for modification, e.g., for use in sequencing technologies, are commercially available. For example, DPO4 polymerase is available from TREVEGAN[®] and New England Biolabs[®].

In addition to wildtype polymerases, chimeric polymerases made from a mosaic of different sources can be used. For example, DPO4-type polymerases made by taking sequences from more than one parental polymerase into account can be used as a starting point for mutation to produce the polymerases of the invention. Chimeras can be produced, e.g., using consideration of similarity regions between the polymerases to define consensus sequences that are used in the chimera, or using gene shuffling technologies in which multiple DPO4-related polymerases are randomly or semi-randomly shuffled via available gene shuffling techniques (e.g., via "family gene shuffling"; see Crameri et al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291; Clackson et al. (1991) "Making antibody fragments using phage display libraries" *Nature* 352:624-628; Gibbs et al. (2001) "Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling" *Gene* 271:13-20; and Hiraga and Arnold (2003) "General method for sequence-independent site-directed chimeragenesis: *J. Mol. Biol.* 330:287-296). In these methods, the recombination points can be predetermined such that the gene fragments assemble in the correct order. However, the combinations, e.g., chimeras, can be formed at random. Appropriate mutations to improve incorporation of bulky nucleotide analog substrates or another desirable property can be introduced into the chimeras.

20 Nucleotide Analogs

As discussed, various polymerases of the invention can incorporate one or more nucleotide analogs into a growing oligonucleotide chain. Upon incorporation, the analog can leave a residue that is the same as or different than a natural nucleotide in the growing oligonucleotide (the polymerase can incorporate any non-standard moiety of the analog, or can cleave it off during incorporation into the oligonucleotide). A "nucleotide analog" herein is a compound, that, in a particular application, functions in a manner similar or analogous to a naturally occurring nucleoside triphosphate (a "nucleotide"), and does not otherwise denote any particular structure. A nucleotide analog is an analog other than a standard naturally occurring nucleotide, i.e., other than

A, G, C, T, or U, though upon incorporation into the oligonucleotide, the resulting residue in the oligonucleotide can be the same as (or different from) an A, G, C, T, or U residue.

Many nucleotide analogs are available and can be incorporated by the
 5 polymerases of the invention. These include analog structures with core similarity to naturally occurring nucleotides, such as those that comprise one or more substituent on a phosphate, sugar, or base moiety of the nucleoside or nucleotide relative to a naturally occurring nucleoside or nucleotide.

In one useful aspect of the invention, nucleotide analogs can also be
 10 modified to achieve any of the improved properties desired. For example, various tethers, linkers, or other substituents can be incorporated into analogs to create a “bulky” nucleotide analog, wherein the term “bulky” is understood to mean that the size of the analog is substantially larger than a natural nucleotide, while not denoting any particular dimension. For example, the analog can include a substituted compound (i.e.,
 15 a “XNTP”, as disclosed in U.S. Patent No. 7,939,259 and PCT Publication No. WO 2016/081871 to Kokoris et al.) of the formula:



As shown in the above formula, the monomeric XNTP construct has a
 nucleobase residue, N, that has two moieties separated by a selectively cleavable bond
 20 (V), each moiety attaching to one end of a tether (T). The tether ends can attach to the linker group modifications on the heterocycle, the ribose group, or the phosphate backbone. The monomer substrate also has an intra-substrate cleavage site positioned within the phosphoribosyl backbone such that cleavage will provide expansion of the constrained tether. For example, to synthesize a XATP monomer, the amino linker on
 25 8-[(6-Amino)hexyl]-amino-ATP or N6-(6-Amino)hexyl-ATP can be used as a first tether attachment point, and, a mixed backbone linker, such as the non-bridging modification (N-1-aminoalkyl) phosphoramidate or (2-aminoethyl) phosphonate, can be

used as a second tether attachment point. Further, a bridging backbone modification such as a phosphoramidate (3' O--P--N 5') or a phosphorothiolate (3' O--P--S 5'), for example, can be used for selective chemical cleavage of the primary backbone. R¹ and R² are end groups configured as appropriate for the synthesis protocol in which the substrate construct is used. For example, R¹=5'-triphosphate and R²=3'-OH for a polymerase protocol. The R¹ 5' triphosphate may include mixed backbone modifications, such as an aminoethyl phosphonate or 3'-O--P--S-5' phosphorothiolate, to enable tether linkage and backbone cleavage, respectively. Optionally, R² can be configured with a reversible blocking group for cyclical single-substrate addition.

10 Alternatively, R¹ and R² can be configured with linker end groups for chemical coupling. R¹ and R² can be of the general type XR, wherein X is a linking group and R is a functional group. Detailed atomic structures of suitable substrates for polymerase variants of the present invention may be found, e.g., in Vaghefi, M. (2005) "Nucleoside Triphosphates and their Analogs" CRC Press Taylor & Francis Group.

15 Applications for Increased Abilities to Incorporate Bulky Nucleotide Analog Substrates

Polymerases of the invention, e.g., modified recombinant polymerases, or variants, may be used in combination with nucleotides and/or nucleotide analogs and nucleic acid templates (DNA or RNA) to copy the template nucleic acid. That is, a mixture of the polymerase, nucleotides/analog, and optionally other appropriate reagents, the template and a replication initiating moiety (e.g., primer) is reacted such that the polymerase synthesizes a daughter nucleic acid strand (e.g., extends the primer) in a template-dependent manner. The replication initiating moiety can be a standard oligonucleotide primer, or, alternatively, a component of the template, e.g., the template can be a self-priming single stranded DNA, a nicked double stranded DNA, or the like.

20 Similarly, a terminal protein can serve as an initiating moiety. At least one nucleotide analog can be incorporated into the DNA. The template DNA can be a linear or circular DNA, and in certain applications, is desirably a circular template (e.g., for rolling circle replication or for sequencing of circular templates). Optionally, the composition can be present in an automated DNA replication and/or sequencing system.

25

In one embodiment, the daughter nucleic acid strand is an Xpandomer intermediate comprised of XNTPs, as disclosed in U.S. Patent No. 7,939,259, and PCT Publication No. WO 2016/081871 to Kokoris et al. and assigned to Stratos Genomics, which are herein incorporated by reference in its entirety. Stratos Genomics has developed a method called Sequencing by Expansion (“SBX”) that uses a DNA polymerase to transcribe the sequence of DNA onto a measurable polymer called an “Xpandomer”. In general terms, an Xpandomer encodes (parses) the nucleotide sequence data of the target nucleic acid in a linearly expanded format, thereby improving spatial resolution, optionally with amplification of signal strength. The transcribed sequence is encoded along the Xpandomer backbone in high signal-to-noise reporters that are separated by ~10 nm and are designed for high-signal-to-noise, well-differentiated responses. These differences provide significant performance enhancements in sequence read efficiency and accuracy of Xpandomers relative to native DNA. Xpandomers can enable several next generation DNA sequencing technologies and are well suited to nanopore sequencing. As discussed above, one method of Xpandomer synthesis uses XNTPs as nucleic acid analogs to extend the template-dependent synthesis and uses a DNA polymerase variant as a catalyst.

Mutating Polymerases

Various types of mutagenesis are optionally used in the present invention, e.g., to modify polymerases to produce variants, e.g., in accordance with polymerase models and model predictions as discussed above, or using random or semi-random mutational approaches. In general, any available mutagenesis procedure can be used for making polymerase mutants. Such mutagenesis procedures optionally include selection of mutant nucleic acids and polypeptides for one or more activity of interest (e.g., the ability to incorporate bulky nucleotide analogs into a daughter nucleic acid strand). Procedures that can be used include, but are not limited to: site-directed point mutagenesis, random point mutagenesis, in vitro or in vivo homologous recombination (DNA shuffling and combinatorial overlap PCR), mutagenesis using uracil containing templates, oligonucleotide-directed mutagenesis, phosphorothioate-modified DNA

mutagenesis, mutagenesis using gapped duplex DNA, point mismatch repair, mutagenesis using repair-deficient host strains, restriction-selection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, degenerate PCR, double-strand break repair, and many others known to persons of skill. The starting polymerase for mutation can be any of those noted herein, including wildtype DPO4 polymerase.

Optionally, mutagenesis can be guided by known information (e.g., "rational" or "semi-rational" design) from a naturally occurring polymerase molecule, or of a known altered or mutated polymerase (e.g., using an existing mutant polymerase as noted in the preceding references), e.g., sequence, sequence comparisons, physical properties, crystal structure and/or the like as discussed above. However, in another class of embodiments, modification can be essentially random (e.g., as in classical or "family" DNA shuffling, see, e.g., Cramer et al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291.

Additional information on mutation formats is found in: Sambrook et al., *Molecular Cloning--A Laboratory Manual* (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 2000 ("Sambrook"); *Current Protocols in Molecular Biology*, F. M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2011) ("Ausubel") and *PCR Protocols A Guide to Methods and Applications* (Innis et al. eds) Academic Press Inc. San Diego, Calif. (1990) ("Innis"). The following publications and references cited within provide additional detail on mutation formats: Arnold, Protein engineering for unusual environments, *Current Opinion in Biotechnology* 4:450-455 (1993); Bass et al., Mutant Trp repressors with new DNA-binding specificities, *Science* 242:240-245 (1988); Bordo and Argos (1991) Suggestions for "Safe" Residue Substitutions in Site-directed Mutagenesis 217:721-729; Botstein & Shortle, Strategies and applications of in vitro mutagenesis, *Science* 229:1193-1201 (1985); Carter et al., Improved oligonucleotide site-directed mutagenesis using M13 vectors, *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, Site-directed mutagenesis, *Biochem. J.* 237:1-7 (1986); Carter, Improved oligonucleotide-

directed mutagenesis using M13 vectors, *Methods in Enzymol.* 154: 382-403 (1987); Dale et al., Oligonucleotide-directed random mutagenesis using the phosphorothioate method, *Methods Mol. Biol.* 57:369-374 (1996); Eghtedarzadeh & Henikoff, Use of oligonucleotides to generate large deletions, *Nucl. Acids Res.* 14: 5115 (1986); Fritz et al., Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro, *Nucl. Acids Res.* 16: 6987-6999 (1988); Grundstrom et al., Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis, *Nucl. Acids Res.* 13: 3305-3316 (1985); Hayes (2002) Combining Computational and Experimental Screening for rapid Optimization of Protein Properties PNAS 99(25) 15926-15931; Kunkel, The efficiency of oligonucleotide directed mutagenesis, in *Nucleic Acids & Molecular Biology* (Eckstein, F. and Lilley, D. M. J. eds., Springer Verlag, Berlin) (1987); Kunkel, Rapid and efficient site-specific mutagenesis without phenotypic selection, *Proc. Natl. Acad. Sci. USA* 82:488-492 (1985); Kunkel et al., Rapid and efficient site-specific mutagenesis without phenotypic selection, *Methods in Enzymol.* 154, 367-382 (1987); Kramer et al., The gapped duplex DNA approach to oligonucleotide-directed mutation construction, *Nucl. Acids Res.* 12: 9441-9456 (1984); Kramer & Fritz Oligonucleotide-directed construction of mutations via gapped duplex DNA, *Methods in Enzymol.* 154:350-367 (1987); Kramer et al., Point Mismatch Repair, *Cell* 38:879-887 (1984); Kramer et al., Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations, *Nucl. Acids Res.* 16: 7207 (1988); Ling et al., Approaches to DNA mutagenesis: an overview, *Anal Biochem.* 254(2): 157-178 (1997); Lorimer and Pastan *Nucleic Acids Res.* 23, 3067-8 (1995); Mandecki, Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis, *Proc. Natl. Acad. Sci. USA*, 83:7177-7181(1986); Nakamaye & Eckstein, Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 14: 9679-9698 (1986); Nambiar et al., Total synthesis and cloning of a gene coding for the ribonuclease S protein, *Science* 223: 1299-1301(1984); Sakamar and Khorana, Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment

guanine nucleotide-binding protein (transducin), *Nucl. Acids Res.* 14: 6361-6372 (1988); Sayers et al., Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 16:791-802 (1988); Sayers et al., Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide, (1988) *Nucl. Acids Res.* 16: 803-814; Sieber, et al., *Nature Biotechnology*, 19:456-460 (2001); Smith, In vitro mutagenesis, *Ann. Rev. Genet.* 19:423-462 (1985); *Methods in Enzymol.* 100: 468-500 (1983); *Methods in Enzymol.* 154: 329-350 (1987); Stemmer, *Nature* 370, 389-91(1994); Taylor et al., The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA, *Nucl. Acids Res.* 13: 8749-8764 (1985); Taylor et al., The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA, *Nucl. Acids Res.* 13: 8765-8787 (1985); Wells et al., Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin, *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986); Wells et al., Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites, *Gene* 34:315-323 (1985); Zoller & Smith, Oligonucleotide-directed mutagenesis using M 13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment, *Nucleic Acids Res.* 10:6487-6500 (1982); Zoller & Smith, Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, *Methods in Enzymol.* 100:468-500 (1983); Zoller & Smith, Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template, *Methods in Enzymol.* 154:329-350 (1987); Clackson et al. (1991) "Making antibody fragments using phage display libraries" *Nature* 352:624-628; Gibbs et al. (2001) "Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling" *Gene* 271:13-20; and Hiraga and Arnold (2003) "General method for sequence-independent site-directed chimeragenesis: *J. Mol. Biol.* 330:287-296. Additional details on many of the above methods can be found in *Methods in Enzymology* Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

Screening Polymerases

Screening or other protocols can be used to determine whether a polymerase displays a modified activity, e.g., for a nucleotide analog, as compared to a parental DNA polymerase. For example, the ability to bind and incorporate bulky nucleotide analogs into a daughter strand during template-dependent DNA synthesis. Assays for such properties, and the like, are described herein. Performance of a recombinant polymerase in a primer extension reaction can be examined to assay properties such as nucleotide analog incorporations etc., as described herein.

In one desirable aspect, a library of recombinant DNA polymerases can be made and screened for these properties. For example, a plurality of members of the library can be made to include one or more mutation that alters incorporations and/or randomly generated mutations (e.g., where different members include different mutations or different combinations of mutations), and the library can then be screened for the properties of interest (e.g., incorporations, etc.). In general, the library can be screened to identify at least one member comprising a modified activity of interest.

Libraries of polymerases can be either physical or logical in nature. Moreover, any of a wide variety of library formats can be used. For example, polymerases can be fixed to solid surfaces in arrays of proteins. Similarly, liquid phase arrays of polymerases (e.g., in microwell plates) can be constructed for convenient high-throughput fluid manipulations of solutions comprising polymerases. Liquid, emulsion, or gel-phase libraries of cells that express recombinant polymerases can also be constructed, e.g., in microwell plates, or on agar plates. Phage display libraries of polymerases or polymerase domains (e.g., including the active site region or interdomain stability regions) can be produced. Likewise, yeast display libraries can be used. Instructions in making and using libraries can be found, e.g., in Sambrook, Ausubel and Berger, referenced herein.

For the generation of libraries involving fluid transfer to or from microtiter plates, a fluid handling station is optionally used. Several "off the shelf" fluid handling stations for performing such transfers are commercially available, including e.g., the Zymate systems from Caliper Life Sciences (Hopkinton, Mass.) and other

stations which utilize automatic pipettors, e.g., in conjunction with the robotics for plate movement (e.g., the ORCA® robot, which is used in a variety of laboratory systems available, e.g., from Beckman Coulter, Inc. (Fullerton, Calif.).

In an alternate embodiment, fluid handling is performed in microchips, e.g., involving transfer of materials from microwell plates or other wells through microchannels on the chips to destination sites (microchannel regions, wells, chambers or the like). Commercially available microfluidic systems include those from Hewlett-Packard/Agilent Technologies (e.g., the HP2100 bioanalyzer) and the Caliper High Throughput Screening System. The Caliper High Throughput Screening System provides one example interface between standard microwell library formats and Labchip technologies. RainDance Technologies' nanodroplet platform provides another method for handling large numbers of spatially separated reactions. Furthermore, the patent and technical literature includes many examples of microfluidic systems which can interface directly with microwell plates for fluid handling.

15 Tags and Other Optional Polymerase Features

The recombinant DNA polymerase optionally includes additional features exogenous or heterologous to the polymerase. For example, the recombinant polymerase optionally includes one or more tags, e.g., purification, substrate binding, or other tags, such as a polyhistidine tag, a His10 tag, a His6 tag, an alanine tag, an Ala16 tag, an Ala16 tag, a biotin tag, a biotin ligase recognition sequence or other biotin attachment site (e.g., a BiTag or a Btag or variant thereof, e.g., BtagV1-11), a GST tag, an S Tag, a SNAP-tag, an HA tag, a DSB (Sso7D) tag, a lysine tag, a NanoTag, a Cmyc tag, a tag or linker comprising the amino acids glycine and serine, a tag or linker comprising the amino acids glycine, serine, alanine and histidine, a tag or linker comprising the amino acids glycine, arginine, lysine, glutamine and proline, a plurality of polyhistidine tags, a plurality of His10 tags, a plurality of His6 tags, a plurality of alanine tags, a plurality of Ala10 tags, a plurality of Ala16 tags, a plurality of biotin tags, a plurality of GST tags, a plurality of BiTags, a plurality of S Tags, a plurality of SNAP-tags, a plurality of HA tags, a plurality of DSB (Sso7D) tags, a plurality of lysine

tags, a plurality of NanoTags, a plurality of Cmyc tags, a plurality of tags or linkers comprising the amino acids glycine and serine, a plurality of tags or linkers comprising the amino acids glycine, serine, alanine and histidine, a plurality of tags or linkers comprising the amino acids glycine, arginine, lysine, glutamine and proline, biotin, 5 avidin, an antibody or antibody domain, antibody fragment, antigen, receptor, receptor domain, receptor fragment, or ligand, one or more protease site (e.g., Factor Xa, enterokinase, or thrombin site), a dye, an acceptor, a quencher, a DNA binding domain (e.g., a helix-hairpin-helix domain from topoisomerase V), or combination thereof. The one or more exogenous or heterologous features at the N- and/or C-terminal regions of 10 the polymerase can find use not only for purification purposes, immobilization of the polymerase to a substrate, and the like, but can also be useful for altering one or more properties of the polymerase.

The one or more exogenous or heterologous features can be included internal to the polymerase, at the N-terminal region of the polymerase, at the C-terminal 15 region of the polymerase, or both the N-terminal and C-terminal regions of the polymerase. Where the polymerase includes an exogenous or heterologous feature at both the N-terminal and C-terminal regions, the exogenous or heterologous features can be the same (e.g., a polyhistidine tag, e.g., a His10 tag, at both the N- and C-terminal regions) or different (e.g., a biotin ligase recognition sequence at the N-terminal region 20 and a polyhistidine tag, e.g., His10 tag, at the C-terminal region). Optionally, a terminal region (e.g., the N- or C-terminal region) of a polymerase of the invention can comprise two or more exogenous or heterologous features which can be the same or different (e.g., a biotin ligase recognition sequence and a polyhistidine tag at the N-terminal region, a biotin ligase recognition sequence, a polyhistidine tag, and a Factor Xa 25 recognition site at the N-terminal region, and the like). As a few examples, the polymerase can include a polyhistidine tag at the C-terminal region, a biotin ligase recognition sequence and a polyhistidine tag at the N-terminal region, a biotin ligase recognition sequence and a polyhistidine tag at the N-terminal region and a polyhistidine tag at the C-terminal region, or a polyhistidine tag and a biotin ligase 30 recognition sequence at the C-terminal region.

Making and Isolating Recombinant Polymerases

Generally, nucleic acids encoding a polymerase of the invention can be made by cloning, recombination, in vitro synthesis, in vitro amplification and/or other available methods. A variety of recombinant methods can be used for expressing an expression vector that encodes a polymerase of the invention. Methods for making recombinant nucleic acids, expression and isolation of expressed products are well known and described in the art. A number of exemplary mutations and combinations of mutations, as well as strategies for design of desirable mutations, are described herein. Methods for making and selecting mutations in the active site of polymerases, including for modifying steric features in or near the active site to permit improved access by nucleotide analogs are found hereinabove and, e.g., in PCT Publication Nos. WO 2007/076057 and WO 2008/051530.

Additional useful references for mutation, recombinant and in vitro nucleic acid manipulation methods (including cloning, expression, PCR, and the like) include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, Calif. (Berger); Kaufman et al. (2003) Handbook of Molecular and Cellular Methods in Biology and Medicine Second Edition Ceske (ed) CRC Press (Kaufman); and The Nucleic Acid Protocols Handbook Ralph Rapley (ed) (2000) Cold Spring Harbor, Humana Press Inc (Rapley); Chen et al. (ed) PCR Cloning Protocols, Second Edition (Methods in Molecular Biology, volume 192) Humana Press; and in Viljoen et al. (2005) Molecular Diagnostic PCR Handbook Springer, ISBN 1402034032.

In addition, a plethora of kits are commercially available for the purification of plasmids or other relevant nucleic acids from cells, (see, e.g., EasyPrep™ FlexiPrep™ both from Pharmacia Biotech; StrataClean™, from Stratagene; and, QIAprep™ from Qiagen). Any isolated and/or purified nucleic acid can be further manipulated to produce other nucleic acids, used to transfect cells, incorporated into related vectors to infect organisms for expression, and/or the like. Typical cloning vectors contain transcription and translation terminators, transcription and translation initiation sequences, and promoters useful for regulation of the expression of the

particular target nucleic acid. The vectors optionally comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the cassette in eukaryotes, or prokaryotes, or both, (e.g., shuttle vectors) and selection markers for both prokaryotic and eukaryotic systems. Vectors are suitable
5 for replication and integration in prokaryotes, eukaryotes, or both.

Other useful references, e.g. for cell isolation and culture (e.g., for subsequent nucleic acid isolation) include Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, third edition, Wiley-Liss, New York and the references cited therein; Payne et al. (1992) *Plant Cell and Tissue Culture in Liquid Systems* John
10 Wiley & Sons, Inc. New York, N.Y.; Gamborg and Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual*, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, Fla.

Nucleic acids encoding the recombinant polymerases of the invention
15 are also a feature of the invention. A particular amino acid can be encoded by multiple codons, and certain translation systems (e.g., prokaryotic or eukaryotic cells) often exhibit codon bias, e.g., different organisms often prefer one of the several synonymous codons that encode the same amino acid. As such, nucleic acids of the invention are optionally "codon optimized," meaning that the nucleic acids are synthesized to include
20 codons that are preferred by the particular translation system being employed to express the polymerase. For example, when it is desirable to express the polymerase in a bacterial cell (or even a particular strain of bacteria), the nucleic acid can be synthesized to include codons most frequently found in the genome of that bacterial cell, for efficient expression of the polymerase. A similar strategy can be employed when it is
25 desirable to express the polymerase in a eukaryotic cell, e.g., the nucleic acid can include codons preferred by that eukaryotic cell.

A variety of protein isolation and detection methods are known and can be used to isolate polymerases, e.g., from recombinant cultures of cells expressing the recombinant polymerases of the invention. A variety of protein isolation and detection
30 methods are well known in the art, including, e.g., those set forth in R. Scopes, *Protein*

Purification, Springer-Verlag, N.Y. (1982); Deutscher, Methods in Enzymology Vol. 182: Guide to Protein Purification, Academic Press, Inc. N.Y. (1990); Sandana (1997) Bioseparation of Proteins, Academic Press, Inc.; Bollag et al. (1996) Protein Methods, 2.sup.nd Edition Wiley-Liss, NY; Walker (1996) The Protein Protocols Handbook
5 Humana Press, NJ, Harris and Angal (1990) Protein Purification Applications: A Practical Approach IRL Press at Oxford, Oxford, England; Harris and Angal Protein Purification Methods: A Practical Approach IRL Press at Oxford, Oxford, England; Scopes (1993) Protein Purification: Principles and Practice 3.sup.rd Edition Springer Verlag, NY; Janson and Ryden (1998) Protein Purification: Principles, High Resolution
10 Methods and Applications, Second Edition Wiley-VCH, NY; and Walker (1998) Protein Protocols on CD-ROM Humana Press, NJ; and the references cited therein. Additional details regarding protein purification and detection methods can be found in Satinder Ahuja ed., Handbook of Bioseparations, Academic Press (2000).

Nucleic Acid and Polypeptide Sequences and Variants

15 As described herein, the invention also features polynucleotide sequences encoding, e.g., a polymerase as described herein. Examples of polymerase sequences that include features found herein, e.g., as in Table 2 are provided. However, one of skill in the art will immediately appreciate that the invention is not limited to the specifically exemplified sequences. For example, one of skill will appreciate that the
20 invention also provides, e.g., many related sequences with the functions described herein, e.g., polynucleotides and polypeptides encoding conservative variants of a polymerase of Tables 2 and 3 or any other specifically listed polymerase herein. Combinations of any of the mutations noted herein are also features of the invention.

Accordingly, the invention provides a variety of polypeptides
25 (polymerases) and polynucleotides (nucleic acids that encode polymerases). Exemplary polynucleotides of the invention include, e.g., any polynucleotide that encodes a polymerase of Table 2 or otherwise described herein. Because of the degeneracy of the genetic code, many polynucleotides equivalently encode a given polymerase sequence. Similarly, an artificial or recombinant nucleic acid that hybridizes to a polynucleotide

indicated above under highly stringent conditions over substantially the entire length of the nucleic acid (and is other than a naturally occurring polynucleotide) is a polynucleotide of the invention. In one embodiment, a composition includes a polypeptide of the invention and an excipient (e.g., buffer, water, pharmaceutically acceptable excipient, etc.). The invention also provides an antibody or antisera specifically immunoreactive with a polypeptide of the invention (e.g., that specifically recognizes a feature of the polymerase that confers decreased branching or increased complex stability.

In certain embodiments, a vector (e.g., a plasmid, a cosmid, a phage, a virus, etc.) comprises a polynucleotide of the invention. In one embodiment, the vector is an expression vector. In another embodiment, the expression vector includes a promoter operably linked to one or more of the polynucleotides of the invention. In another embodiment, a cell comprises a vector that includes a polynucleotide of the invention.

One of skill will also appreciate that many variants of the disclosed sequences are included in the invention. For example, conservative variations of the disclosed sequences that yield a functionally similar sequence are included in the invention. Variants of the nucleic acid polynucleotide sequences, wherein the variants hybridize to at least one disclosed sequence, are considered to be included in the invention. Unique subsequences of the sequences disclosed herein, as determined by, e.g., standard sequence comparison techniques, are also included in the invention.

Conservative Variations

Owing to the degeneracy of the genetic code, "silent substitutions" (i.e., substitutions in a nucleic acid sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of every nucleic acid sequence that encodes an amino acid sequence. Similarly, "conservative amino acid substitutions," where one or a limited number of amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties, are also readily identified as

being highly similar to a disclosed construct. Such conservative variations of each disclosed sequence are a feature of the present invention.

"Conservative variations" of a particular nucleic acid sequence refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or, where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. One of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 4%, 2% or 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the deletion of an amino acid, addition of an amino acid, or substitution of an amino acid with a chemically similar amino acid, while retaining the relevant mutational feature (for example, the conservative substitution can be of a residue distal to the active site region, or distal to an interdomain stability region). Thus, "conservative variations" of a listed polypeptide sequence of the present invention include substitutions of a small percentage, typically less than 5%, more typically less than 2% or 1%, of the amino acids of the polypeptide sequence, with an amino acid of the same conservative substitution group. Finally, the addition of sequences which do not alter the encoded activity of a nucleic acid molecule, such as the addition of a non-functional or tagging sequence (introns in the nucleic acid, poly His or similar sequences in the encoded polypeptide, etc.), is a conservative variation of the basic nucleic acid or polypeptide.

Conservative substitution tables providing functionally similar amino acids are well known in the art, where one amino acid residue is substituted for another amino acid residue having similar chemical properties (e.g., aromatic side chains or positively charged side chains), and therefore does not substantially change the functional properties of the polypeptide molecule. The following sets forth example groups that contain natural amino acids of like chemical properties, where substitutions within a group is a "conservative substitution".

Table 1

Conservative Amino Acid Substitutions				
Nonpolar and/or aliphatic side chains	Polar, uncharged side chains	Aromatic side chains	Positively charged side chains	Negatively charged side chains
Glycine Alanine Valine Leucine Isoleucine Proline	Serine Threonine Cysteine Methionine Asparagine Glutamine	Phenylalanine Tyrosine Tryptophan	Lysine Arginine Histidine	Aspartate Glutamate

Nucleic Acid Hybridization

Comparative hybridization can be used to identify nucleic acids of the invention, including conservative variations of nucleic acids of the invention. In addition, target nucleic acids which hybridize to a nucleic acid of the invention under high, ultra-high and ultra-ultra high stringency conditions, where the nucleic acids encode mutants corresponding to those noted in Tables 2 and 3 or other listed polymerases, are a feature of the invention. Examples of such nucleic acids include those with one or a few silent or conservative nucleic acid substitutions as compared to a given nucleic acid sequence encoding a polymerase of Table 2 (or other exemplified polymerase), where any conservative substitutions are for residues other than those noted in Table 2 or elsewhere as being relevant to a feature of interest (improved nucleotide analog incorporations, etc.).

A test nucleic acid is said to specifically hybridize to a probe nucleic acid when it hybridizes at least 50% as well to the probe as to the perfectly matched complementary target, i.e., with a signal to noise ratio at least half as high as hybridization of the probe to the target under conditions in which the perfectly matched probe binds to the perfectly matched complementary target with a signal to noise ratio that is at least about 5x-10x as high as that observed for hybridization to any of the unmatched target nucleic acids.

Nucleic acids "hybridize" when they associate, typically in solution. Nucleic acids hybridize due to a variety of well characterized physico-chemical forces,

such as hydrogen bonding, solvent exclusion, base stacking and the like. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes part I chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, N.Y.), as well as in Current Protocols in Molecular Biology, Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2011); Hames and Higgins (1995) Gene Probes 1 IRL Press at Oxford University Press, Oxford, England, (Hames and Higgins 1) and Hames and Higgins (1995) Gene Probes 2 IRL Press at Oxford University Press, Oxford, England (Hames and Higgins 2) provide details on the synthesis, labeling, detection and quantification of DNA and RNA, including oligonucleotides.

An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin with 1 mg of heparin at 42°C with the hybridization being carried out overnight. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (see, Sambrook, supra for a description of SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove background probe signal. An example low stringency wash is 2x SSC at 40°C for 15 minutes. In general, a signal to noise ratio of 5x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

"Stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and northern hybridizations are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993), supra. and in Hames and Higgins, 1 and 2. Stringent hybridization and wash conditions can easily be determined empirically for any test nucleic acid. For example, in determining stringent hybridization and wash conditions, the hybridization and wash conditions are gradually increased (e.g., by increasing temperature, decreasing salt concentration, increasing

detergent concentration and/or increasing the concentration of organic solvents such as formalin in the hybridization or wash), until a selected set of criteria are met. For example, in highly stringent hybridization and wash conditions, the hybridization and wash conditions are gradually increased until a probe binds to a perfectly matched complementary target with a signal to noise ratio that is at least 5x as high as that observed for hybridization of the probe to an unmatched target

"Very stringent" conditions are selected to be equal to the thermal melting point (T_m) for a particular probe. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the test sequence hybridizes to a perfectly matched probe. For the purposes of the present invention, generally, "highly stringent" hybridization and wash conditions are selected to be about 5°C lower than the T_m for the specific sequence at a defined ionic strength and pH.

"Ultra high-stringency" hybridization and wash conditions are those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x as high as that observed for hybridization to any of the unmatched target nucleic acids. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio of at least 1/2 that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-high stringency conditions.

Similarly, even higher levels of stringency can be determined by gradually increasing the hybridization and/or wash conditions of the relevant hybridization assay. For example, those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x, 20x, 50x, 100x, or 500x or more as high as that observed for hybridization to any of the unmatched target nucleic acids. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio of at least 1/2 that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-ultra-high stringency conditions.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

5 Sequence Comparison, Identity, and Homology

The terms "identical" or "percent identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as
10 measured using one of the sequence comparison algorithms described below (or other algorithms available to persons of skill) or by visual inspection.

The phrase "substantially identical," in the context of two nucleic acids or polypeptides (e.g., DNAs encoding a polymerase, or the amino acid sequence of a polymerase) refers to two or more sequences or subsequences that have at least about
15 60%, about 80%, about 90-95%, about 98%, about 99% or more nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm or by visual inspection. Such "substantially identical" sequences are typically considered to be "homologous," without reference to actual ancestry. Preferably, the "substantial identity" exists over a
20 region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably, the sequences are substantially identical over at least about 150 residues, or over the full length of the two sequences to be compared.

Proteins and/or protein sequences are "homologous" when they are
25 derived, naturally or artificially, from a common ancestral protein or protein sequence. Similarly, nucleic acids and/or nucleic acid sequences are homologous when they are derived, naturally or artificially, from a common ancestral nucleic acid or nucleic acid sequence. Homology is generally inferred from sequence similarity between two or more nucleic acids or proteins (or sequences thereof). The precise percentage of

similarity between sequences that is useful in establishing homology varies with the nucleic acid and protein at issue, but as little as 25% sequence similarity over 50, 100, 150 or more residues is routinely used to establish homology. Higher levels of sequence similarity, e.g., 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% or more identity, 5 can also be used to establish homology. Methods for determining sequence similarity percentages (e.g., BLASTP and BLASTN using default parameters) are described herein and are generally available.

For sequence comparison and homology determination, typically one sequence acts as a reference sequence to which test sequences are compared. When 10 using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

15 Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms 20 (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see generally *Current Protocols in Molecular Biology*, Ausubel et al., eds., *Current Protocols*, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., supplemented through 2011).

25 One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs 30 (HSPs) by identifying short words of length W in the query sequence, which either

match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then
5 extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in
10 each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W , T , and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide
15 sequences) uses as defaults a word length (W) of 11, an expectation (E) of 10, a cutoff of 100, $M=5$, $N=-4$, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word length (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

20 In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993) Proc. Nat'l. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two
25 nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

30 For reference, the amino acid sequence of a wild-type DPO4 polymerase is presented in Table 2.

Exemplary Mutation Combinations

A list of exemplary polymerase mutation combinations and the amino acid sequences of recombinant DPO4 polymerases harboring the exemplary mutation combinations are provided in Table 2. Positions of amino acid substitutions are identified relative to a wildtype DPO4 DNA polymerase (SEQ ID NO:1). Polymerases of the invention (including those provided in Table 2) can include any exogenous or heterologous feature (or combination of such features) at the N- and/or C-terminal region. For example, it will be understood that polymerase mutants in Table 2 that do not include, e.g., a C-terminal polyhistidine tag can be modified to include a polyhistidine tag at the C-terminal region, alone or in combination with any of the exogenous or heterologous features described herein. The variants set forth herein include a deletion of the last 12 amino acids of the protein (i.e., amino acids 341-352) so as to, e.g., increase protein solubility in bacterial expression systems.

Table 2
DPO4 Variants Identified through Semi-Rational Design

SEQ ID NO	Amino Acid Sequence
1 wt DPO4 DNA polymerase	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGAVA TANYEARKFGVKAGIPIVEAKKILPNAVYLPMPRKEVYQQVSSRI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAKIAADMAKPNGIKVIDDEEVKRLIR ELDIADVPGIGNITAEKLLKLGINKLVDTLSEFDKLGKMIGE KAKYLISLARDEYNEPIRTRVRKSIGRIVTMKRNSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVAVTEDLDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRI GVRFSKFIEAIGLDKFFDT
2 CO1065 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152E_I153M_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAMAGRMKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGE KAKYLISLARDEYNEPIRTRVRKSIGRTV TMKRNSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVAVAWKSYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRI GVRFSKF
3 CO1066 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGE KAKYLISLARDEYNEPIRTRVRKSIGRTV TMKRNSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVAVAWKSYWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRI GVRFSKF

SEQ ID NO	Amino Acid Sequence
<p>4 CO1067 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF</p>
<p>5 C1222 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_R87Q_K152G_A155M _D156W_P184L_I189W_ K196S_I248T_R253N_R256T _K278N_R279Q_V289W_ T290R_E291S_D292Y_ L293W_I295Y_S297H_ G299L_T301RΔ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSEQI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEKLLKSLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNNSNLEEI KPY LFRAIEESYYKLDNQIPKAIHVVAWRS YWDYVHRLRRFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF</p>
<p>6 C1224 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_R116N_K152G_ A155M_D156W_R173N_ R176N_P184L_I189W_I248T _V289W_T290R_E291S_ D292Y_L293W_I295Y_ S297H_G299L_T301R_ K317Q_Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVNDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKNLIN ELDIADV LGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWRS YWDYVHRLRRFPHGISK ETAYSESVQLLQKILEEDERKIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
<p>7 C1226 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152G_A155M_ D156W_P184L_I189W_ I248T_V289W_T290R_ E291S_D292Y_L293W_ I295Y_S297H_G299L_T301R_ _E324K_E327KΔ341-352</p>	<p>MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEK LK KLGINKLVD T L S I E F D K L K G M I G E A KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWDYVHRLRRFPHGISK ETAYSESVKLLQKILKEDKRKIRRIGVRF SKF</p>
<p>8 C1337 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_R116N_K152G_ A155M_D156W_R173N_ R176N_P184L_I189W_I248T_ _V289W_T290R_E291S_ D292Y_L293W_I295Y_ S297H_G299L_T301R_ E324K_E327K_Δ341-352</p>	<p>MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVNDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKNLIN ELDIADV LGIGNWTAEK LK KLGINKLVD T L S I E F D K L K G M I G E A KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWDYVHRLRRFPHGISK ETAYSESVKLLQKILKEDKRKIRRIGVRF SKF</p>
<p>9 C1363 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152G_A155M_ D156W_K164T_P184L_ I189W_I248T_V289W_ T290R_E291S_D292Y_ L293W_I295Y_S297H_ G299L_T301RΔ341-352</p>	<p>MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGITVIDDEEVKRLIR ELDIADV LGIGNWTAEK LK KLGINKLVD T L S I E F D K L K G M I G E A KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWDYVHRLRRFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
10 C1368 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152G_A155M_ D156W_P184L_I189W_ E192R_K196L_R242W_ I248T_K282G_V289W_ T290R_E291S_D292Y_ L293W_I295Y_S297H_ G299L_T301RΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTARKLKLKLLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVWKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPGAIHVVAWRSYWDYVHRLRRFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
11 C1417 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_R116N_K152G_ A155M_D156W_R173N_ R176N_P184L_I189W_ I248T_V289W_T290R_ E291S_D292Y_L293W_ I295Y_S297H_G299L_T301R _K317Q_K321Q_E324K_ E325K_E327K_Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVNDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKNLIN ELDIADVLGIGNWTAEKLKKLLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWDYVHRLRRFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
12 C1418 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_R116N_K152G_ A155M_D156W_R173N_ R176N_P184L_I189W_I248T _V289W_T290R_E291S_ D292Y_L293W_I295Y_ S297H_G299L_T301R_ K317Q_K321Q_E324N_ E325Q_E327Q_Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVNDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKNLIN ELDIADV LGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGE A KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWDYVHRLRRFPHGISK ETAYSESVQLLQQILNQDQRKIRRIGVRF SKF
13 C1428 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_I189W_I248T V289W_T290K_E291S_ D292Y_L293W_E324K_ E327K Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGE A KAKYLISLARDEYNEPIRTRVRKSIGRIVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILKEDKRKIRRIGVRF SKF
14 C1429 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_ S86E_K152G_A155M_ D156W_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_E324K_ E327KΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGE A KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILKEDKRKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
<p>15 C1452 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152G_A155M_ D156W_P184L_I189W_ E192R_K196L_R242W_ I248T_K282G_V289W_ T290R_E291S_D292Y_ L293W_I295Y_S297H_ G299L_T301R_E324R_ E327RA341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTARKLKL LGINKLVD TLSIEFDKLG MIGE A KAKYLISLARDEYNEPIRTRVWKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPGAIHVVAWRSYWDYVHRLRRFPHGISK ETAYSESVKLLQKILREDRRKIRRIGVRF SKF</p>
<p>16 C1454 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_I189W_I248T V289W_T290K_E291S_ D292Y_L293W_K317Q_ K321Q_E324K_E325K_ E327K Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEK LK KL LGINKLVD TLSIEFDKLG MIGE A KAKYLISLARDEYNEPIRTRVRK SIGRIVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRF SKF</p>
<p>17 C1456 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_G187P_N188Y_ I189W_T190Y_I248T_ V289W_T290K_E291S_ D292Y_L293W_K317Q_ K321Q_E324K_E325K_ E327KΔ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEK LK KL LGINKLVD TLSIEFDKLG MIGE A KAKYLISLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
18 C1457 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_G187P_N188Y_ I189W_T190Y_I248T_ V289W_T290K_E291S_ D292Y_L293W_K317Q_ K321Q_E324N_E325Q_ E327QΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVQLLQQILNQDQRKIRRIGVRF SKF
19 C1506 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_I189W_I248T V289W_T290K_E291S_ D292Y_L293W_E325K_ D326K Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRIVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEKKERKIRRIGVRF SKF
20 C1395 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152Q_I153R_A155G_ D156A_P184L_G187P_ N188Y_I189W_T190Y_ I248T_V289W_T290K_ E291S_D292Y_L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVVFQAQRAGAMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
21 C1187 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_ V289W_T290R_E291S_ D292Y_L293W_D294Q_ I295M_V296A_S297T_ R298L_G299F_T301H Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELADIADV LGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
22 CO672 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_V289 W_T290R_E291S_D292Y_ L293W_I295A_S297P_ G299W_T301FΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELADIADV LGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWDVPRWRFFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
23 CO677 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_V289 W_T290R_E291S_D292Y_ L293_I295Y_S297W_ G299W_T301VΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELADIADV LGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWDYVWRWRVFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
24 CO982 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_V289 W_T290R_E291S_D292Y_ L293_I295C_S297W_ G299I_T301WΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEK LKKLGINKLVD TLSIEFDK LKGMIGEA KAKYLISLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRI PKAIHVVAWRSYWD CVWRIRW FPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
25 C1152 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_ K282G_A283L_V289W_ T290K_E291S_D292Y_ L293W_F302E_P303R_ H304GΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEK LKKLGINKLVD TLSIEFDK LKGMIGEA KAKYLISLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRI PGLIHVVAWKS YWDIVSRGRTERGGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
26 C1159 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152G_I153H_ A155G_D156R_P184L_ G187P_N188Y_I189W_ T190Y_I248T_V289W_ T290K_E291S_D292Y_L293 WΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGHAGRMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEK LKKLGINKLVD TLSIEFDK LKGMIGEA KAKYLISLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRI PKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
27 C1185 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_ V289W_T290R_E291S_ D292Y_L293W_D294A_ I295P_V296R_S297L_R298K _G299A_R300C_T301E Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEKLLKLGINKLVD TLSIEFDKLG MIGE A KAKYLISLARDEYNEPIRTRVRK SIGRTVTM KRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWAPRLKACEFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
28 C1190 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_ V289W_T290R_E291S_ D292Y_L293W_D294C_ V296S_S297Q_R298T_ T301LΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEKLLKLGINKLVD TLSIEFDKLG MIGE A KAKYLISLARDEYNEPIRTRVRK SIGRTVTM KRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWCISQTGR LFP HGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
29 C1191 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_V289 W_T290R_E291S_D292Y_ L293W_D294S_I295T_ V296R_R298T_G299H_ R300S_T301YΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEKLLKLGINKLVD TLSIEFDKLG MIGE A KAKYLISLARDEYNEPIRTRVRK SIGRTVTM KRNSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWSTRSTHSYFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
30 C1234 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_V289 W_T290R_E291S_D292Y_ L293W_D294N_I295S_ V296Q_S297Y_G299W_ R300S_T301W_Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWNSQYRWSWFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
31 C1237 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_ V289W_T290R_E291S_ D292Y_L293W_D294L_ I295T_V296R_S297C_T301R Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWLTRCRGRRFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
32 C1238 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_ V289W_T290R_E291S_ D292Y_L293W_D294A_ I295W_V296R_S297R_ R298E_G299S_R300S_ T301MAΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYAWRRESSMFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF

SEQ ID NO	Amino Acid Sequence
33 C1240 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152A_I153R_ A155G_D156L_P184L_ G187P_N188Y_I189W_ T190Y_I248T_V289W_ T290K_E291S_D292Y_ L293WΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVF AARAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
34 C1278 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_ V289W_T290R_E291S_ D292Y_L293W_D294P_ I295P_V296R_G299A_ R300C_T301KΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADVLGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWP PPSRACKFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
35 C1287 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152Y_I153S_A155C_ D156G_P184L_G187P_ N188Y_I189W_T190Y_ I248T_V289W_T290K_ E291S_D292Y_L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAYSACGMKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
36 C1288 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152V_I153Y_A155C _D156G_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293WΔ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAVYACGMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
37 C1289 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152M_I153Q_A155G _D156Y_P184L_G187P_ N188Y_I189W_T190Y_ I248T_V289W_T290K_ E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAVYACGMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
38 C1292 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152L_I153K_A155G_ D156S_P184L_G187P_ N188Y_I189W_T190Y_ I248T_V289W_T290K_ E291S_D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAVYACGMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
39 C1328 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_V289 W_T290R_E291S_D292Y_ L293W_I295Y_S297F_ G299Y_T301SΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEK LKKLGINKLVD TLSIEFDK LKGMIGEA KAKYLISLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEI KPY LFRAIEESYKLDKRI PKAIHVVAWRSYWDYVFRYRSFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
40 C1384 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152G_A155M_ D156W_P184L_I189W_ I248T_V289W_T290R_ E291S_D292Y_L293W_ D294H_I295W_V296Y_ S297QΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEK LKKLGINKLVD TLSIEFDK LKGMIGEA KAKYLISLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEI KPY LFRAIEESYKLDKRI PKAIHVVAWRSYWHWYQGRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
41 C1392 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152E_I153V_A155G_D156 R_P184L_I189W_I248T_ V289W_T290R_E291S_ D292Y_L293W_I295Y_ S297H_G299L_T301R Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAEVAGRMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEK LKKLGINKLVD TLSIEFDK LKGMIGEA KAKYLISLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEI KPY LFRAIEESYKLDKRI PKAIHVVAWRSYWDYVHRLRRFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
42 C1404 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_ V289W_T290R_E291S_ D292Y_L293W_D294S_ I295F_V296A_S297T_ R298G_T301QΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWSFATGGRQFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
43 C1428 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_E324K_ E327K Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILKEDKRKIRRIGVRF SKF
44 C1442 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_A155M_D156W_ P184L_I189W_I248T_ V289W_T290R_E291S_ D292Y_L293W_D294P_ I295P_S297Y_R298L_R300V _T301RΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGIAMWMAKPNGIKVIDDEEVKRLIR ELDIADV LGIGNWTAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWPVYLGVRFP HGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
45 C1462 A42V_K56Y_M76W_R77S_ K78N_E79L_Q82W_Q83G_ S86E_E97S_K152G_I153T_ A155G_D156L_P184L_ G187P_N188Y_I189W_ T190Y_I248T_V289W_ T290K_E291S_D292Y_ L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWSNLVYWGVSERI MNLLREYSSKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVD TLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
46 C1464 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_R298E_ K321N_R328N Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVD TLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSEGRTPHGHGISK ETAYSESVKLLQNILEEDENKIRRIGVRF SKF
47 C1516 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_P184L_G187E_ N188S_I189W_T190V_I248T _V289W_T290K_E291S_ D292Y_L293W_R300I Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADV LGIESWVAEKLLKLGINKLVD TLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGITFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
48 C1517 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152A_I153K_D156I_ P184L_G187P_N188Y_ I189W_T190Y_I248T_ V289W_T290K_E291S_ D292Y_L293WΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFVAAKAAIMAKPNGIKVIDDEEVKRLIR ELADIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
49 C1520 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153Q_A155G_ D156M_P184L_G187E_ N188S_I189W_T190V_I248T_ _V289W_T290K_E291S_ D292Y_L293WΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALQAGMMAKPNGIKVIDDEEVKRLIR ELADIADVLGIESWVAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
50 C1554 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ I248T_V289W_D211P_ K214V_E219W_K223C_ T290K_E291S_D292Y_ L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFACTAGLMAKPNGIKVIDDEEVKRLIR ELADIADVLGIPYWYAEKLLKLGINKLVDTLSEFPKLVGMIGWA KACYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
51 C1555 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ D211S_K214S_E219G_I248T _V289W_T290K_E291S_ D292Y_L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEIEFSKLSGMIGGA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
52 C1556 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ D211P_K214V_E219A_ K223T_I248T_V289W_ T290K_E291S_D292Y_ L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEIEFPKLVGMIGAA KATYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF
53 C1557 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ D211P_K214A_E219A_ K223R_I248T_V289W_ T290K_E291S_D292Y_ L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEIEFPKLAGMIGAA KARYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF

SEQ ID NO	Amino Acid Sequence
54 C1558 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189Y_T190F_I248T _V289W_T290K_E291S_ D292Y_L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYYFAEKLKKGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
55 C1560 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_Q320K_ K321R_I322A_E324A_E325 R Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLKKGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLKRALARDERKIRRIGVRF SKF
56 C1561 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_Q320R_ K321S_I322L_E324G_E325G Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLKKGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLRSLGGDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
57 C1562 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T_ _V289W_T290K_E291S_ D292Y_L293W_ K321R_I322C_E324S_E325G Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQRCLSGDERKIRRIGVRF SKF
58 C1569 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ K214S_G215R_E219A_ K221R_K223R_I248T_ V289W_T290K_E291S_ D292Y_L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLSRMIGAA RARYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
59 C1570 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ K214V_G215S_E219R_ K221R_I248T_ V289W_T290K_E291S_ D292Y_L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLVSMIGRA RAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
60 C1575 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_E170M_R173G_ L174F_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGTAGLMAKPNGIKVIDDEMVKGFIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLVSMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
61 C1577 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153Y_A155G_ D156T_M157R_P184L_ G187P_N188Y_I189W_ T190Y_I248T_V289W_T290 K_E291S_D292Y_L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKKNKVFAGYAGTRAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLVSMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
62 C1581 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ K212D_K214R_G215K_ K221G_K223V_I248T_ V289W_T290K_E291S_ D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELADIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDLDRKMIGEA GAVYLI SLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
63 C1584 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ K212F_K214S_G215T_ K221R_K223R_I248T_ V289W_T290K_E291S_ D292Y_L293W Δ341-352	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELADIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDLSTMIGEA RARYLI SLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
<p>64 C1589 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ K214S_G215E_E219G_ K221G_K223R_I248T_ V289W_T290K_E291S_ D292Y_L293W Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLESEMIGGA GARYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF</p>
<p>65 C1592 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_Q320M_K32 1G_I322V_E324S_E325V Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLESEMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLMGVLSVDERKIRRIGVRF SKF</p>
<p>66 C1593 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_K321R_ E324M_E325SΔ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLESEMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQRILMSDERKIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
67 C1594 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_Q320K_ K321Q_I322V_E324W_ E325G Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLKQVLWGDERKIRRIGVRF SKF
68 C1599 K24S_K26L_A42V_K56Y_ M76W_K78N_E79L_Q82W_ Q83G_S86E_K152G_I153T_ A155G_D156L_P184L_ G187P_N188Y_I189W_ T190Y_I248T_V289W_ T290K_E291S_D292Y_ L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLSGLPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLKQILEEDERKIRRIGVRF SKF
69 C1614 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _R253N_R256N_E259S_ E260Q_K262N_V289W_ T290K_E291S_D292Y_ L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNNSNNLSQINPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLKQILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
70 C1615 D7V_A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_R267Q_E270Q_D277G_K278N_R279Q_V289W_T290K_E291S_D292Y_L293W Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELADIADV LGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFQAIQESYYKLG NQIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF
71 C1616 A42V_K56Y_M76W_K78N_E79L_Q82W_Q83G_S86E_K152G_I153T_A155G_D156L_P184L_G187P_N188Y_I189W_T190Y_I248T_K282S_V289W_T290K_E291S_D292Y_L293W_R300S_K308Q_E309N_ Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELADIADV LGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFR AIEESYYKLDKRIPSAIHVVAWKS YWDIVSRGSTFPHGISQ NTAYSESVKLLQKILEEDERKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
<p>72 C1617 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ _I248T_V289W_ T290K_E291S_D292Y_ L293W_E314Q_K317Q_ E324K_E327K_R328S Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSQSVQLLQKILKEDKSKIRRIGVRF SKF</p>
<p>73 C1618 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _D277G_K278N_R279Q_ V289W_T290K_E291S_ D292Y_L293W_E309D_ E324K_E327K Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLG NQIPKAIHVVAWKS YWDIVSRGRTFPHGISK DTAYSESVKLLQKILKEDKRKIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
<p>74 C1621 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ K214R_G215R_E219G_ K221R_K223T_I248T_ V289W_T290K_E291S_ D292Y_L293W Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLRMIGGA RATYLI SLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF</p>
<p>75 C1622 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ K214A_G215E_E219V_ K221G_K223S_I248T_ V289W_T290K_E291S_ D292Y_L293W Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLAEMIGVA GASYLI SLARDEYNEPIRTRVRK SIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
<p>76 C1623 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ K214Q_G215A_E219L_ K221G_K223R_I248T_ V289W_T290K_E291S_ D292Y_L293W Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLQAMIGLA GARYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF</p>
<p>77 C1626 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_Q320I_ K321Q_I322T_E324D_ E325V Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLKGMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLIQTL DVDERKIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
<p>78 C1627 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_Q320T_ K321R_I322T_E324A_ E325G Δ341-352</p>	<p>MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLTRTLAGDERKIRRIGVRF SKF</p>
<p>79 C1628 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_Q320N_ K321R_E324R_ E325R Δ341-352</p>	<p>MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLNRILRRDERKIRRIGVRF SKF</p>
<p>80 C1630 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_Q320N_ K321S_I322L_E324R_ E325S Δ341-352</p>	<p>MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLNSLLRS DERKIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
<p>81 C1666 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290R_E291S_ D292Y_L293W_D294S_ I295P_S297T_R300G_T301Q Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWRSYWSPVTRGGQFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF</p>
<p>82 C1788 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ D211T_K214A_E219V_ I248T_V289W_T290K_ E291S_D292Y_L293W_ Q320T_K321R_I322C_ E324G_E325G Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFTKLAGMIGVA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLTRCLGVDERKIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
<p>83 C1789 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_P184L_G187P_ N188Y_I189W_T190Y_ D211T_K214A_E219V_ I248T_V289W_T290K_ E291S_D292Y_L293W_ Q320S_K321M_I322C_ E324S_E325A Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSEFTKLAGMIGVA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVKLLSMCL SADERKIRRIGVRF SKF</p>
<p>84 C1800 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152G_I153T_A155G_ D156L_M157R_E170M_ R173G_L174F_P184L_G187P _N188Y_I189W_T190Y_ D211P_K214A_E219A_ K223R_I248T_D277G_ K278N_R279Q_V289W_T29 0K_ E291S_D292Y_L293W_ D294Q_I295M_V296A_ S297T_R298L_G299F_ T301H_E309D_K317Q_ K321Q_E324K_E325K_ E327K Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAGTAGLRAKPNGIKVIDDEMVKGFIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSEFPKLAGMIGAA KARYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLG NQIPKAIHVVAWKS YWQMATLFRHFPHGISK DTAYSESVQLLQQILKKDKR KIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
<p>85 C1801 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_I295A_ S297P_G299W_T301F Δ341-352</p>	<p>MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDAVPRWRFFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF</p>
<p>86 C1802 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_I295A_ S297P_G299W_T301F _K317Q_K321Q_E324K_ E325K_E327KΔ341-352</p>	<p>MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDAVPRWRFFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRFSKF</p>

SEQ ID NO	Amino Acid Sequence
<p>87 C1803 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_I295C_ S297W_G299I_T301W Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELADIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYWDCVWRIRWFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEF</p>
<p>88 C1804 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_I295C_ S297W_G299I_T301W _K317Q_K321Q_E324K_ E325K_E327KΔ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELADIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYWDCVWRIRWFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRFSEF</p>

SEQ ID NO	Amino Acid Sequence
89 C1805 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_I295Y_ S297W_G299W_T301V Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDYVWRVRFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEK
90 C1807 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_I295Y_ S297H_G299L_T301R Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDYVHRLRRFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEK

SEQ ID NO	Amino Acid Sequence
91 C1808 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L293W_I295Y_ S297H_G299L_T301R_K317 Q_K321Q_E324K_E325K_ E327KΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVLYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDYVHRLRRFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRF SKF
92 C1809 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_ E192R_K196L_R242W_I248 T_K282G_V289W_T290K_ E291S_D292Y_L293W_ I295Y_S297H_G299L_T301R _E324K_E327K_Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVLYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYARKLLKLLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVWKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPGAIHVVAWKS YWDYVHRLRRFPHGISK ETAYSESVKLLQKILKEDDKRKIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
<p>93 C1810 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_ E192R_K196L_R242W_I248 T_K282G_V289W_T290K_ E291S_D292Y_L293W_ I295Y_S297H_G299L_T301R _E324K_E327K_Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYARKLKLKLLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVWKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPGAIHVVAWKSYWQMATLFRHFPHGISK ETAYSESVQQLLQQILKKDKRKRIRRIGVRF SKF</p>
<p>94 C1812 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_ E192R_K196L_R242W_I248 T_K282G_V289W_T290K_ E291S_D292Y_L293W_ D294Q_I295M_V296A_ S297T_R298L_G299F_ T301H_K317Q_K321Q_ E324K_E325K_E327K Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYARKLKLKLLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVWKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPGAIHVVAWKSYWQMATLFRHFPHGISK ETAYSESVKLLQKILKEDKRKRIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
<p>95 C1816 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_I248T_ _V289W_T290K_E291S_ D292Y_L293W_D294N_I295 S_V296Q_S297Y_G299W_ R300S_T301W_K317Q_ K321Q_E324K_E325K_ E327K Δ341-352</p>	<p>MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWNSQYRWSWFPHGISK ETAYSESVQLLQQILKKDKRKIRRIGVRF SKF</p>
<p>96 C1817 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152A_I153K_D156I_V183L _P184L_G187P_N188Y_ I189W_T190Y_I248T_ V289W_T290K_E291S_ D292Y_L293W_D294Q_ I295M_V296A_S297T_ R298L_G299F_T301H Δ341-352</p>	<p>MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAAKAAIMAKPNGIKVIDDEEVKRLIR ELDIADLLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
<p>97 C1818 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152Q_I153R_A155G_ D156A_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L 293W_D294Q_ I295M_V296A_S297T_ R298L_G299F_T301H Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAQRAGAMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF</p>
<p>98 C1819 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152V_I153Y_A155C_ D156G_P184L_G187P_ N188Y_I189W_T190Y_I248T _V289W_T290K_E291S_ D292Y_L 293W_D294Q_ I295M_V296A_S297T_ R298L_G299F_T301H Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAVYACGMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRF SKF</p>

SEQ ID NO	Amino Acid Sequence
<p>99 C1820 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_E170M_R173G_ L174F_P184L_G187P_ N188Y_I189W_T190Y_I248T_ _V289W_T290K_E291S_ D292Y_L293W_D294Q_ I295M_V296A_S297T_ R298L_G299F_T301H Δ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNULLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEMVKGFIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF</p>
<p>100 C1821 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_P184L_G187P_ N188Y_I189W_T190Y_ D211P_K214V_E219A_K223 T_I248T_V289W_T290K_ E291S_D292Y_L293W_ D294Q_I295M_V296A_ S297T_R298L_G299F_ T301HΔ341-352</p>	<p>MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNULLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELDIADVLGIPYWYAEKLLKLGINKLVDTLSIEFPKLVGMIGAA KATYLISLARDEYNEPIRTRVRKSIGRTVTMKRNSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSKF</p>

SEQ ID NO	Amino Acid Sequence
101 C1823 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152A_I153K_D156I_ P184L_G187P_N188Y_ I189W_T190Y_I248T_ V289W_T290K_E291S_ D292Y_L293W_K317Q_ K321Q_E324K_E325K_ E327KΔ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAAKAAIMAKPNGIKVIDDEEVKRLIR ELADIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVQLLQQILKKDKRKRIRRIGVRF SKF
102 C1824 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152Q_I153R_A155G_ D156A_P184L_ G187P_N188Y_I189W_ T190Y_I248T_V289W_T290 K_E291S_ D292Y_L293W_K317Q_ K321Q_E324K_E325K_ E327K Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNKVFAQ RAGAMAKPNGIKVIDDEEVKRLIR ELADIADV LGIPYWYAEKLLKLGINKLVDTLSIEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEI KPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDIVSRGRTFPHGISK ETAYSESVQLLQQILKKDKRKRIRRIGVRF SKF

SEQ ID NO	Amino Acid Sequence
103 C1826 A42V_K56Y_A57P_M76W_ K78N_E79L_Q82W_Q83G_ S86E_K152Q_I153R_A155G_ _D156A_P184L_ G187P_N188Y_I189W_ T190Y_I248T_ V289W_T290K_E291S_ D292Y_L293W_I295Y_ S297H_G299L_T301R_ K317Q_K321Q_E324K_ E325K_E327K Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYPGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNVFAQRAGAMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWDYVHRLRRFPHGISK ETAYSESVQLLQQILKKDKRKIRRI GVRFSKF
104 C1827 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152Q_I153R_A155G_ D156A_P184L_ G187P_N188Y_I189W_ T190Y_I248T_ V289W_T290K_E291S_ D292Y_L293W_D294Q_ I295M_V296A_S297T_ R298L_G299F_T301H_ K317Q_K321Q_E324K_ E325K_E327K Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGISKNVFAQRAGAMAKPNGIKVIDDEEVKRLIR ELDIADV LGIPYWYAEKLLKLGINKLVDTLSEFDKLGKMIGEA KAKYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEKPY LFRAIEESYYKLDKRIPKAIHVVAWKS YWQMATLFRHFPHGISK ETAYSESVQLLQQILKKDKRKIRRI GVRFSKF

SEQ ID NO	Amino Acid Sequence
105 C1828 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_P184L_ G187P_N188Y_I189W_ T190Y_I248T_ V289W_D211P_K214A_ E219A_K223R_T290K_ E291S_D292Y_L293W_ K317Q_K321Q_E324K_ E325K_E327K Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGIGSKNKVFALTAGRMAKPNGIKVIDDEEVKRLIR ELADIADVLGIPYWYAEKLLKLGINKLVDTLSEIEFPKLAGMIGAA KARYLISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNLEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYWWDIVSRGRTFPHGISK ETAYSESVQLLQQILKKDKRKRIRRIGVRFSEK
106 C1895 A42V_K56Y_M76W_K78N_ E79L_Q82W_Q83G_S86E_ K152L_I153T_A155G_ D156R_E170M_R173G_ L174F_P184L_G187P_ N188Y_I189W_T190Y_I248T _L258P_V289W_T290K_ E291S_D292Y_L293W_ D294Q_I295M_V296A_ S297T_R298L_G299F_ T301H Δ341-352	MIVLFDVDFDYFYAQVEEVLNPSLKGKPVVVCVFSGRFEDSGVVA TANYEARKFGVYAGIPIVEAKKILPNAVYLPWRNLVYWGVSERI MNLLREYSEKIEIASIDEAYLDISDKVRDYREAYNLGLEIKNKI LEKEKITVTVGIGSKNKVFALTAGRMAKPNGIKVIDDEMVKGFIR ELADIADVLGIPYWYAEKLLKLGINKLVDTLSEIEFDKLGKMGIEA KAKYLIISLARDEYNEPIRTRVRKSIGRTVTMKNRSRNPEEIKPY LFRAIEESYYKLDKRIPKAIHVVAWKSYWQMATLFRHFPHGISK ETAYSESVKLLQKILEEDERKIRRIGVRFSEK

The Examples and polymerase variants provided below further illustrate and exemplify the compositions of the present invention and methods of preparing and using such compositions. It is to be understood that the scope of the present invention is not limited in any way by the scope of the following Examples.

5

EXAMPLES

EXAMPLE 1

IDENTIFICATION OF DPO4 AS A CANDIDATE TRANSLESION DNA POLYMERASE FOR INCORPORATION OF BULKY NUCLEOTIDE ANALOGS DURING TEMPLATE-MEDIATED DNA SYNTHESIS

10 To identify a DNA polymerase with the ability to synthesize daughter strands using “bulky” substrates (i.e., able to bind and incorporate heavily substituted nucleotide analogs into a growing nucleic acid strand), a screen was conducted of several commercially available polymerases. Candidate polymerases were assessed for the ability to extend an oligonucleotide-bound primer using a pool of dNTP analogs
15 substituted with alkyne linkers on both the backbone α -phosphate and the nucleobase moieties (model bulky substrates, referred to herein as, “dNTP-2c”). Polymerases screened for activity included the following: Vent_R (Exo-), Deep Vent_R® (Exo-), Therminator, Therminator II, Therminator III, Therminator Y, 9^oN_m, PWO, PWO SuperYield, PyroPhage 3173 (Exo-), Bst, Large Fragment, Exo- Pfu, Platinum
20 Genotype TSP, Hemo Klen Taq, Taq, MasterAMP Taq, Phi29, Bsu, Large Fragment, Exo-Minus Klenow (D355A, E357A), Sequenase Version 2.0, Transcriptor, Maxima, Thermoscript, M-MuLV (RNase H-), AMV, M-MuLV, Monsterscript, and DPO4. Of the polymerases tested, DPO4 (naturally expressed by the archaea, *Sulfolobus solfataricus*) was most able to effectively extend a template-bound primer with dNTP-
25 2c nucleotide analogs. Without being bound by theory, it was speculated that DPO4, and possibly other members of the translesion DNA polymerase family (i.e., class Y DNA polymerases), may be able to effectively utilize bulky nucleotide analogs owing

to their relatively large substrate binding sites, which have evolved to accommodate naturally occurring, bulky DNA lesions.

EXAMPLE 2

IDENTIFICATION OF “HOT SPOTS” FOR DIRECTED MUTAGENESIS IN THE DPO4 PROTEIN 5 AND SCREEN OF DPO4 MUTANT LIBRARIES TO IDENTIFY OPTIMIZED SEQUENCE MOTIFS

As a first step in generating DPO4 variants with improved polymerase activity when challenged with bulky substrates, the “HotSpot Wizard” web tool was used to identify amino acids in the DPO4 protein to target for mutagenesis. This tool implements a protein engineering protocol that targets evolutionarily variable amino
10 acid positions located in, e.g., the enzyme active site. “Hot spots” for mutation are selected through the integration of structural, functional, and evolutionary information (see, e.g., Pavelka et al., “HotSpot Wizard: a Web Server for Identification of Hot Spots in Protein Engineering” (2009) *Nuc Acids Res* 37 doi:10.1093/nar/gkp410). Applying this tool to the DPO4 protein, it was observed that hot spot residues identified tended to
15 cluster into certain zones, or regions, spread throughout the full amino acid sequence. Arbitrary boundaries were set to distinguish 13 such regions, designated “Mut1” – “Mut13”, in which mutagenesis hot spots are concentrated. These 13 “Mut” regions are illustrated in FIG. 1 with hot spot residues identified by underscoring.

To screen for DPO4 variants with improved polymerase activity based
20 on hot spot mapping, a saturation mutagenesis library was created for each of the 13 Mut regions, in which hot spot amino acids were changed, while conserved amino acids were left unaltered. Screening was conducted using a 96 well plate platform, and polymerase activity was assessed with a primer extension assay using “dNTP-OAc” nucleotide analogs as substrates. These model bulky substrates are substituted with
25 triazole acetate moieties conjugated to alkyne substituents on both the α -phosphate and the nucleobase moieties. Screening results identified two Mut regions in particular that consistently produced DPO4 mutants with enhanced activity. These regions, “Mut_4” and “Mut_11”, correspond to amino acids 76-86 and amino acids 289-304, respectively, of the DPO4 protein. Further analysis of high-performing Mut_4 and Mut_11 variants

lead to the identification of an optimized variant motif sequence for each region. The optimized Mut_4 motif identified herein is as follows: M76W, K78N, E79L, Q82W, Q83G, and S86E, while that for the Mut_11 region is as follows: V289W, T290K, E291S, D292Y, and L293W.

5

EXAMPLE 3

SEMI-RATIONAL APPROACHES TO DESIGNING DPO4 VARIANTS WITH ENHANCED
POLYMERIZATION ACTIVITY

To continue evolving DPO4 variants with improved utilization of bulky substrates, a “semi-rational” design approach was taken following a number of different strategies. In one strategy, a library was created that directed random mutagenesis on 10 positions 152, 153, 155, and 156 (corresponding to the Mut_6 region) on the previously identified high performing variant, CO534 (disclosed in applicants’ co-pending PCT patent application no. PCT/US2016/061661, herein disclosed in its entirety by reference). Screening of this library was conducted as described above and results 15 identified three variants in particular that demonstrated superior primer extension activity using bulky nucleotide analog substrates. These variants are set forth in Table 2 as C1065 (SEQ ID NO: 2), C1066 (SEQ ID NO: 3), and CO1067 (SEQ ID NO: 4).

In another strategy, charged amino acids exposed on the surface of the DPO4 protein were targeted for mutation to reduce, eliminate, or reverse the charge. It was 20 hypothesized that alterations in the overall surface charge profile of the protein might enhance the interaction between the polymerase and nucleotide analog substrates bearing large, positively charged substituents, such as the XNTP substrates disclosed herein. This approach identified several mutants demonstrating potentially enhanced activity in primer extension assays using bulky nucleotide analog substrates. These 25 variants are set forth as SEQ ID NO:s 5- 21 in Table 2.

In yet other strategies, high-performing variants, including CO416 and CO534 (disclosed in applicants’ co-pending PCT application no. PCT/US2016/061661) and C1066 (SEQ ID NO: 3), CO1067 (SEQ ID NO:4), C1454 (SEQ ID NO:16) and C1187 (SEQ ID NO:21) were subject to random or directed mutagenesis as disclosed herein to

create libraries for further screening. Screening of these libraries was conducted as described above and results identified several variants that performed at least as well as the parental variants in primer extension assays using bulky XNTP substrates. These variants are set forth in Table 2 as SEQ ID NO:s 22-106.

5

All of the U.S. patents, U.S. patent application publications, U.S. patent applications, foreign patents, foreign patent applications and non-patent publications referred to in this specification and/or listed in the Application Data Sheet, including but not limited to U.S. Patent No. 7,939,259, U.S. Provisional Patent Application No. 10 62/501,526, and PCT Publication No. WO 2016/081871 are incorporated herein by reference, in their entirety. Such documents may be incorporated by reference for the purpose of describing and disclosing, for example, materials and methodologies described in the publications, which might be used in connection with the presently described invention.

15

Claims

1. An isolated recombinant DNA polymerase, which recombinant DNA polymerase comprises an amino acid sequence that is at least 80% identical to amino acids 1-340 of SEQ ID NO:1, which recombinant polymerase comprises mutations at amino acid positions 42, 56, 76, 78, 79, 82, 83, 86, 184, 189, 248, 289, 290, 291, 292, and 293, and at least one position selected from the group consisting of 7, 24, 26, 57, 77, 87, 97, 116, 152, 153, 155, 156, 157, 164, 170, 173, 174, 176, 183, 187, 188, 190, 191, 192, 196, 211, 212, 214, 215, 219, 221, 223, 242, 253, 256, 258, 259, 260, 262, 267, 270, 277, 278, 279, 282, 283, 290, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 308, 309, 314, 317, 320, 321, 322, 324, 325, 326, 327 and 328, wherein identification of positions is relative to wildtype DPO4 polymerase (SEQ ID NO:1), and which recombinant DNA polymerase exhibits polymerase activity.

2. The polymerase of claim 1, wherein the mutation at position 42 is A42V, the mutation at position 56 is K56Y, the mutation at position 76 is M76W, the mutation at position 78 is K78N, the mutation at position 79 is E79L, the mutation at position 82 is Q82W, the mutation at position 83 is Q83G, the mutation at position 86 is S86E, the mutation at position 184 is P184L, the mutation at position 189 is I189W, the mutation at position 289 is V289W, the mutation at position 290 is T290K or T290R, the mutation at position 291 is E291S, the mutation at position 292 is D292Y, and the mutation at position 293 is L293W.

3. The polymerase of claim 1, wherein the mutation at position 7 is D7V.

4. The polymerase of claim 1, wherein the mutation at position 24 is K24S.

5. The polymerase of claim 1, wherein the mutation at position 26 is K26L.
6. The polymerase of claim 1, wherein the mutation at position 57 is A57P.
7. The polymerase of claim 1, wherein the mutation at position 77 is R77S.
8. The polymerase of claim 1, wherein the mutation at position 87 is R87Q.
9. The polymerase of claim 1, wherein the mutation at position 97 is E97S.
10. The polymerase of claim 1, wherein the mutation at position 116 is R116N.
11. The polymerase of claim 1, wherein the mutation at position 152 is selected from the group consisting of K152E, K152G, K152L, K152Q, K152A, K152Y, K152V, K152M, and K152L.
12. The polymerase of claim 1, wherein the mutation at position 153 is selected from the group consisting of I153M, I153T, I153R, I153H, I153S, I153Y, I153Q, I153K, and I153V.
13. The polymerase of claim 1, wherein the mutation at position 155 is selected from the group consisting of A155G, A155M and A155C.

14. The polymerase of claim 1, wherein the mutation at position 156 is selected from the group consisting of D156R, D156L, D156W, D156A, D156G, D156Y, D156S, D156I, D156M, and D156T..
15. The polymerase of claim 1, wherein the mutation at position 157 is M157R.
16. The polymerase of claim 1, wherein the mutation at position 164 is K164T.
17. The polymerase of claim 1, wherein the mutation at position 170 is E170M.
18. The polymerase of claim 1, wherein the mutation at position 173 is R173N or R173G.
19. The polymerase of claim 1, wherein the mutation at position 176 is R176N.
20. The polymerase of claim 1, wherein the mutation at position 183 is V183L.
21. The polymerase of claim 1, wherein the mutation at position 187 is G187P or G183E.
22. The polymerase of claim 1, wherein the mutation at position 188 is N188Y or N188S.
23. The polymerase of claim 1, wherein the mutation at position at 190 is selected from the group consisting of T190Y, T190V and T190F.

24. The polymerase of claim 1, wherein the mutation at position 191 is A191E.
25. The polymerase of claim 1, wherein the mutation at position 192 is E192R.
26. The polymerase of claim 1, wherein the mutation at position 196 is K196S or K196L.
27. The polymerase of claim 1, wherein the mutation at position 211 is selected from the group consisting of D211P, D211S, and D211T.
28. The polymerase of claim 1, wherein the mutation at position 212 is K212D or K212F.
29. The polymerase of claim 1, wherein the mutation at position 214 is selected from the group consisting of K214V, K214S, K214A, K214R and K214Q.
30. The polymerase of claim 1, wherein the mutation at position 215 is selected from the group consisting of G215R, G215S, G215K, G215T, G215E and G215A.
31. The polymerase of claim 1, wherein the mutation at position 219 is selected from the group consisting of E219W, E219G, E219A, E219R, E219V and E219L.
32. The polymerase of claim 1, wherein the mutation at position 221 is K221R or K221G.

33. The polymerase of claim 1, wherein the mutation at position 223 is selected from the group consisting of K223C, K223T, K223R, K223V and K223S.
34. The polymerase of claim 1, wherein the mutation at position 242 is R242W.
35. The polymerase of claim 1, wherein the mutation at position 248 is I248T.
36. The polymerase of claim 1, wherein the mutation at position 253 is R253N.
37. The polymerase of claim 1, wherein the mutation at position 256 is R256T or R256N.
38. The polymerase of claim 1, wherein the mutation at position at position 258 is L258P.
39. The polymerase of claim 1, wherein the mutation at position 259 is R259S.
40. The polymerase of claim 1, wherein the mutation at position 260 is E260Q.
41. The polymerase of claim 1, wherein the mutation at position 262 is K262N.
42. The polymerase of claim 1, wherein the mutation at position 267 is R267Q.

43. The polymerase of claim 1, wherein the mutation at position 270 is E270Q.
44. The polymerase of claim 1, wherein the mutation at position 277 is D277G.
45. The polymerase of claim 1, wherein the mutation at position 278 is K278N.
46. The polymerase of claim 1, wherein the mutation at position 279 is R279Q.
47. The polymerase of claim 1, wherein the mutation at position 282 is K282G or K282S.
48. The polymerase of claim 1, wherein the mutation at position 290 is R290K.
49. The polymerase of claim 1, wherein the mutation at position 294 is selected from the group consisting of D294Q, D294A, D294C, D294S, D294N, D294L, D294P, and D294H.
50. The polymerase of claim 1, wherein the mutation at position 295 is selected from the group consisting of I295Y, I295C, I295P, I295T, I295S, I295W, I295Y, I295F, I295M, and I295A.
51. The polymerase of claim 1, wherein the mutation at position 296 is selected from the group consisting of V296A, V296R, V296S, V296Q, and V296Y..

52. The polymerase of claim 1, wherein the mutation at position 297 is selected from the group consisting of S297H, S297T, S297P, S297W, S297L, S297Q, S297Y, S297C, S297R and S297F.

53. The polymerase of claim 1, wherein the mutation at position 298 is selected from the group consisting of R298L, R298K, R298T, R298E and R298G.

54. The polymerase of claim 1, wherein the mutation at position 299 is selected from the group consisting of G299L, G299F, G299W, G299I, G299A, G299H, G299S and G299Y.

55. The polymerase of claim 1, wherein the mutation at position 300 is selected from the group consisting of R300C, R300S, R300V and R300G.

56. The polymerase of claim 1, wherein the mutation at position 301 is selected from the group consisting of T301R, T301H, T301F, T301V, T301W, T301E, T301Y, T301M, T301K, T301S and T301Q.

57. The polymerase of claim 1, wherein the mutation at position 302 is F302E.

58. The polymerase of claim 1, wherein the mutation at position 303 is P303R.

59. The polymerase of claim 1, wherein the mutation at position 304 is H304G.

60. The polymerase of claim 1, wherein the mutation at position 308 is K308Q.

61. The polymerase of claim 1, wherein the mutation at position 309 is selected from the group consisting of E309N, E309P and E309D.
62. The polymerase of claim 1, wherein the mutation at position 314 is E314Q.
63. The polymerase of claim 1, wherein the mutation at position 317 is K317Q.
64. The polymerase of claim 1, wherein the mutation at position 320 is selected from the group consisting of Q320K, Q320R, Q320M, Q320I, Q320T, Q320N and Q320S.
65. The polymerase of claim 1, wherein the mutation at position 321 is selected from the group consisting of K321Q, K321N, K321R, K321S, K321G and K321M.
66. The polymerase of claim 1, wherein the mutation at position 322 is selected from the group consisting of I322A, I322L, I322C, I322V and I322T.
67. The polymerase of claim 1, wherein the mutation at position 324 is selected from the group consisting of E324K, E324N, E324R, E324A, E324G, E324S, E324R, E324W, E3234D and E324K..
68. The polymerase of claim 1, wherein the mutation at position 325 is selected from the group consisting of E325K, E325Q, E325R, E325G, E325V, E325S and E325A.
69. The polymerase of claim 1, wherein the mutation at position 326 is D326K.

70. The polymerase of claim 1, wherein the mutation at position 327 is selected from the group consisting of E327K, E337Q, and E327R.

71. The polymerase of claim 1, wherein the mutation at position 328 is R328N or R328S.

72. The polymerase of claim 1, comprising the amino acid sequence as set forth in any one of SEQ ID NOs: 2-106.

73. A composition comprising a recombinant DNA polymerase as set forth in any one of claims 1-71.

74. The composition of claim 73, wherein the composition is present in a DNA sequencing system that comprises at least one non-natural nucleotide analog substrate.

75. A modified nucleic acid encoding a modified DPO4-type DNA polymerase as set forth in any one of claims 1-71.

1 M I V L F V D F Y F Y A Q V E E V L N P S L K G K P V V V C V F S G R F E D S G A V A T A
 Mut1 Mut2 Mut13
 20 30 40
 50 N Y E A R K F G V K A G I P I V E A K K I L P N A V Y L P M R K E V Y Q Q V S S R I M N L L L
 Mut3 Mut4 Mut14
 60 70 80 90
 100 R E Y S E K I E I A S I D E A Y L D I S D K V R D Y R E A Y N L G L E I K N K I L E K E K I
 Mut5 Mut6 Mut7
 110 120 130 140
 150 T V T V G I S K N K V F A K I A A D M A K P N G I K V I D D E E V K R L I R E L D I A D V P
 Mut8 Mut9 Mut10 Mut11 Mut12
 160 170 180 190 200 210 220 230
 240 G I G N I T A E K L K K L G I N K L V D T L S I E F D K L K G M I G E A K A K Y L I S L A R
 Mut7 Mut8 Mut9 Mut10 Mut11 Mut12
 250 260 270 280
 290 D E Y N E P I R T R V R K S I G R I V T M K R N S R N L E E I K P Y L F R A I E E S Y Y K L
 Mut9 Mut10 Mut11 Mut12
 300 310 320
 330 D K R I P K A I H V V A V T E D L D I V S R G R T E P H G I S K E T A Y S E S V K L L Q K I
 Mut10 Mut11 Mut12
 340 350
 360 L E E D E R K I R R I G V R F S K F I E A I G L D K F F D T
 Mut12

FIG. 1B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/30972

A. CLASSIFICATION OF SUBJECT MATTER
 IPC(8) - C12Q 1/68, C12N 9/12 (2018.01)
 CPC - C12N 9/1252

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)

See Search History Document

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

See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2016/0304926 A1 (NOXXON PHARMA AG) 20 October 2016 (20.10.2016) para [0074], [0075], [0180], Table 1(A), SEQ ID NO: 15	1-75
A	UniProtKB accession number M9U6J3 (26 June 2013 [online]. [Retrieved 15 July 2018]. Retrieved from the internet <URL: https://www.uniprot.org/uniprot/M9U6J3.txt > whole doc.	1-75
A	US 7,745,188 B2 (WOODGATE et al.) 29 June 2010 (29.06.2010) col 6, ln 4-14, Fig. 15, SEQ ID NO: 72	1-75
A	US 2014/0127694 A1 (HOLLIGER et al.) 08 May 2014 (08.05.2014) para [0040], [0040], [0179]	1-75
A,P	WO 2017/087281 A1 (STRATOS GENOMICS, INC.) 26 May 2017 (26.05.2017) (hereinafter 'Stratos') pg 4, ln 9-23, Claim 51, SEQ ID NO: 1, 115	1-75

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

16 July 2018

Date of mailing of the international search report

08 AUG 2018

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
 P.O. Box 1450, Alexandria, Virginia 22313-1450
 Facsimile No. 571-273-8300

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
 PCT OSP: 571-272-7774