Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2742594 C 2020/03/24

(11)(21) **2 742 594**

(12) **BREVET CANADIEN**
CANADIAN PATENT

(13) **C**

(86) **Date de dépôt PCT/PCT Filing Date:** 2009/11/03
(87) **Date publication PCT/PCT Publication Date:** 2010/06/03
(45) **Date de délivrance/Issue Date:** 2020/03/24
(85) **Entrée phase nationale/National Entry:** 2011/05/03
(86) **N° demande PCT/PCT Application No.:** US 2009/063167
(87) **N° publication PCT/PCT Publication No.:** 2010/062777
(30) **Priorité/Priority:** 2008/11/03 (US61/110,877)

(51) **Cl.Int./Int.Cl.** ***C12N 9/12*** (2006.01),
C12N 15/10 (2006.01), ***C12N 15/54*** (2006.01)

(72) **Inventeurs/Inventors:**
BOURN, WILLIAM, ZA;
APPEL, MARYKE, ZA;
RUSH, GAVIN, ZA;
FOSKETT, JOHN, ZA;
MCEWAN, PAUL, ZA

(73) **Propriétaire/Owner:**
KAPA BIOSYSTEMS, INC., US

(74) **Agent:** FASKEN MARTINEAU DUMOULIN LLP

(54) **Titre : ADN POLYMERASES DE TYPE A MODIFIEES**
(54) **Title: MODIFIED TYPE A DNA POLYMERASES**

(57) **Abrégé/Abstract:**
The present invention provides improved DNA polymerases, in particular, type A DNA polymerases, that may be better suited for applications in recombinant DNA technologies. Among other things, the present invention provides modified DNA polymerases derived from directed evolution experiments designed to select mutations that confer advantageous phenotypes under conditions used in industrial or research applications.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 June 2010 (03.06.2010)

PCT

(10) International Publication Number
WO 2010/062777 A3

(51) International Patent Classification:
C12N 9/12 (2006.01) *C12N 15/10* (2006.01)
C12N 15/54 (2006.01)

(21) International Application Number:
PCT/US2009/063167

(22) International Filing Date:
3 November 2009 (03.11.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/110,877 3 November 2008 (03.11.2008) US

(71) Applicant *(for all designated States except US)*: **KAPABIOSYSTEMS** [US/US]; 600 West Cummings Park, Suite 5350, Woburn, MA 01801 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BOURN, William** [GB/ZA]; 59 Severn Road, 7801 Plumstead, Western Cape (ZA). **APPLE, Maryke** [ZA/ZA]; 6 Pipit Lane, Century View, 7441 Milnerton, Western Cape (ZA). **RUSH, Gavin** [ZA/ZA]; 48 Liesbeek Road, 7700 Rosebank, Western Cape (ZA). **FOSKETT, John** [US/ZA]; 25 The Drive, 8005 Camps Bay, Western Cape (ZA). **MCEWAN, Paul** [GB/ZA]; 3 Sonnekus, Victoria Road, 8005 Camps Bay, Western Cape (ZA).

(74) Agent: **CHEN, Fangli**; Choate, Hall & Stewart LLP, Two International Place, Boston, MA 02110 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

(88) Date of publication of the international search report:
26 August 2010

WO 2010/062777 A3

(54) Title: MODIFIED TYPE A DNA POLYMERASES

(57) Abstract: The present invention provides improved DNA polymerases, in particular, type A DNA polymerases, that may be better suited for applications in recombinant DNA technologies. Among other things, the present invention provides modified DNA polymerases derived from directed evolution experiments designed to select mutations that confer advantageous phenotypes under conditions used in industrial or research applications.

MODIFIED TYPE A DNA POLYMERASES

[0001] [Deleted]

BACKGROUND OF THE INVENTION

[0002] DNA polymerases are a family of enzymes that use single-stranded DNA as a template to synthesize the complementary DNA strand. In particular, DNA polymerases can add free nucleotides to the 3' end of a newly-forming strand resulting in elongation of the new strand in a 5'-3' direction. Most DNA polymerases are multifunctional proteins that possess both polymerizing and exonucleolytic activities (e.g., 3'->5' exonuclease or 5' -> 3' exonuclease activity).

[0003] DNA polymerases, like other natural enzymes, have evolved over millions of years to be efficient in their natural cellular environment. Many of them are almost perfectly adapted to work in that environment. In such an environment, the way that the protein can evolve is constrained by a number of requirements; the protein has to interact with other cellular components, it has to function in the cytoplasm (i.e., particular pH, ionic strength, in the presence of particular compounds, etc.) and it cannot cause lethal or disadvantageous side effects that detract from the fitness of the parent organism as a whole.

[0004] When DNA polymerases are removed from their natural environment and used in industrial or research applications, the environment and conditions under which the enzyme is operating is inevitably vastly different than those in which it evolved. Many of the constraints that limited the evolutionary direction the protein could take fall away. Therefore, there is vast potential for improvement of DNA polymerases for use in industrial or research applications.

SUMMARY OF THE INVENTION

[0005] The present invention provides improved DNA polymerases, in particular, type A DNA polymerases, that may be better suited for applications in recombinant DNA technologies. Among other things, the present invention provides modified DNA polymerases derived from directed evolution experiments designed to select mutations that confer advantageous phenotypes under conditions used in industrial or research applications.

[0006] Accordingly, in one aspect, the present invention provides modified type A DNA polymerases containing one or more amino acid alterations (e.g., one or more substitutions, deletions, or insertions) corresponding to one or more positions selected from the positions identified in Table 2 relative to the corresponding parental or wild-type enzyme. In some embodiments, such amino acid alterations alter (e.g., increase or decrease) enzyme activity, fidelity, processivity, elongation rate, stability, primer-dimer formation, salt resistance, solubility, expression efficiency, folding robustness, thermostability, polymerization activity, concentration robustness, resistance to impurities, strand-displacement activity, nucleotide selectivity, altered nuclease activity, resistance to nucleic acid intercalating dyes and/or other properties and characteristics involved in the process of DNA polymerization.

[0007] In some embodiments, modified type A DNA polymerases of the invention contain amino acid alterations at one or more positions corresponding to P6, K53, K56, E57, K171, T203, E209, D238, L294, V310, G364, E400, A414, E507, S515, E742 or E797 of Taq polymerase. For example, in some embodiments, the one or more positions includes a position corresponding to E507 of Taq polymerase.

[0008] In some embodiments, the amino acid alterations are amino acid substitutions. In some embodiments, the one or more amino acid substitutions correspond to amino acid substitutions selected from Table 2. In some embodiments, the one or more amino acid substitutions correspond to the substitutions selected from the group consisting of P6S, K53N, K56Q, E57D, K171R, T203I, E209G, E209K, D238N, L294P, V310A, G364D, G364S, E400K, A414T, E507K, S515G, E742K or E797G, and combinations thereof.

[0009] In some embodiments, the DNA polymerase is modified from a naturally-occurring polymerase, e.g., a naturally-occurring polymerase isolated from *Thermus*

aquaticus, Thermus thermophilus, Thermus caldophilus, Thermus filiformis, Thermus flavus, Thermotoga maritima, Bacillus strearothermophilus, or *Bacillus caldotenax*. In some embodiments, modified type A DNA polymerases of the invention are modified from a truncated version of a naturally-occurring polymerase, e.g., KlenTaq which contains a deletion of a portion of the 5' to 3' exonuclease domain (see, Barnes W. M. (1992) *Gene* 112:29-35; and Lawyer F.C. et al. (1993) *PCR Methods and Applications*, 2:275-287). In some embodiments, modified type A DNA polymerases of the invention are modified from a chimeric DNA polymerase.

[0010] In some embodiments, modified type A DNA polymerases of the invention are modified from a fusion polymerase.

[0011] In another aspect, present invention features kits containing modified type A DNA polymerases described herein. In addition, the present invention provides nucleotide sequences encoding modified type A DNA polymerases described herein and vectors and/or cells containing the nucleotide sequences according to the invention.

[0012] In another, related aspect, the present invention features modified Taq DNA polymerases containing one or more amino acid alterations (e.g., one or more substitutions, deletions, or insertions) at one or more positions selected from the positions identified in Table 2 relative to wild-type enzyme. In some embodiments, the one or more amino acid alterations increase enzyme activity, processivity, elongation rate, altered nuclease activity, resistance to salt, resistance to nucleic acid intercalating dyes or other PCR additives.

[0013] In some embodiments, the modified Taq DNA polymerases contain amino acid alterations at one or more positions corresponding to P6, K53, K56, E57, K171, T203, E209, D238, L294, V310, G364,E400, A414, E507, S515, E742, or E797.

[0014] In some embodiments, the one or more amino acid alterations are substitutions. In some embodiments, the one or more amino acid substitutions are selected from Table 2. In some embodiments, the one or more amino acid substitutions are selected from the group consisting of P6S, K53N, K56Q, E57D, K171R, T203I, E209G, E209K, D238N, L294P, V310A, G364D, G364S, E400K, A414T, E507K, S515G, E742K or E797G, and combinations thereof.

[0015] In other, related aspects, the present invention provides modified Taq DNA polymerases containing an amino acid sequence selected from the group consisting of SEQ ID NO:2 (A3E), SEQ ID NO:3 (G9S), SEQ ID NO:4 (D5S), SEQ ID NO:5 (D2), SEQ ID NO:6 (A5E), SEQ ID NO:7 (B6S), SEQ ID NO:8 (E2S), SEQ ID NO:9 (A3), SEQ ID NO:10 (H10), SEQ ID NO:11 (H1S), SEQ ID NO:12 (F9E), SEQ ID NO:13 (A5S), SEQ ID NO:14 (C10E), SEQ ID NO:15 (F5S), SEQ ID NO:16 (E7S), SEQ ID NO:17 (G6S), SEQ ID NO:18 (E1E), SEQ ID NO:19 (C7), SEQ ID NO:20 (E12), SEQ ID NO:21 (D9), SEQ ID NO:22 (F10), SEQ ID NO:23 (H7), SEQ ID NO:24 (A5) and combinations thereof.

[0015a] According to a particular aspect, the invention relates to a modified DNA polymerase that has at least 95% amino acid sequence identity with that of wild type Taq DNA polymerase set out in SEQ ID NO:l, but differs from wild type Taq DNA polymerase in that it includes a lysine at a position that corresponds to E507 of SEQ ID NO:1, and shows increased salt-resistance and ability to catalyze a DNA-template-dependent DNA polymerization activity that is improved, relative to that of the wild type Taq DNA polymerase of SEQ ID NO:1.

[0015b] According to another particular aspect, the invention relates to a modified Taq DNA polymerase that has at least 90% amino acid sequence identity with an amino acid sequence selected from the group consisting of SEQ ID NO:5 (D2), SEQ ID NO:6 (A5E), SEQ ID NO:7 (B6S), SEQ ID NO:8 (E2S), SEQ ID NO:9 (A3), SEQ ID NO : 10 (H10), SEQ ID NO : 11 (H1S), SEQ ID NO : 12 (F9E), SEQ ID NO : 13 (A5S), SEQ ID NO: 14 (C10E), SEQ ID NO: 15(F5S), SEQ ID NO: 16 (E7S), SEQ ID NO: 17 (G6S), SEQ ID NO: 18 (E1E), SEQ ID NO: 19 (C7), SEQ IDNO:20 (E12), SEQ ID NO:21 (D9), SEQ ID NO:22 (F10), SEQ ID NO:23 (H7), and SEQ ID NO:24 (A5), and is different from that of wild type Taq DNA polymerase as set forth in SEQ ID NO:1, which polymerase is characterized by: (i) presence of lysine at a position corresponding to E507 of SEQ ID NO:1 and (ii) improved ability to catalyze a DNA-template-dependent DNA polymerization activity relative to that of the wild type Taq DNA polymerase of SEQ ID NO:1.

[0015c] According to another particular aspect, the invention relates to a modified Taq DNA polymerase having an amino acid sequence that is at least 95% identical to a sequence selected from the group consisting of SEQ ID NO:5 (D2), SEQ ID NO:6 (A5E), SEQ ID NO:7 (B6S), SEQ ID NO:8 (E2S), SEQ ID NO:9 (A3), SEQ ID NO:10 (H10), SEQ ID NO:11 (H1S),

SEQ ID NO:12 (F9E), SEQ ID NO:13 (A5S), SEQ ID NO:14 (C10E), SEQ ID NO:15 (F5S), SEQ ID NO:16 (E7S), SEQ ID NO:17 (G6S), SEQ ID NO:18 (E1E), SEQ ID NO:20 (E12), SEQ ID NO:21 (D9), SEQ ID NO:22 (F10), SEQ ID NO:23 (H7), SEQ ID NO:24 (A5) and is different from that of wild type Taq DNA polymerase as set forth in SEQ ID NO:1, which polymerase (i) has a lysine at a position corresponding to E507 of SEQ ID NO:1 and (ii) shows increased salt tolerance and improved ability to catalyze a DNA-template-dependent DNA polymerization activity relative to that of the wild type Taq DNA polymerase of SEQ ID NO:1.

[0015d] According to another particular aspect, the invention relates to a modified Taq DNA polymerase whose amino acid sequence is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:5 (D2), SEQ ID NO:6 (A5E), SEQ ID NO:7 (B6S), SEQ ID NO:8 (E2S), SEQ ID NO:9 (A3), SEQ ID NO : 10 (H10), SEQ ID NO : 11 (H1S), SEQ ID NO : 12 (F9E), SEQ ID NO : 13 (A5S), SEQ ID NO: 14 (C10E), SEQ ID NO: 15(F5S), SEQ ID NO: 16 (E7S), SEQ ID NO: 17 (G6S), SEQ ID NO: 18 (E1E), SEQ ID NO: 19 (C7), SEQ IDNO:20 (E12), SEQ ID NO:21 (D9), SEQ ID NO:22 (F10), SEQ ID NO:23 (H7), and SEQ ID NO:24 (A5), and is different from that of wild type Taq DNA polymerase as set forth in SEQ ID NO:1, which polymerase (i) has a lysine at a position corresponding to E507 of SEQ ID NO: 1, (ii) is capable of catalyzing a DNA-template-dependent DNA polymerization activity; and (iii) which displays increased heparin binding affinity relative to the wild type Taq DNA polymerase of SEQ ID NO: 1 in that it shows improved conductivity when eluted off a column comprising heparin.

[0015e] According to another particular aspect, the invention relates to a Taq DNA polymerase that:

(a) is a mutant of a wild type Taq polymerase of SEQ ID NO:1 in that its amino acid sequence differs from that of SEQ ID NO:1;

(b) has an amino acid sequence that is at least 95% identical to that of a clone selected from the group consisting of: SEQ ID NO:5 (D2), SEQ ID NO:6 (A5E), SEQ ID NO:7 (B6S), SEQ ID NO:8 (E2S), SEQ ID NO:10 (H10), SEQ ID NO:12 (F9E), SEQ ID NO:13 (A5S), SEQ ID NO:14 (C10E), SEQ ID NO:15 (F5S), SEQ ID NO:16 (E7S), SEQ ID NO:17 (G6S), SEQ ID NO:20 (E12), SEQ ID NO:21 (D9), SEQ ID NO:22 (F10), SEQ ID NO:23 (H7), and SEQ ID NO:24 (A5);

(c) has a lysine at a position corresponding to E507 of SEQ ID NO: 1; and

(d) displays high processivity relative to the wild type Taq DNA polymerase of SEQ ID NO: 1 in that it shows improved ability to amplify DNA fragments of at least 5 kilobases in a PCR reaction.

[0016] The present invention also features kits containing a modified Taq DNA polymerase described herein and uses thereof. In one embodiment the kit comprises: i) a modified Taq DNA polymerase whose sequence is at least 95% identical to that of wild type Taq DNA polymerase of SEQ ID NO:1 and includes a substitution of lysine at a position corresponding to E507 of SEQ ID NO: 1, wherein the modified Taq DNA polymerase has improved ability to catalyze a DNA-template-dependent DNA polymerization activity, relative to the Taq of SEQ ID NO:1; and ii) at least one buffer suitable for use in a PCR reaction. In one embodiment the modified type A DNA polymerase comprises a sequence which includes a substitution of lysine at a position corresponding to E507 of wild type Taq of SEQ ID NO: 1.

[0016a] In addition, the present invention provides nucleotide molecules encoding modified Taq DNA polymerases described herein, and vectors and/or cells that include the nucleotide molecules.

[0017] The invention further provides methods including amplifying a DNA fragment using a modified type A DNA polymerases (e.g., Taq DNA polymerase) as described herein.

[0018] In some embodiments, the DNA fragment amplified according to the present invention is longer than 5kb (e.g., longer than 6kb, 7kb, 8kb, 9kb, 10kb, 12kb, or longer).

[0018a] According to a particular aspect, the invention relates to a method of performing a DNA polymerization reaction, the method comprising:

a) providing a mixture wherein the mixture comprises:

i) a template nucleic acid;

ii) at least one primer;

iii) nucleotides; and

iv) a modified Taq DNA polymerase whose amino acid sequence is at least 95% identical to that of wild type Taq DNA polymerase set forth in SEQ ID NO:1 and includes a lysine at a position corresponding to E507 of SEQ ID NO: 1, wherein the modified Taq DNA polymerase has improved ability to catalyze a DNA-template-dependent DNA polymerization activity, relative to the Taq of SEQ ID NO:1; and

b) incubating the mixture under conditions that:

i) permit hybridization of the at least one primer to the template nucleic acid; and

ii) permit extension of the at least one primer by polymerization of the nucleotides, which polymerization is catalyzed by the modified Taq DNA polymerase.

[0018b] According to another particular aspect, the invention relates to a method of providing an improvement in performance in a polymerase chain reaction (PCR) relative to that of wild type Taq DNA polymerase of SEQ ID NO:1, which improvement in performance is selected from one or more of enzyme activity, processivity, resistance to nucleic acid intercalating dyes, and salt-resistance, the method comprising:

using a modified Taq DNA polymerase whose sequence is at least 95% identical to that of wild type Taq DNA polymerase set forth in SEQ ID NO:1 and includes a lysine at a position corresponding to E507 of SEQ ID NO: 1 to perform the PCR, wherein the modified Taq DNA polymerase has improved ability to catalyze a DNA-template-dependent DNA polymerization activity, relative to the Taq of SEQ ID NO:1.

[0018c] The invention further relates to the use of a modified type A DNA polymerase as defined herein for amplifying a DNA fragment.

BRIEF DESCRIPTION OF THE DRAWING

[0019] The drawing is for illustration purposes only not for limitation.

[0020] Figure 1 depicts an alignment of amino acid sequences of naturally-occurring type A DNA polymerases from thermophilic bacterial species. Exemplary amino acid alterations discovered by directed evolution experiments are shown above each alignment.

DEFINITIONS

[0021] Amino *acid*: As used herein, term "amino acid," in its broadest sense, refers to any compound and/or substance that can be incorporated into a polypeptide chain. In some embodiments, an amino acid has the general structure H_2N–C(H)(R)–COOH. In some ———

embodiments, an amino acid is a naturally-occurring amino acid. In some embodiments, an amino acid is a synthetic amino acid; in some embodiments, an amino acid is a D-amino acid; in some embodiments, an amino acid is an L-amino acid. "Standard amino acid" refers to any of the twenty standard L-amino acids commonly found in naturally occurring peptides. "Nonstandard amino acid" refers to any amino acid, other than the standard amino acids, regardless of whether it is prepared synthetically or obtained from a natural source. As used herein, "synthetic amino acid" encompasses chemically modified amino acids, including but not limited to salts, amino acid derivatives (such as amides), and/or substitutions. Amino acids, including carboxy- and/or amino-terminal amino acids in peptides, can be modified by methylation, amidation, acetylation, and/or substitution with other chemical without adversely affecting their activity. Amino acids may participate in a disulfide bond. The term "amino acid" is used interchangeably with "amino acid residue," and may refer to a free amino acid and/or to an amino acid residue of a peptide. It will be apparent from the context in which the term is used whether it refers to a free amino acid or a residue of a peptide. It should be noted that all amino acid residue sequences are represented herein by formulae whose left and right orientation is in the conventional direction of amino-terminus to carboxy-terminus.

[0022] *Base Pair (bp)*: As used herein, base pair refers to a partnership of adenine (A) with thymine (T), or of cytosine (C) with guanine (G) in a double stranded DNA molecule.

[0023] *Chimeric polymerase*: As used herein, the term "chimeric polymerase" (also referred to as "chimera") refers to any recombinant polymerase containing at least a first amino acid sequence derived from a first DNA polymerase and a second amino acid sequence derived from a second DNA polymerase. Typically, the first and second DNA polymerases are characterized with at least one distinct functional characteristics (e.g., processivity, elongation rate, fidelity). As used herein, a sequence derived from a DNA polymerase of interest refers to any sequence found in the DNA polymerase of interest, or any sequence having at least 70% (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%) identical to an amino acid sequence found in the DNA polymerase of interest. A "chimeric polymerase" according to the invention may contain two or more amino acid sequences from related or similar polymerases (e.g., proteins sharing similar sequences and/or structures), joined to form a new functional protein. A "chimeric polymerase" according to the invention may contain two or more amino acid sequences from

unrelated polymerases, joined to form a new functional protein. For example, a chimeric polymerase of the invention may be an "interspecies" or "intergenic" fusion of protein structures expressed by different kinds of organisms.

[0024] *Complementary*: As used herein, the term "complementary" refers to the broad concept of sequence complementarity between regions of two polynucleotide strands or between two nucleotides through base-pairing. It is known that an adenine nucleotide is capable of forming specific hydrogen bonds ("base pairing") with a nucleotide which is thymine or uracil. Similarly, it is known that a cytosine nucleotide is capable of base pairing with a guanine nucleotide.

[0025] *DNA binding affinity:* As used herein, the term "DNA-binding affinity" typically refers to the activity of a DNA polymerase in binding DNA nucleic acid. In some embodiments, DNA binding activity can be measured in a two band-shift assay. For example, in some embodiments (based on the assay of Guagliardi *et al.* (1997) *J. Mol. Biol.* 267:841-848), double-stranded nucleic acid (the 452-bp HindIII-EcoRV fragment from the *S. solfataricus lacS* gene) is labeled with ^{32}P to a specific activity of at least about 2.5 X 10^7 cpm/μg (or at least about 4000 cpm/fmol) using standard methods. See, e.g., Sambrook *et al.* (2001) Molecular Cloning: A Laboratory Manual ($_3^{rd}$ ed., Cold Spring Harbor Laboratory Press, NY) at 9.63-9.75 (describing end-labeling of nucleic acids). A reaction mixture is prepared containing at least about 0.5 μg of the polypeptide in about 10 μl of binding buffer (50 mM sodium phosphate buffer (pH 8.0), 10% glycerol, 25 mM KCl, 25 mM $MgCl_2$). The reaction mixture is heated to 37 °C for 10 min. About 1 X 10^4 to 5 X 10^4 cpm (or about 0.5-2 ng) of the labeled double-stranded nucleic acid is added to the reaction mixture and incubated for an additional 10 min. The reaction mixture is loaded onto a native polyacrylamide gel in 0.5 X Tris-borate buffer. The reaction mixture is subjected to electrophoresis at room temperature. The gel is dried and subjected to autoradiography using standard methods. Any detectable decrease in the mobility of the labeled double-stranded nucleic acid indicates formation of a binding complex between the polypeptide and the double-stranded nucleic acid. Such nucleic acid binding activity may be quantified using standard densitometric methods to measure the amount of radioactivity in the binding complex relative to the total amount of radioactivity in the initial reaction mixture.

[0026] *Elongation rate*: As used herein, the term "elongation rate" refers to the average speed at which a DNA polymerase extends a polymer chain. As used herein, a high elongation rate refers to an elongation rate higher than 50 nt/s (e.g., higher than 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140 nt/s). As used in this application, the terms "elongation rate" and "speed" are used inter-changeably.

[0027] *Enzyme activity*: As used herein, the term "enzyme activity" refers to the specificity and efficiency of a DNA polymerase. Enzyme activity of a DNA polymerase is also referred to as "polymerase activity," which typically refers to the activity of a DNA polymerase in catalyzing the template-directed synthesis of a polynucleotide. Enzyme activity of a polymerase can be measured using various techniques and methods known in the art. For example, serial dilutions of polymerase can be prepared in dilution buffer (e.g., 20 mM Tris.Cl, pH 8.0, 50 mM KCl, 0.5% NP 40, and 0.5% Tween-20). For each dilution, 5 µl can be removed and added to 45 µl of a reaction mixture containing 25 mM TAPS (pH 9.25), 50 mM KCl, 2 mM $MgCl_2$, 0.2 mM dATP, 0.2 mM dGTP, 0.2 mM dTTP, 0.1 mM dCTP, 12.5 µg activated DNA, 100 µM [α-^{32}P]dCTP (0.05 µCi/nmol) and sterile deionized water. The reaction mixtures can be incubated at 37 °C. (or 74 °C. for thermostable DNA polymerases) for 10 minutes and then stopped by immediately cooling the reaction to 4 °C and adding 10 µl of ice-cold 60 mM EDTA. A 25 µl aliquot can be removed from each reaction mixture. Unincorporated radioactively labeled dCTP can be removed from each aliquot by gel filtration (Centri-Sep™ Princeton Separations, Adelphia, N.J.). The column eluate can be mixed with scintillation fluid (1 ml). Radioactivity in the column eluate is quantified with a scintillation counter to determine the amount of product synthesized by the polymerase. One unit of polymerase activity can be defined as the amount of polymerase necessary to synthesize 10 nmole of product in 30 minutes (Lawyer et al. (1989) *J. Biol. Chem.* 264:6427-647). Other methods of measuring polymerase activity are known in the art (see, e.g. Sambrook et al. (2001) Molecular Cloning: A Laboratory Manual (3.sup.rd ed., Cold Spring Harbor Laboratory Press, NY)).

[0028] *Fidelity*: As used herein, the term "fidelity" refers to the accuracy of DNA polymerization by template-dependent DNA polymerase. The fidelity of a DNA polymerase is typically measured by the error rate (the frequency of incorporating an inaccurate nucleotide, i.e., a nucleotide that is not incorporated at a template-dependent manner). The accuracy or fidelity of DNA polymerization is maintained by both the polymerase activity

and the exonuclease activity of a DNA polymerase. The term "high fidelity" refers to an error rate less than 4.45 X 10^{-6} (e.g., less than 4.0 X 10^{-6}, 3.5 X 10^{-6}, 3.0 X 10^{-6}, 2.5 X 10^{-6}, 2.0 X 10^{-6}, 1.5 X 10^{-6}, 1.0 X 10^{-6}, 0.5 X 10^{-6}) mutations/nt/doubling. The fidelity or error rate of a DNA polymerase may be measured using assays known to the art. For example, the error rates of DNA polymerases can be tested using the *lacI* PCR fidelity assay described in Cline, J. et al. (96) NAR 24: 3546-3551. Briefly, a 1. 9kb fragment encoding the *lacIOlacZα* target gene is amplified from pPRIAZ plasmid DNA using 2.5U DNA polymerase (i.e. amount of enzyme necessary to incorporate 25 nmoles of total dNTPs in 30 min. at 72°C) in the appropriate PCR buffer. The *lacI*-containing PCR products are then cloned into lambda GT10 arms, and the percentage of *lacI* mutants (MF, mutation frequency) is determined in a color screening assay, as described (Lundberg, K. S. , Shoemaker, D. D. , Adams, M. W. W. , Short, J. M. , Sorge, J. A., and Mathur, E. J. (1991) *Gene* 180: 1-8). Error rates are expressed as mutation frequency per bp per duplication (MF/bp/d), where bp is the number of detectable sites in the lacI gene sequence (349) and d is the number of effective target doublings. Similar to the above, any plasmid containing the *lacIOlacZα* target gene can be used as template for the PCR. The PCR product may be cloned into a vector different from lambda GT (e.g., plasmid) that allows for blue/white color screening.

[0029] *Fusion DNA polymerase:* As used herein, the term "fusion DNA polymerase" refers to any DNA polymerase that is combined (e.g., covalently or non-covalently) with one or more protein domains having a desired activity (e.g., DNA-binding, stabilizing template-primer complexes, hydrolyzing dUTP). In some embodiments, the one or more protein domains are derived from a non-polymerase protein. Typically, fusion DNA polymerases are generated to improve certain functional characteristics (e.g., processivity, elongation rate, fidelity, salt-resistance, etc.) of a DNA polymerase.

[0030] *Modified DNA polymerase:* As used herein, the term "modified DNA polymerase" refers to a DNA polymerase originated from another (i.e., parental) DNA polymerase and contains one or more amino acid alterations (e.g., amino acid substitution, deletion, or insertion) compared to the parental DNA polymerase. In some embodiments, a modified DNA polymerases of the invention is originated or modified from a naturally-occurring or wild-type DNA polymerase. In some embodiments, a modified DNA polymerase of the invention is originated or modified from a recombinant or engineered DNA polymerase including, but not limited to, chimeric DNA polymerase, fusion DNA

polymerase or another modified DNA polymerase. Typically, a modified DNA polymerase has at least one changed phenotype compared to the parental polymerase.

[0031] *Mutation*: As used herein, the term "mutation" refers to a change introduced into a parental sequence, including, but not limited to, substitutions, insertions, deletions (including truncations). The consequences of a mutation include, but are not limited to, the creation of a new character, property, function, phenotype or trait not found in the protein encoded by the parental sequence. Herein, the term "mutation" is used interchangeably with "alteration."

[0032] *Mutant*: As used herein, the term "mutant" refers to a modified protein which displays altered characteristics when compared to the parental protein.

[0033] *Joined*: As used herein, "joined" refers to any method known in the art for functionally connecting polypeptide domains, including without limitation recombinant fusion with or without intervening domains, inter-mediated fusion, non-covalent association, and covalent bonding, including disulfide bonding, hydrogen bonding, electrostatic bonding, and conformational bonding.

[0034] *Nucleotide*: As used herein, a monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is a nucleoside. When the nucleoside contains a phosphate group bonded to the 3' or 5' position of the pentose it is referred to as a nucleotide. A sequence of operatively linked nucleotides is typically referred to herein as a "base sequence" or "nucleotide sequence," and is represented herein by a formula whose left to right orientation is in the conventional direction of 5'-terminus to 3'-terminus.

[0035] *Nucleic acid intercalating dyes*: As used herein, the term "nucleic acid intercalating dyes" refers to any molecules that bind to nucleic acids in a reversible, non-covalent fashion, by insertion between the base pairs of the double helix, thereby indicating the presence and amount of nucleic acids. Generally, nucleic acid intercalating dyes are planar, aromatic, ring-shaped chromophore molecules. In some embodiments, intercalating dyes include fluorescent dyes. Numerous intercalating dyes are known in the art. Some non-limiting examples include PICO GREEN™ (P-7581, Molecular Probes), EB (E-8751, Sigma),

propidium iodide (P-4170, Sigma), Acridine orange (A-6014, Sigma), 7-aminoactinomycin D (A-1310, Molecular Probes), cyanine dyes (e.g., TOTO™, YOYO™, BOBO™, and POPO™), SYTO™, SYBR™ Green I, SYBR™ Green II, SYBR™ DX, OliGreen™, CyQuant™ GR, SYTOX™ Green, SYTO9™, SYTO10™, SYTO17™, SYBR14™, FUN-1™, DEAD Red™, Hexidium Iodide, Dihydroethidium, Ethidium Homodimer, 9-Amino-6-Chloro-2-Methoxyacridine, DAPI, DIPI, Indole dye, Imidazole dye, Actinomycin D, Hydroxystilbamidine, and LDS 751™ (U.S. Pat. No. 6,210,885), BOXTO™, LC Green™, Evagreen™, Bebo™.

[0036] *Oligonucleotide or Polynucleotide*: As used herein, the term "oligonucleotide" is defined as a molecule including two or more deoxyribonucleotides and/or ribonucleotides, preferably more than three. Its exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. The oligonucleotide may be derived synthetically or by cloning. As used herein, the term "polynucleotide" refers to a polymer molecule composed of nucleotide monomers covalently bonded in a chain. DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are examples of polynucleotides.

[0037] *Polymerase*: As used herein, a "polymerase" refers to an enzyme that catalyzes the polymerization of nucleotide (i.e., the polymerase activity). Generally, the enzyme will initiate synthesis at the 3'- end of the primer annealed to a polynucleotide template sequence, and will proceed toward the 5' end of the template strand. A "DNA polymerase" catalyzes the polymerization of deoxynucleotides.

[0038] *Primer*: As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally or produced synthetically, which is capable of acting as a point of initiation of nucleic acid synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, *e.g.*, in the presence of four different nucleotide triphosphates and thermostable enzyme in an appropriate buffer ("buffer" includes pH, ionic strength, cofactors, etc.) and at a suitable temperature. The primer is preferably single-stranded for maximum efficiency in amplification, but may alternatively be double-stranded. If double-stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the thermostable enzyme. The

exact lengths of the primers will depend on many factors, including temperature, source of primer and use of the method. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25 nucleotides, although it may contain more or few nucleotides. Short primer molecules generally require colder temperatures to form sufficiently stable hybrid complexes with template.

[0039] *Processivity*: As used herein, "processivity" refers to the ability of a polymerase to remain attached to the template and perform multiple modification reactions. "Modification reactions" include but are not limited to polymerization, and exonucleolytic cleavage. In some embodiments, "processivity" refers to the ability of a DNA polymerase to perform a sequence of polymerization steps without intervening dissociation of the enzyme from the growing DNA chains. Typically, "processivity" of a DNA polymerase is measured by the length of nucleotides (for example 20 nts, 300 nts, 0.5-1 kb, or more) that are polymerized or modified without intervening dissociation of the DNA polymerase from the growing DNA chain. "Processivity" can depend on the nature of the polymerase, the sequence of a DNA template, and reaction conditions, for example, salt concentration, temperature or the presence of specific proteins. As used herein, the term "high processivity" refers to a processivity higher than 20nts (e.g., higher than 40nts, 60nts, 80nts, 100nts, 120nts, 140nts, 160nts, 180nts, 200nts, 220nts, 240nts, 260nts,280nts, 300nts, 320nts, 340nts, 360nts, 380nts, 400nts, or higher) per association/disassociation with the template. Processivity can be measured according the methods defined herein and in WO 01/92501 A1.

[0040] *Synthesis*: As used herein, the term "synthesis" refers to any *in vitro* method for making new strand of polynucleotide or elongating existing polynucleotide (i.e., DNA or RNA) in a template dependent manner. Synthesis, according to the invention, includes amplification, which increases the number of copies of a polynucleotide template sequence with the use of a polymerase. Polynucleotide synthesis (e.g., amplification) results in the incorporation of nucleotides into a polynucleotide (i.e., a primer), thereby forming a new polynucleotide molecule complementary to the polynucleotide template. The formed polynucleotide molecule and its template can be used as templates to synthesize additional polynucleotide molecules. "DNA synthesis," as used herein, includes, but is not limited to, PCR, the labeling of polynucleotide (i.e., for probes and oligonucleotide primers), polynucleotide sequencing.

[0041] *Template DNA molecule*: As used herein, the term "template DNA molecule" refers to a strand of a nucleic acid from which a complementary nucleic acid strand is synthesized by a DNA polymerase, for example, in a primer extension reaction.

[0042] *Template dependent manner*: As used herein, the term "template dependent manner" refers to a process that involves the template dependent extension of a primer molecule (e.g., DNA synthesis by DNA polymerase). The term "template dependent manner" typically refers to polynucleotide synthesis of RNA or DNA wherein the sequence of the newly synthesized strand of polynucleotide is dictated by the well-known rules of complementary base pairing (see, for example, Watson, J. D. et al., In: Molecular Biology of the Gene, 4th Ed. , W. A. Benjamin, Inc., Menlo Park, Calif. (1987)).

[0043] *Thermostable enzyme*: As used herein, the term "thermostable enzyme" refers to an enzyme which is stable to heat (also referred to as heat-resistant) and catalyzes (facilitates) polymerization of nucleotides to form primer extension products that are complementary to a polynucleotide template sequence. Typically, thermostable stable polymerases are preferred in a thermocycling process wherein double stranded nucleic acids are denatured by exposure to a high temperature (e.g., about 95°C) during the PCR cycle. A thermostable enzyme described herein effective for a PCR amplification reaction satisfies at least one criteria, i.e., the enzyme does not become irreversibly denatured (inactivated) when subjected to the elevated temperatures for the time necessary to effect denaturation of double-stranded nucleic acids. Irreversible denaturation for purposes herein refers to permanent and complete loss of enzymatic activity. The heating conditions necessary for denaturation will depend, e.g., on the buffer salt concentration and the length and nucleotide composition of the nucleic acids being denatured, but typically range from about 90 °C to about 96 °C for a time depending mainly on the temperature and the nucleic acid length, typically about 0.5 to ten minutes. Higher temperatures may be tolerated as the buffer salt concentration and/or GC composition of the nucleic acid is increased. In some embodiments, thermostable enzymes will not become irreversibly denatured at about 90 °C -100 °C. Typically, a thermostable enzyme suitable for the invention has an optimum temperature at which it functions that is higher than about 40 °C, which is the temperature below which hybridization of primer to template is promoted, although, depending on (1) magnesium and salt, concentrations and (2) composition and length of primer, hybridization can occur at higher temperature (e.g., 45 °C - 70 °C). The higher the temperature optimum for the enzyme, the greater the specificity

and/or selectivity of the primer-directed extension process. However, enzymes that are active below 40 °C (e.g., at 37 °C) are also with the scope of this invention provided they are heat-stable. In some embodiments, the optimum temperature ranges from about 50 °C to 90 °C (e.g., 60 °C – 80 °C).

[0044] *Wild-type*: As used herein, the term "wild-type" refers to a gene or gene product which has the characteristics of that gene or gene product when isolated from a naturally-occurring source.

DETAILED DESCRIPTION OF THE INVENTION

[0045] The present invention provides, among other things, modified DNA polymerases (e.g., type A DNA polymerases) containing amino acid alterations based on mutations identified in directed evolution experiments designed to select enzymes that are better suited for applications in recombinant DNA technologies.

[0046] As described in the Examples section, the present inventors have successfully developed directed DNA polymerase evolution experiments by mimicking the typical or less-than typical environments and conditions under which an enzyme is usually used or expected to be used in real-life industrial or research applications.

[0047] As discussed in the Examples, various mutations have been observed during the selection process (see Table 2). Many mutations confer advantages relating to enzyme characteristics including, but not limited to, expression efficiency, solubility and folding robustness, thermostability, polymerization activity, processivity, speed (elongation rate), concentration robustness, resistance to impurities, resistance to chemical additives, fidelity, avoidance of primer-dimers, strand-displacement activity, altered nuclease activity, nucleotide selectivity, and other properties and characteristics involved in the process of DNA polymerization.

[0048] It is contemplated that the mutations identified herein confer a variety of phenotypes that can make DNA polymerases better suited for applications in recombinant DNA technologies. For example, mutations identified in accordance with the present invention may confer enzymatic phenotypes related to the selective advantages described

herein. Indeed, the present inventors have identified or expect to identify mutant polymerases that express well, are more soluble, that display higher activity, fidelity, processivity and/or speed, that are active over a wide range of concentrations, that are resistant to salt, PCR additives (e.g., PCR enhancers) and/or inhibitors, that work over a range of concentrations and have a higher fidelity, and other phenotypes that may not be immediately measurable. Since many of these phenotypes may depend on the manner in which the DNA and polymerase interact, it is contemplated that many of the mutations identified in accordance with the present invention may affect DNA-polymerase binding characteristics.

[0049] In addition, it is contemplated that mutations identified according to the present invention may confer enzymatic phenotypes not directly related to the selective advantages described herein. For example, some phenotypes may confer no advantage, but merely be a side effect of the advantageous mutation. In addition, some mutants may display phenotypes that could be considered disadvantageous. For example, some mutations confer an advantage (for example, high activity), but this advantage comes at a cost (for example, high error-rate). If the advantage outweighs the disadvantage, the mutation will still be selected for. Such mutations may have commercial uses. For example, a low fidelity enzyme could be used in error prone PCR (e.g., for mutagenesis).

[0050] Exemplary mutations and mutant clones containing combinations of mutations associated with specific phenotypes are discussed in the Examples section and are shown at least in Tables 3, 4, 5, 8, 12, and 15.

[0051] It is further contemplated that, since many DNA polymerases have similar sequences, structures and functional domains, mutations and/or the positions where mutations occur identified herein can serve as bases for modification of DNA polymerases in general. For example, same or similar mutations, as well as other alterations, may be introduced at the corresponding positions in various DNA polymerases to generate modified enzymes that are better adapted for recombinant use.

DNA polymerases

[0052] DNA polymerases in accordance with the present invention may be modified from any types of DNA polymerases including, but not limited to, naturally-occurring wild-type DNA polymerases, recombinant DNA polymerase or engineered DNA polymerases such as chimeric DNA polymerases, fusion DNA polymerases, or other modified DNA polymerases. In particular embodiments, DNA polymerases suitable for the invention are thermostable DNA polymerases (PCR-able).

Naturally-occurring DNA polymerases

[0053] In some embodiments, naturally-occurring DNA polymerases suitable for the invention are type A DNA polymerases (also known as family A DNA polymerases). Type A DNA polymerases are classified based on amino acid sequence homology to *E. coli* polymerase I (Braithwaite and Ito, *Nuc. Acids. Res.* 21:787-802, 1993), and include *E. coli* pol I, *Thermus aquaticus* DNA pol I (Taq polymerase), *Thermus flavus* DNA pol I, *Streptococcus pneumoniae* DNA pol I, *Bacillus stearothermophilus* pol I, phage polymerase T5, phage polymerase T7, mitochondrial DNA polymerase pol gamma, as well as additional polymerases discussed below.

[0054] Family A DNA polymerases are commercially available, including Taq polymerase (New England BioLabs), *E. coli* pol I (New England BioLabs), *E. coli* pol I Klenow fragment (New England BioLabs), and T7 DNA polymerase (New England BioLabs), and *Bacillus stearothermophilus* (Bst) DNA polymerase (New England BioLabs).

[0055] Suitable DNA polymerases can also be derived from bacteria or other organisms with optimal growth temperatures that are similar to the desired assay temperatures. For example, such suitable bacteria or other organisms may exhibit maximal growth temperatures of > 80-85°C or optimal growth temperatures of > 70-80°C.

[0056] Sequence information of many type A DNA polymerases are publicly available. Table 1 provides a list of GenBank Accession numbers and other GenBank Accession information for exemplary type A DNA polymerases, including species from which they are derived.

Table 1. Sequence Accession Information for Certain Type A DNA Polymerases

Geobacillus stearothermophilus
ACCESSION 3BDP_A
VERSION 3BDP_A GI:4389065
DBSOURCE pdb: molecule 3BDP, chain 65, release Aug 27, 2007.

Natranaerobius thermophilus JW/NM-WN-LF
ACCESSION ACB85463
VERSION ACB85463.1 GI:179351193
DBSOURCE accession CP001034.1

Thermus thermophilus HB8
ACCESSION P52028
VERSION P52028.2 GI:62298349
DBSOURCE swissprot: locus DPO1T_THET8, accession P52028

Thermus thermophilus
ACCESSION P30313
VERSION P30313.1 GI:232010
DBSOURCE swissprot: locus DPO1F_THETH, accession P30313

Thermus caldophilus
ACCESSION P80194
VERSION P80194.2 GI:2506365
DBSOURCE swissprot: locus DPO1_THECA, accession P80194

Thermus filiformis
ACCESSION O52225
VERSION O52225.1 GI:3913510
DBSOURCE swissprot: locus DPO1_THEFI, accession O52225

Thermus filiformis
ACCESSION AAR11876
VERSION AAR11876.1 GI:38146983
DBSOURCE accession AY247645.1

Thermus aquaticus
ACCESSION P19821
VERSION P19821.1 GI:118828
DBSOURCE swissprot: locus DPO1_THEAQ, accession P19821

Thermotoga lettingae TMO
ACCESSION YP_001469790
VERSION YP_001469790.1 GI:157363023
DBSOURCE REFSEQ: accession NC_009828.1

Thermosipho melanesiensis BI429
ACCESSION YP_001307134
VERSION YP_001307134.1 GI:150021780
DBSOURCE REFSEQ: accession NC_009616.1

Thermotoga petrophila RKU-1
ACCESSION YP_001244762
VERSION YP_001244762.1 GI:148270302
DBSOURCE REFSEQ: accession NC_009486.1

Thermotoga maritima MSB8
ACCESSION NP_229419
VERSION NP_229419.1 GI:15644367
DBSOURCE REFSEQ: accession NC_000853.1

Thermodesulfovibrio yellowstonii DSM 11347
ACCESSION YP_002249284
VERSION YP_002249284.1 GI:206889818
DBSOURCE REFSEQ: accession NC_011296.1

Dictyoglomus thermophilum
ACCESSION AAR11877
VERSION AAR11877.1 GI:38146985
DBSOURCE accession AY247646.1

Geobacillus sp. MKK-2005
ACCESSION ABB72056
VERSION ABB72056.1 GI:82395938
DBSOURCE accession DQ244056.1

Bacillus caldotenax
ACCESSION BAA02361
VERSION BAA02361.1 GI:912445
DBSOURCE locus BACPOLYTG accession D12982.1

Thermoanaerobacter thermohydrosulfuricus
ACCESSION AAC85580
VERSION AAC85580.1 GI:3992153
DBSOURCE locus AR003995 accession AAC85580.1

Thermoanaerobacter pseudethanolicus ATCC 33223
ACCESSION ABY95124
VERSION ABY95124.1 GI:166856716
DBSOURCE accession CP000924.1

Enterobacteria phage T5
ACCESSION AAS77168 CAA04580
VERSION AAS77168.1 GI:45775036
DBSOURCE accession AY543070.1

Enterobacteria phage T7 (T7)
ACCESSION NP_041982
VERSION NP_041982.1 GI:9627454
DBSOURCE REFSEQ: accession NC_001604.1

[0001] DNA polymerases suitable for the present invention include DNA polymerases that have not yet been isolated.

Truncated DNA polymerases

[0002] In some embodiments, DNA polymerases suitable for the present invention include truncated versions of naturally-occurring polymerases (e.g., a fragment of a DNA polymerase resulted from an N-terminal, C-terminal or internal deletion that retains polymerase activity). One exemplary truncated DNA polymerase suitable for the invention is KlenTaq which contains a deletion of a portion of the 5' to 3' exonuclease domain (see, Barnes W. M. (1992) *Gene* 112:29-35; and Lawyer F.C. et al. (1993) *PCR Methods and Applications*, 2:275-287).

Chimeric DNA polymerases

[0003] In some embodiments, chimeric DNA polymerases suitable for the invention include any DNA polymerases containing sequences derived from two or more different DNA polymerases. In some embodiments, chimeric DNA polymerases suitable for the invention include chimeric DNA polymerases as described in co-pending application entitled "Chimeric DNA polymerases" filed on even date.

[0004] Chimeric DNA polymerases suitable for the invention also include the chimeric DNA polymerases described in U.S. Publication No. 20020119461, U.S. Patent Nos. 6,228,628 and 7,244,602.

Fusion DNA polymerases

[0005] Suitable fusion DNA polymerases include any DNA polymerases that are combined (e.g., covalently or non-covalently) with one or more protein domains having a desired activity (e.g., DNA-binding, dUTP hydrolysis or stabilizing template-primer complexes). In some embodiments, the one or more protein domains having the desired activity are derived from a non-polymerase protein. Typically, fusion DNA polymerases are generated to improve certain functional characteristics (e.g., processivity, elongation rate, fidelity, salt-resistance, dUTP tolerance etc.) of a DNA polymerase. For example, DNA polymerase has been fused in frame to the helix-hairpin-helix DNA binding motifs from DNA topoisomerase V and shown to increase processivity, salt resistance and thermostability of the fusion DNA polymerase as described in Pavlov et al., 2002, *Proc. Natl. Acad. Sci USA*, 99:13510-13515. Fusion of the thioredoxin binding domain to T7 DNA polymerase enhances the processivity of the DNA polymerase fusion in the presence of thioredoxin as described in WO 97/29209, U.S. Pat. No. 5,972,603 and Bedford et al. *Proc. Natl. Acad. Sci. USA* 94: 479-484 (1997). Fusion of the archaeal PCNA binding domain to Taq DNA polymerase results in a DNA polymerase fusion that has enhanced processivity and produces higher yields of PCR amplified DNA in the presence, of PCNA (Motz, M., et al., *J. Biol. Chem.* May 3, 2002; 277 (18); 16179-88). Also, fusion of the sequence non-specific DNA binding protein *Sso7d* or *Sac7d* from *Sulfolobus sulfataricus* to a DNA polymerase, such as Pfu or Taq DNA polymerase, was shown to greatly increase the processivity of these DNA polymerases as disclosed in WO 01/92501 A1. Additional fusion polymerases are described in US Publication No. 20070190538A1.

[0006] Commercially available exemplary fusion polymerases include, but are not limited to, TopoTaq™ (Fidelity Systems) which is a hybrid of Taq polymerase fused to a sequence non-specific Helix–hairpin–helix (HhH) motif from DNA topoisomerase V (Topo V) (see, US Patent Nos. 5,427,928; 5,656,463; 5,902,879; 6,548,251; Pavlov et al., 2002, *Proc. Natl. Acad. Sci USA*, 99:13510-13515); Phusion™ (Finnzymes and NEB, sold by BioRad as iProof) which is a chimeric Deep Vent™/Pfu DNA polymerase fused to a small basic chromatin-like Sso7d protein (see, U.S. Patent No. 6627424, U.S. Application Publication Nos. 20040191825, 20040081963, 20040002076, 20030162173, 20030148330, and Wang et al. 2004, Nucleic Acids Research,

32(3), 1197-1207.); PfuUltra™ II Fusion (Stratagene) which is a Pfu-based DNA polymerase fused to a double stranded DNA binding protein (U.S. Application No. 20070148671); Herculase II Fusion™ (Stratagene) which is a Herculase II™ enzyme fused to a DNA-binding domain; and Pfx50™ (Invitrogen) which is a DNA polymerase from T. zilligii fused to an accessory protein that stabilizes primer-template complexes.

Generation of modified DNA polymerases of the invention

[0063] Modified DNA polymerases can be generated by introducing one or more amino acid alterations into a DNA polymerase at the positions corresponding to the positions described herein (e.g., positions identified in Tables 2, 3, 4, 5, 8, 12, and 15).

[0064] Corresponding positions in various DNA polymerases can be determined by alignment of amino acid sequences. Alignment of amino acid sequences can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST™, ALIGN™ or Megalign™ (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. Preferably, the WU-BLAST-2™ software is used to determine amino acid sequence identity (Altschul *et al.*, *Methods in Enzymology* 266, 460-480 (1996); URL: //blast.wustl/edu/blast/README.html). WU-BLAST-2™ uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span=1, overlap fraction=0.125, word threshold (T)=11. HSP score (S) and HSP S2 parameters are dynamic values and are established by the program itself, depending upon the composition of the particular sequence, however, the minimum values may be adjusted and are set as indicated above. An example of an alignment is shown in Figure 1.

[0065] Alterations may be a substitution, deletion or insertion of one or more amino acid residues. Appropriate alteration for each position can be determined by examining the nature and the range of mutations at the corresponding position described herein. In some embodiments, appropriate amino acid alterations can be determined by evaluating a three-dimensional structure of a DNA polymerase of interest (e.g., parental DNA polymerase). For

example, amino acid substitutions identical or similar to those described in Tables 2, 3, and 4 can be introduced to a DNA polymerase. Alternative amino acid substitutions can be made using any of the techniques and guidelines for conservative and non-conservative amino acids as set forth, for example, by a standard Dayhoff frequency exchange matrix or BLOSUM matrix. Six general classes of amino acid side chains have been categorized and include: Class I (Cys); Class II (Ser, Thr, Pro, Ala, Gly); Class III (Asn, Asp, Gln, Glu); Class IV (His, Arg, Lys); Class V (Ile, Leu, Val, Met); and Class VI (Phe, Tyr, Trp). For example, substitution of an Asp for another class III residue such as Asn, Gln, or Glu, is a conservative substitution. As used herein, "non-conservative substitution" refers to the substitution of an amino acid in one class with an amino acid from another class; for example, substitution of an Ala, a class II residue, with a class III residue such as Asp, Asn, Glu, or Gln. Insertions or deletions may optionally be in the range of 1 to 5 amino acids.

[0066] Appropriate amino acid alterations allowed in relevant positions may be confirmed by testing the resulting modified DNA polymerases for activity in the *in vitro* assays known in the art or as described in the Examples below.

[0067] The variations can be made using methods known in the art such as oligonucleotide-mediated (site-directed) mutagenesis, and PCR mutagenesis. Site-directed mutagenesis (Carter et al., *Nucl. Acids Res.*, 13:4331 (1986); Zoller et al., *Nucl. Acids Res.*, 10:6487 (1987)), cassette mutagenesis (Wells et al., *Gene*, 34:315 (1985)), restriction selection mutagenesis (Wells et al., Philos. *Trans. R. Soc.* London SerA, 317:415 (1986)), inverse PCR with mutations included in the primer sequence, or other known techniques can be performed on the cloned DNA to produce desired modified DNA polymerases.

[0068] In some embodiments, alterations suitable for the invention also include chemical modification including acetylation, acylation, amidation, ADP-ribosylation, glycosylation, GPI anchor formation, covalent attachment of a lipid or lipid derivative, methylation, myristlyation, pegylation, prenylation, phosphorylation, ubiqutination, or any similar process.

[0069] Modified DNA polymerases according to the invention may contain one or more amino acid alterations at one or more positions corresponding to those described in Tables 2, 3, 4, 5, 8, 12, and 15. Modified DNA polymerases according to the invention may also contain additional substitutions, insertions and/or deletions independent of the mutations

observed or selected in the directed evolution experiments. Thus, in some embodiments, a modified DNA polymerase according to the invention has an amino acid sequence at least 70%, e.g., at least 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99%, identical to a corresponding wild-type (or naturally-occurring) DNA polymerase. In some embodiments, a modified DNA polymerase has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 amino acid substitutions, deletions, insertions, or a combination thereof, relative to a wild type form of the polymerase.

[0070] "Percent (%) amino acid sequence identity" is defined as the percentage of amino acid residues in a modified sequence that are identical with the amino acid residues in the corresponding parental sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity are similar to the alignment for purposes of determining corresponding positions as described above.

[0071] Methods well known in the art may be applied to express and isolate modified DNA polymerases. Many bacterial expression vectors contain sequence elements or combinations of sequence elements allowing high level inducible expression of the protein encoded by a foreign sequence. For example, expression vectors are commercially available from, for example, Novagen (http://www.emdbiosciences.com/html/NVG/AllTables.html#).

[0072] As an example, bacteria expressing an integrated inducible form of the T7 RNA polymerase gene may be transformed with an expression vector bearing a modified DNA polymerase gene linked to the T7 promoter. Induction of the T7 RNA polymerase by addition of an appropriate inducer, for example, isopropyl-p-D-thiogalactopyranoside (IPTG) for a lac-inducible promoter, induces the high level expression of the chimeric gene from the T7 promoter.

[0073] Appropriate host strains of bacteria may be selected from those available in the art by one of skill in the art. As a non-limiting example, *E. coli* strain BL-21 is commonly used for expression of exogenous proteins since it is protease deficient relative to other strains of *E. coli*. For situations in which codon usage for the particular polymerase gene differs from that normally seen in *E. coli* genes, there are strains of BL-21 that are modified to carry tRNA genes encoding tRNAs with rarer anticodons (for example, argU,

ileY, leuW, and proL tRNA genes), allowing high efficiency expression of cloned chimeric genes (several BL21-CODON PLUSTM cell strains carrying rare-codon tRNAs are available from Stratagene, for example). Additionally or alternatively, genes encoding DNA polymerases may be codon optimized to facilitate expression in *E. coli*. Codon optimized sequences can be chemically synthesized.

[0074] There are many methods known to those of skill in the art that are suitable for the purification of a modified DNA polymerase of the invention. For example, the method of Lawyer et al. (1993, *PCR Meth. & App.* 2: 275) is well suited for the isolation of DNA polymerases expressed in *E. coli*, as it was designed originally for the isolation of Taq polymerase. Alternatively, the method of Kong et al. (1993, *J. Biol. Chem.* 268: 1965, incorporated herein by reference) may be used, which employs a heat denaturation step to destroy host proteins, and two column purification steps (over DEAE-Sepharose™ and heparin-Sepharose™ columns) to isolate highly active and approximately 80% pure DNA polymerase.

[0075] Further, modified DNA polymerase may be isolated by an ammonium sulfate fractionation, followed by Q Sepharose™ and DNA cellulose columns, or by adsorption of contaminants on a HiTrap Q™ column, followed by gradient elution from a HiTrap heparin column.

Applications of modified DNA polymerases of the invention

[0076] Modified DNA polymerases of the present invention may be used for any methods involving polynucleotide synthesis. Polynucleotide synthesis methods are well known to a person of ordinary skill in the art and can be found, for example, in Molecular Cloning second edition, Sambrook et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y. (1989). For example, modified DNA polymerases of the present invention have a variety of uses in recombinant DNA technology including, but not limited to, labeling of DNA by nick translation, second-strand cDNA synthesis in cDNA cloning, DNA sequencing, whole-genome amplification and amplifying, detecting, and/or cloning nucleic acid sequences using polymerase chain reaction (PCR).

[0077] In some embodiments, the invention provides enzymes that are better suited for PCR used in industrial or research applications. PCR refers to an *in vitro* method for

amplifying a specific polynucleotide template sequence. The technique of PCR is described in numerous publications, including, PCR: A Practical Approach, M. J. McPherson, et al., IRL Press (1991), PCR Protocols: A Guide to Methods and Applications, by Innis, et al., Academic Press (1990), and PCR Technology : Principals and Applications for DNA Amplification, H. A. Erlich, Stockton Press (1989). PCR is also described in many U. S. Patents, including U. S. Patent Nos. 4,683, 195; 4,683, 202; 4,800, 159; 4,965, 188; 4,889, 818; 5,075, 216 ; 5,079, 352; 5,104, 792; 5,023, 171; 5,091, 310; and 5,066, 584.

[0078] Modified DNA polymerases with higher processivity, elongation rate, salt resistance, and/or fidelity are expected to improve efficiency and success rate of long-range amplification (higher yield, longer targets amplified) and reduce the amount of required DNA template.

[0079] Various specific PCR amplification applications are available in the art (for reviews, see for example, Erlich, 1999, *Rev Immunogenet.*, 1: 127-34; Prediger 2001, *Methods Mol. Biol.* 160: 49-63; Jurecic et al., 2000, *Curr. Opin. Microbiol.* 3: 316-21; Triglia, 2000, *Methods Mol. Biol.* 130: 79-83; MaClelland et al., 1994, *PCR Methods Appl.* 4: S66-81 ; Abramson and Myers, 1993, *Current Opinion in Biotechnology* 4: 41-47).

[0080] As non-limiting examples, modified DNA polymerases described herein can be used in PCR applications including, but are not limited to, i) hot-start PCR which reduces non-specific amplification; ii) touch-down PCR which starts at high annealing temperature, then decreases annealing temperature in steps to reduce non-specific PCR product; iii) nested PCR which synthesizes more reliable product using an outer set of primers and an inner set of primers; iv) inverse PCR for amplification of regions flanking a known sequence. In this method, DNA is digested, the desired fragment is circularized by ligation, then PCR using primer complementary to the known sequence extending outwards; v) AP-PCR (arbitrary primed) /RAPD (random amplified polymorphic DNA). These methods create genomic fingerprints from species with little-known target sequences by amplifying using arbitrary oligonucleotides; vi) RT-PCR which uses RNA-directed DNA polymerase (e.g., reverse transcriptase) to synthesize cDNAs which is then used for PCR. This method is extremely sensitive for detecting the expression of a specific sequence in a tissue or cells. It may also

be use to quantify mRNA transcripts ; vii) RACE (rapid amplification of cDNA ends). This is used where information about DNA/protein sequence is limited. The method amplifies 3' or 5' ends of cDNAs generating fragments of cDNA with only one specific primer each (plus one adaptor primer). Overlapping RACE products can then be combined to produce full length cDNA; viii) DD-PCR (differential display PCR) which is used to identify differentially expressed genes in different tissues. A first step in DD-PCR involves RT-PCR, then amplification is performed using short, intentionally nonspecific primers; ix) Multiplex-PCR in which two or more unique targets of DNA sequences in the same specimen are amplified simultaneously. One DNA sequence can be use as control to verify the quality of PCR; x) Q/C-PCR (Quantitative comparative) which uses an internal control DNA sequence (but of different size) which compete with the target DNA (competitive PCR) for the same set of primers; xi) Recursive PCR which is used to synthesize genes. Oligonucleotides used in this method are complementary to stretches of a gene (>80 bases), alternately to the sense and to the antisense strands with ends overlapping (-20 bases); xii) Asymmetric PCR; xiii) In Situ PCR; xiv) Site-directed PCR Mutagenesis; xv) DOP-PCR that uses partially degenerate primers for whole-genome amplification; xvi) quantitative PCR using SYBR™ green or oligonucleotide probes to detect amplification; and xvii) error-prone PCR in which conditions are optimized to give an increased number of mutations in the PCR product.

[0081] It should be understood that this invention is not limited to any particular amplification system. As other systems are developed, those systems may benefit by practice of this invention.

Kits

[0082] The invention also contemplates kit formats which include a package unit having one or more containers containing modified DNA polymerases of the invention and compositions thereof. In some embodiments, the present invention provides kits further including containers of various reagents used for polynucleotide synthesis, including synthesis in PCR.

[0083] Inventive kits in accordance with the present invention may also contain one or more of the following items: polynucleotide precursors, primers, buffers, instructions, PCR

additives and controls. Kits may include containers of reagents mixed together in suitable proportions for performing the methods in accordance with the invention. Reagent containers preferably contain reagents in unit quantities that obviate measuring steps when performing the subject methods.

EXAMPLES

Example 1. Directed evolution experiments using Taq polymerase

[0007] To select mutated enzymes that would better be suited for recombinant DNA technologies, a directed evolution experiment is designed by simply mimicking the normal conditions under which the enzyme is usually used, or possibly under less than perfect conditions such as are expected in real-life applications. After conducting enough rounds of selection, an enzyme (or multiple enzymes) that is better suited for typical applications in recombinant DNA technologies should appear. Details of directed evolution experiments and exemplary advantages of associated with selected mutations are described in the co-pending application entitled "Modified DNA Polymerases" filed on even date.

[0008] In particular, we have performed directed evolution experiments using a type A DNA polymerase, Taq. Directed evolution experiments were conducted on Taq mutant libraries created by error-prone PCR.

[0009] Several rounds of selection were conducted. During the course of the ongoing selection, it is likely that many different mutations will confer different types of advantage, to different degrees, either alone or in combination. Typically, during the first rounds of selection, there are no obvious dominant clones, while the huge numbers of neutral or disadvantageous mutants are likely to be eliminated. Thereafter, a number of particular mutations typically appear in higher than expected numbers. These mutations are there because they have some advantages.

[0010] Typically, the selections are considered to have worked when the vast pool of mutants that are in the starting material have been eliminated and the pool is dominated by a remaining few types or families of mutants that have out-competed the other mutants and the

wild type. At this stage, it is not necessary to define exactly the nature of the improvement that the mutations confer. The fact that it was selected for is sufficient proof, especially if the same mutation becomes dominant in independently run selections.

[0088] Further selection results in the number of some of these mutations increasing in the pool, while others may be eliminated possibly because they have some advantages but they are not sufficient to compete with better-adapted clones. At the same time, some previously unnoticed mutants may appear. The late appearance of these mutants might be due to the fact that these specific mutations were low in number in the starting pool, or that the mutation required another (or more than one) mutation in the same clone for the advantage to manifest. If selections continue even further, eventually, a few clones will likely to dominate substantially. Typically, it is important to isolate clones before this final point if it is desirable to isolate a wide range of beneficial mutations.

[0089] In particular experiments, high processivity mutants were generated and screened for either (1) resistance to high-salt (KCl) in a PCR reaction and/or (2) resistance to high levels of SYBR Green I intercalating dye in a PCR reaction. Several rounds of selection were conducted on Taq. During the course of the ongoing selections, many different mutations were observed either alone or in combination at various positions. Clones that exhibited higher tolerance than wild-type to either of these pressures were selected and sequenced. Exemplary mutations and corresponding positions are shown in Table 2. Exemplary clones containing various mutations or combinations of mutations are shown in Table 3 (based on resistance to high-salt (KCl)) and Table 4 (based on resistance to high levels of SYBR Green I). The Enzymes containing one or more of these mutations retain the enzymatic activity. A general phenotype of these selected clones has higher specific activity than wild-type Taq and they are further characterized for a variety of phenotypes, as described in further Examples below.

Table 2. Mutations Observed in Taq Mutant Clones Selected for Resistance to High Salt or High Levels of SYBR™ Green

Position	Mutation	Position	Mutation
6	P6S	**360**	A360T
7	L7P	**364**	G364D
9	E9K	**364**	G364S
20	H20Q	**368**	P368L
26	T26M	**400**	E400K
27	F27S	**404**	E404G
30	L30P	**413**	F413L
39	E39K	**414**	A414T
41	V41A	**419**	R419H
50	S50N	**468**	A468T
53	K53N	**471**	E471G
56	K56Q	**507**	E507K
57	E57D	**515**	S515G
116	Y116C	**518**	V518A
151	D151N	**578**	D578N
171	K171R	**604**	W604R
203	T203I	**631**	V631A
209	E209G	**636**	R636H
209	E209K	**649**	V649A
225	K225R	**651**	R651H
238	D238N	**684**	I684V
245	L245M	**690**	Q690R
259	A259V	**717**	R717G
262	R262C	**730**	V730A
274	E274G	**732**	D732G
292	K292I	**742**	E742K
294	L294P	**744**	A744V
305	A305V	**797**	E797G
310	V310A	**804**	K804E
340	K340R	**814**	A814G

Table 3. Exemplary Taq Mutant Clones Selected for Resistance to High Salt

Clone Name	Mutations
A3E	T203I, D578N; E742K
G9S	T203I; G364S, D732G; E742K
D5S	P6S, T203I, K340R, E742K
D2	E9K, S50N, E209K, E507K
A5E	E39K; V41A; K53N, K56Q, E57D, V310A, P368L, A468T, E507K
B6S	F27S, K53N, K56Q, E57D, A259V, V310A, E507K, W604R
E2S	K53N, K56Q, E57D, D151N, E209G, A360T, E507K, Q690R
A3	K53N, K56Q, E57D, L245M, R262C, G364D, E507K, V518A
H10	K53N, K56Q, E57D, E507K, S515G
H1S	K53N, K56Q, E57D, R419H, E507K, S515G; V631A
F9E	K53N, K56Q, E57D, E507K, V649A
A5S	K53N, K56Q, E57D, E404G, E507K
C10E	H20Q, K53N, K56Q, E57D, K171R, K225R, K292I, A305V, E471G, E507K, A814G
F5S	L30P, K53N, K56Q, E57D, K171R, E274G, E507K
E7S	K53N, K56Q, E57D, E507K
G6S	T26M, K53N, K56Q, E57D, E507K, I684V
E1E	K53N, K56Q, E57D, F413L, E507K, R636H

Table 4. Exemplary Tag Mutant Clones Selected for Resistance to SYBR™ Green

Clone Name	Mutations
C7	E507K, A744V, E797G
E12	P6S, L7P, E507K, R651H, V730A, E797G
D9	P6S, E507K, E797G, K804E
F10	D238N, L294P, E400K, E507K
H7	D238N, L294P, E400K, A414T, E507K
A5	Y116C, D238N, L294P, E400K, A414T, E507K, R717G

Example 2. Types of selective advantage

[0090] There are a wide range of advantages that may have been selected for, some of which are listed and discussed below:

1) Expression efficiency:

[0091] The clones that express higher levels of the enzyme will have an advantage over those that express less. The specific activity of the mutated enzyme may not have been improved but the total activity will have. This characteristics is particularly valuable to a manufacture of enzymes because this will allow increased production levels and/or reduced production costs.

2) Solubility and folding robustness:

[0092] When solubility increases, the probability of inclusion bodies forming decreases. Therefore, in these clones, a higher proportion of useful, correctly folded enzyme product is expressed.

3) Thermostability:

[0093] It is well known that, during the thermocycling required for PCR, a certain fraction of the enzyme is inactivated due to the heating. An enzyme that is resistant to heat-inactivation will maintain activity longer. Therefore, less enzyme can be used and/or more cycles can be conducted.

4) Activity:

[0094] Mutants with increased enzymatic activity provide more efficient polymerization.

5) Processivity:

[0095] Mutants with increased processivity are able to synthesize long PCR products and synthesize sequences with complexed secondary structure. Mutant enzymes that can incorporate more nucleotides/extension step are likely to operate efficiently at lower concentrations.

6) Speed:

[0096] Mutants with increased elongation rate provide more efficient polymerization. Enzymes that are fast can also be used with shorter extension times. This is particularly valuable for a high-throughput system.

7) Concentration robustness:

[0097] It is known that PCR reactions may not be carried out appropriately if too much or too little enzyme is used. Under the selection conditions we used, a polymerase that can generate appropriate products whether it is supplied in excess or at low levels will have an advantage.

8) Resistance to salts, PCR additives and other inhibitors:

[0098] The selection was conducted in the presence of salts, PCR additives (e.g., intercalating dyes), and other impurities. The presence of salts may reduce the DNA binding affitnity of polymerases. The presence of impurities may interfere with formation of a desired PCR product. A polymerase that is resistant to salts and inhibitors and can synthesize desired products is advantageous and will be selected for. The characteristic is particularly suited for applications in which PCR is used in crude samples.

9) Fidelity:

[0099] All polymerases make mistakes during replication, either by incorporating the wrong dNTP or by stuttering which causes deletions and insertions. Such mistakes can eliminate functional genes during selection, so there is a pressure for mistakes not to be made. A polymerase with higher fidelity is advantageous and will be selected for.

10) Strand-displacement activity:

[0100] Secondary structure in the DNA due to intramolecular self annealing may inhibit DNA strand-elongation catalyzed by the polymerase. Similarly, partial re-annealing of the complementary DNA in addition to the primer will inhibit PCR. Any enzyme with improved strand-displacement activity will have an advantage in the selection.

11) Pyrophosphate tolerance:

[0101] Pyrophosphate is released during incorporation of nucleotides into the nascent strand by polymerases. Accumulation of pyrophosphate may lead to inhibition of the polymerase activity. Polymerases that were selected for in the Directed evolution example may have evolved to become less affected by pyrophosphate inhibition.

12) Unknown:

[0102] There many other factors involved in the process of PCR. Enzymes that are better adapted to PCR for any reason may be selected under our selection conditions.

[0103] Certain selected clones and mutations are further characterized for a variety of phenotypes. So far, we have conducted tests for a few different phenotypes: processivity, ability to synthesize large fragments, and tolerance to inhibitors. The tests to examine phenotypes are described in the following examples.

Example 3. Heparin binding assays

[0104] To test the processivity of the selected Taq mutants, we used heparin binding assays. Heparin is a member of the glycosaminoglycan family of carbohydrates (which includes the closely related molecule heparan sulfate) and consists of a variably sulfated repeating disaccharide unit. Heparin polymers form a helical structure and it is believed that DNA processing enzymes bind to heparin at the same contact points that bind double stranded DNA. Thus, DNA binding affinity can be measured based on heparin binding assays. Briefly, at physiological pH the sulphate groups are deprotonated. The negative charge and the helical structure mimic the structure and charge of DNA, enabling binding of DNA-binding proteins to heparin. DNA polymerases contain a number of positively charged amino acid residues that are involved in binding of the enzyme to DNA. This property can be utilized during purification of polymerases whereby the polymerase binds to heparin that is covalently coupled to agarose beads. The binding affinity of the polymerase is determined by the number and strength of binding interactions. The polymerase is eluted by increasing the amount of salt in the elution buffer. Ion-bonds between the polymerase and heparin will be disrupted by adding an increasing concentration of salt. The salt concentration at which the

enzyme elutes is, therefore, indicative of the binding affinity of the polymerase for heparin and DNA.

[0105] In particular, pellets of *E. coli* cells containing Taq mutants were lysed in 50 mM Tris-HCl pH 8.0, 150 mM NaCl (binding buffer). The lysates were incubated for 30 min at 75 °C to denature *E. coli* proteins, followed by centrifugation at 20 000 x g for 20 min at 20 °C. The supernatant was loaded onto a HiTrap™ Heparin column (GE Healthcare) and eluted on a 0.15 to 2 M NaCl gradient. The conductivity (mS/cm) at the elution peak was recorded as a measure of salt concentration of the eluate. A high conductivity indicates high affinity of the polymerase for heparin and DNA. The conductivity at the elution peak of Taq polymerase was 38.3 mS/cm (see Table 5). The conductivity for low affinity polymerase mutants was between below 38 mS/cm. The conductivity of certain high affinity polymerase mutants was between 46.7 and 54.4 mS/cm (see Table 5) .

[0106] The conductivity is proportional to the amount of salt in a solution. We empirically determined the correlation between salt concentration and conductivity. We used the binding buffer and elution buffer at various ratios (final concentrations of 200 to 700 mM NaCl) and measured the conductivity. We plotted the conductivity vs. NaCl concentration. Linear regression analysis revealed that the conductivity (Cd) can be expressed as Cd = 0.084 x Cs + 7.26 (R2 = 0.9995), where Cs is concentration of NaCl. From this we calculated that Taq polymerase eluted at around 370 mM NaCl, and the mutants eluted at between around 470 and 561 mM NaCl.

Table 5. Conductivity of Taq Clones

Clone:	Conductivity of main peak (mS/cm):
Wild-type Taq	38.3
A3E	48.2
G9S	46.7
D5S	46.8
D2	54.4
A5E	50.0
B6S	50.2
E2S	52.7
A3	48.9
H10	49.5
H1S	49.6
F9E	49.9
A5S	50.7
C10E	50.4
F5S	49.7
E7S	49.8
G6S	50.4
E1E	50.2
C7	47.4
E12	50.0
D9	49.8
F10	50.4
H7	51.6
A5	51.1

Example 4. Ability to generate long PCR fragments

[0107] Primer pairs were designed to generate either a 5kb, 8kb, or 10kb fragment from a lambda DNA template. Each of the high processivity enzymes, under limiting enzyme concentration, was tested for their ability to amplify each of the amplicon lengths.

[0108] Exemplary primers include forward primer L30350F: 5′-CCTGCTCTGCCGCTTCACGC-3′ (SEQ ID NO:28) and reverse primers as follows:

L-5R: 5'-CGAACGTCGCGCAGAGAAACAGG-3' (SEQ ID NO:29)

L-8R: 5′-GCCTCGTTGCGTTTGTTTGCACG-3′ (SEQ ID NO:30)

L-10R: 5′-GCACAGAAGCTATTATGCGTCCCCAGG-3′ (SEQ ID NO:31)

[0109] The reaction components for the assays are shown in Table 6. The cycling profile for these reactions is shown in Table 7.

Table 6. Exemplary Reaction Components

Reaction component	Concentration		In 20 µL
PCR water	-	-	16.185
10x Kapa Buffer A with loading dye	10	x	2.00
MgCl2 (supplement to 2 mM for increased dNTPs)	25	mM	0.40
dNTPs	10	mM each	0.80
Primer L30350-F	100	uM	0.10
Reverse primer	100	uM	0.10
Lambda DNA	300	ng/uL	0.015
Each DNA polymerase	20	ng/uL	0.40
TOTAL			**20.00**

Table 7. Exemplary Cycling Profile

Cycling profile:

Cycle	No.	Temp ('C)	Time	
Initial denaturation	1	95	2	min
Denaturation	**25**	95	30	sec
Annealing/Extension	**25**	72	6	min
Final elongation	1	72	6	min
HOLD	1	4	Indefinite	

[0110] Reaction products were run on an agarose gel and scored for either a presence or absence of a band at the appropriate fragment size. Exemplary results are shown in Table 8.

Table 8. Fragments Produced by Taq Clones

Clone Name:	5kb	8kb	10kb
Wild type Taq	no	no	no
A3E	yes	yes	no
G9S	no	no	no
D5S	no	no	no
D2	yes	yes	yes
A5E	yes	yes	no
B6S	yes	no	no
E2S	yes	yes	yes
A3	no	no	no
H10	yes	no	no
H1S	no	no	no
F9E	yes	no	no
A5S	yes	no	no
C10E	yes	yes	yes
F5S	yes	no	no
E7S	yes	no	no
G6S	yes	no	no
E1E	no	no	no
C7	no	no	no
E12	yes	yes	no
D9	yes	yes	no
F10	yes	yes	yes
H7	yes	yes	yes
A5	yes	yes	no

Example 5. Tolerance to high salt

[0111] High processivity Taq clones were tested for the ability to amplify a 2kb PCR amplicon in the presence of high salt. Reactions were performed in a buffer containing 10mM Tris-HCl (pH 8.4 @ 25°C) and either 150mM KCl or 150mM NaCl. Exemplary reaction components for assays with 150mM KCl and 150mM NaCl are shown in Tables 9 and 10, respectively. Exemplary cycling profile for these assays is shown in Table 11.

[0112] Exemplary primers include forward primer L30350F: 5′-CCTGCTCTGCCGCTTCACGC-3′ (SEQ ID NO:28) and reverse primer L-2R: 5′-CCATGATTCAGTGTGCCCGTCTGG-3′ (SEQ ID NO:32).

Table 9. Exemplary Reaction Components for Assays with 150 mM KCl

150 mM KCl		Reaction volume =	**25**
Reaction component	**Concentration**		**In 25 uL**
PCR water	-	-	15.967
100x Tris-HCl, pH8.4	100	x	0.250
MgCl2 (supplement to 1.5 mM)	25	mM	1.50
dNTPs	10	mM each	0.50
KCl	2500	mM	1.50
Primer L30350-F	**100**	uM	0.125
Primer L-2R	**100**	uM	0.125
Lambda DNA	300	ng/uL	0.033
Taq DNA polymerase	20	ng/uL	5.00
TOTAL			**25.00**

Table 10. Exemplary Reaction Components for Assays with 150 mM NaCl

150 mM NaCl		Reaction volume =	**25**
Reaction component	**Concentration**		**In 25 uL**
PCR water	-	-	15.967
100x Tris-HCl, pH8.4	100	x	0.25
MgCl2 (supplement to 1.5 mM)	25	mM	1.50
dNTPs	10	mM each	0.50
NaCl	2500	mM	1.50
Primer L30350-F	**100**	uM	0.125
Primer L-2R	**100**	uM	0.125
Lambda DNA	300	ng/uL	0.033
Taq DNA polymerase	20	ng/uL	5.00
TOTAL			**25.00**

Table 11. Exemplary Cycling Profile (High Salt Conditions)

Cycling profile:				
Cycle	**No.**	**Temp ('C)**	**Time**	
Initial denaturation	1	95	2	min
Denaturation	**35**	95	30	sec
Annealing/Extension	**35**	72	2	min
Final elongation	1	72	2	min
HOLD	1	4	Indefinite	

[0113] Reaction products were run on an agarose gel and scored for either a presence or absence of a band at the appropriate fragment size. Exemplary results are shown in Table 12.

Table 12. Fragments Produced by Taq Clones (High Salt Conditions)

Clone Name:	150mM KCl	150mM NaCl
Wild type Taq	no	no
A3E	yes	no
G9S	yes	no
D5S	yes	yes
D2	yes	yes
A5E	yes	yes
B6S	yes	yes
E2S	yes	yes
A3	yes	yes
H10	no	no
H1S	yes	no
F9E	yes	yes
A5S	yes	yes
C10E	yes	yes
F5S	yes	yes
E7S	yes	yes
G6S	yes	yes
E1E	yes	yes
C7	no	no
E12	yes	yes
D9	yes	yes
F10	yes	yes
H7	yes	yes
A5	yes	yes

Example 6. Resistance to Phenol

[0114] High processivity Taq clones were tested for the ability to amplify a 2kb PCR amplicon in the presence of 1% phenol. Reactions were performed in a buffer containing

10mM Tris-HCl (pH 8.4 @ 25°C) and 1% phenol. The reaction components for these assays are shown in Table 13. The cycling profile for these assays is shown in Table 14.

Table 13. Exemplary Reaction Components (High Phenol Conditions)

1% phenol		Reaction volume =	**25**
Reaction component	**Concentration**		**In 25 uL**
PCR water	-	-	16.467
10x Kapa Buffer A with loading dye + MgCl2	10	x	2.50
MgCl2	25	mM	0.00
dNTPs	10	mM each	0.50
Phenol	100	%	0.25
Primer L30350-F	**100**	uM	0.125
Primer L-2R	**100**	uM	0.125
Lambda DNA	300	ng/uL	0.033
Taq DNA polymerase	20	ng/uL	5.00
TOTAL			**25.00**

Table 14. Exemplary Cycling Profile (High Phenol Conditions)

Cycling profile:				
Cycle	**No.**	**Temp ('C)**	**Time**	
Initial denaturation	1	95	2	min
Denaturation	**35**	95	30	sec
Annealing/Extension	**35**	72	2	min
Final elongation	1	72	2	min
HOLD	1	4	Indefinite	

[0115] Reaction products were run on an agarose gel and scored for either a presence or absence of a band at the appropriate fragment size. Exemplary results are shown in Table 15.

Table 15. Fragments Produced by Taq Clones (High Phenol Conditions)

Clone Name:	1% phenol
Wild type Taq	no
A3E	yes
G9S	no
D5S	yes
D2	yes
A5E	yes
B6S	no
E2S	no
A3	yes
H10	yes
H1S	yes
F9E	yes
A5S	yes
C10E	yes
F5S	yes
E7S	yes
G6S	yes
E1E	no
C7	yes
E12	no
D9	no
F10	yes
H7	yes
A5	no

Table 16. Sequences

Amino acid sequences of Taq and modified Taq polymerases

>Wild-type (SEQ ID NO:1)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>A3E (SEQ ID NO:2)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKIARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSNPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVRKAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>G9S (SEQ ID NO:3)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKIARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALRESLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPGLEARVKSVRKAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>D5S (SEQ ID NO:4)
MRGMLSLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKIARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYRALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVRKAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>D2 (SEQ ID NO:5)
MRGMLPLFKPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKNLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLKEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>A5E (SEQ ID NO:6)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGKPAQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFALSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLLPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEITRLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>B6S (SEQ ID NO:7)
MRGMLPLFEPKGRVLLVDGHHLAYRTSHALKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFVKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFALSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGRLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>E2S (SEQ ID NO:8)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
NRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLGEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLTLREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEARAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>A3 (SEQ ID NO:9)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDMAKVRTDLPLEVDFAKRCEPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREDLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAALEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>H10 (SEQ ID NO:10)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTGAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>H1S (SEQ ID NO:11)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGHLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTGAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRAFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>F9E (SEQ ID NO:12)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGAPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>A5S (SEQ ID NO:13)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGGRAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>C10E (SEQ ID NO:14)
MRGMLPLFEPKGRVLLVDGQHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWERYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLR
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPIALEEAPWP
PPEGVFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLGAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLGVPLEVEVGIGE
DWLSAKE*

>F5S (SEQ ID NO:15)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHAPKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWERYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLGRLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>E7S (SEQ ID NO:16)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>G6S (SEQ ID NO:17)
MRGMLPLFEPKGRVLLVDGHHLAYRMFHALKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAVPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>E1E (SEQ ID NO:18)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLNALQDDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLLANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGHDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>C7 (SEQ ID NO:19)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAVERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAGAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>E12 (SEQ ID NO:20)
MRGMLSPFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPHEAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYAPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAGAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>D9 (SEQ ID NO:21)
MRGMLSLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAGAVARLAEEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>F10 (SEQ ID NO:22)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDNLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKAPEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTKEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>H7 (SEQ ID NO:23)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDNLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKAPEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTKEAGERAALSERLFTNLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

>A5 (SEQ ID NO:24)
MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGCEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDNLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKAPEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTKEAGERAALSERLFTNLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTKKTGKRSTSAAVLEALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRGGYVETLFGRRRYVPDLEARVKSVREAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE*

EQUIVALENTS

[0116] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. The scope of the present invention is not intended to be limited to the above Description, but rather is as set forth in the appended claims. The articles "a", "an", and "the" as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to include the plural referents. Claims or descriptions that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention also includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process. Furthermore, it is to be understood that the invention encompasses variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, etc., from one or more of the claims is introduced into another claim dependent on the same base claim (or, as relevant, any other claim) unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise. Where elements are presented as lists, e.g., in Markush group or similar format, it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It should it be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements, features, etc., certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements, features, etc. For purposes of simplicity those embodiments have not in every case been specifically set forth herein. It should also be understood that any embodiment of the invention, e.g., any embodiment found within the prior art, can be explicitly excluded from the claims, regardless of whether the specific exclusion is recited in the specification.

[0117] It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one act, the order of the acts of the method is not necessarily limited to the order in which the acts of the method are recited, but the invention includes embodiments in which the order is so limited. Furthermore, where the

claims recite a composition, the invention encompasses methods of using the composition and methods of making the composition. Where the claims recite a composition, it should be understood that the invention encompasses methods of using the composition and methods of making the composition.

CLAIMS:

1. A modified Taq DNA polymerase whose amino acid sequence is at least 95% identical to that of wild type Taq DNA polymerase set out in SEQ ID NO: 1, but differs from that of wild type Taq DNA polymerase in that it includes a lysine at a position that corresponds to E507 of SEQ ID NO:1, which polymerase shows increased salt-resistance and improved ability to catalyze a DNA-template-dependent DNA polymerization activity relative to that of the wild type Taq DNA polymerase of SEQ ID NO:1.

2. A modified Taq DNA polymerase whose amino acid sequence is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:5 (D2), SEQ ID NO:6 (A5E), SEQ ID NO:7 (B6S), SEQ ID NO:8 (E2S), SEQ ID NO:9 (A3), SEQ ID NO : 10 (H10), SEQ ID NO : 11 (H1S), SEQ ID NO : 12 (F9E), SEQ ID NO : 13 (A5S), SEQ ID NO: 14 (C10E), SEQ ID NO: 15(F5S), SEQ ID NO: 16 (E7S), SEQ ID NO: 17 (G6S), SEQ ID NO: 18 (E1E), SEQ ID NO: 19 (C7), SEQ IDNO:20 (E12), SEQ ID NO:21 (D9), SEQ ID NO:22 (F10), SEQ ID NO:23 (H7), and SEQ ID NO:24 (A5), and is different from that of wild type Taq DNA polymerase as set forth in SEQ ID NO:1, which polymerase is characterized by: (i) presence of lysine at a position corresponding to E507 of SEQ ID NO:1 and (ii) improved ability to catalyze a DNA-template-dependent DNA polymerization activity relative to that of the wild type Taq DNA polymerase of SEQ ID NO:1.

3. A modified Taq DNA polymerase having an amino acid sequence that is at least 95% identical to a sequence selected from the group consisting of SEQ ID NO:5 (D2), SEQ ID NO:6 (A5E), SEQ ID NO:7 (B6S), SEQ ID NO:8 (E2S), SEQ ID NO:9 (A3), SEQ ID NO:10 (H10), SEQ ID NO:11 (H1S), SEQ ID NO:12 (F9E), SEQ ID NO:13 (A5S), SEQ ID NO:14 (C10E), SEQ ID NO:15 (F5S), SEQ ID NO:16 (E7S), SEQ ID NO:17 (G6S), SEQ ID NO:18 (E1E), SEQ ID NO:20 (E12), SEQ ID NO:21 (D9), SEQ ID NO:22 (F10), SEQ ID NO:23 (H7), SEQ ID NO:24 (A5) and is different from that of wild type Taq DNA polymerase as set forth in SEQ ID NO:1, which polymerase (i) has a lysine at a position corresponding to E507 of SEQ ID NO:1 and (ii) shows increased salt tolerance and improved ability to catalyze

a DNA-template-dependent DNA polymerization activity relative to that of the wild type Taq DNA polymerase of SEQ ID NO:1.

4. The modified Taq DNA polymerase of claim 1, wherein the amino acid sequence is at least 96% identical to that of the wild-type Taq DNA polymerase set forth in SEQ ID NO: 1.

5. The modified Taq DNA polymerase of claim 1, wherein the amino acid sequence is at least 97% identical to that of the wild-type Taq DNA polymerase set forth in SEQ ID NO: 1.

6. The modified Taq DNA polymerase of claim 1, wherein the amino acid sequence is at least 98% identical to that of the wild-type Taq DNA polymerase set forth in SEQ ID NO: 1.

7. The modified Taq DNA polymerase of claim 2 or 3, wherein the percent identity is at least 96%.

8. The modified Taq DNA polymerase of claim 2 or 3, wherein the percent identity is at least 97%.

9. The modified Taq DNA polymerase of claim 2 or 3, wherein the percent identity is at least 98%.

10. A modified Taq DNA polymerase having the amino acid sequence of SEQ ID NO: 16 (E7S).

11. The modified Taq DNA polymerase of any one of claims 1 to 10, wherein the polymerase has improved ability relative to that of SEQ ID NO.: 1, when used in a PCR reaction, to amplify DNA fragments of 2 kilobases or more in length.

12. A kit comprising:

i) a modified Taq DNA polymerase according to any one of claims 1 to 11; and
ii) at least one buffer suitable for use in a polymerase chain reaction.

13. A nucleic acid molecule whose nucleotide sequence encodes a modified Taq DNA polymerase according to any one of claims 1 to 11.

14. A vector comprising the nucleic acid molecule of claim 13.

15. A cell comprising the nucleic acid molecule of claim 13.

16. A cell comprising the vector of claim 14.

17. Use of a modified Taq DNA polymerase according to any one of claims 1-11 for amplifying a DNA fragment.

18. A method of performing a DNA polymerization reaction, the method comprising:

a) providing a mixture wherein the mixture comprises:

i) a template nucleic acid;

ii) at least one primer;

iii) nucleotides; and

iv) a modified Taq DNA polymerase whose amino acid sequence is at least 95% identical to that of wild type Taq DNA polymerase set forth in SEQ ID NO:l and includes a lysine at a position corresponding to E507 of SEQ ID NO: 1, wherein the modified Taq DNA polymerase has improved ability to catalyze a DNA-template-dependent DNA polymerization activity, relative to the Taq of SEQ ID NO:1; and

b) incubating the mixture under conditions that:

i) permit hybridization of the at least one primer to the template nucleic acid; and

ii) permit extension of the at least one primer by polymerization of the nucleotides, which polymerization is catalyzed by the modified Taq DNA polymerase.

19. The method of claim 18, wherein the template nucleic acid is DNA.

20. The method of claim 18 or 19, wherein the conditions comprise presence of one or more of a nucleic acid intercalating agent, phenol, high salt concentration, an inhibitor of processivity and an inhibitor of enzyme activity.

21. The method of any one of claims 18 to 20, wherein the product produced is longer than 2 kilobases.

22. The method of any one of claims 18 to 20, wherein the product produced is longer than 5 kilobases.

23. The method of any one of claims 18 to 20, wherein the product produced is longer than 8 kilobases.

24. The method of any one of claims 18 to 20, wherein the product produced is longer than 10 kilobases.

25. The method of any one of claims 18 to 24, wherein the method further comprises: labeling of DNA by nick translation, second-strand cDNA synthesis in cDNA cloning, DNA sequencing, whole-genome amplification, and amplifying, detecting, and/or cloning nucleic acid sequences using polymerase chain reaction.

26. In a method of performing a polymerase chain reaction (PCR) using a modified Taq DNA polymerase, the improvement comprising utilizing a modified Taq DNA polymerase whose amino acid sequence is at least 95% identical to that of wild type Taq DNA polymerase set forth in SEQ ID NO:1 and includes a lysine at a position corresponding to E507 of SEQ ID NO: 1.

27. A method of providing an improvement in performance in a polymerase chain reaction (PCR) relative to that of wild type Taq DNA polymerase of SEQ ID NO:1, which improvement in performance is selected from one or more of enzyme activity, processivity, resistance to nucleic acid intercalating dyes, and salt-resistance, the method comprising:

using a modified Taq DNA polymerase whose sequence is at least 95% identical to that of wild type Taq DNA polymerase set forth in SEQ ID NO:1 and includes a lysine at a position corresponding to E507 of SEQ ID NO: 1 to perform the PCR, wherein the modified Taq DNA polymerase has improved ability to catalyze a DNA-template-dependent DNA polymerization activity, relative to the Taq of SEQ ID NO:1.

28. A kit comprising:

a modified Taq DNA polymerase whose sequence is at least 95% identical to that of wild type Taq DNA polymerase of SEQ ID NO:1 and includes a substitution of lysine at a position corresponding to E507 of SEQ ID NO: 1, wherein the modified Taq DNA polymerase has improved ability to catalyze a DNA-template-dependent DNA polymerization activity, relative to the Taq of SEQ ID NO:1; and

at least one buffer suitable for use in a PCR reaction.

29. A nucleic acid molecule whose nucleotide sequence encodes a modified Taq DNA polymerase whose amino acid sequence has at least 95% overall identity with that of wild type Taq of SEQ ID NO:1 but differs from that of wild type Taq DNA polymerase in that:

a) it includes a lysine at a position corresponding to E507 of wild type Taq of SEQ ID NO: 1; and

b) it is characterized by ability to catalyze a DNA-template-dependent DNA polymerization activity that is improved relative to that of the wild type Taq DNA polymerase.

30. A vector comprising the nucleic acid molecule of claim 29.

31. A cell comprising the nucleic acid molecule of claim 29.

32. A cell comprising the vector of claim 30.

33. A modified Taq DNA polymerase whose amino acid sequence is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:5 (D2), SEQ ID NO:6 (A5E), SEQ ID NO:7 (B6S), SEQ ID NO:8 (E2S), SEQ ID NO:9 (A3), SEQ ID NO : 10 (H10), SEQ ID NO : 11 (H1S), SEQ ID NO : 12 (F9E), SEQ ID NO : 13 (A5S), SEQ ID NO: 14 (C10E), SEQ ID NO: 15(F5S), SEQ ID NO: 16 (E7S), SEQ ID NO: 17 (G6S), SEQ ID NO: 18 (E1E), SEQ ID NO: 19 (C7), SEQ IDNO:20 (E12), SEQ ID NO:21 (D9), SEQ ID NO:22 (F10), SEQ ID NO:23 (H7), and SEQ ID NO:24 (A5), and is different from that of wild type Taq DNA polymerase as set forth in SEQ ID NO:1, which polymerase (i) has a lysine at a position corresponding to E507 of SEQ ID NO: 1, (ii) is capable of catalyzing a DNA-template-dependent DNA polymerization activity; and (iii) which displays increased heparin binding affinity relative to the wild type Taq DNA polymerase of SEQ ID NO: 1 in that it shows improved conductivity when eluted off a column comprising heparin.

34. The modified Taq DNA polymerase of claim 33, wherein the percent identity is at least 96%.

35. The modified Taq DNA polymerase of claim 33, wherein the percent identity is at least 97%.

36. The modified Taq DNA polymerase of claim 33, wherein the percent identity is at least 98%.

37. A kit comprising:

a modified Taq DNA polymerase as set out in claim any one of claims 33 to 36; and

at least one buffer suitable for use in a PCR reaction.

38. A Taq DNA polymerase that:

(a) is a mutant of a wild type Taq polymerase of SEQ ID NO:1 in that its amino acid sequence differs from that of SEQ ID NO:1;

(b) has an amino acid sequence that is at least 95% identical to that of a clone selected from the group consisting of: SEQ ID NO:5 (D2), SEQ ID NO:6 (A5E), SEQ ID NO:7 (B6S), SEQ ID NO:8 (E2S), SEQ ID NO:10 (H10), SEQ ID NO:12 (F9E), SEQ ID NO:13 (A5S), SEQ ID NO:14 (C10E), SEQ ID NO:15 (F5S), SEQ ID NO:16 (E7S), SEQ ID NO:17 (G6S), SEQ ID NO:20 (E12), SEQ ID NO:21 (D9), SEQ ID NO:22 (F10), SEQ ID NO:23 (H7), and SEQ ID NO:24 (A5);

(c) has a lysine at a position corresponding to E507 of SEQ ID NO: 1; and

(d) displays high processivity relative to the wild type Taq DNA polymerase of SEQ ID NO: 1 in that it shows improved ability to amplify DNA fragments of at least 5 kilobases in a PCR reaction.

39. The Taq DNA polymerase of claim 38 , wherein the polymerase displays high processivity relative to the wild type DNA polymerase in that it shows improved ability to amplify long DNA fragments of at least 8 kilobases.

40. The Taq DNA polymerase of claim 38, wherein the polymerase displays high processivity relative to the wild type DNA polymerase in that it shows improved ability to amplify long DNA fragments of at least 10 kilobases.

41. The modified Taq DNA polymerase of any one of claims 38 to 40, wherein the percent identity is at least 96%.

42. The modified Taq DNA polymerase of any one of claims 39 to 40, wherein the percent identity is at least 97%.

43. The modified Taq DNA polymerase of any one of claims 39 to 40, wherein the percent identity is at least 98%.

1/2

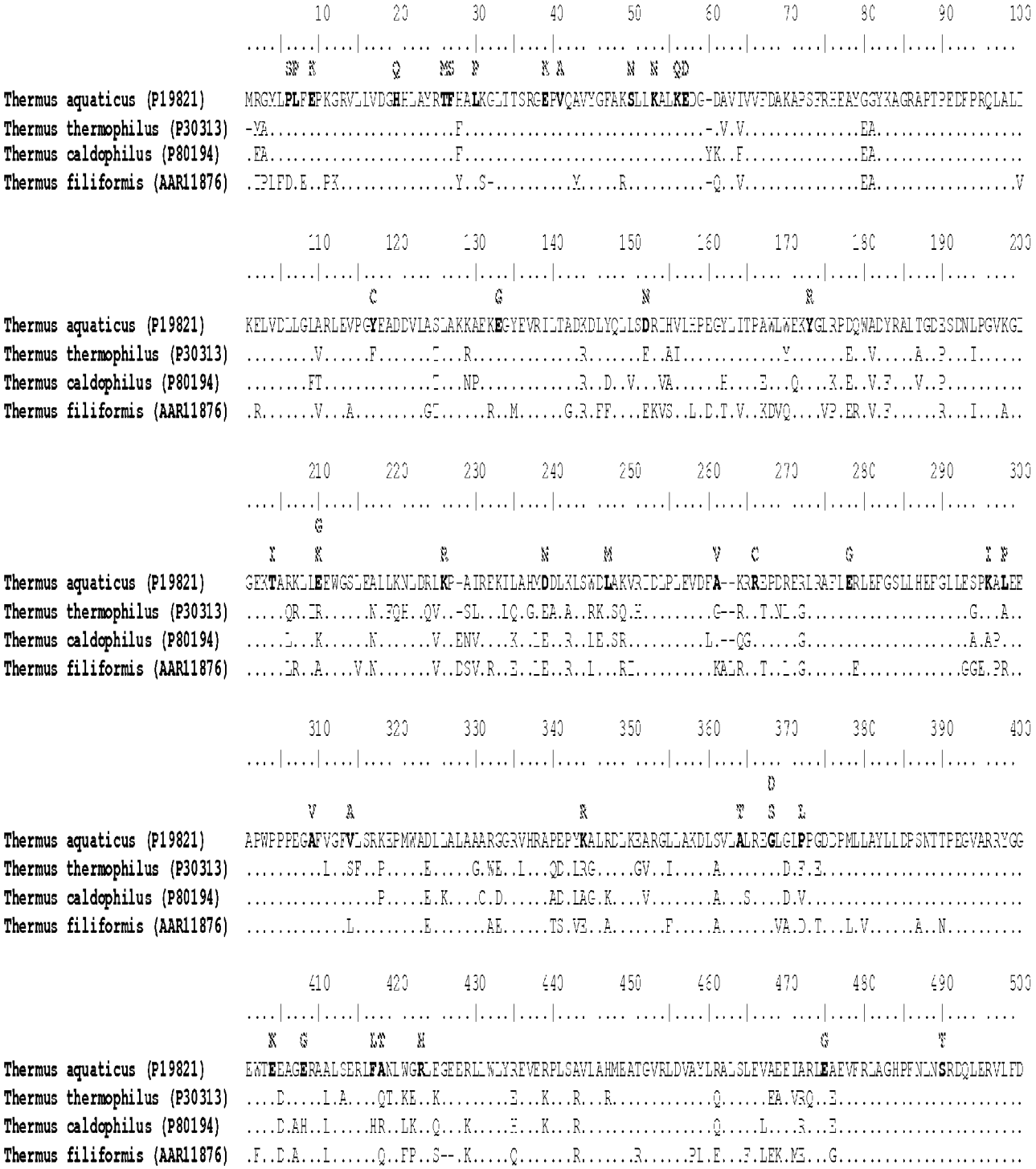
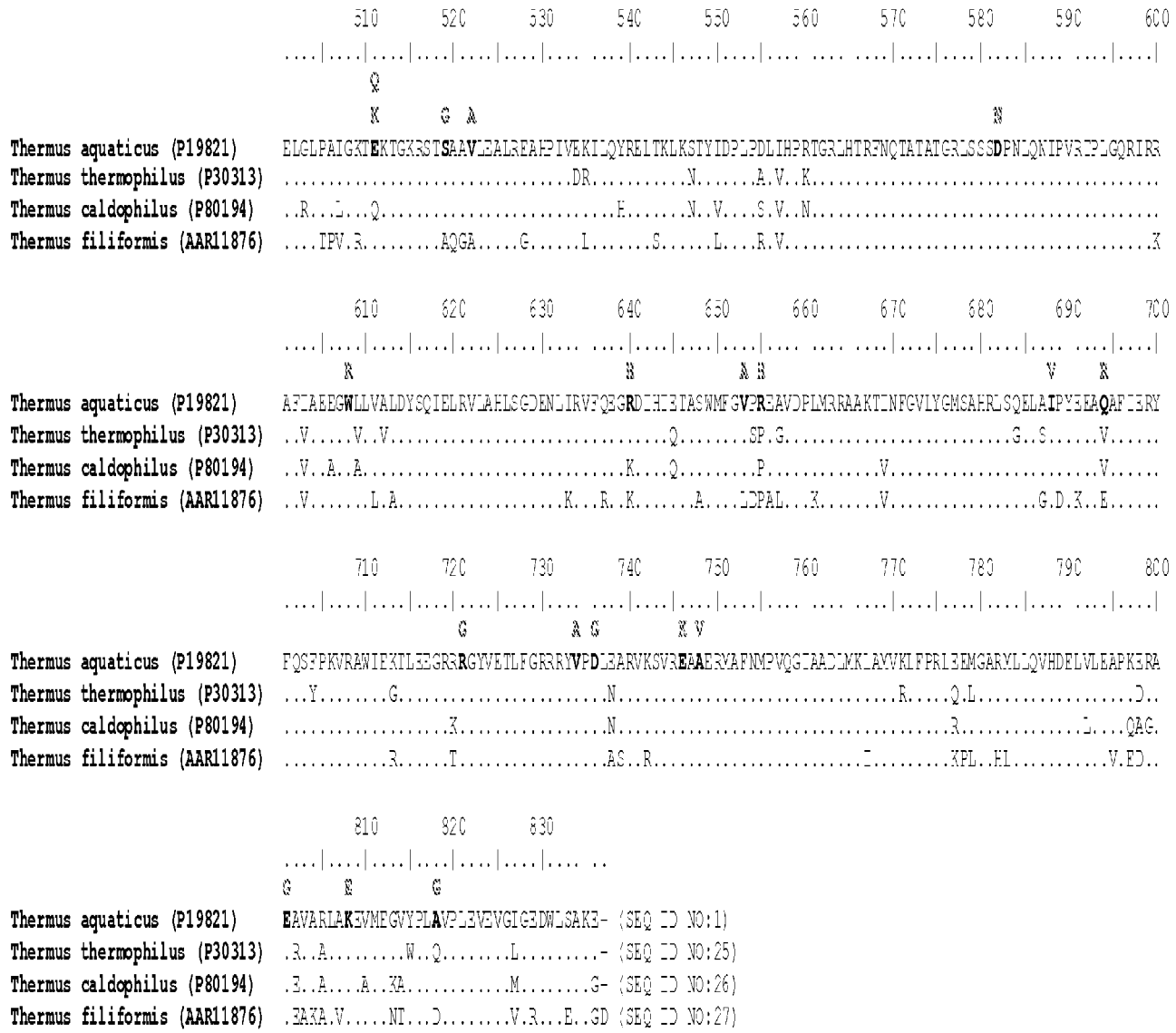

Thermus aquaticus (P19821)
Thermus thermophilus (P30313)
Thermus caldophilus (P80194)
Thermus filiformis (AAR11876)

FIG. 1



CONTINUATION OF FIG. 1