

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
22 December 2011 (22.12.2011)

PCT

(10) International Publication Number  
**WO 2011/157438 A1**

- (51) International Patent Classification:  
C12N 9/12 (2006.01) C12N 15/10 (2006.01)  
C12N 15/54 (2006.01)
- (21) International Application Number:  
PCT/EP2011/003004
- (22) International Filing Date:  
17 June 2011 (17.06.2011)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
61/356,287 18 June 2010 (18.06.2010) US
- (71) Applicant (for DE only): **ROCHE DIAGNOSTICS GMBH** [DE/DE]; Sandhofer Strasse 116, 68305 Mannheim (DE).
- (71) Applicant (for all designated States except DE, US): **F. HOFFMANN-LA ROCHE AG** [CH/CH]; Grenzacherstrasse 124, CH-4070 Basel (CH).

(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, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

- (72) Inventors: **REICHERT, Fred**; 675 Haas Avenue, San Leandro, CA 94577-2929 (US). **BAUER, Keith**; 62 W. Crescent Drive, San Rafael, CA 94901 (US). **MYERS, Thomas, W.**; 600 Loyola Drive, Los Altos, CA 94204 (US).
- (74) Common Representative: **ROCHE DIAGNOSTICS GMBH**; Patent Department, 68298 Mannheim (DE).

**Declarations under Rule 4.17:**

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

**Published:**

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

(54) Title: DNA POLYMERASES WITH INCREASED 3'-MISMATCH DISCRIMINATION

**Figure 1**

	*
Z05	TGRLSSSDPNLQNI PI R T P L GQRIRRAFVAE (SEQ ID NO:12)
Taq	TGRLSSSDPNLQNI PV R T P L GQRIRRAFIAE (SEQ ID NO:13)
Tfi	TGRLSSSDPNLQNI PV R T P L GQRIRKAFIAE (SEQ ID NO:14)
Tfl	TGRLSSSDPNLQNI PV R T P L GQRIRRAFVAE (SEQ ID NO:15)
Sps17	TGRLSSSDPNLQNI PV R T P L GQRIRKAFIAE (SEQ ID NO:16)
Tth	TGRLSSSDPNLQNI PV R T P L GQRIRRAFVAE (SEQ ID NO:17)
Tca	TGRLSSSDPNLQNI PV R T P L GQRIRRAFVAE (SEQ ID NO:18)
Tma	TGRLSSSDPNLQNL PT K S E E GKEIRKAIVPQ (SEQ ID NO:19)
Tne	TGRLSSSDPNLQNL PT K S E E GKEIRKAIVPQ (SEQ ID NO:20)
Taf	TGRLSSSDPNLQNL PT R S E E GKEIRKAVRPQ (SEQ ID NO:21)
Dra	TGRLSSSLNPNLQNI PI R S E L GREIRKGFIAE (SEQ ID NO:23)
Bst	TGRLSSVEPNLQNI PI R L E E GRKIRQAFVPS (SEQ ID NO:24)
Bca	TGRLSSTEPNLQNI PI R L E E GRKIRQAFVPA (SEQ ID NO:25)
	-----PNLQNX <sub>1</sub> PX <sub>2</sub> X <sub>3</sub> X <sub>4</sub> X <sub>5</sub> X <sub>6</sub> C----- (SEQ ID NO:26)

(57) Abstract: Disclosed are mutant DNA polymerases having increased 3'-mismatch discrimination relative to a corresponding, unmodified polymerase. The mutant polymerases are useful in a variety of disclosed primer extension methods. Also disclosed are related compositions, including recombinant nucleic acids, vectors, and host cells, which are useful, e.g., for production of the mutant DNA polymerases.

WO 2011/157438 A1

## DNA POLYMERASES WITH INCREASED 3'-MISMATCH DISCRIMINATION

### FIELD OF THE INVENTION

5 The present invention provides DNA polymerases with increased 3'-mismatch discrimination and their use in various applications, including nucleic acid polynucleotide extension and amplification.

### BACKGROUND OF THE INVENTION

10 DNA polymerases are responsible for the replication and maintenance of the genome, a role that is central to accurately transmitting genetic information from generation to generation. DNA polymerases function in cells as the enzymes responsible for the synthesis of DNA. They polymerize deoxyribonucleoside triphosphates in the presence of a metal activator, such as  $Mg^{2+}$ , in an order dictated by the DNA template or polynucleotide template that is copied. *In vivo*, DNA polymerases participate in a spectrum of DNA synthetic processes including DNA  
15 replication, DNA repair, recombination, and gene amplification. During each DNA synthetic process, the DNA template is copied once or at most a few times to produce identical replicas. In contrast, *in vitro*, DNA replication can be repeated many times such as, for example, during polymerase chain reaction (*see, e.g.*, U.S. Patent No. 4,683,202).

20 In the initial studies with polymerase chain reaction (PCR), the DNA polymerase was added at the start of each round of DNA replication (*see* U.S. Patent No. 4,683,202, *supra*). Subsequently, it was determined that thermostable DNA polymerases could be obtained from bacteria that grow at elevated temperatures, and that these enzymes need to be added only once (*see* U.S. Patent No. 4,889,818 and U.S. Patent No. 4,965,188). At the elevated temperatures used during PCR, these enzymes are not irreversibly inactivated. As a result, one can carry out repetitive cycles of  
25 polymerase chain reactions without adding fresh enzymes at the start of each synthetic addition process. DNA polymerases, particularly thermostable polymerases, are the key to a large number of techniques in recombinant DNA studies and in medical diagnosis of disease. For diagnostic applications in particular, a target nucleic acid sequence may be only a small portion of the DNA

or RNA in question, so it may be difficult to detect the presence of a target nucleic acid sequence without amplification.

The overall folding pattern of DNA polymerases resembles the human right hand and contains three distinct subdomains of palm, fingers, and thumb. (See Beese *et al.*, *Science* 260:352-355, 1993); Patel *et al.*, *Biochemistry* 34:5351-5363, 1995). While the structure of the fingers and thumb subdomains vary greatly between polymerases that differ in size and in cellular functions, the catalytic palm subdomains are all superimposable. For example, motif A, which interacts with the incoming dNTP and stabilizes the transition state during chemical catalysis, is superimposable with a mean deviation of about one Å amongst mammalian pol α and prokaryotic pol I family DNA polymerases (Wang *et al.*, *Cell* 89:1087-1099, 1997). Motif A begins structurally at an antiparallel β-strand containing predominantly hydrophobic residues and continues to an α-helix. The primary amino acid sequence of DNA polymerase active sites is exceptionally conserved. In the case of motif A, for example, the sequence DYSQIELR (SEQ ID NO:28) is retained in polymerases from organisms separated by many millions years of evolution, including, *e.g.*, *Thermus aquaticus*, *Chlamydia trachomatis*, and *Escherichia coli*.

In addition to being well-conserved, the active site of DNA polymerases has also been shown to be relatively mutable, capable of accommodating certain amino acid substitutions without reducing DNA polymerase activity significantly. (See, *e.g.*, U.S. Patent No. 6,602,695) Such mutant DNA polymerases can offer various selective advantages in, *e.g.*, diagnostic and research applications comprising nucleic acid synthesis reactions. Thus, there is a need in the art for identification of amino acid positions amenable to mutation to yield improved polymerase activities. The present invention, as set forth herein, meets these and other needs.

### BRIEF SUMMARY OF THE INVENTION

Provided herein are DNA polymerases having increased 3'-mismatch discrimination relative to a corresponding, unmodified control polymerase, and methods of making and using such DNA polymerases. In some embodiments, the polymerase is a thermostable DNA polymerase. In some embodiments, the DNA polymerase is a thermoactive DNA polymerase. In some embodiments, the DNA polymerase is derived from a *Thermus* species. In some embodiments, the DNA

polymerase is derived from a *Thermotoga* species. In some embodiments, the amino acid of the DNA polymerase corresponding to position 589 of SEQ ID NO:1 is any amino acid other than R or K, and the control DNA polymerase has the same amino acid sequence as the DNA polymerase except that the amino acid of the control DNA polymerase corresponding to position 589 of SEQ ID NO:1 is R or K. For example, in some embodiments, the amino acid at the position corresponding to position 589 of SEQ ID NO:1 is selected from G, A, V, L, I, M, F, W, P, S, T, C, Y, N, Q, D, E or H. In some embodiments, the amino acid at the position corresponding to position 589 of SEQ ID NO:1 is an amino acid having a polar, uncharged side-chain (e.g., N, Q, H, S, T, or Y). In some embodiments, the amino acid at the position corresponding to position 589 of SEQ ID NO:1 is an amino acid having a nonpolar, uncharged side-chain (e.g., G, A, L, M, W, P, F, C, V, or I). In some embodiments, the amino acid at the position corresponding to position 589 of SEQ ID NO:1 is selected from H, L, or S.

In some embodiments, the DNA polymerase having increased 3'-mismatch discrimination comprises a motif in the polymerase domain comprising

15 P-N-L-Q-N-X<sub>1</sub>-P-X<sub>2</sub>-X<sub>3</sub>-X<sub>4</sub>-X<sub>5</sub>-X<sub>6</sub>-G, wherein:  
 X<sub>1</sub> is I or L;  
 X<sub>2</sub> is I, V or T;  
 X<sub>3</sub> is any amino acid other than R or K;  
 X<sub>4</sub> is T, S or L;  
 20 X<sub>5</sub> is P or E; and  
 X<sub>6</sub> is L or E (SEQ ID NO:8).

In some embodiments, X<sub>3</sub> is selected from G, A, L, M, W, P, S, T, F, Y, C, N, Q, D, E, V, I or H (SEQ ID NO:42).

In some embodiments, the DNA polymerase having increased 3'-mismatch discrimination comprises a motif in the polymerase domain comprising

25 P-N-L-Q-N-X<sub>1</sub>-P-X<sub>2</sub>-X<sub>3</sub>-X<sub>4</sub>-X<sub>5</sub>-X<sub>6</sub>-G, wherein:  
 X<sub>1</sub> is I or L;  
 X<sub>2</sub> is I, V or T;  
 X<sub>3</sub> is any amino acid other than R or K;  
 30 X<sub>4</sub> is T or S;  
 X<sub>5</sub> is P or E; and

X<sub>6</sub> is L or E (SEQ ID NO:9).

In some embodiments, the DNA polymerase having increased 3'-mismatch discrimination comprises a motif in the polymerase domain comprising

P-N-L-Q-N-I-P-X<sub>2</sub>-X<sub>3</sub>-T-P-L-G, wherein:

5 X<sub>2</sub> is I or V; and

X<sub>3</sub> is any amino acid other than R (SEQ ID NO:10).

In some embodiments, X<sub>3</sub> is an amino acid having a polar, uncharged side-chain (e.g., N, Q, H, S, T, or Y) or nonpolar, uncharged side-chain (i.e., G, A, L, M, W, P, F, C, V, or I).

In some embodiments, X<sub>3</sub> is H, L or S (SEQ ID NO:11).

10 In some embodiments, the amino acid of the DNA polymerase corresponding to position 580 of SEQ ID NO:1 is any amino acid other than D or E. In some embodiments, the amino acid of the DNA polymerase corresponding to position 580 of SEQ ID NO:1 is any amino acid other than D. In some embodiments, the amino acid of the DNA polymerase corresponding to position 580 of SEQ ID NO:1 is selected from the group consisting of L, G, T, Q, A, S, N, R and K. In some  
15 embodiments, the amino acid of the DNA polymerase corresponding to position 580 of SEQ ID NO:1 is G.

Various DNA polymerases are amenable to mutation according to the present invention.

Particularly suitable are thermostable polymerases, including wild-type or naturally occurring thermostable polymerases from various species of thermophilic bacteria, as well as synthetic  
20 thermostable polymerases derived from such wild-type or naturally occurring enzymes by amino acid substitution, insertion, or deletion, or other modification. Exemplary unmodified forms of polymerase include, e.g., CS5 (SEQ ID NO:29), CS6 (SEQ ID NO:30) or Z05 DNA polymerase (SEQ ID NO:1), or a functional DNA polymerase having at least 80%, preferably at least 90%, more preferably at least 95% sequence identity thereto. Other unmodified polymerases include,  
25 e.g., DNA polymerases from any of the following species of thermophilic bacteria (or a functional DNA polymerase having at least 80%, preferably at least 90%, more preferably at least 95% sequence identity to such a polymerase): *Thermotoga maritima* (SEQ ID NO:38); *Thermus aquaticus* (SEQ ID NO:2); *Thermus thermophilus* (SEQ ID NO:6); *Thermus flavus* (SEQ ID NO:4); *Thermus filiformis* (SEQ ID NO:3); *Thermus* sp. Sps17 (SEQ ID NO:5);  
30 *Thermus* sp. Z05 (SEQ ID NO:1); *Thermotoga neopolitana* (SEQ ID NO:39); *Thermosiphon*

*africanus* (SEQ ID NO:37); *Thermus caldophilus* (SEQ ID NO:7), *Deinococcus radiodurans* (SEQ ID NO:36), *Bacillus stearothermophilus* (SEQ ID NO:40) or *Bacillus caldotenax* (SEQ ID NO:41). Suitable polymerases also include those having reverse transcriptase (RT) activity and/or the ability to incorporate unconventional nucleotides, such as ribonucleotides or other 2'-  
5 modified nucleotides.

While thermostable DNA polymerases possessing efficient 3'-mismatch discrimination activity are particularly suited for performing PCR, thermoactive, but not thermostable DNA polymerases possessing efficient 3'-mismatch discrimination activity also are amenable to mutation according to the present invention.

- 10 In some embodiments, the DNA polymerase is a *Thermus* DNA polymerase. For example, in some embodiments, the DNA polymerase has at least 80%, preferably at least 90%, more preferably at least 95% sequence identity to a polymerase selected from the group consisting of:
- (a) a *Thermus sp.* Z05 DNA polymerase (Z05) (SEQ ID NO:1);
  - (b) a *Thermus aquaticus* DNA polymerase (Taq) (SEQ ID NO:2);
  - 15 (c) a *Thermus filiformis* DNA polymerase (Tfi) (SEQ ID NO:3);
  - (d) a *Thermus flavus* DNA polymerase (Tfl) (SEQ ID NO:4);
  - (e) a *Thermus sp.* Sps17 DNA polymerase (Sps17) (SEQ ID NO:5);
  - (f) a *Thermus thermophilus* DNA polymerase (Tth) (SEQ ID NO:6); and
  - (g) a *Thermus caldophilus* DNA polymerase (Tca) (SEQ ID NO:7).
- 20 In some embodiments, the DNA polymerase is a *Thermotoga* DNA polymerase. For example, in some embodiments, the DNA polymerase has at least 80%, preferably at least 90%, more preferably at least 95% sequence identity to a polymerase selected from the group consisting of:
- (a) a *Thermotoga maritima* DNA polymerase (Tma) (SEQ ID NO:38);
  - (b) a *Thermotoga neopolitana* DNA polymerase (Tne) (SEQ ID NO:39);
- 25 In certain embodiments, the DNA polymerase has at least 80%, preferably at least 90%, more preferably at least 95% sequence identity to SEQ ID NO:1. In some embodiments, the DNA polymerase is a *Thermus sp.* Z05 DNA polymerase (Z05) DNA polymerase (*i.e.*, SEQ ID NO:1), except that the amino acid at position 589 is any amino acid other than R. For example, in some  
30 Y, N, Q, D, E, K, or H. In some embodiments, the DNA polymerase is a Z05 DNA polymerase,

and the amino acid at position 589 is any amino acid other than R or K. In some embodiments, the DNA polymerase is a Z05 DNA polymerase, and the amino acid at position 589 is H, L, or S. In some embodiments, the DNA polymerase is a Z05 DNA polymerase further comprising a substitution at position 580, and the amino acid at position 580 is any amino acid other than D or  
5 E. In some embodiments, the DNA polymerase is a Z05 DNA polymerase, and the amino acid at position 580 is any amino acid other than D. In some embodiments, the DNA polymerase is a Z05 DNA polymerase, and the amino acid at position 580 is selected from the group consisting of L, G, T, Q, A, S, N, R and K. In some embodiments, the DNA polymerase is a Z05 DNA polymerase, and the amino acid at position 580 is G.

10 The mutant or improved polymerase can include other, non-substitutional modifications. One such modification is a thermally reversible covalent modification that inactivates the enzyme, but which is reversed to activate the enzyme upon incubation at an elevated temperature, such as a temperature typically used for polynucleotide extension. Exemplary reagents for such thermally reversible modifications are described in U.S. Patent Nos. 5,773, 258 and 5,677,152 to Birch *et*  
15 *al.*.

In some embodiments, the 3'-mismatch activity is determined using a mutant BRAF V600R target polynucleotide having the nucleic acid sequence of SEQ ID NO:35 (wild type BRAF = SEQ ID NO:34) in the presence of a forward primer that is perfectly matched to the mutant sequence and has a single 3' A:C mismatch to the wild type sequence in one or more reaction  
20 mixtures having a predetermined number of copies of wild-type BRAF V600 target polynucleotide and a predetermined number of copies of mutant BRAF V600R target polynucleotide equal in number or fewer than the number of copies of wild-type target (*e.g.*, 10,000 or fewer copies). Two or more reaction mixtures can have titrated numbers of copies of mutant BRAF V600R target polynucleotide (*e.g.*, 1:5 titrations, 1:10 titrations, *e.g.*, 10,000  
25 copies, 1000 copies, 100 copies, 10 copies, 1 copy, 0 copies in several reaction mixtures). The 3'-mismatch discrimination ability of a polymerase of the invention can be compared to the 3'-mismatch discrimination ability of a reference polymerase (*e.g.*, a naturally occurring or unmodified polymerase), over a preselected unit of time, as described herein. Polymerases with increased 3'-mismatch discrimination ability will not amplify the wild-type sequence when  
30 contacted with a primer that is perfectly matched to a mutant allele, or will require a greater number of PCR cycles to amplify the wild-type sequence using the mutant allele-specific primer

(*i.e.*, exhibit a higher Cp value), in comparison to a naturally occurring or unmodified polymerase.

In various other aspects, the present invention provides a recombinant nucleic acid encoding a mutant or improved DNA polymerase as described herein, a vector comprising the recombinant nucleic acid, and/or a host cell transformed with the vector. In certain embodiments, the vector is an expression vector. Host cells comprising such expression vectors are useful in methods of the invention for producing the mutant or improved polymerase by culturing the host cells under conditions suitable for expression of the recombinant nucleic acid. The polymerases of the invention may be contained in reaction mixtures and/or kits. The embodiments of the recombinant nucleic acids, host cells, vectors, expression vectors, reaction mixtures and kits are as described above and herein.

In yet another aspect, a method for conducting polynucleotide extension is provided. The method generally includes contacting a DNA polymerase having increased 3'-mismatch discrimination as described herein with a primer, a polynucleotide template, and nucleoside triphosphates under conditions suitable for extension of the primer, thereby producing an extended primer. The polynucleotide template can be, for example, an RNA or DNA template. The nucleoside triphosphates can include unconventional nucleotides such as, *e.g.*, ribonucleotides and/or labeled nucleotides. Further, the primer and/or template can include one or more nucleotide analogs. In some variations, the polynucleotide extension method is a method for polynucleotide amplification that includes contacting the mutant or improved DNA polymerase with a primer pair, the polynucleotide template, and the nucleoside triphosphates under conditions suitable for amplification of the polynucleotide. The polynucleotide extension reaction can be, *e.g.*, PCR, isothermal extension, or sequencing (*e.g.*, 454 sequencing reaction).

In some embodiments the primer extension method is a method for conducting polymerase chain reaction (PCR).

The present invention also provides a kit useful in such a polynucleotide extension method. Generally, the kit includes at least one container providing a mutant or improved DNA polymerase as described herein. In certain embodiments, the kit further includes one or more additional containers providing one or more additional reagents. For example, in specific variations, the one or more additional containers provide nucleoside triphosphates; a buffer



suitable for polynucleotide extension; and/or a primer hybridizable, under polynucleotide extension conditions, to a predetermined polynucleotide template.

Further provided are reaction mixtures comprising the polymerases of the invention. The reactions mixtures can also contain a template nucleic acid (DNA and/or RNA), one or more  
5 primer or probe polynucleotides, nucleoside triphosphates (including, *e.g.*, deoxyribonucleotides, ribonucleotides, labeled nucleotides, unconventional nucleotides), buffers, salts, labels (*e.g.*, fluorophores).

Further embodiments of the invention are described herein.

### DEFINITIONS

10 Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Although essentially any methods and materials similar to those described herein can be used in the practice or testing of the present invention, only exemplary methods and materials are described. For purposes of the present invention, the following terms are defined below.

15 The terms “a,” “an,” and “the” include plural referents, unless the context clearly indicates otherwise.

An “amino acid” refers to any monomer unit that can be incorporated into a peptide, polypeptide, or protein. As used herein, the term “amino acid” includes the following twenty natural or  
20 genetically encoded alpha-amino acids: alanine (Ala or A), arginine (Arg or R), asparagine (Asn or N), aspartic acid (Asp or D), cysteine (Cys or C), glutamine (Gln or Q), glutamic acid (Glu or E), glycine (Gly or G), histidine (His or H), isoleucine (Ile or I), leucine (Leu or L), lysine (Lys or K), methionine (Met or M), phenylalanine (Phe or F), proline (Pro or P), serine (Ser or S), threonine (Thr or T), tryptophan (Trp or W), tyrosine (Tyr or Y), and valine (Val or V). In cases  
25 where “X” residues are undefined, these should be defined as “any amino acid.” The structures of these twenty natural amino acids are shown in, *e.g.*, Stryer et al., Biochemistry, 5<sup>th</sup> ed., Freeman and Company (2002). Additional amino acids, such as selenocysteine and pyrrolysine, can also be genetically coded for (Stadtman (1996) “Selenocysteine,” Annu Rev Biochem. 65:83-100 and Ibba et al. (2002) “Genetic code: introducing pyrrolysine,” Curr Biol. 12(13):R464-R466). The term “amino acid” also includes unnatural amino acids, modified

amino acids (e.g., having modified side chains and/or backbones), and amino acid analogs. See, e.g., Zhang et al. (2004) "Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells," Proc. Natl. Acad. Sci. U.S.A. 101(24):8882-8887, Anderson et al. (2004) "An expanded genetic code with a functional quadruplet codon" Proc. Natl. Acad. Sci. U.S.A. 101(20):7566-7571, Ikeda et al. (2003) "Synthesis of a novel histidine analogue and its efficient incorporation into a protein in vivo," Protein Eng. Des. Sel. 16(9):699-706, Chin et al. (2003) "An Expanded Eukaryotic Genetic Code," Science 301(5635):964-967, James et al. (2001) "Kinetic characterization of ribonuclease S mutants containing photoisomerizable phenylazophenylalanine residues," Protein Eng. Des. Sel. 14(12):983-991, Kohrer et al. (2001) "Import of amber and ochre suppressor tRNAs into mammalian cells: A general approach to site-specific insertion of amino acid analogues into proteins," Proc. Natl. Acad. Sci. U.S.A. 98(25):14310-14315, Bacher et al. (2001) "Selection and Characterization of Escherichia coli Variants Capable of Growth on an Otherwise Toxic Tryptophan Analogue," J. Bacteriol. 183(18):5414-5425, Hamano-Takaku et al. (2000) "A Mutant Escherichia coli Tyrosyl-tRNA Synthetase Utilizes the Unnatural Amino Acid Azatyrosine More Efficiently than Tyrosine," J. Biol. Chem. 275(51):40324-40328, and Budisa et al. (2001) "Proteins with {beta}-(thienopyrrolyl)alanines as alternative chromophores and pharmaceutically active amino acids," Protein Sci. 10(7):1281-1292.

To further illustrate, an amino acid is typically an organic acid that includes a substituted or unsubstituted amino group, a substituted or unsubstituted carboxy group, and one or more side chains or groups, or analogs of any of these groups. Exemplary side chains include, e.g., thiol, seleno, sulfonyl, alkyl, aryl, acyl, keto, azido, hydroxyl, hydrazine, cyano, halo, hydrazide, alkenyl, alkylnl, ether, borate, boronate, phospho, phosphono, phosphine, heterocyclic, enone, imine, aldehyde, ester, thioacid, hydroxylamine, or any combination of these groups. Other representative amino acids include, but are not limited to, amino acids comprising photoactivatable cross-linkers, metal binding amino acids, spin-labeled amino acids, fluorescent amino acids, metal-containing amino acids, amino acids with novel functional groups, amino acids that covalently or noncovalently interact with other molecules, photocaged and/or photoisomerizable amino acids, radioactive amino acids, amino acids comprising biotin or a biotin analog, glycosylated amino acids, other carbohydrate modified amino acids, amino acids comprising polyethylene glycol or polyether, heavy atom substituted amino acids, chemically

cleavable and/or photocleavable amino acids, carbon-linked sugar-containing amino acids, redox-active amino acids, amino thioacid containing amino acids, and amino acids comprising one or more toxic moieties.

5 The term "aptamer" refers to a single-stranded DNA that recognizes and binds to DNA polymerase, and efficiently inhibits the polymerase activity as described in U.S. Pat. No. 5,693,502.

The term "mutant," in the context of DNA polymerases of the present invention, means a polypeptide, typically recombinant, that comprises one or more amino acid substitutions relative to a corresponding, naturally-occurring or unmodified DNA polymerase.

10 The term "unmodified form," in the context of a mutant polymerase, is a term used herein for purposes of defining a mutant DNA polymerase of the present invention: the term "unmodified form" refers to a functional DNA polymerase that has the amino acid sequence of the mutant polymerase except at one or more amino acid position(s) specified as characterizing the mutant polymerase. Thus, reference to a mutant DNA polymerase in terms of (a) its unmodified form  
15 and (b) one or more specified amino acid substitutions means that, with the exception of the specified amino acid substitution(s), the mutant polymerase otherwise has an amino acid sequence identical to the unmodified form in the specified motif. The "unmodified polymerase" (and therefore also the modified form having increased 3'-mismatch discrimination) may contain additional mutations to provide desired functionality, e.g., improved incorporation of  
20 dideoxynucleotides, ribonucleotides, ribonucleotide analogs, dye-labeled nucleotides, modulating 5'-nuclease activity, modulating 3'-nuclease (or proofreading) activity, or the like. Accordingly, in carrying out the present invention as described herein, the unmodified form of a DNA polymerase is predetermined. The unmodified form of a DNA polymerase can be, for example, a wild-type and/or a naturally occurring DNA polymerase, or a DNA polymerase that  
25 has already been intentionally modified. An unmodified form of the polymerase is preferably a thermostable DNA polymerases, such as DNA polymerases from various thermophilic bacteria, as well as functional variants thereof having substantial sequence identity to a wild-type or naturally occurring thermostable polymerase. Such variants can include, for example, chimeric DNA polymerases such as, for example, the chimeric DNA polymerases described in U.S. Patent  
30 No. 6,228,628 and U.S. Application Publication No. 2004/0005599. In certain embodiments, the unmodified form of a polymerase has reverse transcriptase (RT) activity.

The term "thermostable polymerase," refers to an enzyme that is stable to heat, is heat resistant, and retains sufficient activity to effect subsequent polynucleotide extension reactions and does not become irreversibly denatured (inactivated) when subjected to the elevated temperatures for the time necessary to effect denaturation of double-stranded nucleic acids. The heating  
5 conditions necessary for nucleic acid denaturation are well known in the art and are exemplified in, e.g., U.S. Patent Nos. 4,683,202, 4,683,195, and 4,965,188. As used herein, a thermostable polymerase is suitable for use in a temperature cycling reaction such as the polymerase chain reaction ("PCR"). Irreversible denaturation for purposes herein refers to permanent and complete loss of enzymatic activity. For a thermostable polymerase, enzymatic activity refers to the  
10 catalysis of the combination of the nucleotides in the proper manner to form polynucleotide extension products that are complementary to a template nucleic acid strand. Thermostable DNA polymerases from thermophilic bacteria include, e.g., DNA polymerases from *Thermotoga maritima*, *Thermus aquaticus*, *Thermus thermophilus*, *Thermus flavus*, *Thermus filiformis*, *Thermus* species Sps17, *Thermus* species Z05, *Thermus caldophilus*, *Bacillus caldotenax*,  
15 *Thermotoga neopolitana*, and *Thermosipho africanus*.

The term "thermoactive" refers to an enzyme that maintains catalytic properties at temperatures commonly used for reverse transcription or anneal/extension steps in RT-PCR and/or PCR reactions (i.e., 45-80 °C). Thermostable enzymes are those which are not irreversibly inactivated or denatured when subjected to elevated temperatures necessary for nucleic acid denaturation.  
20 Thermoactive enzymes may or may not be thermostable. Thermoactive DNA polymerases can be DNA or RNA dependent from thermophilic species or from mesophilic species including, but not limited to, *Escherichia coli*, *Moloney murine leukemia viruses*, and *Avian myoblastosis virus*.

As used herein, a "chimeric" protein refers to a protein whose amino acid sequence represents a fusion product of subsequences of the amino acid sequences from at least two distinct proteins. A  
25 chimeric protein typically is not produced by direct manipulation of amino acid sequences, but, rather, is expressed from a "chimeric" gene that encodes the chimeric amino acid sequence. In certain embodiments, for example, an unmodified form of a mutant DNA polymerase of the present invention is a chimeric protein that consists of an amino-terminal (N-terminal) region derived from a *Thermus* species DNA polymerase and a carboxy-terminal (C-terminal) region  
30 derived from Tma DNA polymerase. The N-terminal region refers to a region extending from the

N-terminus (amino acid position 1) to an internal amino acid. Similarly, the C-terminal region refers to a region extending from an internal amino acid to the C-terminus.

In the context of DNA polymerases, "correspondence" to another sequence (*e.g.*, regions, fragments, nucleotide or amino acid positions, or the like) is based on the convention of numbering according to nucleotide or amino acid position number and then aligning the sequences in a manner that maximizes the percentage of sequence identity. Because not all positions within a given "corresponding region" need be identical, non-matching positions within a corresponding region may be regarded as "corresponding positions." Accordingly, as used herein, referral to an "amino acid position corresponding to amino acid position [X]" of a specified DNA polymerase refers to equivalent positions, based on alignment, in other DNA polymerases and structural homologues and families. In some embodiments of the present invention, "correspondence" of amino acid positions are determined with respect to a region of the polymerase comprising one or more motifs of SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 36, 37, 38, 39, 40, or 41. When a polymerase polypeptide sequence differs from SEQ ID NOS:1, 2, 3, 4, 5, 6, 7, 36, 37, 38, 39, 40, or 41 (*e.g.*, by changes in amino acids or addition or deletion of amino acids), it may be that a particular mutation associated with improved activity as discussed herein will not be in the same position number as it is in SEQ ID NOS:1, 2, 3, 4, 5, 6, 7, 36, 37, 38, 39, 40, or 41. This is illustrated, for example, in Table 1.

"Recombinant," as used herein, refers to an amino acid sequence or a nucleotide sequence that has been intentionally modified by recombinant methods. By the term "recombinant nucleic acid" herein is meant a nucleic acid, originally formed *in vitro*, in general, by the manipulation of a nucleic acid by endonucleases, in a form not normally found in nature. Thus an isolated, mutant DNA polymerase nucleic acid, in a linear form, or an expression vector formed *in vitro* by ligating DNA molecules that are not normally joined, are both considered recombinant for the purposes of this invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell, it will replicate non-recombinantly, *i.e.*, using the *in vivo* cellular machinery of the host cell rather than *in vitro* manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes of the invention. A "recombinant protein" is a protein made using recombinant techniques, *i.e.*, through the expression of a recombinant nucleic acid as depicted above.

A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation.

- 5 The term "host cell" refers to both single-cellular prokaryote and eukaryote organisms (*e.g.*, bacteria, yeast, and actinomycetes) and single cells from higher order plants or animals when being grown in cell culture.

The term "vector" refers to a piece of DNA, typically double-stranded, which may have inserted into it a piece of foreign DNA. The vector or may be, for example, of plasmid origin. Vectors  
10 contain "replicon" polynucleotide sequences that facilitate the autonomous replication of the vector in a host cell. Foreign DNA is defined as heterologous DNA, which is DNA not naturally found in the host cell, which, for example, replicates the vector molecule, encodes a selectable or screenable marker, or encodes a transgene. The vector is used to transport the foreign or heterologous DNA into a suitable host cell. Once in the host cell, the vector can replicate  
15 independently of or coincidental with the host chromosomal DNA, and several copies of the vector and its inserted DNA can be generated. In addition, the vector can also contain the necessary elements that permit transcription of the inserted DNA into an mRNA molecule or otherwise cause replication of the inserted DNA into multiple copies of RNA. Some expression vectors additionally contain sequence elements adjacent to the inserted DNA that increase the  
20 half-life of the expressed mRNA and/or allow translation of the mRNA into a protein molecule. Many molecules of mRNA and polypeptide encoded by the inserted DNA can thus be rapidly synthesized.

The term "nucleotide," in addition to referring to the naturally occurring ribonucleotide or deoxyribonucleotide monomers, shall herein be understood to refer to related structural variants  
25 thereof, including derivatives and analogs, that are functionally equivalent with respect to the particular context in which the nucleotide is being used (*e.g.*, hybridization to a complementary base), unless the context clearly indicates otherwise.

The term "nucleic acid" or "polynucleotide" refers to a polymer that can be corresponded to a ribose nucleic acid (RNA) or deoxyribose nucleic acid (DNA) polymer, or an analog thereof.  
30 This includes polymers of nucleotides such as RNA and DNA, as well as synthetic forms,

modified (*e.g.*, chemically or biochemically modified) forms thereof, and mixed polymers (*e.g.*, including both RNA and DNA subunits). Exemplary modifications include methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, and the like), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, and the like), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids and the like). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Typically, the nucleotide monomers are linked via phosphodiester bonds, although synthetic forms of nucleic acids can comprise other linkages (*e.g.*, peptide nucleic acids as described in Nielsen *et al.* (*Science* 254:1497-1500, 1991). A nucleic acid can be or can include, *e.g.*, a chromosome or chromosomal segment, a vector (*e.g.*, an expression vector), an expression cassette, a naked DNA or RNA polymer, the product of a polymerase chain reaction (PCR), an oligonucleotide, a probe, and a primer. A nucleic acid can be, *e.g.*, single-stranded, double-stranded, or triple-stranded and is not limited to any particular length. Unless otherwise indicated, a particular nucleic acid sequence comprises or encodes complementary sequences, in addition to any sequence explicitly indicated.

The term "oligonucleotide" refers to a nucleic acid that includes at least two nucleic acid monomer units (*e.g.*, nucleotides). An oligonucleotide typically includes from about six to about 175 nucleic acid monomer units, more typically from about eight to about 100 nucleic acid monomer units, and still more typically from about 10 to about 50 nucleic acid monomer units (*e.g.*, about 15, about 20, about 25, about 30, about 35, or more nucleic acid monomer units). The exact size of an oligonucleotide will depend on many factors, including the ultimate function or use of the oligonucleotide. Oligonucleotides are optionally prepared by any suitable method, including, but not limited to, isolation of an existing or natural sequence, DNA replication or amplification, reverse transcription, cloning and restriction digestion of appropriate sequences, or direct chemical synthesis by a method such as the phosphotriester method of Narang *et al.* (*Meth. Enzymol.* 68:90-99, 1979); the phosphodiester method of Brown *et al.* (*Meth. Enzymol.* 68:109-151, 1979); the diethylphosphoramidite method of Beaucage *et al.* (*Tetrahedron Lett.* 22:1859-1862, 1981); the triester method of Matteucci *et al.* (*J. Am. Chem. Soc.* 103:3185-3191, 1981);

automated synthesis methods; or the solid support method of U.S. Pat. No. 4,458,066, or other methods known to those skilled in the art.

The term "primer" as used herein refers to a polynucleotide capable of acting as a point of initiation of template-directed nucleic acid synthesis when placed under conditions in which polynucleotide extension is initiated (*e.g.*, under conditions comprising the presence of requisite nucleoside triphosphates (as dictated by the template that is copied) and a polymerase in an appropriate buffer and at a suitable temperature or cycle(s) of temperatures (*e.g.*, as in a polymerase chain reaction)). To further illustrate, primers can also be used in a variety of other oligonucleotide-mediated synthesis processes, including as initiators of *de novo* RNA synthesis and *in vitro* transcription-related processes (*e.g.*, nucleic acid sequence-based amplification (NASBA), transcription mediated amplification (TMA), etc.). A primer is typically a single-stranded oligonucleotide (*e.g.*, oligodeoxyribonucleotide). The appropriate length of a primer depends on the intended use of the primer but typically ranges from 6 to 40 nucleotides, more typically from 15 to 35 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template for primer elongation to occur. In certain embodiments, the term "primer pair" means a set of primers including a 5' sense primer (sometimes called "forward") that hybridizes with the complement of the 5' end of the nucleic acid sequence to be amplified and a 3' antisense primer (sometimes called "reverse") that hybridizes with the 3' end of the sequence to be amplified (*e.g.*, if the target sequence is expressed as RNA or is an RNA). A primer can be labeled, if desired, by incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include <sup>32</sup>P, fluorescent dyes, electron-dense reagents, enzymes (as commonly used in ELISA assays), biotin, or haptens and proteins for which antisera or monoclonal antibodies are available.

The term "5'-nuclease probe" refers to an oligonucleotide that comprises at least one light emitting labeling moiety and that is used in a 5'-nuclease reaction to effect target nucleic acid detection. In some embodiments, for example, a 5'-nuclease probe includes only a single light emitting moiety (*e.g.*, a fluorescent dye, etc.). In certain embodiments, 5'-nuclease probes include regions of self-complementarity such that the probes are capable of forming hairpin structures under selected conditions. To further illustrate, in some embodiments a 5'-nuclease



probe comprises at least two labeling moieties and emits radiation of increased intensity after one of the two labels is cleaved or otherwise separated from the oligonucleotide. In certain embodiments, a 5'-nuclease probe is labeled with two different fluorescent dyes, e.g., a 5' terminus reporter dye and the 3' terminus quencher dye or moiety. In some embodiments, 5'-nuclease probes are labeled at one or more positions other than, or in addition to, terminal positions. When the probe is intact, energy transfer typically occurs between the two fluorophores such that fluorescent emission from the reporter dye is quenched at least in part. During an extension step of a polymerase chain reaction, for example, a 5'-nuclease probe bound to a template nucleic acid is cleaved by the 5' to 3' nuclease activity of, e.g., a *Taq* polymerase or another polymerase having this activity such that the fluorescent emission of the reporter dye is no longer quenched. Exemplary 5'-nuclease probes are also described in, e.g., U.S. Pat. No. 5,210,015; U.S. Pat. No. 5,994,056; and U.S. Pat. No. 6,171,785. In other embodiments, a 5' nuclease probe may be labeled with two or more different reporter dyes and a 3' terminus quencher dye or moiety.

The term "FRET" or "fluorescent resonance energy transfer" or "Foerster resonance energy transfer" refers to a transfer of energy between at least two chromophores, a donor chromophore and an acceptor chromophore (referred to as a quencher). The donor typically transfers the energy to the acceptor when the donor is excited by light radiation with a suitable wavelength. The acceptor typically re-emits the transferred energy in the form of light radiation with a different wavelength. When the acceptor is a "dark" quencher, it dissipates the transferred energy in a form other than light. Whether a particular fluorophore acts as a donor or an acceptor depends on the properties of the other member of the FRET pair. Commonly used donor-acceptor pairs include the FAM-TAMRA pair. Commonly used quenchers are DABCYL and TAMRA. Commonly used dark quenchers include BlackHole Quenchers™ (BHQ), (Biosearch Technologies, Inc., Novato, Cal.), Iowa Black™ (Integrated DNA Tech., Inc., Coralville, Iowa), and BlackBerry™ Quencher 650 (BBQ-650) (Berry & Assoc., Dexter, Mich.).

The term "conventional" or "natural" when referring to nucleic acid bases, nucleoside triphosphates, or nucleotides refers to those which occur naturally in the polynucleotide being described (*i.e.*, for DNA these are dATP, dGTP, dCTP and dTTP). Additionally, dITP, and 7-deaza-dGTP are frequently utilized in place of dGTP and 7-deaza-dATP can be utilized in place

of dATP in *in vitro* DNA synthesis reactions, such as sequencing. Collectively, these may be referred to as dNTPs.

The term "unconventional" or "modified" when referring to a nucleic acid base, nucleoside, or nucleotide includes modification, derivations, or analogues of conventional bases, nucleosides, or nucleotides that naturally occur in a particular polynucleotide. Certain unconventional nucleotides are modified at the 2' position of the ribose sugar in comparison to conventional dNTPs. Thus, although for RNA the naturally occurring nucleotides are ribonucleotides (*i.e.*, ATP, GTP, CTP, UTP, collectively rNTPs), because these nucleotides have a hydroxyl group at the 2' position of the sugar, which, by comparison is absent in dNTPs, as used herein, ribonucleotides are unconventional nucleotides as substrates for DNA polymerases. As used herein, unconventional nucleotides include, but are not limited to, compounds used as terminators for nucleic acid sequencing. Exemplary terminator compounds include but are not limited to those compounds that have a 2',3' dideoxy structure and are referred to as dideoxynucleoside triphosphates. The dideoxynucleoside triphosphates ddATP, ddTTP, ddCTP and ddGTP are referred to collectively as ddNTPs. Additional examples of terminator compounds include 2'-PO<sub>4</sub> analogs of ribonucleotides (*see, e.g.*, U.S. Application Publication Nos. 2005/0037991 and 2005/0037398). Other unconventional nucleotides include phosphorothioate dNTPs ([[α]-S]dNTPs), 5'-[α]-borano-dNTPs, [α]-methyl-phosphonate dNTPs, and ribonucleoside triphosphates (rNTPs). Unconventional bases may be labeled with radioactive isotopes such as <sup>32</sup>P, <sup>33</sup>P, or <sup>35</sup>S; fluorescent labels; chemiluminescent labels; bioluminescent labels; hapten labels such as biotin; or enzyme labels such as streptavidin or avidin. Fluorescent labels may include dyes that are negatively charged, such as dyes of the fluorescein family, or dyes that are neutral in charge, such as dyes of the rhodamine family, or dyes that are positively charged, such as dyes of the cyanine family. Dyes of the fluorescein family include, *e.g.*, FAM, HEX, TET, JOE, NAN and ZOE. Dyes of the rhodamine family include Texas Red, ROX, R110, R6G, and TAMRA. Various dyes or nucleotides labeled with FAM, HEX, TET, JOE, NAN, ZOE, ROX, R110, R6G, Texas Red and TAMRA are marketed by Perkin-Elmer (Boston, MA), Applied Biosystems (Foster City, CA), or Invitrogen/Molecular Probes (Eugene, OR). Dyes of the cyanine family include Cy2, Cy3, Cy5, and Cy7 and are marketed by GE Healthcare UK Limited (Amersham Place, Little Chalfont, Buckinghamshire, England).

As used herein, "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the sequence in the comparison window can comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

10 The terms "identical" or "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same. Sequences are "substantially identical" to each other if they have a specified percentage of nucleotides or amino acid residues that are the same (*e.g.*, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identity over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. These definitions also refer to the complement of a test sequence. Optionally, the identity exists over a region that is at least about 50 nucleotides in length, or more typically over a region that is 100 to 500 or 1000 or more nucleotides in length.

25 The terms "similarity" or "percent similarity," in the context of two or more polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of amino acid residues that are either the same or similar as defined by a conservative amino acid substitutions (*e.g.*, 60% similarity, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% similar over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Sequences are "substantially similar" to each other if they are at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, or at least 55% similar to each other. Optionally, this similarity exists over a region that is at least about 50 amino acids in length, or more typically over a region that is at least about 100 to 500 or 1000 or more amino acids in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters are commonly used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities or similarities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window," as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, for example, by the local homology algorithm of Smith and Waterman (*Adv. Appl. Math.* 2:482, 1970), by the homology alignment algorithm of Needleman and Wunsch (*J. Mol. Biol.* 48:443, 1970), by the search for similarity method of Pearson and Lipman (*Proc. Natl. Acad. Sci. USA* 85:2444, 1988), by computerized implementations of these algorithms (*e.g.*, GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection (*see, e.g.*, Ausubel *et al.*, *Current Protocols in Molecular Biology* (1995 supplement)).

Algorithms suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (*Nuc. Acids Res.* 25:3389-402, 1977), and Altschul *et al.* (*J. Mol. Biol.* 215:403-10, 1990), respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can

be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) or 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff and Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915, 1989) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin and Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873-87, 1993). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, typically less than about 0.01, and more typically less than about 0.001.

The term “mismatch discrimination” refers to the ability of a biocatalyst (e.g., an enzyme, such as a polymerase, ligase, or the like) to distinguish a fully complementary sequence from a mismatch-containing sequence when extending a nucleic acid (e.g., a primer or other oligonucleotide) in a template-dependent manner by attaching (e.g., covalently) one or more nucleotides to the nucleic acid. The term “3'-mismatch discrimination” refers to the ability of a biocatalyst to distinguish a fully complementary sequence from a mismatch-containing (nearly complementary) sequence where the nucleic acid to be extended (e.g., a primer or other oligonucleotide) has a mismatch at the nucleic acid's 3' terminus compared to the template to which the nucleic acid hybridizes. In some embodiments, the nucleic acid to be extended comprises a mismatch at the 3' end relative to the fully complementary sequence. In some

embodiments, the nucleic acid to be extended comprises a mismatch at the penultimate (N-1) 3' position and/or at the N-2 position relative to the fully complementary sequence.

The term "Cp value" or "crossing point" value refers to a value that allows quantification of input target nucleic acids. The Cp value can be determined according to the second-derivative maximum method (Van Luu-The, et al., "Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction," BioTechniques, Vol. 38, No. 2, February 2005, pp. 287–293). In the second derivative method, a Cp corresponds to the first peak of a second derivative curve. This peak corresponds to the beginning of a log-linear phase. The second derivative method calculates a second derivative value of the real-time fluorescence intensity curve, and only one value is obtained. The original Cp method is based on a locally defined, differentiable approximation of the intensity values, e.g., by a polynomial function. Then the third derivative is computed. The Cp value is the smallest root of the third derivative. The Cp can also be determined using the fit point method, in which the Cp is determined by the intersection of a parallel to the threshold line in the log-linear region (Van Luu-The, et al., BioTechniques, Vol. 38, No. 2, February 2005, pp. 287–293). These computations are easily carried out by any person skilled in the art.

The term "PCR efficiency" refers to an indication of cycle to cycle amplification efficiency for the perfectly matched primer template. PCR efficiency is calculated for each condition using the equation: % PCR efficiency =  $(10^{(-\text{slope})} - 1) \times 100$ , wherein the slope was calculated by linear regression with the log copy number plotted on the y-axis and Cp plotted on the x-axis.

The term "multiplex" refers to amplification with more than one set of primers, or the amplification of more than one polymorphism site in a single reaction.

### BRIEF DESCRIPTION OF THE DRAWINGS

**Figure 1** depicts an amino acid sequence alignment of a region from the polymerase domain of exemplary DNA polymerases from various species of bacteria: *Thermus* species Z05 (Z05) (SEQ ID NO:12), *Thermus aquaticus* (Taq) (SEQ ID NO:13), *Thermus filiformis* (Tfi) (SEQ ID NO:14), *Thermus flavus* (Tfl) (SEQ ID NO:15), *Thermus* species Sps17 (Sps17) (SEQ ID NO:16), *Thermus thermophilus* (Tth) (SEQ ID NO:17), *Thermus caldophilus* (Tca) (SEQ ID NO:18), *Thermotoga maritima* (Tma) (SEQ ID NO:19), *Thermotoga neopolitana* (Tne) (SEQ ID

NO:20), *Thermosipho africanus* (Taf) (SEQ ID NO:21), *Escherichia coli* (E) (SEQ ID NO:22),  
*Deinococcus radiodurans* (Dra) (SEQ ID NO:23), *Bacillus stearothermophilus* (Bst) (SEQ ID  
NO:24), and *Bacillus caldotenax* (Bca) (SEQ ID NO:25). In addition, the polypeptide regions  
shown comprise the amino acid motif P-N-L-Q-N-X<sub>1</sub>-P-X<sub>2</sub>-X<sub>3</sub>-X<sub>4</sub>-X<sub>5</sub>-X<sub>6</sub>-G (SEQ ID NO:26),  
5 the variable positions of which are further defined herein. This motif is highlighted in bold type  
for each polymerase sequence. Amino acid positions amenable to mutation in accordance with  
the present invention are indicated with an asterisk (\*).

**Figure 2** provides sequence identities among the following DNA Polymerase I enzymes:

*Thermus sp.* Z05 DNA polymerase (Z05); *Thermus aquaticus* DNA polymerase (Taq); *Thermus*  
10 *filiformis* DNA polymerase (Tfi); *Thermus flavus* DNA polymerase (Tfl); *Thermus sp.* Sps17  
DNA polymerase (Sps17); *Thermus thermophilus* DNA polymerase (Tth); *Thermus caldophilus*  
DNA polymerase (Tca); *Deinococcus radiodurans* DNA polymerase (Dra); *Thermotoga*  
*maritima* DNA polymerase (Tma); *Thermotoga neopolitana* DNA polymerase (Tne);  
*Thermosipho africanus* DNA polymerase (Taf); *Bacillus stearothermophilus* DNA polymerase  
15 (Bst); and *Bacillus caldotenax* DNA polymerase (Bca). (A) sequence identities over the entire  
polymerase I enzyme (corresponding to amino acids 1-834 of Z05); and (B) sequence identities  
over the polymerase sub domain corresponding to amino acids 420-834 of Z05.

## DETAILED DESCRIPTION

The present invention provides improved DNA polymerases in which one or more amino acids  
20 in the polymerase domain have been identified as improving one or more polymerase activity or  
characteristics. The DNA polymerases of the invention are active enzymes having increased 3'-  
mismatch discrimination activity (i.e., the inventive polymerases described herein are less likely  
to extend primers that are mismatched to template at or near the 3' end of the primer) relative to  
the unmodified form of the polymerase otherwise identical except for the amino acid difference  
25 noted herein. The DNA polymerases are useful in a variety of applications involving  
polynucleotide extension or amplification of polynucleotide templates, including, for example,  
applications in recombinant DNA studies and medical diagnosis of disease.

***Polymerases of the Invention***

In some embodiments, the DNA polymerases of the invention can be characterized by having the following motif:

5 Pro-Asn-Leu-Gln-Asn-X<sub>1</sub>-Pro-X<sub>2</sub>-X<sub>3</sub>-X<sub>4</sub>-X<sub>5</sub>-X<sub>6</sub>-Gly (also referred to  
 herein in the one-letter code as P-N-L-Q-N-X<sub>1</sub>-P-X<sub>2</sub>-X<sub>3</sub>-X<sub>4</sub>-  
 X<sub>5</sub>-X<sub>6</sub>-G); wherein  
 X<sub>1</sub> is Ile (I), or Leu (L);  
 X<sub>2</sub> is Ile (I), Val (V), or Thr (T);  
 X<sub>3</sub> is any amino acid other than Arg (R) or Lys (K);  
 10 X<sub>4</sub> is Thr (T), Ser (S) or Leu (L);  
 X<sub>5</sub> is Pro (P) or Glu (E); and  
 X<sub>6</sub> is Leu (L) or Glu (E) (SEQ ID NO:8).

In some embodiments, X<sub>3</sub> is selected from G, A, L, M, W, P, S, T, F, Y, C, N, Q, D, E, V, I or H (SEQ ID NO:42).

15 In some embodiments, DNA polymerases of the invention can be characterized by having the following motif (corresponding to *Thermus* and *Thermotoga*):

Pro-Asn-Leu-Gln-Asn- X<sub>1</sub>-Pro-X<sub>2</sub>-X<sub>3</sub>-X<sub>4</sub>-X<sub>5</sub>-X<sub>6</sub>-Gly (also  
 referred to herein in the one-letter code as P-N-L-Q-N-X<sub>1</sub>-P-  
 X<sub>2</sub>-X<sub>3</sub>-X<sub>4</sub>-X<sub>5</sub>-X<sub>6</sub>-G); wherein  
 20 X<sub>1</sub> is Ile (I), or Leu (L);  
 X<sub>2</sub> is Ile (I), Val (V), or Thr (T);  
 X<sub>3</sub> is any amino acid other than Arg (R) or Lys (K);  
 X<sub>4</sub> is Thr (T) or Ser (S);  
 X<sub>5</sub> is Pro (P) or Glu (E); and  
 25 X<sub>6</sub> is Leu (L) or Glu (E) (SEQ ID NO:9).

In some embodiments, DNA polymerases of the invention can be characterized by having the following motif:

Pro-Asn-Leu-Gln-Asn-Ile-Pro-X<sub>2</sub>-X<sub>3</sub>-Thr-Pro-Leu-Gly (also  
 referred to herein in the one-letter code as P-N-L-Q-N-I-P-X<sub>2</sub>-



X<sub>3</sub>-T-P-L-G); wherein

X<sub>2</sub> is Ile (I) or Val (V);

X<sub>3</sub> is any amino acid other than Arg (R) (SEQ ID NO:10).

5 In some embodiments, DNA polymerases of the invention can be characterized by having the following motif:

Pro-Asn-Leu-Gln-Asn-Ile-Pro-Ile -X<sub>a3</sub>-Thr-Pro-Leu-Gly (also referred to herein in the one-letter code as P-N-L-Q-N-I-P-I-

X<sub>3</sub>-T-P-L-G); wherein

X<sub>3</sub> is His (H), Leu (L), or Ser (S) (SEQ ID NO:11).

10 This motif is present within the “fingers” domain of many Family A type DNA-dependent DNA polymerases, particularly thermostable DNA polymerases from thermophilic bacteria (Li et al., *EMBO J.* 17:7514-7525, 1998). For example, Figure 1 shows an amino acid sequence alignment comprising the native sequence corresponding to the motif above in DNA polymerases from several species of bacteria: *Escherichia coli*, *Bacillus caldotenax*, *Bacillus stearothermophilus*,  
15 *Deinococcus radiodurans*, *Thermosiphon africanus*, *Thermotoga maritima*, *Thermotoga neopolitana*, *Thermus aquaticus*, *Thermus caldophilus*, *Thermus filiformis*, *Thermus flavus*, *Thermus* sp. Sps17, *Thermus* sp. Z05, and *Thermus thermophilus*. As shown, the motif of SEQ ID NO:8 (except where X<sub>3</sub> is R or K) is present in each of these polymerases, indicating a conserved function for this region of the polymerase. Figure 2 provides sequence identities  
20 among these DNA polymerases.

Accordingly, in some embodiments, the invention provides for a polymerase comprising SEQ ID NO:8, 9, 10, or 11 (e.g., where X<sub>3</sub> is selected, as appropriate based on the consensus sequence, from G, A, V, L, I, M, F, W, P, S, T, C, Y, N, Q, D, E, K, or H), having the improved activity and/or characteristics described herein, and wherein the DNA polymerase is otherwise a wild-  
25 type or a naturally occurring DNA polymerase, such as, for example, a polymerase from any of the species of thermophilic bacteria listed above, or is substantially identical to such a wild-type or a naturally occurring DNA polymerase. For example, in some embodiments, the polymerase of the invention comprises SEQ ID NO:8, 9, 10, or 11 and is at least 80%, 85%, 90%, or 95% identical to SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 36, 37, 38, 39, 40, or 41. In one variation, the  
30 unmodified form of the polymerase is from a species of the genus *Thermus*. In some

embodiments of the invention, the unmodified polymerase is from a thermophilic species other than *Thermus*, e.g., *Thermotoga*. The full nucleic acid and amino acid sequence for numerous thermostable DNA polymerases are available. The sequences each of *Thermus aquaticus* (Taq) (SEQ ID NO:2), *Thermus thermophilus* (Tth) (SEQ ID NO:6), *Thermus* species Z05 (SEQ ID NO:1), *Thermus* species Sps17 (SEQ ID NO:5), *Thermotoga maritima* (Tma) (SEQ ID NO:38), and *Thermosiphon africanus* (Taf) (SEQ ID NO:37) polymerase have been published in PCT International Patent Publication No. WO 92/06200. The sequence for the DNA polymerase from *Thermus flavus* (SEQ ID NO:4) has been published in Akhmetzjanov and Vakhitov (*Nucleic Acids Research* 20:5839, 1992). The sequence of the thermostable DNA polymerase from *Thermus caldophilus* (SEQ ID NO:7) is found in EMBL/GenBank Accession No. U62584. The sequence of the thermostable DNA polymerase from *Thermus filiformis* can be recovered from ATCC Deposit No. 42380 using, e.g., the methods provided in U.S. Pat. No. 4,889,818, as well as the sequence information provided in Table 1. The sequence of the *Thermotoga neapolitana* DNA polymerase (SEQ ID NO:39) is from GeneSeq Patent Data Base Accession No. R98144 and PCT WO 97/09451. The sequence of the thermostable DNA polymerase from *Bacillus caldotenax* (SEQ ID NO:41) is described in, e.g., Uemori et al. (*J Biochem (Tokyo)* 113(3):401-410, 1993; see also, Swiss-Prot database Accession No. Q04957 and GenBank Accession Nos. D12982 and BAA02361). Examples of unmodified forms of DNA polymerases that can be modified as described herein are also described in, e.g., U.S. Pat. Nos. 6,228,628; 6,346,379; 7,030,220; 6,881,559; 6,794,177; 6,468,775; and U.S. Pat. Appl. Nos. 20040005599; 20020012970; 20060078928; 20040115639. Representative full length polymerase sequences are also provided in the sequence listing.

In some embodiments, the polymerase of the invention, as well as having a polymerase domain comprising SEQ ID NOS:8, 9, 10, or 11, also comprises a nuclease domain (e.g., corresponding to positions 1 to 291 of Z05).

In some embodiments, a polymerase of the invention is a chimeric polymerase, i.e., comprising polypeptide regions from two or more enzymes. Examples of such chimeric DNA polymerases are described in, e.g., U.S. Patent No. 6,228,628. Particularly suitable are chimeric CS-family DNA polymerases, which include the CS5 (SEQ ID NO:29) and CS6 (SEQ ID NO:30) polymerases and variants thereof having substantial sequence identity or similarity to SEQ ID NO:29 or SEQ ID NO:30 (typically at least 80% sequence identity, more typically at least 90%,

and most typically at least 95% sequence identity) and can thus be modified to contain SEQ ID NO:8. The CS5 and CS6 DNA polymerases are chimeric enzymes derived from *Thermus* sp. Z05 and *Thermotoga maritima* (*Tma*) DNA polymerases. They comprise the N-terminal 5'-nuclease domain of the *Thermus* enzyme and the C-terminal 3'-5' exonuclease and the polymerase domains of the *Tma* enzyme. These enzymes have efficient reverse transcriptase activity, can extend nucleotide analog-containing primers, and can utilize alpha-phosphorothioate dNTPs, dUTP, dITP, and also fluorescein- and cyanine-dye family labeled dNTPs. The CS5 and CS6 polymerases are also efficient Mg<sup>2+</sup>-activated PCR enzymes. The CS5 and CS6 chimeric polymerases are further described in, e.g., U.S. Pat. Application Publication No. 2004/0005599.

5 In some embodiments, the polymerase of the invention comprises SEQ ID NO:8, 9, 10, or 11 and further comprises one or more additional amino acid changes (e.g., by amino acid substitution, addition, or deletion) compared to a native polymerase. In some embodiments, such polymerases retain the amino acid motif of SEQ ID NO:8 (or a motif of SEQ ID NO:9, 10 or 11), and further comprise the amino acid motif of SEQ ID NO:27 (corresponding to the D580X mutation of Z05 (SEQ ID NO:1)) as follows:

15

T-G-R-L-S-S-X<sub>7</sub>-X<sub>8</sub>-P-N-L-Q-N; wherein  
X<sub>7</sub> is Ser (S) or Thr (T); and  
X<sub>8</sub> is any amino acid other than D or E (SEQ ID NO:27)

The mutation characterized by SEQ ID NO:27 is discussed in more detail in, e.g., US Patent Publication No. 2009/0148891. In some embodiments, such functional variant polymerases typically will have substantial sequence identity or similarity to the wild-type or naturally occurring polymerase (e.g., SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 39, 40, 41, 42, 43, or 44), typically at least 80% sequence identity, more typically at least 90%, and most typically at least 95% sequence identity.

20

In some embodiments, the amino acid at position X<sub>3</sub> is substituted with an amino acid as set forth in SEQ ID NO:8, 9, 10 or 11, and the amino acid at position X<sub>8</sub> is substituted with an amino acid as set forth in SEQ ID NO:27. Thus, in some embodiments, the amino acid at position X<sub>3</sub> is any amino acid other than Arg (R) or Lys (K) and the amino acid at position X<sub>8</sub> is any amino acid other than Asp (D) or Glu (E). In some embodiments, amino acid substitutions include Leucine (L), Glycine (G), Threonine (T), Glutamine (Q), Alanine (A), Serine (S), Asparagine (N),

30

Arginine (R), and Lysine (K) at position X<sub>8</sub> of SEQ ID NO:27. In certain embodiments, amino acid substitutions independently include Histidine (H), Lysine (K) or Serine (S) at position X<sub>3</sub>, and Glycine (G) at position X<sub>8</sub>. Other suitable amino acid substitution(s) at one or more of the identified sites can be determined using, *e.g.*, known methods of site-directed mutagenesis and determination of polynucleotide extension performance in assays described further herein or otherwise known to persons of skill in the art.

Because the precise length of DNA polymerases vary, the precise amino acid positions corresponding to each of X<sub>3</sub> and X<sub>8</sub> can vary depending on the particular polymerase used. Amino acid and nucleic acid sequence alignment programs are readily available (*see, e.g.*, those referred to *supra*) and, given the particular motifs identified herein, serve to assist in the identification of the exact amino acids (and corresponding codons) for modification in accordance with the present invention. The positions corresponding to each of X<sub>3</sub> and X<sub>8</sub> are shown in Table 1 for representative chimeric thermostable DNA polymerases and thermostable DNA polymerases from exemplary thermophilic species.

**Table 1.** Amino Acid Positions Corresponding to Motif Positions X<sub>3</sub> (*e.g.*, of SEQ ID NOS:8, 9, 10, and 11) and X<sub>8</sub> (of SEQ ID NO:27) in Exemplary Polymerases.

<u>Organism or Chimeric Sequence</u>	<u>Amino Acid Position</u>	
	X <sub>3</sub>	X <sub>8</sub> (of SEQ ID NO:27)
Consensus (SEQ ID NO:)	X <sub>3</sub>	X <sub>8</sub> (of SEQ ID NO:27)
<i>T. thermophilus</i> (6)	589	580
<i>T. caldophilus</i> (7)	589	580
<i>T. sp. Z05</i> (1)	589	580
<i>T. aquaticus</i> (2)	587	578
<i>T. flavus</i> (4)	586	577
<i>T. filiformis</i> (3)	585	576
<i>T. sp. Sps17</i> (5)	585	576
<i>T. maritima</i> (38)	649	640
<i>T. neapolitana</i> (39)	649	640
<i>T. africanus</i> (37)	648	639
<i>B. caldotenax</i> (41)	630	621

<u>Organism or Chimeric Sequence</u>	<u>Amino Acid Position</u>	
Consensus (SEQ ID NO:)	X <sub>3</sub>	X <sub>8</sub> (of SEQ ID NO:27)
<i>B. stearothermophilus</i> (40)	629	620
CS5 (29)	649	640
CS6 (30)	649	640

- In some embodiments, the DNA polymerase of the present invention is derived from *Thermus sp.* Z05 DNA polymerase (SEQ ID NO:1) or a variant thereof (*e.g.*, carrying the D580G mutation or the like). As referred to above, in *Thermus sp.* Z05 DNA polymerase, position X<sub>3</sub> corresponds to Arginine (R) at position 589; position X<sub>8</sub> corresponds to Aspartate (D) at position 580. Thus, in certain variations of the invention, the mutant polymerase comprises at least one amino acid substitution, relative to a *Thermus sp.* Z05 DNA polymerase, at R589 and D580. Thus, in some embodiments, the amino acid at position 589 is not R. In some embodiments, the amino acid at position 589 is selected from G, A, V, L, I, M, F, W, P, S, T, C, Y, N, Q, D, E, K, or H. In certain embodiments, the amino acid residue at position R589 is H, L, or S. The amino acid residues at position D580 can be selected from Leucine (L), Glycine (G), Threonine (T), Glutamine (Q), Alanine (A), Serine (S), Asparagine (N), Arginine (R), and Lysine (K). Exemplary *Thermus sp.* Z05 DNA polymerase mutants include those comprising the amino acid substitution(s) R589H, R589L, or R589S, and D580G.
- In some embodiments, the Z05 DNA polymerase further comprises additional amino acid substitutions. For example, in some embodiments, the amino acid at position 517 of SEQ ID NO:1 is any amino acid other than S. In some embodiments, the amino acid at position 517 of SEQ ID NO:1 is selected from G, A, V, L, I, M, F, W, P, R, K, T, C, Y, N, Q, D, E or H. In some embodiments, the amino acid at position 517 of SEQ ID NO:1 is G. In some embodiments, the amino acid at position 770 of SEQ ID NO:1 is any amino acid other than L. In some embodiments, the amino acid at position 770 of SEQ ID NO:1 is selected from G, A, V, S, I, M, F, W, P, R, K, T, C, Y, N, Q, D, E or H. In some embodiments, the amino acid at position 770 of SEQ ID NO:1 is F. In some embodiments, the amino acid at position 794 of SEQ ID NO: 1 is any amino acid other than P. In some embodiments, the amino acid at position 794 of SEQ ID

NO:1 is selected from G, A, V, L, I, M, F, W, S, R, K, T, C, Y, N, Q, D, E or H. In some embodiments, the amino acid at position 794 of SEQ ID NO:1 is T.

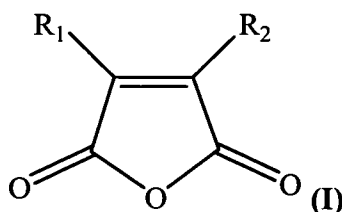
Exemplary *Thermus* sp. Z05 DNA polymerase mutants include those comprising the amino acid substitution(s) R589H, R589L, R589S, S517G, L770F, P794T, and/or D580G. In some  
5 embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, e.g., amino acid residue substitutions R589H and D580G, or R589L and D580G, or R589S and D580G. In some  
embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, e.g., amino acid residue  
substitutions R589H and S517G, or R589L and S517G, or R589S and S517G. In some  
embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, e.g., amino acid residue  
10 substitutions R589H and L770F, or R589L and L770F, or R589S and L770F. In some  
embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, e.g., amino acid residue  
substitutions R589H and P794T, or R589L and P794T, or R589S and P794T. In certain  
embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, e.g., amino acid residue  
substitutions independently selected from R589H, S517G, L770F, P794T, and/or D580G. In  
15 certain embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, e.g., amino acid  
residue substitutions independently selected from R589L, S517G, L770F, P794T, and/or D580G.  
In certain embodiments, the mutant *Thermus* sp. Z05 DNA polymerase comprises, e.g., amino  
acid residue substitutions independently selected from R589S, S517G, L770F, P794T, and/or  
D580G.

20 In some embodiments, the DNA polymerase of the invention comprises an amino acid at the  
position corresponding to position 589 of SEQ ID NO:1 that does not have a polar, positively-  
charged side-chain (e.g., K or R) at neutral pH (e.g., about pH 7.4). In some embodiments, the  
DNA polymerase of the invention is derived from a *Thermus* species, and the amino acid at the  
position corresponding to position 589 of SEQ ID NO:1 is an amino acid that does not have a  
25 polar, positively-charged side-chain (e.g., R) at neutral pH (e.g., about pH 7.4). In some  
embodiments, the amino acid at the position corresponding to position 589 of SEQ ID NO:1 is  
an amino acid having a polar or nonpolar, uncharged side-chain (e.g., H, L, or S) at neutral pH  
(e.g., about pH 7.4). In some embodiments, the amino acid at the position corresponding to  
position 589 of SEQ ID NO:1 having a polar, uncharged side-chain is H or S. It will be  
30 understood that at about pH 7.4, the side-chain of H is predominantly uncharged, with a small  
percentage of positively charged side-chains. In some embodiments, the amino acid at the

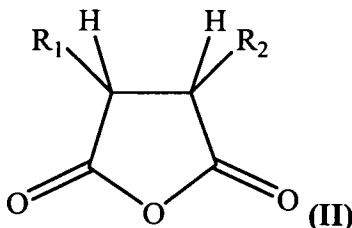
position corresponding to position 589 of SEQ ID NO:1 having a nonpolar, uncharged side-chain is L.

In some embodiments, the DNA polymerases of the present invention can also include other, non-substitutional modification(s). Such modifications can include, for example, covalent  
 5 modifications known in the art to confer an additional advantage in applications comprising polynucleotide extension. For example, in certain embodiments, the mutant DNA polymerase further includes a thermally reversible covalent modification. DNA polymerases comprising such thermally reversible modifications are particularly suitable for hot-start applications, such as, *e.g.*, various hot-start PCR techniques. Thermally reversible modifier reagents amenable to  
 10 use in accordance with the mutant DNA polymerases of the present invention are described in, for example, U.S. Patent No. 5,773,258 to Birch *et al.*

For example, particularly suitable polymerases comprising a thermally reversible covalent modification are produced by a reaction, carried out at alkaline pH at a temperature which is less than about 25 °C, of a mixture of a thermostable enzyme and a dicarboxylic acid anhydride  
 15 having a general formula as set forth in the following formula I:



where R<sub>1</sub> and R<sub>2</sub> are hydrogen or organic radicals, which may be linked; or having the following formula II:



20 where R<sub>1</sub> and R<sub>2</sub> are organic radicals, which may linked, and the hydrogens are *cis*, essentially as described in Birch *et al, supra*.

The DNA polymerases of the present invention can be constructed by mutating the DNA sequences that encode the corresponding unmodified polymerase (*e.g.*, a wild-type polymerase or a corresponding variant from which the polymerase of the invention is derived), such as by using techniques commonly referred to as site-directed mutagenesis. Nucleic acid molecules encoding the unmodified form of the polymerase can be mutated by a variety of polymerase chain reaction (PCR) techniques well-known to one of ordinary skill in the art. (*See, e.g., PCR Strategies* (M. A. Innis, D. H. Gelfand, and J. J. Sninsky eds., 1995, Academic Press, San Diego, CA) at Chapter 14; *PCR Protocols : A Guide to Methods and Applications* (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White eds., Academic Press, NY, 1990).

5

10 By way of non-limiting example, the two primer system, utilized in the Transformer Site-Directed Mutagenesis kit from Clontech, may be employed for introducing site-directed mutants into a polynucleotide encoding an unmodified form of the polymerase. Following denaturation of the target plasmid in this system, two primers are simultaneously annealed to the plasmid; one of these primers contains the desired site-directed mutation, the other contains a mutation at another

15 point in the plasmid resulting in elimination of a restriction site. Second strand synthesis is then carried out, tightly linking these two mutations, and the resulting plasmids are transformed into a mutS strain of *E. coli*. Plasmid DNA is isolated from the transformed bacteria, restricted with the relevant restriction enzyme (thereby linearizing the unmutated plasmids), and then retransformed into *E. coli*. This system allows for generation of mutations directly in an expression plasmid,

20 without the necessity of subcloning or generation of single-stranded phagemids. The tight linkage of the two mutations and the subsequent linearization of unmutated plasmids result in high mutation efficiency and allow minimal screening. Following synthesis of the initial restriction site primer, this method requires the use of only one new primer type per mutation site. Rather than prepare each positional mutant separately, a set of "designed degenerate"

25 oligonucleotide primers can be synthesized in order to introduce all of the desired mutations at a given site simultaneously. Transformants can be screened by sequencing the plasmid DNA through the mutagenized region to identify and sort mutant clones. Each mutant DNA can then be restricted and analyzed by electrophoresis, such as for example, on a Mutation Detection Enhancement gel (Mallinckrodt Baker, Inc., Phillipsburg, NJ) to confirm that no other alterations

30 in the sequence have occurred (by band shift comparison to the unmutagenized control).



Alternatively, the entire DNA region can be sequenced to confirm that no additional mutational events have occurred outside of the targeted region.

Verified mutant duplexes in pET (or other) overexpression vectors can be employed to transform *E. coli* such as, *e.g.*, strain *E. coli* BL21 (DE3) pLysS, for high level production of the mutant protein, and purification by standard protocols. The method of FAB-MS mapping, for example, can be employed to rapidly check the fidelity of mutant expression. This technique provides for sequencing segments throughout the whole protein and provides the necessary confidence in the sequence assignment. In a mapping experiment of this type, protein is digested with a protease (the choice will depend on the specific region to be modified since this segment is of prime interest and the remaining map should be identical to the map of unmutagenized protein). The set of cleavage fragments is fractionated by, for example, microbore HPLC (reversed phase or ion exchange, again depending on the specific region to be modified) to provide several peptides in each fraction, and the molecular weights of the peptides are determined by standard methods, such as FAB-MS. The determined mass of each fragment are then compared to the molecular weights of peptides expected from the digestion of the predicted sequence, and the correctness of the sequence quickly ascertained. Since this mutagenesis approach to protein modification is directed, sequencing of the altered peptide should not be necessary if the MS data agrees with prediction. If necessary to verify a changed residue, CAD-tandem MS/MS can be employed to sequence the peptides of the mixture in question, or the target peptide can be purified for subtractive Edman degradation or carboxypeptidase Y digestion depending on the location of the modification.

Mutant DNA polymerases with more than one amino acid substituted can be generated in various ways. In the case of amino acids located close together in the polypeptide chain, they may be mutated simultaneously using one oligonucleotide that codes for all of the desired amino acid substitutions. If however, the amino acids are located some distance from each other (separated by more than ten amino acids, for example) it is more difficult to generate a single oligonucleotide that encodes all of the desired changes. Instead, one of two alternative methods may be employed. In the first method, a separate oligonucleotide is generated for each amino acid to be substituted. The oligonucleotides are then annealed to the single-stranded template DNA simultaneously, and the second strand of DNA that is synthesized from the template will encode all of the desired amino acid substitutions. An alternative method involves two or more

rounds of mutagenesis to produce the desired mutant. The first round is as described for the single mutants: DNA encoding the unmodified polymerase is used for the template, an oligonucleotide encoding the first desired amino acid substitution(s) is annealed to this template, and the heteroduplex DNA molecule is then generated. The second round of mutagenesis utilizes the mutated DNA produced in the first round of mutagenesis as the template. Thus, this template already contains one or more mutations. The oligonucleotide encoding the additional desired amino acid substitution(s) is then annealed to this template, and the resulting strand of DNA now encodes mutations from both the first and second rounds of mutagenesis. This resultant DNA can be used as a template in a third round of mutagenesis, and so on. Alternatively, the multi-site mutagenesis method of Seyfang & Jin (*Anal. Biochem.* 324:285-291. 2004) may be utilized.

Accordingly, also provided are recombinant nucleic acids encoding any of the DNA polymerases of the present invention (e.g., polymerases comprising any of SEQ ID NOS:8, 9, 10, or 11). Using a nucleic acid of the present invention, encoding a DNA polymerase of the invention, a variety of vectors can be made. Any vector containing replicon and control sequences that are derived from a species compatible with the host cell can be used in the practice of the invention. Generally, expression vectors include transcriptional and translational regulatory nucleic acid regions operably linked to the nucleic acid encoding the mutant DNA polymerase. The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. In addition, the vector may contain a Positive Retroregulatory Element (PRE) to enhance the half-life of the transcribed mRNA (*see Gelfand et al. U.S. Patent No. 4,666,848*). The transcriptional and translational regulatory nucleic acid regions will generally be appropriate to the host cell used to express the polymerase. Numerous types of appropriate expression vectors, and suitable regulatory sequences are known in the art for a variety of host cells. In general, the transcriptional and translational regulatory sequences may include, e.g., promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. In typical embodiments, the regulatory sequences include a promoter and transcriptional start and stop sequences. Vectors also typically include a polylinker region containing several restriction sites for insertion of foreign DNA. In certain embodiments, "fusion flags" are used to facilitate purification and, if desired, subsequent

removal of tag/flag sequence, e.g., "His-Tag". However, these are generally unnecessary when purifying an thermoactive and/or thermostable protein from a mesophilic host (e.g., *E. coli*) where a "heat-step" may be employed. The construction of suitable vectors containing DNA encoding replication sequences, regulatory sequences, phenotypic selection genes, and the mutant polymerase of interest are prepared using standard recombinant DNA procedures.

Isolated plasmids, viral vectors, and DNA fragments are cleaved, tailored, and ligated together in a specific order to generate the desired vectors, as is well-known in the art (*see, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, NY, 2nd ed. 1989)).

10 In certain embodiments, the expression vector contains a selectable marker gene to allow the selection of transformed host cells. Selection genes are well known in the art and will vary with the host cell used. Suitable selection genes can include, for example, genes coding for ampicillin and/or tetracycline resistance, which enables cells transformed with these vectors to grow in the presence of these antibiotics.

15 In one aspect of the present invention, a nucleic acid encoding a DNA polymerase of the invention is introduced into a cell, either alone or in combination with a vector. By "introduced into" or grammatical equivalents herein is meant that the nucleic acids enter the cells in a manner suitable for subsequent integration, amplification, and/or expression of the nucleic acid. The method of introduction is largely dictated by the targeted cell type. Exemplary methods include  
20 CaPO<sub>4</sub> precipitation, liposome fusion, LIPOFECTIN®, electroporation, viral infection, and the like.

In some embodiments, prokaryotes are used as host cells for the initial cloning steps of the present invention. They are particularly useful for rapid production of large amounts of DNA, for production of single-stranded DNA templates used for site-directed mutagenesis, for screening  
25 many mutants simultaneously, and for DNA sequencing of the mutants generated. Suitable prokaryotic host cells include *E. coli* K12 strain 94 (ATCC No. 31,446), *E. coli* strain W3110 (ATCC No. 27,325), *E. coli* K12 strain DG116 (ATCC No. 53,606), *E. coli* X1776 (ATCC No. 31,537), and *E. coli* B; however many other strains of *E. coli*, such as HB101, JM101, NM522, NM538, NM539, and many other species and genera of prokaryotes including bacilli such as  
30 *Bacillus subtilis*, other enterobacteriaceae such as *Salmonella typhimurium* or *Serratia marcesans*, and various *Pseudomonas* species can all be used as hosts. Prokaryotic host cells or

other host cells with rigid cell walls are typically transformed using the calcium chloride method as described in section 1.82 of Sambrook *et al.*, *supra*. Alternatively, electroporation can be used for transformation of these cells. Prokaryote transformation techniques are set forth in, for example Dower, in *Genetic Engineering, Principles and Methods* 12:275-296 (Plenum Publishing Corp., 1990); Hanahan *et al.*, *Meth. Enzymol.*, 204:63, 1991. Plasmids typically used for transformation of *E. coli* include pBR322, pUCI8, pUCI9, pUCII8, pUC119, and Bluescript M13, all of which are described in sections 1.12-1.20 of Sambrook *et al.*, *supra*. However, many other suitable vectors are available as well.

In some embodiments, the DNA polymerases of the present invention are produced by culturing a host cell transformed with an expression vector containing a nucleic acid encoding the DNA polymerase, under the appropriate conditions to induce or cause expression of the DNA polymerase. Methods of culturing transformed host cells under conditions suitable for protein expression are well-known in the art (*see, e.g.*, Sambrook *et al.*, *supra*). Suitable host cells for production of the polymerases from lambda pL promotor-containing plasmid vectors include *E. coli* strain DG116 (ATCC No. 53606) (*see* US Pat. No. 5,079,352 and Lawyer, F.C. *et al.*, *PCR Methods and Applications* 2:275-87, 1993). Following expression, the polymerase can be harvested and isolated. Methods for purifying the thermostable DNA polymerase are described in, for example, Lawyer *et al.*, *supra*.

Once purified, a DNA polymerase's 3' mismatch discrimination can be assayed. For example, in some embodiments, 3' mismatch discrimination activity is determined by comparing the amplification of a target sequence perfectly matched to the primer to amplification of a target that has a single base mismatch at the 3' end of the primer. Amplification can be detected, for example, in real time by use of TaqMan™ probes. Ability of a polymerase to distinguish between the two target sequences can be estimated by comparing the Cps of the two reactions. Optionally, simultaneous amplification of a second target gene in each well can be performed and detected in a second optical channel as a control. "Delta Cp values" refer to the difference in value between the Cp associated with the mismatched template minus the Cp of the matched target (*see, e.g.*, the Examples). In some embodiments, the improved polymerases of the invention have a delta Cp value of at least 1, 2, 3, 4, 5, or more compared to an otherwise identical control polymerase having a native amino acid (*e.g.*, R) at position X<sub>3</sub> of SEQ ID NO:8.

In some embodiments, this determination is made with the precise materials and conditions set forth in the Examples.

### ***Methods of the invention***

The improved DNA polymerases of the present invention may be used for any purpose in which such enzyme activity is necessary or desired. The improved DNA polymerase can be a thermoactive or thermostable DNA polymerase, as described herein. Accordingly, in one aspect of the invention, methods of polynucleotide extension, including PCR, using the polymerases of the invention are provided. In some embodiments, the invention provides a thermoactive DNA polymerase that is useful to extend an RNA or DNA template when amplification of the template nucleic acid is not required, for example, when it is desired to immediately detect the presence of a target nucleic acid. In some embodiments, the invention provides a thermostable DNA polymerase that is useful when it is desired to extend and/or amplify a target nucleic acid. Conditions suitable for polynucleotide extension are known in the art. (*See, e.g., Sambrook et al., supra. See also Ausubel et al., Short Protocols in Molecular Biology* (4th ed., John Wiley & Sons 1999). Generally, a primer is annealed, *i.e.*, hybridized, to a target nucleic acid to form a primer-template complex. The primer-template complex is contacted with the mutant DNA polymerase and nucleoside triphosphates in a suitable environment to permit the addition of one or more nucleotides to the 3' end of the primer, thereby producing an extended primer complementary to the target nucleic acid. The primer can include, *e.g.*, one or more nucleotide analog(s). In addition, the nucleoside triphosphates can be conventional nucleotides, unconventional nucleotides (*e.g.*, ribonucleotides or labeled nucleotides), or a mixture thereof. In some variations, the polynucleotide extension reaction comprises amplification of a target nucleic acid. Conditions suitable for nucleic acid amplification using a DNA polymerase and a primer pair are also known in the art (*e.g.*, PCR amplification methods). (*See, e.g., Sambrook et al., supra; Ausubel et al., supra; PCR Applications: Protocols for Functional Genomics* (Innis et al. eds., Academic Press 1999).

In some embodiments, use of the present polymerases, which provide increased 3' mismatch discrimination, allow for, *e.g.*, rare allele detection. For example, the fidelity of 3' mismatch discrimination of a particular polymerase sets its sensitivity (ability to accurately detect small quantities of a target sequence in the presence of larger quantities of a different but related non-target sequence). Thus, increased 3'- mismatch discrimination results in greater sensitivity for

detection of rare alleles. Rare allele detection is useful, for example, when screening biopsies or other samples for rare genetic changes, e.g., a cell carrying a cancer allele in a mass of normal cells.

- In some embodiments, the improved polymerases are used for polynucleotide extension in the context of allele specific PCR or single nucleotide polymorphism (SNP) detection. Exemplary SNP detection methods are described in Chen et al., "Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput" *Pharmacogenomics J.* 3(2):77-96 (2003); Kwok et al., "Detection of single nucleotide polymorphisms" *Curr. Issues Mol. Biol.* 5(2):43-60 (April 2003); Shi, "Technologies for individual genotyping: detection of genetic polymorphisms in drug targets and disease genes" *Am. J. Pharmacogenomics* 2(3):197-205 (2002); and Kwok, "Methods for genotyping single nucleotide polymorphisms" *Annu. Rev. Genomics Hum. Genet.* 2:235-58 (2001). Exemplary techniques for high-throughput SNP detection are described in Marnellos, "High-throughput SNP analysis for genetic association studies" *Curr. Opin. Drug Discov. Devel.* 6(3):317-21 (May 2003). Common SNP detection methods include, but are not limited to, TaqMan assays, molecular beacon assays, nucleic acid arrays, allele-specific primer extension, allele-specific PCR, arrayed primer extension, homogeneous primer extension assays, primer extension with detection by mass spectrometry, pyrosequencing, multiplex primer extension sorted on genetic arrays, ligation with rolling circle amplification, homogeneous ligation, OLA (U.S. Pat. No. 4,988,167), multiplex ligation reaction sorted on genetic arrays, restriction-fragment length polymorphism, single base extension-tag assays, and the Invader assay. Such methods may be used in combination with detection mechanisms such as, for example, luminescence or chemiluminescence detection, fluorescence detection, time-resolved fluorescence detection, fluorescence resonance energy transfer, fluorescence polarization, mass spectrometry, and electrical detection.
- Detection of multiple different alleles can also be accomplished using multiplex reactions, which allow the detection of multiple different alleles in a single reaction. In multiplex reactions, two or more allele-specific primers are used to extend and amplify SNPs or multiple nucleotide polymorphisms or alleles. Exemplary methods for multiplex detection of single and multiple nucleotide polymorphisms are described in U.S. Patent Publication No. 2006/0172324.
- Other methods for detecting extension products or amplification products using the improved polymerases described herein include the use of fluorescent double-stranded nucleotide binding

dyes or fluorescent double-stranded nucleotide intercalating dyes. Examples of fluorescent double-stranded DNA binding dyes include SYBR-green (Molecular Probes). Examples of fluorescent double-stranded intercalating dyes include ethidium bromide. The double stranded DNA binding dyes can be used in conjunction with melting curve analysis to measure primer extension products and/or amplification products. The melting curve analysis can be performed on a real-time PCR instrument, such as the ABI 5700/7000 (96 well format) or ABI 7900 (384 well format) instrument with onboard software (SDS 2.1). Alternatively, the melting curve analysis can be performed as an end point analysis. Exemplary methods of melting point analysis are described in U.S. Patent Publication No. 2006/0172324.

10 In yet other embodiments, the polymerases of the invention are used for primer extension in the context of DNA sequencing, DNA labeling, or labeling of primer extension products. For example, DNA sequencing by the Sanger dideoxynucleotide method (Sanger *et al.*, *Proc. Natl. Acad. Sci. USA* 74: 5463, 1977) is improved by the present invention for polymerases capable of incorporating unconventional, chain-terminating nucleotides. Advances in the basic Sanger *et al.* method have provided novel vectors (Yanisch-Perron *et al.*, *Gene* 33:103-119, 1985) and base analogues (Mills *et al.*, *Proc. Natl. Acad. Sci. USA* 76:2232-2235, 1979; and Barr *et al.*, *Biotechniques* 4:428-432, 1986). In general, DNA sequencing requires template-dependent primer extension in the presence of chain-terminating base analogs, resulting in a distribution of partial fragments that are subsequently separated by size. The basic dideoxy sequencing procedure involves (i) annealing an oligonucleotide primer, optionally labeled, to a template; (ii) extending the primer with DNA polymerase in four separate reactions, each containing a mixture of unlabeled dNTPs and a limiting amount of one chain terminating agent such as a ddNTP, optionally labeled; and (iii) resolving the four sets of reaction products on a high-resolution denaturing polyacrylamide/urea gel. The reaction products can be detected in the gel by autoradiography or by fluorescence detection, depending on the label used, and the image can be examined to infer the nucleotide sequence. These methods utilize DNA polymerase such as the Klenow fragment of *E. coli* Pol I or a modified T7 DNA polymerase.

The availability of thermostable polymerases, such as Taq DNA polymerase, resulted in improved methods for sequencing with thermostable DNA polymerase (*see* Innis *et al.*, *Proc. Natl. Acad. Sci. USA* 85:9436, 1988) and modifications thereof referred to as "cycle sequencing" (Murray, *Nuc Acids Res.* 17:8889, 1989). Accordingly, thermostable polymerases of the present

invention can be used in conjunction with such methods. As an alternative to basic dideoxy sequencing, cycle sequencing is a linear, asymmetric amplification of target sequences complementary to the template sequence in the presence of chain terminators. A single cycle produces a family of extension products of all possible lengths. Following denaturation of the extension reaction product from the DNA template, multiple cycles of primer annealing and primer extension occur in the presence of terminators such as ddNTPs. Cycle sequencing requires less template DNA than conventional chain-termination sequencing. Thermostable DNA polymerases have several advantages in cycle sequencing; they tolerate the stringent annealing temperatures which are required for specific hybridization of primer to nucleic acid targets as well as tolerating the multiple cycles of high temperature denaturation which occur in each cycle, *e.g.*, 90-95 °C. For this reason, AMPLITAQ® DNA Polymerase and its derivatives and descendants, *e.g.*, AmpliTaq CS DNA Polymerase and AmpliTaq FS DNA Polymerase have been included in Taq cycle sequencing kits commercialized by companies such as Perkin-Elmer (Norwalk, CT) and Applied Biosystems (Foster City, CA).

The improved polymerases find use in 454 sequencing (Roche) (Margulies, M et al. 2005, *Nature*, 437, 376-380). 454 sequencing involves two steps. In the first step, DNA is sheared into fragments of approximately 300-800 base pairs, and the fragments are blunt ended. Oligonucleotide adaptors are then ligated to the ends of the fragments. The adaptors serve as primers for amplification and sequencing of the fragments. The fragments can be attached to DNA capture beads, *e.g.*, streptavidin-coated beads using, *e.g.*, Adaptor B, which contains 5'-biotin tag. The fragments attached to the beads are PCR amplified within droplets of an oil-water emulsion. The result is multiple copies of clonally amplified DNA fragments on each bead. In the second step, the beads are captured in wells (pico-liter sized). Pyrosequencing is performed on each DNA fragment in parallel. Addition of one or more nucleotides generates a light signal that is recorded by a CCD camera in a sequencing instrument. The signal strength is proportional to the number of nucleotides incorporated.

Pyrosequencing makes use of pyrophosphate (PPi) which is released upon nucleotide addition. PPi is converted to ATP by ATP sulfurylase in the presence of adenosine 5' phosphosulfate. Luciferase uses ATP to convert luciferin to oxyluciferin, and this reaction generates light that is detected and analyzed.



Variations of chain termination sequencing methods include dye-primer sequencing and dye-terminator sequencing. In dye-primer sequencing, the ddNTP terminators are unlabeled, and a labeled primer is utilized to detect extension products (Smith *et al.*, *Nature* 32:674-679, 1986). In dye-terminator DNA sequencing, a DNA polymerase is used to incorporate dNTPs and fluorescently labeled ddNTPs onto the end of a DNA primer (Lee *et al.*, *Nuc. Acids. Res.* 20:2471, 1992). This process offers the advantage of not having to synthesize dye labeled primers. Furthermore, dye-terminator reactions are more convenient in that all four reactions can be performed in the same tube.

Both dye-primer and dye-terminator methods may be automated using an automated sequencing instrument produced by Applied Biosystems (Foster City, CA) (U.S. Pat. No. 5,171,534). When using the instrument, the completed sequencing reaction mixture is fractionated on a denaturing polyacrylamide gel or capillaries mounted in the instrument. A laser at the bottom of the instrument detects the fluorescent products as they are electrophoretically separated according to size through the gel.

Two types of fluorescent dyes are commonly used to label the terminators used for dye-terminator sequencing—negatively charged and zwitterionic fluorescent dyes. Negatively charged fluorescent dyes include those of the fluorescein and BODIPY families. BODIPY dyes (4,4-difluoro-4-bora-3a,4a-diaza-s-indacene) are described in International Patent Publication WO 97/00967. Zwitterionic fluorescent dyes include those of the rhodamine family. Commercially available cycle sequencing kits use terminators labeled with rhodamine derivatives. However, the rhodamine-labeled terminators are rather costly and the product must be separated from unincorporated dye-ddNTPs before loading on the gel since they co-migrate with the sequencing products. Rhodamine dye family terminators seem to stabilize hairpin structures in GC-rich regions, which causes the products to migrate anomalously. This can involve the use of dITP, which relaxes the secondary structure but also affects the efficiency of incorporation of terminator.

In contrast, fluorescein-labeled terminators eliminate the separation step prior to gel loading since they have a greater net negative charge and migrate faster than the sequencing products. In addition, fluorescein-labeled sequencing products have better electrophoretic migration than sequencing products labeled with rhodamine. Although wild-type Taq DNA polymerase does not efficiently incorporate terminators labeled with fluorescein family dyes, this can now be

accomplished efficiently by use of the modified enzymes as described in U.S. Patent Application Publication No. 2002/0142333. Accordingly, modifications as described in US 2002/0142333 can be used in the context of the present invention to produce fluorescein-family-dye-incorporating thermostable polymerases having improved primer extension rates. For example, in certain embodiments, the unmodified DNA polymerase in accordance with the present invention is a modified thermostable polymerase as described in US 2002/0142333 and having the motif set forth in SEQ ID NO:8 (or a motif of SEQ ID NO:9, 10 or 11), and optionally the motif of SEQ ID NO:27.

Other exemplary nucleic acid sequencing formats in which the mutant DNA polymerases of the invention can be used include those involving terminator compounds that include 2'-PO<sub>4</sub> analogs of ribonucleotides (*see, e.g.*, U.S. Application Publication Nos. 2005/0037991 and 2005/0037398, and US Patent Application No. 12/174,488).

### ***Kits***

In another aspect of the present invention, kits are provided for use in primer extension methods described herein. In some embodiments, the kit is compartmentalized for ease of use and contains at least one container providing a DNA polymerase of the invention having increased 3' mismatch discrimination in accordance with the present invention. One or more additional containers providing additional reagent(s) can also be included. Such additional containers can include any reagents or other elements recognized by the skilled artisan for use in primer extension procedures in accordance with the methods described above, including reagents for use in, *e.g.*, nucleic acid amplification procedures (*e.g.*, PCR, RT-PCR), DNA sequencing procedures, or DNA labeling procedures. For example, in certain embodiments, the kit further includes a container providing a 5' sense primer hybridizable, under primer extension conditions, to a predetermined polynucleotide template, or a primer pair comprising the 5' sense primer and a corresponding 3' antisense primer. In some embodiments, the kit includes one or more containers containing one or more primers that are fully complementary to single nucleotide polymorphisms or multiple nucleotide polymorphisms, wherein the primers are useful for multiplex reactions, as described above. In other, non-mutually exclusive variations, the kit includes one or more containers providing nucleoside triphosphates (conventional and/or unconventional). In specific embodiments, the kit includes alpha-phosphorothioate dNTPs, dUTP, dITP, and/or labeled dNTPs such as, *e.g.*, fluorescein- or cyanin-dye family dNTPs. In still other, non-mutually

exclusive embodiments, the kit includes one or more containers providing a buffer suitable for a primer extension reaction. In some embodiments, the kit includes one or more labeled or unlabeled probes. Examples of probes include dual-labeled FRET (fluorescence resonance energy transfer) probes and molecular beacon probes. In another embodiment, the kit contains an aptamer, e.g., for hot start PCR assays.

### ***Reaction Mixtures***

In another aspect of the present invention, reaction mixtures are provided comprising the polymerases with increased 3'-mismatch discrimination activity, as described herein. The reaction mixtures can further comprise reagents for use in, e.g., nucleic acid amplification procedures (e.g., PCR, RT-PCR), DNA sequencing procedures, or DNA labeling procedures. For example, in certain embodiments, the reaction mixtures comprise a buffer suitable for a primer extension reaction. The reaction mixtures can also contain a template nucleic acid (DNA and/or RNA), one or more primer or probe polynucleotides, nucleoside triphosphates (including, e.g., deoxyribonucleotides, ribonucleotides, labeled nucleotides, unconventional nucleotides), salts (e.g.,  $Mn^{2+}$ ,  $Mg^{2+}$ ), and labels (e.g., fluorophores). In some embodiments, the reaction mixture further comprises double stranded DNA binding dyes, such as SYBR green, or double stranded DNA intercalating dyes, such as ethidium bromide. In some embodiments, the reaction mixtures contain a 5'-sense primer hybridizable, under primer extension conditions, to a predetermined polynucleotide template, or a primer pair comprising the 5'-sense primer and a corresponding 3' antisense primer. In certain embodiments, the reaction mixture further comprises a fluorogenic FRET hydrolysis probe for detection of amplified template nucleic acids, for example a Taqman® probe. In some embodiments, the reaction mixture contains two or more primers that are fully complementary to single nucleotide polymorphisms or multiple nucleotide polymorphisms. In some embodiments, the reaction mixtures contain alpha-phosphorothioate dNTPs, dUTP, dITP, and/or labeled dNTPs such as, e.g., fluorescein- or cyanin-dye family dNTPs.

## EXAMPLES

The following examples are offered to illustrate, but not to limit the claimed invention.

### **Example 1: Identification of mutant DNA polymerases with increased 3'-mismatch discrimination**

The control DNA polymerase of this example is a *Thermus sp.* Z05 DNA polymerase of SEQ ID NO:1 except that the amino acid at position 580 is glycine (e.g., a D580G substitution) (hereinafter Z05 D580G polymerase).

Mutations in Z05 D580G polymerase were identified that provide a reduced ability to extend an oligonucleotide primer with a 3'-mismatch to a template. In brief, the steps in this screening process included library generation, expression and partial purification of the mutant enzymes, screening of the enzymes for the desired property, DNA sequencing, clonal purification, and further characterization of selected candidate mutants. Each of these steps is described further below.

**Clonal Library generation:** A nucleic acid encoding the polymerase domain of Z05 D580G DNA polymerase was subjected to error-prone (mutagenic) PCR between Bln I and Bgl II restriction sites of a plasmid including this nucleic acid sequence. The amplified sequence is provided as SEQ ID NO:33. The primers used for this are given below:

Forward Primer: 5'- CTACCTCCTGGACCCCTCCAA-3' (SEQ ID NO:31); and,

Reverse Primer: 5'- ATAACCAACTGGTAGTGGCGTGTA-3' (SEQ ID NO:32).

PCR was performed using a range of  $Mg^{2+}$  concentrations from 1.8-3.6 mM, in order to generate libraries with a range of mutation rates. Buffer conditions were 50 mM Bicine pH 8.2, 115 mM KOAc, 8% w/v glycerol, and 0.2 mM each dNTPs. A GeneAmp® AccuRT Hot Start PCR enzyme was used at 0.15 U/  $\mu$ L. Starting with  $5 \times 10^5$  copies of linearized Z05 D580G plasmid DNA per reaction volume of 50  $\mu$ L, reactions were denatured using a temperature of 94 °C for 60 seconds, then 30 cycles of amplification were performed, using a denaturation temperature of 94 °C for 15 seconds, an annealing temperature of 60 °C for 15 seconds, an extension

temperature of 72 °C for 120 seconds, and followed by a final extension at a temperature of 72 °C for 5 minutes.

The resulting amplicon was purified with a QIAquick PCR Purification Kit (Qiagen, Inc., Valencia, CA, USA) and cut with Bsp I and Bgl II, and then re-purified with a QIAquick PCR Purification Kit. A Z05 D580G vector plasmid was prepared by cutting with the same two restriction enzymes and treating with alkaline phosphatase, recombinant (RAS, cat# 03359123001) and purified with a QIAquick PCR Purification Kit. The cut vector and the mutated insert were mixed at a 1:3 ratio and treated with T4 DNA ligase for 5 minutes at room temperature (NEB Quick Ligation™ Kit). The ligations were purified with a QIAquick PCR Purification Kit and transformed into an *E.coli* host strain by electroporation.

Aliquots of the expressed cultures were plated on ampicillin-selective medium in order to determine the number of unique transformants in each transformation. Transformations were stored at -70 °C to -80 °C in the presence of glycerol as a cryo-protectant.

Each library was then spread on large format ampicillin-selective agar plates. Individual colonies were transferred to 384-well plates containing 2X Luria broth with ampicillin and 10% w/v glycerol using an automated colony picker (QPix2, Genetix Ltd). These plates were incubated overnight at 30 °C to allow the cultures to grow and then stored at -70 °C to -80 °C. The glycerol added to the 2X Luria broth was low enough to permit culture growth and yet high enough to provide cryo-protection. Several thousand colonies at several mutagenesis ( $Mg^{2+}$ ) levels were prepared in this way for later use.

***Extract library preparation Part 1—Fermentation:*** From the clonal libraries described above, a corresponding library of partially purified extracts suitable for screening purposes was prepared. The first step of this process was to make small-scale expression cultures of each clone. These cultures were grown in 96-well format; therefore there were 4 expression culture plates for each 384-well library plate. 0.5 µL was transferred from each well of the clonal library plate to a well of a 96 well seed plate, containing 150 µL of Medium A (see Table 3 below). This seed plate was shaken overnight at 1150 rpm at 30 °C, in an iEMS plate incubator/shaker (ThermoElectron). These seed cultures were then used to inoculate the same medium, this time inoculating 20 µL into 250 µL Medium A in large format 96 well plates (Nunc # 267334). These plates were incubated overnight at 37 °C with shaking. The expression plasmid contained

transcriptional control elements, which allow for expression at 37 °C but not at 30 °C. After overnight incubation, the cultures expressed the clone protein at typically 1-10% of total cell protein. The cells from these cultures were harvested by centrifugation. These cells were either frozen (-20 °C) or processed immediately, as described below.

5

**Table 2.** Medium A (Filter-sterilized prior to use)

Component	Concentration
MgSO <sub>4</sub> ·7H <sub>2</sub> O	0.2 g/L
Citric acid·H <sub>2</sub> O	2 g/L
K <sub>2</sub> HPO <sub>4</sub>	10 g/L
NaNH <sub>4</sub> PO <sub>4</sub> ·4H <sub>2</sub> O	3.5 g/L
MgSO <sub>4</sub>	2 mM
Casamino acids	2.5 g/L
Glucose	2 g/L
Thiamine·HCl	10 mg/L
Ampicillin	100 mg/L

**Extract library preparation Part 2—Extraction:** Cell pellets from the fermentation step were resuspended in 25 µL Lysis buffer (Table 3 below) and transferred to 384-well thermocycler plates and sealed. Note that the buffer contained lysozyme to assist in cell lysis, and DNase to remove DNA from the extract. To lyse the cells the plates were incubated at 37 °C for 15 minutes, frozen overnight at -20 °C, and incubated again at 37 °C for 15 minutes. Ammonium sulfate was added (1.5 µL of a 2 M solution) and the plates incubated at 75 °C for 15 minutes in order to precipitate and inactivate contaminating proteins, including the exogenously added nucleases. The plates were centrifuged at 3000 x g for 15 minutes at 4 °C and the supernatants transferred to a fresh 384-well thermocycler plate. These extract plates were frozen at -20 °C for later use in screens. Each well contained about 0.5-3 µM of the mutant library polymerase enzyme.

10

15

**Table 3. Lysis Buffer**

Component	Concentration or Percentage
Tris pH 7.5	50 mM
EDTA	1 mM
MgCl <sub>2</sub>	6 mM
Tween 20	0.5% v/v
Lysozyme (from powder)	1 mg/mL
DNase I	0.05 Units/ $\mu$ L

**Screening extract libraries for reduced 3' primer mismatch extension rate:** The extract library was screened by comparing the extension rate of a primer perfectly matched to an oligonucleotide template vs. the extension rate of a primer with a 3' G:T mismatch.

The enzyme extracts above were diluted 10-fold for primer extension reactions by combining 2.5  $\mu$ l extract with 22.5  $\mu$ l of a buffer containing 20 mM Tris-HCl, pH 8, 100 mM KCl, 0.1 mM EDTA, and 0.2% Tween-20 in a 384-well thermocycler plate, covering and heating for 10 minutes at 90 °C. Control reactions with perfect match primer combined 0.5  $\mu$ l of the diluted extract with 15  $\mu$ l master mix in 384-well PCR plates. Extension of the primed template was monitored every 10 seconds in a modified kinetic thermal cycler using a CCD camera (see, Watson, *supra*). Master mix contained 50 nM primed primer template, 25 mM Tricine, pH 8.3, 100 mM KOAc, 0.6X SYBR Green I, 200  $\mu$ M each dNTP, 100 nM Aptamer, and 2.5 mM Magnesium Acetate. In order to distinguish extension-derived fluorescence from background fluorescence, parallel wells were included in the experiment in which primer strand extension was prevented by leaving out the nucleotides from the reaction master mix. Reactions with the 3'-mismatched primer were performed as above except 1.5  $\mu$ l of the diluted extract was added to each reaction and 1.5 mM Manganese Acetate was substituted for the Magnesium Acetate. Increasing the amount of extract three fold and using Manganese as the metal activator both make mismatch extension more likely and therefore improve the selectivity of the screen for those enzymes with the greatest ability to discriminate against 3'-mismatch extension.

Approximately 5000 mutant extracts were screened using the above protocol. Approximately 7% of the original pool was chosen for rescreening based on a perfect match primer extension value above an arbitrary cutoff and low mismatch to perfect match extension ratio. Culture wells corresponding to the top extracts were sampled to fresh growth medium and re-grown to produce a new culture plates containing the best mutants, as well as a number of parental cultures to be used for comparison. These culture plates were then used to make fresh extracts which were rescreened to confirm the original screen phenotype. The primer extension rates for the reactions with the perfect 3'-matched and the 3'-mismatched primers were calculated as the slope of the rise in fluorescence over time for the linear portion of the curve. The ratio of mismatched extension slope divided by the perfect matched extension slope was used to rank and select the best candidates. Selected clones from the rescreening, plus for comparison the parental clone Z05 D580G, with their respective genotypes and phenotypes are included in the table below.

**Table 4.**

Enzyme	Perfect Match Slope	Mismatch Slope	MM Slope / PM Slope
Z05 D580G	8.29	8.04	0.97
Z05 D580G R589H	8.13	0.50	0.06
Z05 D580G S517G R589L	19.80	3.08	0.16
Z05 D580G R589S L770F P794T	12.00	0.83	0.07

15 **Example 2: Amplification of mutant BRAF plasmid template in a background of wild-type BRAF human genomic template**

The control DNA polymerase of this example is a *Thermus sp.* Z05 DNA polymerase of SEQ ID NO:1 except that the amino acid at position 580 is Glycine (e.g., a D580G substitution) (hereinafter Z05 D580G polymerase).

20 Purified Z05 D580G R589H, Z05 D580G S517G R589L, and Z05 D580G R589S L770F P794T were compared to the parental enzyme Z05 D580G in TaqMan PCR for improved discrimination of a mutant BRAF V600R target in a background of wild-type Human Genomic DNA.



Primers were used that amplify a region of the human BRAF gene and are perfectly matched to the target when said target carries a mutation in codon 600 of BRAF, V600K. Against wild-type BRAF target, present in human genomic DNA, the allele selective primer results in a single A:C mismatch at the 3' end. The common primer is perfectly matched to the BRAF gene, as is the probe sequence, which allows for real-time, TaqMan detection of amplification. Each reaction had 10,000 copies (33 ng) of wild-type Human Genomic cell line DNA, or either 10,000 or 100 copies of a linearized plasmid containing the BRAF V600R mutant sequence in a final volume of 16  $\mu$ l. To allow for the different salt optima of the enzymes, amplifications were performed using a range of KCl concentrations from 25 to 130 mM. Buffer conditions were 50 mM Tris-HCl pH 8.0, 2.5 mM MgCl<sub>2</sub>, 0.2 mM each dNTP, 0.02 U/ $\mu$ l UNG, and 200 nM Aptamer. Forward and Reverse primers were at 100 nM and the probe was at 25 nM. All DNA polymerases were assayed at 20 nM and add 2 % (v/v) enzyme storage buffer (50% v/v glycerol, 100 mM KCl, 20 mM Tris pH 8.0, 0.1 mM EDTA, 1 mM DTT, 0.5% Tween 20) to the reactions. The reactions were performed in a Roche LightCycler 480 thermal cycler and denatured using a temperature of 95 °C for 60 seconds, then 99 cycles of amplification were performed, using a denaturation temperature of 92 °C for 10 seconds and an annealing temperature of 62 °C for 30 seconds.

Reactions were run in duplicate, crossing points ("Cps") were calculated by the Abs Quant/2<sup>nd</sup> derivative Max method and the Cps were averaged. The averaged Cp values are shown in the table below as well as calculated PCR efficiency and discrimination factor values at the KCl concentration for each enzyme which resulted in the earliest high copy mutant Cp. High Copy delta Cp is equal to the difference between the average Cp values of the reactions with 10,000 copy of 3'-mismatched wild-type genomic target and the average Cp values of the reactions with 10,000 copy of perfect match plasmid target in a background of 10,000 copy of 3'-mismatched wild-type genomic target. All reactions have a background of 10,000 copy wild type BRAF target, therefore the Cps of the reactions with no mutant plasmid represent breakthrough amplification of the mismatched primer template and the limit of discrimination for that enzyme under the condition tested. Z05 D580G R589H, Z05 D580G S517G R589L, and Z05 D580G R589S L770F P794T showed better discrimination than the parental Z05 D580G.

**Table 5. Cps of Amplification of BRAF V600K mutant plasmid vs. Human genomic DNA**

Enzyme	Optimum KCL (mM)	0 copies mutant plasmid	100 copies mutant plasmid	10,000 copies mutant plasmid	% PCR Efficie ncy	Discrimin ation Factor	High copy $\Delta$ Cp
Z05 D580G	120	34.0	32.2	26.1	110	2.6	8
Z05 D580G R589H	100	38.1	32.9	26.5	105	3.6	12
Z05 D580G S517G R589L	120	45.7	33.1	26.4	99	5.7	19
Z05 D580G R589S L770F P794T	80	42.2	33.7	26.2	86	4.3	16

This example demonstrates that the R589H, R589L, and R589S mutant enzymes have improved rare allele detection relative to the control parental enzyme, Z05 D580G.

- 5 It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art.

**INFORMAL SEQUENCE LISTING**

**SEQ ID NO:1 *Thermus sp.* Z05 DNA polymerase (Z05)**

MKAMLPLEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQAVYGF AKSLLKALKEDGYKAVF  
 VVFDKAPSF RHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGFEADDVLATLAK  
 5 KAEREGYEVRIILTADRDL YQLVSDRVAVLHPEGHLITPEWLWEKYGLKPEQWVDFRALVGDPSD  
 NLPGVKIG EKTALKLLKEWGSLENILKNLDRVKPESVRERIKAHLEDLKL SLELSRVRSDLPL  
 EVDFARRREPDREGLRAFLE RLEFGSLLHEFGLL EAPAPLEEAPWPPPEGAFVGFVLSRPEPMW  
 AELKALAA CKEGRVHRAKDPLAGL KDLKEVRGLLAKDLAVLALREGLDLAPSDDPMLLAYLLDP  
 10 SNTTPEGVARRYGGEWTE DAAHRALLAERLQQNLLERLKGE EKLLWLYQEVEKPLSRVLAHMEA  
 TGVRLDVAYL KALSLELAE EIRRL EEEVFR LAGHPFNLSRDQLERVL FDELRLPALGKTQKTG  
 KRSTSAAVLEALREAHPIVEK I LQHRELTKLKNTYVDPLPGLVHPRTGRLHTRFNQTATATGRL  
 SSSDPNLQNIPIRTP LGQIRRAFVAEAGWALVALDYSQIELRVLAHLSG DENLIRVFQEGKDI  
 HTQTASWMFGV SPEAVDPLMRRAAKTVNFGVLYGMSAHRLSQELAI PYEEAVAFIER YFQSF PK  
 15 VRAWIEKTLEEGRKR GYVETLFGRRRYVPDLNARVKS VREAAERMAFNMPVQGT AADLMK LAMV  
 KLFPHLREMGARMLLQVHDELL EAPQARAE EVAALAKEAMEKAYPLAVPLEVEV GIGEDWLSA  
 KG

**SEQ ID NO:2 *Thermus aquaticus* DNA polymerase (Taq)**

MRGMLPLEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQAVYGF AKSLLKALKEDGDAVIV  
 VVFDKAPSF RHEAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKK  
 20 AEKEGYEVRIILTADKDL YQLLSDRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDN  
 LPGVKIG EKTARKLLEEWGSLEALLKNLDR LKPAIREKILAHMDDLKLSWDLAKVRTDLPLEV  
 DFAKRREPDRE RLRRAFLE RLEFGSLLHEFGLL ESPKALEEAPWPPPEGAFVGFVLSRKEPMWAD  
 LLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSN  
 25 TTPEGVARRYGGEWTEE EAGERAA LSERL FANLWGRLEGEERLLWLYREVERPLSAVLAHMEATG  
 VRLDVAYLRALSLEVAEEIARLEAEVFR LAGHPFNLSRDQLERVL FDELGLPAIGKTEKTGKR  
 STSAAVLEALREAHPIVEK I LQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSS  
 SDPNLQNI PVRTPLGQRI RRAFIAEEGWLLVALDYSQIELRVLAHLSG DENLIRVFQEGRDIHT  
 ETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIER YFQSF PKVR  
 30 AWIEKTLEEGRRR RGYVETLFGRRRYVPDL EARVKS VREAAERMAFNMPVQGT AADLMK LAMVKL  
 FPRLEEMGARMLLQVHDEL VLEAPKERAEAVARLAKEVM EGVYPLAVPLEVEV GIGEDWLSAKE

**SEQ ID NO:3 *Thermus filiformis* DNA polymerase (Tfi)**

MLP LLEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQAVYGF AKSLLKALKEDGEVAIVVFD  
 35 AKAPSF RHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGLVRLLEVPGFEADDVLATLARKAER  
 EGYEVRI LSADRDL YQLLSDRIHLLHPEGEV LTPGWLQERYGLSPERWVEYRALVGDPSDNLPG  
 VPGIG EKTALKLLKEWGSLEAILKNLDQVKPERVWEA IRNNLDK LQMSLELSRLRTDLP LEVDF  
 AKRREPTGKGLKAFLE RLEFGSLLHEFGLL EAPKEAEEAPWPPPGGAF LGFLLSRPEPMWAE LL  
 ALAGAKEGRVHRAEDPVGALKDLKEIRGLLAKDLSVLALREGREIPPGDDPMLLAYLLDPGNTN  
 40 PEGVARRYGGEWKEDAAARALLSERLWQALYPRVAEEERLLWLYREVERPLAQVLAHMEATGVR  
 LDVPYLEALSQEVAFELERLEAEVHRLAGHPFNLSRDQLERVL FDELGLPPIGKTEKTGKRST  
 SAAVLELLREAHPIVGRILEYRELMKLKSTYIDPLPRLVHPKTGRLHTRFNQTATATGRLSSSD  
 PNLQNI PVRTPLGQRI RKAFAIEEGHLLVALDYSQIELRVLAHLSG DENLIRVFRREGKDIHTET  
 AAWMFGVPP EGV DGAMRRAAKTVNFGVLYGMSAHRLSQELSI PYEEAAAFIER YFQSF PKVRAW  
 45 IAKTLEEGRKKGYVETLFGRRRYVPDLNARVKS VREAAERMAFNMPVQGT AADLMK LAMVKLFP  
 RLRPLGVRI LLQVHDEL VLEAPKARAE EAAQLAKETMEGVYPLSVPLEVEV GMGEDWLSAKE

**SEQ ID NO:4 *Thermus flavus* DNA polymerase (Tff)**

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQAVYGFSAKSLKALKEDGDVVVVV  
 FDAKAPSRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGLVRLVLEVPVGFEEADDVLATLAKRA  
 EKEGYEVRILTADRDL YQLLSERIAILHPEGYLITPAWLYEKYGLRPEQWVDYRALAGDPSDNI  
 5 PGVKGIGEKTAQRLIREWGSLENLFQHL DQVKPSLREKLQAGMEALALSRKLSQVHTDLPLEVD  
 FGRRRTPNLEGLRAFLEERLEFGSLLHEFGLLLEGPKAAEEAPWPPPEGAF LGFSFSRPEPMWAE  
 LALAGAWEGRLHRAQDPLRGLRDLKGVGILAKDLAVLALREGLDLFPEDDPM LLAYLLDPSNT  
 TPEGVARRYGGEWTE DAGERALLAERLFQTLKERLKGEEERLLWLYEEVEKPLSRVLARMEATGV  
 RLDVAYLQALSLEVEAEVRQLEEEVFR LAGHPFNLSRDQLERVLFDDELGLPAIGKTEKTGKRS  
 10 TSAAVLEALREAHPIVDRILO YRELTKLKNTYIDPLPALVHPKTGRLHTRFNQTATATGRLSSS  
 DPNLQNI PVRTPLGQIRRAFVAEEGWLVVLDYSQIELRVLAHLSG DENLIRVFQEGRDIHTQ  
 TASWMFGVSP EGVDP LMRRAAKTINFGVLYGMSAHRLSGELSI PYEEAVAFIER YFQSYPKVRA  
 WIEGTLEEGRRRGYVETLFGRRRYVPDLNARVKS VREAAERMAFNMPVQGT AADLMK LAMVRLF  
 PRLQELGARM L LQVHDEL VLEAPKDRAERVAALAKEVM EGVWPLQVPLEVEVGLGEDWLSAKE

15

**SEQ ID NO:5 *Thermus sp.* Sps17 DNA polymerase (Sps17)**

MLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQAVYGFSAKSLKALKEDGEVAIVVFD  
 AKAPSRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGLVRLVLEVPVGFEEADDVLATLAKKAER  
 EGYEVRILSADRDL YQLLSDRIHLLHPEGEV LTPGWLQERYGLSPERWVEYRALVGDPSDNLPG  
 20 VPGIGEK TALKLLKEWGSLEAILKNLDQVKPERVREAIRNNLDK LQMSLELSRLRTDLPLEVDF  
 AKRREP DW EGLKAFLEERLEFGSLLHEFGLL EAPKEAEEAPWPPPGGAF LGFLLSRPEPMWAE LL  
 ALAGAKEGRVHRAEDPVGALKDLKEIRGLLAKDLSVLALREGREI PPGDDPML LAYLLDPGNTN  
 PEGVARRYGGEWKEDAAARALLSERLWQALYPRVAEEERLLWLYREVERPLA QVLAHMEATGVR  
 LDVPYLEALSQEVAFELERLEAEVHRLAGHPFNLSRDQLERVLFDDELGLPPIGKTEKTGKRST  
 25 SAAVLELLREAHPIVGRILEYRELMKLKSTYIDPLPRLVHPKTGRLHTRFNQTATATGRLSSSD  
 PNLQNI PVRTPLGQIRKAFIAEEGHLLVALDYSQIELRVLAHLSG DENLIRVFRREGKDIHTET  
 AAWMFVGPPEGV D G AMRRAAKTVNFGVLYGMSAHRLSQELSI PYEEAAAFIER YFQSF PKVRAW  
 IAKTLEEGRKKGYVETLFGRRRYVPDLNARVKS VREAAERMAFNMPVQGT AADLMK LAMVKLFP  
 RLRPLGVRI LLQVHDEL VLEAPKARAEAAQLAKETMEGVYPLSVPLEVEVGMGEDWLSAKA

30

**SEQ ID NO:6 *Thermus thermophilus* DNA polymerase (Tth)**

MEAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQAVYGFSAKSLKALKEDGYKAVF  
 VVFDKAPSRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPVGYEADDVLATLAK  
 KAEKEGYEVRILTADRDL YQLVSDRVAVLHPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSD  
 35 NLPVKGIGEK TALKLLKEWGSLENLLKNLDRVKPENVREKIKAHLEDLRLSLELSRVRTDLP  
 EVDLAQGREPDREGLRAFLEERLEFGSLLHEFGLL EAPAPLEEAPWPPPEGAFVGFVLSRPEPMW  
 AELKALAACRDGRVHRAADPLAGLKD LKEVRGLLAKDLAVLASREGLDLVPGDDPML LAYLLDP  
 SNTTPEGVARRYGGEWTE DAAHRALLSERLHRNLLKRLEGEKLLWLYHEVEKPLSRVLAHMEA  
 TGVRD VAYLQALSLELAEIIRLEEEVFR LAGHPFNLSRDQLERVLFDDELRLPALGKTQKTG  
 40 KRSTSAAVLEALREAHPIVEKILQHRELTKLKNTYVDPLPSLVHPRTGRLHTRFNQTATATGRL  
 SSSDPNLQNI PVRTPLGQIRRAFVAEAGWALVALDYSQIELRVLAHLSG DENLIRVFQEGKDI  
 HTQTASWMFGV PPEAVDPLMRRAAKTVNFGVLYGMSAHRLSQELAI PYEEAVAFIER YFQSF PK  
 VRAWIEKTLEEGRKRGYVETLFGRRRYVPDLNARVKS VREAAERMAFNMPVQGT AADLMK LAMV  
 KLFPR LREMGARML LQVHDELLLEAPQARAEVAALAKEAMEKAYPLAVPLEVEVGMGEDWLSA  
 45 KG

**SEQ ID NO:7 *Thermus caldophilus* DNA polymerase (Tca)**

MEAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPVQAVYGFSAKSLKALKEDGYKAVF  
 VVFDKAPSFRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAK  
 NPEKEGYEVRILTADRDLQLVSDRVAVLHPEGLITPEWLWQKYGLKPEQWVDFRALVGDPSD  
 5 NLPGVKGIGEKTKLLKEWGSLENLLKNLDRVKPENVREKIKAHLEDLRLSLELSRVRTDLPL  
 EVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLEAPAPLEEAPWPPPEGAFVGFVLSRPEPMW  
 AELKALAACRDRGVHRAADPLAGLKDLKEVRGLLAKDLAVLASREGLDLVPGDDPMLLAYLLDP  
 SNTTPEGVARRYGGEWTEAAHRALLSERLHRNLLKRLQGEEKLLWLHYHEVEKPLSRVLAHMEA  
 TGVRLDVAYLQALSLELAEEIRRLEEEVFRLAGHPFNLSRDQLERVLFDLRLPALGKTQKTG  
 10 KRSTSAAVLEALREAHPIVEKILQHRELTKLKNTYVDPLPSLVHPNTGRLHTRFNQTATATGRL  
 SSSDPNLQNI PVRTPLGQRIRRAFVAEAGWALVALDYSQIELRVLAHLSGDENLIRVFQEGKDI  
 HTQTASWMFGVPPEAVDPLMRRAAKTVNFGVLYGMSAHRLSQELAIPIYEEAVAFIERYFQSFPK  
 VRAWIEKTL EEGRKRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAAADLMKLAMV  
 15 KLFPRLREMGARMLLQVHDELLLEAPQAGAEVAALAKEAMEKAYPLAVPLEVEVGMGEDWLSA  
 KG

**SEQ ID NO:8**

Pro Asn Leu Gln Asn X6 Pro X8 X9 X10 X11 X12 Gly, wherein X6 is  
 Ile or Leu; X8 is Ile, Val or Thr; X9 is any amino acid other  
 20 than Arg or Lys; X10 is Thr, Ser or Leu; X11 is Pro or Glu; and  
 X12 is Leu or Glu.

**SEQ ID NO:9**

Pro Asn Leu Gln Asn X6 Pro X8 X9 X10 X11 X12 Gly, wherein X6 is  
 25 Ile or Leu; X8 is Ile, Val or Thr; X9 is any amino acid other  
 than Arg or Lys; X10 is Thr or Ser; X11 is Pro or Glu; and X12  
 is Leu or Glu.

**SEQ ID NO:10**

30 Pro Asn Leu Gln Asn Ile Pro X8 X9 Thr Pro Leu Gly, wherein X8 is  
 Ile or Val; and X9 is any amino acid other than Arg.

**SEQ ID NO:11**

35 Pro Asn Leu Gln Asn Ile Pro Ile X9 Thr Pro Leu Gly, wherein X9  
 is His, Leu or Ser.

**SEQ ID NO:12 Z05**

TGRLSSSDPNLQNIPIRTPLGQRIRRAFVAE

40 **SEQ ID NO:13 Taq**

TGRLSSSDPNLQNI PVRTPLGQRIRRAFIAE

**SEQ ID NO:14 Tfi**

45 TGRLSSSDPNLQNI PVRTPLGQRIRKAFIAE

**SEQ ID NO:15 Tfi**

TGRLSSSDPNLQNI PVRTPLGQRIRRAFVAE

**SEQ ID NO:16 Sps17**

TGRLSSSDPNLQNI PVRTPLGQRIRKAFIAE

**SEQ ID NO:17 Tth**

5 TGRLSSSDPNLQNI PVRTPLGQRIRRAEVAE

**SEQ ID NO:18 Tca**

TGRLSSSDPNLQNI PVRTPLGQRIRRAEVAE

**SEQ ID NO:19 Tma**

10 TGRLSSSDPNLQNLPTKSEEGKEIRKAIVPQ

**SEQ ID NO:20 Tne**

TGRLSSSDPNLQNLPTKSEEGKEIRKAIVPQ

15

**SEQ ID NO:21 Taf**

TGRLSSSNPNLQNLPTKSEEGKEIRKAVRPQ

**SEQ ID NO:22**

20 Ala Ala Ala Ala Ala Ala

**SEQ ID NO:23 Dra**

TGRLSSLNPNLQNIPIRSELGREIRKGFIAE

**SEQ ID NO:24 Bst**

25 TGRLSSVEPNLQNIPIRLEEGRKIRQAFVPS

**SEQ ID NO:25 Bca**

TGRLSSTEPNLQNIPIRLEEGRKIRQAFVPA

30

**SEQ ID NO:26 native consensus motif**

Pro Asn Leu Gln Asn X6 Pro X8 X9 X10 X11 X12 Gly, wherein X6 is Ile or Leu; X8 is Ile, Val or Thr; X9 is Arg or Lys; X10 is Thr, Ser or Leu; X11 is Pro or Glu; and X12 is Leu or Glu.

35

**SEQ ID NO:27 modified Z05 D580 motif**

T-G-R-L-S-S-X7-X8-P-N-L-Q-N

wherein X7 is Ser (S) or Thr (T); and X8 is any amino acid other than Asp (D), or Glu (E)

40

**SEQ ID NO:28 Conserved DNA polymerase active site**

DYSQIELR

**SEQ ID NO:29 CS5 DNA polymerase**

45 MKAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQAVYGFASLLKALKEDGYKAVF  
VVFDAKAPSRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGFEEADDVLATLAK  
KAEREGYEVRIILTADRDLVQLVSDRVAVLHPEGLITPEWLWEKYGLKPEQWVDFRALVGDPSD  
NLPGVKGIGKKTALKLLKEWGSLENILKNLDRVKPESVRERIKAHLEDLKLSELSRVRSDLPL

EVDFARRREPDREGLRAFLERLEFGSLLHEFGLLSEEPVGYRIVKDLVEFEKLIIEKLRESPSF  
 AIDLETSSLDPFDCDIVGIVSVSFKPKEAYYIPLHHRNAQNLEKEVLLKKEILEDPGAKIVGQ  
 NLKFDYKVLVMVKGVEPVPPYFDTMIAAYLLEPNEKKFNLDDLALKFLGYKMTSYQELMSFSFPL  
 FGFSFADVPVEKAANYSCEDADITYRLYKTLCLKLHEADLENVYKIEMLVNVLARMELNGVY  
 5 VDTEFLKKLSEEYGKKLEELAEIYRIAGEPFNINSPKQVSRILFEKLGKIKPRGKTTKTGDYST  
 RIEVLEELAGEHEIIPLILEYRKIQKLSKSTYIDALPKMVNPKTGRIHASFNQGTATGRLSSSD  
 PNLQNLPTKSEEGKEIRKAIVPQDPNWWIVSADYSQIELRILAHLSGDENLLRAFEEGIDVHTL  
 TASRIFNVKPEEVTEEMRRAGKMNFSIIYGVTPYGLSVRLGVPVKEAEKMIVNYFVLYPKVRD  
 YIQRVVSEAKEKGYVRTLFGRKRDI PQLMARDRNTQAEGERIAINTPIQGTAA DI IKLAMIEID  
 10 RELKERKMRSKMIIQVHDELVFEVPNEEKDALVELVKDRMTNVVKLSVPLEVDVTIGKTWS

**SEQ ID NO:30 CS6 DNA polymerase**

MKAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQAVYGFAKSLLKALKEDGYKAVF  
 VVFDKAPSFREHAYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGFEADDVLATLAK  
 15 KAEREGYEVRIILTADRDLVQLVSDRVAVLHPEGHLITPEWLWEKYGLKPEQWVDFRALVGDPSD  
 NLPGVKGIGEKTALKLLKEWGSLENILKNLDRVKPESVRERIKAHLEDLKLSELSRVRSDLPL  
 EVDFARRREPDREGLRAFLERLEFGSLLHEFGLLSEEPVGYRIVKDLVEFEKLIIEKLRESPSF  
 AIALATSSLDPFDCDIVGIVSVSFKPKEAYYIPLHHRNAQNLEKEVLLKKEILEDPGAKIVGQ  
 NLKFDYKVLVMVKGVEPVPPYFDTMIAAYLLEPNEKKFNLDDLALKFLGYKMTSYQELMSFSFPL  
 20 FGFSFADVPVEKAANYSCEDADITYRLYKTLCLKLHEADLENVYKIEMLVNVLARMELNGVY  
 VDTEFLKKLSEEYGKKLEELAEIYRIAGEPFNINSPKQVSRILFEKLGKIKPRGKTTKTGDYST  
 RIEVLEELAGEHEIIPLILEYRKIQKLSKSTYIDALPKMVNPKTGRIHASFNQGTATGRLSSSD  
 PNLQNLPTKSEEGKEIRKAIVPQDPNWWIVSADYSQIELRILAHLSGDENLLRAFEEGIDVHTL  
 TASRIFNVKPEEVTEEMRRAGKMNFSIIYGVTPYGLSVRLGVPVKEAEKMIVNYFVLYPKVRD  
 25 YIQRVVSEAKEKGYVRTLFGRKRDI PQLMARDRNTQAEGERIAINTPIQGTAA DI IKLAMIEID  
 RELKERKMRSKMIIQVHDELVFEVPNEEKDALVELVKDRMTNVVKLSVPLEVDVTIGKTWS

**SEQ ID NO:31 Forward Primer**

5'- CTACCTCCTGGACCCCTCCAA-3'  
 30

**SEQ ID NO:32 Reverse Primer**

5'- ATAACCAACTGGTAGTGGCGTGTA-3'

**SEQ ID NO:33 Polymerase domain of Z05 D580G DNA polymerase**

CTACCTCCTGGACCCCTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGGGGGGAGTGG  
 ACGGAGGACGCCGCCACCGGGCCCTCCTCGCTGAGCGGCTCCAGCAAACCTCTTGGAACGCC  
 TCAAGGGAGAGGAAAAGCTCCTTTGGCTCTACCAAGAGGTGGAAAAGCCCTCTCCCGGGTCC  
 GGCCACATGGAGGCCACCGGGTAAGGCTGGACGTGGCCTATCTAAAGGCCCTTTCCCTGGAG  
 40 CTTGCGGAGGAGATTCGCCGCTCGAGGAGGAGTCTTCCGCTGGCGGGCCACCCCTTCAACC  
 TGAACCTCCCGTGACCAGCTAGAGCGGGTCTCTTTGACGAGCTTAGGCTTCCCGCCCTGGGCAA  
 GACGCAAAGACGGGAAGCGCTCCACCAGCGCCGCGGTGCTGGAGGCCCTCAGGGAGGCCAC  
 CCCATCGTGGAGAAGATCCTCCAGCACCGGGAGCTCACCAAGCTCAAGAACACCTACGTAGACC  
 CCTCCCGGGCTCGTCCACCCGAGGACGGGCCGCTCCACACCCGCTTCAACCAGACAGCCAC  
 GGCCACGGGAAGGCTCTCTAGCTCCGGGCCAACCTGCAGAACATCCCCATCCGCACCCCTTG  
 45 GGCCAGAGGATCCGCCGGGCTTCGTGGCCGAGCGGGATGGGCGTTGGTGGCCCTGGACTATA  
 GCCAGATAGAGCTCCGGGTCCTCGCCACCTCTCCGGGACGAGAACCTGATCAGGGTCTTCCA  
 GGAGGGGAAGGACATCCACACCCAGACCGCAAGCTGGATGTTCCGGCGTCTCCCGGAGGCCGTG  
 GACCCCTGATGCGCCGGGCGGCCAAGACGGTGAACCTCGGCGTCTCTACGGCATGTCCGCC

ATAGGCTCTCCCAGGAGCTTGCCATCCCCTACGAGGAGGCGGTGGCCTTTATAGAGCGCTACTT  
 CCAAAGCTTCCCCAAGGTGCGGGCCTGGATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGC  
 TACGTGGAACCCCTCTTCGGAAGAAGGCGCTACGTGCCCGACCTCAACGCCCGGGTGAAGAGCG  
 TCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCCGACCTCAT  
 5 GAAGCTCGCCATGGTGAAGCTCTTCCCCACCTCCGGGAGATGGGGGCCCGCATGCTCCTCCAG  
 GTCCACGACGAGCTCCTCCTGGAGGCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCA  
 AGGAGGCCATGGAGAAGGCCTATCCCCTCGCCGTGCCCTGGAGGTGGAGGTGGGGATCGGGGA  
 GGACTGGCTTTCCGCCAAGGGCTGATATCAGATCTCCCTGATTATGCGTCAGTCTATGAAGAAA  
 AATCGTATACAGATGGACGAAGAGAGAATCCTTGTGAATTTAACAGAGGGGTATAGGGATTACAC  
 10 GCCACTACCAGTTGGTTAT

**SEQ ID NO:34 - BRAF wild type sequence**

AGTAAAAATAGGTGATTTTGGTCTAGCTACAGTGAAATCTCGATGGAGTGGGTCCCATCAGTTT  
 GAACAGTTGTCTGGATCCATTTTGTGGATGGTAAGAATTGAGGCTA  
 15

**SEQ ID NO:35 - BRAF V600R mutant sequence**

AGTAAAAATAGGTGATTTTGGTCTAGCTACAAGGAAATCTCGATGGAGTGGGTCCCATCAGTTT  
 GAACAGTTGTCTGGATCCATTTTGTGGATGGTAAGAATTGAGGCTA

**SEQ ID NO:36 *Deinococcus radiodurans* DNA polymerase (Dra)**

MADASPDPSKPDALVLIIDGHALAFRSYFALPPLNNSKGMTDAIVGFMKLLLRARQKSNQVIV  
 VFDPPVKTLRHEQYEGYKSGRAQTPEDLRGQINRIRALVDALGFPRLEEPGYEADDVIAASLTRM  
 AEGKGYEVRIVTSRDAYQLLDEHVKVIANDFSLIGPAQVEEKYGVTVRQWVDYRALTGDasDN  
 IPGAKGIGPKTAAKLLQEYGTLEKVYEAHAGTLKPDGTRKKLLDSEENVKFSHDLSCMVTDL P  
 25 LDIEFGVRRPLPNPLVTEDLLTELELHSLRPMILGLNGPEQDGHAPDDLLEHAQTPEEDEAA  
 ALPAFSAPELAEWQTPAEGAVWGYVLSREDDLTAALLAAATFEDGVARPARVSEPDEWAQAEAP  
 ENLFGELLPSPDKPLTKKEQKALEKAQKDAEKARAKLREQFPATVDEAEFVGQRTVTAATAAKALA  
 AHL SVRGTVVEPGDDPLLYAYLLDPANTNMPVVAKRYLDREWPADAPTRAAITGHLVREL PPLL  
 DDARRKMYDEMEKPLSGVLGRMEVRGVQVDSDFLQTL SIQAGVRLADLESQIHEYAGEEFHIRS  
 30 PKQLETVLYDKLELASSKTKLTGQRSTAVSALEPLRDAHP I IPLVLEFRELDKLRGTYLDP I P  
 NLVNPHTGRLHTTFAQTAVATGRLSSLNPNLQNIPIRSELGREIRKGFIAEDGFTLIAADYSQI  
 ELRLLAHIADDPLMQQAFVEGADIHRRTAAQVLGLDEATVDANQRRAAKTVNFGVLYGMSAHLR  
 SNDLGI PYAEAATFIEIYFATYPGIRRYINHTLDFGRTHGYVETLYGRRRYV PGLSSRN RVQRE  
 AEERLAYNMP IQGTAADIMKLAMVQLDPQLDAIGARMLLQVHDELLIEAPLDKAEQVAALTKKV  
 35 MENVVQLKVPLAVEVGTGPNWFDTK

**SEQ ID NO:37 *Thermosipho africanus* DNA polymerase (Taf)**

MGKMFLFDGTGLVYRAFYAIDQSLQTS SGLHTNAVYGLTKMLIKFLKEHISIGKDACVFVLD SKG  
 GSKRRKDILETYKANRPSTPDLLEQIPYVEELVDALGIKVLKIEGFEADDI IATLSK KFESDFE  
 40 KVNIITGDKDLLQLVSDKVFVWRVERGITDLVLYDRNKVIEKYGIYPEQFKDYLSLVGDQIDNIP  
 GVKGIGKKTAVSLLKKYNSLENV LKNINLLTEKLRRLLED SKEDLQKSIELVELIYDVPMDVEKD  
 EIIYRGYNPKLLKVLKKYEFSSIIKELNLQEKLEKEYILVDNEDKLLKLAEEIEKYKTF SIDTE  
 TTSLDPFEAKLVGISISTMEGKAYYIPVSHFGAKNISKSLIDKFLKQILQEKDYNIVGQNLKFDY  
 EIFKSMGFSPNVP HFDTMIAAYLLNPDEKRFNLEELSLKYLGYKMI SFDELV NENVPLFGNDFSY  
 45 VPLERAVEYSCEDADV TYRIFRKLGRKIYENEMEKL FYEIEEMPLIDVLSE MELNGVYFDEEYLKE  
 LSKKYQEKMDGIKEKVF EIAGETFNLSSTQVAYILFEKLN IAPYKKTATGKFSTNAEVL EELSK  
 EHEIAKLLLE YRKYQKLKSTYIDS I PLSINRKTNRVHTTFHQ TGTSTGRLSSSNPNLQNLPT RSE  
 EGKEIRKAVRPQRQDWILGADYSQIELRVLAVHVKDENLLKAFKEDLDIHTITAAKIFGVSEMF



VSEQMRRVGMVNFALIIYGVSPYGLSKRIGLSVSETKKIIDNYFRYYKGVFEYLKRMKDEARKKG  
 YVTTLFGRRRIYIPQLRSKNGNRVQEGERIAVNTPIQGTAADIIKIAMINIHNRLKKNLRSKMIL  
 QVHDELVFEVDPNELEIVKDLVRDEMENAVKLDVPLKVDVYVYKKEWE

5 **SEQ ID NO:38 *Thermotoga maritima* DNA polymerase (Tma)**

MARLFLFDGTALAYRAYYALDRSLSTSTGIPTNATYGVARMLVRFIKDHIIVGKDYVAVAFDKKA  
 ATFRHKLLETYKAQRPKTPDLLIQQLPYIKKLVEALGMKVLEVEGYEADDIIATLAVKGLPLFDE  
 IFIVTGDKDMLQLVNEKIKVWRIVKGISDLELYDAQVKEKYGVPEPQQIPDLLALTGDEIDNIPG  
 10 VTGIGEKTAVQQLLEKYKDLEDILNHVRELPOKVRKALLRDRENAILSKKLAILETNVPIEINWEE  
 LRYQGYDREKLLPLLKELEFASIMKELQLYEESEPVGYRIVKDLVEFEKLEIKLRESPSFAIDLE  
 TSSLDPFDCDIVGISVSFKPKPEAYYIPLHHRNAQNLDKEEVLKKLKEILEDPGAKIVGQNLKFDY  
 KVLVVGKVEPVPPYFDTMIAAYLLEPNEKKNLDDLALKFLGYKMTSYQELMSFSPLFGFSFAD  
 VPVEKAANYSCEDADITYRLYKTLCLKHEADLENVFIKIEMPLVNVLARMELNGVYVDTEFLKK  
 LSEEYGKLEELAEIYRIAGEPFNINSPKQVSRILFEKLGKPRGKTTKTGDIYSTRIEVLEELA  
 15 GEHEIIPLILEYRKIQKLSSTYIDALPKMVNPKTGRIHASFNQTGTATGRLSSSDPNLQNLPTKS  
 EEGKEIRKAIVPQDPNWWIVSADYSQIELRILAHLSGDNLLRAFEEGIDVHTLTASRIYFNKPE  
 EVTEEMRRAGKMNFSIIYGVTPYGLSVRLGVPVKEAEKMIYVNYFVLYPKVRDYIQRVVSEAKEK  
 GYVRTLFGRKRDIQOLMARDRNTQAEGERIAINTPIQGTAADIIKLAMIEIDRELKERKMRSKMI  
 IQVHDELVFEVPEEKDALVELVKDRMTNVVVKLSVPLEVDVTIGKTWS

20

**SEQ ID NO:39 *Thermotoga neopolitana* DNA polymerase (Tne)**

MARLFLFDGTALAYRAYYALDRSLSTSTGIPTNAVYGVARMLVKFIKEHIPEKDYA AVAFDKKA  
 ATFRHKLLVSDKAQRPKTPALLVQQLPYIKRLIEALGFKVLELEGEYADDIIATLAVRAARFLMR  
 FSLITGDKDMLQLVNEKIKVWRIVKGISDLELYDSKKVKERYGVPEHQIPDLLALTGDDIDNIPG  
 25 VTGIGEKTAVQLLGKYRNLEYILEHARELPQRVRKALLRDREVAIILSKKLATLVTNAPVEVDWEE  
 MKYRGYDKRKLPLILKELEFASIMKELQLYEEAEPTGYEIVKDHKTFEDLIEKLKEVPSFALDLE  
 TSSLDPFNCEIVGISVSFKPKPTAYYIPLHHRNAHNLDLTLVLSKLKEILEDPSSKIVGQNLKYDY  
 KVLVVGKISPVYPHFDTMIAAYLLEPNEKKNLEDLCLKFLGYKMTSYQELMSFSSPLFGFSFAD  
 VPVDKAAEYSCEDADITYRLYKILSMKLHEAELENVFIYRIEMPLVNVLARMFENWVYVDTEFLKK  
 30 LSEEYGKLEELAEKIYQIAGEPFNINSPKQVSNILFEKLGKPRGKTTKTGDIYSTRIEVLEEIA  
 NEHEIVPLILEFRKILKLSSTYIDTLPKLVNPKTGRFHASFHQTGTATGRLSSSDPNLQNLPTKS  
 EEGKEIRKAIVPQDPDWWIVSADYSQIELRILAHLSGDNLVKAFEEGIDVHTLTASRIYFNKPE  
 EVNEEMRRVGMVNFSSIYGVTPYGLSVRLGIPVKEAEKMIISYFTLYPKVRSYIQQVVAEAKEK  
 GYVRTLFGRKRDIQOLMARDKNTQSEGERIAINTPIQGTAADIIKLAMIDIDEELRKRNMKSRI  
 35 IQVHDELVFEVPDEEKEELVDLVKNKMTNVVVKLSVPLEVDISIGKSW

**SEQ ID NO:40 *Bacillus stearothermophilus* DNA polymerase (Bst)**

MKNKLVLIDGNSVAYRAFFALPLLHNDKGIHTNAVYGFMTMLNKILAEEQPTHILVAFDAGKTF  
 RHETFQDYKGRQQTPELSEQFPLLRELLKAYRIPAYELDHYEADDIIGTMAARAEREGFAVKV  
 40 ISGDRDLTQLASPQVTVEITKKGITDIESYTPETVVEKYGLTPEQIVDLKGLMGDKSDNIPGVP  
 IGEKTAVKLLKQFGTVENVLASIDEIKGEKLNENLRQYRDLALLSKQLAAICRDAPVELTLDDIV  
 YKGEDREKVVALFQELGFQSFQFLDKMAVQVQTEGEKPLAGMDFAIADSVTDEMLADKAALVVEVVD  
 NYHHAPIVGIALANERGRFFLRPETALADPKFLAWLGDETKKKTMFDSKRAAVALKWKGIELRGV  
 VFDLLLAAYLLDPAQAAGDVA AVAKMHQYEAVRSDEAVYKGAARTVPDEPTLAEHLARKAAAIW  
 45 ALEEPLMDELRRNEQDRLLTELEQPLAGILANMEFTGVKVDTKRLEQMGAELEQLOAVERRIYE  
 LAGQEFNINSPKQLGTVLFDKLLQPLVKKTKTGYSSTADVLEKLAPHHEIVEHILHYRQLGKLQS  
 TYIEGLLKVVHPVTGKVHTMFNQALTQTGRLSSVEPNLQNIPIRLEEGRKIRQAFVPSEPDWLI  
 AADYSQIELRVLAHIAEDDNLIEAFRRGLDIHTKTAMDI FHVSEEDVTANMRRQAKAVNFGIVY

ISDYGLAQNLNITRKEAAEFIERYFASFPGVKQYMDNIVQEAKQKGYVTLLHRRRYLPDITSRN  
FNVRSFAERTAMNTPIQGSAADIKKAMIDLSVRLREERLQARLLLQVHDELILEAPKEEIERLC  
RLVPEVMEQAVALRVPLKVDYHYGPTWYDAK

5 **SEQ ID NO:41 *Bacillus caldotenax* DNA polymerase (Bca)**

MKKKLVLIDGSSVAYRAFFALPLLHNDKGIHTNAVYGFMMMLNKILAEETPHMLVAFDAGKTF  
RHEAFQEYKGGRRQQTPELSEQFPLLRELLRAYRIPAYELNVEADDIIGTLAARAEQEGFEVKV  
ISGDRDLTQLASPHVTVDITKKGITDIEPYTPEAVREKYGLTPEQIVDLKGLMGDKSDNIPGVPG  
10 IGEKTAVKLLRQFGTVENVLASIDEIKGEKLETLRQHREALLSKKLAAIRRDAPVELSLDDIA  
YQGEDREKVVALFKELGFQSFLEKMESPSSEEEKPLAKMAFTLADRVTEEMLADKAALVVEVVEE  
NYHDAPIVGIHAVNEHGRFFLRPETALADPQFVAWLGDETKKKSMFDSKRAAVALKWKGIELCGV  
SFDLLLAAYLLDPAQGVDDVAAAAMKQYEAVRPDEAVYGGKAKRAVPDEPVLAEHLVRKAAAIW  
ALERPFLDELRRNEQDRLLVELEQPLSSILAEMEFAGVKVDTKRLEQMGEELAEQLRTVEQRIYE  
LAGQEFNINSPKQLGVILFEKLQLPVLKKSKTGYSTSADVLEKLAPYHEIVENILQHYRQLGKLO  
15 STYIEGLLKVVPRPDTKKVHTIFNQALTQTGRLSSTEPNLQNIPIRLEEGRKIRQAFVPSSEDLI  
FAADYSQIELRVLAHIAEDDNLMEAFRRDLDIHTKTAMDIFQVSEDEVTPNMRRQAKAVNFGIVY  
GISDYGLAQNLNISRKEAAEFIERYFESFPGVKRYMENIVQEAKQKGYVTLLHRRRYLPDITSR  
NFNVRSAERMAMNTPIQGSAADIKKAMIDLNARLKEERLQARLLLQVHDELILEAPKEEMERL  
20 CRLVPEVMEQAVTLRVPLKVDYHYGSTWYDAK

20 **SEQ ID NO:42**

Pro Asn Leu Gln Asn X6 Pro X8 X9 X10 X11 X12 Gly, wherein X6 is  
Ile or Leu; X8 is Ile, Val or Thr; X9 is Gly, Ala, Leu, Met, Trp,  
Pro, Ser, Thr, Phe, Tyr, Cys, Asn, Gln, Asp, Glu, Val, Ile or His;  
25 X10 is Thr, Ser or Leu; X11 is Pro or Glu; and X12 is Leu or Glu.

## WHAT IS CLAIMED IS:

1. A DNA polymerase having increased 3'-mismatch discrimination activity compared with a control DNA polymerase, wherein the amino acid of the DNA polymerase corresponding to position 589 of SEQ ID NO:1 is any amino acid other than R or K, and wherein the control DNA polymerase has the same amino acid sequence as the DNA polymerase except that the amino acid of the control DNA polymerase corresponding to position 589 of SEQ ID NO:1 is R or K.
2. The DNA polymerase of claim 1, comprising a motif in the polymerase domain comprising  
P-N-L-Q-N-X<sub>1</sub>-P-X<sub>2</sub>-X<sub>3</sub>-X<sub>4</sub>-X<sub>5</sub>-X<sub>6</sub>-G, wherein:  
X<sub>1</sub> is I or L;  
X<sub>2</sub> is I, V or T;  
X<sub>3</sub> is any amino acid other than R or K;  
X<sub>4</sub> is T, S or L;  
X<sub>5</sub> is P or E; and  
X<sub>6</sub> is L or E (SEQ ID NO:8).
3. The DNA polymerase of claim 1, comprising a motif in the polymerase domain comprising  
P-N-L-Q-N-X<sub>1</sub>-P-X<sub>2</sub>-X<sub>3</sub>-X<sub>4</sub>-X<sub>5</sub>-X<sub>6</sub>-G, wherein:  
X<sub>1</sub> is I or L;  
X<sub>2</sub> is I, V or T;  
X<sub>3</sub> is any amino acid other than R or K;  
X<sub>4</sub> is T or S;

X<sub>5</sub> is P or E; and

X<sub>6</sub> is L or E (SEQ ID NO:9).

4. The DNA polymerase of claim 1, comprising a motif in the polymerase domain comprising  
5 P-N-L-Q-N-I-P-X<sub>2</sub>-X<sub>3</sub>-T-P-L-G, wherein:  
X<sub>2</sub> is I or V;  
X<sub>3</sub> is any amino acid other than R (SEQ ID NO:10).
5. The DNA polymerase of claim 4, wherein X<sub>3</sub> is H, L or S (SEQ ID NO:11).
6. The DNA polymerase of any one of claims 1 to 5, wherein the amino acid  
10 corresponding to position 580 of SEQ ID NO:1 is any amino acid other than D or E.
7. The DNA polymerase of any one of claims 1 to 6, wherein the amino acid  
corresponding to position 580 of SEQ ID NO:1 is selected from the group consisting of  
L, G, T, Q, A, S, N, R, and K.
8. The DNA polymerase of any one of claims 1 to 7, wherein the DNA polymerase  
15 comprises an amino acid sequence at least 80%, preferably at least 90%, more  
preferably at least 95% sequence identical to SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 36, 37, 38,  
39, 40 or 41.
9. The DNA polymerase of any one of claims 6 or 7, wherein the polymerase has at least  
20 80%, preferably at least 90%, more preferably at least 95% sequence identity to SEQ ID  
NO:1.
10. The DNA polymerase of claim 9, wherein the DNA polymerase is a Z05 DNA  
polymerase, and the amino acid at position 580 is selected from the group consisting of  
L, G, T, Q, A, S, N, R, and K.

11. A recombinant nucleic acid encoding the DNA polymerase according to any one of claims 1 to 10.
12. A method for conducting primer extension, comprising:  
contacting a DNA polymerase according to any one of claims 1 to 10 with a primer, a  
5 polynucleotide template, and nucleoside triphosphates under conditions suitable for  
extension of the primer, thereby producing an extended primer.
13. A kit for producing an extended primer, comprising:  
at least one container providing a DNA polymerase according to any one of claims 1 to  
10.
- 10 14. The kit according to claim 13, further comprising one or more additional containers  
selected from the group consisting of:
  - (a) a container providing a primer hybridizable, under primer extension conditions, to  
a predetermined polynucleotide template;
  - (b) a container providing nucleoside triphosphates; and
  - 15 (c) a container providing a buffer suitable for primer extension.
15. A reaction mixture comprising a DNA polymerase according to any one of claims 1 to  
10, at least one primer, a polynucleotide template, and nucleoside triphosphates.

*Figure 1*

\*  
 Z05 TGR LSSSDP N L Q N I P I R T P L G Q R I R R A F V A E (SEQ ID NO: 12)  
 T a q TGR LSSSDP N L Q N I P V R T P L G Q R I R R A F I A E (SEQ ID NO: 13)  
 T f i TGR LSSSDP N L Q N I P V R T P L G Q R I R K A F I A E (SEQ ID NO: 14)  
 T f l TGR LSSSDP N L Q N I P V R T P L G Q R I R R A F V A E (SEQ ID NO: 15)  
 S p s 1 7 TGR LSSSDP N L Q N I P V R T P L G Q R I R K A F I A E (SEQ ID NO: 16)  
 T t h TGR LSSSDP N L Q N I P V R T P L G Q R I R R A F V A E (SEQ ID NO: 17)  
 T c a TGR LSSSDP N L Q N I P V R T P L G Q R I R R A F V A E (SEQ ID NO: 18)  
 T m a TGR LSSSDP N L Q N L P T K S E E G K E I R K A I V P Q (SEQ ID NO: 19)  
 T n e TGR LSSSDP N L Q N L P T K S E E G K E I R K A I V P Q (SEQ ID NO: 20)  
 T a f TGR LSSSNP N L Q N L P T R S E E G K E I R K A V R P Q (SEQ ID NO: 21)  
 D r a TGR LSSLN P N L Q N I P I R S E L G R E I R K G F I A E (SEQ ID NO: 23)  
 B s t TGR LSSV E P N L Q N I P I R L E E G R K I R Q A F V P S (SEQ ID NO: 24)  
 B c a TGR LSS T E P N L Q N I P I R L E E G R K I R Q A F V P A (SEQ ID NO: 25)  
 -----P N L Q N X<sub>1</sub> P X<sub>2</sub> X<sub>3</sub> X<sub>4</sub> X<sub>5</sub> X<sub>6</sub> G----- (SEQ ID NO: 26)

**FIGURE 2**

A. Sequence identities over the entire polymerase I enzyme (corresponding to amino acids 1-834 of Z05)													
Name	Z05	Taq	Tfi	Tfi	Sps17	Tth	Tca	Dra	Tma	Tne	Taf	Bst	Bca
Z05		0.864	0.833	0.859	0.839	0.962	0.958	0.459	0.374	0.368	0.359	0.407	0.408
Taq	0.864		0.831	0.854	0.836	0.872	0.864	0.468	0.382	0.368	0.351	0.397	0.397
Tfi	0.833	0.831		0.82	0.991	0.829	0.824	0.45	0.371	0.375	0.353	0.405	0.397
Tfi	0.859	0.854	0.82		0.824	0.853	0.848	0.462	0.381	0.374	0.356	0.397	0.398
Sps17	0.839	0.836	0.991	0.824		0.835	0.83	0.452	0.375	0.377	0.355	0.407	0.399
Tth	0.962	0.872	0.829	0.853	0.835		0.989	0.463	0.373	0.367	0.358	0.406	0.406
Tca	0.958	0.864	0.824	0.848	0.83	0.989		0.46	0.371	0.365	0.356	0.404	0.404
Dra	0.459	0.468	0.45	0.462	0.452	0.463	0.46		0.334	0.325	0.314	0.338	0.339
Tma	0.374	0.382	0.371	0.381	0.375	0.373	0.371	0.334		0.854	0.567	0.37	0.377
Tne	0.368	0.368	0.375	0.374	0.377	0.367	0.365	0.325	0.854		0.558	0.377	0.376
Taf	0.359	0.351	0.353	0.356	0.355	0.358	0.356	0.314	0.567	0.558		0.356	0.364
Bst	0.407	0.397	0.405	0.397	0.407	0.406	0.404	0.338	0.37	0.377	0.356		0.881
Bca	0.408	0.397	0.397	0.398	0.399	0.406	0.404	0.339	0.377	0.376	0.364	0.881	
B. Sequence identities over polymerase sub domain only (corresponding to amino acids 420-834 of Z05)													
Name	Z05	Taq	Tfi	Tfi	Sps17	Tth	Tca	Dra	Tma	Tne	Taf	Bst	Bca
Z05		0.901	0.845	0.891	0.845	0.975	0.973	0.563	0.483	0.478	0.44	0.498	0.49
Taq	0.901		0.879	0.901	0.877	0.906	0.901	0.561	0.488	0.473	0.44	0.503	0.495
Tfi	0.845	0.879		0.857	0.997	0.853	0.853	0.566	0.495	0.49	0.449	0.512	0.49
Tfi	0.891	0.901	0.857		0.855	0.889	0.889	0.571	0.492	0.48	0.444	0.494	0.485
Sps17	0.845	0.877	0.997	0.855		0.853	0.853	0.566	0.495	0.49	0.449	0.512	0.49
Tth	0.975	0.906	0.853	0.889	0.853		0.99	0.563	0.478	0.473	0.437	0.496	0.488
Tca	0.973	0.901	0.853	0.889	0.853	0.99		0.563	0.478	0.473	0.437	0.496	0.488
Dra	0.563	0.561	0.566	0.571	0.566	0.563	0.563		0.45	0.448	0.426	0.474	0.454
Tma	0.483	0.488	0.495	0.492	0.495	0.478	0.478	0.45		0.883	0.622	0.474	0.475
Tne	0.478	0.473	0.49	0.48	0.49	0.473	0.473	0.448	0.883		0.615	0.476	0.473
Taf	0.44	0.44	0.449	0.444	0.449	0.437	0.437	0.426	0.622	0.615		0.46	0.473
Bst	0.498	0.503	0.512	0.494	0.512	0.496	0.496	0.474	0.474	0.476	0.46		0.898
Bca	0.49	0.495	0.49	0.485	0.49	0.488	0.488	0.454	0.475	0.473	0.473	0.898	

INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2011/003004

A. CLASSIFICATION OF SUBJECT MATTER  
INV. C12N9/12 C12N15/54 C12N15/10  
ADD.  
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED  
Minimum documentation searched (classification system followed by classification symbols)  
C12N  
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)  
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2008/046612 A1 (ROCHE DIAGNOSTICS GMBH [DE]; HOFFMANN LA ROCHE [CH]) 24 April 2008 (2008-04-24) tables 12, 13, 14 -----	6,7,9,10
A	THOMAS E EXNER: "Insights into the high fidelity of a DNA polymerase I mutant", JOURNAL OF MOLECULAR MODELING, SPRINGER VERLAG, DE, vol. 15, no. 10, 31 March 2009 (2009-03-31), pages 1271-1280, XP019742687, ISSN: 0948-5023, DOI: 10.1007/S00894-009-0491-4 -----	1
X	US 2009/134042 A1 (NOMOTO TSUYOSHI [JP] ET AL) 28 May 2009 (2009-05-28) paragraph [0208] -----	1-5,8, 11-15
-/--		

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search  22 September 2011	Date of mailing of the international search report  30/09/2011
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer  Van der Schaal, C



## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2011/003004

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 01/90337 A2 (THIRD WAVE TECH INC [US]; ALLAWI HATIM [US]; BARTHOLOMAY CHRISTIAN TOR) 29 November 2001 (2001-11-29) page 13, line 7 - line 13 page 134, lines 8-13 page 280; table 3; sequence 40 -----	1-4,8, 11-15

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2011/003004

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008046612	A1	24-04-2008	CA 2666758 A1
			CN 101528919 A
			EP 2079834 A1
			JP 2011502467 A
			US 2009280539 A1
			US 2009148891 A1
-----			
US 2009134042	A1	28-05-2009	WO 2007148809 A1
-----			
WO 0190337	A2	29-11-2001	AU 6344901 A
			CA 2410072 A1
			EP 1294863 A2
			JP 2004521606 A
-----			