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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 March 2005 (03.03.2005)

PCT

(10) International Publication Number
WO 2005/019476 A1

(51) International Patent Classification[7]: C12Q 1/68, C12P 19/34, C12N 9/12

(21) International Application Number: PCT/US2003/026412

(22) International Filing Date: 20 August 2003 (20.08.2003)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **APPLERA CORPORATION** [US/US]; 850 Lincoln Centre Drive, Foster City, CA 94404 (US).

(72) Inventors: **BOND-NUTTER, Diane, R.**; 11177 Elvessa Street, Oakland, CA 94605 (US). **HEINER, Cheryl, R.**; 1800 Portola State Park Road, La Honda, CA 94020 (US). **ROZZELLE, James, E.**; 761 Second Avenue, San Francisco, CA 94118 (US).

(74) Agents: **GARRETT, Arthur, S.** et al.; Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P., 1300 I Street, N.W., Washington, DC 20005-3315 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: POLYMERASE COMPOSITIONS

(57) Abstract: Compositions comprising at least two polymerases are provided. Methods for producing primer extension products using at least two polymerases are also provided. Kits for producing primer extension products comprising at least two polymerases are also provided.

WO 2005/019476 A1

POLYMERASE COMPOSITIONS

Field of the Invention

[001] The present invention generally relates to compositions comprising different polymerases and methods that employ such compositions.

Background of the Invention

[002] DNA polymerases are enzymes that synthesize DNA molecules from deoxynucleotide triphosphates (dNTPs) using a template DNA strand and a complementary oligonucleotide primer annealed to a portion of the template DNA strand. A detailed description of certain DNA polymerases and their characterization can be found, e.g., in Kornberg, *DNA Replication Second Edition*, W.H. Freeman (1989).

[003] DNA polymerases have a variety of uses in molecular biology techniques. Such techniques include primer extension reactions, DNA sequencing, and nucleic acid amplification techniques such as the polymerase chain reaction (PCR).

Summary of the Invention

[004] In certain embodiments, a composition comprising at least two polymerases selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), *Thermus thermophilus* GK24 (G46D; F669Y), AmpliTaq FS (E681I), and AmpliTaq FS (T664N; R660G) is provided.

[005] In certain embodiments, a composition comprising *Thermus thermophilus* GK24 (G46D; F669Y) and *Thermus oshimai* (G43D; F665Y) is provided.

[006] In certain embodiments, the composition further comprises at least one primer. In certain embodiments, the composition further comprises at least one extendable nucleotide. In certain embodiments, the composition further comprises at least one terminator.

[007] In certain embodiments, a method of generating at least one primer extension product is provided. In certain embodiments, methods comprise forming a reaction composition comprising at least one target nucleic acid template, at least one primer, at least one extendable nucleotide, and at least two polymerases selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), *Thermus thermophilus* GK24 (G46D; F669Y), AmpliTaq FS (E681I), and AmpliTaq FS (T664N; R660G); and incubating the composition under appropriate conditions to generate at least one primer extension product.

[008] In certain embodiments, a method of generating at least one primer extension product comprises forming a reaction composition comprising at least one target nucleic acid template, at least one primer, at least one extendable nucleotide, and *Thermus thermophilus* GK24 (G46D; F669Y) and *Thermus oshimai* (G43D; F665Y); and incubating the composition under appropriate conditions to generate at least one primer extension product.

[009] In certain embodiments, the method further comprises separating the at least one primer extension product. In certain embodiments, the method further comprises detecting the at least one primer extension product.

[010] In certain embodiments, methods of sequencing a target nucleic acid template are provided. In certain embodiments, a method of sequencing a target nucleic acid template comprises forming a reaction composition comprising at least one target nucleic acid template, at least one primer, at least one extendable nucleotide, at least one terminator, and at least two polymerases selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus*

scotoductus (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), *Thermus thermophilus* GK24 (G46D; F669Y), AmpliTaq FS (E681I), and AmpliTaq FS (T664N; R660G); incubating the composition under appropriate conditions to generate at least one primer extension product; separating the at least one primer extension product; detecting the at least one primer extension product; and determining the sequence of the target nucleic acid template.

[011] In certain embodiments, a kit is provided. In certain embodiments, the kit comprises at least two polymerases selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), *Thermus thermophilus* GK24 (G46D; F669Y), AmpliTaq FS (E681I), and AmpliTaq FS (T664N; R660G).

[012] In certain embodiments, the kit comprises *Thermus thermophilus* GK24 (G46D; F669Y) and *Thermus oshimai* (G43D; F665Y).

[013] In certain embodiments, the kit further comprises at least one terminator. In certain embodiments, the kit further comprises at least one extendable nucleotide. In certain embodiments, the kit further comprises at least one primer.

Brief Description of the Figures

[014] Figure 1A shows the amino acid sequence of *Thermus aquaticus* polymerase (SEQ ID NO: 1). Figure 1B shows the nucleic acid sequence of *Thermus aquaticus* polymerase (SEQ ID NO: 10).

[015] Figure 2 shows the amino acid and nucleic acid sequences of *Thermus thermophilus* HB8 polymerase (SEQ ID NO: 2 and SEQ ID NO: 11, respectively).

[016] Figure 3A shows the amino acid sequence of *Thermus oshimai* polymerase (SEQ ID NO: 3). Figure 3B shows the nucleic acid sequence of *Thermus oshimai* polymerase (SEQ ID NO: 12).

[017] Figure 4 shows the nucleic acid and amino acid sequences of three strains of *Thermus scotoductus* polymerase. Figure 4A shows the amino acid sequence of a polymerase from *Thermus scotoductus* strain X-1 (SEQ ID NO: 4). Figure 4B shows the amino acid sequence of a polymerase from *Thermus scotoductus* strain X-1 lacking the first three amino terminal amino acids (SEQ ID NO: 5). Figure 4C shows the amino acid sequence of a polymerase from *Thermus scotoductus* strain SM3 (SEQ ID NO: 6). Figure 4D shows the amino acid sequence of a polymerase from *Thermus scotoductus* strain Vi7a (SEQ ID NO: 7). Figure 4E shows the nucleic acid sequence of a polymerase from *Thermus scotoductus* strain X-1 (SEQ ID NO: 13). Figure 4F shows the nucleic acid sequence of a polymerase from *Thermus scotoductus* strain SM3 (SEQ ID NO: 14). Figure 4G shows the nucleic acid sequence of a polymerase from *Thermus scotoductus* strain Vi7a (SEQ ID NO: 15).

[018] Figure 5A shows the amino acid sequence of *Thermus thermophilus* 1B21 polymerase (SEQ ID NO: 8). Figure 5B shows the nucleic acid sequence of *Thermus thermophilus* 1B21 polymerase (SEQ ID NO: 16).

[019] Figure 6A shows the amino acid sequence of *Thermus thermophilus* GK24 polymerase (SEQ ID NO: 9). Figure 6B shows the nucleic acid sequence of *Thermus thermophilus* GK24 polymerase (SEQ ID NO: 17).

Detailed Description of Certain Exemplary Embodiments

[020] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed. In this application, the use of the singular includes the plural unless specifically stated otherwise. In this application, the use of "or" means "and/or" unless stated otherwise. Furthermore, the use of the term "including", as well as other forms, such as "includes" and "included", is not limiting.

[021] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described. All documents, or portions of documents, cited in this application,

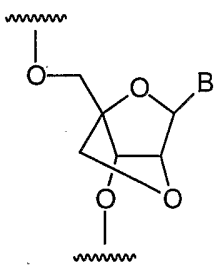
including but not limited to patents, patent applications, articles, books, and treatises, are hereby expressly incorporated by reference in their entirety for any purpose.

Definitions

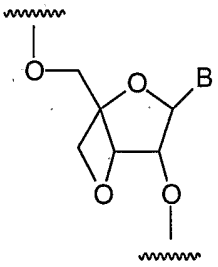
[022] The term "nucleotide base", as used herein, refers to a substituted or unsubstituted aromatic ring or rings. In certain embodiments, the aromatic ring or rings contain at least one nitrogen atom. In certain embodiments, the nucleotide base is capable of forming Watson-Crick and/or Hoogsteen hydrogen bonds with an appropriately complementary nucleotide base. Exemplary nucleotide bases and analogs thereof include, but are not limited to, naturally occurring nucleotide bases adenine, guanine, cytosine, uracil, thymine, and analogs of the naturally occurring nucleotide bases, e.g., 7-deazaadenine, 7-deazaguanine, 7-deaza-8-azaguanine, 7-deaza-8-azaadenine, N6 -Δ2 -isopentenyladenine (6iA), N6 -Δ2 -isopentenyl-2-methylthioadenine (2ms6iA), N2 -dimethylguanine (dmG), 7-methylguanine (7mG), inosine, nebularine, 2-aminopurine, 2-amino-6-chloropurine, 2,6-diaminopurine, hypoxanthine, pseudouridine, pseudocytosine, pseudoisocytosine, 5-propynylcytosine, isocytosine, isoguanine, 7-deazaguanine, 2-thiopyrimidine, 6-thioguanine, 4-thiothymine, 4-thiouracil, O^6-methylguanine, N^6-methyladenine, O^4-methylthymine, 5,6-dihydrothymine, 5,6-dihydrouracil, pyrazolo[3,4-D]pyrimidines (see, e.g., U.S. Patent Nos. 6,143,877 and 6,127,121 and PCT published application WO 01/38584), ethenoadenine, indoles such as nitroindole and 4-methylindole, and pyrroles such as nitropyrrole. Certain exemplary nucleotide bases can be found, e.g., in Fasman, 1989, Practical Handbook of Biochemistry and Molecular Biology, pp. 385-394, CRC Press, Boca Raton, Fla., and the references cited therein.

[023] The term "nucleotide", as used herein, refers to a compound comprising a nucleotide base linked to the C-1' carbon of a sugar, such as ribose, arabinose, xylose, and pyranose, and sugar analogs thereof. The term nucleotide also encompasses nucleotide analogs. The sugar may be substituted or unsubstituted. Substituted ribose sugars include, but are not

limited to, those riboses in which one or more of the carbon atoms, for example the 2'-carbon atom, is substituted with one or more of the same or different Cl, F, –R, –OR, $–NR_2$ or halogen groups, where each R is independently H, C_1-C_6 alkyl or C_5-C_{14} aryl. Exemplary riboses include, but are not limited to, 2'-(C1 -C6)alkoxyribose, 2'-(C5 -C14)aryloxyribose, 2',3'-didehydroribose, 2'-deoxy-3'-haloribose, 2'-deoxy-3'-fluororibose, 2'-deoxy-3'-chlororibose, 2'-deoxy-3'-aminoribose, 2'-deoxy-3'-(C1 -C6)alkylribose, 2'-deoxy-3'-(C1 -C6)alkoxyribose and 2'-deoxy-3'-(C5 -C14)aryloxyribose, ribose, 2'-deoxyribose, 2',3'-dideoxyribose, 2'-haloribose, 2'-fluororibose, 2'-chlororibose, and 2'-alkylribose, e.g., 2'-O-methyl, 4'-α-anomeric nucleotides, 1'-α-anomeric nucleotides, 2'-4'- and 3'-4'-linked and other "locked" or "LNA", bicyclic sugar modifications (see, e.g., PCT published application nos. WO 98/22489, WO 98/39352;, and WO 99/14226). Exemplary LNA sugar analogs within a polynucleotide include, but are not limited to, the structures:



2'-4' LNA



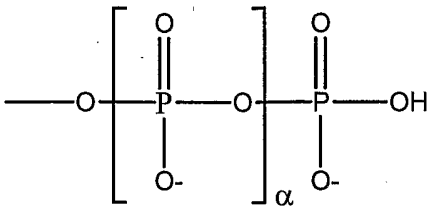
3'-4' LNA

where B is any nucleotide base.

[024] Modifications at the 2'- or 3'-position of ribose include, but are not limited to, hydrogen, hydroxy, methoxy, ethoxy, allyloxy, isopropoxy, butoxy, isobutoxy, methoxyethyl, alkoxy, phenoxy, azido, amino, alkylamino, fluoro, chloro and bromo. Nucleotides include, but are not limited to, the natural D optical isomer, as well as the L optical isomer forms (see, e.g., Garbesi (1993) Nucl. Acids Res. 21:4159-65; Fujimori (1990) J. Amer. Chem. Soc. 112:7435; Urata, (1993) Nucleic Acids Symposium Ser. No. 29:69-70). When the nucleotide base is purine, e.g. A or G, the ribose sugar is attached to the N^9-position of the nucleotide base. When the nucleotide base is pyrimidine, e.g. C, T or U, the pentose sugar is attached to the N^1-position of the nucleotide

base, except for pseudouridines, in which the pentose sugar is attached to the C5 position of the uracil nucleotide base (see, e.g., Kornberg and Baker, (1992) *DNA Replication*, 2nd Ed., Freeman, San Francisco, CA).

[025] One or more of the pentose carbons of a nucleotide may be substituted with a phosphate ester having the formula:



where α is an integer from 0 to 4. In certain embodiments, α is 2 and the phosphate ester is attached to the 3'- or 5'-carbon of the pentose. In certain embodiments, the nucleotides are those in which the nucleotide base is a purine, a 7-deazapurine, a pyrimidine, or an analog thereof. "Nucleotide 5'-triphosphate" refers to a nucleotide with a triphosphate ester group at the 5' position, and are sometimes denoted as "NTP", or "dNTP" and "ddNTP" to particularly point out the structural features of the ribose sugar. The triphosphate ester group may include sulfur substitutions for the various oxygens, e.g. α-thio-nucleotide 5'-triphosphates. For a review of nucleotide chemistry, see, e.g., Shabarova, Z. and Bogdanov, A. *Advanced Organic Chemistry of Nucleic Acids*, VCH, New York, 1994.

[026] The term "nucleotide analog", as used herein, refers to embodiments in which the pentose sugar and/or the nucleotide base and/or one or more of the phosphate esters of a nucleotide may be replaced with its respective analog. In certain embodiments, exemplary pentose sugar analogs are those described above. In certain embodiments, the nucleotide analogs have a nucleotide base analog as described above. In certain embodiments, exemplary phosphate ester analogs include, but are not limited to, alkylphosphonates, methylphosphonates, phosphoramidates, phosphotriesters, phosphorothioates, phosphorodithioates, phosphoroselenoates, phosphorodiselenoates, phosphoroanilothioates,

phosphoroanilidates, phosphoroamidates, boronophosphates, etc., and may include associated counterions.

[027] Also included within the definition of "nucleotide analog" are nucleotide analog monomers which can be polymerized into polynucleotide analogs in which the DNA/RNA phosphate ester and/or sugar phosphate ester backbone is replaced with a different type of internucleotide linkage. Exemplary polynucleotide analogs include, but are not limited to, peptide nucleic acids, in which the sugar phosphate backbone of the polynucleotide is replaced by a peptide backbone.

[028] An "extendable nucleotide" is a nucleotide which is: (i) capable of being enzymatically or synthetically incorporated onto the terminus of a polynucleotide chain, and (ii) capable of supporting further enzymatic or synthetic extension. Extendable nucleotides include nucleotides that have already been enzymatically or synthetically incorporated into a polynucleotide chain, and have either supported further enzymatic or synthetic extension, or are capable of supporting further enzymatic or synthetic extension. Extendable nucleotides include, but are not limited to, nucleotide 5'-triphosphates, e.g., dNTP and NTP, phosphoramidites suitable for chemical synthesis of polynucleotides, and nucleotide units in a polynucleotide chain that have already been incorporated enzymatically or chemically.

[029] The term "nucleotide terminator" or "terminator", as used herein, refers to an enzymatically-incorporable nucleotide, which does not support incorporation of subsequent nucleotides in a primer extension reaction. A terminator is therefore not an extendable nucleotide. In certain embodiments, terminators are those in which the nucleotide is a purine, a 7-deaza-purine, a pyrimidine, or a nucleotide analog, and the sugar moiety is a pentose which includes a 3'-substituent that blocks further synthesis, such as a dideoxynucleotide triphosphate (ddNTP). In certain embodiments, substituents that block further synthesis include, but are not limited to, amino, deoxy, halogen, alkoxy and aryloxy groups. Exemplary terminators include, but are not limited to, those in which the sugar-phosphate ester moiety is 3'-

(C1 -C6)alkylribose-5'-triphosphate, 2'-deoxy-3'-(C1 -C6)alkylribose-5'-triphosphate, 2'-deoxy-3'-(C1 -C6)alkoxyribose-5-triphosphate, 2'-deoxy-3'-(C5 -C14)aryloxyribose-5'-triphosphate, 2'-deoxy-3'-haloribose-5'-triphosphate, 2'-deoxy-3'-aminoribose-5'-triphosphate, 2',3'-dideoxyribose-5'-triphosphate or 2',3'-didehydroribose-5'-triphosphate. Terminators include, but are not limited to, T terminators, including ddTTP and dUTP, which incorporate opposite an adenine, or adenine analog, in a template; A terminators, including ddATP, which incorporate opposite a thymine, uracil, or an analog of thymine or uracil, in the template; C terminators, including ddCTP, which incorporate opposite a guanine, or guanine analog, in the template; and G terminators, including ddGTP and ddITP, which incorporate opposite a cytosine, or cytosine analog, in the template.

[030] The term "label" refers to any moiety which can be attached to a molecule and: (i) provides a detectable signal; (ii) interacts with a second label to modify the detectable signal provided by the second label, e.g. FRET (Fluorescent Resonance Energy Tansfer); (iii) stabilizes hybridization, e.g., duplex formation; or (iv) provides a member of a binding complex or affinity set, e.g., affinity, antibody/antigen, ionic complexation, hapten/ligand, e.g. biotin/avidin. Labeling can be accomplished using any one of a large number of known techniques employing known labels, linkages, linking groups, reagents, reaction conditions, and analysis and purification methods. Labels include, but are not limited to, light-emitting or light-absorbing compounds which generate or quench a detectable fluorescent, chemiluminescent, or bioluminescent signal (see, e.g., Kricka, L. in *Nonisotopic DNA Probe Techniques* (1992), Academic Press, San Diego, pp. 3-28). Fluorescent reporter dyes useful for labelling biomolecules include, but are not limited to, fluoresceins (see, e.g., U.S. Patent Nos. 5,188,934; 6,008,379; and 6,020,481), rhodamines (see, e.g., U.S. Patent Nos. 5,366,860; 5,847,162; 5,936,087; 6,051,719; and 6,191,278), benzophenoxazines (see, e.g., U.S. Patent No. 6,140,500), energy-transfer fluorescent dyes (ETFDs), comprising pairs of donors and acceptors (see, e.g., U.S. Patent Nos. 5,863,727; 5,800,996; and 5,945,526),

and cyanines (see, e.g., Kubista, WO 97/45539), as well as any other fluorescent label capable of generating a detectable signal. Examples of fluorescein dyes include, but are not limited to, 6-carboxyfluorescein; 2',4',1,4,-tetrachlorofluorescein; and 2',4',5',7',1,4-hexachlorofluorescein. Labels also include, but are not limited to, semiconductor nanocrystals, or quantum dots (see, e.g., U.S. Patent Nos. 5,990,479 and 6,207,392 B1; Han et al. *Nature Biotech.* 19: 631-635).

[031] A class of labels are hybridization-stabilizing moieties which serve to enhance, stabilize, or influence hybridization of duplexes, e.g. intercalators, minor-groove binders, and cross-linking functional groups (see, e.g., Blackburn, G. and Gait, M. Eds. "DNA and RNA structure" in *Nucleic Acids in Chemistry and Biology*, 2nd Edition, (1996) Oxford University Press, pp. 15-81). Yet another class of labels effect the separation or immobilization of a molecule by specific or non-specific capture, for example biotin, digoxigenin, and other haptens (see, e.g., Andrus, A. "Chemical methods for 5' non-isotopic labeling of PCR probes and primers" (1995) in *PCR 2: A Practical Approach*, Oxford University Press, Oxford, pp. 39-54). Non-radioactive labelling methods, techniques, and reagents are reviewed in: *Non-Radioactive Labelling, A Practical Introduction*, Garman, A.J. (1997) Academic Press, San Diego.

[032] Labels may be "detectably different", which means that they are distinguishable from one another by at least one detection method. Detectably different labels include, but are not limited to, labels that emit light of different wavelengths, labels that absorb light of different wavelengths, labels that have different fluorescent decay lifetimes, labels that have different spectral signatures, labels that have different radioactive decay properties, labels of different charge, and labels of different size.

[033] The term "labeled terminator", as used herein, refers to a terminator that is physically joined to a label. The linkage to the label is at a site or sites on the terminator that do not prevent the incorporation of the terminator by a polymerase into a polynucleotide.

[034] As used herein, the term "target nucleic acid template" refers to a nucleic acid sequence that serves as a template for a primer extension reaction. Target nucleic acid templates include, but are not limited to, genomic DNA, including mitochondrial DNA and nucleolar DNA, cDNA, synthetic DNA, plasmid DNA, yeast artificial chromosomal DNA (YAC), bacterial artificial chromosomal DNA (BAC), and other extrachromosomal DNA, and primer extension products. Target nucleic acid templates also include, but are not limited to, RNA, synthetic RNA, mRNA, tRNA, and analogs of both RNA and DNA, such as peptide nucleic acids (PNA).

[035] Different target nucleic acid templates may be different portions of a single contiguous nucleic acid or may be on different nucleic acids. Different portions of a single contiguous nucleic acid may overlap.

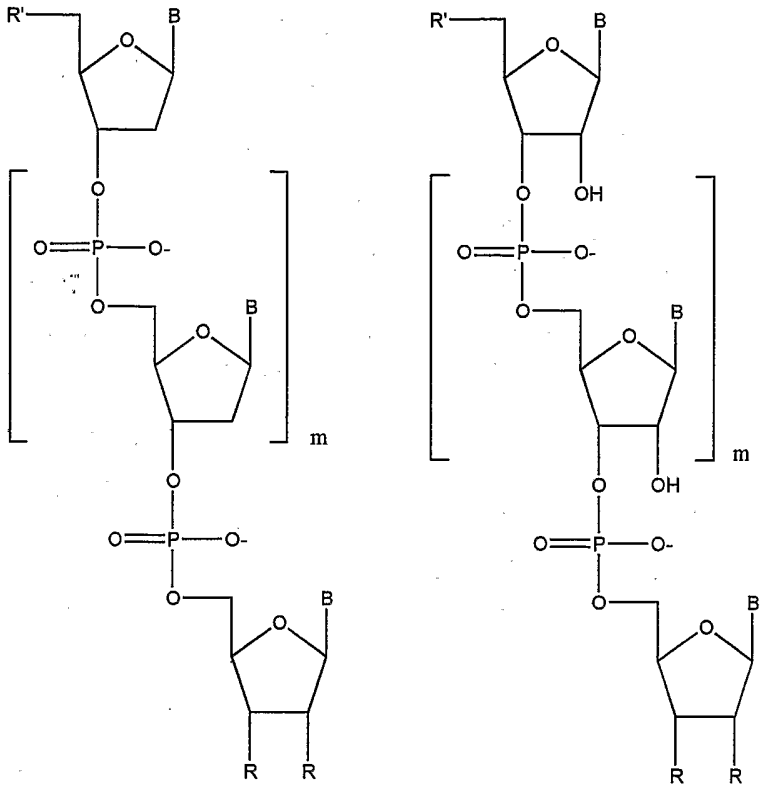
[036] "Primer" as used herein refers to a polynucleotide or oligonucleotide that has a free 3'-OH (or functional equivalent thereof) that can be extended by at least one nucleotide in a primer extension reaction catalyzed by a polymerase. In certain embodiments, primers may be of virtually any length, provided they are sufficiently long to hybridize to a polynucleotide of interest in the environment in which primer extension is to take place. In certain embodiments, primers are at least 14 nucleotides in length. Primers may be specific for a particular sequence, or, alternatively, may be degenerate, e.g., specific for a set of sequences.

[037] The terms "primer extension" and "primer extension reaction" are used interchangeably, and refer to a process of adding one or more nucleotides to a nucleic acid primer, or to a primer extension product, using a polymerase, a template, and one or more nucleotides.

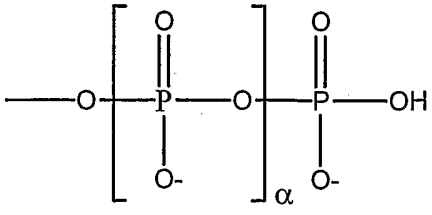
[038] A "primer extension product" is produced when one or more nucleotides has been added to a primer in a primer extension reaction. A primer extension product may serve as a target nucleic acid template in subsequent extension reactions. A primer extension product may include a terminator.

[039] As used herein, the terms "polynucleotide", "oligonucleotide", and "nucleic acid" are used interchangeably and mean single-stranded and double-stranded polymers of nucleotide monomers, including 2'-deoxyribonucleotides (DNA) and ribonucleotides (RNA) linked by internucleotide phosphodiester bond linkages, or internucleotide analogs, and associated counter ions, e.g., H^+, NH_4^+, trialkylammonium, Mg^{2+}, Na^+ and the like. A polynucleotide may be composed entirely of deoxyribonucleotides, entirely of ribonucleotides, or chimeric mixtures thereof. The nucleotide monomer units may comprise any of the nucleotides described herein, including, but not limited to, nucleotides and nucleotide analogs. Polynucleotides typically range in size from a few monomeric units, e.g. 5-40 when they are sometimes referred to in the art as oligonucleotides, to several thousands of monomeric nucleotide units. Unless denoted otherwise, whenever a polynucleotide sequence is represented, it will be understood that the nucleotides are in 5' to 3' order from left to right and that "A" denotes deoxyadenosine or an analog thereof, "C" denotes deoxycytidine or an analog thereof, "G" denotes deoxyguanosine or an analog thereof, and "T" denotes thymidine or an analog thereof, unless otherwise noted.

[040] Polynucleotides may be composed of a single type of sugar moiety, e.g., as in the case of RNA and DNA, or mixtures of different sugar moieties, e.g., as in the case of RNA/DNA chimeras. In certain embodiments, nucleic acids are ribopolynucleotides and 2'-deoxyribopolynucleotides according to the structural formulae below:



wherein each B is independently the base moiety of a nucleotide, e.g., a purine, a 7-deazapurine, a pyrimidine, or an analog thereof; each m defines the length of the respective nucleic acid and can range from zero to thousands, tens of thousands, or even more; each R is independently selected from the group comprising hydrogen, hydroxyl, halogen, --R", --OR", and --NR"R", where each R" is independently (C_1 -C_6) alkyl or (C_5 -$C1_4$) aryl, or two adjacent Rs may be taken together to form a bond such that the ribose sugar is 2',3'-didehydroribose, and each R' may be independently hydroxyl or



where α is zero, one or two.

[041] In certain embodiments of the ribopolynucleotides and 2'-deoxyribopolynucleotides illustrated above, the nucleotide bases B are

covalently attached to the C1' carbon of the sugar moiety as previously described.

[042] The terms "nucleic acid", "polynucleotide", and "oligonucleotide" may also include nucleic acid analogs, polynucleotide analogs, and oligonucleotide analogs. The terms "nucleic acid analog", "polynucleotide analog" and "oligonucleotide analog" are used interchangeably and, as used herein, refer to a polynucleotide that contains at least one nucleotide analog and/or at least one phosphate ester analog and/or at least one pentose sugar analog. Also included within the definition of polynucleotide analogs are polynucleotides in which the phosphate ester and/or sugar phosphate ester linkages are replaced with other types of linkages, such as N-(2-aminoethyl)-glycine amides and other amides (see, e.g., Nielsen et al., 1991, *Science* **254**: 1497-1500; WO 92/20702; U.S. Pat. No. 5,719,262; U.S. Pat. No. 5,698,685;); morpholinos (see, e.g., U.S. Pat. No. 5,698,685; U.S. Pat. No. 5,378,841; U.S. Pat. No. 5,185,144); carbamates (see, e.g., Stirchak & Summerton, 1987, *J. Org. Chem.* **52**: 4202); methylene(methylimino) (see, e.g., Vasseur et al., 1992, *J. Am. Chem. Soc.* **114**: 4006); 3'-thioformacetals (see, e.g., Jones et al., 1993, *J. Org. Chem.* **58**: 2983); sulfamates (see, e.g., U.S. Pat. No. 5,470,967); 2-aminoethylglycine, commonly referred to as PNA (see, e.g., Buchardt, WO 92/20702; Nielsen (1991) Science 254:1497-1500); and others (see, e.g., U.S. Pat. No. 5,817,781; Frier & Altman, 1997, *Nucl. Acids Res.* **25**:4429 and the references cited therein). Phosphate ester analogs include, but are not limited to, (i) C_1–C_4 alkylphosphonate, e.g. methylphosphonate; (ii) phosphoramidate; (iii) C_1–C_6 alkyl-phosphotriester; (iv) phosphorothioate; and (v) phosphorodithioate.

[043] The terms "annealing" and "hybridization" are used interchangeably and mean the base-pairing interaction of one nucleic acid with another nucleic acid that results in formation of a duplex, triplex, or other higher-ordered structure. In certain embodiments, the primary interaction is base specific, e.g., A/T and G/C, by Watson/Crick and Hoogsteen-type

hydrogen bonding. Base-stacking and hydrophobic interactions may also contribute to duplex stability.

[044] The term "variant" as used herein refers to any alteration of a protein, including, but not limited to, changes in amino acid sequence, substitutions of one or more amino acids, addition of one or more amino acids, deletion of one or more amino acids, and alterations to the amino acids themselves. In certain embodiments, the changes involve conservative amino acid substitutions. Conservative amino acid substitution may involve replacing one amino acid with another that has, e.g., similar hydrophobicity, hydrophilicity, charge, or aromaticity. In certain embodiments, conservative amino acid substitutions may be made on the basis of similar hydropathic indices. A hydropathic index takes into account the hydrophobicity and charge characteristics of an amino acid, and in certain embodiments, may be used as a guide for selecting conservative amino acid substitutions. The hydropathic index is discussed, e.g., in Kyte *et al.*, *J. Mol. Biol.*, 157:105-131 (1982). It is understood in the art that conservative amino acid substitutions may be made on the basis of any of the aforementioned characteristics.

[045] Alterations to the amino acids may include, but are not limited to, glycosylation, methylation, phosphorylation, biotinylation, and any covalent and noncovalent additions to a protein that do not result in a change in amino acid sequence. "Amino acid" as used herein refers to any amino acid, natural or nonnatural, that may be incorporated, either enzymatically or synthetically, into a polypeptide or protein.

[046] "Thermostable", as used herein, refers to a polymerase that remains active at a temperature greater than about 37°C. In certain embodiments, the nucleic acid polymerases of the invention remain active at a temperature greater than about 42 °C. In certain embodiments, the nucleic acid polymerases of the invention remain active at a temperature greater than about 50 °C. In certain embodiments, the nucleic acid polymerases of the invention remain active at a temperature greater than about 60 °C. In certain

embodiments, the nucleic acid polymerases of the invention remain active at a temperature greater than about 70 °C.

[047] As used herein, a "unit" of polymerase is defined as the amount of polymerase that will catalyze the incorporation of 10 nmoles of total nucleotide into acid-insoluble form in 30 minutes at 74°C.

[048] As used herein, the "ratio" of one polymerase to another in a composition is determined based on the percentage of the units of the polymerase when the polymerase is used alone. Thus, as a nonlimiting example, if the ratio is 60:40, one calculates the amounts of polymerase A and polymerase B as follows. In this example, polymerase A, when used as the only polymerase in a composition, is used at a concentration of 10 units per 20 µl reaction. In this example, polymerase B, when used as the only polymerase in a composition, is used at a concentration of 20 units per 20 µl reaction. If polymerase A and polymerase B are used together in a 20 µl composition at a ratio of 60:40, then the composition contains 6 units of polymerase A (60% of 10 units) and 8 units of polymerase B (40% of 20 units).

[049] As used herein, "mobility-dependent analysis technique" or "MDAT" means an analytical technique based on differential rates of migration among different analyte types. In certain embodiments, the primer extension products may be separated based on, e.g., mobility, molecular weight, length, sequence, and/or charge. Any method that allows two or more nucleic acid sequences in a mixture to be distinguished, e.g., based on mobility, length, molecular weight, sequence and/or charge, is within the scope of the invention. Exemplary mobility-dependent analysis techniques include, without limitation, electrophoresis, including gel or capillary electrophoresis; chromatography, including as HPLC; mass spectroscopy, including MALDI-TOF; sedimentation, including gradient centrifugation; gel filtration; field-flow fractionation; multi-stage extraction techniques; and the like. In certain embodiments, the MDAT is electrophoresis or chromatography.

Certain Exemplary Embodiments of the Invention

[050] In certain embodiments, the present invention is directed to compositions and methods for generating at least one primer extension product. According to certain embodiments, the present invention provides methods for generating a primer extension product using at least two polymerases. In certain embodiments, the methods employ compositions comprising at least one target nucleic acid template, at least one primer, at least one extendable nucleotide, and at least two polymerases. In certain embodiments, a duplex (double stranded polynucleotide) is formed between a target nucleic acid template and a primer in the composition. In certain embodiments, the primer hybridizes to a predetermined location on the target nucleic acid template.

[051] The composition is then incubated under appropriate reaction conditions, such that one or more extendable nucleotides are incorporated sequentially onto the 3' end of the primer. In certain embodiments, the primer extension products generated by the primer extension reaction may then be separated based on size.

[052] Polymerases may or may not be thermostable. In certain embodiments, polymerases have mutations that reduce discrimination against 3'-dideoxynucleotide terminators as compared with nucleotide triphosphates. In certain embodiments, a polymerase bearing one or more of these mutations may incorporate 3'-deoxynucleotide terminators with greater efficiency than does the wild type polymerase (see, e.g., U.S. Patent 5,885,813 and U.S. Patent 6,265,193, which are herein incorporated by reference for any purpose). In certain embodiments, mutations that reduce discrimination against 3'-dideoxynucleotide terminators are in the nucleotide-binding region of the polymerase. In certain embodiments, the nucleotide-binding region is located from about amino acid 520 to about amino acid 832 of the polymerase.

[053] In certain embodiments, polymerases have mutations that reduce discrimination against fluorescent-labeled nucleotides. In certain

embodiments, a polymerase bearing one or more of these mutations may incorporate fluorescent-labeled nucleotides with greater efficiency than does the wild type polymerase (see, e.g., U.S. Patent 5,885,813 and U.S. Patent 6,265,193, which are herein incorporated by reference for any purpose). In certain embodiments, mutations that reduce discrimination against fluorescent-labeled nucleotides are in the nucleotide-binding region of the polymerase.

[054] In certain embodiments, polymerases have mutations that reduce discrimination against ETFD-labelled terminators.

[055] In certain embodiments, DNA polymerases possess exonuclease activity that may allow them to remove incorporated 3'-deoxynucleotide terminators. In certain embodiments, a polymerase bearing one or more mutations or deletions may have reduced 3'-5' exonuclease activity. In certain embodiments, such mutations or deletions are made in the amino-terminal region of the DNA polymerase. Certain examples of such mutations are described, e.g., in U.S. Patent No. 4,795,699; U.S. Patent No. 5,541,099; and U.S. Patent No. 5,489,523; which are herein incorporated by reference for any purpose. In certain embodiments, such mutations or deletions are made in the region of DNA polymerase that confers 3'-5' exonuclease activity. In certain embodiments, that region is located from about amino acid 1 to about amino acid 272 of the DNA polymerase.

[056] In certain embodiments, the phenylalanine at position 667 of *Thermus aquaticus* DNA polymerase is replaced with a tyrosine. A detailed description of this mutant can be found, e.g., in U.S. Pat. No. 5,614,365, which is herein incorporated by reference for any purpose. This mutant polymerase may conveniently be referred to as *Thermus aquaticus* F667Y, or Taq F667Y (because the phenylalanine (F) at position 667 is replaced with a tyrosine(Y)).

[057] In certain embodiments, the glycine at position 46 of *Thermus aquaticus* DNA polymerase is replaced with an aspartic acid. This mutant

polymerase may conveniently be referred to as *Thermus aquaticus* G46D, or *Taq* G46D.

[058] According to certain embodiments, polymerases include, but are not limited to, AmpliTaq FS, *Thermus thermophilus* HB8, *Thermus oshimai*, *Thermus scotoductus*, *Thermus thermophilus* 1B21, and *Thermus thermophilus* GK24. *Thermus aquaticus* polymerase is shown in Figure 1 (SEQ ID NO: 1). *Thermus thermophilus* HB8 polymerase is shown in Figure 2 (SEQ ID NO: 2). *Thermus oshimai* polymerase is shown in Figure 3 (SEQ ID NO: 3). Polymerases from three strains of *Thermus scotoductus* are shown in Figure 4 (SEQ ID NOs: 4-7). *Thermus thermophilus* 1B21 polymerase is shown in Figure 5 (SEQ ID NO: 8). *Thermus thermophilus* GK24 polymerase is shown in Figure 6 (SEQ ID NO: 9).

[059] AmpliTaq FS is described, e.g., in U.S. Patent No. 5,614,365, which is herein incorporated by reference for any purpose. AmpliTaq FS has a mutation at position 46, where a glycine is replaced with an aspartic acid (G46D), and a mutation at position 667, where a phenylalanine is replaced with a tyrosine (F667Y). Certain embodiments comprise a mutant AmpliTaq FS that has the G46D and F667Y mutations of AmpliTaq FS, and additionally has a mutation at position 681, where the glutamine residue is replaced with an isoleucine. This mutant AmpliTaq FS is referred to as AmpliTaq FS (E681I). Certain embodiments comprise a mutant AmpliTaq FS that has the G46D and F667Y mutations of AmpliTaq FS and has two additional mutations as follows: the threonine residue at position 664 is replaced with an asparagine and the arginine residue at position 660 is replaced with a glycine. This mutant AmpliTaq FS is referred to as AmpliTaq FS (T664N; R660G).

[060] *Thermus thermophilus* HB8 is described, e.g., in U.S. Patent No. 5,789,224, which is herein incorporated by reference for any purpose. Certain embodiments comprise a mutant *Thermus thermophilus* HB8 that has three mutations as follows: the aspartic acid residue at position 18 is replaced with an alanine; the phenylalanine residue at position 669 is replaced with a tyrosine; and the glutamine at position 683 is replaced with an arginine. This

mutant *Thermus thermophilus* HB8 is referred to as *Thermus thermophilus* HB8 (D18A; F669Y; E683R).

[061] Certain embodiments comprise a *Thermus thermophilus* HB8 that has an amino-terminal deletion that removes the first 271 amino acids of the protein, and that has two mutations as follows: the phenylalanine residue at position 669 is replaced with a tyrosine and the glutamine at position 683 is replaced with a tryptophan. This mutant *Thermus thermophilus* HB8 is referred to as *Thermus thermophilus* HB8 (Δ271; F669Y; E683W).

[062] Certain embodiments comprise a mutant *Thermus thermophilus* HB8 that has two mutations as follows: the aspartic acid residue at position 18 is replaced with an alanine and the phenylalanine residue at position 669 is replaced with a tyrosine. This mutant *Thermus thermophilus* HB8 is referred to as *Thermus thermophilus* HB8 (D18A; F669Y).

[063] *Thermus oshimai* is described, e.g., in U.S. Provisional Application No. 60/334,798, filed November 30, 2001, which is herein incorporated by reference for any purpose. Certain embodiments comprise a mutant *Thermus oshimai* that has two mutations as follows: the glycine residue at position 43 is replaced with an aspartic acid and the phenylalanine residue at position 665 is replaced with a tyrosine. This mutant of *Thermus oshimai* is referred to as *Thermus oshimai* (G43D; F665Y).

[064] *Thermus thermophilus* GK24 is described, e.g., in U.S. Provisional Application No. 60/336,046, filed November 30, 2001, which is herein incorporated by reference for any purpose. Certain embodiments comprise a mutant *Thermus thermophilus* GK24 that has two mutations as follows: the glycine residue at position 46 is replaced with an aspartic acid and the phenylalanine residue at position 669 is replaced with a tyrosine. This mutant of *Thermus thermophilus* GK24 is referred to as *Thermus thermophilus* GK24 (G46D; F669Y).

[065] *Thermus thermophilus* 1B21 is described, e.g., in U.S. Provisional Application No. 60/336,046, filed November 30, 2001, which is herein incorporated by reference for any purpose. Certain embodiments

comprise a mutant *Thermus thermophilus* 1B21 that has two mutations as follows: the glycine residue at position 46 is replaced with an aspartic acid and the phenylalanine residue at position 669 is replaced with a tyrosine. This mutant of *Thermus thermophilus* 1B21 is referred to as *Thermus thermophilus* 1B21 (G46D; F669Y).

[066] *Thermus scotoductus* is described, e.g., in U.S. Provisional Application No. 60/334,489, filed November 30, 2001, which is herein incorporated by reference for any purpose. In certain embodiments, *Thermus scotoductus* polymerase may be from any of strains X-1 (SEQ ID NOs: 4 and 5), SM3 (SEQ ID NO: 6), and Vi7a (SEQ ID NO: 7). In certain embodiments, *Thermus scotoductus* polymerase from strain X-1 may lack the first three amino terminal amino acids (SEQ ID NO: 5). Certain embodiments of the invention comprise a mutant *Thermus scotoductus* that has two mutations as follows: the glycine residue at position 46 is replaced with an aspartic acid and the phenylalanine residue at position 668 is replaced with a tyrosine. This mutant of *Thermus scotoductus* is referred to as *Thermus scotoductus* (G46D; F668Y). In certain embodiments, when the *Thermus scotoductus* polymerase from strain X-1 lacking the amino-terminal three amino acids is used, the glycine residue at position 43 is replaced with an aspartic acid and the phenylalanine residue at position 665 is replaced with a tyrosine. For convenience of nomenclature, this mutant is also referred to as *Thermus scotoductus* (G46D; F668Y).

[067] In certain embodiments, a composition comprises at least two polymerases selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), *Thermus thermophilus* GK24 (G46D; F669Y), AmpliTaq FS (E681I), and AmpliTaq FS (T664N; R660G).

[068] In certain embodiments, the at least two polymerases are two polymerases. The two polymerases may be present in a composition at any

ratio. In certain embodiments, the two polymerases are present in a composition at a ratio of 1:99. In certain embodiments, the ratio is 10:90. In certain embodiments, the ratio is 20:80. In certain embodiments, the ratio is 30:70. In certain embodiments, the ratio is 40:60. In certain embodiments, the ratio is 50:50.

[069] In certain embodiments, a composition comprises *Thermus thermophilus* GK24 (G46D; F669Y) and *Thermus oshimai* (G43D; F665Y).

[070] In certain embodiments, a composition comprises at least three polymerases selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), *Thermus thermophilus* GK24 (G46D; F669Y), AmpliTaq FS (E681I), and AmpliTaq FS (T664N; R660G). In certain embodiments, the at least three polymerases are three polymerases. The three polymerases may be present in a composition at any ratio.

[071] In certain embodiments, the three polymerases are AmpliTaq FS (E681I), AmpliTaq FS (T664N; R660G), and at least one polymerase selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), and *Thermus thermophilus* GK24 (G46D; F669Y).

[072] In certain embodiments, the combination of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) is referred to as FS-I/FS-NG. The ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG may be any ratio. In certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 2:1. In certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 99:1. In certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 90:10. In

certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 80:20. In certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 70:30. In certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 60:40. In certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 50:50. In certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 40:60. In certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 30:70. In certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 20:80. In certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 10:90. In certain embodiments, the ratio of AmpliTaq FS (E681I) and AmpliTaq FS (T664N; R660G) in FS-I/FS-NG is 1:99.

[073] In certain embodiments, FS-I/FS-NG is combined with at least one polymerase selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), and *Thermus thermophilus* GK24 (G46D; F669Y). In certain embodiments, FS-I/FS-NG is combined with one polymerase selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), and *Thermus thermophilus* GK24 (G46D; F669Y).

[074] FS-I/FS-NG and the polymerase selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus*

thermophilus 1B21 (G46D; F669Y), and *Thermus thermophilus* GK24 (G46D; F669Y) may be combined at any ratio. In certain embodiments, FS-I/FS-NG and the polymerase selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), and *Thermus thermophilus* GK24 (G46D; F669Y) are combined at a ratio of 1:99. In certain embodiments, the ratio is 10:90. In certain embodiments, the ratio is 20:80. In certain embodiments, the ratio is 30:70. In certain embodiments, the ratio is 40:60. In certain embodiments, the ratio is 50:50. In certain embodiments, the ratio is 60:40. In certain embodiments, the ratio is 70:30. In certain embodiments, the ratio is 80:20. In certain embodiments, the ratio is 90:10. In certain embodiments, the ratio is 99:1.

[075] In certain embodiments, a composition comprises FS-I/FS-NG and *Thermus oshimai* (G43D; F665Y). In certain embodiments, a composition comprises FS-I/FS-NG and *Thermus thermophilus* GK24 (G46D; F669Y).

[076] In certain embodiments, a composition comprises more than three polymerases selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), *Thermus thermophilus* GK24 (G46D; F669Y), AmpliTaq FS (E681I), and AmpliTaq FS (T664N; R660G).

[077] In certain embodiments, the sequence of a nucleic acid may be determined by generating primer extension products. For example, in certain embodiments, one may employ the method of Sanger (see, e.g., Sanger et al. *Proc. Nat. Acad. Sci* 74: 5463-5467 (1977)). According to certain embodiments, the present invention provides methods for sequencing a target nucleic acid template using at least two polymerases. In certain embodiments, the methods employ a composition comprising at least one

target nucleic acid template, at least one primer, at least one extendable nucleotide, at least one terminator, and at least two polymerases. In certain embodiments, a duplex (double stranded polynucleotide) is formed between a target nucleic acid template and a primer in the composition. In certain embodiments, the primer hybridizes to a predetermined location on the target nucleic acid template.

[078] The composition is incubated under appropriate reaction conditions, such that one or more extendable nucleotides are incorporated sequentially onto the 3' end of the primer. In certain embodiments, a terminator may be incorporated into the primer extension product, and once incorporated, prevents further incorporation of nucleotides to the 3' end of the primer extension product by polymerase. In certain embodiments, the primer extension products generated by the primer extension reaction may then be separated based on size. In certain embodiments, the sequence of the nucleic acid template may be determined from the particular sizes of the products and the identity of the terminator on each product.

[079] In certain embodiments, a composition of the invention comprises at least two polymerases, at least one extendable nucleotide, and at least one terminator. In certain embodiments, the at least one extendable nucleotide is selected from dATP, dCTP, dITP, dGTP, dUTP, and dTTP. In certain embodiments, the composition comprises extendable nucleotides dATP, dCTP, dITP, and dUTP. In certain embodiments, the composition comprises extendable nucleotides dATP, dCTP, dITP, and dTTP. In certain embodiments, the at least one terminator is selected from A terminators, C terminators, G terminators, and T terminators. In certain embodiments, the at least one terminator further comprises a label. In certain embodiments, the at least one terminator further comprises an energy-transfer fluorescent dye (ETFD) label. In certain embodiments, the composition comprises an A terminator, a C terminator, a G terminator, and a T terminator. In certain embodiments, each of the different terminators further comprises a detectably different label. In certain embodiments, each of the different terminators

further comprises a detectably different ETFD label. In certain embodiments, the composition contains four different ETFD-labeled terminators, e.g., an ETFD-labeled A terminator, an ETFD-labeled C terminator, an ETFD-labeled G terminator, and an ETFD-labeled T terminator, where each ETFD is detectably different.

[080] In certain embodiments, a composition further comprises at least one buffering agent. In certain embodiments, the at least one buffering agent is selected from Tris and Tricine. In certain embodiments, a composition further comprises at least one type of divalent cation. In certain embodiments, the at least one type of divalent cation is selected from Mg^{2+} and Mn^{2+}. In certain embodiments, a composition further comprises at least one additive. In certain embodiments, the at least one additive is selected from glycerol and DMSO.

[081] In certain embodiments, a composition comprises 80 mM Tris having a pH in the range of 8-9; 5 mM $MgCl_2$; 0-10% glycerol; 200 µM dATP; 200 µM dCTP; 300 µM dITP; 200 µM dUTP; 25 nM-1225 nM of each of an ETFD-labeled A terminator, an ETFD-labeled C terminator, an ETFD-labeled G terminator, and an ETFD-labeled T terminator; and 1.5-60 units of each of at least two polymerases in a 20 µl reaction volume. In certain embodiments, the composition further comprises thermoplasma acidophilum inorganic pyrophosphatase (TAP). In certain embodiments, one uses the buffer, extendable nucleotides, and terminators from the ABI PRISM BigDye™ Terminators v. 3.0 Cycle Sequencing Kit (Applied Biosystems, Cat. No. 4390236), and replaces the kit's polymerase with at least two polymerases selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), *Thermus thermophilus* GK24 (G46D; F669Y), AmpliTaq FS (E681I), and AmpliTaq FS (T664N; R660G).

[082] In certain embodiments, methods are provided for sequencing a target nucleic acid template. In certain embodiments, such methods comprise: forming a composition comprising a target nucleic acid template, at least one primer, at least one extendable nucleotide, at least one terminator, and at least two polymerases selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), *Thermus thermophilus* GK24 (G46D; F669Y), AmpliTaq FS (E681I), and AmpliTaq FS (T664N; R660G); and incubating the composition under appropriate conditions to generate at least one primer extension product.

[083] In certain embodiments, the methods include cycle sequencing, in which, following the primer extension reaction and termination, the primer extension product is released from the target nucleic acid template, and a new primer is annealed, extended, and terminated. Cycle sequencing is but one example of amplification of primer extension products. In certain embodiments, cycle sequencing is performed using a thermocycler apparatus. Certain cycle sequencing reactions are described, e.g., in U.S. Patent No. 5,741,640; U.S. Patent No. 5,741,676; U.S. Patent No. 5,756,285; U.S. Patent No. 5,674,679; and U.S. Patent No. 5,998,143; which are herein incorporated by reference for any purpose.

[084] In cycle sequencing, an incubation cycle comprises two or more incubations, each incubation comprising a certain temperature for a certain period of time. In certain embodiments, one such incubation cycle comprises 95°C for 20 seconds, followed by 50°C for 15 seconds, followed by 60°C for 4 minutes. In certain embodiments, cycle sequencing comprises repeating the incubation cycle 25 times.

[085] In certain embodiments, the primer extension products may be separated by a mobility-dependent analysis technique, or MDAT. In certain embodiments, the MDAT is electrophoresis. In certain embodiments, by

separating the primer extension products, one can determine the sequence of the template nucleic acid based on the size of each product and the identity of the terminator at its 3' end. In certain embodiments, when the terminator is a labeled terminator, the identity of the terminator at the 3' end is determined by the identity of the label.

Kits

[086] The invention also provides kits. In certain embodiments, kits serve to expedite the performance of the methods of interest by assembling two or more components used to carry out the methods. In certain embodiments, kits contain components in pre-measured unit amounts to minimize the need for measurements by end-users. In certain embodiments, kits include instructions for performing one or more methods. In certain embodiments, the kit components are optimized to operate in conjunction with one another.

[087] In certain embodiments, kits comprise at least two polymerases selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R), *Thermus oshimai* (G43D; F665Y), *Thermus thermophilus* HB8 (Δ271; F669Y; E683W), *Thermus thermophilus* HB8 (D18A; F669Y), *Thermus scotoductus* (G46D; F668Y), *Thermus thermophilus* 1B21 (G46D; F669Y), *Thermus thermophilus* GK24 (G46D; F669Y), AmpliTaq FS (E681I), and AmpliTaq FS (T664N; R660G).

[088] In certain embodiments, the kits may be used to sequence at least one target nucleic acid template. In certain embodiments, the kits further comprise at least one terminator. In certain embodiments, the at least one terminator is a labeled terminator. In certain embodiments, the at least one terminator is selected from an ETFD-labeled A terminator, an ETFD-labeled C terminator, an ETFD-labeled G terminator, and an ETFD-labeled T terminator. In certain embodiments, kits may contain additional components, including, but not limited to, at least one primer and/or at least one extendable nucleotide. In certain embodiments, kits may also include reagents for

performing a control reaction, which may include one or more of the above components, and at least one target nucleic acid template.

[089] Other embodiments of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. It is intended that the specification and examples be considered as exemplary only, with the true scope and spirit of the invention being indicated by the following claims.

What is claimed is:

1. A composition comprising

a) at least one first polymerase selected from *Thermus thermophilus* HB8 polymerase, *Thermus thermophilus* HB8 mutant polymerase, *Thermus oshimai* polymerase, *Thermus oshimai* mutant polymerase, *Thermus scotoductus* polymerase, *Thermus scotoductus* mutant polymerase, *Thermus thermophilus* 1B21 polymerase, *Thermus thermophilus* 1B21 mutant polymerase, *Thermus thermophilus* GK24 polymerase, and *Thermus thermophilus* GK24 mutant polymerase; and

b) at least one second polymerase selected from *Thermus aquaticus* polymerase and *Thermus aquaticus* mutant polymerase.

2. The composition of claim 1, wherein the composition comprises at least one polymerase comprising at least one mutation that reduces 3'-5' exonuclease activity.

3. The composition of claim 1, wherein the composition comprises at least one polymerase comprising at least one deletion that reduces 3'-5' exonuclease activity.

4. The composition of claim 1, wherein the at least one first polymerase is selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R) mutant polymerase, *Thermus thermophilus* HB8 (Δ271; F669Y; E683W) mutant polymerase, *Thermus oshimai* (G43D; F665Y) mutant polymerase, *Thermus scotoductus* (G46D; F668Y) mutant polymerase, *Thermus thermophilus* 1B21 (G46D; F669Y) mutant polymerase, and *Thermus thermophilus* GK24 (G46D; F669Y) mutant polymerase.

5. The composition of any of claims 1-4, wherein the at least one second polymerase is selected from *Thermus aquaticus* (G46D; F667Y) mutant polymerase, *Thermus aquaticus* (G46D; F667Y; E681I) mutant polymerase, and *Thermus aquaticus* (G46D; F667Y; T664N; R660G) mutant polymerase.

6. A composition comprising at least two polymerases selected from *Thermus thermophilus* HB8 polymerase, *Thermus thermophilus* HB8

mutant polymerase, *Thermus oshimai* polymerase, *Thermus oshimai* mutant polymerase, *Thermus scotoductus* polymerase, *Thermus scotoductus* mutant polymerase, *Thermus thermophilus* 1B21 polymerase, *Thermus thermophilus* 1B21 mutant polymerase, *Thermus thermophilus* GK24 polymerase, and *Thermus thermophilus* GK24 mutant polymerase.

7. The composition of claim 6, wherein at least one of the at least two polymerases comprises at least one mutation that reduces 3'-5' exonuclease activity.

8. The composition of claim 6, wherein at least one of the at least two polymerases comprises at least one deletion that reduces 3'-5' exonuclease activity.

9. The composition of claim 6, wherein the at least two polymerases are selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R) mutant polymerase, *Thermus thermophilus* HB8 (Δ271; F669Y; E683W) mutant polymerase, *Thermus oshimai* (G43D; F665Y) mutant polymerase, *Thermus scotoductus* (G46D; F668Y) mutant polymerase, *Thermus thermophilus* 1B21 (G46D; F669Y) mutant polymerase, *Thermus thermophilus* GK24 (G46D; F669Y) mutant polymerase.

10. The composition of claim 9, wherein the at least two polymerases are *Thermus thermophilus* GK24 (G46D; F669Y) mutant polymerase and *Thermus oshimai* (G43D; F665Y) mutant polymerase.

11. The composition of any of claims 6-10, further comprising at least one second polymerase selected from *Thermus aquaticus* polymerase and *Thermus aquaticus* mutant polymerase.

12. The composition of claim 11, wherein the at least one second polymerase is selected from *Thermus aquaticus* (G46D; F667Y) mutant polymerase, *Thermus aquaticus* (G46D; F667Y; E681I) mutant polymerase, and *Thermus aquaticus* (G46D; F667Y; T664N; R660G) mutant polymerase.

13. A composition comprising *Thermus aquaticus* (G46D; F667Y; E681I) mutant polymerase, *Thermus aquaticus* (G46D; F667Y; T664N; R660G) mutant polymerase, and at least one polymerase selected from

Thermus thermophilus HB8 polymerase, *Thermus thermophilus* HB8 mutant polymerase, *Thermus oshimai* polymerase, *Thermus oshimai* mutant polymerase, *Thermus scotoductus* polymerase, *Thermus scotoductus* mutant polymerase, *Thermus thermophilus* 1B21 polymerase, *Thermus thermophilus* 1B21 mutant polymerase, *Thermus thermophilus* GK24 polymerase, and *Thermus thermophilus* GK24 mutant polymerase.

14. The composition of claim 13, wherein at least one of the at least one polymerase comprises at least one mutation that reduces 3'-5' exonuclease activity.

15. The composition of claim 13, wherein at least one of the at least one polymerase comprises at least one deletion that reduces 3'-5' exonuclease activity.

16. The composition of claim 13, wherein the at least one polymerase is selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R) mutant polymerase, *Thermus thermophilus* HB8 (Δ271; F669Y; E683W) mutant polymerase, *Thermus oshimai* (G43D; F665Y) mutant polymerase, *Thermus scotoductus* (G46D; F668Y) mutant polymerase, *Thermus thermophilus* 1B21 (G46D; F669Y) mutant polymerase, and *Thermus thermophilus* GK24 (G46D; F669Y) mutant polymerase.

17. A method of sequencing a target nucleic acid template comprising:

(a) forming a composition comprising a target nucleic acid template; at least one primer; at least one extendable nucleotide; at least one terminator; at least one first polymerase selected from *Thermus thermophilus* HB8 polymerase, *Thermus thermophilus* HB8 mutant polymerase, *Thermus oshimai* polymerase, *Thermus oshimai* mutant polymerase, *Thermus scotoductus* polymerase, *Thermus scotoductus* mutant polymerase, *Thermus thermophilus* 1B21 polymerase, *Thermus thermophilus* 1B21 mutant polymerase, *Thermus thermophilus* GK24 polymerase, and *Thermus thermophilus* GK24 mutant polymerase; and at least one second polymerase

selected from *Thermus aquaticus* polymerase and *Thermus aquaticus* mutant polymerase;

(b) incubating the composition under appropriate conditions to generate at least one primer extension product;

(c) separating the at least one primer extension product, wherein the separating comprises at least one mobility-dependent analysis technique (MDAT);

(d) detecting the at least one primer extension product; and

(e) determining the sequence of the target nucleic acid template.

18. The method of claim 17, wherein the composition comprises at least one polymerase comprising at least one mutation that reduces 3'-5' exonuclease activity.

19. The method of claim 17, wherein the composition comprises at least one polymerase comprising at least one deletion that reduces 3'-5' exonuclease activity.

20. The method of claim 17, wherein the at least one first polymerase is selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R) mutant polymerase, *Thermus thermophilus* HB8 (Δ271; F669Y; E683W) mutant polymerase, *Thermus oshimai* (G43D; F665Y) mutant polymerase, *Thermus scotoductus* (G46D; F668Y) mutant polymerase, *Thermus thermophilus* 1B21 (G46D; F669Y) mutant polymerase, and *Thermus thermophilus* GK24 (G46D; F669Y) mutant polymerase.

21. The method of claim 17, wherein the at least one second polymerase is selected from *Thermus aquaticus* (G46D; F667Y) mutant polymerase, *Thermus aquaticus* (G46D; F667Y; E681I) mutant polymerase, and *Thermus aquaticus* (G46D; F667Y; T664N; R660G) mutant polymerase.

22. A method of sequencing a target nucleic acid template comprising:

(a) forming a composition comprising a target nucleic acid template; at least one primer; at least one extendable nucleotide; at least two first polymerases selected from *Thermus thermophilus* HB8 polymerase, *Thermus*

thermophilus HB8 mutant polymerase, *Thermus oshimai* polymerase, *Thermus oshimai* mutant polymerase, *Thermus scotoductus* polymerase, *Thermus scotoductus* mutant polymerase, *Thermus thermophilus* 1B21 polymerase, *Thermus thermophilus* 1B21 mutant polymerase, *Thermus thermophilus* GK24 polymerase, and *Thermus thermophilus* GK24 mutant polymerase;

(b) incubating the composition under appropriate conditions to generate at least one primer extension product;

(c) separating the at least one primer extension product, wherein the separating comprises at least one mobility-dependent analysis technique (MDAT);

(d) detecting the at least one primer extension product; and

(e) determining the sequence of the target nucleic acid template.

23. The method of claim 22, wherein the at least two first polymerases are selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R) mutant polymerase, *Thermus thermophilus* HB8 (Δ271; F669Y; E683W) mutant polymerase, *Thermus oshimai* (G43D; F665Y) mutant polymerase, *Thermus scotoductus* (G46D; F668Y) mutant polymerase, *Thermus thermophilus* 1B21 (G46D; F669Y) mutant polymerase, and *Thermus thermophilus* GK24 (G46D; F669Y) mutant polymerase.

24. The method of claim 22 or 23 wherein the composition further comprises at least one second polymerase selected from *Thermus aquaticus* polymerase and *Thermus aquaticus* mutant polymerase.

25. The method of claim 24, wherein the at least one second polymerase is selected from *Thermus aquaticus* (G46D; F667Y) mutant polymerase, *Thermus aquaticus* (G46D; F667Y; E681I) mutant polymerase, and *Thermus aquaticus* (G46D; F667Y; T664N; R660G) mutant polymerase.

26. A kit comprising

a) at least one first polymerase selected from *Thermus thermophilus* HB8 polymerase, *Thermus thermophilus* HB8 mutant polymerase, *Thermus oshimai* polymerase, *Thermus oshimai* mutant polymerase, *Thermus*

scotoductus polymerase, *Thermus scotoductus* mutant polymerase, *Thermus thermophilus* 1B21 polymerase, *Thermus thermophilus* 1B21 mutant polymerase, *Thermus thermophilus* GK24 polymerase, and *Thermus thermophilus* GK24 mutant polymerase; and

b) at least one second polymerase selected from *Thermus aquaticus* polymerase and *Thermus aquaticus* mutant polymerase. .

27. The kit of claim 26, wherein the at least one first polymerase is selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R) mutant polymerase, *Thermus thermophilus* HB8 (Δ271; F669Y; E683W) mutant polymerase, *Thermus oshimai* (G43D; F665Y) mutant polymerase, *Thermus scotoductus* (G46D; F668Y) mutant polymerase, *Thermus thermophilus* 1B21 (G46D; F669Y) mutant polymerase, and *Thermus thermophilus* GK24 (G46D; F669Y) mutant polymerase.

28. A kit comprising at least two first polymerases selected from *Thermus thermophilus* HB8 polymerase, *Thermus thermophilus* HB8 mutant polymerase, *Thermus oshimai* polymerase, *Thermus oshimai* mutant polymerase, *Thermus scotoductus* polymerase, *Thermus scotoductus* mutant polymerase, *Thermus thermophilus* 1B21 polymerase, *Thermus thermophilus* 1B21 mutant polymerase, *Thermus thermophilus* GK24 polymerase, and *Thermus thermophilus* GK24 mutant polymerase.

29. The kit of claim 28, wherein the at least two first polymerases are selected from *Thermus thermophilus* HB8 (D18A; F669Y; E683R) mutant polymerase, *Thermus thermophilus* HB8 (Δ271; F669Y; E683W) mutant polymerase, *Thermus oshimai* (G43D; F665Y) mutant polymerase, *Thermus scotoductus* (G46D; F668Y) mutant polymerase, *Thermus thermophilus* 1B21 (G46D; F669Y) mutant polymerase, and *Thermus thermophilus* GK24 (G46D; F669Y) mutant polymerase.

30. The kit of claim 28 or 29, further comprising at least one second polymerase selected from *Thermus aquaticus* polymerase and *Thermus aquaticus* mutant polymerase.

31. The kit of claim 30, wherein the at least one second polymerase is selected from *Thermus aquaticus* (G46D; F667Y) mutant polymerase, *Thermus aquaticus* (G46D; F667Y; E681I) mutant polymerase, and *Thermus aquaticus* (G46D; F667Y; T664N; R660G) mutant polymerase.

32. A kit comprising *Thermus aquaticus* (G46D; F667Y; E681I) mutant polymerase, *Thermus aquaticus* (G46D; F667Y; T664N; R660G) mutant polymerase, and at least one polymerase selected from *Thermus thermophilus* HB8 polymerase, *Thermus thermophilus* HB8 mutant polymerase, *Thermus oshimai* polymerase, *Thermus oshimai* mutant polymerase, *Thermus scotoductus* polymerase, *Thermus scotoductus* mutant polymerase, *Thermus thermophilus* 1B21 polymerase, *Thermus thermophilus* 1B21 mutant polymerase, *Thermus thermophilus* GK24 polymerase, and *Thermus thermophilus* GK24 mutant polymerase.

```
1    mrgmlplfep kgrvllvdgh hlayrtfhal kglttsrgep vqavygfaks llkalkedgd
61   avivvfdaka psfrheaygg ykagraptpe dfprqlalik elvdllglar levpgyeadd
121  vlaslakkae kegyevrilt adkdlyqlls drihvlhpeg ylitpawlwe kyglrpdqwa
181  dyraltgdes dnlpgvkgig ektarkllee wgsleallkn ldrlkpaire kilahmddlk
241  lswdlakvrt dlplevdfak rrepdrerlr aflerlefgs llhefglles pkaleeapwp
301  ppegafvgfv lsrkepmwad llalaaargg rvhrapepyk alrdlkearg llakdlsvla
361  lreglglppg ddpmllayll dpsnttpegv arryggewte eageraalse rlfanlwgrl
421  egeerllwly reverplsav lahmeatgvr ldvaylrals levaeeiarl eaevfrlagh
481  pfnlnsrdql ervlfdelgl paigktektg krstsaavle alreahpive kilqyreltk
541  lkstyidplp dlihprtgrl htrfnqtata tgrlsssdpn lqnipvrtpl gqrirrafia
601  eegwllvald ysqielrvla hlsgdenlir vfqegrdiht etaswmfgvp reavdplmrr
661  aaktinfgvl ygmsahrlsq elaipyeeaq afieryfqsf pkvrawiekt leegrrrgyv
721  etlfgrrryv pdlearvksv reaaermafn mpvqgtaadl mklamvklfp rleemgarml
781  lqvhdelvle apkeraeava rlakevmegv yplavpleve vgigedwlsa ke
```

FIG. 1A

```
   1  atgaggggga tgctgcccct ctttgagccc aagggccggg tcctcctggt ggacggccac
  61  cacctggcct accgcacctt ccacgccctg aagggcctca ccaccagccg gggggagccg
 121  gtgcaggcgg tctacggctt cgccaagagc ctcctcaagg ccctcaagga ggacggggac
 181  gcggtgatcg tggtctttga cgccaaggcc ccctccttcc gccacgaggc ctacgggggg
 241  tacaaggcgg gccgggcccc cacgccggag gactttcccc ggcaactcgc cctcatcaag
 301  gagctggtgg acctcctggg gctggcgcgc ctcgaggtcc cgggctacga ggcggacgac
 361  gtcctggcca gcctggccaa gaaggcggaa aaggagggct acgaggtccg catcctcacc
 421  gccgacaaag acctttacca gctcctttcc gaccgcatcc acgccctcca ccccgagggg
 481  tacctcatca ccccggcctg gctttgggaa aagtacggcc tgaggcccga ccagtgggcc
 541  gactaccggg ccctgaccgg ggacgagtcc gacaaccttc ccggggtcaa gggcatcggg
 601  gagaagacgg cgaggaagct tctggaggag tgggggagcc tggaagccct cctcaagaac
 661  ctggaccggc tgaagcccgc catccgggag aagatcctgg cccacatgga cgatctgaag
 721  ctctcctggg acctggccaa ggtgcgcacc gacctgcccc tggaggtgga cttcgccaaa
 781  aggcgggagc ccgaccggga gaggcttagg gcctttctgg agaggcttga gtttggcagc
 841  ctcctccacg agttcggcct tctggaaagc cccaaggccc tggaggaggc cccctggccc
 901  ccgccggaag gggccttcgt gggctttgtg ctttcccgca aggagcccat gtgggccgat
 961  cttctggccc tggccgccgc caggggggcc cgggtccacc gggcccccga gccttataaa
1021  gccctcaggg acctgaagga ggcgcggggg cttctcgcca aagacctgag cgttctggcc
1081  ctgagggaag gccttggcct cccgcccggc gacgacccca tgctcctcgc ctacctcctg
1141  gacccttcca acaccacccc cgaggggtg gcccggcgct acggcgggga gtggacggag
1201  gaggcggggg agcgggccgc cctttccgag aggctcttcg ccaacctgtg ggggaggctt
1261  gaggggggagg agaggctcct ttggctttac cgggaggtgg agaggcccct ttccgctgtc
1321  ctggcccaca tggaggccac gggggtgcgc ctggacgtgg cctatctcag ggccttgtcc
1381  ctggaggtgg ccgaggagat cgcccgcctc gaggccgagg tcttccgcct ggccggccac
1441  cccttcaacc tcaactcccg ggaccagctg gaaagggtcc tctttgacga gctagggctt
1501  cccgccatcg gcaagacgga gaagaccggc aagcgctcca ccagcgccgc cgtcctggag
1561  gccctccgcg aggcccaccc catcgtggag aagatcctgc agtaccggga gctcaccaag
1621  ctgaagagca cctacattga ccccttgccg gacctcatcc accccaggac gggccgcctc
1681  cacacccgct tcaaccagac ggccacggcc acgggcaggc taagtagctc cgatcccaac
1741  ctccagaaca tccccgtccg caccccgctt gggcagagga tccgccgggc cttcatcgcc
1801  gaggaggggt ggctattggt ggccctggac tatagccaga tagagctcag ggtgctggcc
1861  cacctctccg gcgacgagaa cctgatccgg gtcttccagg aggggcggga catccacacg
1921  gagaccgcca gctggatgtt cggcgtcccc cgggaggccg tggacccccc gatgcgccgg
1981  gcggccaaga ccatcaactt cggggtcctc tacggcatgt cggcccaccg cctctcccag
2041  gagctagcca tcccttacga ggaggcccag gccttcattg agcgctactt tcagagcttc
2101  cccaaggtgc gggcctggat tgagaagacc ctggaggagg gcaggaggcg ggggtacgtg
2161  gagaccctct tcggccgccg ccgctacgtg ccagacctag aggcccgggt gaagagcgtg
2221  cgggaggcgg ccgagcgcat ggccttcaac atgcccgtcc agggcaccgc cgccgacctc
```

FIG. 1B-1

```
2281  atgaagctgg ctatggtgaa gctcttcccc aggctggagg aaatgggggc caggatgctc
2341  cttcaggtcc acgacgagct ggtcctcgag gccccaaaag agagggcgga ggccgtggcc
2401  cggctggcca aggaggtcat ggaggggtg tatcccctgg ccgtgccct ggaggtggag
2461  gtggggatag gggaggactg gctctccgcc aaggagtgat accaccccat gctggcccaa
2521  gccagcatgg gggccccggc aaaaggtttc tggggaagtt accaggcatg gtgggccgaa
2581  ggaaacagga aacaagggta tgagggtttt ttgccctaaa gaaaggccag ggggtcctcc
2641  cgaaggaagg cttccagggg gataccccct gggcccagga gtagcccctt tcctccaaag
2701  gcctgggtga aggcttttag ccctttggtc ctttgggaag gggcgctttt gacctccaaa
2761  gccagaaggc gccttccctt cttcaagacg aagtcaacct cctggtccct ttcccgccag
2821  tagtacacct caaagccccc gtgggggccg tgggccagaa ggtgggcgcc cacggcggtt
2881  tccacgagcc gccccatgag ggcggcgtcc gcccacacct cctttgggga aagccccaag
2941  accgccgtga tgagccccgt attaagggcc aggagcttgg ggctcgaggc gcggcgccgg
3001  aagggctcgg cggcgtactt ctgcag
```

FIG. 1B-2

-61
CCGAGGAGACCTACACCCCGGAGGGCGTGCGCTTCGCCCTCCTCCTCCCCAAGCCCGAGC

-1
GGGAAGGTTTCCTCAGGGCGCTCCTGGACGCCACCCGGGGACAGGTGGCCCTGGAGTAGC

1 60
ATGGAGGCGATGCTTCCGCTCTTTGAACCCAAAGGCCGGGTCCTCCTGGTGGACGGCCAC
Met Glu Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val Asp Gly His
1 20

120
CACCTGGCCTACCGCACCTTCTTCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCG
His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Leu Thr Thr Ser Arg Gly Glu Pro
40

180
GTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGGCCCTGAAGGAGGACGGGTAC
Val Gln Ala Val Tyr Gly Phe Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr
60

240
AAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACGAGGCCTACGAG
Lys Ala Val Phe Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu
80

300
GCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATC
Ala Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu Ala Leu Ile
100

360
AAGGAGCTGGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGAC
Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu Glu Val Pro Gly Tyr Glu Ala Asp
120

420
GACGTTCTCGCCACCCTGGCCAAGAAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTC
Asp Val Leu Ala Thr Leu Ala Lys Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu
140

FIG. 2-1

480
ACCGCCGACCGCGACCTCTACCAACTCGTCTCCGACCGCGTCGCCGTCCTCCACCCCGAG
Thr Ala Asp Arg Asp Leu Tyr Gln Leu Val Ser Asp Arg Val Ala Val Leu His Pro Glu
160

540
GGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGG
Gly His Leu Ile Thr Pro Glu Trp Leu Trp Glu Lys Tyr Gly Leu Arg Pro Glu Gln Trp
180

600
GTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATC
Val Asp Phe Arg Ala Leu Val Gly Asp Pro Ser Asp Asn Leu Pro Gly Val Lys Gly Ile
200

660
GGGGAGAAGACCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAG
Gly Glu Lys Thr Ala Leu Lys Leu Leu Lys Glu Trp Gly Ser Leu Glu Asn Leu Leu Lys
220

720
AACCTGGACCGGGTAAAGCCAGAAAACGTCCGGGAGAAGATCAAGGCCCACCTGGAAGAC
Asn Leu Asp Arg Val Lys Pro Glu Asn Val Arg Glu Lys Ile Lys Ala His Leu Glu Asp
240

780
CTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCCCCCTGGAGGTGGACCTC
Leu Arg Leu Ser Leu Glu Leu Ser Arg Val Arg Thr Asp Leu Pro Leu Glu Val Asp Leu
260

840
GCCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTCCTGGAGAGGCTGGAGTTC
Ala Gln Gly Arg Glu Pro Asp Arg Glu Gly Leu Arg Ala Phe Leu Glu Arg Leu Glu Phe
280

900
GGCAGCCTCCTCCACGAGTTCGGCCTCCTGGAGGCGCCCGCCCCCCTGGAGGAGGCCCCC
Gly Ser Leu Leu His Glu Phe Gly Leu Leu Glu Ala Pro Ala Pro Leu Glu Glu Ala Pro
300

FIG. 2-2

960
TGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTCGTCCTCTCCCGCCCCGAGCCCATGTGG
Trp Pro Pro Pro Glu Gly Ala Phe Val Gly Phe Val Leu Ser Arg Pro Glu Pro Met Trp
320

1020
GCGGAGCTTAAAGCCCTGGCCGCCTGCAGGGACGGCCGGGTGCACCGGGCAGCAGACCCC
Ala Glu Leu Lys Ala Leu Ala Ala Cys Arg Asp Gly Arg Val His Arg Ala Ala Asp Pro
340

1080
TTGGCGGGGCTAAAGGACCTCAAGGAGGTCCGGGGCCTCCTCGCCAAGGACCTCGCCGTC
Leu Ala Gly Leu Lys Asp Leu Lys Glu Val Arg Gly Leu Leu Ala Lys Asp Leu Ala Val
360

1140
TTGGCCTCGAGGGAGGGGCTAGACCTCGTGCCCGGGGACGACCCCATGCTCCTCGCCTAC
Leu Ala Ser Arg Glu Gly Leu Asp Leu Val Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr
380

1200
CTCCTGGACCCCTCCAACACCACCCCCGAGGGGGTGGCGCGGCGCTACGGGGGGGAGTGG
Leu Leu Asp Pro Ser Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp
400

1260
ACGGAGGACGCCGCCCACCGGGCCCTCCTCTCGGAGAGGCTCCATCGGAACCTCCTTAAG
Thr Glu Asp Ala Ala His Arg Ala Leu Leu Ser Glu Arg Leu His Arg Asn Leu Leu Lys
420

1320
CGCCTCGAGGGGGAGGAGAAGCTCCTTTGGCTCTACCACGAGGTGGAAAAGCCCCTCTCC
Arg Leu Glu Gly Glu Glu Lys Leu Leu Trp Leu Tyr His Glu Val Glu Lys Pro Leu Ser
440

1380
CGGGTCCTGGCCCACATGGAGGCCACCGGGGTACGGCTGGACGTGGCCTACCTTCAGGCC
Arg Val Leu Ala His Met Glu Ala Thr Gly Val Arg Leu Asp Val Ala Tyr Leu Gln Ala
460

FIG. 2-3

1440
CTTTCCCTGGAGCTTGCGGAGGAGATCCGCCGCCTCGAGGAGGAGGTCTTCCGCTTGGCG
Leu Ser Leu Glu Leu Ala Glu Glu Ile Arg Arg Leu Glu Glu Glu Val Phe Arg Leu Ala
480

1500
GGCCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTGCTCTTTGACGAGCTT
Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp Glu Leu
500

1560
AGGCTTCCCGCCTTGGGGAAGACGCAAAAGACAGGCAAGCGCTCCACCAGCGCCGCGGTG
Arg Leu Pro Ala Leu Gly Lys Thr Gln Lys Thr Gly Lys Arg Ser Thr Ser Ala Ala Val
520

1620
CTGGAGGCCCTACGGGAGGCCCACCCCATCGTGGAGAAGATCCTCCAGCACCGGGAGCTC
Leu Glu Ala Leu Arg Glu Ala His Pro Ile Val Glu Lys Ile Leu Gln His Arg Glu Leu
540

1680
ACCAAGCTCAAGAACACCTACGTGGACCCCCTCCCAAGCCTCGTCCACCCGAGGACGGGC
Thr Lys Leu Lys Asn Thr Tyr Val Asp Pro Leu Pro Ser Leu Val His Pro Arg Thr Gly
560

1740
CGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGGAGGCTTAGTAGCTCCGAC
Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser Ser Asp
580

1800
CCCAACCTGCAGAACATCCCCGTCCGCACCCCCTTGGGCCAGAGGATCCGCCGGGCCTTC
Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln Arg Ile Arg Arg Ala Phe
600

1860
GTGGCCGAGGCGGGTTGGGCGTTGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGCGTC
Val Ala Glu Ala Gly Trp Ala Leu Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val
620

FIG. 2-4

1920
CTCGCCCACCTCTCCGGGGACGAAAACCTGATCAGGGTCTTCCAGGAGGGGAAGGACATC
Leu Ala His Leu Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Lys Asp Ile
640

1980
CACACCCAGACCGCAAGCTGGATGTTCGGCGTCCCCCCGGAGGCCGTGGACCCCCTGATG
His Thr Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val Asp Pro Leu Met
660

2040
CGCCGGGCGGCCAAGACGGTGAACTTCGGCGTCCTCTACGGCATGTCCGCCCATAGGCTC
Arg Arg Ala Ala Lys Thr Val Asn Phe Gly Val Leu Tyr Gly Met Ser Ala His Arg Leu
680

2100
TCCCAGGAGCTTGCCATCCCCTACGAGGAGGCGGTGGCCTTTATAGAGCGCTACTTCCAA
Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln
700

2160
AGCTTCCCCAAGGTGCGGGCCTGGATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGC
Ser Phe Pro Lys Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Lys Arg Gly
720

2220
TACGTGGAAACCCTCTTCGGAAGAAGGCGCTACGTGCCCGACCTCAACGCCCGGGTGAAG
Tyr Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala Arg Val Lys
740

2280
AGCGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro Val Gln Gly Thr Ala Ala
760

2340
GACCTCATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGC
Asp Leu Met Lys Leu Ala Met Val Lys Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg
780

FIG. 2-5

2400
ATGCTCCTCCAGGTCCACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAG
Met Leu Leu Gln Val His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu
800
2460
GTGGCGGCTTTGGCCAAGGAGGCCATGGAGAAGGCCTATCCCCTCGCCGTGCCCCTGGAG
Val Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala Val Pro Leu Glu
820

2520
GTGGAGGTGGGGATGGGGGAGGACTGGCTTTCCGCCAAGGGTTAGGGGGGCCCTGCCGTT
Val Glu Val Gly Met Gly Glu Asp Trp Leu Ser Ala Lys Gly
834

FIG. 2-6

```
  1 MLPLFEPKGR VLLVDGHHLA YRTFFALKGL TTSRGEPVQA
VYGFAKSLLK   50
 51 ALKEDGEVAI VVFDAKAPSF RHEAYEAYKA GRAPTPEDFP
RQLALIKELV  100
101 DLLGLVRLEV PGFEADDVLA TLAKKAEREG YEVRILSADR
DLYQLLSDRI  150
151 HLLHPEGEVL TPGWLQERYG LSPERWVEYR ALVGDPSDNL
PGVPGIGEKT  200
201 ALKLLKEWGS LEAILKNLDQ VKPERVREAI RNNLDKLQMS
LELSRLRTDL  250
251 PLEVDFAKRR EPDWEGLKAF LERLEFGSLL HEFGLLEAPK
EAEEAPWPPP  300
301 GGAFLGFLLS RPEPMWAELL ALAGAKEGRV HRAEDPVGAL
KDLKEIRGLL  350
351 AKDLSVLALR EGREIPPGDD PMLLAYLLDP GNTNPEGVAR
RYGGEWKEDA  400
401 AARALLSERL WQALYPRVAE EERLLWLYRE VERPLAQVLA
HMEATGVRLD  450
451 VPYLEALSQE VAFELERLEA EVHRLAGHPF NLNSRDQLER
VLFDELGLPP  500
501 IGKTEKTGKR STSAAVLELL REAHPIVGRI LEYRELMKLK
STYIDPLPRL  550
551 VHPKTGRLHT RFNQTATATG RLSSSDPNLQ NIPVRTPLGQ
RIRKAFIAEE  600
601 GHLLVALDYS QIELRVLAHL SGDENLIRVF REGKDIHTET
AAWMFGVPPE  650
651 GVDGAMRRAA KTVNFGVLYG MSAHRLSQEL SIPYEEAAAF
IERYFQSFPK  700
701 VRAWIAKTLE EGRKKGYVET LFGRRRYVPD LNARVKSVRE
AAERMAFNMP  750
751 VQGTAADLMK LAMVKLFPRL RPLGVRILLQ VHDELVLEAP
KARAEEAAQL  800
801 AKETMEGVYP LSVPLEVEVG MGEDWLSAKA
830
```

FIG. 3A

1 ATGCTGCCCC TCTTTGAGCC CAAGGGCCGG GTCCTCCTGG
TGGACGGCCA
51 CCACCTGGCC TACCGCACCT TTTTCGCCCT CAAGGGCCTC
ACCACCAGCC
101 GGGGCGAGCC CGTGCAGGCG GTTTATGGCT TCGCCAAAAG
CCTCCTCAAG
151 GCCCTGAAGG AGGATGGGGA GGTGGCCATC GTGGTCTTTG
ACGCCAAGGC
201 CCCCTCCTTC CGCCACGAGG CCTACGAGGC CTACAAGGCG
GGCCGGGCCC
251 CCACCCCGGA GGACTTTCCC CGGCAGCTCG CCCTCATCAA
GGAGCTGGTG
301 GACCTTTTGG GCCTCGTGCG CCTTGAGGTC CCGGGCTTTG
AGGCGGACGA
351 TGTCCTCGCC ACCCTGGCCA AGAAGGCAGA AAGGGAGGGG
TACGAGGTGC
401 GCATCCTGAG CGCGGACCGC GACCTCTACC AGCTCCTTTC
CGACCGGATC
451 CACCTCCTCC ACCCCGAGGG GGAGGTCCTG ACCCCCGGGT
GGCTCCAGGA
501 GCGCTACGGC CTCTCCCCGG AGAGGTGGGT GGAGTACCGG
GCCCTGGTGG
551 GGGACCCTTC GGACAACCTC CCCGGGGTGC CCGGCATCGG
GGAGAAGACC
601 GCCCTGAAGC TCCTGAAGGA GTGGGGTAGC CTGGAAGCGA
TTCTAAAGAA
651 CCTGGACCAG GTGAAGCCGG AAAGGGTGCG GGAGGCCATC
CGGAATAACC
701 TGGATAAGCT CCAGATGTCC CTGGAGCTTT CCCGCCTCCG
CACCGACCTC
751 CCCCTGGAGG TGGACTTCGC CAAGAGGCGG GAGCCCGACT
GGGAGGGGCT
801 TAAGGCCTTT TTGGAGCGGC TTGAGTTCGG AAGCCTCCTC
CACGAGTTCG
851 GCCTTCTGGA GGCCCCCAAG GAGGCGGAGG AGGCCCCCTG
GCCCCCGCCT
901 GGAGGGGCCT TTTTGGGCTT CCTCCTCTCC CGCCCCGAGC
CCATGTGGGC
951 GGAGCTTTTG GCCCTGGCGG GGGCCAAGGA GGGGCGGGTC
CATCGGGCGG

FIG. 3B-1

1001 AAGACCCCGT GGGGGCCCTA AAGGACCTGA AGGAGATCCG
GGGCCTCCTC
1051 GCCAAGGACC TCTCGGTCCT GGCCCTGAGG GAGGGCCGGG
AGATCCCGCC
1101 GGGGGACGAC CCCATGCTCC TCGCCTACCT CCTGGACCCG
GGGAACACCA
1151 ACCCCGAGGG GGTGGCCCGG CGGTACGGGG GGGAGTGGAA
GGAGGACGCC
1201 GCCGCCCGGG CCCTCCTTTC GGAAAGGCTC TGGCAGGCCC
TTTACCCCCG
1251 GGTGGCGGAG GAGGAAAGGC TCCTTTGGCT CTACCGGGAG
GTGGAGCGGC
1301 CCCTCGCCCA GGTCCTCGCC CACATGGAGG CCACGGGGGT
GCGGCTGGAT
1351 GTGCCCTACC TGGAGGCCCT TTCCCAGGAG GTGGCCTTTG
AGCTGGAGCG
1401 CCTCGAGGCC GAGGTCCACC GCCTGGCGGG CCACCCCTTC
AACCTGAACT
1451 CTAGGGACCA GCTGGAGCGG GTCCTCTTTG ACGAGCTCGG
CCTACCCCCC
1501 ATCGGCAAGA CGGAGAAGAC GGGCAAGCGC TCCACCAGCG
CCGCCGTCCT
1551 GGAGCTCTTA AGGGAGGCCC ACCCCATCGT GGGGCGGATC
CTGGAGTACC
1601 GGGAGCTCAT GAAGCTCAAG AGCACCTACA TAGACCCCCT
CCCCAGGCTG
1651 GTCCACCCCA AAACCGGCCG GCTCCACACC CGCTTCAACC
AGACGGCCAC
1701 CGCCACGGGC CGCCTCTCCA GCTCCGACCC CAACCTGCAG
AACATCCCCG
1751 TGCGCACCCC CTTAGGCCAG CGCATCCGCA AGGCCTTCAT
TGCCGAGGAG
1801 GGCCATCTCC TGGTGGCCCT GGACTATAGC CAGATCGAGC
TCCGGGTCCT
1851 CGCCCACCTC TCGGGGGACG AGAACCTCAT CCGGGTCTTC
CGGGAAGGGA
1901 AGGACATCCA CACCGAGACC GCCGCCTGGA TGTTCGGCGT
GCCCCCCGAG
1951 GGGGTGGACG GGGCCATGCG CCGGGCGGCC AAGACGGTGA
ACTTCGGGGT

FIG. 3B-2

2001 GCTCTACGGG ATGTCCGCCC ACCGCCTCTC CCAGGAGCTC
TCCATCCCCT
2051 ACGAGGAGGC GGCGGCCTTC ATCGAGCGCT ACTTCCAGAG
CTTCCCCAAG
2101 GTGCGGGCCT GGATCGCCAA AACCTTGGAG GAGGGGCGGA
AGAAGGGGTA
2151 CGTGGAGACC CTCTTCGGCC GCCGCCGCTA CGTGCCCGAC
CTCAACGCCC
2201 GGGTGAAGAG CGTGCGGGAG GCGGCGGAGC GCATGGCCTT
CAACATGCCC
2251 GTGCAGGGCA CCGCCGCGGA CCTCATGAAG CTGGCCATGG
TGAAGCTCTT
2301 CCCCAGGCTC AGGCCCTTGG GCGTTCGCAT CCTCCTCCAG
GTGCACGACG
2351 AGCTGGTCTT GGAGGCCCCA AAGGCGCGGG CGGAGGAGGC
CGCCCAGTTG
2401 GCCAAGGAGA CCATGGAAGG GGTTTACCCC CTCTCCGTCC
CCCTGGAGGT
2451 GGAGGTGGGG ATGGGGGAGG ACTGGCTTTC CGCCAAGGCC TAG

FIG. 3B-3

```
1    MRAMLPLFEP KGRVLLVDGH HLAYRTFFAL KGLTTSRGEP VQAVYGFAKS   50
51   LLKALREDGD VVIVVFDAKA PSFRHQTYEA YKAGRAPTPE DFPRQLALIK  100
101  EMVDLLGLER LEVPGFEADD VLATLAKKAE KEGYEVRILT ADRDLYQLLS  150
151  ERISILHPEG YLITPEWLWE KYGLKPSQWV DYRALAGDPS DNIPGVKGIG  200
201  EKTAAKLIRE WGSLENLLKH LEQVKPASVR EKILSHMEDL KLSLELSRVR  250
251  TDLPLQVDFA RRREPDREGL KAFLERLEFG SLLHEFGLLE SPVAAEEAPW  300
301  PPPEGAFVGY VLSRPEPMWA ELNALAAAWE GRVYRAEDPL EALRGLGEVR  350
351  GLLAKDLAVL ALREGIALAP GDDPMLLAYL LDPSNTAPEG VARRYGGEWT  400
401  EEAGERALLS ERLYAALLER LKGEERLLWL YEEVEKPLSR VLAHMEATGV  450
451  RLDVAYLKAL SLEVEAELRR LEEEVHRLAG HPFNLNSRDQ LERVLFDELG  500
501  LPAIGKTEKT GKRSTSAAVL EALREAHPIV DRILQYRELS KLKGTYIDPL  550
551  PALVHPKTNR LHTRFNQTAT ATGRLSSSDP NLQNIPVRTP LGQRIRRAFV  600
601  AEEGWRLVVL DYSQIELRVL AHLSGDENLI RVFQEGQDIH TQTASWMFGV  650
651  PPEAVDSLMR RAAKTINFGV LYGMSAHRLS GELAIPYEEA VAFIERYFQS  700
701  YPKVRAWIEK TLAEGRERGY VETLFGRRRY VPDLASRVKS IREAAERMAF  750
751  NMPVQGTAAD LMKLAMVKLF PRLQELGARM LLQVHDELVL EAPKEQAEEV  800
801  AQEAKRTMEE VWPLKVPLEV EVGIGEDWLS AKA                   833
```

FIG. 4A

```
1      MLPLFEP KGRVLLVDGH HLAYRTFFAL KGLTTSRGEP VQAVYGFAKS   50
51   LLKALREDGD VVIVVFDAKA PSFRHQTYEA YKAGRAPTPE DFPRQLALIK  100
101  EMVDLLGLER LEVPGFEADD VLATLAKKAE KEGYEVRILT ADRDLYQLLS  150
151  ERISILHPEG YLITPEWLWE KYGLKPSQWV DYRALAGDPS DNIPGVKGIG  200
201  EKTAAKLIRE WGSLENLLKH LEQVKPASVR EKILSHMEDL KLSLELSRVR  250
251  TDLPLQVDFA RRREPDREGL KAFLERLEFG SLLHEFGLLE SPVAAEEAPW  300
301  PPPEGAFVGY VLSRPEPMWA ELNALAAAWE GRVYRAEDPL EALRGLGEVR  350
351  GLLAKDLAVL ALREGIALAP GDDPMLLAYL LDPSNTAPEG VARRYGGEWT  400
401  EEAGERALLS ERLYAALLER LKGEERLLWL YEEVEKPLSR VLAHMEATGV  450
451  RLDVAYLKAL SLEVEAELRR LEEEVHRLAG HPFNLNSRDQ LERVLFDELG  500
501  LPAIGKTEKT GKRSTSAAVL EALREAHPIV DRILQYRELS KLKGTYIDPL  550
551  PALVHPKTNR LHTRFNQTAT ATGRLSSSDP NLQNIPVRTP LGQRIRRAFV  600
601  AEEGWRLVVL DYSQIELRVL AHLSGDENLI RVFQEGQDIH TQTASWMFGV  650
651  PPEAVDSLMR RAAKTINFGV LYGMSAHRLS GELAIPYEEA VAFIERYFQS  700
701  YPKVRAWIEK TLAEGRERGY VETLFGRRRY VPDLASRVKS IREAAERMAF  750
751  NMPVQGTAAD LMKLAMVKLF PRLQELGARM LLQVHDELVL EAPKEQAEEV  800
801  AQEAKRTMEE VWPLKVPLEV EVGIGEDWLS AKA                   833
```

FIG. 4B

1 50
MRAMLPLFEP KGRVLLVDGH HLAYRTFFAL KGLTTSRGEP VQAVYGFAKS

51 100
LLKALREDGD VVIVVFDAKA PSFRHQTYEA YKAGRAPTPE DFPRQLALIK

101 150
EMVDLLGLER LEVPGFEADD VLATLAKKAE KEGYEVRILT ADRDLYQLLS

151 200
DRISILHPEG YLITPEWLWE KYGLKPSQWV DYRALAGDPS DNIPGVKGIG

201 250
EKTAAKLIRE WGSLENLLKH LEQVKPASVR EKILSHMEDL KLSLELSRVH

251 300
TELPLQVDFA RRREPDREGL KAFLERLEFG SLLHEFGLLE SPVAAEEAPW

301 350
PPPEGAFVGY VLSRPEPMWA ELNALAAAWE GRVYRAEDPL EALRGLGEVR

351 400
GLLAKDLAVL ALREGIALAQ GDDPMLLAYL LDPSNTAPEG VARRYGGEWT

401 450
EEAGERALLS ERLYAALLER LKGEERLLWL YEEVEKPLSR VLAHMEATGV

451 500
WLDVAYLKAL SLEVEAELRR LEEEVHRLAG HPFNLNSRDQ LERVLFDELG

501 550
LPAIGKTEKT GKRSTSAAVL EALREAHPIV DRILQYRELS KLKGTYIDPL

551 600
PALVHPKTNR LHTRFNQTAT ATGRLSSSDP NLQNIPVRTP LGQRIRRAFV

601 650
AEEGWRLVVL DYSQIELRVL AHLSGDENLI RVFQEGQDIH TQTASWMFGV

FIG. 4C-1

651 700
PPEAVDSLMR RAAKTINFGV LYGMSAHRLS GELAIPYEEA VAFIERYFQS

701 750
YPKVRAWIEK TLAEGRERGY VETLFGRRRY VPDLASRVKS IREAAERMAF

751 800
NMPVQGTAAD LMKLAMVKLF PRLQELGARM LLQVHDELVL EAPKEQAEEV

801 833
AQEAKRTMEE VWPLKVPLEV EVGIGEDWLS AKA

FIG. 4C-2

1 50
MRAMLPLFEP KGRVLLVDGH HLAYRTFFAL KGLTTSRGEP VQAVYGFAKS

51 100
LLKALREDGD VVIVVFDAKA PSFRHQTYEA YKAGRAPTPE DFPRQLALIK

101 150
EMVDLLGLER LEVPGFEADD VLATLAKKAE KEGYEVRILT ADRDLYQLLS

151 200
DRISILHPEG YLITPEWLWE KYGLKPSQWV DYRALAGDPS DNIPGVKGIG

201 250
EKTAAKLIRE WGSLENLLKH LEQVKPASVR EKILSHMEDL KLSLELSRVH

251 300
TELPLQVDFA RRREPDREGL KAFLERLEFG SLLHEFGLLE SPVAAEEAPW

301 350
PPPEGAFVGY VLSRPEPMWA ELNALAAAWE GRVYRAEDPL EALRGLGEVR

351 400
GLLAKDLAVL ALREGIALAP GDDPMLLAYL LDPSNTAPEG VARRYGGEWT

401 450
EEAGERALLS ERLYAALLER LKGEERLLWL YEEVEKPLSR VLAHMEATGV

451 500
WLDVAYLKAL SLEVEAELRR LEEEVHRLAG HPFNLNSRDQ LERVLFDELG

501 550
LPAIGKTEKT GKRSTSAAVL EALREAHPIV DRILQYRELS KLKGTYIDPL

551 600
PALVHPKTNR LHTRFNQTAT ATGRLSSSDP NLQNIPVRTP LGQRIRRAFV

601 650
AEEGWRLVVL DYSQIELRVL AHLSGDENLI RVFQEGQDIH TQTASWMFGV

FIG. 4D-1

651 700
PPEAVDSLMR RAAKTINFGV LYGMSAHRLS GELAIPYEEA VAFIERYFQS

701 750
YPKVRAWIEK TLAEGRERGY VETLFGRRRY VPDLASRVKS IREAAERMAF

751 800
NMPVQGTAAD LMKLAMVKLF PRLQELGARM LLQVHDELVL EAPKEQAEEV

801 833
AQEAKRTMEE VWPLKVPLEV EVGIGEDWLS AKA

FIG. 4D-2

```
ATGAGGGCGA TGCTGCCCCT CTTTGAGCCC AAGGGCCGGG TGCTTCTGGT 50
GGACGGCCAC CACCTGGCCT ACCGTACCTT TTTTGCCCTG AAGGGCCTCA 100
CCACCAGCCG CGGGGAGCCG GTCCAGGCGG TGTACGGGTT TGCCAAGAGC 150
CTTTTGAAGG CGCTAAGGGA AGACGGGGAT GTGGTGATCG TGGTGTTTGA 200
CGCCAAGGCC CCCTCCTTCC GCCACCAGAC CTACGAGGCC TACAAGGCGG 250
GGCGGGCTCC CACCCCCGAG GACTTTCCCC GGCAGCTTGC CCTTATCAAG 300
GAGATGGTGG ACCTTTTGGG CCTGGAGCGC CTCGAGGTGC CGGGCTTTGA 350
GGCGGATGAC GTCCTGGCTA CCCTGGCCAA GAAGGCGGAA AAGGAAGGCT 400
ACGAGGTGCG CATCCTCACC GCGGACCGGG ACCTTTACCA GCTTCTTTCG 450
GAGCGAATCT CCATCCTTCA CCCGGAGGGT TACCTGATCA CCCCGGAGTG 500
GCTTTGGGAG AAGTATGGGC TTAAGCCTTC CCAGTGGGTG GACTACCGGG 550
CCTTGGCCGG GGACCCTTCC GACAACATCC CCGGCGTGAA GGGCATCGGG 600
GAGAAGACGG CGGCCAAGCT GATCCGGGAG TGGGGAAGCC TGGAAAACCT 650
TCTTAAGCAC CTGGAACAGG TGAAACCTGC CTCCGTGCGG GAGAAGATCC 700
TTAGCCACAT GGAGGACCTC AAGCTATCCC TGGAGCTATC CCGGGTGCGC 750
ACGGACTTGC CCCTTCAGGT GGACTTCGCC CGGCGCCGGG AGCCGGACCG 800
GGAGGGGCTT AAGGCCTTTT TGGAGAGGCT GGAGTTCGGA AGCCTCCTCC 850
ACGAGTTCGG CCTGTTGGAA AGCCCGGTGG CGGCGGAGGA AGCTCCCTGG 900
CCGCCCCCCG AGGGAGCCTT CGTGGGGTAC GTTCTTTCCC GCCCCGAGCC 950
CATGTGGGCG GAGCTTAACG CCTTGGCCGC CGCCTGGGAG GGAAGGGTTT 1000
ACCGGGCGGA GGATCCCTTG GAGGCCTTGC GGGGGCTTGG GGAGGTGAGG 1050
GGGCTTTTGG CCAAGGACCT GGCGGTGCTG GCCCTGAGGG AAGGGATTGC 1100
CCTGGCACCG GGCGACGACC CCATGCTCCT CGCCTACCTC CTGGATCCTT 1150
CCAACACCGC CCCCGAAGGG GTAGCCCGGC GCTACGGGGG GGAGTGGACC 1200
GAGGAGGCGG GGGAAAGGGC GTTGCTTTCC GAAAGGCTTT ACGCCGCCCT 1250
CCTGGAGCGG CTTAAGGGGG AGGAGAGGCT TCTTTGGCTT TACGAGGAGG 1300
TGGAAAAGCC CCTTTCGCGG GTCCTGGCCC ACATGGAGGC CACGGGGGTA 1350
CGGTTGGATG TGGCCTACTT AAAGGCCCTT TCCCTGGAGG TGGAGGCGGA 1400
GCTCAGGCGC CTCGAGGAGG AGGTCCACCG CCTGGCCGGG CATCCTTTCA 1450
ACCTGAACTC CCGGGACCAG CTGGAAAGGG TCCTCTTTGA CGAGCTTGGG 1500
CTTCCCGCCA TCGGCAAGAC GGAGAAGACG GGCAAGCGCT CCACCAGCGC 1550
CGCCGTTTTG GAGGCCTTGC GGGAGGCTCA TCCCATCGTG GACCGCATCC 1600
TTCAGTACCG GGAGCTTTCC AAGCTCAAGG GAACCTACAT CGATCCCTTG 1650
CCTGCCCTGG TCCACCCCAA GACGAACCGC CTCCACACCC GTTTCAACCA 1700
GACGGCCACC GCCACGGGGA GGCTTAGCAG CTCGGATCCC AACCTGCAAA 1750
ATATCCCCGT GCGCACCCCT TTGGGCCAGC GGATCCGCCG GGCCTTCGTG 1800
GCCGAGGAGG GGTGGAGGCT GGTGGTTTTG GACTACAGCC AGATTGAGCT 1850
CAGGGTCCTG GCGCACCTTT CCGGGGACGA GAACCTAATC CGGGTCTTCC 1900
AGGAGGGCCA GGACATCCAC ACCCAGACGG CCAGCTGGAT GTTCGGCGTG 1950
```

FIG. 4E-1

```
CCCCCAGAGG CCGTGGATTC CCTGATGCGT CGGGCGGCCA AGACCATCAA 2000
CTTCGGCGTC CTCTACGGCA TGTCCGCCCA CCGGCTTTCG GGAGAGCTGG 2050
CCATCCCCTA CGAGGAGGCG GTGGCCTTCA TCGAGCGGTA TTTCCAGAGC 2100
TACCCCAAGG TGCGGGCCTG GATTGAGAAA ACCCTGGCGG AAGGACGGGA 2150
ACGGGGCTAT GTGGAAACCC TCTTTGGCCG CCGGCGCTAC GTGCCCGACT 2200
TGGCTTCCCG GGTGAAGAGC ATCCGGGAGG CAGCGGAGCG CATGGCCTTC 2250
AACATGCCGG TCCAGGGGAC CGCCGCGGAT TTGATGAAAC TGGCCATGGT 2300
GAAGCTCTTT CCCAGGCTTC AGGAGCTGGG GGCCAGGATG CTTTTGCAGG 2350
TGCACGACGA ACTGGTCCTC GAGGCTCCCA AGGAGCAAGC GGAGGAAGTC 2400
GCCCAGGAGG CCAAGCGGAC CATGGAGGAG GTGTGGCCCC TGAAGGTGCC 2450
CTTGGAGGTG GAAGTGGGCA TCGGGGAGGA CTGGCTTTCC GCCAAGGCCT 2500
AG                                                    2502
```

FIG. 4E-2

```
ATGAGGGCGA TGCTGCCCCT CTTTGAGCCC AAGGGCCGGG TGCTTCTGGT   50
GGACGGCCAC CACCTGGCCT ACCGTACCTT TTTTGCCCTG AAGGGCCTCA  100
CCACCAGCCG CGGGGAGCCG GTCCAGGCGG TGTACGGGTT TGCCAAGAGC  150
CTTTTGAAGG CGCTAAGGGA AGACGGGGAT GTGGTGATCG TGGTGTTTGA  200
CGCCAAGGCC CCCTCCTTCC GCCACCAGAC CTACGAGGCC TACAAGGCGG  250
GGCGGGCTCC CACCCCCGAG GACTTTCCCC GGCAGCTTGC CCTTATCAAG  300
GAGATGGTGG ACCTTTTGGG CCTGGAGCGC CTCGAAGTGC CGGGTTTTGA  350
GGCGGATGAC GTCCTGGCCA CCCTGGCCAA GAAGGCGGAA AAGGAAGGCT  400
ACGAGGTGCG CATCCTCACC GCGGACCGGG ACCTTTACCA GCTTCTTTCG  450
GACCGAATCT CCATCCTTCA CCCGGAGGGT TACCTGATCA CCCCGGAGTG  500
GCTTTGGGAG AAGTATGGGC TTAAGCCTTC CCAGTGGGTG GACTACCGGG  550
CCTTGGCCGG GGACCCTTCC GACAACATCC CCGGCGTGAA GGGCATCGGG  600
GAGAAGACGG CGGCCAAGCT GATCCGGGAG TGGGGAAGCC TGGAAAACCT  650
TCTTAAGCAC CTGGAACAGG TGAAACCTGC CTCCGTGCGG GAGAAGATCC  700
TTAGCCACAT GGAGGACCTC AAGCTATCCC TGGAGCTTTC CCGGGTGCAC  750
ACGGAGTTGC CCCTTCAGGT GGACTTCGCC CGGCGCCGGG AGCCGGACCG  800
GGAAGGGCTT AAGGCCTTTT TGGAGAGGCT GGAGTTCGGA AGCCTCCTCC  850
ACGAGTTCGG CCTGTTGGAA AGCCCGGTGG CGGCGGAGGA AGCTCCCTGG  900
CCGCCCCCCG AGGGAGCCTT CGTGGGGTAC GTTCTTTCCC GCCCCGAGCC  950
CATGTGGGCG GAGCTTAACG CCTTGGCCGC CGCCTGGGAG GGAAGGGTTT 1000
ACCGGGCGGA GGATCCCTTG GAGGCCTTGC GGGGGCTTGG GGAGGTGAGG 1050
GGGCTTTTGG CCAAGGACCT GGCGGTGCTG GCCCTGAGGG AAGGGATTGC 1100
CCTGGCACAG GGCGACGACC CCATGCTCCT CGCCTACCTC CTGGATCCTT 1150
CCAACACCGC CCCCGAAGGG GTAGCCCGGC GCTACGGGGG GGAGTGGACC 1200
GAGGAGGCGG GGGAAAGGGC GCTGCTTTCC GAAAGGCTTT ACGCCGCCCT 1250
CCTGGAGCGG CTTAAGGGGG AGGAGAGGCT TCTTTGGCTT TACGAGGAGG 1300
TGGAAAAGCC CCTTTCGCGG GTCCTGGCCC ACATGGAGGC CACGGGGGTA 1350
TGGTTGGATG TGGCCTACTT GAAGGCCCTT TCCCTGGAGG TGGAGGCGGA 1400
GCTCAGGCGC CTCGAGGAGG AGGTCCACCG ACTGGCCGGG CATCCTTTCA 1450
ACCTGAACTC CCGGGACCAG CTGGAAAGGG TCCTCTTTGA CGAGCTTGGG 1500
CTTCCCGCCA TCGGCAAGAC GGAGAAGACG GGTAAGCGTT CCACCAGCGC 1550
CGCCGTTTTG GAGGCTTTGA GGGAGGCTCA TCCCATAGTG GACCGCATCC 1600
TCCAGTACCG GGAGCTTTCC AAGCTCAAGG GAACGTACAT CGATCCCTTG 1650
CCCGCCCTGG TCCACCCCAA GACGAACCGC CTCCACACCC GTTTCAACCA 1700
GACGGCCACC GCCACGGGGA GGCTTAGCAG CTCGGATCCC AACCTGCAAA 1750
ATATCCCCGT GCGCACCCCT TTAGGCCAGC GGATCCGCCG GGCCTTCGTG 1800
GCCGAGGAGG GGTGGAGGCT GGTGGTTTTG GACTACAGCC AGATTGAGCT 1850
CAGGGTCCTG GCGCACCTTT CCGGGGACGA GAACCTGATC CGGGTCTTCC 1900
AAGAGGGCCA GGACATCCAC ACCCAGACGG CCAGCTGGAT GTTCGGCGTG 1950
CCCCCAGAGG CCGTGGATTC CCTGATGCGC CGGGCGGCCA AGACCATCAA 2000
```

FIG. 4F-1

CTTCGGCGTC CTCTACGGCA TGTCCGCCCA CCGGCTTTCG GGAGAGCTGG 2050
CCATCCCCTA CGAGGAAGCG GTGGCCTTCA TCGAGCGGTA TTTCCAGAGC 2100
TACCCCAAGG TACGGGCCTG GATTGAGAAA ACCCTGGCGG AAGGACGGGA 2150
GCGGGGCTAT GTGGAAACCC TCTTTGGCCG CCGGCGCTAT GTGCCCGACT 2200
TGGCTTCCCG GGTGAAGAGC ATCCGGGAGG CAGCGGAGCG CATGGCCTTC 2250
AACATGCCGG TCCAGGGGAC CGCCGCGGAT TTGATGAAAC TGGCCATGGT 2300
GAAGCTCTTT CCCAGGCTTC AGGAGCTGGG GGCCAGGATG CTTTTGCAGG 2350
TGCACGACGA ACTGGTCCTC GAGGCTCCCA AGGAGCAAGC GGAGGAAGTC 2400
GCCCAGGAGG CCAAGCGGAC CATGGAGGAG GTGTGGCCCC TGAAGGTGCC 2450
CTTGGAGGTG GAGGTGGGTA TCGGGGAGGA CTGGCTTTCC GCCAAGGCCT 2500
AGTCGAC 2507

```
ATGAGGGCGA TGCTGCCCCT CTTTGAGCCC AAGGGCCGGG TGCTTCTGGT   50
GGACGGCCAC CACCTGGCCT ACCGTACCTT TTTTGCCCTG AAGGGCCTCA  100
CCACCAGCCG CGGGGAGCCG GTCCAGGCGG TGTACGGGTT TGCCAAGAGC  150
CTTTTGAAGG CGCTAAGGGA AGACGGGGAT GTGGTGATCG TGGTGTTTGA  200
CGCCAAGGCC CCCTCCTTCC GCCACCAGAC CTACGAGGCC TACAAGGCGG  250
GGCGGGCTCC CACCCCCGAG GACTTTCCCC GGCAGCTTGC CCTTATCAAG  300
GAGATGGTGG ACCTTTTGGG CCTGGAGCGC CTCGAAGTGC CGGGTTTTGA  350
GGCGGATGAC GTCCTGGCCA CCCTGGCCAA GAAGGCGGAA AAGGAAGGCT  400
ACGAGGTGCG CATCCTCACC GCGGACCGGG ACCTTTACCA GCTTCTTTCG  450
GACCGAATCT CCATCCTTCA CCCGGAGGGT TACCTGATTA CCCCGGAGTG  500
GCTTTGGGAG AAGTATGGGC TTAAGCCTTC CCAGTGGGTG GACTACCGGG  550
CCTTGGCCGG GGACCCTTCC GACAACATCC CCGGCGTGAA GGGCATCGGG  600
GAGAAGACGG CGGCCAAGCT GATCCGGGAG TGGGGAAGCC TGGAAAACCT  650
TCTTAAGCAC CTGGAACAGG TGAAACCTGC CTCCGTGCGG GAGAAGATCC  700
TTAGCCACAT GGAGGACCTC AAGCTATCCC TGGAGCTTTC CCGGGTGCAC  750
ACGGAGTTGC CCCTTCAGGT GGACTTCGCC CGGCGCCGGG AGCCGGACCG  800
GGAAGGGCTT AAGGCCTTTT TGGAGAGGCT GGAGTTCGGA AGCCTCCTCC  850
ACGAGTTCGG CCTGTTGGAA AGCCCGGTGG CGGCGGAGGA AGCTCCCTGG  900
CCGCCCCCCG AGGGAGCCTT CGTGGGGTAC GTTCTTTCCC GCCCCGAGCC  950
CATGTGGGCG GAGCTTAACG CCTTGGCCGC CGCCTGGGAG GGAAGGGTTT 1000
ACCGGGCGGA GGATCCCTTG GAGGCCTTGC GGGGGCTTGG GGAGGTGAGG 1050
GGGCTTTTGG CCAAGGACCT GGCGGTGCTG GCCCTGAGGG AAGGGATTGC 1100
CCTGGCACCG GGCGACGACC CCATGCTCCT CGCCTACCTC CTGGATCCTT 1150
CCAACACCGC CCCCGAAGGG GTAGCCCGGC GCTACGGGGG GGAGTGGACC 1200
GAGGAGGCGG GGGAAAGGGC GCTGCTTTCC GAAAGGCTTT ACGCCGCCCT 1250
CCTGGAGCGG CTTAAGGGGG AGGAGAGGCT TCTTTGGCTT TACGAGGAGG 1300
TGGAAAAGCC CCTTTCGCGG GTCCTGGCCC ACATGGAGGC CACGGGGGTA 1350
TGGTTGGATG TGGCCTACTT GAAGGCCCTT TCCCTGGAGG TGGAGGCGGA 1400
GCTCAGGCGC CTCGAGGAGG AGGTCCACCG ACTGGCCGGG CATCCTTTCA 1450
ACCTGAACTC CCGGGACCAG CTGGAAAGGG TCCTCTTTGA CGAGCTTGGG 1500
CTTCCCGCCA TCGGCAAGAC GGAGAAGACG GGTAAGCGTT CCACCAGCGC 1550
CGCCGTTTTG GAGGCTTTGA GGGAGGCTCA TCCCATAGTG GACCGCATCC 1600
TCCAGTACCG GGAGCTTTCC AAGCTCAAGG GAACGTACAT CGATCCCTTG 1650
CCCGCCCTGG TCCACCCCAA GACGAACCGC CTCCACACCC GTTTCAACCA 1700
GACGGCCACC GCCACGGGGA GGCTTAGCAG CTCGGATCCC AACCTGCAAA 1750
ATATCCCCGT GCGCACCCCT TTAGGCCAGC GGATCCGCCG GGCCTTCGTG 1800
GCCGAGGAGG GGTGGAGGCT GGTGGTTTTG GACTACAGCC AGATTGAGCT 1850
CAGGGTCCTG GCGCACCTTT CCGGGGACGA GAACCTGATC CGGGTCTTCC 1900
AAGAGGGCCA GGACATCCAC ACCCAGACGG CCAGCTGGAT GTTCGGCGTG 1950
CCCCCAGAGG CCGTGGATTC CCTGATGCGC CGGGCGGCCA AGACCATCAA 2000
```

FIG. 4G-1

```
CTACGGCGTC CTCTACGGCA TGTCCGCCCA CCGGCTTTCG GGAGAGCTGG 2050
CCATCCCCTA CGAGGAAGCG GTGGCCTTCA TCGAGCGGTA TTTCCAGAGC 2100
TTCCCCAAGG TACGGGCCTG GATTGAGAAA ACCCTGGCGG AAGGACGGGA 2150
GCGGGGCTAT GTGGAAACCC TCTTTGGCCG CCGGCGCTAT GTGCCCGACT 2200
TGGCTTCCCG GGTGAAGAGC ATCCGGGAGG CAGCGGAGCG CATGGCCTTC 2250
AACATGCCGG TCCAGGGGAC CGCCGCGGAT TTGATGAAAC TGGCCATGGT 2300
GAAGCTCTTT CCCAGGCTTC AGGAGCTGGG GGCCAGGATG CTTTTGCAGG 2350
TGCACGACGA ACTGGTCCTC GAGGCTCCCA AGGAGCAAGC GGAGGAAGTC 2400
GCCCAGGAGG CCAAGCGGAC CATGGAGGAG GTGTGGCCCC TGAAGGTGCC 2450
CTTGGAGGTG GAGGTGGGTA TCGGGGAGGA CTGGCTTTCC GCCAAGGCCT 2500
AGTCGAC                                               2507
```

FIG. 4G-2

```
1    MEAMLPLFEP KGRVLLVDGH HLAYRTFFAL KGLTTSRGEP VQAVYGFAKS  50
51   LLKALKEDGY KAVFVVFDAK APSFRHEAYE AYKAGRAPTP EDFPRQLALI 100
101  KELVDLLGFT RLEVPGYEAD DVLATLAKKA EKEGYEVRIL TADRDLYQLV 150
151  SDRVAVLHPE GHLITPEWLW EKYGLKPEQW VDFRALVGDP SDNLPGVKGI 200
201  GEKTALKLLK EWGSLENLLK NLDRVKPENV REKIKAHLED LRLSLELSRV 250
251  RTDLPLEVDL AQGREPDREG LRAFLERLEF GSLLHEFGLL EAPAPLEEAP 300
301  WPPPEGAFVG FVLSRPEPMW AELKALAACR DGRVHRAADP LAGLKDLKEV 351
351  RGLLAKDLAV LASREGLDLV PGDDPMLLAY LLDPSNTTPE GVARRYGGEW 400
401  TEDAAHRALL SERLHRNLLK RLEGEEKLLW LYHEVEKPLS RVLAHMEATG 450
451  VRLDVAYLQA LSLELAEEIR RLEEEVFRLA GHPFNLNSRD QLERVLFDEL 500
501  RLPALGKTQK TGKRSTSAAV LEALREAHPI VEKILQHREL TKLKNTYVDP 550
551  LPSLVHPRTG RLHTRFNQTA TATGRLSSSD PNLQNIPVRT PLGQRIRRAF 600
601  VAEAGWALVA LDYSQIELRV LAHLSGDENL IRVFQEGKDI HTQTASWMFG 650
651  VPPEAVDPLM RRAAKTVNFG VLYGMSAHRL SQELAIPYEE AVAFIERYFQ 700
701  SFPKVRAWIE KTLEEGRKRG YVETLFGRRR YVPDLNARVK SVREAAERMA 750
751  FNMPVQGTAA DLMKLAMVKL FPRLREMGAR MLLQVHDELL LEAPQARAEE 800
801  VAALAKEAME KAYPLAVPLE VEVGMGEDWL SAKG                  834
```

FIG. 5A

```
1    ATGGAGGCGA TGCTTCCGCT CTTTGAACCC AAAGGCCGGG TCCTCCTGGT
51   GGACGGCCAC CACCTGGCCT ACCGCACCTT CTTCGCCCTG AAGGGCCTCA
101  CCACGAGCCG GGGCGAACCG GTGCAGGCGG TCTACGGCTT CGCCAAGAGC
151  CTCCTCAAGG CCCTGAAGGA GGACGGGTAC AAGGCCGTCT TCGTGGTCTT
201  TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGAG GCCTACAAGG
251  CGGGGAGGGC CCCGACCCCC GAGGACTTCC CCCGGCAGCT CGCCCTCATC
301  AAGGAGCTGG TGGACCTCCT GGGGTTTACC CGCCTCGAGG TCCCCGGCTA
351  CGAGGCGGAC GACGTCCTCG CCACCCTGGC CAAGAAGGCG GAAAAGGAGG
401  GGTACGAGGT GCGCATCCTC ACCGCCGACC GCGACCTCTA CCAACTCGTC
451  TCCGACCGCG TCGCCGTCCT CCACCCCGAG GGCCACCTCA TCACCCCGGA
501  GTGGCTTTGG GAGAAGTACG GCCTCAAGCC GGAGCAGTGG GTGGACTTCC
551  GCGCCCTCGT GGGGGACCCC TCCGACAACC TCCCCGGGGT CAAGGGCATC
601  GGGGAGAAGA CCGCCCTCAA GCTCCTCAAG GAGTGGGGAA GCCTGGAAAA
651  CCTCCTCAAG AACCTGGACC GGGTAAAGCC AGAAAACGTC CGGGAGAAGA
701  TCAAGGCCCA CCTGGAAGAC CTCAGGCTTT CCTTGGAGCT CTCCCGGGTG
751  CGCACCGACC TCCCCCTGGA GGTGGACCTC GCCCAGGGGC GGGAGCCCGA
801  CCGGGAGGGG CTTAGGGCCT TCCTGGAGAG GCTGGAGTTC GGCAGCCTCC
851  TCCACGAGTT CGGCCTCCTG GAGGCCCCCG CCCCCCTGGA GGAGGCCCCC
901  TGGCCCCCGC CGGAAGGGGC CTTCGTGGGC TTCGTCCTCT CCCGCCCCGA
951  GCCCATGTGG GCGGAGCTTA AAGCCCTGGC CGCCTGCAGG GACGGCCGGG
1001 TGCACCGGGC AGCAGACCCC TTGGCGGGGC TAAAGGACCT CAAGGAGGTC
1051 CGGGGCCTCC TCGCCAAGGA CCTCGCCGTC TTGGCCTCGA GGGAGGGGCT
1101 AGACCTCGTG CCCGGGGACG ACCCCATGCT CCTCGCCTAC CTCCTGGACC
1151 CCTCCAACAC CACCCCCGAG GGGGTGGCGC GGCGCTACGG GGGGGAGTGG
1201 ACGGAGGACG CCGCCCACCG GGCCCTCCTC TCGGAGAGGC TCCATCGGAA
1251 CCTCCTTAAG CGCCTCGAGG GGGAGGAGAA GCTCCTTTGG CTCTACCACG
1301 AGGTGGAAAA GCCCCTCTCC CGGGTCCTGG CCCACATGGA GGCCACCGGG
1351 GTACGGCTGG ACGTGGCCTA CCTTCAGGCC CTTTCCCTGG AGCTTGCGGA
1401 GGAGATCCGC CGCCTCGAGG AGGAGGTCTT CCGCTTGGCG GGCCACCCCT
1451 TCAACCTCAA CTCCCGGGAC CAGCTGGAAA GGGTGCTCTT TGACGAGCTT
1501 AGGCTTCCCG CCTTGGGGAA GACGCAAAAG ACGGGCAAGC GCTCCACCAG
1551 CGCCGCGGTG CTGGAGGCCC TACGGGAGGC CCACCCCATC GTGGAGAAGA
1601 TCCTCCAGCA CCGGGAGCTC ACCAAGCTCA AGAACACCTA CGTGGACCCC
1651 CTCCCAAGCC TCGTCCACCC GAGGACGGGC CGCCTCCACA CCCGCTTCAA
1701 CCAGACGGCC ACGGCCACGG GGAGGCTTAG TAGCTCCGAC CCCAACCTGC
1751 AGAACATCCC CGTCCGCACC CCCTTGGGCC AGAGGATCCG CCGGGCCTTC
1801 GTGGCCGAGG CGGGATGGGC GTTGGTGGCC CTGGACTATA GCCAGATAGA
1851 GCTCCGCGTC CTCGCCCACC TCTCCGGGGA CGAGAACCTG ATCAGGGTCT
1901 TCCAGGAGGG GAAGGACATC CACACCCAGA CCGCAAGCTG GATGTTCGGC
```

```
1951 GTCCCCCCGG AGGCCGTGGA CCCCCTGATG CGCCGGGCGG CCAAGACGGT
2001 GAACTTCGGC GTCCTCTACG GCATGTCCGC CCATAGGCTC TCCCAGGAGC
2051 TTGCCATCCC CTACGAGGAG GCGGTGGCCT TTATAGAGCG CTACTTCCAA
2101 AGCTTCCCCA AGGTGCGGGC CTGGATAGAA AAGACCCTGG AGGAGGGGAG
2151 AAAGCGGGGC TACGTGGAAA CCCTCTTCGG AAGAAGGCGC TACGTGCCCG
2201 ACCTCAACGC CCGGGTGAAG AGCGTCAGGG AGGCCGCGGA GCGCATGGCC
2251 TTCAACATGC CCGTCCAGGG CACCGCCGCC GACCTCATGA AGCTCGCCAT
2301 GGTGAAGCTC TTCCCCCGCC TCCGGGAGAT GGGGGCCCGC ATGCTCCTCC
2351 AGGTCCACGA CGAGCTCCTC CTGGAGGCCC CCCAAGCGCG GGCCGAGGAG
2401 GTGGCGGCTT TGGCCAAGGA GGCCATGGAG AAGGCCTATC CCCTCGCCGT
2451 GCCCCTGGAG GTGGAGGTGG GGATGGGGGA GGACTGGCTT TCCGCCAAGG
2501 GTTAG
```

FIG. 5B-2

```
1    MEAMLPLFEP KGRVLLVDGH HLAYRTFFAL KGLTTSRGEP VQAVYGFAKS  50
51   LLKALKEDGY KAVFVVFDAK APSFRHEAYE AYKAGRAPTP EDFPRQLALI 100
101  KELVDLLGFT RLEVPGYEAD DVLATLAKKA EKEGYEVRIL TADRDLYQLV 150
151  SDRVAVLHPE GHLITPEWLW QKYGLKPEQW VDFRALVGDP SDNLPGVKGI 200
201  GEKTALKLLK EWGSLENLLK NLDRVKPENV REKIKAHLED LRLSLELSRV 250
251  RTDLPLEVDL AQGREPDREG LRAFLERLEF GSLLHEFGLL EAPAPLEEAP 300
301  WPPPEGAFVG FVLSRPEPMW AELKALAACR DGRVHRAADP LAGLKDLKEV 350
351  RGLLAKDLAV LASREGLDLV PGDDPMLLAY LLDPSNTTPE GVARRYGGEW 400
401  TEDAAHRALL SERLHRNLLK RLQGEEKLLW LYHEVEKPLS RVLAHMEATG 450
451  VRLDVAYLQA LSLELAEEIR RLEEEVFRLA GHPFNLNSRD QLERVLFDEL 500
501  RLPALGKTQK TGKRSTSAAV LEALREAHPI VEKILQHREL TKLKNTYVDP 550
551  LPSLVHPNTG RLHTRFNQTA TATGRLSSSD PNLQNIPVRT PLGQRIRRAF 600
601  VAEAGWALVA LDYSQIELRV LAHLSGDENL IRVFQEGKDI HTQTASWMFG 650
651  VPPEAVDPLM RRAAKTVNFG VLYGMSAHRL SQELAIPYEE AVAFIERYFQ 700
701  SFPKVRAWIE KTLEEGRKRG YVETLFGRRR YVPDLNARVK SVREAAERMA 750
751  FNMPVQGTAA DLMKLAMVKL FPRLREMGAR MLLQVHDELL LEAPQARAEE 800
801  VAALAKEAME KAYPLAVPLE VEVGMGEDWL SAKG                  834
```

FIG. 6A

```
1    ATGGAGGCGA TGCTTCCGCT CTTTGAACCC AAAGGCCGGG TCCTCCTGGT
51   GGACGGCCAC CACCTGGCCT ACCGCACCTT CTTCGCCCTG AAGGGCCTCA
101  CCACGAGCCG GGGCGAACCG GTGCAGGCGG TCTACGGCTT CGCCAAGAGC
151  CTCCTCAAGG CCCTGAAGGA GGACGGGTAC AAGGCCGTCT TCGTGGTCTT
201  TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGAG GCCTACAAGG
251  CGGGGAGGGC CCCGACCCCC GAGGACTTCC CCCGGCAGCT CGCCCTCATC
301  AAGGAGCTGG TGGACCTCCT GGGGTTTACC CGCCTCGAGG TCCCCGGCTA
351  CGAGGCGGAC GACGTCCTCG CCACCCTGGC CAAGAAGGCG GAAAAGGAGG
401  GGTACGAGGT GCGCATCCTC ACCGCCGACC GCGACCTCTA CCAACTCGTC
451  TCCGACCGCG TCGCCGTCCT CCACCCCGAG GGCCACCTCA TCACCCCGGA
501  GTGGCTTTGG CAGAAGTACG GCCTCAAGCC GGAGCAGTGG GTGGACTTCC
551  GCGCCCTCGT GGGGGACCCC TCCGACAACC TCCCCGGGGT CAAGGGCATC
601  GGGGAGAAGA CCGCCCTCAA GCTCCTCAAG GAGTGGGGAA GCCTGGAAAA
651  CCTCCTCAAG AACCTGGACC GGGTAAAGCC AGAAAACGTC CGGGAGAAGA
701  TCAAGGCCCA CCTGGAAGAC CTCAGGCTTT CCTTGGAGCT CTCCCGGGTG
751  CGCACCGACC TCCCCCTGGA GGTGGACCTC GCCCAGGGGC GGGAGCCCGA
801  CCGGGAGGGG CTTAGGGCCT TCCTGGAGAG GCTGGAGTTC GGCAGCCTCC
851  TCCACGAGTT CGGCCTCCTG GAGGCCCCCG CCCCCCTGGA GGAGGCCCCC
901  TGGCCCCCGC CGGAAGGGGC CTTCGTGGGC TTCGTCCTCT CCCGCCCCGA
951  GCCCATGTGG GCGGAGCTTA AAGCCCTGGC CGCCTGCAGG GACGGCCGGG
1001 TGCACCGGGC AGCGGACCCC TTGGCGGGGC TAAAGGACCT CAAGGAGGTC
1051 CGGGGCCTCC TCGCCAAGGA CCTCGCCGTC TTGGCCTCGA GGGAGGGGCT
1101 AGACCTCGTG CCCGGGGACG ACCCCATGCT CCTCGCCTAC CTCCTGGACC
1151 CCTCCAACAC CACCCCCGAG GGGGTGGCGC GGCGCTACGG GGGGGAGTGG
1201 ACGGAGGACG CCGCCCACCG GGCCCTCCTC TCGGAGAGGC TCCATCGGAA
1251 CCTCCTTAAG CGCCTCCAGG GGGAGGAGAA GCTCCTTTGG CTCTACCACG
1301 AGGTGGAAAA GCCCCTCTCC CGGGTCCTGG CCCACATGGA GGCCACCGGG
1351 GTACGGCTGG ACGTGGCCTA CCTGCAGGCC CTTTCCCTGG AGCTTGCGGA
1401 GGAGATCCGC CGCCTCGAGG AGGAGGTCTT CCGCTTGGCG GGCCACCCCT
1451 TCAACCTCAA CTCCCGGGAC CAGCTGGAGA GGGTGCTCTT TGACGAGCTT
1501 AGGCTTCCCG CCTTGGGGAA GACGCAAAAG ACGGGCAAGC GCTCCACCAG
1551 CGCCGCGGTG CTGGAGGCCC TACGGGAGGC CCACCCCATC GTGGAGAAGA
1601 TCCTCCAGCA CCGGGAGCTC ACCAAGCTCA AGAACACCTA CGTGGACCCC
1651 CTCCCAAGCC TCGTCCACCC GAATACGGGC CGCCTCCACA CCCGCTTCAA
1701 CCAGACGGCC ACGGCCACGG GGAGGCTTAG TAGCTCCGAC CCCAACCTGC
1751 AGAACATCCC CGTCCGCACC CCCTTGGGCC AGAGGATCCG CCGGGCCTTC
1801 GTGGCCGAGG CGGGTTGGGC GTTGGTGGCC CTGGACTATA GCCAGATAGA
1851 GCTCCGCGTC CTCGCCCACC TCTCCGGGGA CGAGAACCTG ATCAGGGTCT
1901 TCCAGGAGGG GAAGGACATC CACACCCAGA CCGCAAGCTG GATGTTCGGC
1951 GTCCCCCCGG AGGCCGTGGA TCCCCTGATG CGCCGGGCGG CCAAGACGGT
```

FIG. 6B-1

```
2001 GAACTTCGGC GTCCTCTACG GCATGTCCGC CCATAGGCTC TCCCAGGAGC
2051 TTGCCATCCC CTACGAGGAG GCGGTGGCCT TTATAGAGCG CTACTTCCAA
2101 AGCTTCCCCA AGGTGCGGGC CTGGATAGAA AAGACCCTGG AGGAGGGGAG
2151 GAAGCGGGGC TACGTGGAAA CCCTCTTCGG AAGAAGGCGC TACGTGCCCG
2201 ACCTCAACGC CCGGGTGAAG AGCGTCAGGG AGGCCGCGGA GCGCATGGCC
2251 TTCAACATGC CCGTCCAGGG CACCGCCGCC GACCTCATGA AGCTCGCCAT
2301 GGTGAAGCTC TTCCCCCGCC TCCGGGAGAT GGGGGCCCGC ATGCTCCTCC
2351 AGGTCCACGA CGAGCTCCTC CTGGAGGCCC CCCAAGCGCG GGCCGAGGAG
2401 GTGGCGGCTT TGGCCAAGGA GGCCATGGAG AAGGCCTATC CCCTCGCCGT
2451 GCCCCTGGAG GTGGAGGTGG GGATGGGGGA GGACTGGCTT TCCGCCAAGG
2501 GTTAG
```

FIG. 6B-2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/26412

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68; C12P 19/34; C12N 9/12
US CL : 435/6, 91.1, 91.2, 194

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)
U.S. : 435/6, 91.1, 91.2, 194

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 6,225,092 B1 (KILGER et al.) 01 May 2001 (01.05.2001), entire document.	1-32
A	US 5,928,906 A (KOSTER et al.) 27 July 1999 (27.07.1999), entire document.	1-32
A	US 6,365,375 B1 (DIETMAIER et al.) 02 April 2002 (02.04.2002), entire document.	1-32
A	US 5,976,842 A (WURST) 02 November 1999 (02.11.1999), entire document.	1-32
A	US 5,436,149 A (BARNES) 25 July 1995 (25.07.1995), entire document.	1-32
A	US 6,077,664 A (SLATER et al.) 20 June 2000 (20.06.2000), entire document.	1-32

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent 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	Date of mailing of the international search report
22 April 2004 (22.04.2004)	21 MAY 2004
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230	Authorized officer Kenneth R Horlick Telephone No. (571) 272-1600

Form PCT/ISA/210 (second sheet) (July 1998)

Continuation of B. FIELDS SEARCHED Item 3:
USPAT, DERWENT WPI, MEDLINE, BIOSIS
search terms: polymerase, HB8, oshimai, scotoductus, 1B21, GK24