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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 October 2009 (29.10.2009)

PCT

(10) International Publication Number
WO 2009/131919 A2

(51) **International Patent Classification**:
C12Q 1/68 (2006.01) ***C12P 19/34*** (2006.01)
C07H 21/04 (2006.01) ***C12N 9/22*** (2006.01)
C12P 21/06 (2006.01)

(21) **International Application Number**:
PCT/US2009/041065

(22) **International Filing Date**:
20 April 2009 (20.04.2009)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/046,987 22 April 2008 (22.04.2008) US

(71) **Applicant** *(for all designated States except US)*: **NEW ENGLAND BIOLABS, INC.** [US/US]; 240 County Road, Ipswich, MA 01938 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **GARDNER, Andrew** [US/US]; 13 School Street, Manchester, MA 01944 (US). **GREENOUGH, Lucia** [US/US]; 32 Fellows Road, Ipswich, MA 01938 (US). **JACK, William, E.** [US/US]; 31 Mayflower Drive, Wenham, MA 01984 (US).

(74) **Agent**: **STRIMPEL, Harriet, M.**; New England Biolabs, Inc., 240 County Road, Ipswich, MA 01938 (US).

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

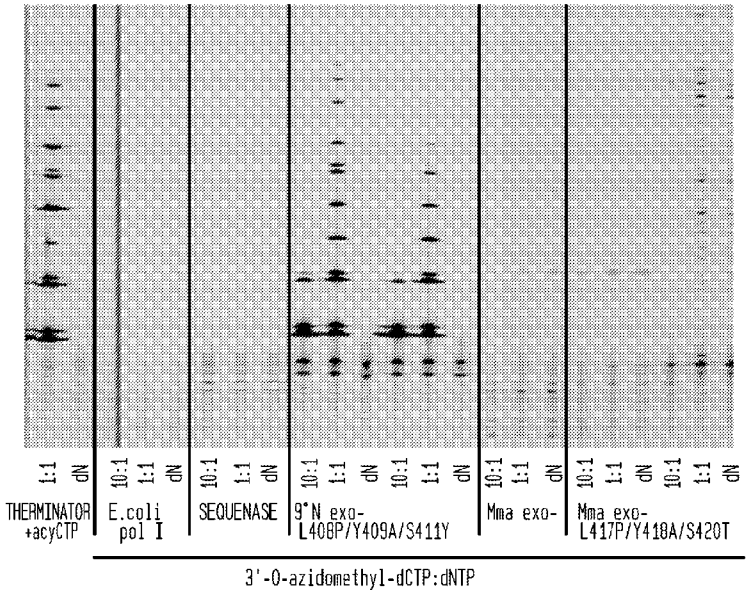
(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR),

[Continued on next page]

(54) **Title**: POLYMERASES FOR INCORPORATING MODIFIED NUCLEOTIDES

(57) **Abstract**: Compositions and methods are provided that relate to a recombinant protein with DNA polymerase activity in which one or more amino acids are mutated compared with the corresponding wild type protein. The recombinant protein is capable of incorporating one or more modified nucleotides into a nucleic acid substrate with a specific activity greater than 200.

FIG. 8



WO 2009/131919 A2

OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

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

Polymerases for Incorporating Modified Nucleotides

BACKGROUND

DNA polymerases have played a central role in the development of molecular biology. Their use is central in a wide range of laboratory protocols, including DNA sequencing (Sanger, *et al., Proc. Natl. Acad. Sci., USA* 74:5463-5467 (1977)), strand displacement amplification (SDA; Walker, *et al., Proc. Natl. Acad. Sci., USA* 89:392-396 (1992)), probe-labeling, site-directed mutagenesis, polymerase chain reaction (PCR) (Saiki, *et al., Science*, 230:1350-1354 (1985)) and cloning. These applications depend critically on the ability of polymerases to faithfully replicate DNA.

A number of applications require polymerases that are able to incorporate modified nucleotides. For example, chain terminator DNA sequencing utilizes incorporation of a chain-terminating nucleotide, most often a ddNTP, to deduce the pattern of bases in a sequencing sample (Sanger, *et al., Proc. Natl. Acad. Sci., USA* 74:5463-5467 (1977)). Additional applications rely on incorporation of nucleotides with modified bases, often to aid in detection of the polymerized product. One such application is the incorporation of nucleotides with fluorescent bases, allowing analysis of the products of chain-terminating DNA sequencing reactions (Prober, *et al., Science* 238:336-341 (1987)). Modifications of this method have also been described that focus on single nucleotide loci, allowing detection of single nucleotide polymorphisms (Ellison, *et al., Biotechniques* 17:742-753 (1994); Chen and Kwok, *Nucleic Acids Res.* 15:347-353 (1997); Chen *et al., Genome Research* 9:492-498 (1999)). Such single-base detection methods have been instructive in genetic testing and analysis.

Sequencing-by-synthesis methods (Metzker *Genome Res* 15(12): 1767-76 (2005)) have emerged as an important technology that have enabled DNA sequencing on a massive scale while reducing costs compared to conventional methods. Sequencing-by-synthesis methods rely on incorporation and detection of "reversible terminators" by a DNA polymerase. Reversible 3'-modified nucleotide terminators can be used in massively parallel sequencing-by-synthesis methods. Inadequate terminal incorporation of 3'-modified nucleotides by naturally occurring DNA polymerases is a limiting factor in the success of these methods. It would be desirable to create a novel DNA polymerase that provides improved efficiency of incorporation of 3'-modified nucleotides to overcome this limitation.

Another difficulty encountered in the methods stated above arises from the inherent fidelity of naturally occurring DNA polymerases. This results in a bias against modified nucleotide incorporation that can sometimes be overcome by increasing the concentration of modified nucleotides in the reaction mixture. Even then, incorporation may be incomplete and non-uniform, limiting the sensitivity of detection and producing patterns that may not accurately reflect the actual nucleotide sequence being replicated. This complicates determinations such as DNA sequence analysis.

Accordingly, there is a need in the art for DNA polymerases that more readily incorporate modified nucleotides. Since a number of methods require a step in which the DNA is denatured at temperatures up to 95°C, thermostable DNA polymerases are preferable. One thermostable enzyme that has been extensively used is *Taq* DNA polymerase, along with a variety of engineered versions of this enzyme. Extensive studies have characterized the

ability of this enzyme to incorporate nucleotides that act as terminators (*e.g.*, ddNTPs) and nucleotides with modified bases (*e.g.*, dye-labeled). Such modifications can affect polymerization. For example, terminator DNA sequencing reactions with the F667Y version of *Taq* DNA polymerase (also know by the trade name Thermosequenase™, USB, Inc., Cleveland, OH) show "…less uniform peak height patterns when compared to primer chemistry profiles, suggesting that the dyes and/or their linker arms affect enzyme selectivity." (Brandis, *Nucleic Acids Res.* 27:1912-1918 (1999)).

Taq DNA polymerase is a Family A DNA polymerase. Amino acid similarities allow the classification of most DNA polymerases into three Families, A, B and C, according to similarities with *Escherichia coli* polymerases I, II and III, respectively (Ito and Braithwaite, *Nucleic Acids Res.* 19:4045-4057 (1991); Heringa and Argos, *The Evolutionary Biology of Viruses* , Morse, S.S., ed., pp. 87-103, Raven Press, New York (1992)).

Family B polymerases include thermostable polymerases from thermophilic archaea. One such example is Vent® (New England Biolabs (NEB), Inc., Ipswich, MA) DNA polymerase, originally isolated from *Thermococcus litoralis* (Perler, *et al.*, *Proc. Natl. Acad. Sci. USA* 89:5577-5581 (1992)). Vent® DNA polymerase has a relatively high K_m for nucleotides (Kong, *et al.*, *J. Biol. Chem.* 268:1965-1975 (1993)), and functions poorly in incorporating dye-substituted terminators in DNA sequencing reactions ("CircumVent™: Questions and Answers," *The NEB Transcript*, September 1992, p. 12-13). Incorporation of unsubstituted didexynucleoside triphosphate (ddNTP) terminators is also inefficient with this enzyme (Gardner and Jack, *Nucleic Acids Res.* 27:2545-2553 (1999)). Thus, Vent® DNA polymerase does not

appear to be a promising candidate for applications requiring incorporation of modified nucleotides.

Archaeon DNA polymerase mutants have been described that somewhat increase the incorporation efficiency of specific classes of chain terminators, namely ddNTPs and 3'-dNTPs. 9°N exo-Y409A/A485L (Therminator™ II, NEB, Inc., Ipswich, MA) has been used in massive parallel sequencing with 3'-modified reversible terminators Seo et al. *J Org Chem* 68(2): 609-12 (2003)). However, the incorporation of the 3' modified reversible terminator by this polymerase mutant was relatively inefficient and required long incubation times and high concentrations of 3'-modified nucleotide reversible terminators to complete the reaction.

Because DNA polymerases discriminate against nucleotide analogs such as ddNTPs and rNTPs resulting in reduced binding affinity and slowed rates of catalysis (Gardner et al. *J Biol Chem* 279(12): 11834-42 (2004)), it has proved extremely challenging to engineer DNA polymerases that will incorporate non-natural nucleotide analogs with fidelity while maintaining high reaction efficiency. Canard et al. (*Proc Natl Acad Sci USA* 92(24): 10859-63 (1995)) demonstrated incorporation of 3'-esterified nucleotides by DNA polymerases but noted that incorporation of these modifed nucleotides was inefficient. In addition, DNA polymerases could use the 3'-esterified linkage as a template to add a subsequent deoxynucleotide triphosphosphate (dNTP) on the 3' end (Canard et al. *Proc Natl Acad Sci USA* 92(24): 10859-63 (1995)). In light of the above, it would be desirable to design a DNA polymerase with a higher efficiency of 3'-modified nucleotide incorporation.

SUMMARY

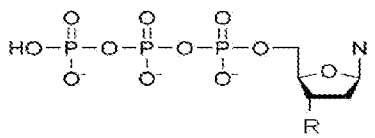
In an embodiment of the invention, a recombinant protein with DNA polymerase activity is described that may be characterized as containing an amino acid sequence that has at least 90% amino acid sequence identity with SEQ ID NO:1. One or more amino acids in the recombinant protein are mutated compared with the corresponding wild type protein. The mutation for example may be located in SEQ ID NO:1 or in Region III. The recombinant protein is capable of (i) incorporating one or more nucleotides into a nucleic acid substrate with a specific activity greater than 200, more specifically a specific activity of greater than 1000, more specifically a specific activity of greater than 5000; and (ii) incorporating one or more modified nucleotides into the nucleic acid substrate with at least two fold greater efficiency than for corresponding wild type DNA polymerase.

Examples of the recombinant protein include: a 9°N archael polymerase, and the mutated amino acids comprising D141A and E143A and an additional mutation selected from the group consisting of: P410V; S411T; L408S/Y409A/P410V; L408P/Y409A/S411T; P410R/S411T; L408S/Y409A/P410V/S411T; L408P/Y409A/P410V/S411T; N491L/Y494S; N491V/Y494H; R406S/L408R; R406L/L408E; R406T/L408R; R406V/L408R; R406T/L408E; R406V/L408R; R406E/L408G; R406P/L408G; Y409A/R406V; Y409A/R406S/L408K; Y409A/R406S/L408R; Y409A/R406T/L408K; Y409A/R406T/L408R; Y409A/R406H/L408G; Y409A/R406Y/L408G; Y409A/R406L/L408G; Y409A/R406P/L408C; Y409A/R406S/L408I; Y409A/R406V/L408Y; Y409A/R406V/A485L; N491L/Y494S; N491V/Y494H; Y409A/R406(nucleophilic amino acid)/L408(basic amino acid); Y409A/R406(hydrophobic amino

acid)/L408(small amino acid); Y409A/R406/L408(hydrophobic amino acid)/L408(small amino acid); P410V/A485L; L408S/P410V/A485L; Y409A/S411T/A485L; L408S/Y409A/P410V/A485L; L408P/Y409A/S411T/A485L; Y409A/P410R/S411T/A485L; L408S/Y409A/P410V/S411T/A485L; L408P/Y409A/P410V/S411T/A485L; N491L/Y494S/A485L; N491V/Y494H/A485L; R406S/L408R/A485L; R406L/L408E/A485L; R406T/L408R/A485L; R406V/L408R/A485L; R406T/L408E/A485L; R406V/L408R/A485L; R406E/L408G/A485L; R406P/L408G/A485L; Y409A/R406S/L408K/A485L; Y409A/R406S/L408R/A485L; Y409A/R406T/L408K/A485L; Y409A/R406T/L408R/A485L; Y409A/R406H/L408G/A485L; Y409A/R406Y/L408G/A485L; Y409A/R406L/L408G/A485L; Y409A/R406P/L408C/A485L; Y409A/R406S/L408I/A485L; Y409A/R406V/L408Y/A485L.

Additional examples include a recombinant protein from *Methanococcus maripolludis* archael DNA polymerase, where the mutated amino acids are selected from D153A/E155A/L417S/P419V and D153A/E155A/L417P/Y418A/S420T.

Examples of modified nucleotides include nucleotides that are selected from 3' terminators and 3' reversible terminators, for example, according to the following chemical structure, when N is a nucleoside, and the R group on the 3' position of the ribose may be larger than a hydroxyl group. In particular, the R group may be substituted by the groups as listed here



R= -H, -SH, -N_3, -F, -Cl, -azidomethyl, -NH_2, -anthranyloyl -fluothioureido, -chain, -amd, -O-allyl, -O-aminoallyl, -O-azidomethyl, -O-methyl, -O-phophate, -O-diphophate, -O-(2-nitrobenzyl), -O-[N6(anthranyl)amidohex

Additionally, modified nucleotides may be selected from the group consisting of:
2'-deoxy-3'-anthranyloyl-dNTPs (3'-ant-dNTPs) 3'-{N3-[3-carboxylato-4-(3-oxido- 6-oxo-6H-xanthen-9-yl)phenyl]thioureido}-3 '-deoxythymidine 5'-triphosphate (3'-fluothioureido-dTTP), 3'-deoxy-3'-(N-methylanthranyloylamino)thymidine 5'-triphosphate (3'-amd-dTTP), 3'-O-[N6(N-methylanthranyl)amidohexanoyl]-dGTP (3 -chain-dGTP), and 3'-O-[N6(anthranyl)amidohex (3'-chain-dATP).

Examples of modified nucleotides include labelled modified nucleotides including fluorescent labels.

In an embodiment of the invention, a method is provided for incorporating modified nucleotides into a nucleic acid by reacting a nucleic acid with a recombinant protein as described above and at least one modified nucleotide.

In an additional embodiment of the invention, a kit is provided that contains a recombinant protein as described above and instructions for use. The kit may further include a modified nucleotide.

In a further embodiment of the invention, a method is provided of screening for a recombinant protein as described above wherein the method includes: (a) determining the size of a substrate after

addition of a modified nucleotide during a polymerization reaction; and (b) measuring at least one of an increase in chain terminator incorporation, and a decrease in average reaction product size, to determine efficiency of incorporation by the composition.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A-1D show titration assays for chain terminator incorporation efficiency. The relative efficiency of incorporation of modified nucleotide terminators was assessed using the titration assay described in Gardner and Jack *Nucleic Acids Researc*h 30: 605-613 (2002). A dye-labeled oligonucleotide primer 5'-AGTGAATTCG AGCTCGGTAC CCGGGGATCC TCTAGAGTCG ACCTGCAGGC-3' was annealed to a single-stranded M13mp18 DNA template (Accession No. X02513) and extended by a DNA polymerase in the presence of various ratios of 3'-azido-ddCTP:dNTP (10:1, 2.5:1, 1:2.5) or dNTP ("dN" in the figure) alone in 1 X ThermoPol buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$,10 mM KCl, 2 mM $MgSO_4$, 0.1 % Triton X-100, pH 8.8 @ 25°C). Reactions were incubated and termination products were resolved on 20% denaturing polyacrylamide gel electrophoresis (PAGE).

Bands correspond to DNA with an added 3'-azido-ddCMP terminator. This showed that DNA polymerase mutants that increased incorporation efficiency required less 3'-azido-ddCTP to produce termination fragments than did the corresponding non-mutated DNA polymerase. The results show that non-mutated DNA polymerases failed to produce termination fragments even at a 10:1 ratio of 3'-azido-ddCTP terminator:dNTP. Figure 1A is 9°N exo- . Figure 1B is 9°N exo- P410V. Figure 1C is 9°N exo- L408S/ P410V

and Figure 1D is 9°N exo- L408S/Y409A/P410V. 9°N exo- L408S/P410V DNA polymerase (Figure 1C) incorporated 3'-azido-ddCTP at least 2-fold more efficiently than wild-type 9°N exo-DNA polymerase (Figure 1A) and at least 4-fold more efficiently than 9°N exo- P410V DNA polymerase (Figure 1B). 9°N exo- L408S/Y409A/P410V DNA polymerase (Figure 1D) incorporated 3'-azido-ddCTP at least 4-fold more efficiently than 9°N exo- L408S/P410V DNA polymerase (Figure 1C).

Figure 2 shows a kinetic analysis of 3'-azido-ddCTP incorporation by 9°N exo- L408S/Y409A/P410V DNA polymerase. An IR800-dye-labeled synthetic primer 5'-AGTGAATTCG AGCTCGGTAC CCGGGGATCC TCTAGAGTCG ACCTGCAGGC-3' (SEQ ID NO:18) was annealed to a template 3'-TCACTTAAGC TCGAGCCATG GGCCCCTAGG AGATCTCAGC TGGACGTCCG GATCCTATAC TAATCCC-5' (SEQ ID NO:19) and used as a substrate to measure rates of 3'-azido-ddCTP incorporation. Incorporation of varying concentrations of 3'-azido-ddCTP (100, 50, 25, 12.5, 6.25, and 3.125 μM) by 9°N exo- L408S/Y409A/P410V DNA polymerase was monitored over a three-minute time course. Reaction aliquots were sampled at 10, 20, 30, 45, 60, and 180 seconds and analyzed by 20% denaturing PAGE. "S" is the unextended primer and "P" is the fully extended product corresponding to 3'-azido-ddCMP addition at the +1 position. The gel shows 2 bands, the larger band corresponding to the extension of the primer by the terminator. The extent of the incorporation was shown to be dependent on the concentration of modified nucleotide such that the greater the concentration of the primer, the faster the reaction.

Figure 3 shows a kinetic analysis of 3'-amino-ddCTP and 3'-azido-ddCTP incorporation by 9°N exo- L408S/Y409A/P410V DNA

polymerase. An IR800-dye-labeled synthetic primer 5'-AGTGAATTCG AGCTCGGTAC CCGGGGATCC TCTAGAGTCG ACCTGCAGGC-3' (SEQ ID NO:18) was annealed to a template 3'-TCACTTAAGC TCGAGCCATG GGCCCCTAGG AGATCTCAGC TGGACGTCCG GATCCTATAC TAATCCC-5' (SEQ ID NO:19) and used as a substrate. Reaction aliquots were sampled at 10, 30, 60, 120 and 300 seconds and analyzed by 20% denaturing PAGE. Unextended primer (0) and fully extended product (+1) are indicated. The gel shows that 10 μM of each of the modified nucleotides is sufficient for efficient incorporation by 9°N exo-L408S/Y409A/P410V DNA polymerase.

Figure 4 shows a time course plot of 3'-azido-ddCTP incorporation by 9°N exo- L408S/Y409A/P410V DNA polymerase. Linear rates of 3'-azido-ddCTP incorporation derived from these curves are 100 μM: 0.05 s^{-1}; 50 μM: 0.05 s^{-1}; 25 μM: 0.05s^{-1}; 12.5 μM: 0.03s^{-1}; 6.25 μM: 0.03s^{-1}; 3.125 μM: 0.02 s^{-1}.

Figure 5 shows a sequence similarity search for Family B DNA polymerase Region II. Region II in 9°N DNA polymerase (DFRSLYPSIIITH) (SEQ ID NO:1) was used to query Genbank for highly similar sequences (>90% identity; e-values less that 0.003) using BLAST (Altschul et al. *Nucleic Acids Res* 25(17): 3389-402 (1997)).

Figures 6-1 to 6-11 show the results of a Clustal W sequence in which Family B DNA polymerases from archaea and bacteriophage were aligned. Family B DNA polymerases from *Thermococcus* sp. 9°N (9N), *Thermococcus* sp. 9°N mutant (9N DNAP), *Thermococcus gorgonarius* (TGO), *Thermococcus kodakarensis* (KOD), *Pyrococcus horikoshii* (P_horikoshii),

Thermococcus aggregans (T_aggregans), *Thermococcus litoralis* (Vent_T.litoralis), *Methanococcus maripaludis* (Mma), *Methanococcus jannaschii* (M_jannaschii), *Methanothermobacter thermautotrophicus* str. Delta H (Mth_PolB1), bacteriophage RB69 (RB69), bacteriophage T4 (T4), and *Methanoculleus marisnigri* JR1 (M_marisigri) were aligned using software Clustal W 2.0 (http://www.ebi.ac.uk/Tools/clustalw2/index.html). Amino acids in Region II and Region III are highlighted in boxes. Conserved amino acids are noted with an asterisk and similar amino acids by a dot (.) or colon (:).

Figure 7 shows specific activities for 9°N D141A/E143A/L408S/Y409A/P410V and 9°N D141A/E143A/Y409V/A485L DNA polymerase dNTP incorporation.

Figure 8 shows the results of extending a 5'-dye-labeled oligonucleotide primer 5'-AGTGAATTCG AGCTCGGTAC CCGGGGATCC TCTAGAGTCG ACCTGCAGGC-3' (SEQ ID NO:18) annealed to a single-stranded M13mp18 DNA template (Accession No. X02513) by various DNA polymerases and mutants. The polymerases used were an *E. coli* DNA polymerase I (pol I) (NEB, Inc., Ipswich, MA), Sequenase™ (USB, Inc., Cleveland, OH), 9°N exo- L408P/Y409A/S411T (SEQ ID NO:6), Mma exo- (SEQ ID NO:22) (Figure 10A), or Mma exo- L417P/Y418A/S420T (SEQ ID NO:23) (Figure 10B) in the presence of various ratios with 10:1 or 1:1 3'-O-azidomethyl-dCTP:dNTP. Reactions were also performed in the absence of terminators to ensure that synthesis by each DNA polymerase was sufficient to extend primers without premature termination ("dNTP" lanes). Reactions with Therminator™ (NEB, Inc., Ipswich, MA) and 9°N exo- L408P/Y409A/S411T DNA polymerase (SEQ ID NO:6) were incubated at 72°C for 30 minutes.

Reactions with *E. coli* Polymerase I (pol I) (NEB, Inc., Ipswich, MA), Sequenase™, (USB, Inc., Cleveland, OH), Mma exo- and Mma exo- L417P/Y418A/S420T were incubated at 37°C for 30 minutes. Each band corresponds to a DNA fragment terminated by a 3'-azido-dCMP. Using a titration assay as described in Gardner and Jack (2002), the relative 3'-azido-dCTP incorporation efficiency was determined for a series of DNA polymerases. The incorporation efficiency for 9°N exo- L408P/Y409A/S411T was greater than Mma exo- L417P/Y418A/S420T, which was greater than Mma exo-, which was greater than Sequenase™ (USB, Inc., Cleveland, OH), which was greater than *E. coli* DNA polymerase I (pol I) (NEB, Inc., Ipswich, MA).

Figure 9 shows examples of nucleotide terminators modified at the 3' position (R). "N" can be adenine, cytosine, guanosine or thymine.

Figure 10A is the amino acid sequence (SEQ ID NO:22) of an exonuclease minus DNA polymerase from Mma exo- created by site-directed mutagenesis to change D141A and E143A. Mma exo- was used to evaluate incorporation of 3'-O-azidomethyl-dCTP in Figure 8 and Example 4.

Figure 10B is the amino acid sequence (SEQ ID NO:23) of an exonuclease minus mutant DNA polymerase from Mma exo- created by site-directed mutagenesis to change D141A and E143A and L417P/Y418A/S420T. Mma exo- L417P/Y418A/S420T was used to evaluate incorporation of 3'-O-azidomethyl-dCTP in Figure 8 and Example 4.

DETAILED DESCRIPTION OF THE EMBODIMENTS

Embodiments of the present invention describe modified Family B archaeon DNA polymerases and related codon-substituted mutants capable of incorporating selected modified nucleotides into nucleic acids with improved efficiency. Examples include: Vent® DNA polymerase (Kong, *et al.*, *J. Biol. Chem*. 268:1965-1975 (1993); and U.S. Patents 5,500,363, 5,834,285, and 5,352,778); *Pyrococcus furiosus* (*Pfu*) DNA polymerase (U.S. Patents 5,489,523 and 5,827,716); *Thermococcus barossii* (*Tba*) DNA polymerase (U.S. Patent 5,882,904); and 9°N7 DNA polymerase (Southworth et al. *Proc. Natl. Acad. Sci. USA* 93: 5281-5285 (1996)).

Some of the above polymerases have 3-5' exonuclease activity. One function of this activity is "proofreading," wherein the polymerase can remove 3' nucleotides before proceeding with polymerization. Incorrectly base-paired nucleotides, or aberrant nucleotides are preferentially removed by this activity, thus increasing the fidelity of replication (Kornberg, *DNA Replication* p. 127 (1980)). Modified nucleotides might reasonably be expected to be identified as "aberrant," and, even if incorporated, be subject to removal by this activity. To avoid this possibility, mutants can be created that lack or have diminished exonuclease activity. Such mutants include Family B archeal DNA polymerases with ≥90% identity to 9°N DNA polymerase and with ≥30% and <90% identity to 9°N DNA polymerase as follows:

Family B archaeal DNA polymerases with ≥90% identity to 9°N DNA polymerase can be obtained from host cells such as: *Pyrococcus kodakaraensis, Pyrococcus furiosus*, *Pyrococcus woesei*, *Pyrococcus glycovorans, Pyrococcus abyssi*, *Pyrococcus* sp. GB-D, *Pyrococcus* sp. ST700, *Pyrococcus horikoshii* OT3, *Thermococcus litoralis, Thermococcus gorgonarius, Thermococcus* sp AM4,

Thermococcus sp. GE8, *Thermococcus thioreducens, Thermococcus onnurineus NA1, Thermococcus sp. GT, Thermococcus zilligii, Thermococcus hydrothermalis, Thermococcus fumicolans, Thermococcus barophilus* MP, and *Thermococcus* sp. TY.

Family B archaeal DNA polymerases with ≥30% and <90% identity to 9°N DNA polymerase can be obtained from host cells such as: *Aciduliprofundum boonei, Aeropyrum pernix, Archaeoglobus fulgidus, Caldivirga maquilingensis, Candidatus korarchaeum cryptofilum, Desulfurococcus kamchatkensis, Hyperthermus butylicus, Ignicoccus hospitalis* KIN4/I, *Methanosphaera stadtmanae, Metallosphaera sedula, Methanobrevibacter smithii, Methanocaldococcus jannaschii, Methanococcoides burtonii, Methanococcus maripaludis, Methanococcus vannielii, Methanococcus aeolicus* Nankai-3, *Methanococcus voltae* A3, *Methanopyrus kandleri* AV19, *Methanosaeta thermophila, Methanosarcina mazei, Methanosarcina acetivorans, Methanothermobacter thermoautotrophicus, Pyrobaculum calidifontis, Pyrobaculum aerophilum, Pyrobaculum arsenaticum, Pyrobaculum islandicum, Pyrodictium occultum,* uncultured methanogenic archaeon (YP_687422.1), *Staphylothermus marinus, Sulfolobus tokodaii, Sulfophobococcus zilligii, Sulfurisphaera ohwakuensis, Thermodesulfovibrio yellowstonii, Thermofilum pendens* Hrk 5, *Thermoproteus neutrophilus* and uncultured euryarchaeote Alv-FOS1.

Reversible terminators for use in embodiments of the invention contain a protecting group attached to the 3'-OH ribose position that terminates DNA synthesis. Removal of the protecting group restores the unblocked natural nucleotide substrate, allowing subsequent addition of reversible terminators.

Examples of reversible terminators include 3'-O-azidomethyl-2'-deoxynucleoside-5'-triphosphate and 3'-O-(2-nitrobenzyl)-2'-deoxynucleoside-5'-triphosphate (Ruparel et al. *Proc Natl Acad Sci USA* 102(17): 5932-7 (2005); Wu et al. *Proc Natl Acad Sci USA* 104(42): 16462-7 (2007) and Guo et al. *Proc Natl Acad Sci USA* 105(27): 9145-50 (2008)). Other examples of modified nucleotides suitable for incorporation into nucleic acids by archaeal DNA polymerases include: 3'-modified nucleotide analogs in which the 3'-position of the deoxyribose in the nucleotide analogue can be: azidomethyl; O-azidomethyl, azido, sulfhydral, amino, fluorine, chlorine, -O-methyl, O-phosphate, O-diphosphate, aminoallyl, O-aminoallyl, hydrogen (Bi et al. *J Am Chem Soc* 128(8): 2542-3 (2006); Kim et al. Nat Rev Genet 4(12): 1001-8 (2006); Turcatti et al. *Nucleic Acids Res* 36(4): e25 (2008); and Foldesi et al. *Nucleosides Nucleotides Nucleic Acids* 26(3): 271-5 (2007)). See also catalogs of Trilink Biotechnologies (San Diego, CA) and Jena Biosciences (Jena, Germany).

Other examples include, 2'-deoxy-3'-anthranyloyl-dNTPs (3'-ant-dNTPs) 3'-{N3-[3-carboxylato-4-(3-oxido-6-oxo-6H-xanthen-9-yl)phenyl]thioureido}-3 '-deoxythymidine 5'-triphosphate (3'-fluothioureido-dTTP), 3'-deoxy-3'-(N-methylanthranyloylamino)thymidine 5'-triphosphate (3'-amd-dTTP), 3'-O-[N6(N-methylanthranyl)amidohexanoyl]-dGTP (3 -chain-dGTP), and 3'-O-[N6(anthranyl)amidohex (3'-chain-dATP) Canard et al. *Proc Natl Acad Sci USA* 92(24): 10859-63 (1995)). In addition, reversible terminators can be conjugated with dyes including JOE, TAMRA, ROX, FAM, Fluorescein or other moieties for detection (Ju et al. *Proc Natl Acad Sci USA* 103(52): 19635-40 (2006)). In addition, 3'-azido-ddNTPs can be incorporated by a DNA

polymerase and then dye-labeled by "CLICK" chemistry methods (Seo et al. *J Org Chem* 68(2): 609-12 (2003)). Additional dye terminators incorporated by reference are those described in the catalog of PerkinElmer, Waltham, MA (JOE-ddATP; JOE-ddCTP; JOE-ddGTP; JOE-ddUTP; TAMRA-ddATP; TAMRA-ddCTP; TAMRA-ddGTP; TAMRA-ddUTP; FAM-ddATP; FAM-ddCTP; FAM-ddGTP; FAM-ddUTP; ROX-ddATP; ROX-ddCTP; ROX-ddGTP; ROX-ddUTP; Fluorescein-12-ddATP; Fluorescein-12-ddCTP; Fluorescein-12-ddGTP; Fluorescein-12-ddUTP). Other dye terminators include: ROX-acycloNTP; TAMRA-acycloNTP; R6G-ddNTP; R110-ddNTP; Fl-12-acycloNTP; IRD40-ddNTP; IRD700-ddNTP; IRD700-acycloNTP; Cyanine 3-ddNTP; Cyanine 5-ddNTP; Bodipy TR-ddNTP; Bodipy TMR-ddNTP; Bodipy R6G-ddNTP and Bodipy Fl-ddNTP (Gardner and Jack *Nucleic Acids Researc*h 30: 605-613 (2002)).

In order to determine the extent of 3'-modified nucleotide incorporation by archaeon DNA polymerases, the titration assay described by Gardner and Jack (2002) was used. In this assay, the efficiency of incorporation of chain terminators was judged by the size of the reaction products in a polymerization reaction. As the efficiency of chain-terminator incorporation increased, the average reaction product size decreased because polymerization was more often halted by terminator addition. By comparing the amount of terminator required to give the same spectrum of reaction products, the relative efficiency of incorporation of the test compounds with the different polymerases was determined.

Several innovations are exploited in novel combinations in present embodiments of the invention to overcome previously noted limitations in chain terminator incorporation. Modified 3'-ddNTP and modified 3'-O-dNTP terminators were identified that are more

efficiently incorporated (see Example 2). Methods have been described to identify additional compounds of this type (Example 3). Such compounds offer a marked advantage over previously tested ddNTPs whose incorporation was disfavored.

The efficient production of chain terminator products is useful for genotyping and DNA sequence determination. These methods require traditional chain terminator sequencing, and automated procedures where detection is via incorporation of dye-labeled terminators. Furthermore, reversible terminators allow massively parallel sequencing-by-synthesis strategies. The present invention is applicable to both long-range DNA sequence determination where hundreds of base pairs of contiguous sequence are revealed and to short-range sequencing, defined as little as one base pair of sequence. In the case of short-range sequencing, the present invention is useful in analyzing sequence polymorphisms, for example in genetic testing and screening for specific single nucleotide polymorphisms (SNPs).

All references cited herein, including U.S. provisional application serial number 61/046,987 filed April 22, 2008, are hereby incorporated by reference.

EXAMPLES

Example 1: Titration assay for chain terminator incorporation efficiency

Nucleotides: dCTP, ddCTP and acycloCTP were from NEB (Ipswich, MA). 3'-amino-ddCTP and 3'-azido-ddCTP were purchased from TriLink Biotech (San Diego, CA).

In order to select DNA polymerase mutants with increased reversible terminator incorporation efficiency, mutations were created in 9°N DNA polymerase (Southworth et al. *Proc. Natl. Acad. Sci. USA* 93: 5281-5285 (1996)) active site residues residing in Region II and Region III (Figures 1, 5, and 6).

A complex library of 9°N DNA polymerase mutants was screened for enhanced incorporation of 3'-azido-ddCTP. Several classes of 9°N DNA polymerase mutants were identified with enhanced terminator incorporation. 9°N exo- L408S/Y409A/P410V (Therminator III™, NEB, Inc., Ipswich, MA) was purified and characterized in more detail. 9°N DNA polymerase single mutants P410V, L408S, and Y409A and double mutants L408S/P410V, L408S/Y409A, and Y409A/P410V were also purified for comparison. Purification and characterization of DNA polymerase mutants was as described in Gardner and Jack, *Nucleic Acids Res.* 27:2545-2553 (1999).

In order to compare relative modified nucleotide analog incorporation efficiency, DNA polymerase mutants were tested (Figure 1) using a titration assay as described by Gardner and Jack (2002). Although the titration assay for chain terminator incorporation efficiency was originally developed to compare incorporation efficiency of ddNTPs, it was here used to monitor incorporation of 3'-modified nucleotides (Figure 1).

A dye-labeled oligonucleotide primer 5'-CACGACGTTGTAAAACGAC-3' (SEQ ID NO. 20) was annealed to a single-stranded M13mp18 DNA template (Accession No. X02513) and extended by a DNA polymerase in the presence of various

ratios of modified nucleotide:dNTP (10:1, 2.5:1, 1:2.5 or dNTP (no terminator)) in 1 X ThermoPol buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$,10 mM KCl, 2 mM $MgSO_4$, 0.1 % Triton X-100, pH 8.8 at 25°C). Reactions were incubated and termination products were resolved on 20% denaturing polyacrylamide gel electrophoresis (PAGE).

Once the spectrum of termination products was determined, a comparison of the length, uniformity and clarity of these patterns was used to evaluate incorporation of the terminator. Reaction conditions producing shorter products at a given ratio of terminator to normal nucleotides were defined by improved efficiency of incorporation of terminator by the DNA polymerase. Conversely, when comparisons revealed identical banding patterns at different terminator ratios, the lower ratio identified conditions more favorable to terminator incorporation.

The size distribution of termination products was determined by the relative rates of dNTP and terminator incorporation. These competing reactions utilized the same pool of template and continued until replication was halted, either by incorporation of a terminator or by extension to the end of the template. The incorporation efficiency of two terminators was compared using parallel reactions differing only in the type and concentration of terminator. The relative incorporation efficiency of the two terminators was reflected in the concentrations of terminators in the two reactions. For example, if a first reaction contained 10-fold more terminator than a second reaction to generate the same distribution of terminator fragments, then the first terminator was 10-fold less efficient than that of the second.

A library of 9°N DNA polymerase mutants was constructed by PCR amplification of the polymerase genes using gene-specific primers or codon optimization and gene synthesis Czar et al. *Trends Biotechnol* 27(2): 63-72 (2009)). The ability to incorporate 3'-modified nucleotide terminators was evaluated by the titration assay for chain terminator incorporation efficiency as described by Gardner and Jack (2002).

Example 2: Enhanced incorporation of 3'-azido-ddCTP by a 9°N DNA polymerase mutant

A 5'-dye-labeled oligonucleotide primer 5'-CACGACGTTGTAAAACGAC-3' (SEQ ID NO. 20) was annealed to a single-stranded M13mp18 DNA template (Accession No. X02513) and extended by a 9°N exo- DNA polymerase mutant in the presence of various ratios of modified nucleotide:dNTP (10:1, 2.5:1, 1:2.5 or dNTP (no terminator)). Using the titration assay as described by Gardner and Jack (2002), the relative 3'-azido-dCTP incorporation efficiency was determined for a series of 9°N mutations (9°N exo- L408S/Y409A/P410V > 9°N exo- L408S/P410V > 9°N exo- P410V > 9°N exo-) (Figure 1).

Example 3: Kinetic analysis of 3'-azido-ddCTP incorporation by a 9°N DNA polymerase mutant

An IR800-dye-labeled synthetic primer 5'-AGTGAATTCG AGCTCGGTAC CCGGGGATCC TCTAGAGTCG ACCTGCAGGC-3' (SEQ ID NO:18 was annealed to a template 3'-TCACTTAAGC TCGAGCCATG GGCCCCTAGG AGATCTCAGC TGGACGTCCG GATCCTATAC TAATCCC-5' (SEQ ID NO:19) and used as a substrate to measure rates of 3'-azido-ddCTP incorporation over a three-

minute time course. The results are shown in Figure 2. Reaction aliquots were sampled at 10, 20, 30, 45, 60, and 180 seconds and analyzed by 20% denaturing PAGE where "S" indicates the unextended primer and "P" the fully extended product corresponding to 3'-azido-ddCMP addition at the +1 position(B). 9°N exo- L408S/Y409A/P410V incorporation of varying concentrations of 3'-azido-ddCTP was measured over a three-minute time course. 9°N exo- L408S/Y409A/P410V was found to incorporate 3'-azido-ddCTP efficiently with almost 100% incorporation after three minutes.

Furthermore, the same assay was used to measure 3'-amino-ddCTP incorporation by 9°N exo- L408S/Y409A/P410V (Figure 3) and to measure rates of incorporation (Figure 4). Examples of mutants that increased modified nucleotide incorporation by at least two-fold were:

9°N D141A/E143A/P410V
9°N D141A/E143A/L408S/P410V
9°N D141A/E143A/Y409A/S411T
9°N D141A/E143A/L408S/Y409A/P410V
9°N D141A/E143A/L408P/Y409A/S411T
9°N D141A/E143A/Y409A/P410R/S411T
9°N D141A/E143A/L408S/Y409A/P410V/S411T
9°N D141A/E143A/L408P/Y409A/P410V/S411T
9°N D141A/E143A/N491L/Y494S
9°N D141A/E143A/N491V/Y494H
9°N D141A/E143A/R406S/L408R
9°N D141A/E143A/R406L/L408E
9°N D141A/E143A/R406T/L408R
9°N D141A/E143A/R406V/L408R

9°N D141A/E143A/R406T/L408E
9°N D141A/E143A/R406V/L408R
9°N D141A/E143A/R406E/L408G
9°N D141A/E143A/R406P/L408G
9°N D141A/E143A/Y409A/R406V
9°N D141A/E143A/Y409A/R406S/L408K
9°N D141A/E143A/Y409A/R406S/L408R
9°N D141A/E143A/Y409A/R406T/L408K
9°N D141A/E143A/Y409A/R406T/L408R
9°N D141A/E143A/Y409A/R406H/L408G
9°N D141A/E143A/Y409A/R406Y/L408G
9°N D141A/E143A/Y409A/R406L/L408G
9°N D141A/E143A/Y409A/R406P/L408C
9°N D141A/E143A/Y409A/R406S/L408I
9°N D141A/E143A/Y409A/R406V/L408Y
9°N D141A/E143A/Y409A/R406V/A485L
9°N D141A/E143A/N491L/Y494S
9°N D141A/E143A/N491V/Y494H
Mma D153A/E155A/L417S/P419V
Mma D153A/E155A/L417P/Y418A/S420T

Y409A mutation in 9°N, or a corresponding mutation, can be incorporated to increase 3'-modified nucleotide terminator incorporation when combined with additional mutations as described below.

In general, combining the mutations 9°N D141A/E143A/Y409A with a change in R406 to a nucleophilic amino acid (serine, threonine) and L408 to a basic amino acid (arginine or lysine) resulted in an increased 3'-modified nucleotide terminator incorporation.

In general, combining the mutations 9°N D141A/E143A/Y409A and R406 to hydrophobic amino acid (leucine, isoleucine, valine) and L408 to a small amino acid (glycine or alanine) resulted in an increase 3'-modified nucleotide terminator incorporation.

In general, combining the mutations 9°N D141A/E143A/Y409A and R406 to hydrophobic amino acid (leucine, isoleucine, valine) and L408 to a hydrophobic amino acid (leucine, isoleucine, valine) resulted in an increase 3'-modified nucleotide terminator incorporation.

Additionally, enhancements in nucleotide analog incorporation efficiency could be achieved by adding an additional mutation A485L to the mutants described above.

9°N D141A/E143A/P410V/A485L
9°N D141A/E143A/L408S/P410V/A485L
9°N D141A/E143A/Y409A/S411T/A485L
9°N D141A/E143A/L408S/Y409A/P410V/A485L
9°N D141A/E143A/L408P/Y409A/S411T/A485L
9°N D141A/E143A/Y409A/P410R/S411T/A485L
9°N D141A/E143A/L408S/Y409A/P410V/S411T/A485L
9°N D141A/E143A/L408P/Y409A/P410V/S411T/A485L
9°N D141A/E143A/N491L/Y494S/A485L
9°N D141A/E143A/N491V/Y494H/A485L
9°N D141A/E143A/R406S/L408R/A485L
9°N D141A/E143A/R406L/L408E/A485L
9°N D141A/E143A/R406T/L408R/A485L
9°N D141A/E143A/R406V/L408R/A485L

9°N D141A/E143A/R406T/L408E/A485L
9°N D141A/E143A/R406V/L408R/A485L
9°N D141A/E143A/R406E/L408G/A485L
9°N D141A/E143A/R406P/L408G/A485L
9°N D141A/E143A/Y409A/R406S/L408K/A485L
9°N D141A/E143A/Y409A/R406S/L408R/A485L
9°N D141A/E143A/Y409A/R406T/L408K/A485L
9°N D141A/E143A/Y409A/R406T/L408R/A485L
9°N D141A/E143A/Y409A/R406H/L408G/A485L
9°N D141A/E143A/Y409A/R406Y/L408G/A485L
9°N D141A/E143A/Y409A/R406L/L408G/A485L
9°N D141A/E143A/Y409A/R406P/L408C/A485L
9°N D141A/E143A/Y409A/R406S/L408I/A485L
9°N D141A/E143A/Y409A/R406V/L408Y/A485L

Example 4: Incorporation of modified nucleotide, 3'-O-azidomethyl-dCTP

Incorporation of 3'-O-azidomethyl-dCTP by various DNA polymerases was measured using the titration assay as described in Example 1. For each reaction, a ratio of either 10:1 or 1:1 3'-O-azidomethyl-dCTP:dNTP was used to generate a termination pattern. Each band corresponded to a DNA fragment terminated by a 3'-azido-dCMP. A control reaction lacking terminator (dNTP) which yielded large extension products gave assurance that the DNA polymerase was active, and that termination products resulted from terminator incorporation rather than incomplete polymerization. Reactions with Therminator™ (NEB, Inc., Ipswich, MA) and 9°N exo-L408P/Y409A/S411T DNA polymerase were incubated at 72°C for 30 minutes. Reactions with *E. coli* Polymerase I (pol I) (NEB, Inc., Ipswich, MA), Sequenase™ (USB, Inc., Cleveland, OH), Mma exo-

and Mma exo- L417P/Y418A/S420T were conducted at 37°C for 30 mintues. Therminator™ (NEB, Inc., Ipswich, MA) DNA polymerase was assayed with a 1:1 ratio of acyCTP as control.

E. coli DNA polymerase I (pol I) (NEB, Inc., Ipswich, MA) and Sequenase™ (USB, Inc., Cleveland, OH) discriminated against 3'-O-azidomethyl-dCTP and failed to terminate synthesis. 9°N exo- L408P/Y409A/S411T incorporated 3'-O-azidomethyl-dCTP efficiently and generated a termination pattern with 10:1 or 1:1 3'-O-azidomethyl-dCTP:dNTP. *Methanococcus maripaludis* DNA polymerase exo- (Mma exo-) discriminated against 3'-O-azidomethyl-dCTP and failed to terminate synthesis. Mma exo- L417P/Y418A/S420T incorporated 3'-O-azidomethyl-dCTP efficiently and generated a termination pattern with 10:1 or 1:1 3'-O-azidomethyl-dCTP:dNTP.

The equivalent mutations in 9°N exo- (L408P/Y409A/S411T) and Mma exo- (L417P/Y418A/S420T) resulted in increased 3'-O-azidomethyl-dCTP incorporation suggesting functional conservation despite differences in optimum temperature.

Example 5. Determining specific activities of DNA polymerases

9°N D141A/E143A/L408S/Y409A/P410V and 9°N D141A/E143A/Y409V/A485L DNA polymerases were purified by a method described by Gardner and Jack (*Nucleic Acids Res* 27(12): 2545-53 (1999)). Specific activities for 9°N D141A/E143A/L408S/Y409A/P410V and 9°N D141A/E143A/Y409V/A485L DNA polymerase were determined by measuring the DNA polymerase activity as well as the protein concentration. Briefly, a primer (5'-

CGCCAGGGTTTTCCCAGTCACGAC-3') (SEQ ID NO:21) was annealed to single-stranded M13mp18 (Accession Number: X02513) in 1X Thermopol Buffer (20 mM Tris-HCl, 10 mM (NH4)2SO4, 10 mM KCl, 2 mM MgSO4, 0.1 % Triton X-100, pH 8.8 at 25°C). 9°N D141A/E143A/L408S/Y409A/P410V and 9°N D141A/E143A/Y409V/A485L DNA polymerase activity was measured using a primed M13 substrate as described in Kong, et al. (*J. Biol. Chem*. 268:1965-1975 (1993)). DNA polymerase activity was converted to units (one unit was the amount of enzyme that incorporated 10 nmol of dNTP into acid-insoluble material 30 minutes at 75°C.) 9°N D141A/E143A/L408S/Y409A/P410V and 9°N D141A/E143A/Y409V/A485L DNA polymerase protein concentration was determined as described in Bradford *Anal Biochem* 72: 248-54 ((1976). Specific activity of the DNA polymerase is defined by units/mg protein where a unit is the amount of enzyme that will incorporate 10 nM of dNTP into acid insoluble material.

Claims

1. A recombinant protein with DNA polymerase activity, comprising:
an amino acid sequence that has at least 90% amino acid sequence identity with SEQ ID NO:1 wherein one or more amino acids in the recombinant protein are mutated compared with the corresponding wild type protein such that the recombinant protein is capable of (i) incorporating one or more nucleotides into a nucleic acid substrate with a specific activity greater than 200, and (ii) incorporating modified nucleotides with at least two fold greater efficiency than a corresponding wild type DNA polymerase.

2. The recombinant protein according to claim 1, wherein the specific activity is greater than 1000.

3. The recombinant protein according to claim 1, wherein the specific activity is greater than 5000.

4. The recombinant protein according to claim 1, wherein at least one mutation is located in SEQ ID NO:1.

5. The recombinant protein according to claim 1, wherein at least one mutation is located in a conserved region identified as Region III.

6. The recombinant protein according to claim 1, wherein at least one mutation is located in the amino acid sequence outside of SEQ ID NO:1.

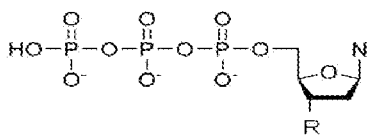
7. The recombinant protein according to claim 1, wherein the composition is a 9°N archael polymerase, and the mutated amino acids comprise D141A and E143A and an additional mutation selected from the group consisting of: P410V; S411T; L408S/Y409A/P410V; L408P/Y409A/S411T; P410R/S411T; L408S/Y409A/P410V/S411T; L408P/Y409A/P410V/S411T; N491L/Y494S; N491V/Y494H; R406S/L408R; R406L/L408E; R406T/L408R; R406V/L408R; R406T/L408E; R406V/L408R; R406E/L408G; R406P/L408G; Y409A/R406V; Y409A/R406S/L408K; Y409A/R406S/L408R; Y409A/R406T/L408K; Y409A/R406T/L408R; Y409A/R406H/L408G; Y409A/R406Y/L408G; Y409A/R406L/L408G; Y409A/R406P/L408C; Y409A/R406S/L408I; Y409A/R406V/L408Y; Y409A/R406V/A485L; N491L/Y494S; N491V/Y494H; Y409A/R406(nucleophilic amino acid)/L408(basic amino acid); Y409A/R406(hydrophobic amino acid)/L408(small amino acid); Y409A/R406/L408(hydrophobic amino acid)/L408(small amino acid); P410V/A485L; L408S/P410V/A485L; Y409A/S411T/A485L; L408S/Y409A/P410V/A485L; L408P/Y409A/S411T/A485L; Y409A/P410R/S411T/A485L; L408S/Y409A/P410V/S411T/A485L; L408P/Y409A/P410V/S411T/A485L; N491L/Y494S/A485L; N491V/Y494H/A485L; R406S/L408R/A485L; R406L/L408E/A485L; R406T/L408R/A485L; R406V/L408R/A485L; R406T/L408E/A485L; R406V/L408R/A485L; R406E/L408G/A485L; R406P/L408G/A485L; Y409A/R406S/L408K/A485L; Y409A/R406S/L408R/A485L; Y409A/R406T/L408K/A485L; Y409A/R406T/L408R/A485L; Y409A/R406H/L408G/A485L; Y409A/R406Y/L408G/A485L; Y409A/R406L/L408G/A485L;

Y409A/R406P/L408C/A485L; Y409A/R406S/L408I/A485L; Y409A/R406V/L408Y/A485L.

8. The recombinant protein according to claim 1, wherein the composition is a *Methanococcus maripaludis* (Mma) archaeal polymerase, and the mutated amino acids are selected from D153A/E155A/L417S/P419V and D153A/E155A/L417P/Y418A/S420T.

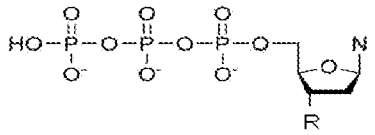
9. The recombinant protein according to claim 1, wherein the one or more modified nucleotides are selected from 3' terminators and 3' reversible terminators.

10. The recombinant protein according to claim 1, wherein N is a nucleoside and the R group on the 3' position of the ribose is substituted by one of the following:



R= -H, -SH, $-N_3$, -F, -Cl, -azidomethyl, $-NH_2$, -anthranyloyl -fluothioureido, -chain, -amd, -O-allyl, -O-aminoallyl, -O-azidomethyl ,-O-methyl, -O-phophate, -O-diphophate, -O-(2-nitrobenzyl), -O-[N6(anthranyl)amidohex

11. The recombinant protein according to claim 1, wherein N is a nucleoside and the R group on the 3' position of the ribose is substituted by one of the following:



wherein R is larger than a hydroxyl group.

12. The recombinant protein according to claim 10 or 11 wherein R may further coprise a marker.

13. The recombinant protein according to claim 12 wherein the marker is a fluorescent label.

14. The recombinant protein according to claim 1, wherein the modified nucleotides are selected from the group consisting of: 2'-deoxy-3'-anthranyloyl-dNTPs (3'-ant-dNTPs) 3'-{N3-[3-carboxylato-4-(3-oxido- 6-oxo-6H-xanthen-9-yl)phenyl]thioureido}-3 '-deoxythymidine 5'-triphosphate (3'-fluothioureido-dTTP), 3'-deoxy-3'-(N-methylanthranyloylamino)thymidine 5'-triphosphate (3'-amd-dTTP), 3'-O-[N6(N-methylanthranyl)amidohexanoyl]-dGTP (3 -chain-dGTP), and 3'-O-[N6(anthranyl)amidohex (3'-chain-dATP).

15. A method of incorporating modified nucleotides into a nucleic acid, comprising: reacting a nucleic acid with the recombinant protein according to claim 1 and at least one modified nucleotide.

16. A kit comprising the recombinant protein according to claim 1 and instruction for use.

17. A kit according to claim 16, further comprising at least one modified nucleotide.

18. A method of screening for the recombinant protein according to claim 1, comprising:

(a) determining a size of a substrate incorporating a modified nucleotide after a polymerization reaction; and

(b) measuring at least one of an increase in chain-terminator incorporation, and a decrease in average reaction product size, to determine efficiency of incorporation by the composition.

FIG. 1

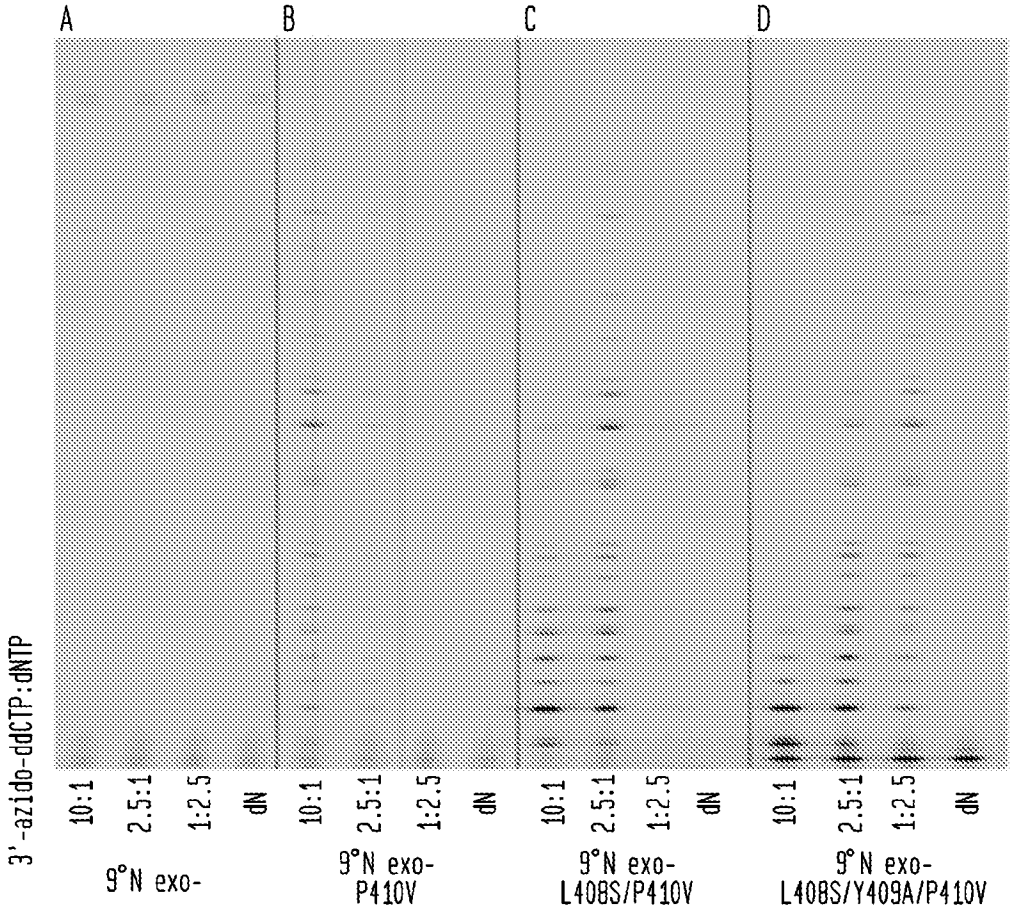

A
B
C
D
3'-azido-ddCTP:dNTP
10:1
2.5:1
1:2.5
dN
9°N exo-
9°N exo-
P410V
9°N exo-
L408S/P410V
9°N exo-
L408S/Y409A/P410V

FIG. 2

[3'-azido-ddCTP]

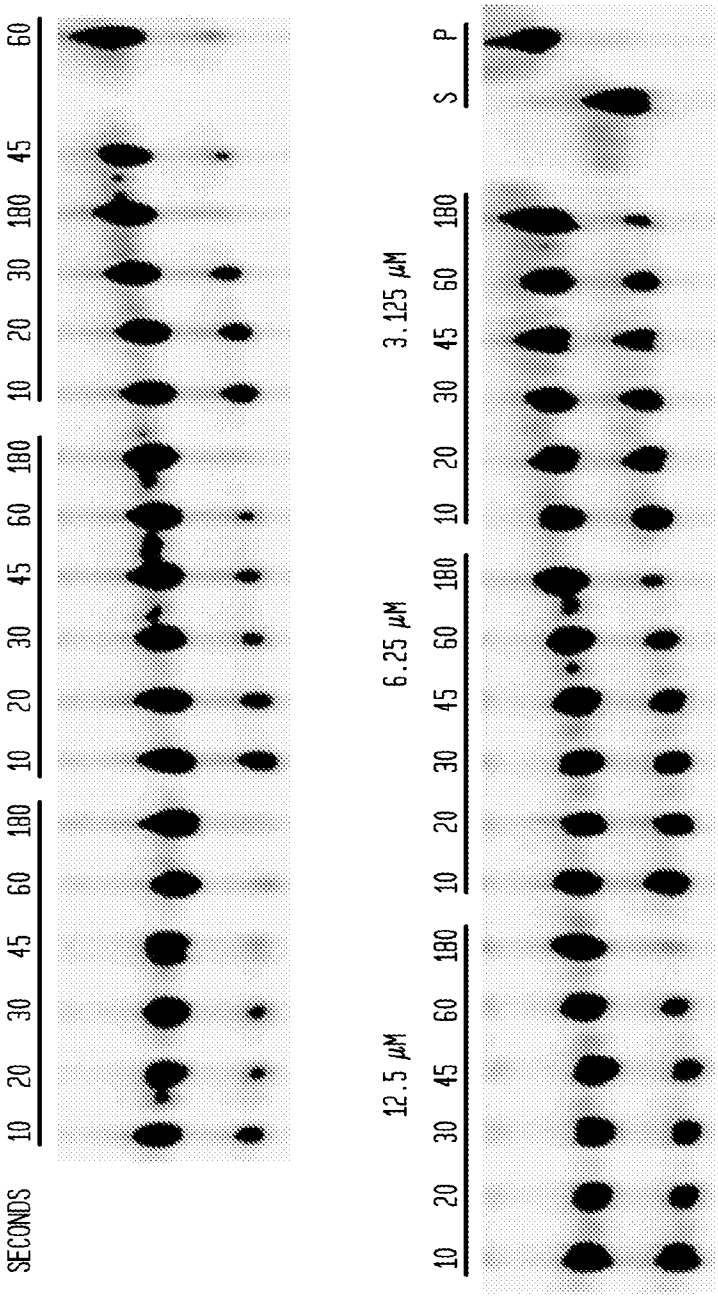


FIG. 3

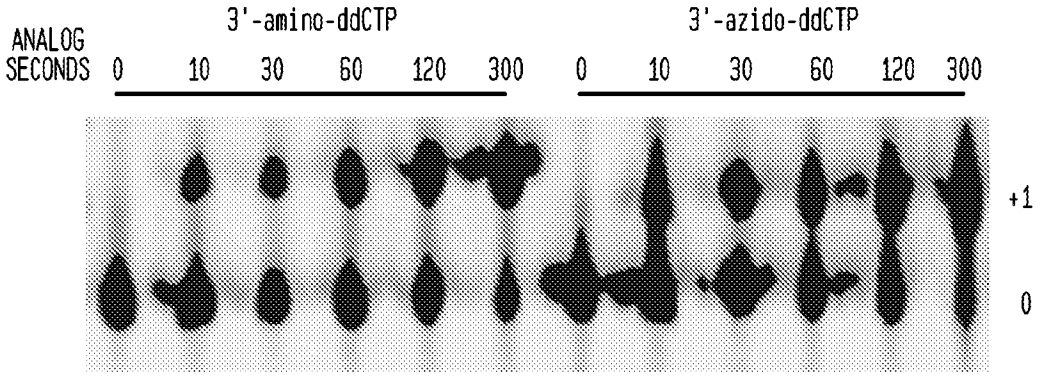

3'-amino-ddCTP
3'-azido-ddCTP
ANALOG
SECONDS
0
10
30
60
120
300
0
10
30
60
120
300
+1
0

FIG. 4

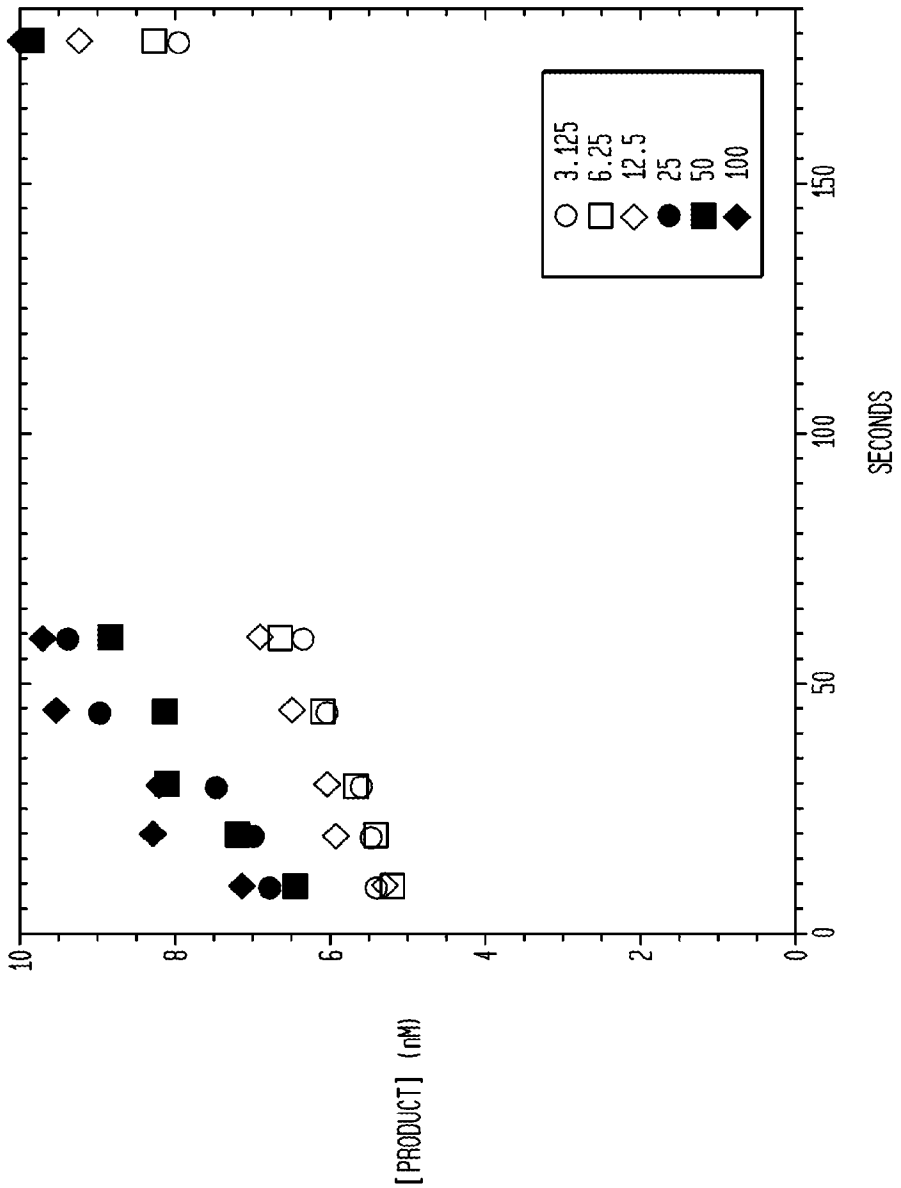


FIG. 5

Accession Number	Organism	Expect (e)	Identity	Region II sequence
Q56366.1	Thermococcus sp. 9°N	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
EEB74380.1	Thermococcus sp. AM4	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
ABK59374.1	Thermococcus thioreducens	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
CAA90888.1	Pyrococcus abyssi	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
CAB81809.1	Pyrococcus glycovorans	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
CAC12847.1	Pyrococcus sp. ST700	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
YP_002306381.1	Thermococcus onnurineus NA1	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
Q9HH84.1	Thermococcus sp. GE8	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
Q51334.1	Deep vent DNA polymerase	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
NP_143776.1	Pyrococcus horikoshii OT3	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
1WNS	Pyrococcus kodakaraensis Kod1	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
CAA90887.1	Pyrococcus sp.	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
NP_127396.1	Pyrococcus abyssi GE5	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
P56689.1	Thermococcus Gorgonarius	5e-06	100%	DFRSLYPSIITH (SEQ ID NO:1)
NP_577941.1	Pyrococcus furiosus DSM 3638	3e-05	92%	DFRALYPSIITH (SEQ ID NO:2)
P30317.1	Thermococcus literalis	3e-05	92%	DFRSLYPSIVTH (SEQ ID NO:3)
ABD14869.1	Thermococcus sp. GT	2e-04	92%	DFMSLYPSIITH (SEQ ID NO:4)
Q9HH05.1	Thermococcus hydrothermalis	2e-04	92%	DFMSLYPSIITH (SEQ ID NO:4)

FIG. 6-1

Abbreviation	Organism
9N	Thermococcus sp. 9°N
9N DNAPM	Thermococcus sp. 9°N
TGO	Thermococcus gorgonarius
KOD	Thermococcus kodakarensis
P_horikoshii	Pyrococcus horikoshii
T_aggregans	Thermococcus aggregans
Vent_T.litoralis	Thermococcus litoralis
Mma	Methanococcus maripaludis
M_jannaschii	Methanococcus jannaschii
Mth_PolB1	Methanothermobacter thermautotrophicus str. Delta H
RB69	bacteriophage RB69
T4	bacteriophage T4
M_marisigri	Methanoculleus marisnigri JR1

FIG. 6-2

```
9N                  ----------MILDTDYITENGKPVIRVFKKENG---EFKIEYDRTFEPYFYALLKDDS 46
9N DNAP M           ----------MILDTDYITENGKPVIRVFKKENG---EFKIEYDRTFEPYFYALLKDDS 46
TGO                 ----------MILDTDYITEDGKPVIRIFKKENG---EFKIDYDRNFEPYIYALLKDDS 46
KOD                 ----------MILDTDYITEDGKPVIRIFKKENG---EFKIEYDRTFEPYFYALLKDDS 46
P_horikoshii        ----------MILDADYITEDGKPIIRIFKKENG---EFKVEYDRNFRPYIYALLRDDS 46
T_aggregans         ----------MILDTDYITKDGKPIIRIFKKENG---EFKIELDPHFQPYIYALLKDDS 46
Vent_T.litoralis    ----------MILDTDYITKDGKPIIRIFKKENG---EFKIELDPHFQPYIYALLKDDS 46
Mma                 ---------MESLIDLDYN--SDDLCIYLYLINS-------IIKEKDFKPYFYVNSTDKE 42
M_jannaschii        MGMSMGKIKIDALIDNTYKTIEDKAVIYLYLINS-------ILKDRDFKPYFYVELH-KE 52
Mth_PolB1           -----MEDYRMVLLDIDYVTVDEVPVIRLFGKDKSGGNEPIIAHDRSFRPYIYAIPTDLD 55
RB69                ----------MKEFYLTVEQIGDSIFERYIDSNG----RERTREVEYKPSLFAHCPESQ 45
T4                  ----------MKEFYISIETVGNNIVERYIDENG----KERTREVEYLPTMFRHCKE-- 43
M_marisigri         ----------MSVPATLEDFGKIRVGIHQVEYG--------NVGADTPVVYIFGRDAS 40
                                                  .  .  .                 * .:

9N                  AIEDVKKVTAKRH---GTVVKVKRAEKVQKKFLGRP----IEVWKLYFNHPQDVPAIRDR 99
9N DNAP M           AIEDVKKVTAKRH---GTVVKVKRAEKVQKKFLGRP----IEVWKLYFNHPQDVPAIRDR 99
TGO                 AIEDVKKITAERH---GTTVRVVRAEKVKKKFLGRP----IEVWKLYFTHPQDVPAIRDK 99
KOD                 AIEEVKKITAERH---GTVVTVKRVEKVQKKFLGRP----VEVWKLYFTHPQDVPAIRDK 99
P_horikoshii        AIDEIKKITAQRH---GKVVRIVETEKIQRKFLGRP----IEVWKLYLEHPQDVPAIRDK 99
T_aggregans         AIDEIKAIKGERH---GKIVRVVDAVKVKKKFLGRD----VEVWKLIFEHPQDVPALRGK 99
Vent_T.litoralis    AIEEIKAIKGERH---GKTVRVLDAVKVRKKFLGRE----VEVWKLIFEHPQDVPAMRGK 99
Mma                 QILEFLKDYEKKHKLDSEISKMIENIETVKKIVFDENYQEKELSKVTVKYPNNVKTVR-E 101
```

FIG. 6-3

```
M_jannaschii        KVENEDIEKIKEFLLKNDLLKFVENIEVVKKIILRK---EKEVIKIIATHPQKVPKLR-K 108
Mth_PolB1           ECLRELEELELEK---------LEVKEMRD--LGRP----TEVIRIEFRHPQDVPKIRDR 100
RB69                ATKYFDIYGKPCTR--KLFANMRDASQWIKRMEDIG---LEALGMDDFKLAYLSDTYNYE 100
T4                  ESKYKDIYGKNCAP--QKFPSMKDARDWMKRMEDIG---LEALGMNDFKLAYISDTYGSE 98
M_marisigri         GKAIQVRVSGFRP-YFYAPADLVDGRSLPQEIVGVEENTTYRSIQGVPLRRLYTRRPGDV 99

9N                  IRAHPAVVDIYEYDIPFAKRYLIDKGLIP-----MEGDE---------------------- 133
9N DNAP M           IRAHPAVVDIYEYDIPFAKRYLIDKGLIP-----MEGDE---------------------- 133
TGO                 IKEHPAVVDIYEYDIPFAKRYLIDKGLIP-----MEGDE---------------------- 133
KOD                 IREHGAVIDIYEYDIPFAKRYLIDKGLVP-----MEGDE---------------------- 133
P_horikoshii        IREHPAVVDIFEYDIPFAKRYLIDKGLTP-----MEGNE---------------------- 133
T_aggregans         IREHPAVIDIYEYDIPFAKRYLIDKGLIP-----MEGDE---------------------- 133
Vent_T.litoralis    IREHPAVVDIYEYDIPFAKRYLIDKGLIP-----MEGDE---------------------- 133
Mma                 I--LMEFERLYEYDIPFVRRYLIDNSVIPTSTWDFENNKKIDN------------------ 142
M_jannaschii        IKECEIVKEIYEHDIPFAKRYLIDNEIIPMTYWDFENKKPVSI------------------ 151
Mth_PolB1           IRDLESVRDIREHDIPFYRRYLIDKSIVPMEELEFQGVEVDSAPSVTTDVRTVEVTGRVQ 160
RB69                IKYDHTKIRVANFDIEVTSPDGFPEPSQAKHPIDAITHYD---------------------- 140
T4                  IVYDRKFVRVANCDIEVTG-DKFPDPMKAEYEIDAITHYD---------------------- 137
M_marisigri         RAVRDVFSHHYEADIPFTTRFMIDCGLTAGVELPAGAVESFDG--------AFEIECCEL 151

9N                  -------ELTMLAFDIETLYHEGEEFGT-GPILMISYA-------------DGSEARVI 171
9N DNAP M           -------ELTMLAFAIATLYHEGEEFGT-GPILMISYA-------------DGSEARVI 171
TGO                 -------ELKMLAFDIETLYHEGEEFAE-GPILMISYA-------------DEEGARVI 171
KOD                 -------ELKMLAFDIQTLYHEGEEFAE-GPILMISYA-------------DEEGARVI 171
```

FIG. 6-4

```
P_horikoshii        -------KLTFLAVDIETLYHEGEEFGK-GPVIMISYA-------------DEEGAKVI 171
T_aggregans         -------ELKLMAFDIETFYHEGDEFGK-GEIIMISYA-------------DEEEARVI 171
Vent_T.litoralis    -------ELKLLAFDIETFYHEGDEFGK-GEIIMISYA-------------DEEEARVI 171
Mma                 ----KIPDFKTVSFDIEVY-CNKEPNPKKDPIIMASFS-------------SKDFNTVV 183
M_jannaschii        ----EIPKLKSVAFDMEVYNRDTEPNPERDPILMASFW-------------DENGGKVI 193
Mth_PolB1           STGSGAHGLDILSFDIEVRNPHGMPDPEKDEIVMIGVAG-------------NMGYESVI 207
RB69                -------SIDDRFYVFDLLNSPYGNVEEWSIEIAAKLQ--------------EQGGDEV 178
T4                  -------SIDDRFYVFDLLNSMYGSVSKWDAKLAAKLD--------------CEGGDEV 175
M_marisigri         APAEIKAPARTCIMDIECVDEQGFPEPERDPIICVTCWDSFDDDYTTLLWQPGEAAGDAP 211
                                     :            .  :

9N                  TWKKIDLPYVDVVSTEKEMIKRFLRVVREKDPDVLITYNGDNFDFAYLKKRCEELGIKFT 231
9N DNAP M           TWKKIDLPYVDVVSTEKEMIKRFLRVVREKDPDVLITYNGDNFDFAYLKKRCEELGIKFT 231
TGO                 TWKNIDLPYVDVVSTEKEMIKRFLKVVKEKDPDVLITYNGDNFDFAYLKKRSEKLGVKFI 231
KOD                 TWKNVDLPYVDVVSTEREMIKRFLRVVKEKDPDVLITYNGDNFDFAYLKKRCEKLGINFA 231
P_horikoshii        TWKKIDLPYVEVVSSEREMIKRLIRVIKEKDPDVIITYNGDNFDFPYLLKRAEKLGIKLL 231
T_aggregans         TWKNIDLPYVDVVSNEREMIKRFVQIVREKDPDVLITYNGDNFDLPYLIKRAEKLGVTLL 231
Vent_T.litoralis    TWKNIDLPYVDVVSNEREMIKRFVQVVKEKDPDVIITYNGDNFDLPYLIKRAEKLGVRLV 231
Mma                 STKKFDHEKLEYVKDEKELIKRIIEILKEYD--IIYTYNGDNFDFPYLKKRAESFGLELK 241
M_jannaschii        TYKEFNHPNIEVVKNEKELIKKIIETLKEYD--VIYTYNGDNFDFPYLKARAKIYGIDIN 251
Mth_PolB1           STAGDHLDFVEVVEDERELLERFAEIVIDKKPDILVGYNSDNFDFPYITRRAAILGAELD 267
RB69                PSEIIDKIIYMPFDNEKELLMEYLNFWQQKTPVILTGWNVESFDIPYVYNRIKNIFGEST 238
T4                  PQEILDRVIYMPFDNERDMLMEYINLWEQKRPAIFTGWNIEGFDVPYIMNRVKMILGERS 235
M_marisigri         DLCVQERHRVVRYPDEIAMLKGLVDYVKKRDPDILSGWNFVEFDIPYIVKRMGALGLK-- 269
                         .        * ::       .   :: :*  **..*: *
```

FIG. 6-5

```
9N                  LGRD------GSEPKIQRMGDRFAVEVKGRIHFDLYPVIRRT--INLPTYTLEAVYEAVF 283
9N DNAP M           LGRD------GSEPKIQRMGDRFAVEVKGRIHFDLYPVIRRT--INLPTYTLEAVYEAVF 283
TGO                 LGRE------GSEPKIQRMGDRFAVEVKGRIHFDLYPVIRRT--INLPTYTLEAVYEAIF 283
KOD                 LGRD------GSEPKIQRMGDRFAVEVKGRIHFDLYPVIRRT--INLPTYTLEAVYEAVF 283
P_horikoshii        LGRD------NSEPKMQKMGDSLAVEIKGRIHFDLFPVIRRT--INLPTYTLEAVYEAIF 283
T_aggregans         LGRDK----EHPEPKIHRMGDSFAVEIKGRIHFDLFPVVRRT--INLPTYTLEAVYEAVL 285
Vent_T.litoralis    LGRDK----EHPEPKIQRMGDSFAVEIKGRIHFDLFPVVRRT--INLPTYTLEAVYEAVL 285
Mma                 LGKN------DEKIKITKGGMNSKSYIPGRVHIDLYPIARRL--LNLTKYRLENVTEALF 293
M_jannaschii        LGKD------GEELKIKRGGMEYRSYIPGRVHIDLYPISRRL--LKLTKYTLEDVVYNLF 303
Mth_PolB1           LGWD------GSKIRTMRRGFANATAIKGTVHVDLYPVMRRY--MNLDRYTLERVYQELF 319
RB69                AKRLSPHRKTRVKVIENMYGSREIITLFGISVLDYIDLYKKFSFTNQPSYSLDYISEFEL 298
T4                  MKRFSPIGRVKSKLIQNMYGSKEIYSIDGVSILDYLDLYKKFAFTNLPSFSLESVAQHET 295
M_marisigri         ----------AEDLARIPGQTERNAVRGRSIFDLLGAYRKMHQAQKESYRLDAIAGEEL 318
                              :     *     : *  .*    ::   :  : *: :

9N                  GKPKEKVYAEEIAQAWESGEGLERVARYSMEDAKVTYELGREFFPMEAQLSRLIGQSLWD 343
9N DNAP M           GKPKEKVYAEEIAQAWESGEGLERVARYSMEDAKVTYELGREFFPMEAQLSRLIGQSLWD 343
TGO                 GQPKEKVYAEEIAQAWETGEGLERVARYSMEDAKVTYELGKEFFPMEAQLSRLVGQSLWD 343
KOD                 GQPKEKVYAEEITPAWETGENLERVARYSMEDAKVTYELGKEFLPMEAQLSRLIGQSLWD 343
P_horikoshii        GKPKEKVYADEIAKAWETGEGLERVAKYSMEDAKVTYELGREFFPMEAQLARLVGQPVWD 343
T_aggregans         GKTKSKLGAEEIAAIWETEESMKKLAQYSMEDARATYELGKEFFPMEAELAKLIGQSVWD 345
Vent_T.litoralis    GKTKSKLGAEEIAAIWETEESMKKLAQYSMEDARATYELGKEFFPMEAELAKLIGQSVWD 345
Mma                 DVKKVDVGHENIPKMWDNLD--ETLVEYSHQDAYYTQRIGEQFLPLEIMFSRVVNQSLYD 351
M_jannaschii        GIEKLKIPHTKIVDYWANND--KTLIEYSLQDAKYTYKIGKYFFPLEVMFSRIVNQTPFE 361
Mth_PolB1           GEEKIDLPGDRLWEYWDRDELRDELFRYSLDDVVATHRIAEKILPLNLELTRLVGQPLFD 379
RB69                NVGKLKYDG-PISKLRESNH--QRYISYNIIDVYRVLQIDAKRQFINLSLDMGYYAKIQI 355
```

FIG. 6-6

```
T4                  KKGKLPYDG-PINKLRETNH--QRYISYNIIDVESVQAIDKIRGFIDLVLSMSYYAKMPF 352
M_marisigri         GVTKVRYTG-TITDLWRTDP--KRLVEYNYRDVELCVGIDQKNNIIEFYREIARYVGCPL 375
                       *      :          .   *. *.    :     ::

9N                  VSRSSTGNLVEWFLLRKAYKRNELAPNKPDERELARR-RGGYAGGYVKEPERGLWDNIVY 402
9N DNAP M           VSRSSTGNLVEWFLLRKAYKRNELAPNKPDERELARR-RGGYAGGYVKEPERGLWDNIVY 402
TGO                 VSRSSTGNLVEWFLLRKAYERNELAPNKPDERELARR-RESYAGGYVKEPERGLWENIVY 402
KOD                 VSRSSTGNLVEWFLLRKAYERNELAPNKPDEKELARR-RQSYEGGYVKEPERGLWENIVY 402
P_horikoshii        VSRSSTGNLVEWFLLRKAYERNELAPNKPDEKEYERRLRESYEGGYVKEPEKGLWEGIVS 403
T_aggregans         VSRSSTGNLVEWYLLRVAYERNELAPNKPDEEEYRRRLRTTYLGGYVKEPERGLWENIAY 405
Vent_T.litoralis    VSRSSTGNLVEWYLLRVAYARNELAPNKPDEEEYKRRLRTTYLGGYVKEPEKGLWENIIY 405
Mma                 INRMSSSQMVEYLLLKNSYKMGVIAPNRPSGKEYQKRIRSSYEGGYVKEPLKGIHEDIVS 411
M_jannaschii        ITRMSSGQMVEYLLMKRAFKXNMIVPNKPDEEEYRRRVLTTYEGGYVKEPEKGMFEDIIS 421
Mth_PolB1           ISRMATGQQAEWFLVRKAYQYGELVPNKPSQSDFSSRRGRRAVGGYVKEPEKGLHENIVQ 439
RB69                QSVFSPIKTWDAIIFNSLKEQNKVIPQG------RSHPVQPYPGAFVKEPIPNRYKYVMS 409
T4                  SGVMSPIKTWDAIIFNSLKGEHKVIPQQ------GSHVKQSFPGAFVFEPKPIARRYIMS 406
M_marisigri         DRTLNSSNVIDIFVLRKASG-TFVLPSK------GLAAGDEFEGATVFEPATGLRENVVV 428
                        . :  :  :..      : *.              *. * **      :
                       REGION II
9N                  LDFRSLYPSIIITHNVSPDTLN----------REGCKEYD---VAPEVGHKFCKDFPG 447
9N DNAP M           LDFRSPAPTIIITHNVSPDTLN----------REGCKEYD---VAPEVGHKFCKDFPG 447
TGO                 LDFRSLYPSIIITHNVSPDTLN----------REGCEEYD---VAPQVGHKFCKDFPG 447
KOD                 LDFRSLYPSIIITHNVSPDTLN----------REGCKEYD---VAPQVGHRFCKDFPG 447
P_horikoshii        LDFRSLYPSIIITHNVSPDTLN----------REGCEEYD---VAPKVGHRFCKDFPG 448
T_aggregans         LDF-SLYPSIIVTHNVSPDTLE----------REGCKNYD---VAPIVGYKFCKDFPG 449
```

FIG. 6-7

```
Vent_T.litoralis    LDFRSLYPSIIVTHNVSPDTLE----------KEGCKNYD---VAPIVGYRFCKDFPG 450
Mma                 MDFLSLYPSIIMSHNLSPETID----------CTCCSDEENGENEEILGHKFCKKSIG 459
M_jannaschii        MDFRSLYPSIIISYNISPDTLD----------CECCKDVS----EKILGHWFCKKKEG 465
Mth_PolB1           FDFRSLYPSIIISKNISPDTLT----------DDEESECY---VAPEYGYRFRKSPRG 484
RB69                FDLTSLYPSIIRQVNISPETIAGTFKVAPLHDYINAVAERPSDVYSCSPNGMMYYKDRDG 469
T4                  FDLTSLYPSIIRQVNISPETIRGQFKVHPIHEYIAGTAPKPSDEYSCSPNGWMYDKHQEG 466
M_marisigri         LDLKSLYPMAMMTINASPETKN----------------PDGE---LRAPNGIRFSREPDG 469
                    :*: *  *  :   * **:*                          *  : :   *

9N                  FIPSLLGDLLEERQKIKRKMKATVD---------------------------------- 472
9N DNAP M           FIPSLLGDLLEERQKIKRKMKATVD---------------------------------- 472
TGO                 FIPSLLGDLLEERQKVKKKMKATID---------------------------------- 472
KOD                 FIPSLLGDLLEERQKIKKKMKATID---------------------------------- 472
P_horikoshii        FIPSLLGQLLEERQKIKKRMKESKD---------------------------------- 473
T_aggregans         FIPSILGELITMRQEIKKKMKATID---------------------------------- 474
Vent_T.litoralis    FIPSILGDLIAMRQDIKKKMKSTID---------------------------------- 475
Mma                 IIPKTLMDLINRRKKVKKVLREKAEKG-------------------------------- 486
M_jannaschii        LIPKTLRNLIERRINIKRRMKKMAEIG-------------------------------- 492
Mth_PolB1           FVPSVIGEILSERVRIKEEMKGSDD---------------------------------- 509
RB69                VVPTEITKVFNQRKEHKGYMLAAQRNGEIIKEALHNPNLSVDEPLDVDYRFDFSDEIKEK 529
T4                  IIPKEIAKVFFQRKDWKKKMFAEEMNAEAIKKIIMKGAGSCSTKPEVERYVKFSDDFLNE 526
M_marisigri         LTRSIIAELLEERDERKRLRNLYPFG---------------------------------- 495
                    .  . : .:: *  *

                                                REGION III
9N                  ----------------PLEKKLLDYRQRAIKILANSFYGYYGYAKARWYCKECAESVTAW 516
```

FIG. 6-8

```
9N DNAP M          ----------PLEKKLLDYRQRAIKILANSFYGYYGYAKARWYCKECAESVTAW 516
TGO                ----------PIEKKLLDYRQRAIKILANSFYGYYGYAKARWYCKECAESVTAW 516
KOD                ----------PIERKLLDYRQRAIKILANSYYGYYGYARARWYCKECAESVTAW 516
P_horikoshii       ----------PVEKKLLDYRQRAIKILANSYYGYYGYAKARWYCKECAESVTAW 517
T_aggregans        ----------PIEKKMLDYRQRAVKLLANSYYGYMGYPKARWYSKECAESVTAW 518
Vent_T.litoralis   ----------PIEKKMLDYRQRAIKLLANSYYGYMGYPKARWYSKECAESVTAW 519
Mma                ----------EFDEEYQILDYEQRSIKVLANSHYGYLAFPMARWYSRDCAEITTHL 532
M_jannaschii       ----------EINEEYNLLDYEQKSLKILANSVYGYLAFPRARFYSRECAEIVTYL 538
Mth_PolB1          ----------PMERKILNVQQEALKRLANTMYGVYGYSRFRWYSMECAEAITAW 553
RB69               IKKLSAKSLNEMLFRAQRTEVAGMTAQINRKLLINSLYGALGNVWFRYYDLRNATAITTF 589
T4                 LSNYTESVLNSLIEECEKAATLANTNQLNRKILINSLYGALGNIHFRYYDLRNATAITIF 586
M_marisigri        ----------SPEYVLYDLQQNVLKVIMNSYYGVSGYTRFRLYDREIGSAVTSV 539
                                      *  * : *: **  .   * *   .  *

9N                 GREYIEMVIRELEEKF--------GFKVLYADTDGLHATIPG----ADAETVKKK----- 559
9N DNAP M          GREYIEMVIRELEEKF--------GFKVLYADTDGLHATIPG----ADAETVKKK----- 559
TGO                GRQYIETTIREIEEKF---------GFKVLYADTDGFFATIPG-----ADAETVKKK----- 559
KOD                GREYITMTIKEIEEKY--------GFKVIYSDTDGFFATIPG----ADAETVKKK----- 559
P_horikoshii       GRQYIDLVRRELEAR---------GFKVLYIDTDGLYATIPGV---KDWEEVKRR----- 560
T_aggregans        GRHYIEMTIKEIEEKF--------GFKVLYADTDGFYATIPG----EKPETIKKK----- 561
Vent_T.litoralis   GRHYIEMTIREIEEKF--------GFKVLYADTDGFYATIPG----EKPELIKKK----- 562
Mma                GRQYIQKTIEE-AENF--------GFKVIYADTDGFYSKWADDKEKLSKYELLEK----- 578
M_jannaschii       GRKYILETVKE-AEKF--------GFKVLYIDTDGFYAIWK---EKISKEELIKK----- 581
Mth_PolB1          GRDYIKKTIKT-AEEF--------GFHTVYADTDGFYATYRG----------------- 586
```

(SEQ ID NO:14)

FIG. 6-9

```
RB69                GQMALQWIERKVNEYLNEVCGTEGEAFVLYGDTDSIYVSADKIIDKVGESKFRDTNHWVD 649
T4                  GQVGIQWIARKINEYLNKVCGTNDEDFIAAGDTDSVYVCVDKVIEKVGLDRFKEQNDLVE 646
M_marisigri         GRAIIRHTRDIITNLG---------YTVLYGDTDSCMIEVPPGDLEATIARAREIEAKLN 590
                    *:  :                         ***.

9N                  ----------------AKEFLKYINPKLPGLLELEYEGFYVR--------GFFVTKKKY 594
9N DNAP M           ----------------AKEFLKYINPKLPGLLELEYEGFYVR--------GFFVTKKKY 594
TGO                 ----------------AKEFLDYINAKLPGLLELEYEGFYKR--------GFFVTKKKY 594
KOD                 ----------------AMEFLNYINAKLPGALELEYEGFYKR--------GFFVTKKKY 594
P_horikoshii        ----------------ALEFVDYINSKLPGVLELEYEGFYAR--------GFFVTKKKY 595
T_aggregans         ----------------AKEFLKYINSKLPGLLELEYEGFYLR--------GFFVAKKRY 596
Vent_T.litoralis    ----------------AKEFLNYINSKLPGLLELEYEGFYLR--------GFFVTKKRY 597
Mma                 ----------------TREFLKNINNTLPGEMELEFEGYFKR--------GIFVTKKKY 613
M_jannaschii        ----------------AMEFVEYINSKLPGTMELEFEGYFKR--------GIFVTKKRY 616
Mth_PolB1           ------------------------------------------------------------
RB69                FLDKFARERMEPAIDRGFREMCEYMNNKQ-HLMFMDREAIAGPPLGSKGIGGFWTGKKRY 708
T4                  FMNQFGKKKMEPMIDVAYRELCDYMNNRE-HLMHMDREAISCPPLGSKGVGGFWKAKKRY 705
M_marisigri         ---------------ASYGDFAKTELNADTHYFSIKFEKVYRR---------FFQAGKKK 626


9N                  AVIDEEG--------KITTRGLEIVRRDWSEIAKETQARVLEAILKHGDVEEAVRIVKEV 646
9N DNAP M           AVIDEEG--------KITTRGLEIVRRDWSEIAKETQARVLEAILKHGDVEEAVRIVKEV 646
TGO                 AVIDEED--------KITTRGLEIVRRDWSEIAKETQARVLEAILKHGDVEEAVRIVKEV 646
KOD                 AVIDEEG--------KITTRGLEIVRRDWSEIAKETQARVLEALLKDGDVEKAVRIVKEV 646
P_horikoshii        ALIDEEG--------KIVTRGLEIVRRDWSEIAKETQARVLEAILKHGNVEEAVKIVKDV 647
T_aggregans         AVIDEEG--------RITTRGLEVVRRDWSEIAKETQAKVLEAILKEDSVEKAVEIVKDV 648
```

FIG. 6-10

```
Vent_T.litoralis    AVIDEEG--------RITTRGLEVVRRDWSEIAKETQAKVLEAILKEGSVEKAVEVVRDV 649
Mma                 ALIDENE--------KITVKGLEVVRRDWSNVSKNTQKNVLNALLKEGSVENAKKVIQDT 665
M_jannaschii        ALIDENG--------RVTVKGLEFVRRDWSNIAKITQRRVLEALLVEGSIEKAKKIIQDV 668
Mth_PolB1           ------------------------------------------------------------
RB69                ALNVWDMEGTRYAEPKLKIMGLETQKSSTPKAVQKALKECIRRMLQEG-EESLQEYFKEF 767
T4                  ALNVYDMEDKRFAEPHLKIMGMETQQSSTPKAVQEALEESIRRILQEG-EESVQEYYKNF 764
M_marisigri         RYAGHLVWKEGKDVDEVDVVGFEIRRSDSPQITREVQRAVIEMILRGDAFSDVQAYLRDV 686

9N                  TEKLSKYEVPPEKLVIHEQITRDLRDYKATGPHVAVAKRLAARG-----VKIRPGTVISY 701
9N DNAP M           TEKLSKYEVPPEKLVIHEQITRDLRDYKATGPHVAVAKRLAARG-----VKIRPGTVISY 701
TGO                 TEKLSKYEVPPEKLVIYEQITRDLKDYKATGPHVAVAKRLAARG-----IKIRPGTVISY 701
KOD                 TEKLSKYEVPPEKLVIHEQITRDLKDYKATGPHVAVAKRLAARG-----VKIRPGTVISY 701
P_horikoshii        TEKLTNYEVPPEKLVIYEQITRPINEYKAIGPHVAVAKRLMARG-----IKVKPGMVIGY 702
T_aggregans         VEEIAKYQVPLEKLVIHEQITKDLSEYKAIGPHVAIAKRLAAKG-----IKVRPGTIISY 703
Vent_T.litoralis    VEKIAKYRVPLEKLVIHEQITRDLKDYKAIGPHVAIAKRLAARG-----IKVKPGTIISY 704
Mma                 IKELKDGKVNNEDLLIHTQLTKRIEDYKTTAPHVEVAKKILKSG-----NRVNTGDVISY 720
M_jannaschii        IKDLREKKIKKEDLIIYTQLTKDPKEYKTTAPHVEIAKKLMREG-----KRIKVGDIIGY 723
Mth_PolB1           ------------------------------------------------------------
RB69                EKEFRQLNYISIASVSSANNIAKYDVGGFPGPKCPFHIRGILTY-----NRAIKGNIDAP 822
T4                  EKEYRQLDYKVIAEVKTANDIAKYDDKGWPGFKCPFHIRGVLTY-----RRAVSG-LGVA 818
M_marisigri         IRKYRRGEYSLDEAGIPGGIGKSLDSYENDDAHIRGAKYSNMHLGTDFKRGSKPKRVYIK 746

9N                  IVLKGSGRIGDRAIPADEFDPTKHRYDAEYYIENQVLPAVERILKAFGYRKEDLRYQKTK 761
9N DNAP M           IVLKGSGRIGDRAIPADEFDPTKHRYDAEYYIENQVLPAVERILKAFGYRKEDLRYQKTK 761
TGO                 IVLKGSGRIGDRAIPFDEFDPAKHKYDAEYYIENQVLPAVERILRAFGYRKEDLRYQKTR 761
```

FIG. 6-11

```
KOD                IVLKGSGRIGDRAIPFDEFDPTKHKYDAEYYIENQVLPAVERILRAFGYRKEDLRYQKTR 761
P_horikoshii       IVLRGDGPISKRAISIEEFDPRKHKYDAEYYIENQVLPAVERILKAFGYKREDLRWQKTK 762
T_aggregans        IVLRGSGKISDRVILLSEYDPKKHKYDPDYYIENQVLPAVLRILEAFGYRKEDLKYQSSK 763
Vent_T.litoralis   IVLKGSGKISDRVILLTEYDPRKHKYDPDYYIENQVLPAVLRILEAFGYRKEDLRYQSSK 764
Mma                IITSGNKSISERAEILEN----AKNYDTNYYIENQILPPVIRLMEALGITKDELKDSK-K 775
M_jannaschii       IIVKGTKSISERAKLPEEVD--IDDIDVNYYIDNQILPPVLRIMEAVGVSKNELKKEG-A 780
Mth_PolB1          ------------------------------------------------------------
RB69               QVVEGEKVYVLPLREGNPFGDKCIAWPSGTEITDLIKDDVLHWMDYTVLLEKTFIKPLEG 882
T4                 PILDGNKVMVLPLREGNPFGDKCIAWPSGTELPKEIRSDVLSWIDHSTLFQKSFVKPLAG 878
M_marisigri        AVTAKYPRTDVVCFEYADQVPPEFVVDWETMLEKTLKGPLSRIIEPLGWDWHDVDPSRTT 806

9N                 QVGLGAWLKVKGKK------- 775 (SEQ ID NO:5)
9N DNAP M          QVGLGAWLKVKGKK------- 775 (SEQ ID NO:6)
TGO                QVGLGAWLKPKT--------- 773 (SEQ ID NO:7)
KOD                QVGLSAWLKPKGT-------- 774 (SEQ ID NO:8)
P_horikoshii       QVGLGAWIKVKKS-------- 775 (SEQ ID NO:9)
T_aggregans        QVGLDAWLKK---------- 773 (SEQ ID NO:10)
Vent_T.litoralis   QTGLDAWLKR---------- 774 (SEQ ID NO:11)
Mma                QYTLHHFLK---------- 784 (SEQ ID NO:12)
M_jannaschii       QLTLDKFFK---------- 789 (SEQ ID NO:13)
Mth_PolB1          ------------------
RB69               FTSAAKLDYEKKASLFDMFDF 903 (SEQ ID NO:15)
T4                 MCESAGMDYEEKASLDFLFG- 898 (SEQ ID NO:16)
M_marisigri        LFDFGM--------------- 812 (SEQ ID NO:17)
```

FIG. 7

DNA polymerase	Specific activity
9°N D141A/E143A/Y409V/A485L	3,000 units/mg
9°N D141A/E143A/L408S/Y409A/P410V	5,000 units/mg

FIG. 8

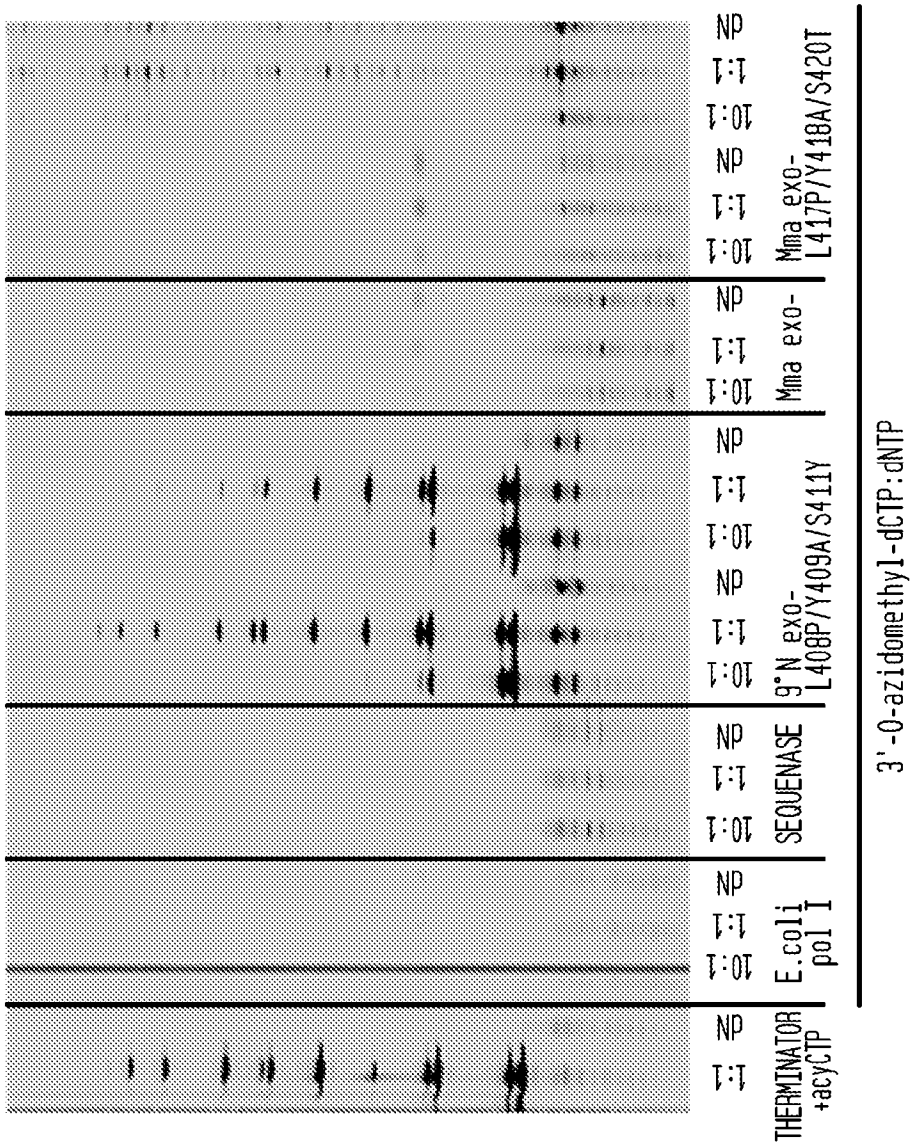
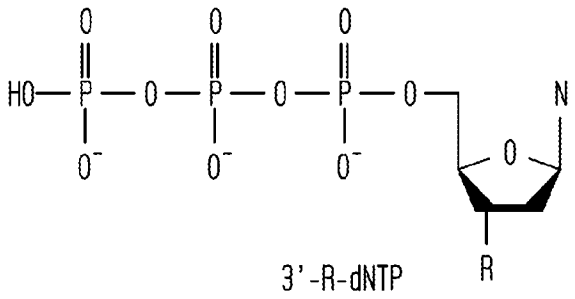

THERMINATOR +acyCTP
E.coli pol I
SEQUENASE
9°N exo- L408P/Y409A/S411Y
Mma exo-
Mma exo- L417P/Y418A/S420T
NP
1:1
10:1
3'-O-azidomethyl-dCTP:dNTP

FIG. 9



3'-R-dNTP

R=

-H	2',3'-dideoxynucleoside-5'-triphosphate
-SH	3'-sulfhydral-2'-deoxynucleoside-5'-triphosphate
-N3	3'-azido-2',3'-dideoxynucleoside-5'-triphosphate
-F	3'-fluoro-2',3'-dideoxynucleoside-5'-triphosphate
-Cl	3'-chloro-2',3'-dideoxynucleoside-5'-triphosphate
-azidomethyl	3'-azidomethyl-2'-deoxynucleoside-5'-triphosphate
-NH2	3'-amino-2',3'-dideoxynucleoside-5'-triphosphate
-anthranyloyl	3'-anthranyloyl-2'-deoxynucleoside-5'-triphosphate
-fluothioureido	3'-{N3-[3-carboxylato-4-(3-oxido-6-oxo-6H-xanthen-9 yl)phenyl]thioureido}-3'-deoxythymidine 5'-triphosphate
-chain	3'-O-[N6(N-methylanthranyl)amidohexanoyl]-2'-deoxyguanosine-5'-triphosphate
-amd	3'-deoxy-3'-(N-methylanthranyloylamino)thymidine 5'-triphosphate
-O-allyl	3'-O-allyl-2'-deoxynucleoside-5'-triphosphate
-O-aminoallyl	3'-O-aminoallyl-2'-deoxynucleoside-5'-triphosphate
-O-azidomethyl	3'-O-azidomethyl-2'-deoxynucleoside-5'-triphosphate
-O-methyl	3'-O-methyl-2'-deoxynucleoside-5'-triphosphate
-O-phophate	3'-O-phosphate-2'-deoxynucleoside-5'-triphosphate
-O-diphophate	3'-O-diphosphate-2'-deoxynucleoside-5'-triphosphate
-O-(2-nitrobenzyl)	3'-O-(2-nitrobenzyl)-2'-deoxynucleoside-5'-triphosphate
-O-[N6(anthranyl)amidohex	3'-O-[N6(anthranyl)amidohex-2'-deoxyadenosine-5'-triphosphate

FIG. 10A

>Mma exo- (SEQ ID NO:22)
MESLIDLDYNSDDLCIYLYLINSIIKEKDFKPYFYVNSTDKEQILEFLKDYEKKHK
DSEISKMIENIETVKKIVFDENYQEKELSKVTVKYPNNVKTVREILMEFERLYEY
IPFVRRYLIDNSVIPTSTWDFENNKKIDNKIPDFKTVSFAIAVYCNKEPNPKKDPI
MASFSSKDFNTVVSTKKFDHEKLEYVKDEKELIKRIIEILKEYDIIYTYNGDNFDF
YLKKRAESFGLELKLGKNDEKIKITKGGMNSKSYIPGRVHIDLYPIARRLLNLTK
RLENVTEALFDVKKVDVGHENIPKMWDNLDETLVEYSHQDAYYTQRIGEQFLP
EIMFSRVVNQSLYDINRMSSSQMVEYLLLKNSYKMGVIAPNRPSGKEYQKRIR
SYEGGYVKEPLKGIHEDIVSMDFLSLYPSIIMSHNLSPETIDCTCCSDEENGEN
EILGHKFCKKSIGIIPKTLMDLINRRKKVKKVLREKAEKGEFDEEYQILDYEQRSI
VLANSHYGYLAFPMARWYSRDCAEITTHLGRQYIQKTIEEAENFGFKVIYADT
GFYSKWADDKEKLSKYELLEKTREFLKNINNTLPGEMELEFEGYFKRGIFVTK
KYALIDENEKITVKGLEVVRRDWSNVSKNTQKNVLNALLKEGSVENAKKVIQDT
KELKDGKVNNEDLLIHTQLTKRIEDYKTTAPHVEVAKKILKSGNRVNTGDVISYII
SGNKSISERAEILENAKNYDTNYYIENQILPPVIRLMEALGITKDELKDSKKQYT
HHFLK

FIG. 10B

>Mma exo- L417P/Y418A/S420T (SEQ ID NO:23)
MESLIDLDYNSDDLCIYLYLINSIIKEKDFKPYFYVNSTDKEQILEFLKDYEKKHK
DSEISKMIENIETVKKIVFDENYQEKELSKVTVKYPNNVKTVREILMEFERLYEY
IPFVRRYLIDNSVIPTSTWDFENNKKIDNKIPDFKTVSFAIAVYCNKEPNPKKDPI
MASFSSKDFNTVVSTKKFDHEKLEYVKDEKELIKRIIEILKEYDIIYTYNGDNFDF
YLKKRAESFGLELKLGKNDEKIKITKGGMNSKSYIPGRVHIDLYPIARRLLNLTK
RLENVTEALFDVKKVDVGHENIPKMWDNLDETLVEYSHQDAYYTQRIGEQFLP
EIMFSRVVNQSLYDINRMSSSQMVEYLLLKNSYKMGVIAPNRPSGKEYQKRIR
SYEGGYVKEPLKGIHEDIVSMDFLSPAPTIIMSHNLSPETIDCTCCSDEENGEN
EILGHKFCKKSIGIIPKTLMDLINRRKKVKKVLREKAEKGEFDEEYQILDYEQRSI
VLANSHYGYLAFPMARWYSRDCAEITTHLGRQYIQKTIEEAENFGFKVIYADT
GFYSKWADDKEKLSKYELLEKTREFLKNINNTLPGEMELEFEGYFKRGIFVTK
KYALIDENEKITVKGLEVVRRDWSNVSKNTQKNVLNALLKEGSVENAKKVIQDT
KELKDGKVNNEDLLIHTQLTKRIEDYKTTAPHVEVAKKILKSGNRVNTGDVISYII
SGNKSISERAEILENAKNYDTNYYIENQILPPVIRLMEALGITKDELKDSKKQYT
HHFLK