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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 July 2007 (05.07.2007)

PCT

(10) International Publication Number
WO 2007/076057 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2006/049122

(22) International Filing Date:
21 December 2006 (21.12.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/753,670 22 December 2005 (22.12.2005) US

(71) Applicant *(for all designated States except US)*: **PACIFIC BIOSCIENCES OF CALIFORNIA, INC.** [US/US]; 1505 Adams Drive, Menlo Park, CA 94025 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HANZEL, David, K.** [US/US]; 988 Loma Verde Ave, Palo Alto, CA 94303 (US). **OTTO, Geoff** [US/US]; 681 Woodhams Rd, Santa Clara, CA 95051 (US). **MURPHY, Devon** [US/US]; 1760 California Street, Apt #1, Mountain View, CA 94040 (US). **PELUSO, Paul** [US/US]; 25200 Carlos Bee Blvd. Apt 167, Hayward, CA 94542 (US). **PHAM, Thang** [US/US]; 1101 Lincoln Drive, Mountain View, CA 94040-4022 (US). **RANK, David, R.** [US/US]; 485 El Capitan Pl., Palo Alto, CA 94306 (US). **MITSIS, Paul** [US/US]; 100 Abernethy Dr., Trenton, NJ 08618-4904 (US).

(74) Agents: **LEBLANC, Stephen, J.** et al.; Quine Intellectual Property Law Group, P.c., P.o. Box 458, Alameda, CA 94501 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, 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, LV, LY, MA, MD, 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, 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, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), 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*

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.

WO 2007/076057 A2

(54) Title: POLYMERASES FOR NUCLEOTIDE ANALOGUE INCORPORATION

(57) Abstract: Compositions that include polymerases with features for improving entry of nucleotide analogues into active site regions and for coordinating with the nucleotide analogues in the active site region are provided. Methods of making the polymerases and of using the polymerases in sequencing and DNA replication and amplification as well as kinetic models of polymerase activity and computer-implemented methods of using the models are also provided.

POLYMERASES FOR NUCLEOTIDE ANALOGUE INCORPORATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a non-provisional utility patent application claiming priority to and benefit of the following prior provisional patent application: USSN 60/753,670, filed December 22, 2005, entitled "POLYMERASES FOR NUCLEOTIDE ANALOGUE INCORPORATION" by David K. Hanzel et al., which is incorporated herein by reference in its entirety for all purposes.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0002] Portions of the invention were made with government support under NHGRI Grant No. R01 HG003710-01, and the government may have certain rights to the invention.

FIELD OF THE INVENTION

[0003] The invention relates to polymerases with features for improving entry of nucleotide analogues into active site regions and for coordinating with the nucleotide analogues in the active site region. Methods of making the polymerases and of using the polymerases in sequencing and DNA replication and amplification, as well as kinetic models of polymerase activity and computer-implemented methods of using the models, are also described.

BACKGROUND OF THE INVENTION

[0004] DNA polymerases replicate the genomes of living organisms. In addition to this central role in biology, DNA polymerases are also ubiquitous tools of biotechnology. They are widely used, e.g., for reverse transcription, amplification, labeling, and sequencing, which are central technologies for a variety of applications such as sequencing, nucleic acid amplification, cloning, protein engineering, diagnostics, molecular medicine and many other technologies.

[0005] Because of the significance of DNA polymerases, they have been extensively studied. This study has focused, e.g., on phylogenetic relationships among polymerases, structure of polymerases, structure-function features of polymerases, and the role of polymerases in DNA replication and other basic biology, as well as ways of using DNA

polymerases in biotechnology. For a review of polymerases, see, e.g., Hübscher et al. (2002) EUKARYOTIC DNA POLYMERASES Annual Review of Biochemistry Vol. 71: 133-163; Alba (2001) "Protein Family Review: Replicative DNA Polymerases" Genome Biology 2(1):reviews 3002.1-3002.4; Steitz (1999) "DNA polymerases: structural diversity and common mechanisms" J Biol Chem 274:17395-17398 and Burgers et al. (2001) "Eukaryotic DNA polymerases: proposal for a revised nomenclature" J Biol Chem. 276(47):43487-90. Crystal structures have been solved for many polymerases, which often share a similar architecture. The basic mechanisms of action for many polymerases have been determined.

[0006] A fundamental application of DNA technology involves various labeling strategies for labeling a DNA that is produced by a DNA polymerase. This is useful in microarray technology, DNA sequencing, SNP detection, cloning, PCR analysis, and many other applications. Labeling is often performed in various post-synthesis hybridization or chemical labeling schemes, but DNA polymerases have also been used to directly incorporate various labeled nucleotides in a variety of applications, e.g., via nick translation, reverse transcription, random priming, amplification, the polymerase chain reaction, etc. See, e.g., Giller et al. (2003) "Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. I. Chemical synthesis of various reporter group-labeled 2′-deoxyribonucleoside-5′-triphosphates" Nucleic Acids Res. 31(10): 2630–2635; Augustin et al. (2001) "Progress towards single-molecule sequencing: enzymatic synthesis of nucleotide-specifically labeled DNA" J. Biotechnol., 86:289–301; Tonon et al. (2000) "Spectral karyotyping combined with locus-specific FISH simultaneously defines genes and chromosomes involved in chromosomal translocations" Genes Chromosom. Cancer 27:418–423; Zhu and Waggoner (1997) "Molecular mechanism controlling the incorporation of fluorescent nucleotides into DNA by PCR." Cytometry, 28:206–211. Yu et al. (1994) "Cyanine dye dUTP analogs for enzymatic labeling of DNA probes" Nucleic Acids Res., 22:3226–3232; Zhu et al. (1994) "Directly labeled DNA probes using fluorescent nucleotides with different length linkers." Nucleic Acids Res. 22:3418–3422; Ried et al. (1992) "Simultaneous visualization of seven different DNA probes by in situ hybridization using combinatorial fluorescence and digital imaging microscopy" Proc. Natl Acad. Sci. USA, 89:1388–1392.

[0007] DNA polymerase mutants have been identified that have altered nucleotide analogue incorporation properties relative to wild-type counterpart enzymes. For example, Vent^{A488L} DNA polymerase can incorporate certain non-standard nucleotides with a higher efficiency than native Vent DNA polymerase. *See* Gardner et al. (2004) "Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase" J. Biol. Chem., 279(12), 11834-11842; Gardner and Jack "Determinants of nucleotide sugar recognition in an archaeon DNA polymerase" Nucleic Acids Research, 27(12) 2545-2553. The altered residue in this mutant, A488, is predicted to be facing away from the nucleotide binding site of the enzyme. The pattern of relaxed specificity at this position roughly correlates with the size of the substituted amino acid side chain and affects incorporation by the enzyme of a variety of modified nucleotide sugars.

[0008] The ability to improve specificity, processivity, or other features of DNA polymerases towards labeled nucleotide analogues would be highly desirable in a variety of contexts where, e.g., nucleic acid labeling is desired, including DNA amplification, sequencing, labeling, detection, cloning, and many others. The present invention provides new DNA polymerases with modified properties for labeled nucleotide analogues, methods of making such polymerases, methods of using such polymerases, and many other features that will become apparent upon a complete review of the following.

SUMMARY OF THE INVENTION

[0009] The invention includes polymerases that incorporate nucleotide analogues, such as phosphate analogues, into a growing template copy, during DNA amplification. Without being bound to any particular theory of operation, these polymerases are optionally modified such that the active site of the polymerase is modified to reduce steric entry inhibition of the analogue into the active site and/or to provide complementarity with one or more non-natural features of the nucleotide analogue. Such polymerases are particularly well-suited for DNA amplification and/or sequencing applications, including real-time applications, *e.g.*, in the context of amplification or sequencing protocols that include incorporation of analogue residues into DNA by the polymerase. The analogue residue that is incorporated can be the same as a natural residue, e.g., where a label or other moiety of the analogue is removed by action of the polymerase during incorporation, or the analogue residue can have one or more feature that distinguishes it from a natural nucleotide residue.

[0010] Accordingly, the invention includes compositions that include a recombinant DNA polymerase. The recombinant DNA polymerase includes a modified active site region that is homologous to a wild-type active site region of a wild-type DNA polymerase. The modified active site region includes one or more structural modifications relative to the wild type active site region that improve the desired activity of the enzyme, e.g., toward naturally occurring nucleotides and/or nucleotide analogues. In certain aspects, and without being bound to a particular theory of operation, such modifications include those that reduce steric inhibition for entry of a natural nucleotide or nucleotide analogue into the modified active site region and/or that make the active site region complementary with one or more non-natural features of the natural nucleotide and/or nucleotide analogue. The recombinant DNA polymerase displays a modified property for the nucleotide analogue as compared to the wild-type polymerase.

[0011] A variety of DNA polymerases are optionally modified to include the modified active site region. For example, the recombinant DNA polymerase is optionally homologous to a Φ29 DNA polymerase or mutant thereof, a Taq polymerase, an exonuclease deficient Taq polymerase, a DNA Pol I polymerase, a T7 polymerase, an RB69 polymerase, a T5 polymerase, or a polymerase corresponding to a Klenow fragment of a DNA Pol I polymerase. For example, the recombinant DNA polymerase can be homologous to a wild-type or exonuclease deficient Φ29 DNA polymerase, e.g., as described in U.S. Patent Nos. 5,001,050, 5,198,543, or 5,576,204. Similarly, the recombinant DNA polymerase can be homologous to Φ29 , B103, GA-1, PZA, Φ15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, or L17, or the like. For nomenclature, *see also*, Meijer et al. (2001) "Φ29 Family of Phages" Microbiology and Molecular Biology Reviews, 65(2):261-287.

[0012] The modified active site region can include any of a variety of different modifications to reduce steric inhibition and/or to make the region complementary with one or more non-natural features of the nucleotide analogue. For example, structural modifications within or proximal to the active site relative to the wild-type Φ29 DNA polymerase are selected from: a Δ505-525 deletion, a deletion within Δ505-525, a K135A mutation, an L384R mutation in combination with another mutation herein (when an L384R mutation is present, it will generally be in combination with one or more additional mutation that reduces steric inhibition for entry of the nucleotide analogue), an E375H

mutation, an E375S mutation, an E375K mutation, an E375R mutation, an E375A mutation, an E375Q mutation, an E375W mutation, an E375Y mutation, an E375F mutation, an E486A mutation, an E486D mutation, a K512A mutation, and combinations thereof. The polymerase can also include an additional mutation or combination of mutations selected from those listed in Table 8.

[0013] The polymerase optionally further includes one or more mutations/deletions relative to the wild-type polymerase that reduce or eliminate endogenous exonuclease activity. For example, relative to the wild-type Φ29 DNA polymerase, N62 is optionally mutated or deleted to reduce exonuclease activity; e.g., the polymerase can include an N62D mutation. Other example mutations that reduce exonuclease activity include: D12A, T15I, E14I, and/or D66A; accordingly, the polymerases of the invention optionally comprise one or more of these mutations.

[0014] The recombinant DNA polymerase optionally includes additional features exogenous or heterologous to a corresponding DNA polymerase such as a wild-type or nuclease deficient polymerase. For example, the recombinant polymerase optionally includes one or more exogenous affinity tags, e.g., purification or substrate binding tags, such as a 6 His tag sequence, a GST tag, an HA tag sequence, a plurality of 6 His tag sequences, a plurality of GST tags, a plurality of HA tag sequences, a SNAP-tag, or the like. These may be inserted into any of a variety of positions within the protein, and are preferably at one or more termini, e.g., C terminus or N terminus of the protein, and are more preferably, at the terminus that is most distal to the active site in the 3D structure of the protein.

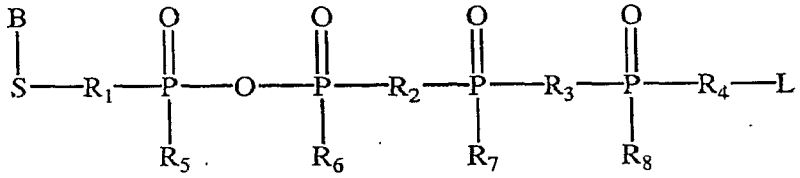
[0015] Example polymerases of the invention include those listed in Table 3.

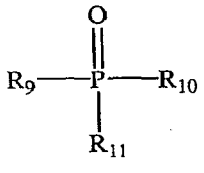
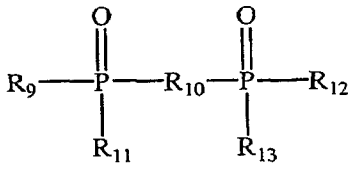
[0016] The compositions optionally include the nucleotide analogue. Example nucleotide analogues include those that include fluorophore and/or dye moieties. For example, the nucleotide analogue can be a labeled nucleotide, e.g., a base, sugar and/or phosphate labeled nucleotide. The analogue can be a mono-deoxy or a dideoxy nucleotide analogue.

[0017] One example class of nucleotide analogues are phosphate-labeled nucleotide analogues, including mono-deoxy phosphate-labeled nucleotide analogues and/or dideoxy phosphate-labeled nucleotide analogues. For example, the nucleotide analogue can be a

labeled nucleotide analogue having from 3 to 6 phosphate groups (e.g., where the nucleotide analogue is a triphosphate, a tetraphosphate, a pentaphosphate or a hexaphosphate).

[0018] For example, the composition can include a labeled compound of the formula:



wherein B is a nucleobase (note that B optionally includes a label); S is selected from a sugar moiety, an acyclic moiety or a carbocyclic moiety (note that S optionally includes a label); L is an optional detectable label; R_1 is selected from O and S; R_2, R_3 and R_4 are independently selected from O, NH, S, methylene, substituted methylene, C(O), $C(CH_2)$, CNH_2, CH_2CH_2, $C(OH)CH_2R$ where R is 4-pyridine or 1-imidazole, provided that R_4 may additionally be selected from  and  ; R_5, R_6, R_7, R_8, R_{11} and R_{13} are, when present, each independently selected from O, BH_3, and S; and R_9, R_{10} and R_{12} are independently selected from O, NH, S, methylene, substituted methylene, CNH_2, CH_2CH_2, $C(OH)CH_2R$ where R is 4-pyridine or 1-imidazole. In some cases, phosphonate analogs may be employed as the analogs, e.g., where one of R_2, R_3, R_4, R9, R10 or R12 are not O, e.g., they are methyl etc.

[0019] The recombinant DNA polymerase displays a modified property for the nucleotide analogue as compared to the wild-type polymerase. For example, the modified property can be, e.g., K_m, k_{cat}, V_{max}, recombinant polymerase processivity in the presence of the nucleotide analogue (or of a naturally occurring nucleotide), average template read-length by the recombinant polymerase in the presence of the nucleotide analogue, specificity of the recombinant polymerase for the nucleotide analogue, rate of binding of the nucleotide analogue, rate of product (pyrophosphate, triphosphate, etc.) release, and/or branching rate. In one desirable embodiment, the modified property is a reduced K_m for the nucleotide analogue and/or an increased k_{cat}/K_m or V_{max}/K_m for the nucleotide analogue. Similarly, the recombinant polymerase optionally has an increased rate of binding of the

nucleotide analogue, an increased rate of product release, and/or a decreased branching rate, as compared to the wild-type polymerase.

[0020] At the same time, the recombinant DNA polymerase can incorporate natural nucleotides (e.g., A, C, G and T) into a growing copy nucleic acid. For example, the recombinant polymerase optionally displays a specific activity for a natural nucleotide that is at least about 5% as high (e.g., 5%, 10%, 25%, 50%, 75%, 100% or higher), as a corresponding wild-type polymerase and a processivity with natural nucleotides in the presence of a template that is at least 5% as high (e.g., 5%, 10%, 25%, 50%, 75%, 100% or higher) as the wild-type polymerase in the presence of the natural nucleotide. Optionally, the recombinant polymerase displays a k_{cat}/K_m or V_{max}/K_m for a naturally occurring nucleotide that is at least about 5% as high (e.g., about 5%, 10%, 25%, 50%, 75% or 100% or higher) as the wild-type polymerase.

[0021] The nucleotide analogue and a DNA template are optionally included in compositions of the invention, e.g., in which the recombinant polymerase incorporates the nucleotide analogue into a copy nucleic acid in response to the template DNA. The template DNA can be linear or circular DNA, and in certain sequencing applications is desirable a circular template. Thus, the composition can be present in a DNA amplification and/or sequencing system. Optionally, in one class of embodiments, the sequencing system comprises a Zero Mode Waveguide.

[0022] Methods of making and using the compositions are also features of the invention. For example, in one aspect, methods of making a DNA e.g., comprising one or more nucleotide analogue residues are provided. In these methods, a reaction mixture is provided. The reaction mixture typically includes those components that can at least partially replicate a template, e.g., a template, nucleotides, the polymerase and a replication initiating moiety that complexes with the template, or is integral to it, to prime the polymerase. The replication initiating moiety in this context is any moiety that can serve as a site to initiate the polymerase, e.g., a separate oligonucleotide complementary to the template, a hairpin or other self-complementary region of a template (e.g., a hairpin in a single-stranded template), a terminal protein, or the like. The polymerase is a recombinant polymerase capable of at least partially replicating the template in a template-dependent polymerase extension reaction (e.g., using the replication initiation moiety as a site of initiation). Typically, the one or more nucleotides comprise a nucleotide analogue. In

preferred aspects, at least one, preferably two or more, three or more or at least four nucleotides are nucleotide analogues. The recombinant DNA polymerase has a modified active site (a region of the polymerase that, when modified, results in an alteration in an activity of the polymerase) that is homologous to a wild-type active site of a wild-type DNA polymerase. As discussed in the context of the compositions above, the modified active site can include one or more structural modification relative to the wild type active site that improves the activity of the enzyme toward one or more natural nucleotides and/or nucleotide analogues. In at least one example, and without being bound to any particular theory of operation, the modification to the active site reduces steric inhibition for entry of the nucleotide analogue into the modified active site and/or the modification is complementary with one or more non-natural features of the nucleotide analogue.

[0023] The mixture is reacted such that the recombinant polymerase replicates at least a portion of the template in a template-dependent manner, whereby at least one nucleotide analogue residue is incorporated into the resulting DNA. Incorporation of the analogue can result in the incorporation of a non-standard residue into the extended DNA (e.g., as a labeled nucleotide residue), or action of the polymerase can modify the analogue such that the nucleotide analogue residue incorporated into the extended DNA is structurally the same as a standard nucleotide residue. For example, in the latter embodiment, a variety of labels are cleaved by action of the polymerase, e.g., certain phosphate labels discussed in more detail herein are cleaved from the nucleotide analogue as it is incorporated into the growing DNA (typically providing a signal upon release of the label).

[0024] In a related class of methods, a reaction mixture is provided that includes a template, a replication initiating moiety, a template-dependent recombinant polymerase and one or more nucleotides. The one or more nucleotides include a phosphate labeled nucleotide. A K_m value of the recombinant polymerase for the nucleotide analogue is lower than a K_m for a corresponding homologous wild-type polymerase for the nucleotide analogue. The mixture is reacted such that the polymerase at least partially replicates the template in a template-dependent manner, e.g., whereby at least one nucleotide analogue residue is incorporated into the resulting DNA. As noted previously, once incorporated, the residue can be the same as a natural nucleotide residue, or can be different from a natural nucleotide residue.

[0025] In another related class of methods of making a DNA, a reaction mixture that includes a template, a replication initiating moiety that complexes with or is integral to the template, a polymerase capable of replicating at least a portion of the template using the moiety in a template-dependent polymerase extension reaction, and one or more nucleotide is provided. Here again, the one or more nucleotide typically includes a labeled phosphate nucleotide analogue. The polymerase in this class of embodiments is homologous to a Φ29 DNA polymerase. The polymerase has a K_m for 488dC4P, A568dC4P, or both, that is less than about 75% of a K_m of a GST-N62D Φ29 DNA polymerase for 488dC4P, A568dC4P or both. For example, the K_m for 488dC4P, A568dC4P can be about 40% or less than GST-N62D Φ29 DNA polymerase, or, e.g., about 15% or less. The mixture is reacted such that the polymerase replicates at least a portion of the template.

[0026] The polymerases used in the methods can be any of those noted above with reference to the compositions. The properties of the polymerases used in the methods can be any of those noted in reference to compositions. For example, the polymerase optionally has a k_{cat}/K_m for the nucleotide analogue that is higher than a k_{cat}/K_m of a wild-type Φ29 for the nucleotide analogue. Similarly, the nucleotide analogues used in the methods can be any of those noted in reference to the compositions herein. The recombinant polymerases herein can have a K_m for the nucleotide analogue that is e.g., about 90% as high, about 80% as high, about 75% as high, about 60% as high, about 50% as high, about 40% as high, about 25% as high, about 15% as high, about 10% as high, or less than about 5% as high as a K_m of a naturally occurring polymerase homologous to the recombinant polymerase. The recombinant polymerase optionally has an increased rate of binding of the nucleotide analogue, an increased rate of product release, and/or a decreased branching rate, as compared to the corresponding wild-type polymerase.

[0027] In addition to methods of using the compositions herein, the present invention also includes methods of making the compositions. For example, in one aspect, a method of making a recombinant DNA polymerase (e.g., any of those discussed with respect to the compositions herein) is provided. For example, the methods can include structurally modeling a first polymerase, e.g., using any available crystal structure and molecular modeling software or system. Based on the modeling, one or more steric inhibition feature or complementarity feature affecting nucleotide access to the active site and/or binding of a nucleotide analogue within the active site region is identified, e.g., in the

active site or proximal to it. The first DNA polymerase is mutated to reduce or remove at least one steric inhibition feature or to add the complementarity feature.

[0028] The method can additionally include screening or other protocols to determine whether the resulting recombinant polymerase displays a modified activity for a nucleotide analogue as compared to the first DNA polymerase. For example, k_{cat}, K_m, V_{max}, or k_{cat}/K_m of the recombinant DNA polymerase for the nucleotide analogue can be determined. Further, k_{cat}, K_m, V_{max}, or k_{cat}/K_m of the recombinant DNA polymerase for a natural nucleotide can also be determined (e.g., where the polymerase desirably includes both analogue and natural nucleotide incorporation activity).

[0029] A library of recombinant DNA polymerases can be made and screened for these properties. For example, a plurality of members of the library can be made to include one or more steric inhibition feature mutation and/or a mutation to produce complementary with one or more non-natural features of the nucleotide analogue, that is then screened for the properties of interest. In general, the library can be screened to identify at least one member comprising a modified activity of interest.

[0030] In an additional aspect, the invention includes computer-implemented methods, e.g., for modeling enzyme kinetics. The methods include, e.g., defining a plurality of polymerase state transitions for discrete time steps during a template-based polymerization reaction; defining a plurality of rate transition rates between the states; generating a multidimensional probability matrix of possible states, based upon a given nucleic acid template sequence, nucleotides in a reaction mixture and the polymerase state transitions; and, storing the multidimensional probability matrix in a computer readable medium.

[0031] A variety of features of the method can vary. For example, the polymerase state transitions are optionally user-selectable. The rate transition rates between the states optionally vary depending on nucleotide concentration, template sequence and position of the polymerase along the template. The nucleotides in the reaction mixture optionally comprise one or more nucleotide analogues. The rate transition rates between states optionally include a conformational transition rate for the polymerase during use of the nucleotide analogues by the polymerase, with the rate set to be equal to a conformational transition rate for a natural nucleotide. The multidimensional probability matrix is

optionally automatically generated based upon the template sequence, a standardized matrix of probability states, and the nucleotides in the reaction mixture. The probability matrix is optionally simplified by assuming that all possible Watson-Crick base pairings are equivalent in all state transitions.

[0032] Similarly, a second reagent concentration matrix is optionally generated to account for reagent concentration changes that result from position of the polymerase along a template, based on an output of the probability matrix. The probability matrix is optionally vectorized for multiple templates and the resulting vectorized probability matrix can be multiplied by the multidimensional probability matrix to provide a state distribution matrix. An exponential time factor for the probability matrix can be used to account for repeated sequences within the template sequence. A polymerase nucleotide mismatch fraction using either a continuum model or a counting model can be defined.

BRIEF DESCRIPTION OF THE DRAWINGS

[0033] Figure 1 schematically depicts a vector for expression of tagged N62D Phi 29 DNA polymerase.

[0034] Figure 2 Panel A presents a sequence alignment for Phi 29-like polymerases in the region surrounding residues 505-525 (Phi29 SEQ ID NO:35, B103 SEQ ID NO:36, PZA SEQ ID NO:37, M2 SEQ ID NO:38, G1 SEQ ID NO:39, cp-1 SEQ ID NO:40). Panel B illustrates the structure of Phi 29 with (top) and without (bottom) residues 505-525. Views of the structures from three different angles are shown.

[0035] Figure 3 Panel A presents a sequence alignment for Phi 29-like polymerases in the region surrounding E375 of Phi 29 (Phi29 SEQ ID NO:41, B103 SEQ ID NO:42, PZA SEQ ID NO:43, M2 SEQ ID NO:44, G1 SEQ ID NO:45, cp-1 SEQ ID NO:46). Panels B illustrates the structure of Phi 29 (top) and an E375H mutant (bottom). Views of the structures from three different angles are shown.

[0036] Figure 4 Panel A presents a sequence alignment for Phi 29-like polymerases in the region surrounding E486 of Phi 29 (Phi29 SEQ ID NO:47, B103 SEQ ID NO:48, PZA SEQ ID NO:49, M2 SEQ ID NO:50, G1 SEQ ID NO:51, cp-1 SEQ ID NO:52). Panels B illustrates the structure of Phi 29 (top) and an E486A mutant (bottom). Views of the structures from three different angles are shown.

[0037] Figure 5 Panel A shows a sequence alignment for Phi 29-like polymerases in the region surrounding K512 of Phi 29 (Phi29 SEQ ID NO:53, B103 SEQ ID NO:54, PZA SEQ ID NO:55, M2 SEQ ID NO:56, G1 SEQ ID NO:57, cp-1 SEQ ID NO:58). Panels B illustrates the structure of Phi 29 (top) and a K512A mutant (bottom). Views of the structures from three different angles are shown.

[0038] Figure 6 Panel A shows a sequence alignment for Phi 29-like polymerases in the region surrounding K135 of Phi 29 (Phi29 SEQ ID NO:59, B103 SEQ ID NO:60, PZA SEQ ID NO:61, M2 SEQ ID NO:62, G1 SEQ ID NO:63, cp-1 SEQ ID NO:64). Panels B illustrates the structure of Phi 29 (top) and a K135A mutant (bottom). Views of the structures from three different angles are shown..

[0039] Figure 7 Panel A schematically illustrates a FRET stopped flow assay used to determine rates of binding and product release. Results of the assay are shown in Panels B-D, for Phi29 N62D (Panel B), N62D:E375Y (Panel C), and N62D:E375W (Panel D).

[0040] Figure 8 Panel A schematically illustrates a FRET stopped flow assay used to determine branching rate. Results of the assay are shown in Panels B-D, for Phi29 N62D (Panel B), N62D:E375Y (Panel C), and N62D:E375W (Panel D).

[0041] Figure 9 depicts a plot of kinetic matrix jump size vs. concentration drop.

DETAILED DISCUSSION OF THE INVENTION

OVERVIEW

[0042] A variety of technologies rely on the incorporation of labels into nucleic acids to observe the results of an experiment. For example, the outcome of sequencing, nucleic acid amplification and nick translation reactions are all typically monitored by labeling product nucleic acids. This is often done by covalently or non-covalently binding labels to the product nucleic acids, e.g., by binding labeled probes to the product nucleic acid. In other approaches, nucleotide analogues are incorporated into product nucleic acids during synthesis of the product nucleic acid. This typically occurs, e.g., in sequencing by incorporation methods, and certain real-time PCR (RT-PCR) and real-time LCR reactions (RT-LCR). A label present on the analogue can be incorporated into the DNA, or it can be released by action of the polymerase. Incorporation or release of the label can be monitored to monitor incorporation of an analogue residue into the product nucleic acid.

[0043] The present invention provides new polymerases that incorporate nucleotide analogues, such as dye labeled phosphate labeled analogues, into a growing template copy, during DNA amplification. These polymerases are modified such that the active site of the polymerase is modified to reduce steric entry inhibition of the analogue into the active site (facilitating entry of the nucleotide analogue into the active site) and/or to provide complementarity with one or more non-natural features of the nucleotide analogue.

[0044] These new polymerases are particularly well-suited to DNA amplification (e.g., RT-PCR and RT-LCR) and/or sequencing applications, e.g., in the context of amplification or sequencing protocols that include incorporation of labeled analogues into DNA amplicons.

DNA POLYMERASES

[0045] DNA polymerases that can be modified to interact with nucleotide analogues by reducing steric entry inhibition into the active site, or by adding features complementary to the analogues, are generally available. DNA polymerases have relatively recently been classified into six main groups based upon various phylogenetic relationships, e.g., with E. coli Pol I (class A), E. coli Pol II (class B), E. coli Pol III (class C), Euryarchaeotic Pol II (class D), human Pol beta (class X), and E. coli UmuC/DinB and eukaryotic RAD30/xeroderma pigmentosum variant (class Y). For a review of recent nomenclature, see, e.g., Burgers et al. (2001) "Eukaryotic DNA polymerases: proposal for a revised nomenclature" J Biol Chem. 276(47):43487-90. For a review of polymerases, see, e.g., Hübscher et al. (2002) EUKARYOTIC DNA POLYMERASES Annual Review of Biochemistry Vol. 71: 133-163; Alba (2001) "Protein Family Review: Replicative DNA Polymerases" Genome Biology 2(1):reviews 3002.1-3002.4; and Steitz (1999) "DNA polymerases: structural diversity and common mechanisms" J Biol Chem 274:17395-17398. The basic mechanisms of action for many polymerases have been determined. The sequences of literally hundreds of polymerases are publicly available, and the crystal structures for many of these have been determined, or can be inferred based upon similarity to solved crystal structures for homologous polymerases. For example, the crystal structure of Φ29 is available.

[0046] Available DNA polymerase enzymes have also been modified in any of a variety of ways, e.g., to reduce or eliminate exonuclease activities (many native DNA

polymerases have a proof-reading exonuclease function that interferes with, e.g., sequencing applications), to simplify production by making protease digested enzyme fragments such as the Klenow fragment recombinant, etc. Any of these available polymerases can be modified in accordance with the invention to reduce steric inhibition to analogue entry into the active site, or to provide features complementary to the analogue. Many such polymerases that are suitable for modification are available, e.g., for use in sequencing, labeling and amplification technologies. For example, Human DNA Polymerase Beta is available from R&D systems. DNA polymerase I is available from Epicenter, GE Health Care, Invitrogen, New England Biolabs, Promega, Roche Applied Science, Sigma Aldrich and many others. The Klenow fragment of DNA Polymerase I is available in both recombinant and protease digested versions, from, e.g., Ambion, Chimerx, eEnzyme LLC, GE Health Care, Invitrogen, New England Biolabs, Promega, Roche Applied Science, Sigma Aldrich and many others. Φ29 DNA polymerase is available from e.g., Epicenter. Poly A polymerase, reverse transcriptase, Sequenase, SP6 DNA polymerase, T4 DNA polymerase, T7 DNA polymerase, and a variety of thermostable DNA polymerases (Taq, hot start, titanium Taq, etc.) are available from a variety of these and other sources. Recent commercial DNA polymerases include Phusion™ High-Fidelity DNA Polymerase is available from New England Biolabs; GoTaq® Flexi DNA Polymerase available from Promega; RepliPHI™ Φ29 DNA Polymerase from EPICENTRE; PfuUltra™ Hotstart DNA Polymerase available from Stratagene; KOD HiFi DNA Polymerase is available from Novagen and many others. Biocompare(dot)com provides comparisons of many different commercially available polymerases.

[0047] DNA polymerases that are preferred substrates for mutation to reduce steric inhibition or to incorporate features complementary to the nucleotide analogue include Taq polymerases, exonuclease deficient Taq polymerases, E. coli DNA Polymerase 1, Klenow fragment, reverse transcriptases, Φ29 related polymerases including wild type Φ29 polymerase derivatives of such polymerases such as exonuclease deficient forms, T7 DNA Polymerase, T5 DNA Polymerase, an RB69 polymerase, etc. For example, the recombinant DNA polymerase can be homologous to a wild-type or exonuclease deficient Φ29 DNA polymerase, e.g., as described in U.S. Patent Nos. 5,001,050, 5,198,543, or 5,576,204. Similarly, the recombinant DNA polymerase can be homologous to Φ29 , B103, GA-1,

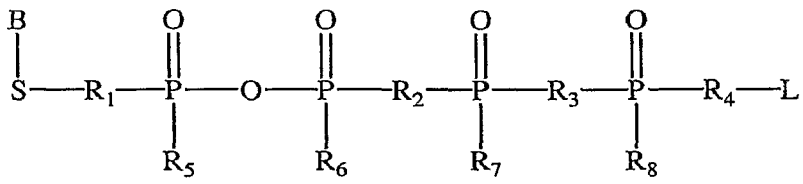
PZA, Φ15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, or L17, or the like.

NUCLEOTIDE ANALOGUES

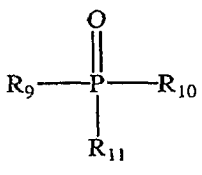
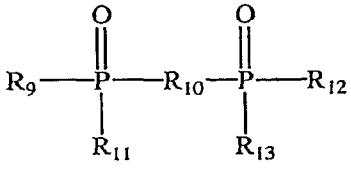
[0048] As discussed, various polymerases of the invention can incorporate one or more nucleotide analogues into a growing oligonucleotide chain. Upon incorporation, the analogue can leave a residue that is the same or different than a natural nucleotide in the growing oligonucleotide (the polymerase can incorporate any non-standard moiety of the analogue, or can cleave it off during incorporation into the oligonucleotide). A "nucleotide analogue" herein is a compound, that, in a particular application, functions in a manner similar or analogous to a naturally occurring nucleoside triphosphate (a "nucleotide"), and does not otherwise denote any particular structure. A nucleotide analogue is an analogue other than a standard naturally occurring nucleotide, i.e., other than A, G, C, T, or U, though upon incorporation into the oligonucleotide, the resulting residue in the oligonucleotide can be the same as (or different from) an A, G, C, T or U residue.

[0049] Many nucleotide analogues are available. These include analogue structures with core similarity to naturally occurring nucleotides, such as those that comprise one or more substituent on a phosphate, sugar or base moiety of the nucleoside or nucleotide relative to a naturally occurring nucleoside or nucleotide. In one embodiment, a nucleotide analogue can include one or more extra phosphate containing groups, relative to a nucleoside triphosphate. For example, a variety of nucleotide analogues that comprise, e.g., from 4-6 phosphates are described in detail in U.S. Patent Application No. 11/241,809, filed September 29, 2005, and incorporated herein by reference in its entirety for all purposes.

[0050] For example, the analogue can include a labeled compound of the formula:



wherein B is a nucleobase (and optionally includes a label); S is selected from a sugar moiety, an acyclic moiety or a carbocyclic moiety (and optionally includes a label); L is an optional detectable label; R_1 is selected from O and S; R_2, R_3 and R_4 are independently selected from O, NH, S, methylene, substituted methylene, C(O), $C(CH_2)$, CNH_2, CH_2CH_2,

$C(OH)CH_2R$ where R is 4-pyridine or 1-imidazole, provided that R_4 may additionally be selected from  , and  ; R_5, R_6, R_7, R_8, R_{11} and R_{13} are, when present, each independently selected from O, BH_3, and S; and R_9, R_{10} and R_{12} are independently selected from O, NH, S, methylene, substituted methylene, CNH_2, CH_2CH_2, $C(OH)CH_2R$ where R is 4-pyridine or 1-imidazole. In some cases, phosphonate analogs may be employed as the analogs, e.g., where one of R_2, R_3, R_4, R9, R10 or R12 are not O, e.g., they are methyl etc. See, e.g., U.S. Patent Application No. 11/241,809, previously incorporated herein by reference in its entirety for all purposes.

[0051] The base moiety incorporated into the analogue is generally selected from any of the natural or non-natural nucleobases or nucleobase analogs, including, e.g., purine or pyrimidine bases that are routinely found in nucleic acids and available nucleic acid analogs, including adenine, thymine, guanine, cytidine, uracil, and in some cases, inosine. As noted, the base optionally includes a label moiety. For convenience, nucleotides and nucleotide analogs are generally referred to based upon their relative analogy to naturally occurring nucleotides. As such, an analogue that operates, functionally, like adenosine triphosphate, may be generally referred to herein by the shorthand letter A. Likewise, the standard abbreviations of T, G, C, U and I, may be used in referring to analogs of naturally occurring nucleosides and nucleotides typically abbreviated in the same fashion. In some cases, a base may function in a more universal fashion, e.g., functioning like any of the purine bases in being able to hybridize with any pyrimidine base, or vice versa. The base moieties used in the present invention may include the conventional bases described herein or they may include such bases substituted at one or more side groups, or other fluorescent bases or base analogs, such as 1,N6 ethenoadenosine or pyrrolo C, in which an additional ring structure renders the B group neither a purine nor a pyrimidine. For example, in certain cases, it may be desirable to substitute one or more side groups of the base moiety with a labeling group or a component of a labeling group, such as one of a donor or acceptor fluorophore, or other labeling group. Examples of labeled nucleobases and processes for labeling such groups are described in, e.g., U.S. Patent Nos. 5,328,824 and 5,476,928, each of which is incorporated herein by reference in its entirety for all purposes.

[0052] In the analogues, the S group is optionally a sugar moiety that provides a suitable backbone for a synthesizing nucleic acid strand. For example, the sugar moiety is optionally selected from a *D*-ribosyl, 2' or 3' *D*-deoxyribosyl, 2',3'-*D*-dideoxyribosyl, 2', 3'-*D*-didehydrodideoxyribosyl, 2' or 3' alkoxyribosyl, 2' or 3' aminoribosyl, 2' or 3' mercaptoribosyl, 2' or 3' alkothioribosyl, acyclic, carbocyclic or other modified sugar moieties. A variety of carbocyclic or acyclic moieties can be incorporated as the "S" group in place of a sugar moiety, including, e.g., those described in published U.S. Patent Application No. 2003/0124576, previously incorporated herein by reference in its entirety for all purposes.

[0053] For most cases, the phosphorus containing chain in the analogues, e.g., a triphosphate in conventional NTPs, is preferably coupled to the 5' hydroxyl group, as in natural nucleoside triphosphates. However, in some cases, the phosphorus containing chain is linked to the S group by the 3' hydroxyl group.

[0054] L generally refers to a detectable labeling group that is coupled to the terminal phosphorus atom via the R_4 (or R_{10} or R_{12}) group. The labeling groups employed in the analogs of the invention may comprise any of a variety of detectable labels. Detectable labels generally denote a chemical moiety that provides a basis for detection of the analogue compound separate and apart from the same compound lacking such a labeling group. Examples of labels include, e.g., optical labels, e.g., labels that impart a detectable optical property to the analogue, electrochemical labels, e.g., labels that impart a detectable electrical or electrochemical property to the analogue, physical labels, e.g., labels that impart a different physical or spatial property to the analogue, e.g., a mass tag or molecular volume tag. In some cases individual labels or combinations may be used that impart more than one of the aforementioned properties to the analogs of the invention.

[0055] Optionally, the labeling groups incorporated into the analogs comprise optically detectable moieties, such as luminescent, chemiluminescent, fluorescent, fluorogenic, chromophoric and/or chromogenic moieties, with fluorescent and/or fluorogenic labels being preferred. A variety of different label moieties are readily employed in nucleotide analogs. Such groups include fluorescein labels, rhodamine labels, cyanine labels (i.e., Cy3, Cy5, and the like, generally available from the Amersham Biosciences division of GE Healthcare), the Alexa family of fluorescent dyes and other fluorescent and fluorogenic dyes available from Molecular Probes/Invitrogen, Inc., and

described in 'The Handbook — A Guide to Fluorescent Probes and Labeling Technologies, Tenth Edition' (2005) (available from Invitrogen, Inc./Molecular Probes). A variety of other fluorescent and fluorogenic labels for use with nucleoside polyphosphates, and which would be applicable to the nucleotide analogues incorporated by the polymerases of the present invention are described in, e.g., Published U.S. Patent Application No. 2003/0124576, the full disclosure of which is incorporated herein in its entirety for all purposes.

[0056] Additional details regarding analogues and methods of making such analogues can be found in U.S. Patent Application No. 11/241,809, filed September 29, 2005, and incorporated herein by reference in its entirety for all purposes.

[0057] Thus, in one illustrative example, the analogue can be a phosphate analogue (e.g., an analogue that has more than the typical number of phosphates found in nucleoside triphosphates) that include, e.g., an Alexa dye label. For example, an Alexa488 dye can be labeled on a deltaphosphate (denoted, e.g., A488dC4P), or an Alexa568 or Alexa633 dye can be used (e.g., A568dC4P, and A633dC4P respectively), or an Alexa546 dye can be used (e.g., A546dG4P), or an Alexa594 dye can be used (e.g., A594dT4P). Similarly, to facilitate color separation, a pair of fluorophores exhibiting FRET (fluorescence resonance energy transfer) can be labeled on a delta phosphate of a tetraphosphate analog (denoted, e.g., FAM-amb-A532dG4P or FAM-amb-A594dT4P).

MODIFYING DNA POLYMERASES TO REDUCE STERIC HINDRANCE FEATURES AND/OR TO ADD COMPLEMENTARITY FEATURES

Structure-based design of recombinant polymerases

[0058] Structural data for a polymerase can be used to conveniently identify amino acid residues as candidates for mutagenesis to create recombinant polymerases having modified active site regions. For example, analysis of the three-dimensional structure of a polymerase can identify residues that sterically hinder access to the active site by a natural nucleotide or nucleotide analogue or analogue thereof or that can be mutated to introduce a feature complementary to a non-natural feature of the analogue, e.g., by adding or altering charge, hydrophobicity, size, or the like.

[0059] The three-dimensional structures of a large number of DNA polymerases have been determined by x-ray crystallography and nuclear magnetic resonance (NMR)

spectroscopy, including the structures of polymerases with bound templates, nucleotides, and/or nucleotide analogues. Many such structures are freely available for download from the Protein Data Bank, at (www(dot)rcsb(dot)org/pdb. Structures, along with domain and homology information, are also freely available for search and download from the National Center for Biotechnology Information's Molecular Modeling DataBase, at www(dot)ncbi(dot)nlm(dot)nih(dot)gov/Structure/MMDB/mmdb(dot)shtml. The structures of additional polymerases can be modeled, for example, based on homology of the polymerases with polymerases whose structures have already been determined. Alternatively, the structure of a given polymerase, optionally complexed with a nucleotide analogue, or the like, can be determined.

[0060] Techniques for crystal structure determination are well known. See, for example, McPherson (1999) Crystallization of Biological Macromolecules Cold Spring Harbor Laboratory; Bergfors (1999) Protein Crystallization International University Line; Mullin (1993) Crystallization Butterwoth-Heinemann; Stout and Jensen (1989) X-ray structure determination: a practical guide, 2nd Edition Wiley Publishers, New York; Ladd and Palmer (1993) Structure determination by X-ray crystallography, 3rd Edition Plenum Press, NewYork; Blundell and Johnson (1976) Protein Crystallography Academic Press, New York; Glusker and Trueblood (1985) Crystal structure analysis: A primer, 2nd Ed. Oxford University Press, NewYork; International Tables for Crystallography, Vol. F. Crystallography of Biological Macromolecules; McPherson (2002) Introduction to Macromolecular Crystallography Wiley-Liss; McRee and David (1999) Practical Protein Crystallography, Second Edition Academic Press; Drenth (1999) Principles of Protein X-Ray Crystallography (Springer Advanced Texts in Chemistry) Springer-Verlag; Fanchon and Hendrickson (1991) Chapter 15 of Crystallographic Computing, Volume 5 IUCr/Oxford University Press; Murthy (1996) Chapter 5 of Crystallographic Methods and Protocols Humana Press; Dauter et al. (2000) "Novel approach to phasing proteins: derivatization by short cryo-soaking with halides" Acta Cryst.D56:232-237; Dauter (2002) "New approaches to high-throughput phasing" Curr. Opin. Structural Biol. 12:674–678; Chen et al. (1991) "Crystal structure of a bovine neurophysin-II dipeptide complex at 2.8 Å determined from the single-wavelength anomalous scattering signal of an incorporated iodine atom" Proc. Natl Acad. Sci. USA, 88:4240–4244; and Gavira et al. (2002) "*Ab initio*

crystallographic structure determination of insulin from protein to electron density without crystal handling" Acta Cryst.D58:1147-1154.

[0061] In addition, a variety of programs to facilitate data collection, phase determination, model building and refinement, and the like are publicly available. Examples include, but are not limited to, the HKL2000 package (Otwinowski and Minor (1997) "Processing of X-ray Diffraction Data Collected in Oscillation Mode" Methods in Enzymology 276:307-326), the CCP4 package (Collaborative Computational Project (1994) "The CCP4 suite: programs for protein crystallography" Acta Crystallogr D 50:760-763), SOLVE and RESOLVE (Terwilliger and Berendzen (1999) Acta Crystallogr D 55 (Pt 4):849-861), SHELXS and SHELXD (Schneider and Sheldrick (2002) "Substructure solution with SHELXD" Acta Crystallogr D Biol Crystallogr 58:1772-1779), Refmac5 (Murshudov et al. (1997) "Refinement of Macromolecular Structures by the Maximum-Likelihood Method" Acta Crystallogr D 53:240-255), PRODRG (van Aalten et al. (1996) "PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules" J Comput Aided Mol Des 10:255-262), and O (Jones et al. (1991) "Improved methods for building protein models in electron density maps and the location of errors in these models" Acta Crystallogr A 47 (Pt 2):110-119).

[0062] Techniques for structure determination by NMR spectroscopy are similarly well described in the literature. See, e.g., Cavanagh et al. (1995) Protein NMR Spectroscopy: Principles and Practice, Academic Press; Levitt (2001) Spin Dynamics: Basics of Nuclear Magnetic Resonance, John Wiley & Sons; Evans (1995) Biomolecular NMR Spectroscopy, Oxford University Press; Wüthrich (1986) NMR of Proteins and Nucleic Acids (Baker Lecture Series), Kurt Wiley-Interscience; Neuhaus and Williamson (2000) The Nuclear Overhauser Effect in Structural and Conformational Analysis, 2nd Edition, Wiley-VCH; Macomber (1998) A Complete Introduction to Modern NMR Spectroscopy, Wiley-Interscience; Downing (2004) Protein NMR Techniques (Methods in Molecular Biology), 2nd edition, Humana Press; Clore and Gronenborn (1994) NMR of Proteins (Topics in Molecular and Structural Biology), CRC Press; Reid (1997) Protein NMR Techniques, Humana Press; Krishna and Berliner (2003) Protein NMR for the Millenium (Biological Magnetic Resonance), Kluwer Academic Publishers; Kiihne and De Groot (2001) Perspectives on Solid State NMR in Biology (Focus on Structural Biology, 1),

Kluwer Academic Publishers; Jones et al. (1993) Spectroscopic Methods and Analyses: NMR, Mass Spectrometry, and Related Techniques (Methods in Molecular Biology, Vol. 17), Humana Press; Goto and Kay (2000) Curr. Opin. Struct. Biol. 10:585; Gardner (1998) Annu. Rev. Biophys. Biomol. Struct. 27:357; Wüthrich (2003) Angew. Chem. Int. Ed. 42:3340; Bax (1994) Curr. Opin. Struct. Biol. 4:738; Pervushin et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94:12366; Fiaux et al. (2002) Nature 418:207; Fernandez and Wider (2003) Curr. Opin. Struct. Biol. 13:570; Ellman et al. (1992) J. Am. Chem. Soc. 114:7959; Wider (2000) BioTechniques 29:1278-1294; Pellecchia et al. (2002) Nature Rev. Drug Discov. (2002) 1:211-219; Arora and Tamm (2001) Curr. Opin. Struct. Biol. 11:540-547; Flaux et al. (2002) Nature 418:207-211; Pellecchia et al. (2001) J. Am. Chem. Soc. 123:4633-4634; and Pervushin et al. (1997) Proc. Natl. Acad. Sci. USA 94:12366-12371.

[0063] The structure of a polymerase with a given nucleotide analogue incorporated into the active site can, as noted, be directly determined, e.g., by x-ray crystallography or NMR spectroscopy, or the structure can be modeled based on the structure of the polymerase and/or a structure of a polymerase with a natural nucleotide bound. The active site region of the polymerase can be identified, for example, by homology with other polymerases, examination of polymerase-template or polymerase-nucleotide co-complexes, biochemical analysis of mutant polymerases, and/or the like. The position of a nucleotide analogue in the active site can be modeled, for example, by projecting the location of non-natural features of the analogue (e.g., additional phosphate or phosphonate groups in the phosphorus containing chain linked to the nucleotide, e.g., tetra, penta or hexa phosphate groups, detectable labeling groups, e.g., fluorescent dyes, or the like) based on the previously determined location of another nucleotide or nucleotide analogue in the active site.

[0064] Such modeling of the nucleotide analogue in the active site can involve simple visual inspection of a model of the polymerase, for example, using molecular graphics software such as the PyMOL viewer (open source, freely available on the World Wide Web at www(dot)pymol(dot)org) or Insight II (commercially available from Accelrys at (www (dot) accelrys (dot) com/products/insight). Alternatively, modeling of the nucleotide analogue in the active site of the polymerase or a putative mutant polymerase, for example, can involve computer-assisted docking, molecular dynamics, free energy minimization, and/or like calculations. Such modeling techniques have been well described

in the literature; see, e.g., Babine and Abdel-Meguid (eds.) (2004) Protein Crystallography in Drug Design, Wiley-VCH, Weinheim; Lyne (2002) "Structure-based virtual screening: An overview" Drug Discov. Today 7:1047-1055; Molecular Modeling for Beginners, at (www (dot) usm (dot) maine (dot) edu/~rhodes/SPVTut/index (dot) html; and Methods for Protein Simulations and Drug Design at (www (dot) dddc (dot) ac (dot) cn/embo04; and references therein. Software to facilitate such modeling is widely available, for example, the CHARMm simulation package, available academically from Harvard University or commercially from Accelrys (at www (dot) accelrys (dot) com), the Discover simulation package (included in Insight II, *supra*), and Dynama (available at (www(dot) cs (dot) gsu (dot) edu/~cscrwh/progs/progs (dot) html). See also an extensive list of modeling software at (www (dot) netsci (dot) org/Resources/Software/Modeling/MMMD/top (dot) html.

[0065] Visual inspection and/or computational analysis of a polymerase model can identify relevant features of the active site region, including, for example, residues that can sterically inhibit entry of a nucleotide analogue into the active site (e.g., residues undesirably close to the projected location of one or more atoms within the analogue when the analogue is bound to the polymerase). Such a residue can, for example, be deleted or replaced with a residue having a smaller side chain; for example, many residues can be conveniently replaced with a residue having similar characteristics but a shorter amino acid side chain, or, e.g., with alanine. Similarly, residues that can be altered to introduce desirable interactions with the nucleotide analogue can be identified. Such a residue can be replaced with a residue that is complementary with a non-natural feature of the analogue, for example, with a residue that can hydrogen bond to the analogue (e.g., serine, threonine, histidine, asparagine, or glutamine), a hydrophobic residue that can interact with a hydrophobic group on the analogue, an aromatic residue that can provide favorable hydrophobic interactions with a group on the analogue (e.g., a fluorophore), an aromatic residue that can engage in a π-π or edge-face stacking interaction with an aromatic group in the analogue, a residue that can engage in a cation-π interaction with the analogue, or a charged residue (e.g., aspartic or glutamic acid, or lysine, arginine, or histidine) that can electrostatically interact with an oppositely charged moiety on the analogue (e.g., an additional phosphate group).

[0066] As just one specific example of such structure-based design, inspection of a model of the Φ29 polymerase identified the Δ505-525 domain and residues K135, E486,

and K512 as potentially sterically inhibiting entry of an analogue into the active site, and suggested that mutation of E375 to histidine, lysine, or arginine would introduce a positive charge complementary to a non-natural tetra phosphate on the analogue. Similarly, inspection of the model suggested that mutation of E375 to an aromatic residue such as tryptophan, tyrosine, or phenylalanine would improve hydrophobic interactions with a fluorophore on the analogue. *See* Examples 2 and 3 below for additional details.

[0067] Thus, in addition to methods of using the polymerases and other compositions herein, the present invention also includes methods of making the polymerases. As described, methods of making a recombinant DNA polymerase can include structurally modeling a first polymerase, e.g., using any available crystal structure and molecular modeling software or system. Based on the modeling, one or more steric inhibition feature or complementarity feature affecting nucleotide access to the active site and/or binding of a nucleotide analogue within the active site region is identified, e.g., in the active site or proximal to it. The first DNA polymerase is mutated to reduce or remove at least one steric inhibition feature or to add the complementarity feature.

Mutating active site regions

[0068] Various types of mutagenesis are optionally used in the present invention, e.g., to modify polymerases to produce variants comprising complementarity features and or to reduce steric hindrance features, e.g., in accordance with polymerase models and model predictions as discussed above. In general, any available mutagenesis procedure can be used for making such mutants. Such mutagenesis procedures optionally include selection of mutant nucleic acids and polypeptides for one or more activity of interest (e.g., improved K_m, V_{max}, k_{cat} etc., for a nucleotide analogue). Procedures that can be used include, but are not limited to: site-directed point mutagenesis, random point mutagenesis, in vitro or in vivo homologous recombination (DNA shuffling), mutagenesis using uracil containing templates, oligonucleotide-directed mutagenesis, phosphorothioate-modified DNA mutagenesis, mutagenesis using gapped duplex DNA, point mismatch repair, mutagenesis using repair-deficient host strains, restriction-selection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, degenerate PCR, double-strand break repair, and many others known to persons of skill.

[0069] Optionally, mutagenesis can be guided by known information from a naturally occurring polymerase molecule, or of a known altered or mutated polymerase

(e.g., using an existing mutant polymerase that displays reduced exonuclease activity), e.g., sequence, sequence comparisons, physical properties, crystal structure and/or the like as discussed above. However, in another class of embodiments, modification can be essentially random (e.g., as in classical DNA shuffling).

[0070] Additional information on mutation formats is found in: Sambrook et al., Molecular Cloning - A Laboratory Manual (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2000 ("Sambrook"); Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2006) ("Ausubel")) and PCR Protocols A Guide to Methods and Applications (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) (Innis). The following publications and references cited within provide still additional detail on mutation formats: Arnold, *Protein engineering for unusual environments*, Current Opinion in Biotechnology 4:450-455 (1993); Bass et al., *Mutant Trp repressors with new DNA-binding specificities*, Science 242:240-245 (1988); Botstein & Shortle, *Strategies and applications of in vitro mutagenesis*, Science 229:1193-1201(1985); Carter et al., *Improved oligonucleotide site-directed mutagenesis using M13 vectors*, Nucl. Acids Res. 13: 4431-4443 (1985); Carter, *Site-directed mutagenesis*, Biochem. J. 237:1-7 (1986); Carter, *Improved oligonucleotide-directed mutagenesis using M13 vectors*, Methods in Enzymol. 154: 382-403 (1987); Dale et al., *Oligonucleotide-directed random mutagenesis using the phosphorothioate method*, Methods Mol. Biol. 57:369-374 (1996); Eghtedarzadeh & Henikoff, *Use of oligonucleotides to generate large deletions*, Nucl. Acids Res. 14: 5115 (1986); Fritz et al., *Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro*, Nucl. Acids Res. 16: 6987-6999 (1988); Grundström et al., *Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis*, Nucl. Acids Res. 13: 3305-3316 (1985); Kunkel, *The efficiency of oligonucleotide directed mutagenesis*, in Nucleic Acids & Molecular Biology (Eckstein, F. and Lilley, D.M.J. eds., Springer Verlag, Berlin)) (1987); Kunkel, *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Proc. Natl. Acad. Sci. USA 82:488-492 (1985); Kunkel et al., *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Methods in Enzymol. 154, 367-382 (1987); Kramer et al., *The gapped duplex DNA approach to oligonucleotide-directed mutation construction*, Nucl. Acids Res. 12: 9441-9456 (1984); Kramer & Fritz

Oligonucleotide-directed construction of mutations via gapped duplex DNA, Methods in Enzymol. 154:350-367 (1987); Kramer et al., *Point Mismatch Repair*, Cell 38:879-887 (1984); Kramer et al., *Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations*, Nucl. Acids Res. 16: 7207 (1988); Ling et al., *Approaches to DNA mutagenesis: an overview*, Anal Biochem. 254(2): 157-178 (1997); Lorimer and Pastan Nucleic Acids Res. 23, 3067-8 (1995); Mandecki, *Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis*, Proc. Natl. Acad. Sci. USA, 83:7177-7181 (1986); Nakamaye & Eckstein, *Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis*, Nucl. Acids Res. 14: 9679-9698 (1986); Nambiar et al., *Total synthesis and cloning of a gene coding for the ribonuclease S protein*, Science 223: 1299-1301 (1984); Sakamar and Khorana, *Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)*, Nucl. Acids Res. 14: 6361-6372 (1988); Sayers et al., *Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis*, Nucl. Acids Res. 16:791-802 (1988); Sayers et al., *Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide*, (1988) Nucl. Acids Res. 16: 803-814; Sieber, et al., Nature Biotechnology, 19:456-460 (2001); Smith, *In vitro mutagenesis*, Ann. Rev. Genet. 19:423-462(1985); Methods in Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); Stemmer, Nature 370, 389-91 (1994); Taylor et al., *The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA*, Nucl. Acids Res. 13: 8749-8764 (1985); Taylor et al., *The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA*, Nucl. Acids Res. 13: 8765-8787 (1985); Wells et al., *Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin*, Phil. Trans. R. Soc. Lond. A 317: 415-423 (1986); Wells et al., *Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites*, Gene 34:315-323 (1985); Zoller & Smith, *Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment*, Nucleic Acids Res. 10:6487-6500 (1982); Zoller & Smith, *Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors*, Methods in Enzymol. 100:468-500 (1983); and Zoller & Smith, *Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded*

DNA template, Methods in Enzymol. 154:329-350 (1987). Additional details on many of the above methods can be found in Methods in Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

Determining Kinetic Parameters

[0071] The polymerases of the invention can be screened or otherwise tested to determine whether the polymerase displays a modified activity for or with a nucleotide analogue as compared to the first DNA polymerase (e.g., a corresponding wild-type polymerase from which the recombinant polymerase was derived). For example, k_{cat}, K_m, V_{max}, k_{cat}/K_m, V_{max}/K_m, k_{pol}, and/or K_d of the recombinant DNA polymerase for the nucleotide analogue can be determined. Further, k_{cat}, K_m, V_{max}, V_{max}/K_m, k_{cat}/K_m, k_{pol}, and/or K_d of the recombinant DNA polymerase for a natural nucleotide can also be determined (e.g., where the polymerase desirably includes both analogue and natural nucleotide incorporation activity).

[0072] As is well-known in the art, for enzymes obeying simple Michaelis-Menten kinetics, kinetic parameters are readily derived from rates of catalysis measured at different substrate concentrations. The Michaelis-Menten equation, $V=V_{max}[S]([S]+K_m)^{-1}$, relates the concentration of uncombined substrate ([S], approximated by the total substrate concentration), the maximal rate (V_{max}, attained when the enzyme is saturated with substrate), and the Michaelis constant (K_m, equal to the substrate concentration at which the reaction rate is half of its maximal value), to the reaction rate (V).

[0073] For many enzymes, K_m is equal to the dissociation constant of the enzyme-substrate complex and is thus a measure of the strength of the enzyme-substrate complex. For such an enzyme, in a comparison of K_ms, a lower K_m represents a complex with stronger binding, while a higher Km represents a complex with weaker binding. The ratio k_{cat}/K_m, sometimes called the specificity constant, represents the apparent rate constant for combination of substrate with free enzyme. The larger the specificity constant, the more efficient the enzyme is in binding the substrate and converting it to product.

[0074] The k_{cat} (also called the turnover number of the enzyme) can be determined if the total enzyme concentration ($[E_T]$, i.e., the concentration of active sites) is known, since $V_{max}=k_{cat}[E_T]$. For situations in which the total enzyme concentration is difficult to measure, the ratio V_{max}/K_m is often used instead as a measure of efficiency. K_m and V_{max}

can be determined, for example, from a Lineweaver-Burk plot of 1/V against 1/[S], where the y intercept represents $1/V_{max}$, the x intercept $-1/K_m$, and the slope K_m/V_{max}, or from an Eadie-Hofstee plot of V against V/[S], where the y intercept represents V_{max}, the x intercept V_{max}/K_m, and the slope $-K_m$. Software packages such as KinetAsyst™ or Enzfit (Biosoft, Cambridge, UK) can facilitate the determination of kinetic parameters from catalytic rate data.

[0075] For enzymes such as polymerases that have multiple substrates, varying the concentration of only one substrate while holding the others in suitable excess (e.g., effectively constant) concentration typically yields normal Michaelis-Menten kinetics.

[0076] In one embodiment, using presteady-state kinetics, the nucleotide concentration dependence of the rate k_{obs} (the observed first-order rate constant for dNTP incorporation) provides an estimate of the K_m for a ground state binding and the maximum rate of polymerization (k_{pol}). The k_{obs} is measured using a burst assay. The results of the assay are fitted with the burst equation; Product = $A[1-\exp(-k_{obs}*t)]+k_{ss}*t$ where A represents amplitude an estimate of the concentration of the enzyme active sites, k_{ss} is the observed steady-state rate constant and t is the reaction incubation time. The K_m for dNTP binding to the polymerase-DNA complex and the k_{pol} are calculated by fitting the dNTP concentration dependent change in the k_{obs} using the equation $k_{obs} = (k_{pol}*[S])*(K_m+[S])-1$ where [S] is the substrate concentration. Results are optionally obtained from a rapid-quench experiment (also called a quench-flow measurement), for example, based on the methods described in Johnson (1986) "Rapid kinetic analysis of mechanochemical adenosinetriphosphatases" Methods Enzymol. 134:677-705, Patel et al. (1991) "Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant" Biochemistry 30(2):511-25, and Tsai and Johnson (2006) "A new paradigm for DNA polymerase specificity" Biochemistry 45(32):9675-87.

[0077] Parameters such as rate of binding of a nucleotide analogue by the recombinant polymerase, rate of product release by the recombinant polymerase, or branching rate of the recombinant polymerase (the "branching rate" is the rate of dissociation of a nucleotide or nucleotide analogue from the polymerase active site without incorporation of the nucleotide or nucleotide analogue, where the nucleotide or nucleotide analogue if it were incorporated would correctly base-pair with a complementary nucleotide or nucleotide analogue in the template) can also be determined, and optionally compared to

that of the first polymerase (e.g., a corresponding wild-type polymerase). See, e.g., Example 3 herein.

[0078] For a more thorough discussion of enzyme kinetics, see, e.g., Berg, Tymoczko, and Stryer (2002) Biochemistry, Fifth Edition, W. H. Freeman; Creighton (1984) Proteins: Structures and Molecular Principles, W. H. Freeman; and Fersht (1985) Enzyme Structure and Mechanism, Second Edition, W. H. Freeman.

[0079] As discussed above, the relevant DNA polymerase has a modified active site region that is homologous to a wild-type active site region of a wild-type DNA polymerase e.g., that includes one or more structural modification relative to the wild type active site region that increases the relative activity of the enzyme to one or more of natural nucleotides and/or nucleotide analogues, with increases in activity to nucleotide analogues being a preferred goal. In at least one aspect, without being bound to any particular theory of operation, the modifications are targeted to reduce steric inhibition for entry of the nucleotide analogue into the modified active site and/or that is complementary with one or more non-natural features of the nucleotide analogue. A K_m value of the recombinant polymerase for the nucleotide analogue is typically lower than a K_m for a corresponding homologous wild-type polymerase for the nucleotide analogue.

[0080] In one aspect, the improved activity of the enzymes of the invention is measured with reference to a model analogue or analogue set and compared with a given parental enzyme. For example, in the case of enzymes derived from a Φ29 parental enzyme, an improved enzyme of the invention would have a lower Km than the parental enzyme, e.g., wild type Φ29 or N62D Φ29, toward a given analogue. In general, for purposes of discussion, examples of improved enzymes of the invention will be characterizable as having lower K_ms toward A488dC4P and/or A568dC4P, two analogs that have been reasonably well processed and reasonably poorly processed by Φ29 derived enzymes, respectively, that are, e.g., from about 5% or less to about 90% or less of the Km possessed by N62D Φ29 toward the same analogs. For example, as set forth in more detail in the examples below, e.g., at Table 2, His-375H-N62D Φ29 displays a K_m that is about 40% of K_m of N62D Φ29 for A488dC4P, while His-375S-N62D Φ29 displays a K_m that is about 75% of the K_m of N62D Φ29 for A488dC4P. Similarly, His-375H-N62D Φ29 displays a K_m that is about 15% of the K_m of N62D Φ29 for A568dC4P, while His-375S-N62D Φ29 displays a K_m that is about 38% of the K_m of N62D Φ29 for A568dC4P. While

the foregoing may be used as a characterization tool, it in no way is intended as a specifically limiting reaction of the invention.

Screening Polymerases

[0081] Screening or other protocols can be used to determine whether a polymerase displays a modified activity for a nucleotide analogue as compared to the first DNA polymerase. For example, k_{cat}, K_m, V_{max}, or k_{cat}/K_m of the recombinant DNA polymerase for the nucleotide analogue can be determined as discussed above. Further, k_{cat}, K_m, V_{max}, or k_{cat}/K_m of the recombinant DNA polymerase for a natural nucleotide can also be similarly determined (e.g., where the polymerase desirably includes both analogue and natural nucleotide incorporation activity).

[0082] In one desirable aspect, a library of recombinant DNA polymerases can be made and screened for these properties. For example, a plurality of members of the library can be made to include one or more putative steric inhibition feature mutation an/or a mutation to putatively produce complementary with one or more non-natural features of the nucleotide analogue, that is then screened for the properties of interest. In general, the library can be screened to identify at least one member comprising a modified activity of interest.

[0083] Libraries of polymerases can be either physical or logical in nature. Moreover, any of a wide variety of library formats can be used. For example, polymerases can be fixed to solid surfaces in arrays of proteins. Similarly, liquid phase arrays of polymerases (e.g., in microwell plates) can be constructed for convenient high-throughput fluid manipulations of solutions comprising polymerases. Liquid, emulsion, or gel-phase libraries of cells that express recombinant polymerases can also be constructed, e.g., in microwell plates, or on agar plates. Phage display libraries of polymerases or polymerase domains (e.g., including the active site region) can be produced. Instructions in making and using libraries can be found, e.g., in Sambrook, Ausubel and Berger, referenced herein.

[0084] For the generation of libraries involving fluid transfer to or from microtiter plates, a fluid handling station is optionally used. Several "off the shelf" fluid handling stations for performing such transfers are commercially available, including e.g., the Zymate systems from Caliper Life Sciences (Hopkinton, MA) and other stations which utilize automatic pipettors, e.g., in conjunction with the robotics for plate movement (e.g.,

the ORCA® robot, which is used in a variety of laboratory systems available, e.g., from Beckman Coulter, Inc. (Fullerton, CA).

[0085] In an alternate embodiment, fluid handling is performed in microchips, e.g., involving transfer of materials from microwell plates or other wells through microchannels on the chips to destination sites (microchannel regions, wells, chambers or the like). Commercially available microfluidic systems include those from Hewlett-Packard/Agilent Technologies (e.g., the HP2100 bioanalyzer) and the Caliper High Throughput Screening System. The Caliper High Throughput Screening System provides one example interface between standard microwell library formats and Labchip technologies . Furthermore, the patent and technical literature includes many examples of microfluidic systems which can interface directly with microwell plates for fluid handling.

Desirable Properties

[0086] The polymerases of the invention can include any of a variety of modified properties towards natural or nucleotide analogues or analogues, depending on the application, including increased speed, increased retention time (or decreased speed) for incorporated bases, greater processivity, etc. For example, where a higher level of nucleotide analogue incorporation is desired, the polymerase of the invention is selected to have a lower K_m, a higher V-max and/or a higher k_{cat} than a corresponding homologous wild-type polymerase with respect to a given nucleotide analogue. In certain embodiments, it is desirable to slow or quicken the overall nucleotide incorporation speed of the polymerase (e.g., depending on the resolution of the equipment used to monitor incorporation), or to improve processivity, specificity, or the like. In certain embodiments, the recombinant polymerase has an increased rate of binding of a nucleotide analogue, an increased rate of product release, and/or a decreased branching rate, as compared to a corresponding homologous wild-type polymerase. Any of these features can be screened for or against in selecting a polymerase of the invention.

[0087] For example, the polymerases of the invention can typically incorporate natural nucleotides (e.g., A, C, G and T) into a growing copy nucleic acid. For example, the recombinant polymerase optionally displays a specific activity for a natural nucleotide that is at least about 5% as high (e.g., 5%, 10%, 25%, 50%, 75%, 100% or higher) as a corresponding homologous wild-type polymerase and a processivity with natural nucleotides in the presence of a template that is at least 5% as high (e.g., 5%, 10%, 25%,

50%, 75%, 100% or higher) as the wild-type polymerase in the presence of the natural nucleotide. Optionally, the recombinant polymerase also displays a k_{cat}/K_m or V_{max}/K_m for a naturally occurring nucleotide that is at least about 10% as high (e.g., 10%, 25%, 50%, 75% or 100% or higher) as the wild-type polymerase.

Additional Example Details

[0088] A number of specific examples of modified active site regions are described herein. An "active site region" is a portion of the polymerase that includes or is proximal to the active site (e.g., within about 2 nm of the active site) in a three dimensional structure of a folded polymerase. Specific examples of structural modifications within or proximal to the active site of Φ29 DNA polymerase are described herein. For example, relative to a wild-type Φ29 DNA polymerase, these modification can include any of: a deletion of Δ505-525, a deletion within the Δ505-525 domain, a K135A mutation, an L384R mutation (e.g., in combination with another mutation herein), an E375H mutation, an E375S mutation, an E375K mutation, an E375R mutation, an E375A mutation, an E375Q mutation, an E375W mutation, an E375Y mutation, an E375F mutation, an E486A mutation, an E486D mutation, a K512A mutation, a mutation listed in Table 8, and combinations thereof. For example, the polymerase can include a combination of mutations selected from the list of combinations in Table 8.

[0089] The polymerase optionally further includes one or more mutations/deletions relative to the wild-type polymerase that reduce or eliminate endogenous exonuclease activity. For example, relative to the wild-type Φ29 DNA polymerase, N62 is optionally mutated or deleted to reduce exonuclease activity; e.g., the polymerase can include an N62D mutation. Other example mutations that reduce exonuclease activity include D12A, T15I, E14I, and/or D66A; accordingly, the polymerases of the invention optionally comprise one or more of these mutations.

[0090] As will be appreciated, the numbering of amino acid residues is with respect to the wild-type sequence of the Φ29 polymerase, and actual position within a molecule of the invention may vary based upon the nature of the various modifications that the enzyme includes relative to the wild type Φ29 enzyme, e.g., deletions and/or additions to the molecule, either at the termini or within the molecule itself.

Affinity Tags And Other Optional Polymerase Features

[0091] The recombinant DNA polymerase optionally includes additional features exogenous or heterologous to the polymerase. For example, the recombinant polymerase optionally includes one or more exogenous affinity tags, e.g., purification or substrate binding tags, such as a 6 His tag sequence, a GST tag, an HA tag sequence, a plurality of 6 His tag sequences, a plurality of GST tags, a plurality of HA tag sequences, a SNAP-tag, or the like. These and other features useful in the context of binding a polymerase to a surface are optionally included, e.g., to orient and/or protect the polymerase active site when the polymerase is bound to a surface. Other useful features include recombinant dimer domains of the enzyme, and, e.g., large extraneous polypeptide domains coupled to the polymerase distal to the active site. For example, for Φ29, the active site is in the C terminal region of the protein, and added surface binding elements (extra domains, His tags, etc.) are typically located in the N-terminal region to avoid interfering with the active site when the polymerase is coupled to a surface.

[0092] In general, surface binding elements and purification tags that can be added to the polymerase (recombinantly or, e.g., chemically) include, e.g., polyhistidine tags, HIS-6 tags, biotin, avidin, GST sequences, BiTag sequences, S tags, SNAP-tags, enterokinase sites, thrombin sites, antibodies or antibody domains, antibody fragments, antigens, receptors, receptor domains, receptor fragments, ligands, dyes, acceptors, quenchers, or combinations thereof.

[0093] Multiple surface binding domains can be added to orient the polypeptide relative to a surface and/or to increase binding of the polymerase to the surface. By binding a surface at two or more sites, through two or more separate tags, the polymerase is held in a relatively fixed orientation with respect to the surface. Additional details on fixing a polymerase to a surface are found in U.S. patent application 60/753,446 "PROTEIN ENGINEERING STRATEGIES TO OPTIMIZE ACTIVITY OF SURFACE ATTACHED PROTEINS" by Hanzel et al. and U.S. patent application 60/753,515"ACTIVE SURFACE COUPLED POLYMERASES" by Hanzel et al., both filed December 22, 2005 and incorporated herein by reference for all purposes, and in Attorney Docket number 105-001210US "PROTEIN ENGINEERING STRATEGIES TO OPTIMIZE ACTIVITY OF SURFACE ATTACHED PROTEINS" by Hanzel et al. and Attorney docket 105-00810US

"ACTIVE SURFACE COUPLED POLYMERASES" by Hanzel et al. both co-filed herewith and incorporated herein by reference for all purposes.

APPLICATIONS FOR ENHANCED INCORPORATION OF NUCLEOTIDE ANALOGUES BY A DNA POLYMERASE

[0094] Polymerases of the invention, natural and/or nucleotide analogues and nucleic acid templates (DNA or RNA) are optionally used to copy the template nucleic acid. That is, a mixture of the polymerase, nucleotide analogues, and optionally natural nucleotides and other reagents, the template and a replication initiating moiety is reacted such that the polymerase extends the primer in a template-dependent manner. The moiety can be a standard oligonucleotide primer, or, alternatively, a component of the template, e.g., the template can be a self-priming single stranded DNA, a nicked double stranded DNA, or the like. Similarly, a terminal protein can serve as a initiating moiety. At least one nucleotide analogue can be incorporated into the DNA. The template DNA can be a linear or circular DNA, and in certain applications, is desirably a circular template (e.g., for rolling circle replication or for sequencing of circular templates). Optionally, the composition can be present in an automated DNA replication and/or sequencing system.

[0095] Incorporation of labeled nucleotide analogues by the polymerases of the invention are particularly useful in a variety of different nucleic acid analyses, including real-time monitoring of DNA polymerization. The label can itself be incorporated, or more preferably, can be released during incorporation. For example, analogue incorporation can be monitored in real-time by monitoring label release during incorporation of the analogue by the polymerase. The portion of the analogue that is incorporated can be the same as a natural nucleotide, or can include features of the analogue that differ from a natural nucleotide.

[0096] In general, label incorporation or release can be used to indicate the presence and composition of a growing nucleic acid strand, e.g., providing evidence of template replication/amplification and/or sequence of the template. Signaling from the incorporation can be the result of detecting labeling groups that are liberated from the incorporated analogue, e.g., in a solid phase assay, or can arise upon the incorporation reaction. For example, in the case of FRET labels where a bound label is quenched and a free label is not, release of a label group from the incorporated analogue can give rise to a fluorescent signal.

Alternatively, the enzyme may be labeled with one member of a FRET pair proximal to the active site, and incorporation of an analogue bearing the other member will allow energy transfer upon incorporation. The use of enzyme bound FRET components in nucleic acid sequencing applications is described, e.g., in Published U.S. Patent application No. 2003-0044781, incorporated herein by reference.

[0097] In one example reaction of interest, a polymerase reaction can be isolated within an extremely small observation volume that effectively results in observation of individual polymerase molecules. As a result, the incorporation event provides observation of an incorporating nucleotide analogue that is readily distinguishable from non-incorporated nucleotide analogs. In a preferred aspect, such small observation volumes are provided by immobilizing the polymerase enzyme within an optical confinement, such as a Zero Mode Waveguide. For a description of ZMWs and their application in single molecule analyses, and particularly nucleic acid sequencing, see, e.g., Published U.S. Patent Application No. 2003/0044781, and U.S. Patent No. 6,917,726, each of which is incorporated herein by reference in its entirety for all purposes.

[0098] In general, a polymerase enzyme is complexed with the template strand in the presence of one or more nucleotides and/or one or more nucleotide analogue of the invention. For example, in certain embodiments, labeled analogues are present representing analogous compounds to each of the four natural nucleotides, A, T, G and C, e.g., in separate polymerase reactions, as in classical Sanger sequencing, or multiplexed together in a single reaction, as in multiplexed sequencing approaches. When a particular base in the template strand is encountered by the polymerase during the polymerization reaction, it complexes with an available analogue that is complementary to such nucleotide, and incorporates that analogue into the nascent and growing nucleic acid strand. In one aspect, incorporation can result in a label being released, e.g., in polyphosphate analogues, cleaving between the α and β phosphorus atoms in the analogue, and consequently releasing the labeling group (or a portion thereof). The incorporation event is detected, either by virtue of a longer presence of the analogue and, thus, the label, in the complex, or by virtue of release of the label group into the surrounding medium. Where different labeling groups are used for each of the types of analogs, e.g., A, T, G or C, identification of a label of an incorporated analogue allows identification of that analogue and consequently, determination of the complementary nucleotide in the template strand being processed at

that time. Sequential reaction and monitoring permits a real-time monitoring of the polymerization reaction and determination of the sequence of the template nucleic acid. As noted above, in particularly preferred aspects, the polymerase enzyme/template complex is provided immobilized within an optical confinement that permits observation of an individual complex, e.g., a Zero Mode Waveguide. In addition to their use in sequencing, the analogs of the invention are also equally useful in a variety of other genotyping analyses, e.g., SNP genotyping using single base extension methods, real time monitoring of amplification, e.g., RT-PCR methods, and the like.

[0099] Further details regarding sequencing and nucleic acid amplification can be found in Sambrook et al., Molecular Cloning - A Laboratory Manual (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2000 ("Sambrook"); Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2006) ("Ausubel")) and PCR Protocols A Guide to Methods and Applications (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) ("Innis").

MAKING AND ISOLATING RECOMBINANT POLYMERASES

[0100] Generally, nucleic acids encoding a polymerase of the invention can be made by cloning, recombination, in vitro synthesis, in vitro amplification and/or other available methods. A variety of recombinant methods can be used for expressing an expression vector that encodes a polymerase of the invention, e.g., a mutant polymerase that, without being bound to a particular theory, reduces steric hindrance for a nucleotide analogue of the invention and/or that includes a complementarity feature. Recombinant methods for making nucleic acids, expression and isolation of expressed products are described, e.g., in Sambrook, Ausubel and Innis.

[0101] In addition, a plethora of kits are commercially available for the purification of plasmids or other relevant nucleic acids from cells, (*see*, e.g., EasyPrep™, FlexiPrep™, both from Pharmacia Biotech; StrataClean™, from Stratagene; and, QIAprep™ from Qiagen). Any isolated and/or purified nucleic acid can be further manipulated to produce other nucleic acids, used to transfect cells, incorporated into related vectors to infect organisms for expression, and/or the like. Typical cloning vectors contain transcription and translation terminators, transcription and translation initiation sequences, and promoters

useful for regulation of the expression of the particular target nucleic acid. The vectors optionally comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the cassette in eukaryotes, or prokaryotes, or both, (e.g., shuttle vectors) and selection markers for both prokaryotic and eukaryotic systems. Vectors are suitable for replication and integration in prokaryotes, eukaryotes, or both. *See,* Giliman & Smith, *Gene* 8:81 (1979); Roberts, *et al.*, *Nature*, 328:731 (1987); Schneider, B., *et al.*, *Protein Expr. Purif.* 6435:10 (1995); Ausubel, Sambrook, Berger (*above*). A catalogue of Bacteria and Bacteriophages useful for cloning is provided, e.g., by the ATCC, e.g., *The ATCC Catalogue of Bacteria and Bacteriophage* published yearly by the ATCC. Additional basic procedures for sequencing, cloning and other aspects of molecular biology and underlying theoretical considerations are also found in Watson *et al.* (1992) *Recombinant DNA Second Edition, Scientific American Books, NY.*

[0102] In addition, systems of orthogonal components are available that can incorporate any of a variety of unnatural amino acids into a recombinant protein (e.g., polymerase of the invention). In brief, a cell or other translation system (e.g., an in vitro translation system) is constructed that includes an orthogonal tRNA ("OtRNA"; a tRNA not recognized by the cell's endogenous translation machinery, such as an amber or 4-base tRNA) and an orthogonal tRNA synthetase ("ORS"; this is a synthetase that does not aminoacylate any endogenous tRNA of the cell, but which can aminoacylate the OtRNA in response to a selector codon). A nucleic acid encoding the enzyme is constructed to include a selector codon at a selected site that is specifically recognized by the OtRNA. The ORS specifically incorporates an unnatural amino acid with a desired chemical functionality at one or more selected site(s) (e.g., distal to the active site). This chemical functional group can be unique as compared to those ordinarily found on amino acids, e.g., that incorporate keto or other functionalities. Further information on orthogonal systems can be found, e.g., in Wang et al., (2001), Science 292:498-500; Chin et al., (2002) Journal of the American Chemical Society 124:9026-9027; Chin and Schultz, (2002), ChemBioChem 11:1135-1137; Chin, et al., (2002), PNAS United States of America 99:11020-11024; and Wang and Schultz, (2002), Chem. Comm., 1-10. See also, International Publications WO 2002/086075, entitled "METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA AMINOACYL-tRNA SYNTHETASE PAIRS;" WO 2002/085923, entitled "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS;" WO

2004/094593, entitled "EXPANDING THE EUKARYOTIC GENETIC CODE;" WO 2005/019415, filed July 7, 2004; WO 2005/007870, filed July 7, 2004; and WO 2005/007624, filed July 7, 2004.

[0103] Other useful references, e.g. for cell isolation and culture (e.g., for subsequent nucleic acid isolation) include Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, third edition, Wiley- Liss, New York and the references cited therein; Payne *et al.* (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY; Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL.

[0104] In addition, essentially any nucleic acid can be custom or standard ordered from any of a variety of commercial sources, such as Operon Technologies Inc. (Alameda, CA).

[0105] A variety of protein isolation and detection methods are known and can be used to isolate polymerases, e.g., from recombinant cultures of cells expressing the recombinant polymerases of the invention. A variety of protein isolation and detection methods are well known in the art, including, e.g., those set forth in R. Scopes, Protein Purification, Springer-Verlag, N.Y. (1982); Deutscher, Methods in Enzymology Vol. 182: Guide to Protein Purification, Academic Press, Inc. N.Y. (1990); Sandana (1997) Bioseparation of Proteins, Academic Press, Inc.; Bollag et al. (1996) Protein Methods, 2nd Edition Wiley-Liss, NY; Walker (1996) The Protein Protocols Handbook Humana Press, NJ, Harris and Angal (1990) Protein Purification Applications: A Practical Approach IRL Press at Oxford, Oxford, England; Harris and Angal Protein Purification Methods: A Practical Approach IRL Press at Oxford, Oxford, England; Scopes (1993) Protein Purification: Principles and Practice 3rd Edition Springer Verlag, NY; Janson and Ryden (1998) Protein Purification: Principles, High Resolution Methods and Applications, Second Edition Wiley-VCH, NY; and Walker (1998) Protein Protocols on CD-ROM Humana Press, NJ; and the references cited therein. Additional details regarding protein purification and detection methods can be found in Satinder Ahuja ed., Handbook of Bioseparations, Academic Press (2000).

KITS

[0106] The present invention also provides kits that incorporate the polymerases of the invention, e.g., with one or more nucleotide analogues, e.g., for sequencing, nucleic acid amplification, or the like. Such kits can include the polymerase of the invention packaged in a fashion to enable use of the polymerase, a set of different nucleotide analogs of the invention, e.g., those that are analogous to A, T, G, and C, e.g., where at least one of the analogues bears a detectable moiety, and in preferred aspects more than one, and in many cases, each bears a detectably different labeling group, optionally to permit identification in the presence of the other analogues. Depending upon the desired application, the kits of the invention optionally include additional reagents, such as natural nucleotides, a control template, and other reagents, such as buffer solutions and/or salt solutions, including, e.g., divalent metal ions, i.e., Mg^{++}, Mn^{++} and/or Fe^{++}, standard solutions, e.g., dye standards for detector calibration. Such kits also typically include instructions for use of the compounds and other reagents in accordance with the desired application methods, e.g., nucleic acid sequencing, amplification and the like.

NUCLEIC ACID AND POLYPEPTIDE SEQUENCE AND VARIANTS

[0107] As described herein, the invention provides polynucleotide sequences encoding, e.g., a polymerase as described herein. Examples of polymerase sequences that include steric hindrance or complementarity features are found herein, e.g., in Table 3. However, one of skill in the art will immediately appreciate that the invention is not limited to those sequences. For example, one of skill will appreciate that the invention also provides, e.g., many related sequences with the functions described herein, e.g., polynucleotides and polypeptides encoding conservative variants of a polymerase of Table 3.

[0108] Accordingly, the invention provides a variety of polypeptides (polymerases) and polynucleotides (nucleic acids that encode polymerases). Example polynucleotides of the invention include, e.g., a polynucleotide comprising a nucleotide sequence as set forth in Table 3 or a polynucleotide that is complementary to or that encodes a polynucleotide sequence thereof (e.g., where the given sequence is a DNA, an RNA is one example of a sequence that encodes the DNA, e.g., via reverse transcription). A polynucleotide of the invention also optionally includes any polynucleotide that encodes a polymerase of Table 3.

Because of the degeneracy of the genetic code, many polynucleotides equivalently encode a given polymerase sequence. Similarly, an artificial or recombinant nucleic acid that hybridizes to a polynucleotide indicated above under highly stringent conditions over substantially the entire length of the nucleic acid (and is other than a naturally occurring polynucleotide) is a polynucleotide of the invention. In one embodiment, a composition includes a polypeptide of the invention and an excipient (e.g., buffer, water, pharmaceutically acceptable excipient, etc.). The invention also provides an antibody or antisera specifically immunoreactive with a polypeptide of the invention (e.g., that specifically recognizes an altered steric hindrance or nucleotide analogue complementarity feature.

[0109] In certain embodiments, a vector (e.g., a plasmid, a cosmid, a phage, a virus, etc.) comprises a polynucleotide of the invention. In one embodiment, the vector is an expression vector. In another embodiment, the expression vector includes a promoter operably linked to one or more of the polynucleotides of the invention. In another embodiment, a cell comprises a vector that includes a polynucleotide of the invention.

[0110] One of skill will also appreciate that many variants of the disclosed sequences are included in the invention. For example, conservative variations of the disclosed sequences that yield a functionally similar sequence are included in the invention. Variants of the nucleic acid polynucleotide sequences, wherein the variants hybridize to at least one disclosed sequence, are considered to be included in the invention. Unique subsequences of the sequences disclosed herein, as determined by, e.g., standard sequence comparison techniques, are also included in the invention.

Conservative variations

[0111] Owing to the degeneracy of the genetic code, "silent substitutions" (*i.e.*, substitutions in a nucleic acid sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of *every* nucleic acid sequence that encodes an amino acid sequence. Similarly, "conservative amino acid substitutions," where one or a limited number of amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties, are also readily identified as being highly similar to a disclosed construct. Such conservative variations of each disclosed sequence are a feature of the present invention.

[0112] "Conservative variations" of a particular nucleic acid sequence refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or, where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. One of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 4%, 2% or 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the deletion of an amino acid, addition of an amino acid, or substitution of an amino acid with a chemically similar amino acid, while retaining the relevant reduced steric hindrance or nucleotide analogue complementarity feature (for example, the conservative substitution can be of a residue distal to the active site region). Thus, "conservative variations" of a listed polypeptide sequence of the present invention include substitutions of a small percentage, typically less than 5%, more typically less than 2% or 1%, of the amino acids of the polypeptide sequence, with an amino acid of the same conservative substitution group. Finally, the addition of sequences which do not alter the encoded activity of a nucleic acid molecule, such as the addition of a non-functional or tagging sequence (introns in the nucleic acid, poly His or similar sequences in the encoded polypeptide, etc.), is a conservative variation of the basic nucleic acid or polypeptide.

[0113] In one aspect, the conservative substitution includes one or more deletion or substitution of a residue at an amino acid residue of the polymerase corresponding to amino acid residue 375.

[0114] Conservative substitution tables providing functionally similar amino acids are well known in the art, where one amino acid residue is substituted for another amino acid residue having similar chemical properties (e.g., aromatic side chains or positively charged side chains), and therefore does not substantially change the functional properties of the polypeptide molecule. The following sets forth example groups that contain natural amino acids of like chemical properties, where substitutions within a group is a "conservative substitution".

TABLE A
Conservative Amino Acid Substitutions

Nonpolar and/or Aliphatic Side Chains	Polar, Uncharged Side Chains	Aromatic Side Chains	Positively Charged Side Chains	Negatively Charged Side Chains
Glycine	Serine			
Alanine	Threonine	Phenylalanine	Lysine	Aspartate
Valine	Cysteine	Tyrosine	Arginine	Glutamate
Leucine	Methionine	Tryptophan	Histidine	
Isoleucine	Asparagine			
Proline	Glutamine			

Nucleic Acid Hybridization

[0115] Comparative hybridization can be used to identify nucleic acids of the invention, including conservative variations of nucleic acids of the invention. In addition, target nucleic acids which hybridize to a nucleic acid represented in Table 3 under high, ultra-high and ultra-ultra high stringency conditions, where the nucleic acids are other than a naturally occurring Φ29, or an N62D mutant, are a feature of the invention. Examples of such nucleic acids include those with one or a few silent or conservative nucleic acid substitutions as compared to a given nucleic acid sequence of Table 3.

[0116] A test nucleic acid is said to specifically hybridize to a probe nucleic acid when it hybridizes at least 50% as well to the probe as to the perfectly matched complementary target, i.e., with a signal to noise ratio at least half as high as hybridization of the probe to the target under conditions in which the perfectly matched probe binds to the perfectly matched complementary target with a signal to noise ratio that is at least about 5x-10x as high as that observed for hybridization to any of the unmatched target nucleic acids.

[0117] Nucleic acids "hybridize" when they associate, typically in solution. Nucleic acids hybridize due to a variety of well characterized physico-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* part I chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, New York), as well as in Current Protocols in Molecular Biology,

Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2004) ("Ausubel"); Hames and Higgins (1995) Gene Probes 1 IRL Press at Oxford University Press, Oxford, England, (Hames and Higgins 1) and Hames and Higgins (1995) Gene Probes 2 IRL Press at Oxford University Press, Oxford, England (Hames and Higgins 2) provide details on the synthesis, labeling, detection and quantification of DNA and RNA, including oligonucleotides.

[0118] An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see,* Sambrook, *supra* for a description of SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove background probe signal. An example low stringency wash is 2x SSC at 40°C for 15 minutes. In general, a signal to noise ratio of 5x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

[0119] "Stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and northern hybridizations are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993), *supra.* and in Hames and Higgins, 1 and 2. Stringent hybridization and wash conditions can easily be determined empirically for any test nucleic acid. For example, in determining stringent hybridization and wash conditions, the hybridization and wash conditions are gradually increased (e.g., by increasing temperature, decreasing salt concentration, increasing detergent concentration and/or increasing the concentration of organic solvents such as formalin in the hybridization or wash), until a selected set of criteria are met. For example, in highly stringent hybridization and wash conditions, the hybridization and wash conditions are gradually increased until a probe binds to a perfectly matched complementary target with a signal to noise ratio that is at least 5x as high as that observed for hybridization of the probe to an unmatched target.

[0120] "Very stringent" conditions are selected to be equal to the thermal melting point (T_m) for a particular probe. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the test sequence hybridizes to a perfectly matched probe. For the purposes of the present invention, generally, "highly stringent" hybridization and wash conditions are selected to be about 5° C lower than the T_m for the specific sequence at a defined ionic strength and pH.

[0121] "Ultra high-stringency" hybridization and wash conditions are those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x as high as that observed for hybridization to any of the unmatched target nucleic acids. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio of at least ½ that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-high stringency conditions.

[0122] Similarly, even higher levels of stringency can be determined by gradually increasing the hybridization and/or wash conditions of the relevant hybridization assay. For example, those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x, 20X, 50X, 100X, or 500X or more as high as that observed for hybridization to any of the unmatched target nucleic acids. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio of at least ½ that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-ultra-high stringency conditions.

[0123] Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Unique subsequences

[0124] In some aspects, the invention provides a nucleic acid that comprises a unique subsequence in a nucleic acid that encodes a polymerase of Table 3. The unique subsequence may be unique as compared to a nucleic acid corresponding to wild type Φ29, or to an N62D mutation thereof. Alignment can be performed using, e.g., BLAST set to

default parameters. Any unique subsequence is useful, e.g., as a probe to identify the nucleic acids of the invention.

[0125] Similarly, the invention includes a polypeptide which comprises a unique subsequence in a polymerase of Table 3. Here, the unique subsequence is unique as compared to, e.g., wild type Φ29, or to an N62D mutation thereof.

[0126] The invention also provides for target nucleic acids which hybridize under stringent conditions to a unique coding oligonucleotide which encodes a unique subsequence in a polypeptide selected from the sequences of Table 3, wherein the unique subsequence is unique as compared to a polypeptide corresponding to wild type Φ29, or to an N62D mutation (e.g., parental sequences from which polymerases of the invention were derived, e.g., by mutation). Unique sequences are determined as noted above.

Sequence comparison, identity, and homology

[0127] The terms "identical" or "percent identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (or other algorithms available to persons of skill) or by visual inspection.

[0128] The phrase "substantially identical," in the context of two nucleic acids or polypeptides (*e.g.*, DNAs encoding a polymerase, or the amino acid sequence of a polymerase) refers to two or more sequences or subsequences that have at least about 60%, about 80%, about 90-95%, about 98%, about 99% or more nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm or by visual inspection. Such "substantially identical" sequences are typically considered to be "homologous," without reference to actual ancestry. Preferably, the "substantial identity" exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably, the sequences are substantially identical over at least about 150 residues, or over the full length of the two sequences to be compared.

[0129] Proteins and/or protein sequences are "homologous" when they are derived, naturally or artificially, from a common ancestral protein or protein sequence. Similarly,

nucleic acids and/or nucleic acid sequences are homologous when they are derived, naturally or artificially, from a common ancestral nucleic acid or nucleic acid sequence. Homology is generally inferred from sequence similarity between two or more nucleic acids or proteins (or sequences thereof). The precise percentage of similarity between sequences that is useful in establishing homology varies with the nucleic acid and protein at issue, but as little as 25% sequence similarity over 50, 100, 150 or more residues is routinely used to establish homology. Higher levels of sequence similarity, e.g., 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% or more, can also be used to establish homology. Methods for determining sequence similarity percentages (e.g., BLASTP and BLASTN using default parameters) are described herein and are generally available.

[0130] For sequence comparison and homology determination, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0131] Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally* Current Protocols in Molecular Biology, Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., supplemented through 2004).

[0132] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying

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

[0133] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

COMPUTER-IMPLEMENTED METHODS OF MODELING KINETICS

[0134] In an additional aspect, the invention includes computer-implemented methods, e.g., for modeling enzyme kinetics. In the methods, a plurality of polymerase

state transitions are defined for discrete time steps during a template-based polymerization reaction. In the smallest discrete time step, many polymerase state transitions are forbidden according to the enzymatic kinetics being modeled. A plurality of rate transition rates are defined between the states and a multidimensional probability matrix of possible state transitions is defined for the smallest discrete time step, based upon a given nucleic acid template sequence, nucleotides in a reaction mixture and the polymerase state transitions. The resulting multidimensional probability matrix is stored in a computer readable medium.

[0135] A variety of features of the method can vary. For example, the polymerase state transitions are optionally user-selectable. The transition rates between the states optionally vary depending on nucleotide concentration, polymerase concentration, template concentration, template sequence, position of the polymerase along the template, characteristics of the current Watson-Crick template-nucleotide pair, characteristics of the previous Watson-Crick template-nucleotide pair, or characteristics of the nucleotide being incorporated. The nucleotides in the reaction mixture optionally comprise one or more analogue nucleotides. The transition rates between states optionally include complete orthogonality between every combination of multidimensional dependencies listed above. The multidimensional probability matrix is optionally automatically generated based upon the template sequence, a standardized matrix of probability states, and the nucleotides in the reaction mixture. The probability matrix is optionally simplified by assuming that all possible Watson-Crick base pairings are equivalent in all state transitions. The probability matrix is further optionally simplified by assuming that certain state transitions (eg. polymerase translocation along DNA) are equivalent between different dimensions of the probability matrix (eg. certain characteristics of nucleotide previously incorporated).

[0136] Similarly, a second reagent concentration matrix is optionally generated to account for reagent concentration changes that result from position of the polymerase along a template, based on an output of the probability matrix. The probability matrix is optionally vectorized for multiple templates and the resulting vectorized probability matrix can be multiplied by the multidimensional probability matrix to provide a state distribution matrix. An exponential time factor for the probability matrix can be used to account for repeated sequences within the template sequence. A polymerase nucleotide mismatch fraction using either a continuum model or a counting model can be defined.

EXAMPLES

[0137] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art. Accordingly, the following examples are offered to illustrate, but not to limit, the claimed invention.

[0138] The following sets forth a series of experiments that demonstrate construction and characterization of a variety of recombinant DNA polymerases having modified active site regions and modified properties for nucleotide analogues.

EXAMPLE 1: EXPRESSION OF RECOMBINANT POLYMERASE

[0139] A vector for expression of Phi 29 polymerase was constructed and is schematically illustrated in Figure 1. An N62D mutation was introduced into wild-type Phi 29 (SEQ ID NO:1) to reduce exonuclease activity, and GST (glutathione-S-transferase), His, and S tags were added. The resulting tagged N62D Phi 29 amino acid sequence is presented as SEQ ID NO:2. The sequence of the vector is presented as SEQ ID NO:14. The tagged N62D Phi 29 polymerase is encoded by nucleotides 4839-7428 of the vector sequence, with the polymerase at nucleotides 5700-7428 and the N62D mutation at nucleotides 5883-5885. Other features of the vector include the GST-His-S tag sequences (nucleotides 4838-5699), ribosome binding site (nucleotides 4822-4829), T7 promoter (nucleotides 4746-4758), and kanamycin resistance marker (complement of nucleotides 563-1375).

[0140] Additional mutations are readily introduced into this construct as desired, for example, to facilitate expression of recombinant Phi 29 polymerases having modified active site regions. See, e.g., SEQ ID NOs:15-23. The recombinant proteins can be expressed in *E. coli*, for example, and purified using the GST, His, and/or S tags and standard techniques. The tags are optionally removed by digestion with an appropriate protease (e.g., thrombin or enterokinase).

EXAMPLE 2: EXEMPLARY RECOMBINANT POLYMERASES

[0141] A variety of recombinant Phi 29 polymerases with modified active site regions have been constructed. Without intending to be limited to any particular mechanism, the following examples illustrate structural modifications that can reduce steric inhibition for entry of nucleotide analogues into the modified active site regions, coordinate

extra phosphate groups by providing features that complement these groups (e.g., positively charged amino acid side chains), and/or otherwise enhance the ability of the polymerase to incorporate nucleotide analogues.

[0142] Figure 2 Panel A shows a sequence alignment of Phi 29-like polymerases in the region surrounding residues 505-525, whose position is indicated by the bracket. Amino acid residues differing from Phi 29 are underlined. The majority of this domain is missing in the cp-1 DNA polymerase (which, like G1, is more distantly related to Phi 29). In addition, there is notably less sequence conservation within the domain than in the flanking sequence. These observations suggest that removal of the domain is unlikely to be deleterious.

[0143] The top three views in Figure 2 Panel B illustrate the structure of the Phi 29 polymerase (see, e.g., Kamtekar et al. (2004) "Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage Φ29 " Mol. Cell 16(4): 609-618). The bottom three views show the polymerase with residues 505-525 removed, illustrating that removal of this domain opens up the nucleotide binding pocket. See, e.g., SEQ ID NOs:12 and 13 or 33 and 34, which remove this domain using different turns.

[0144] Figure 3 Panel A shows a sequence alignment of Phi 29-like polymerases in the region surrounding E375 of Phi 29. The top three views in Panel B illustrate the structure of the Phi 29 polymerase. The glutamate at position 375 (indicated by the arrow) is located proximal to the positively charged residues (K371, K379, K383; depicted in medium gray with dark gray knobs) that contact the triphosphate moiety of the incoming dNTP. As illustrated in the bottom three views in Panel B, this negatively charged amino acid (E) was replaced with a positive one (H) in an attempt to coordinate the extra phosphate in the tetraphosphate nucleotide analogues. Additionally, the extra positive charge at this site may help coordinate triphosphate analogs. Analysis of the recombinant polymerase suggests that the E375H mutation has improved the kinetics of the enzyme for incorporating phosphate labeled nucleotide analogues (see Example 3 below). Mutant E375S was also constructed to introduce a neutral residue at this location and/or, e.g., to facilitate conformational change to enable function. See also SEQ ID NOs:4-7 and 25-28.

[0145] Figure 4 Panel A shows a sequence alignment of Phi 29-like polymerases in the region surrounding E486 of Phi 29. The top three views in Panel B illustrate the structure of the Phi 29 polymerase; the location of E486 is indicated by an arrow. As illustrated in the bottom three views, replacement of E486 by an alanine residue creates more room in the active site region near the catalytic carboxylates (D249 and D458, depicted in white) and removes a negative charge. As another example, replacement of E486 by an aspartic acid residue removes a carbon, decreasing steric interference with nucleotide analogue binding while retaining the negative charge. See also SEQ ID NOs:9-10 and 30-31.

[0146] Figure 5 Panel A shows a sequence alignment of Phi 29-like polymerases in the region surrounding K512 of Phi 29. The top three views in Panel B illustrate the structure of the Phi 29 polymerase. K512 (indicated by an arrow) juts out from the residue 505-525 domain and partially blocks the opening to the incoming dNTP binding site. As illustrated in the bottom three views, replacement of K512 by an alanine residue reduces steric inhibition for entry of nucleotide analogues into the active site region, providing more space for them to get into the binding pocket. See also SEQ ID NOs:11 and 32.

[0147] Figure 6 Panel A shows a sequence alignment of Phi 29-like polymerases in the region surrounding K135 of Phi 29. The top three views in Panel B illustrate the structure of the Phi 29 polymerase. K135 (indicated by an arrow) juts into the opening to the incoming dNTP binding site. As illustrated in the bottom three views, replacement of K135 by an alanine residue reduces steric inhibition for entry of nucleotide analogues into the active site region, providing more space for them to get into the binding pocket. See also SEQ ID NOs:3 and 24.

EXAMPLE 3: SCREENING AND CHARACTERIZATION OF RECOMBINANT POLYMERASES

[0148] Recombinant polymerases generated as in Example 2, or through essentially any other rational or random mutagenesis strategy, are optionally characterized to determine their properties for various natural and/or nucleotides. One exemplary five-step protocol for characterizing recombinant polymerases follows.

[0149] The recombinant polymerase is initially evaluated on the quality of the protein preparation and basic catalytic activity. The polymerase's activity is analyzed with

natural (native) nucleotides, and its specific activity (units/mg) is determined. Only catalytically competent mutants are selected for the next steps.

[0150] The processivity (dissociation/kb) of the polymerase is estimated in a primer extension reaction performed in the presence of "Trap" (unlabeled competitor DNA or heparin). The processivity assay is designed to select mutants that retain the capability to synthesize a long DNA product in a continuous polymerization run (without polymerization reinitiation) with natural nucleotides. Mutants with a significant decrease in processivity are not selected for the next step.

[0151] Polymerization rate (bases/min) with four analogues at 10 μM (A488dA4P, A633dC4P, A546dG4P and A594dT4P) and circular template (AGTC, a 72mer circular template largely consisting of a repeating AGTC motif) is determined.

[0152] The most promising polymerase mutants are characterized by determination of the polymerization rate and Km for A488dC4P and A568dC4P and a subset of natural nucleotides (dATP, dGTP and dTTP), using a circular template (AGTC). Velocity is measured at several different concentrations of the analogs, A488dC4P (a representative good substrate) and A568dC4P (a representative less preferred substrate).

[0153] An initial selection for polymerase mutants with improved kinetics for terminal phosphate labeled nucleotide analogues is performed, using a primer extension assay with nucleotide analogues to determine rate with analogues under experimental conditions. Two separate experiments are typically performed, one in the presence of 10μM A488dC4P, 20μM 3dNTPs –dCTP, and circular template (AGTC), and one in the presence of 10μM A568dC4P, 20μM 3dNTPs –dCTP, and circular template (AGTC).

[0154] Other characteristics of the recombinant polymerase are optionally examined, including, for example, fidelity, residence time ($1/V_{max}$), exonuclease activity (e.g., at 10uM, via extension of mismatched primer), active fraction (burst frequency), rate with dNTPs, dN5Ps, linker-only analogs, and/or FRET analogs, kinetics (ability to utilize analogs) with Mg^{2+} vs. Mn^{2+}, sensitivity to photodamage, single-stranded DNA binding, monomeric state (e.g., using gel filtration or native gels), and/or shelf-life.

[0155] Results of protein quality evaluation and polymerization rate and kinetic constant determination for exemplary recombinant Phi 29 polymerases are presented in Tables 1 and 2, respectively.

Table 1. Initial characterization.

Description	Concentration; Yield of Purified Polymerase	Specific Activity (units/mg)
His-K135A-N62D	3.7 uM; 1 mg	12,454,000
His-E375H-N62D	7.4 uM; 1 mg	10,945,000
His-E375S-N62D	109 uM; 7 mg	10,961,000
His-E486A-N62D	40 uM; 3.5 mg	4,133,000
His-E486D-N62D	36 uM; 3.1 mg	11,634,000
His-K512A-N62D	34 uM; 10 mg	16,073,000
His-NipTuck_1-N62D	32 uM; 2.5 mg	12,400,000
His-NipTuck_2-N62D	4.4 uM; 0.3 mg	7,960,000

Table 2. Characterization of polymerization rate with natural and analogue nucleotides.

A	B	C	D	E	F	G	H	I	J
GST-N62D	780	1200	20	1660	74	346	236	65	0.9799
His-N62D	750	1020	21			391	237	68	0.9754
His-K135A-N62D	840	880	24			292	154	43	0.9801
His-E375H-N62D	780	950	8	930	11	411	366	123	0.9510
His-E375S-N62D	940	1190	15	1300	28	420	332	74	0.9815
His-E486A-N62D	1690					303	118	15	0.9875
His-E486D-N62D						220	134	15	0.9885
His-K512A-N62D	1590 (630)					359	196	34	0.9821
His-NipTuck_1-N62D	660	520	24			153	116	24	0.9585
His-NipTuck_2-N62D	540 (1840)					147	129	28	0.9520

Column A: Description.
Column B: dTTP, dATP, dGTP (no G fork) V at 20 μM; determined by an assay with three native nucleotides (dGTP, dTTP and dATP).
Column C: A488dC4P, k_{el} (bp/min); determined by examining the nucleotide analogue concentration dependence of the polymerization rate.
Column D: A488dC4P, Km; determined by examining the nucleotide analogue concentration dependence of the polymerization rate.
Column E: A568dC4P, k_{el} ; determined by examining the nucleotide analogue concentration dependence of the polymerization rate.
Column F: A568dC4P, Km; determined by examining the nucleotide analogue concentration dependence of the polymerization rate.
Column G: A488dC4P, V at 10 μM; determined by an assay with a single analogue at low concentration (10uM) and three native nucleotides.
Column H: A568dC4P, V at 10 μM; determined by an assay with a single analogue at low concentration (10uM) and three native nucleotides.
Column I: A488dA4P, A633dC4P, A546dG4P, A594dT4P, V at 10 μM; determined by an assay with four terminally labeled nucleotide analogs.

Column J: Processivity (kb^{-1}); determined by a processivity assay.

Assay with a single analogue at low concentration (10 μM) and three native nucleotides

[0156] The Φ29 DNA polymerase (parental enzyme or mutant) was preincubated with DNA template (72 nucleotide circular DNA including mostly repetitive sequence AGTC) with annealed DNA primer. The preincubation mix includes three native nucleotides (dTTP, dATP and dGTP) and a terminal labeled nucleotide analogue (A488dC4P or A568dC4P) at 10 μM concentration. After a short preincubation, the reaction was started with $MnCl_2$. The reaction was stopped with EDTA, and the products were separated using agarose gel electrophoresis and stained with SYBR Gold (Invitrogen). The average length of the DNA generated with DNA polymerase was determined and used to estimate the polymerization rate. See, e.g., Table 2 Columns G and H.

Assay with four terminally labeled nucleotide analogs

[0157] The procedure is basically as described above in the section entitled "Assay with a single analogue at low concentration (10 μM) and three native nucleotides," with the exception that in this assay all nucleotides are terminally labeled (A488dA4P, A633dC4P, A546dG4P, A594dT4P all at 10 μM). See, e.g., Table 2 Column I.

Assay with three native nucleotides (dGTP, dTTP and dATP)

[0158] The Φ29 DNA polymerase (parental enzyme or mutant) was preincubated with DNA template (circular DNA including mostly repetitive sequence CAT, no G residues) with annealed DNA primer; the preincubation mix includes three native nucleotides (dTTP, dATP and dGTP). All subsequent steps were basically as described above in the section entitled "Assay with a single analogue at low concentration (10 μM) and three native nucleotides." See, e.g., Table 2 Column B.

The nucleotide analogue concentration dependence of the polymerization rate

[0159] The Φ29 DNA polymerase (parental enzyme or mutant) was preincubated with a DNA template (72 nucleotide circular DNA including mostly repetitive sequence AGTC) with annealed DNA primer. The preincubation mix includes also three native nucleotides (dTTP, dATP and dGTP 20 μM each) and various concentrations of the terminally labeled analogue (A488dC4P or A568dC4P). All subsequent steps were basically as described above in the section entitled "Assay with a single analogue at low

concentration (10 μM) and three native nucleotides." An average length of the DNA products generated with DNA polymerase at an individual analogue concentration was determined, and the results were fitted with the equation $k=k_{el}*[S]*(K_d+[S])^{-1}$ where k is the observed polymerization rate, k_{el} is the polymerization rate at saturating substrate concentration (k_{el} measures incorporation of multiple residues), and [S] is substrate concentration. See, e.g., Table 2 Columns C, D, E, and F.

Processivity assay

[0160] The Φ29 DNA polymerase (parental enzyme or mutant) was preincubated with DNA template (72 nucleotide circular DNA including mostly repetitive sequence AGTC) with annealed DNA primer. After a short preincubation, the reaction was started with a starting mix including $MnCl_2$, dNTP and heparin. Including the heparin in the reaction prevents polymerization from reinitiating after the polymerase dissociates from the template/primer, so that all generated DNA products are a result of continuous polymerization runs. After 20 min incubation, the reaction was stopped with EDTA and the products were separated using agarose gel electrophoresis and stained with SYBR Gold (Invitrogen). The DNA products were analyzed basically as described in Bibillo A, Eickbush TH. J Biol Chem. 2002 Sep 20; 277(38):34836-45, Epub 2002 Jul5. The results were fitted with single exponential equation $A*\exp(-P_{off}*kb)$ where A is amplitude, P_{off} is the probability of premature polymerase dissociation, and kb is DNA length (1000 nucleotides). The probability of chain elongation (processivity) can be readily calculated by subtracting the P_{off} value from 1.0. See, e.g., Table 2 Column J.

Sequences of exemplary recombinant polymerases

[0161] Amino acid and polynucleotide sequences of wild-type Phi 29 and exemplary recombinant polymerases are presented in Table 3.

Table 3. Sequences.

SEQ ID NO:	Notes	Sequence
1	wild-type Phi 29 amino acid sequence	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hnlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiakdfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv nslypaqmys rllpygepiv fegkyvwded yplhighirc efelkegyip

		tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttsegaikq laklmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahestfk rakylrqkty iqdiymkevd gklvegspdd ytdikfsvkc agmtdkikke vtfenfkvgf srkmkpkpvq vpggvvlvdd tftik
2	N62D amino acid sequence (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgdddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki akdfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt segaikqlak lmlnslygkf asnpdvtgkv pylkengalg frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahestfkrak ylrqktyiqd iymkevdgkl vegspddytd ikfsvkcagm tdkikkevtf enfkvgfsrk mkpkpvqvpg gvvlvddtft ik
3	K135A-N62D amino acid sequence (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgdddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki aadfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt segaikqlak lmlnslygkf asnpdvtgkv pylkengalg frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahestfkrak ylrqktyiqd iymkevdgkl vegspddytd ikfsvkcagm tdkikkevtf enfkvgfsrk mkpkpvqvpg gvvlvddtft ik
4	E375H-N62D amino acid sequence (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgdddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki akdfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt shgaikqlak lmlnslygkf asnpdvtgkv pylkengalg

		frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahestfkrak ylrqktyiqd iymkevdgkl vegspddytd ikfsvkcagm tdkikkevtf enfkvgfsrk mkpkpvqvpg gvvlvddtft ik
5	E375S-N62D amino acid sequence (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgdddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki akdfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt ssgaikqlak lmlnslygkf asnpdvtgkv pylkengalg frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahestfkrak ylrqktyiqd iymkevdgkl vegspddytd ikfsvkcagm tdkikkevtf enfkvgfsrk mkpkpvqvpg gvvlvddtft ik
6	E375K-N62D amino acid sequence (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgdddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki akdfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt skgaikqlak lmlnslygkf asnpdvtgkv pylkengalg frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahestfkrak ylrqktyiqd iymkevdgkl vegspddytd ikfsvkcagm tdkikkevtf enfkvgfsrk mkpkpvqvpg gvvlvddtft ik
7	E375R-N62D amino acid sequence (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgdddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki akdfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt srgaikqlak lmlnslygkf asnpdvtgkv pylkengalg frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahestfkrak ylrqktyiqd iymkevdgkl

		vegspddytd ikfsvkcagm tdkikkevtf enfkvgfsrk mkpkpvqvpg gvvlvddtft ik
8	L384R-N62D amino acid sequence (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgddddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki akdfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt segaikqlak rmlnslygkf asnpdvtgkv pylkengalg frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahestfkrak ylrqktyiqd iymkevdgkl vegspddytd ikfsvkcagm tdkikkevtf enfkvgfsrk mkpkpvqvpg gvvlvddtft ik
9	E486A-N62D amino acid sequence (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgddddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki akdfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt segaikqlak lmlnslygkf asnpdvtgkv pylkengalg frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahastfkrak ylrqktyiqd iymkevdgkl vegspddytd ikfsvkcagm tdkikkevtf enfkvgfsrk mkpkpvqvpg gvvlvddtft ik
10	E486D-N62D amino acid sequence (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgddddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki akdfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt segaikqlak lmlnslygkf asnpdvtgkv pylkengalg frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahdstfkrak ylrqktyiqd iymkevdgkl vegspddytd ikfsvkcagm tdkikkevtf enfkvgfsrk mkpkpvqvpg gvvlvddtft ik

11	K512A-N62D amino acid sequence (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgdddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki akdfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt segaikqlak lmlnslygkf asnpdvtgkv pylkengalg frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahestfkrak ylrqktyiqd iymkevdgal vegspddytd ikfsvkcagm tdkikkevtf enfkvgfsrk mkpkpvqvpg gvvlvddtft ik
12	NipTuck_1-N62D amino acid sequence (deletion of residues 505-525) (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgdddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki akdfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt segaikqlak lmlnslygkf asnpdvtgkv pylkengalg frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahestfkrak ylrqktyiqd **ik**dgefsvkc agmtdkikke vtfenfkvgf srkmkpkpvq vpggvvlvdd tftik
13	NipTuck_2-N62D amino acid sequence (deletion of residues 505-525) (tagged)	mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia wplqgwqatf gggdhppksd gstsgsghhh hhhsaglvpr gstaigmket aaakferqhm dspdlgtggg sgdddkspm gyrgsefmkh mprkmyscdf etttkvedcr vwaygymnie dhseykigns ldefmawvlk vqadlyfhdl kfdgafiinw lerngfkwsa dglpntynti isrmgqwymi diclgykgkr kihtviydsl kklpfpvkki akdfkltvlk gdidyhkerp vgykitpeey ayikndiqii aealliqfkq gldrmtagsd slkgfkdiit tkkfkkvfpt lslgldkevr yayrggftwl ndrfkekeig egmvfdvnsl ypaqmysrll pygepivfeg kyvwdedypl hiqhircefe lkegyiptiq ikrsrfykgn eylkssggei adlwlsnvdl elmkehydly nveyisglkf kattglfkdf idkwtyiktt segaikqlak lmlnslygkf asnpdvtgkv pylkengalg frlgeeetkd pvytpmgvfi tawaryttit aaqacydrii ycdtdsihlt gteipdvikd ivdpkklgyw ahestfkrak ylrqktyiqd **id**gfsvkcag mtdkikkevt fenfkvgfsr kmkpkpvqvp ggvvlvddtf tik
14	N62D nucleotide	tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtgg tggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgct

	sequence - pET41 N62D 1 plasmid	cctttcgctttcttcccttcctttctcgccacgttcgccggctttccccg tcaagctctaaatcgggggctcccttagggttccgatttagtgctttac ggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtggg ccatcgccctgatagacggtttttcgccctttgacgttggagtccacgtt ctttaatagtggactcttgttccaaactggaacaacactcaaccctatct cggtctattcttttgatttataagggattttgccgatttcggcctattgg ttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaat attaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaa cccctatttgtttatttttctaaatacattcaaatatgtatccgctcatg aattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttat tcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaat gaaggagaaaactcaccgaggcagttccataggatggcaagatcctggta tcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcc cctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgact gaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttc aacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaac cgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctg ttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacac tgccagcgcatcaacaatattttcacctgaatcaggatattcttctaata cctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatca tcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgt cagccagtttagtctgaccatctcatctgtaacatcattggcaacgctac ctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaat cgatagattgtcgcacctgattgcccgacattatcgcgagcccatttata cccatataaatcagcatccatgttggaatttaatcgcggcctagagcaag acgtttcccgttgaatatggctcataacacccttgtattactgtttatg taagcagacagttttattgttcatgaccaaaatcccttaacgtgagtttt cgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttga gatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccacc gctaccagcggtggtttgtttgccggatcaagagctaccaactcttttc cgaaggtaactggcttcagcagagcgcagataccaaatactgtccttcta gtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctac atacctcgctctgctaatcctgttaccagtggctgctgccagtggcgata agtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcg cagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcg aacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcg ccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagg gtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggta tctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttt tgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcg gccttttacggttcctggccttttgctggccttttgctcacatgttctt tcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagt gagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtg agcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatct gtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctc tgatgccgcatagttaagccagtatacactccgctatcgctacgtgactg ggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgac gggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctcc gggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgag gcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgt ctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaat gtctggcttctgataaagcgggccatgttaagggcggttttttcctgttt ggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaatga taccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaa

		catgcccggttactggaacgttgtgagggtaaacaactggcggtatggat gcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgtta atacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcag atccggaacataatggtgcagggcgctgacttccgcgtttccagacttta cgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacg ttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattc tgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacag gagcacgatcatgctagtcatgccccgcgcccaccggaaggagctgactg ggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtg agctaacttacattaattgcgttgcgctcactgcccgctttccagtcggg aaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagag gcggtttgcgtattgggcgccagggtggtttttctttcaccagtgagac gggcaacagctgattgcccttcaccgcctggccctgagagagttgcagca agcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtg gttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccac taccgagatgtccgcaccaacgcgcagcccggactcggtaatggcgcgca ttgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacg atgccctcattcagcatttgcatggtttgttgaaaaccggacatggcact ccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagat atttatgccagccagccagacgcagacgcgccgagacagaacttaatggg cccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccac gcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtg tctggtcagagacatcaagaaataacgccggaacattagtgcaggcagct tccacagcaatggcatcctggtcatccagcggatagttaatgatcagccc actgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcga cgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcg gcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccag actggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgtt gtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccact ttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcggga aacggtctgataagagacaccggcatactctgcgacatcgtataacgtta ctggtttcacattcaccaccctgaattgactctcttccgggcgctatcat gccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgac gctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttg aggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggc gcccaacagtcccccggccacggggcctgccaccatacccacgccgaaac aagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatg tcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggc cacgatgcgtccggcgtagaggatcgagatcgatctcgatcccgcgaaat taatacgactcactatagggaattgtgagcggataacaattcccctcta gaaataattttgtttaactttaagaaggagatatacatatgtcccctata ctaggttattggaaaattaagggccttgtgcaacccactcgacttctttt ggaatatcttgaagaaaaatatgaagagcatttgtatgagcgcgatgaag gtgataaatggcgaaacaaaaagtttgaattgggtttggagtttcccaat cttccttattatattgatggtgatgttaaattaacacagtctatggccat catacgttatatagctgacaagcacaacatgttgggtggttgtccaaaag agcgtgcagagatttcaatgcttgaaggagcggttttggatattagatac ggtgtttcgagaattgcatatagtaaagactttgaaactctcaaagttga ttttcttagcaagctacctgaaatgctgaaaatgttcgaagatcgtttat gtcataaaacatatttaaatggtgatcatgtaacccatcctgacttcatg ttgtatgacgctcttgatgttgttttatacatggacccaatgtgcctgga tgcgttcccaaaattagtttgttttaaaaaacgtattgaagctatcccac aaattgataagtacttgaaatccagcaagtatatagcatggcctttgcag ggctggcaagccacgtttggtggtggcgaccatcctccaaaatcggatgg

		ttcaactagtggttctggtcatcaccatcaccatcactccgcgggtctgg tgccacgcggtagtactgcaattggtatgaaagaaaccgctgctgctaaa ttcgaacgccagcacatggacagcccagatctgggtaccggtggtggctc cggtgatgacgacgacaagagtcccatgggatatcggggatccgaattca tgaagcatatgccgagaaagatgtatagttgtgactttgagacaactact aaagtggaagactgtagggtatgggcgtatggttatatgaatatagaaga tcacagtgagtacaaaatagqtaatagcctggatgagtttatggcgtggg tgttgaaggtacaagctgatctatatttccatgatctcaaatttgacgga gcttttatcattaactggttggaacgtaatggttttaagtggtcggctga cggattgccaaacacatataatacgatcatatctcgcatgggacaatggt acatgattgatatatgtttaggctacaaagggaaacgtaagatacataca gtgatatatgacagcttaaagaaactaccgtttcctgttaagaagatagc taaagactttaaactaactgttcttaaaggtgatattgattaccacaaag aaagaccagtcggctataagataacacccgaagaatacgcctatattaaa aacgatattcagattattgcggaagctctgttaattcagtttaagcaagg tttagaccggatgacagcaggcagtgacagtctaaaaggtttcaaggata ttataaccactaagaaattcaaaaaggtgtttcctacattgagtcttgga ctcgataaggaagtgagatacgcctatagaggtggttttacatggttaaa tgataggttcaaagaaaaagaaatcggagaaggcatggtcttcgatgtta atagtctatatcctgcacagatgtatagtcgtctccttccatatggtgaa cctatagtattcgagggtaaatacgtttgggacgaagattacccactaca catacagcatatcagatgtgagttcgaattgaaagagggctatataccca ctatacagataaaaagaagtaggttttataaaggtaatgagtacctaaaa agtagcggcggggagatagccgacctctggttgtcaaatgtagacctaga attaatgaaagaacactacgatttatataacgttgaatatatcagcggct taaaatttaaagcaactacaggtttgtttaaagattttatagataaatgg acgtacatcaagacgacatcagaaggagcgatcaagcaactagcaaaact gatgttaaacagtctatacggtaaattcgctagtaaccctgatgttacag ggaaagtcccttatttaaaagagaatggggcgctaggtttcagacttgga gaagaggaaacaaaagaccctgtttatacacctatgggcgttttcatcac tgcatgggctagatacacgacaattacagcggcacaggcttgttatgatc ggataatatactgtgatactgacagcatacatttaacgggtacagagata cctgatgtaataaaagatatagttgacgctaagaaattgggatactgggc acatgaaagtacattcaaaagagctaaatatctgagacagaagacctata tacaagacatctatatgaaagaagtagatggtaagttagtagaaggtagt ccagatgattacactgatataaaatttagtgttaaatgtgcgggaatgac tgacaagattaagaaagaggttacgtttgagaatttcaaagtcggattca gtcggaaaatgaagcctaagcctgtgcaagtgccgggcggggtggttctg gttgatgacacattcacaatcaaataagaattctgtacaggccttggcgc gcctgcaggcgagctccgtcgacaagcttgcggccgcactcgagcaccac caccaccaccaccaccactaattgattaatacctaggctgctaaacaaag cccgaaaggaagctgagttggctgctgccaccgctgagcaataactagca taacccctttggggcctctaaacgggtcttgaggggttttttgctgaaagg aggaactatatccggat
15	K135A-N62D nucleotide sequence	atgaagcacatgccgagaaagatgtatagttgtgactttgagacaactactaaagtggaagactgtaggg tatgggcgtatggttatatgaatatagaagatcacagtgagtacaaaataggtaatagcctggatgagtt tatggcgtgggtgttgaaggtacaagctgatctatatttccatgatctcaaatttgacggagcttttatc attaactggttggaacgtaatggttttaagtggtcggctgacggattgccaaacacatataatacgatca tatctcgcatgggacaatggtacatgattgatatatgtttaggctacaaagggaaacgtaagatacatac agtgatatatgacagcttaaagaaactaccgtttcctgttaagaagatagctgccgactttaaactaact gttcttaaaggtgatattgattaccacaaagaaagaccagtcggctataagataacacccgaagaatacg cctatattaaaaacgatattcagattattgcggaagctctgttaattcagtttaagcaaggtttagaccg gatgacagcaggcagtgacagtctaaaaggtttcaaggatattataaccactaagaaattcaaaaaggtg tttcctacattgagtcttggactcgataaggaagtgagatacgcctatagaggtggttttacatggttaa atgataggttcaaagaaaaagaaatcggagaaggcatggtcttcgatgttaatagtctatatcctgcaca gatgtatagtcgtctccttccatacggtgaacctatagtattcgagggtaaatacgtttgggacgaagat

		tacccactacacatacagcatatcagatgtgagttcgaattgaaagagggctatatacccactatacaga taaaaagaagtaggttttataaaggtaatgagtacctaaaaagtagcggcggggagatagccgacctctg gttgtcaaatgtagacctagaattaatgaaagaacactacgatttatataacgttgaatatatcagcggc ttaaaatttaaagcaactacaggtttgtttaaagattttatagataaatggacgtacatcaagacgacat cagaaggagcgatcaagcaactagcaaaactgatgttaaacagtctatacggtaaattcgctagtaaccc tgatgttacagggaaagtcccttatttaaaagagaatggggcgctaggtttcagacttggagaagaggaa acaaaagaccctgtttatacacctatgggcgttttcatcactgcatgggctagatacacgacaattacag cggcacaggcttgttatgatcggataatatactgtgatactgacagcatacatttaacgggtacagagat acctgatgtaataaaagatatagttgaccctaagaaattgggatactgggcacatgaaagtacattcaaa agagctaaatatctgagacagaagacctatatacaagacatctatatgaaagaagtagatggtaagttag tagaaggtagtccagatgattacactgatataaaatttagtgttaaatgtgcgggaatgactgacaagat taagaaagaggttacgtttgagaatttcaaagtcggattcagtcggaaaatgaagcctaagcctgtgcaa gtgccgggcggggtggttctggttgatgacacattcacaatcaaataa
16	E375H-N62D nucleotide sequence	atgaagcacatgccgagaaagatgtatagttgtgactttgagacaactactaaagtggaagactgtaggg tatgggcgtatggttatatgaatatagaagatcacagtgagtacaaaataggtaatagcctggatgagtt tatggcgtgggtgttgaaggtacaagctgatctatatttccatgatctcaaatttgacggagcttttatc attaactggttggaacgtaatggttttaagtggtcggctgacggattgccaaacacatataatacgatca tatctcgcatgggacaatggtacatgattgatatatgtttaggctacaaagggaaacgtaagatacatac agtgatatatgacagcttaaagaaactaccgtttcctgttaagaagatagctaaagactttaaactaact gttcttaaaggtgatattgattaccacaaagaaagaccagtcggctataagataacacccgaagaatacg cctatattaaaaacgatattcagattattgcggaagctctgttaattcagtttaagcaaggtttagaccg gatgacagcaggcagtgacagtctaaaaggtttcaaggatattataaccactaagaaattcaaaaaggtg tttcctacattgagtcttggactcgataaggaagtgagatacgcctatagaggtggttttacatggttaa atgataggttcaaagaaaaagaaatcggagaaggcatggtcttcgatgttaatagtctatatcctgcaca gatgtatagtcgtctccttccatacggtgaacctatagtattcgagggtaaatacgtttgggacgaagat tacccactacacatacagcatatcagatgtgagttcgaattgaaagagggctatatacccactatacaga taaaaagaagtaggttttataaaggtaatgagtacctaaaaagtagcggcggggagatagccgacctctg gttgtcaaatgtagacctagaattaatgaaagaacactacgatttatataacgttgaatatatcagcggc ttaaaatttaaagcaactacaggtttgtttaaagattttatagataaatggacgtacatcaagacgacat caagcggagcgatcaagcaactagcaaaactgatgttaaacagtctatacggtaaattcgctagtaaccc tgatgttacagggaaagtcccttatttaaaagagaatggggcgctaggtttcagacttggagaagaggaa acaaaagaccctgtttatacacctatgggcgttttcatcactgcatgggctagatacacgacaattacag cggcacaggcttgttatgatcggataatatactgtgatactgacagcatacatttaacgggtacagagat acctgatgtaataaaagatatagttgaccctaagaaattgggatactgggcacatgaaagtacattcaaa agagctaaatatctgagacagaagacctatatacaagacatctatatgaaagaagtagatggtaagttag tagaaggtagtccagatgattacactgatataaaatttagtgttaaatgtgcgggaatgactgacaagat taagaaagaggttacgtttgagaatttcaaagtcggattcagtcggaaaatgaagcctaagcctgtgcaa gtgccgggcggggtggttctggttgatgacacattcacaatcaaataa
17	E375S-N62D nucleotide sequence	atgaagcacatgccgagaaagatgtatagttgtgactttgagacaactactaaagtggaagactgtaggg tatgggcgtatggttatatgaatatagaagatcacagtgagtacaaaataggtaatagcctggatgagtt tatggcgtgggtgttgaaggtacaagctgatctatatttccatgatctcaaatttgacggagcttttatc attaactggttggaacgtaatggttttaagtggtcggctgacggattgccaaacacatataatacgatca tatctcgcatgggacaatggtacatgattgatatatgtttaggctacaaagggaaacgtaagatacatac agtgatatatgacagcttaaagaaactaccgtttcctgttaagaagatagctaaagactttaaactaact gttcttaaaggtgatattgattaccacaaagaaagaccagtcggctataagataacacccgaagaatacg cctatattaaaaacgatattcagattattgcggaagctctgttaattcagtttaagcaaggtttagaccg gatgacagcaggcagtgacagtctaaaaggtttcaaggatattataaccactaagaaattcaaaaaggtg tttcctacattgagtcttggactcgataaggaagtgagatacgcctatagaggtggttttacatggttaa atgataggttcaaagaaaaagaaatcggagaaggcatggtcttcgatgttaatagtctatatcctgcaca gatgtatagtcgtctccttccatacggtgaacctatagtattcgagggtaaatacgtttgggacgaagat tacccactacacatacagcatatcagatgtgagttcgaattgaaagagggctatatacccactatacaga taaaaagaagtaggttttataaaggtaatgagtacctaaaaagtagcggcggggagatagccgacctctg gttgtcaaatgtagacctagaattaatgaaagaacactacgatttatataacgttgaatatatcagcggc ttaaaatttaaagcaactacaggtttgtttaaagattttatagataaatggacgtacatcaagacgacat caagcggagcgatcaagcaactagcaaaactgatgttaaacagtctatacggtaaattcgctagtaaccc tgatgttacagggaaagtcccttatttaaaagagaatggggcgctaggtttcagacttggagaagaggaa acaaaagaccctgtttatacacctatgggcgttttcatcactgcatgggctagatacacgacaattacag cggcacaggcttgttatgatcggataatatactgtgatactgacagcatacatttaacgggtacagagat acctgatgtaataaaagatatagttgaccctaagaaattgggatactgggcacatgaaagtacattcaaa agagctaaatatctgagacagaagacctatatacaagacatctatatgaaagaagtagatggtaagttag tagaaggtagtccagatgattacactgatataaaatttagtgttaaatgtgcgggaatgactgacaagat taagaaagaggttacgtttgagaatttcaaagtcggattcagtcggaaaatgaagcctaagcctgtgcaa gtgccgggcggggtggttctggttgatgacacattcacaatcaaataa
18	L384R-N62D nucleotide sequence	atgaagcacatgccgagaaagatgtatagttgtgactttgagacaactactaaagtggaagactgtaggg tatgggcgtatggttatatgaatatagaagatcacagtgagtacaaaataggtaatagcctggatgagtt tatggcgtgggtgttgaaggtacaagctgatctatatttccatgatctcaaatttgacggagcttttatc attaactggttggaacgtaatggttttaagtggtcggctgacggattgccaaacacatataatacgatca tatctcgcatgggacaatggtacatgattgatatatgtttaggctacaaagggaaacgtaagatacatac agtgatatatgacagcttaaagaaactaccgtttcctgttaagaagatagctaaagactttaaactaact gttcttaaaggtgatattgattaccacaaagaaagaccagtcggctataagataacacccgaagaatacg

		cctatattaaaaacgatattcagattattgcggaagctctgttaattcagtttaagcaaggtttagaccg gatgacagcaggcagtgacagtctaaaaggtttcaaggatattataaccactaagaaattcaaaaaggtg tttcctacattgagtcttggactcgataaggaagtgagatacgcctatagaggtggttttacatggttaa atgataggttcaaagaaaaagaaatcggagaaggcatggtcttcgatgttaatagtctatatcctgcaca gatgtatagtcgtctccttccatacggtgaacctatagtattcgagggtaaatacgtttgggacgaagat tacccactacacatacagcatatcagatgtgagttcgaattgaaagagggctatatacccactatacaga taaaaagaagtaggttttataaaggtaatgagtacctaaaaagtagcggcggggagatagccgacctctg gttgtcaaatgtagacctagaattaatgaaagaacactacgatttatataacgttgaatatatcagcggc ttaaaatttaaagcaactacaggtttgtttaaagattttatagataaatggacgtacatcaagacgacat cagaaggagcgatcaagcaactagcaaaacggatgttaaacagtctatacggtaaattcgctagtaaccc tgatgttacagggaaagtcccttatttaaaagagaatggggcgctaggtttcagacttggagaagaggaa acaaaagaccctgtttatacacctatgggcgttttcatcactgcatgggctagatacacgacaattacag cggcacaggcttgttatgatcggataatatactgtgatactgacagcatacatttaacgggtacagagat acctgatgtaataaaagatatagttgaccctaagaaattgggatactgggcacatgaaagtacattcaaa agagctaaatatctgagacagaagacctatatacaagacatctatatgaaagaagtagatggtaagttag tagaaggtagtccagatgattacactgatataaaatttagtgttaaatgtgcgggaatgactgacaagat taagaaagaggttacgtttgagaatttcaaagtcggattcagtcggaaaatgaagcctaagcctgtgcaa gtgccgggcggggtggttctggttgatgacacattcacaatcaaataa
19	E486A-N62D nucleotide sequence	atgaagcacatgccgagaaagatgtatagttgtgactttgagacaactactaaagtggaagactgtaggg tatgggcgtatggttatatgaatatagaagatcacagtgagtacaaaataggtaatagcctggatgagtt tatggcgtgggtgttgaaggtacaagctgatctatatttccatgatctcaaatttgacggagcttttatc attaactggttggaacgtaatggttttaagtggtcggctgacggattgccaaacacatataatacgatca tatctcgcatgggacaatggtacatgattgatatatgtttaggctacaaagggaaacgtaagatacatac agtgatatatgacagcttaaagaaactaccgtttcctgttaagaagatagctaaagactttaaactaact gttcttaaaggtgatattgattaccacaaagaaagaccagtcggctataagataacacccgaagaatacg cctatattaaaaacgatattcagattattgcggaagctctgttaattcagtttaagcaaggtttagaccg gatgacagcaggcagtgacagtctaaaaggtttcaaggatattataaccactaagaaattcaaaaaggtg tttcctacattgagtcttggactcgataaggaagtgagatacgcctatagaggtggttttacatggttaa atgataggttcaaagaaaaagaaatcggagaaggcatggtcttcgatgttaatagtctatatcctgcaca gatgtatagtcgtctccttccatacggtgaacctatagtattcgagggtaaatacgtttgggacgaagat tacccactacacatacagcatatcagatgtgagttcgaattgaaagagggctatatacccactatacaga taaaaagaagtaggttttataaaggtaatgagtacctaaaaagtagcggcggggagatagccgacctctg gttgtcaaatgtagacctagaattaatgaaagaacactacgatttatataacgttgaatatatcagcggc ttaaaatttaaagcaactacaggtttgtttaaagattttatagataaatggacgtacatcaagacgacat cagaaggagcgatcaagcaactagcaaaactgatgttaaacagtctatacggtaaattcgctagtaaccc tgatgttacagggaaagtcccttatttaaaagagaatggggcgctaggtttcagacttggagaagaggaa acaaaagaccctgtttatacacctatgggcgttttcatcactgcatgggctagatacacgacaattacag cggcacaggcttgttatgatcggataatatactgtgatactgacagcatacatttaacgggtacagaaat acctgatgtaataaaagatatagttgaccctaagaaattgggatactgggcacatgccagtacattcaaa agagctaaatatctgagacagaagacctatatacaagacatctatatgaaagaagtagatggtaagttag tagaaggtagtccagatgattacactgatataaaatttagtgttaaatgcgggaatgactgacaagat taagaaagaggttacgtttgagaatttcaaagtcggattcagtcggaaaatgaagcctaagcctgtgcaa gtgccgggcggggtggttctggttgatgacacattcacaatcaaataa
20	E486D-N62D nucleotide sequence	atgaagcacatgccgagaaagatgtatagttgtgactttgagacaactactaaagtggaagactgtaggg tatgggcgtatggttatatgaatatagaagatcacagtgagtacaaaataggtaatagcctggatgagtt tatggcgtgggtgttgaaggtacaagctgatctatatttccatgatctcaaatttgacggagcttttatc attaactggttggaacgtaatggttttaagtggtcggctgacggattgccaaacacatataatacgatca tatctcgcatgggacaatggtacatgattgatatatgtttaggctacaaagggaaacgtaagatacatac agtgatatatgacagcttaaagaaactaccgtttcctgttaagaagatagctaaagactttaaactaact gttcttaaaggtgatattgattaccacaaagaaagaccagtcggctataagataacacccgaagaatacg cctatattaaaaacgatattcagattattgcggaagctctgttaattcagtttaagcaaggtttagaccg gatgacagcaggcagtgacagtctaaaaggtttcaaggatattataaccactaagaaattcaaaaaggtg tttcctacattgagtcttggactcgataaggaagtgagatacgcctatagaggtggttttacatggttaa atgataggttcaaagaaaaagaaatcggagaaggcatggtcttcgatgttaatagtctatatcctgcaca gatgtatagtcgtctccttccatacggtgaacctatagtattcgagggtaaatacgtttgggacgaagat tacccactacacatacagcatatcagatgtgagttcgaattgaaagagggctatatacccactatacaga taaaaagaagtaggttttataaaggtaatgagtacctaaaaagtagcggcggggagatagccgacctctg gttgtcaaatgtagacctagaattaatgaaagaacactacgatttatataacgttgaatatatcagcggc ttaaaatttaaagcaactacaggtttgtttaaagattttatagataaatggacgtacatcaagacgacat cagaaggagcgatcaagcaactagcaaaactgatgttaaacagtctatacggtaaattcgctagtaaccc tgatgttacagggaaagtcccttatttaaaagagaatggggcgctaggtttcagacttggagaagaggaa acaaaagaccctgtttatacacctatgggcgttttcatcactgcatgggctagatacacgacaattacag cggcacaggcttgttatgatcggataatatactgtgatactgacagcatacatttaacgggtacagagat acctgatgtaataaaagatatagttgaccctaagaaattgggatactgggcacatgacagtacattcaaa agagctaaatatctgagacagaagacctatatacaagacatctatatgaaagaagtagatggtaagttag tagaaggtagtccagatgattacactgatataaaatttagtgttaaatgtgcgggaatgactgacaagat taagaaagaggttacgtttgagaatttcaaagtcggattcagtcggaaaatgaagcctaagcctgtgcaa gtgccgggcggggtggttctggttgatgacacattcacaatcaaataa
21	K512A-	atgaagcacatgccgagaaagatgtatagttgtgactttgagacaactactaaagtggaagactgtaggg tatgggcgtatggttatatgaatatagaagatcacagtgagtacaaaataggtaatagcctggatgagtt

	N62D nucleotide sequence	tatggcgtgggtgttgaaggtacaagctgatctatatttccatgatctcaaatttgacggagcttttatc attaactggttggaacgtaatggttttaagtggtcggctgacggattgccaaacacatataatacgatca tatctcgcatgggacaatggtacatgattgatatatgtttaggctacaaagggaaacgtaagatacatac agtgatatatgacagcttaaagaaactaccgtttcctgttaagaagatagctaaagactttaaactaact gttcttaaaggtgatattgattaccacaaagaaagaccagtcggctataagataacacccgaagaatacg cctatattaaaaacgatattcagattattgcggaagctctgttaattcagtttaagcaaggtttagaccg gatgacagcaggcagtgacagtctaaaaggtttcaaggatattataaccactaagaaattcaaaaaggtg tttcctacattgagtcttggactcgataaggaagtgagatacgcctatagaggtggttttacatggttaa atgataggttcaaagaaaaagaaatcggagaaggcatggtcttcgatgttaatagtctatatcctgcaca gatgtatagtcgtctccttccatacggtgaacctatagtattcgagggtaaatacgtttgggacgaagat tacccactacacatacagcatatcagatgtgagttcgaattgaaagagggctatatacccactatacaga taaaaagaagtaggttttataaaggtaatgagtacctaaaaagtagcggcggggagatagccgacctctg gttgtcaaatgtagacctagaattaatgaaagaacactacgatttatataacgttgaatatatcagcggc ttaaaatttaaagcaactacaggtttgtttaaagattttatagataaatggacgtacatcaagacgacat cagaaggagcgatcaagcaactagcaaaactgatgttaaacagtctatacggtaaattcgctagtaaccc tgatgttacagggaaagtcccttatttaaaagagaatggggcgctaggtttcagacttggagaagaggaa acaaaagaccctgtttatacacctatgggcgttttcatcactgcatgggctagatacacgacaattacag cggcacaggcttgttatgatcggataatatactgtgatactgacagcatacatttaacgggtacagagat acctgatgtaataaaagatatagttgaccctaagaaattgggatactgggcacatgaaagtacattcaaa agagctaaatatctgagacagaagacctatatacaagacatctatatgaaagaagtagatggtgccttag tagaaggtagtccagatgattacactgatataaaatttagtgttaaatgtgcgggaatgactgacaagat taagaaagaggttacgtttgagaatttcaaagtcggattcagtcggaaaatgaagcctaagcctgtgcaa gtgccgggcggggtggttctggttgatgacacattcacaatcaaataa
22	NipTuck_1-N62D nucleotide sequence	atgaagcacatgccgagaaagatgtatagttgtgactttgagacaactactaaagtggaagactgtaggg tatgggcgtatggttatatgaatatagaagatcacagtgagtacaaaatagg taatagcctggatgagtt tatggcgtgggtgttgaaggtacaagctgatctatatttccatgatctcaaatttgacggagcttttatc attaactggttggaacgtaatggttttaagtggtcggctgacggattgccaaacacatataatacgatca tatctcgcatgggacaatggtacatgattgatatatgtttaggctacaaagggaaacgtaagatacatac agtgatatatgacagcttaaagaaactaccgtttcctgttaagaagatagctaaagactttaaactaact gttcttaaaggtgatattgattaccacaaagaaagaccagtcggctataagataacacccgaagaatacg cctatattaaaaacgatattcagattattgcggaagctctgttaattcagtttaagcaaggtttagaccg gatgacagcaggcagtgacagtctaaaaggtttcaaggatattataaccactaagaaattcaaaaaggtg tttcctacattgagtcttggactcgataaggaagtgagatacgcctatagaggtggttttacatggttaa atgataggttcaaagaaaaagaaatcggagaaggcatggtcttcgatgttaatagtctatatcctgcaca gatgtatagtcgtctccttccatacggtgaacctatagtattcgagggtaaatacgtttgggacgaagat tacccactacacatacagcatatcagatgtgagttcgaattgaaagagggctatatacccactatacaga taaaaagaagtaggttttataaaggtaatgagtacctaaaaagtagcggcggggagatagccgacctctg gttgtcaaatgtagacctagaattaatgaaagaacactacgatttatataacgttgaatatatcagcggc ttaaaatttaaagcaactacaggtttgtttaaagattttatagataaatggacgtacatcaagacgacat cagaaggagcgatcaagcaactagcaaaactgatgttaaacagtctatacggtaaattcgctagtaaccc tgatgttacagggaaagtcccttatttaaaagagaatggggcgctaggtttcagacttggagaagaggaa acaaaagaccctgtttatacacctatgggcgttttcatcactgcatgggctagatacacgacaattacag cggcacaggcttgttatgatcggataatatactgtgatactgacagcatacatttaacgggtacagagat acctgatgtaataaaagatatagttgaccctaagaaattgggatactgggcacatgaaagtacattcaaa agagctaaatatctgagacagaagacctatatacaagacatcaaggatggagagtttagtgttaaatgtg cgggaatgactgacaagattaagaaagaggttacgtttgagaatttcaaagtcggattcagtcggaaaat gaagcctaagcctgtgcaagtgccgggcggggtggttctggttgatgacacattcacaatcaaataa
23	NipTuck_2-N62D nucleotide sequence	atgaagcacatgccgagaaagatgtatagttgtgactttgagacaactactaaagtggaagactgtaggg tatgggcgtatggttatatgaatatagaagatcacagtgagtacaaaatagg taatagcctggatgagtt tatggcgtgggtgttgaaggtacaagctgatctatatttccatgatctcaaatttgacggagcttttatc attaactggttggaacgtaatggttttaagtggtcggctgacggattgccaaacacatataatacgatca tatctcgcatgggacaatggtacatgattgatatatgtttaggctacaaagggaaacgtaagatacatac agtgatatatgacagcttaaagaaactaccgtttcctgttaagaagatagctaaagactttaaactaact gttcttaaaggtgatattgattaccacaaagaaagaccagtcggctataagataacacccgaagaatacg cctatattaaaaacgatattcagattattgcggaagctctgttaattcagtttaagcaaggtttagaccg gatgacagcaggcagtgacagtctaaaaggtttcaaggatattataaccactaagaaattcaaaaaggtg tttcctacattgagtcttggactcgataaggaagtgagatacgcctatagaggtggttttacatggttaa atgataggttcaaagaaaaagaaatcggagaaggcatggtcttcgatgttaatagtctatatcctgcaca gatgtatagtcgtctccttccatacggtgaacctatagtattcgagggtaaatacgtttgggacgaagat tacccactacacatacagcatatcagatgtgagttcgaattgaaagagggctatatacccactatacaga taaaaagaagtaggttttataaaggtaatgagtacctaaaaagtagcggcggggagatagccgacctctg gttgtcaaatgtagacctagaattaatgaaagaacactacgatttatataacgttgaatatatcagcggc ttaaaatttaaagcaactacaggtttgtttaaagattttatagataaatggacgtacatcaagacgacat cagaaggagcgatcaagcaactagcaaaactgatgttaaacagtctatacggtaaattcgctagtaaccc tgatgttacagggaaagtcccttatttaaaagagaatggggcgctaggtttcagacttggagaagaggaa acaaaagaccctgtttatacacctatgggcgttttcatcactgcatgggctagatacacgacaattacag cggcacaggcttgttatgatcggataatatactgtgatactgacagcatacatttaacgggtacagagat acctgatgtaataaaagatatagttgaccctaagaaattgggatactgggcacatgaaagtacattcaaa agagctaaatatctgagacagaagacctatatacaagacatcgacggctttagtgttaaatgtgcgggaa tgactgacaagattaagaaagaggttacgtttgagaatttcaaagtcggattcagtcggaaaatgaagcc taagcctgtgcaagtgccgggcggggtggttctggttgatgacacattcacaatcaaataa

24	K135A-N62D amino acid sequence	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hdlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiaadfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv nslypaqmys rllpygepiv fegkyvwded yplhiqhirc efelkegyip tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttsegaikq laklmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahestfk rakylrqkty iqdiymkevd gklvegspdd ytdikfsvkc agmtdkikke vtfenfkvgf srkmkpkpvq vpggvvlvdd tftik
25	E375H-N62D amino acid sequence	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hdlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiakdfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv nslypaqmys rllpygepiv fegkyvwded yplhiqhirc efelkegyip tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttshgaikq laklmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahestfk rakylrqkty iqdiymkevd gklvegspdd ytdikfsvkc agmtdkikke vtfenfkvgf srkmkpkpvq vpggvvlvdd tftik
26	E375S-N62D amino acid sequence	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hdlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiakdfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv nslypaqmys rllpygepiv fegkyvwded yplhiqhirc efelkegyip tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttssgaikq laklmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahestfk rakylrqkty iqdiymkevd gklvegspdd ytdikfsvkc agmtdkikke vtfenfkvgf srkmkpkpvq vpggvvlvdd tftik
27	E375K-N62D amino acid sequence	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hdlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiakdfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv nslypaqmys rllpygepiv fegkyvwded yplhiqhirc efelkegyip tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttskgaikq laklmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahestfk rakylrqkty iqdiymkevd gklvegspdd ytdikfsvkc agmtdkikke vtfenfkvgf srkmkpkpvq vpggvvlvdd tftik
28	E375R-N62D amino acid sequence	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hdlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiakdfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv

		nslypaqmys rllpygepiv fegkyvwded yplhiqhirc efelkegyip tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttsrgaikq laklmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahestfk rakylrqkty iqdiymkevd gklvegspdd ytdikfsvkc agmtdkikke vtfenfkvgf srkmkpkpvq vpggvvlvdd tftik
29	L384R-N62D amino acid sequence	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hdlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiakdfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv nslypaqmys rllpygepiv fegkyvwded yplhiqhirc efelkegyip tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttsegaikq lakrmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahestfk rakylrqkty iqdiymkevd gklvegspdd ytdikfsvkc agmtdkikke vtfenfkvgf srkmkpkpvq vpggvvlvdd tftik
30	E486A-N62D amino acid sequence	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hdlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiakdfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv nslypaqmys rllpygepiv fegkyvwded yplhiqhirc efelkegyip tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttsegaikq laklmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahastfk rakylrqkty iqdiymkevd gklvegspdd ytdikfsvkc agmtdkikke vtfenfkvgf srkmkpkpvq vpggvvlvdd tftik
31	E486D-N62D amino acid sequence	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hdlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiakdfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv nslypaqmys rllpygepiv fegkyvwded yplhiqhirc efelkegyip tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttsegaikq laklmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahdstfk rakylrqkty iqdiymkevd gklvegspdd ytdikfsvkc agmtdkikke vtfenfkvgf srkmkpkpvq vpggvvlvdd tftik
32	K512A-N62D amino acid sequence	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hdlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiakdfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv nslypaqmys rllpygepiv fegkyvwded yplhiqhirc efelkegyip tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttsegaikq laklmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahestfk rakylrqkty iqdiymkevd galvegspdd ytdikfsvkc agmtdkikke vtfenfkvgf srkmkpkpvq vpggvvlvdd tftik

33	NipTuck_1-N62D amino acid sequence (deletion of residues 505-525)	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hdlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiakdfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv nslypaqmys rllpygepiv fegkyvwded yplhighirc efelkegyip tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttsegaikq laklmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahestfk rakylrqkty iqdikdgefs vkcagmtdki kkevtfenfk vgfsrkmkpk pvqvpggvvl vddtftik
34	NipTuck_2-N62D amino acid sequence (deletion of residues 505-525)	mkhmprkmys cdfetttkve dcrvwaygym niedhseyki gnsldefmaw vlkvqadlyf hdlkfdgafi inwlerngfk wsadglpnty ntiisrmgqw ymidiclgyk gkrkihtviy dslkklpfpv kkiakdfklt vlkgdidyhk erpvgykitp eeyayikndi qiiaealliq fkqgldrmta gsdslkgfkd iittkkfkkv fptlslgldk evryayrggf twlndrfkek eigegmvfdv nslypaqmys rllpygepiv fegkyvwded yplhighirc efelkegyip tiqikrsrfy kgneylkssg geiadlwlsn vdlelmkehy dlynveyisg lkfkattglf kdfidkwtyi kttsegaikq laklmlnsly gkfasnpdvt gkvpylkeng algfrlgeee tkdpvytpmg vfitawaryt titaaqacyd riiycdtdsi hltgteipdv ikdivdpkkl gywahestfk rakylrqkty iqdidgfsvk cagmtdkikk evtfenfkvg fsrkmkpkpv qvpggvvlvd dtftik

Characterization of recombinant polymerases with nucleotide analogues

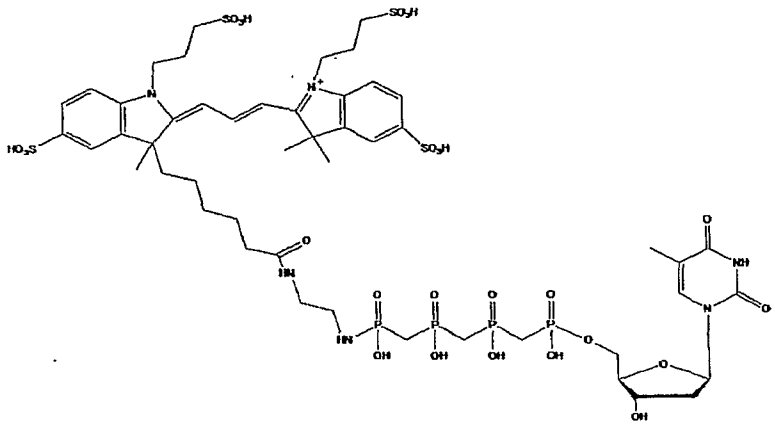
[0162] K_m and V_{max} were determined for exemplary recombinant Phi29 polymerases and various nucleotide analogues. Results are presented in Table 4.

Table 4. K_m and V_{max} versus analogues.

Mutation	Km^1	$Vmax^1$	Km^2	$Vmax^2$	Km^3	$Vmax^3$	Km^4	$Vmax^4$
N62D	23	610	20	540	838	2500	68	1620
N62D:E375H	17	800	15	526	433	1250		
N62D:E375S	16.5	1158					40	1981
N62D:E375K	12	595						
N62D:E375Y	2.5	773	6.6	471	440	1430	18	1292
N62D:E375W	1.8	889	5.0	595	248	1428	16	1585

[1]Measured for Alexa633-O-dC4P (also referred to as A633dC4P herein)

[2]Measured for Alexa555-C2-dT4P. This analogue has a 2-carbon linker ("C2") between the delta phosphate and the label moiety and has the following structure:



[3]Measured for Alexa555-C2-dTTP
[4]Measured for Alexa532-O-dG4P

[0163] A set of exemplary recombinant Phi29 polymerases were characterized with various nucleotides and/or nucleotide analogues. Results are presented in Table 5.

Table 5. Screening data.

Mutation	Ratio (Low/High)[1]	Rate High[2]	Hz Rate[3]	Mutation 2[4]	Tag(s)[5]
E375W	0.677	699.4	19.1	N62D	His
E375Y	0.694	498.5	12.1	N62D	His
E375H	0.445	510.1	9.4	N62D	His
E375Q	0.356	531.6	6.3	N62D	His
E375K	0.425	516.1	6.2	N62D	His
E375S	0.335	528.4	5.9	N62D	His
E375A	0.383	465.9	5.8	N62D	His
T15I	0.355	416.5	3.0		His
N62D	0.355	349.3	2.8		GST-His
N62D	0.362	373.2	2.7		His
K135A	0.412	272.0	1.6	N62D	His
K512A	0.335	347.7	1.4	N62D	His
NipTuck1	0.508	192.4	1.3	N62D	His
D12A	0.888	55.3	1.2		GST-His
E486A	0.441	152.5	1.0	N62D	His
E486D	0.467	142.0	0.9	N62D	His
T15I	0.726	105.9	0.0	N62D	His
NipTuck2	0.635	156.3	0.0	N62D	His
L384R	0.768	79.6	-0	N62D	His

[1]Ratio = (rate at 5 μM A633dC4P with 20 μM dA,dG,dTTP)/(rate at 25 μM A633dC4P with 20 μM dA,dG,dTTP) at 1 mM $MnCl_2$. A higher ratio corresponds to a lower Km.
[2]Rate at 25 μM A633dC4P with 20 μM dA,dG,dTTP
[3]Rate at 10 μM Alexa488-O-dA4P, 10 μM FAM-Alexa532-O-dG4P, 10 μM FAM-Alexa594-O-dT4P, 10 μM Alexa633-O-dC4P with 1 mM $MnCl_2$. Provides a measure of both Km and Vmax, with a representative set of four nucleotide analogues.

[4]Background mutation (if any). The recombinant polymerase corresponds to wild type Phi29 polymerase plus mutation 1 plus mutation 2.
[5]Tag for immobilization and or purification

[0164] Rates of binding and product release were determined for exemplary recombinant Phi29 polymerases with nucleotide analogue A594-dT4P using a FRET stopped flow assay as schematically illustrated in Figure 7 Panel A. Results are depicted graphically in Figure 7 for Phi29 N62D (Panel B), N62D:E375Y (Panel C), and N62D:E375W (Panel D). Product release rates are shown in Table 6.

[0165] The E375Y and E375W mutant polymerases demonstrated increased rates of binding and product release, indicating they utilize the analogue better than does the parent enzyme.

Table 6. Product release rate

Enzyme	Product Release Rate
N62D	55 s^{-1}
N62D:E375Y	117 s^{-1}
N62D:E375W	76 s^{-1}

[0166] Relative branching rate (dissociation of the analogue without incorporation, i.e., substrate dissociation) was also determined for exemplary recombinant Phi29 polymerases with nucleotide analogue Alexa568-dA4P (also called A568-dA4P), using a FRET stopped flow assay as schematically illustrated in Figure 8 Panel A. In this technique, a template with a FRET donor dye compatible for FRET with the corresponding dye on the nucleotide analogue is employed. The primer has a dideoxy-termination at the 3' end to prevent incorporation. The analogue is pre-mixed with the enzyme-template-dideoxyprimer complex. In the stopped flow apparatus, this preformed complex is rapidly mixed with the corresponding native nucleotide (native dATP, in this example) in excess which serves as a "trap" to prevent rebinding of the analogue after it dissociates. The increase in donor dye fluorescence is monitored as a means of monitoring the dissociation/branching rate of the analogue.

[0167] Results are depicted graphically in Figure 8 for Phi29 N62D (Panel B), N62D:E375Y (Panel C), and N62D:E375W (Panel D). Branching rates are shown in Table 7.

Table 7. Branching rate.

Enzyme	Branching Rate
N62D	90 s^{-1}
N62D:E375Y	31 s^{-1}
N62D:E375W	43 s^{-1}

Additional exemplary recombinant polymerases

[0168] Polymerases of the invention can include a Phi29 polymerase (or homolog thereof) including any of the mutations listed in Table 8, singly or in combination with other mutations (e.g., other mutations described herein). For example, polymerases of the invention optionally include a Phi29 polymerase (or homolog thereof) that includes a combination of mutations as specified in Table 8.

Table 8. Exemplary mutations.

D12A E375W T372D
D12A E375W T372E
D12A E375W T372R K478D
D12A E375W T372R K478E
D12A E375W T372K K478D
D12A E375W T372K D478E
D12A E375W K135D
D12A E375W K135E
D12A E375W K512D
D12A E375W K512E
D12A E375W E408K
D12A E375W E408R
D12A E375W T368D L480K
D12A E375W T368E L480K
D12A D456N
N62D D456N
D12A D456A
N62D D456A
D12A D456S
N62D D456S
N62D E375M
N62D E375L
N62D E375I
N62D E375F
N62D E375D
D12A K512W
N62D K512W
D12A K512Y
N62D K512Y

D12A K512F
N62D K512F
D12A E375W K512L
N62D E375W K512L
D12A E375W K512Y
N62D E375W K512Y
D12A E375W K512F
N62D E375W K512F
D12A E375Y K512L
N62D E375Y K512L
D12A E375Y K512Y
N62D E375Y K512Y
D12A E375Y K512F
N62D E375Y K512F
D12A E375W K512H
N62D E375W K512H
D12A E375Y K512H
N62D E375Y K512H
D12A D510F
N62D D510F
D12A D510Y
N62D D510Y
D12A D510W
N62D D510W
D12A E375W D510F
N62D E375W D510F
D12A E375W D510Y
N62D E375W D510Y
D12A E375W D510W
N62D E375W D510W
D12A E375W D510W K512L
N62D E375W D510W K512L
D12A E375W D510W K512F
N62D E375W D510W K512F
D12A E375W D510H
N62D E375W D510H
D12A E375W D510H K512H
N62D E375W D510H K512H
D12A E375W D510H K512F
N62D E375W D510H K512F
D12A V509Y
N62D V509Y
D12A V509W
N62D V509W
D12A V509F
N62D V509F
D12A V514Y
N62D V514Y
D12A V514W

N62D V514W
D12A V514F
N62D V514F
D12S
D12N
D12Q
D12K
D12A
N62D Y254F
N62D Y254V
N62D Y254A
N62D Y390F
N62D Y390A
N62D S252A
N62D N387A
N62D K157E
N62D I242H
N62D Y259S
N62D G320C
N62D L328V
N62D T368M
N62D T368G
N62D Y369R
N62D Y369H
N62D Y369E
N62D I370V
N62D I370K
N62D K371Q
N62D T372N
N62D T372D
N62D T372R
N62D T372L
N62D T373A
N62D T373H
N62D S374E
N62D I378K
N62D K379E
N62D K379T
N62D N387D
N62D Y405V
N62D L408D
N62D G413D
N62D D423V
N62D I442V
N62D Y449F
N62D D456V
N62D L480M
N62D V509K
N62D V509I

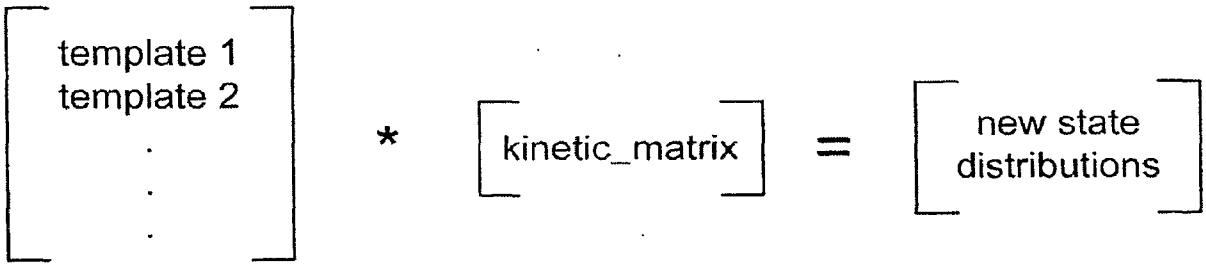
N62D D510A
N62D V514I
N62D V514K
N62D E515K
N62D D523T
N62D H149Y E375W M554S
M8S N62D M102S H116Y M188S E375W
N62D M97S E375W
M8S N62D M97S M102S M188S E375W M554S
M8A N62D M97A M102A M188A E375W M554A

[0169] A few mutations in the Phi29 polymerase have been previously described. For the N62D mutation, see de Vega et al. (1996) "Primer-terminus stabilization at the 3'-5' exonuclease active site of phi29 DNA polymerase. Involvement of two amino acid residues highly conserved in proofreading DNA polymerases" EMBO J. 15(5):1182-92. For the D12A mutation and mutations at positions E14, 66, 165, 169, 12 and 66, and 14 and 66, see Esteban et al. (1994) "3'-->5' exonuclease active site of phi 29 DNA polymerase. Evidence favoring a metal ion-assisted reaction mechanism" J Biol Chem. 269(50):31946-54. For mutation of S252, see Blasco et al. (1993) "Phi 29 DNA polymerase active site. Residue ASP249 of conserved amino acid motif 'Dx2SLYP' is critical for synthetic activities" J Biol Chem. 268(32):24106-13. For mutation of Y254, see Blasco et al. (1992) "Phi 29 DNA polymerase active site. Mutants in conserved residues Tyr254 and Tyr390 are affected in dNTP binding" J Biol Chem. 267(27):19427-34. For mutation of K371, see Truniger et al. (2002) "A positively charged residue of phi29 DNA polymerase, highly conserved in DNA polymerases from families A and B, is involved in binding the incoming nucleotide" Nucleic Acids Res. 30(7):1483-92. For mutation of K379, see Truniger et al. (2004) "Two Positively Charged Residues of φ29 DNA Polymerase, Conserved in Protein-primed DNA Polymerases, are Involved in Stabilisation of the Incoming Nucleotide" Journal of Molecular Biology 335(2):481-494. For mutation of N387, see Blasco et al. (1993) "Phi 29 DNA polymerase active site. The conserved amino acid motif 'Kx3NSxYG' is involved in template-primer binding and dNTP selection" J Biol Chem. 268(22):16763-70. For mutation of Y390, see Blasco et al (1992) "Phi 29 DNA polymerase active site. Mutants in conserved residues Tyr254 and Tyr390 are affected in dNTP binding" J Biol Chem. 267(27):19427-34. For mutation of D456, see Bernad et al. (1990) "The highly conserved amino acid sequence motif Tyr-Gly-Asp-Thr-Asp-Ser in alpha-like DNA polymerases is

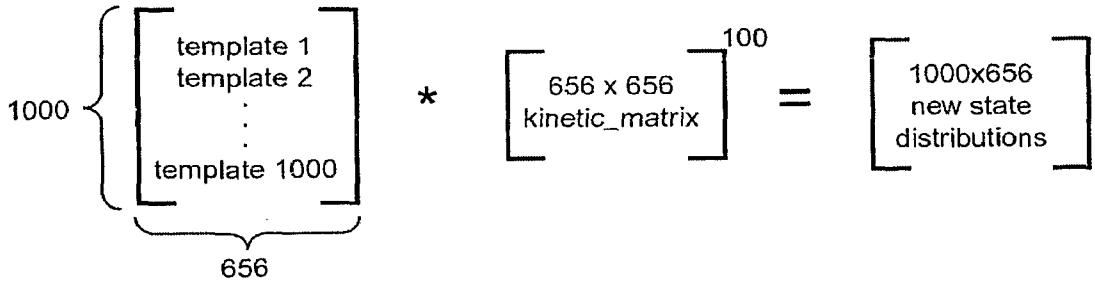
required by phage phi 29 DNA polymerase for protein-primed initiation and polymerization" Proc Natl Acad Sci U S A. 87(12):4610-4.

EXAMPLE 4: A COMPUTATIONAL FRAMEWORK FOR MODELING AND TESTING THE ENZYMATIC KINETICS OF DNA POLYMERASE, ADDRESSING ALL KINETIC PROCESSES AND FREE VARIABLES SIMULTANEOUSLY.

[0170] Polymerase kinetic state transitions are stored in a probability matrix for discrete time steps. A vector of probabilistic state distributions may describe the probability of finding a particular polymerase in a number of polymerase states according to a continuum model. Linear algebra multiplication of the state distribution vector with the state transition probability matrix gives a new vector of polymerase state distributions, describing the effect of the passage of time equal to the discrete time step of the state transition probability matrix.



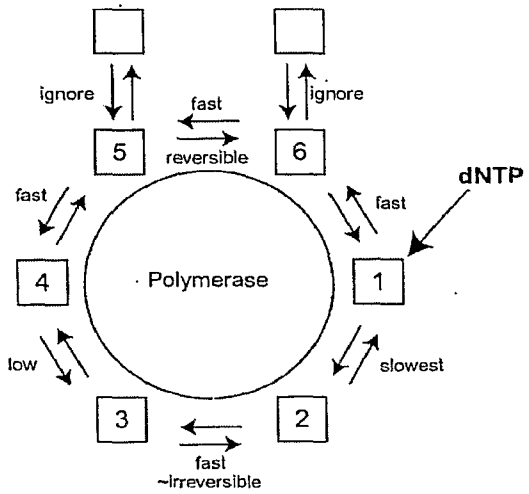
By raising the state transition probability matrix to a particular exponential power (eg. 100), we simulate the passage of time of a particular number of discrete time steps (eg. 100 time steps). Using many discrete time steps we simulate DNA polymerization. Steady State Model.



The transition rates are user-defined. The probability matrix is automatically generated using the template sequence and hard-coded state transition rules. A variety of parameters,

such as reagent concentrations, kinetic rate values, and probability matrix organization can vary from those described in this example.

[0171] The following is an example of a steady state polymerase kinetic model.



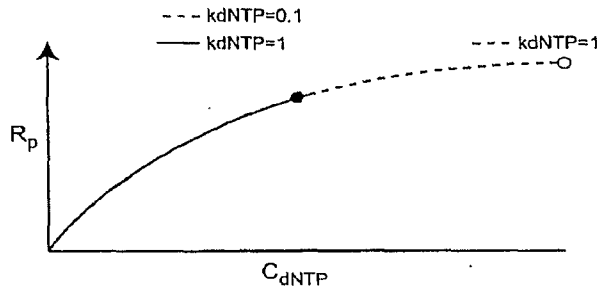
$$R_p = C_6K_{61} - C_1K_{16} = C_1K_{12} - C_2K_{21}$$

$$= C_2K_{23} - C_3K_{32}$$
$$= C_3K_{34} - C_4K_{43}$$
$$= C_4K_{45}$$
$$= C_5K_{56} - C_6K_{65}$$

R_p = rate of catalysis
C_6 = probability of finding polymerase in state 6
K_{61} = transition rate of polymerase in state 6 to state 1
k_{ij} = reaction rated
$P_{ij} = k_{ij}\Delta t$ reaction rated
P_{ij} = i $\rightarrow$ j probability

* $K_{54} \approx 0$ as concentration of pyrophosphate
$\downarrow$
$R_p = C_6K_{61} - C_1K_{16}$ = rate of catalysis
$R_p = (R_p)_{max}$ @ $K_{61} \rightarrow \infty$, $C_6 \rightarrow 0$ as a condition of nucleotide concentration increasing to saturation

To find $(R_p)_{max}$

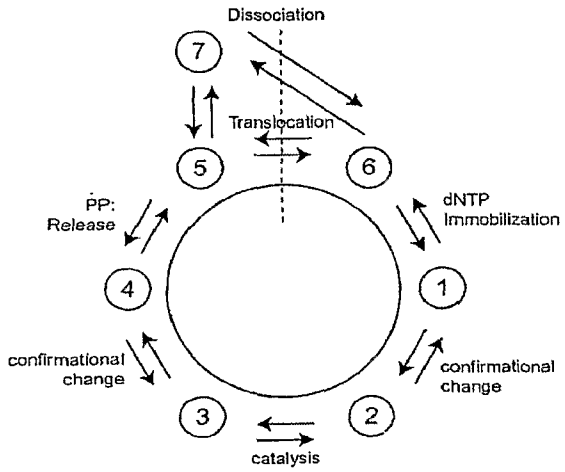


—— $\overline{P}$~ 0.01, $\overline{\Delta t}$ ~0.01
---- $\overline{P}$~ 0.001, $\overline{\Delta t}$ ~0.001
} k_{ij} const., $P_{61} = 1$

* As $\Delta t \downarrow$, find assymptote of R_p

Mega Matrix

[0172] The following is a single 2-D matrix to capture all possible kinetic states of a polymerase-template-dNTP system:



Variables				
Pol State	Template Base	Nucleo. Base	Native/Analog	Previous Nucleo. Base
1-4	A-T	A-T	0-1	A-T
5	A-T	A-T	X	A-T
6	A-T	X	X	A-T
7	A-T	A-T	X	A-T

[0173] * This results in a 656-state matrix, where the states are as follows:

1.	1	A	A	0	A
2.	1	A	A	0	C
3.	1	A	A	0	G
4.	1	A	A	0	T
5.	1	A	A	1	A
6.	1	A	A	1	C
7.	1	A	A	1	G
8.	1	A	A	1	T
9.	1	A	C	0	A
652.	7*	T	G	X	T
653.	7*	T	T	X	A
654.	7*	T	T	X	C
655.	7*	T	T	X	G
656.	7*	T	T	X	T

*In this case the state 7 is dissociation of the polymerase from the template, which may optionally be simplified to never happen.

[0174] In this case the DNA template is the repeated sequence (ACGT.) For a longer template repeated sequence there will be proportionally more states, to the extent that the longer template repeated sequence does not contain the original template sequence. For example, the probability transition matrix generated for the sequence

...[ACGT]ACGT...

would be equivalent to the matrix generated for the sequence

...[ACGTACGT]ACGT....

However, the probability transition matrix generated for the sequence

...[AACCGGTT]AACC...

would be different, as it contains many state transitions not allowed in the original matrix (eg. polymerase translocation from an "A" to another "A" in the template sequence. Furthermore, since this repeated sequence contains eight Watson-Crick bases instead of four, it would generate a matrix of 1,312 states instead of 656..

[0175] Some states do not require all variables to be defined (see above table). For example, characteristics of a nucleotide which has not yet been incorporated in state 6 do not affect the identity of state 6.

577.	6	A	X	X	A

[0176] * The Transition rate between two states will be defined as such:

562.	5	T	A	X	C

[0177] P56TAxC = k56TAxC * time_step

Where P56TAxC is the probability of the polymerase completing translocation from state 5 to state 6 with the additional nucleotide-template conditions described by "TAxC". K56TAxC is the transition rate of this translocation.

[0178] Currently in this 656 state system, there are 1568 transition rates to define. There are a number of approximations that can be made to reduce the number of inputs the user needs to enter.

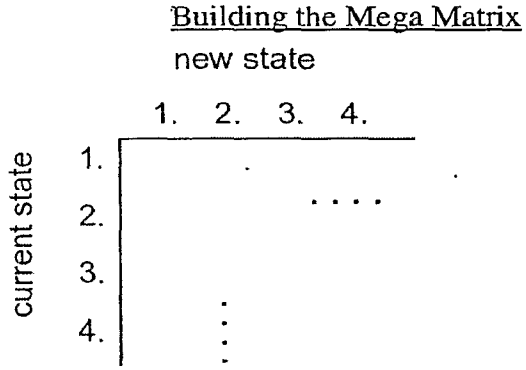
[0179] The following combinations may be treated equivalently in all states transitions:

Template nucleotide: ACGT
TGCA
Likewise, all mismatches may be treated the same

K12AT0A = k12xZ0Z
K12CG0T = k12xZ0Z
K12CT0T = k12xY0Z
K12CT1C = k12xY1Y

X= any variable
Y= any mismatch
Z = any match

[0180] In this way the user input selection is reduced to ~100 unique transition rate variables. All the explicitly defined rates are automatically assigned the appropriate user inputs.



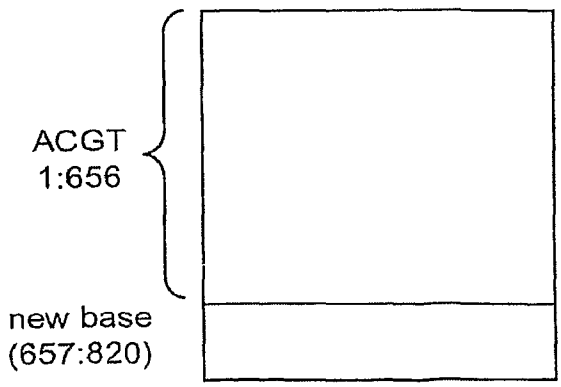
[0181] To capitalize on symmetry for the purpose of inserting user defined transition rates into the matrix automatically, the organization of the 656-state matrix can be changed:

Old					
1.	1	A	A	O	A
2.	1	A	A	O	C
...	...	...	...	...	...
656.	7	T	T	1	T

New					
1.	A	1	A	O	A
2.	A	1	A	O	C
...	...	...	...	...	...
656.	T	7	T	1	T

[0182] This has two advantages:

(1.) the template can be extended with only slight modifications to the matrix. Every Template base in repeated sequence brings an additional 164 states. Previously, new states would have to be interwoven into matrix.



(2) The matrix has a higher degree of symmetry that before, making it easier to construct the matrix using automated code:

```
for ii = 1:164

eval (['. . .']);
. . .

end % ii
```

Seven "eval" statements (a function which evaluates an artificially constructed command) construct seven polymerase states.

[0183] This has been further enhanced to build the matrix for any given template sequence automatically.

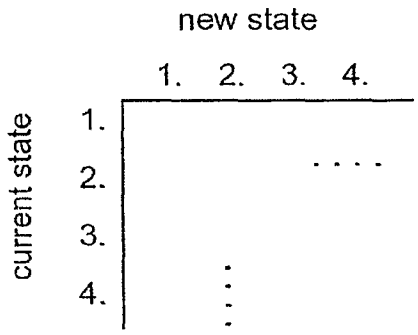
[0184] A further automation of the generation of the state transition probability matrix is through the building of a concentration matrix, which contains the concentrations of all relevant reagents (polymerase, template, nucleotides, etc) . This concentration matrix compliments the rate transition matrix such that (in the linear concentration limit).

kinetic_matrix = rate_transition_matrix .* conc_matrix

state_transition_probability_matrix = kinetic_matrix * time_step

where each element of the rate transition matrix has been multiplied by its corresponding dependent variable in the concentration matrix. In this way we capture the concentration dependent state transitions (eg. the rate of incorporation of nucleotides is dependent upon the concentration of nucleotides). Elements of the matrix which are not concentration dependent are not changed. Non-linear concentration dependencies may be addressed using a nonlinear formula defining the kinetic matrix.

[0185] The following describes the state transition probability matrix:



Matrix = zeros(656,656);

Matrix(1, [1,139,577]) = [1-p12AA0A – p16AA0A, p12AA0A, p16AA0A];
Matrix(2, [2,130,578]) = [1-p12AA0C – p16AA0C, p12AA0C, p16AA0C];

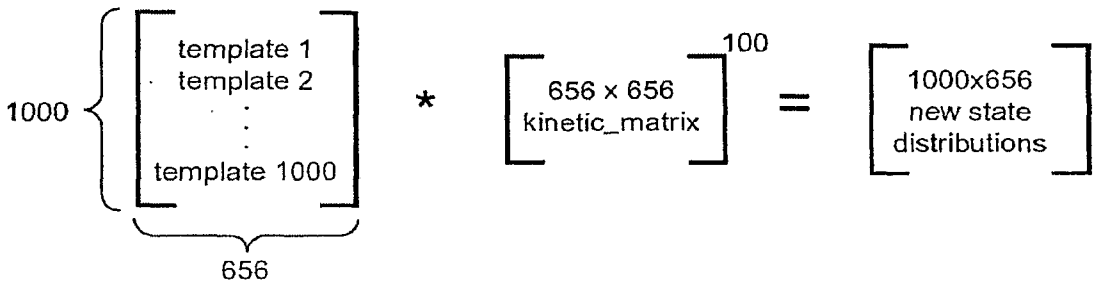
....

Matrix(129, [129,257,1]) = [1-p23AA0A – p21AA0A, p23AA0A, p21AA0A];
....
Matrix(656, [656,580,576]) = [1-p76TTxT – p75TTxT, p76TTxT, p75TTxT];

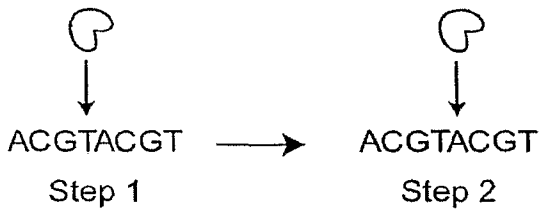
where each of the probability values inserted in the matrix have been calculated using user defined transition rates, concentration values, and a discrete time step. Note that the first element of a row is the probability of having no transition between states, and is thus the difference between 100% and the probabilities of all state transitions out of that particular state.

[0186] Increasing efficiency of simulation:
By raising the state transition probability matrix to a particular exponential power (eg. 100), we simulate the passage of time of a particular number of discrete time steps (eg. 100 time steps).

Further improvements to the efficiency of the simulation may be made through vectorization of many polymerase-template complexes simultaneously.



[0187] Speed Limit: DNA synthesis can be tracked by looking at where pol is on the template.



[1 0 0 0] = "A"
[0 1 0 0] = "C"
etc. . .

[0188] If we move too fast (i.e. too many time steps in the transition matrix exponential), the polymerase may go from "A" straight to "G", making it unclear whether this was forward or reverse translocation. Therefore an error limit (~ 1e-6) is set that defines an exponential time factor on the kinetic_matrix. The speed limit is such that neither the probability of reverse translocation from "A" to "G" nore the probability of forward translocation from "A" to "T" exceeds the error rate limit. A longer DNA repeat sequence will allow us to move faster, but a repeat sequence which is too long will be computationally intensive.

[0189] A further application of this program can be the simulation of reagent consumption rate. Moving at very large step sizes, polymerase movment is simulated along template. This approach uses only one template in a continuous distribution of states (instead of 1000^{+} templates in discrete states). This tracks reagent consumption over time.

[0190] Find the concentration change of reagents based on the current population of the system and based on the transition rate constants:

$d(dTAP_o)_{per\ pol} = C_1\Delta tk16AA0A + C_2\Delta tk16AA0C + \ldots C_{520}\Delta tk16TT0T$

$- C_{145}\Delta tk61AA0A - C_{146}\Delta tk61AA0C - \ldots = C_1p61AA0A + C_{146}p61AA0C + ..$

$C_{145}p61AA0A - C_{146}p61AA0C - \ldots$

[0191] Where these probabilities are for a 1e-6 sec time step from kinetic_matrix:

concentration change (Molar) of reagent dTAP (native) in 1 loop cycle where elapsed time = num_steps * 1e-6 sec

[fast_matrix] = [kinetic_matrix)$^{\wedge num\_steps}$

Speed limit: $\frac{\Delta C_{max}}{C} < 1\%$?

Fast_matrix = kinetic_matrix^n

As N becomes large, the adjustment to concentrations each loop cycle becomes large and inaccurate. This is used to set an exponential time factor on the kinetic_matrix.

See Figure 9, which plots the kinetic matrix jump size vs. concentration drop.

[0192] Even taking num_steps = 1e6 may give accurate "enough" concentration curves (see the approach to smoothness as step size decreases).

[0193] The resulting (4096 x 4096 double matrix is a reasonable memory limit).

[0194] A further application of this program can be the estimation of the polymerase mismatch fraction using either a continuum model or counting model. Currently we say that the 2nd previous template – nucleotide pair is always a match. (This is to reduce size of matrix by 4x . . . the error should be small unless there is lots of exonuclease activity).

[0195] Therefore, any forward translocation from state 5 with a previous mismatch becomes a permanent mismatch (it just won't look that way if we back up).

forward total translocation rate = $C_{5\cdot} * C_{6\cdot} - C_{6\cdot} * k_{65}$

reaction = (mismatch rate) / (total rate)

C_5 represents concentration of all matrix states with pol in state 5 (see pg. 128)

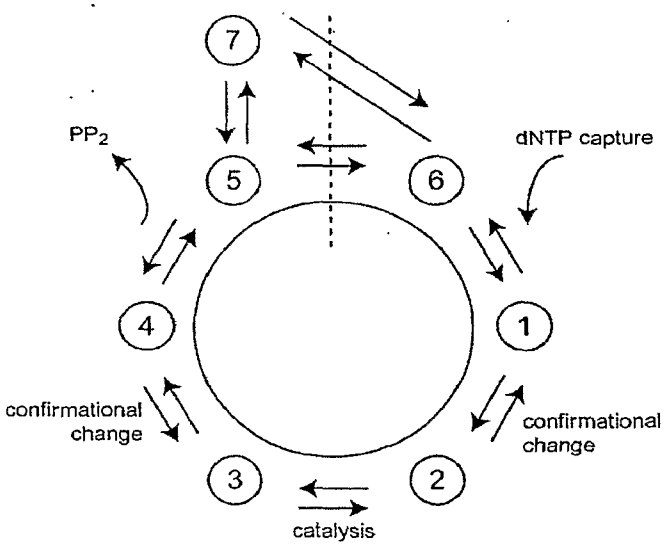
k_{56} is the full set of all corresponding rates for forward translocation

forward mismatch translocation = $C_5^{(m)} . * k_{56}^{(m)} - C_6^{(m)} . * k_{65}^{(m)}$

(In reverse translocation, we never end up in pol state 5 with previous mismatch, *see above*).

[0196] We can also make a counting model which counts number of polymerase/template complexes which have previous template/nucleotide mismatch and which also do forward translocation (making mismatch permanent), and average this over all polymerase to get a mismatch fraction. This should be in the same ballpark as continuum model estimate above.

1) First set all rate constants equal to T7 polymerase as shown by Patel, et al. (1991) "Pre-Steady-State Kinetic Analysis of Processive DNA Replication Including Complete Characterization of an Exonuclease-Deficient Mutant" Biochemistry 30:511-525.



[0197] Specific rate constants, etc.

$K_{61} \geq 50\mu m^{-1} s^{-1}$
$K_{12} = 300\mu m^{-1} s^{-1}$
$K_{23} \geq 9000\mu m^{-1} s^{-1}$
$K_{34} = 1200\mu m^{-1} s^{-1}$

$K_{645} \geq 1000\mu m^{-1} s^{-1}$

$K_{16} \geq 1000\mu m^{-1} s^{-1}$
$K_{21} = 100\mu m^{-1} s^{-1}$
$K_{32} = 18{,}000\mu m^{-1} s^{-1}$
$K_{43} = 18\mu m^{-1} s^{-1}$
$K_{54} \geq 0.5\mu m^{-1} s^{-1}$

$(V_{max})_{native} = 50$ bps
$(V_{max})_{analog} = 5$ bps
$(k_m)_{native} = 0.2\ \mu m$
$(k_m)_{analog} = 6\ \mu m$

2) Using dNTP concentration saturation ($\geq$ 1mM), set $V_{max} = 50$ bps by changing k_{12} (primarily) and other rate constants (if necessary). Keep all analog transition rates the same as native dNTP transition rates. For now cut dissociation (rate $\rightarrow$ 0)

3) Using analog – dNTP concentration saturation ($\geq$ 1mM), set $V_{max} = 5$ bp by changing k_{45} for analogs only.

4) Set $(k_m)_{native} = 0.2\ \mu m$ by setting native dNTP concentration to 0.2μm and changing k_{61} (natives only) such that V = 25bps.

5) Set $(k_m)_{native} = 6\ \mu m$ by setting analog dNTP concentration to 6 μm and changing k_{61} (analogs only) such that V- 2.5 bps.

native dNTP's
$k_{61} = 365\mu m^{-1} s^{-1}$
$k_{12} = 60\mu m^{-1} s^{-1}$
$k_{23} = 9000\mu m^{-1} s^{-1}$
$k_{34} = 1200\mu m^{-1} s^{-1}$
$k_{45} = 1000\mu m^{-1} s^{-1}$
$k_{56} = 500\mu m^{-1} s^{-1}$
$k_{16} = 10\mu m^{-1} s^{-1}$
$k_{21} = 100\mu m^{-1} s^{-1}$
$k_{32} = 1800\mu m^{-1} s^{-1}$
$k_{43} = 18\mu m^{-1} s^{-1}$
$k_{54} = 0.5\mu m^{-1} s^{-1}$
$k_{65} = 100\mu m^{-1} s^{-1}$

analog dNTP's
$k_{61} = 1.1\mu m^{-1} s^{-1}$
$k_{12} = 60\mu m^{-1} s^{-1}$
$k_{23} = 9000\mu m^{-1} s^{-1}$
$k_{34} = 5.5\mu m^{-1} s^{-1}$

$k_{45} = 5.5\mu m^{-1} s^{-1}$
$k_{56} = 500\mu m^{-1} s^{-1}$
$k_{16} = 10\mu m^{-1} s^{-1}$
$k_{21} = 100\mu m^{-1} s^{-1}$
$k_{32} = 1800\mu m^{-1} s^{-1}$
$k_{43} = 18\mu m^{-1} s^{-1}$
$k_{54} = 0.1\mu m^{-1} s^{-1}$
$k_{65} = 100\mu m^{-1} s^{-1}$

All rates will be subject to calibration by future experiments as well.

pol_index.m: Initializes all necessary matrix index lists and pointers based on DNA sequence.

Pol_ratematrix.m: Takes excel file as input, which contains a list of all unique rate constants, produces transition rate matrix based on DNA sequence.

Pol_conmatrix.m: Takes reagent concentrations, builds concentration matrix such that:

Probability matrix = time_step * rate_matrix * conc_matrix

(for all non-diagonal elements)

Pol_dntp_concumption.m: Calculates reagent consumption rates based on continuum model.

POL_dna.m: Combines all former functions of POL_DNA, POL_REAGENTS, POL_CURVEMAP,
tracks all former consumption,
tracts length distribution of DNA synthesis,
tracks free template, completed dsDNA template, template currently being worked on, multiple concentration runs possible
user defined repeating DNA sequence, finite length templates

pol_metal.m: Full embodiment of Mg+ depletion experiment, using stripped down version of POL_DNA.

[0198] While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and apparatus described above can be used in various combinations. All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application,

and/or other document were individually and separately indicated to be incorporated by reference for all purposes.

WHAT IS CLAIMED IS:

1. A composition comprising a recombinant DNA polymerase, the recombinant DNA polymerase comprising:

a modified active site region that is homologous to a wild-type active site region of a wild-type DNA polymerase, the modified active site region comprising one or more structural modification relative to the wild type active site region that reduces steric inhibition for entry of a nucleotide analogue into the modified active site region, or that is complementary with one or more non-natural features of the nucleotide analogue, wherein the recombinant DNA polymerase displays a modified property for the nucleotide analogue as compared to the wild-type polymerase.

2. The composition of claim **1**, wherein the recombinant DNA polymerase is homologous to a Φ29 DNA polymerase, a Taq polymerase, an exonuclease deficient Taq polymerase, a DNA Pol I polymerase, a T7 polymerase, a T5 Polymerase, an RB69 polymerase, a T5 polymerase or a polymerase corresponding to a Klenow fragment of a DNA Pol I polymerase.

3. The composition of claim **1**, wherein the recombinant DNA polymerase is homologous to a wild-type or exonuclease deficient Φ29 DNA polymerase.

4. The composition of claim **1**, wherein the recombinant DNA polymerase is homologous to Φ29, B103, GA-1, PZA, Φ15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, or L17.

5. The composition of claim **3**, wherein the recombinant DNA polymerase comprises a structural modification within or proximal to its active site relative to the wild-type or to an exonucleoase deficient Φ29 DNA polymerase selected from: a deletion of residues 505-525, a deletion within residues 505-525, a K135A mutation, an E375H mutation, an E375S mutation, an E375K mutation, an E375R mutation, an E375A mutation, an E375Q mutation, an E375W mutation, an E375Y mutation, an E375F mutation, an E486A mutation, an E486D mutation, a K512A mutation, and combinations thereof.

6. The composition of claim **5**, wherein the recombinant DNA polymerase further comprises an L384R mutation.

7. The composition of claim **5**, wherein the recombinant DNA polymerase further comprises an additional mutation or combination of mutations selected from Table 8.

8. The composition of claim **1**, wherein the recombinant DNA polymerase comprises a structural modification that reduces exonuclease activity of the recombinant polymerase relative to the wild-type polymerase.

9. The composition of claim **8**, wherein the polymerase is homologous to a Φ29 DNA polymerase and the structural modification is a deletion or alteration of an amino acid that results in reduced exonuclease activity.

10. The composition of claim **9**, wherein the alteration is corresponds to a mutation to N62 relative to wild-type Φ29 DNA polymerase.

11. The composition of claim **1**, wherein the recombinant DNA polymerase comprises one or more exogenous affinity tag sequence.

12. The composition of claim **11**, wherein the affinity tag sequence is selected from: a 6 His tag sequence, a GST tag, an HA tag sequence, a plurality of 6 His tag sequences, a plurality of GST tags, a plurality of HA tag sequences and combinations thereof.

13. The composition of claim **1**, wherein the recombinant DNA polymerase is selected from Table 3.

14. The composition of claim **1**, comprising the nucleotide analogue.

15. The composition of claim **14**, wherein the nucleotide analogue comprises a fluorophore, or a dye moiety.

16. The composition of claim **14**, wherein the nucleotide analogue is a phosphate-labelled nucleotide analogue.

17. The composition of claim **14**, wherein the nucleotide analogue is a mono-deoxy or dideoxy nucleotide analogue.

18. The composition of claim **14**, wherein the nucleotide analogue is a labeled nucleotide analogue having from 3 to 6 phosphate groups.

19. The composition of claim **14**, wherein the nucleotide analogue is a triphosphate, a tetra phosphate, a penta phosphate or a hexaphosphate.

20. The composition of claim **18**, wherein the structural modification includes the addition of a positively charged amino acid residue that binds to a phosphate residue of the nucleotide analogue.

21. The composition of claim **1**, wherein the modified property is selected from: K_m, k_{cat}, V_{max}, recombinant polymerase processivity in the presence of the nucleotide analogue, average template read-length by the recombinant polymerase in the presence of the nucleotide analogue, specificity of the recombinant polymerase for the nucleotide analogue, rate of binding of the nucleotide analogue, rate of product release, and branching rate.

22. The composition of claim **1**, wherein the modified property comprises a reduced K_m for the nucleotide analogue.

23. The composition of claim **1**, wherein the modified property comprises an increased k_{cat}/K_m or V_{max}/K_m for the nucleotide analogue.

24. The composition of claim **1**, wherein the recombinant polymerase displays a specific activity for a natural nucleotide that is at least about 5% as high as the wild-type polymerase and a processivity in the presence of a template that is at least 5% as high as the wild-type polymerase in the presence of the natural nucleotide.

25. The composition of claim **1**, wherein the recombinant polymerase displays a k_{cat}/K_m or V_{max}/K_m for a naturally occuring nucleotide that is at least about 5% as high as the wild-type polymerase.

26. The composition of claim **1**, wherein the recombinant polymerase displays a k_{cat}/K_m or V_{max}/K_m for a naturally occuring nucleotide that is at least about 25% as high as the wild-type polymerase.

27. The composition of claim **1**, comprising the nucleotide analogue and a DNA template, wherein the recombinant polymerase incorporates the nucleotide analogue into a copy nucleic acid in response to the template DNA.

28. The composition of claim **27**, wherein the template is a circular template.

29. The composition of claim **1**, wherein the composition is present in a DNA sequencing system.

30. The composition of claim **29**, wherein the sequencing system comprises a Zero Mode Waveguide.

31. A method of making a DNA, the method comprising:

providing a reaction mixture comprising a template, a replication initiating moiety that complexes with or is integral to the template, a polymerase capable of replicating at least a portion of the template using the moiety in a template-dependent polymerase reaction, and one or more nucleotide, wherein the one or more nucleotide comprises an analogue nucleotide,

wherein the recombinant DNA polymerase comprises a modified active site that is homologous to a wild-type active site of a wild-type DNA polymerase, the modified active site comprising one or more structural modification relative to the wild type active site that reduces steric inhibition for entry of the nucleotide analogue into the modified active site or that is complementary with one or more non-natural features of the nucleotide analogue; and,

reacting the mixture such that the recombinant polymerase replicates at least a portion of the template in a template-dependent manner.

32. A method of making a DNA, the method comprising:

providing a reaction mixture comprising a template, a replication initiating moiety that complexes with or is integral to the template, a polymerase capable of replicating at least a portion of the template using the moiety in a template-dependent polymerase reaction, and one or more nucleotide, wherein the one or more nucleotide comprises an phosphate labeled analogue nucleotide,

wherein a K_m value of the recombinant polymerase for the nucleotide analogue is lower than a K_m for a corresponding homologous wild-type polymerase for the nucleotide analogue and,

reacting the mixture such that the polymerase replicates at least portion of the template in a template-dependent manner, whereby at least one nucleotide analogue residue is incorporated into the resulting DNA.

33. A method of making a DNA, the method comprising:

providing a reaction mixture comprising a template, a replication initiating moiety that complexes with or is integral to the template, a polymerase capable of replicating at least a portion of the template using the moiety in a template-dependent polymerase

reaction, and one or more nucleotide, wherein the one or more nucleotide comprises a labeled phosphate nucleotide analogue,

wherein the polymerase is homologous to a Φ29 DNA polymerase and has a K_m for A488dC4P, A568dC4P, or both, that is less than about 75% of a K_m of GST-N62D Φ29 DNA polymerase mutant for A488dC4P, A568dC4P or both; and,

reacting the mixture such that the polymerase replicates at least a portion of the template.

34. The method of claim **31**, **32**, or **33**, wherein the replication initiating moiety comprises an oligonucleotide primer, a region of self-complementarity in the template, or a polypeptide that binds to the template.

35. The method of claim **31**, **32**, or **33**, wherein a K_m of the polymerase for the analogue is less than about 75% of a K_m of a corresponding wild-type polymerase.

36. The method of claim **31**, **32**, or **33**, wherein a K_m of the polymerase for the analogue is less than about 40% of a K_m of a corresponding wild-type polymerase.

37. The method of claim **31**, **32**, or **33**, wherein a K_m of the polymerase for the analogue is less than about 15% of a K_m of a corresponding wild-type polymerase.

38. The method of claim **31**, **32**, or **33**, wherein the polymerase has a k_{cat}/K_m or V_{max}/K_m for the nucleotide analogue that is higher than a k_{cat}/K_m or V_{max}/K_m of a wild-type Φ29 for the nucleotide analogue.

39. The method of claim **31**, **32**, or **33**, wherein the polymerase is a recombinant DNA polymerase comprising a structural modification relative to a wild-type Φ29 DNA polymerase selected from: a deletion of residues 505-525, a deletion within residues 505-525, a K135A mutation, an E375H mutation, an E375S mutation, an E375K mutation, an E375R mutation, an E375A mutation, an E375Q mutation, an E375W mutation, an E375Y mutation, an E375F mutation, an E486A mutation, an E486D mutation, a K512A mutation, and combinations thereof.

40. A method of making a recombinant DNA polymerase, the method comprising:

structurally modeling a first polymerase;

identifying one or more steric inhibition feature affecting nucleotide access to the active site or complementarity feature of a nucleotide analogue at the active site;

mutating the first DNA polymerase to reduce or remove at least one steric inhibition feature or to add at least one nucleotide analogue complementarity feature; and,

determining whether the resulting recombinant polymerase displays a modified activity for a nucleotide analogue as compared to the first DNA polymerase.

41. The method of claim **40**, comprising determining k_{cat}, K_m, V_{max}, or k_{cat}/K_m of the recombinant DNA polymerase for the nucleotide analogue.

42. The method of claim **40**, comprising determining k_{cat}, K_m, V_{max}, or k_{cat}/K_m of the recombinant DNA polymerase for a natural nucleotide.

43. The method of claim **40**, comprising making a library of recombinant DNA polymerases, a plurality of members of the library comprising one or more steric inhibition feature mutation or complementarity feature mutation.

44. The method of claim **43**, comprising screening the library to identify at least one member comprising the modified activity.

45. A computer-implemented method, comprising:

defining a plurality of polymerase state transitions for discrete time steps during a template-based polymerization reaction;

defining a plurality of rate transition rates between the states;

generating a multidimensional probability matrix of possible states, based upon a given nucleic acid template sequence, nucleotides in a reaction mixture and the polymerase state transitions; and,

storing the multidimensional probability matrix in a computer readable medium.

46. The method of claim **45**, wherein the polymerase state transitions are user-selectable.

47. The method of claim **45**, wherein the rate transition rates between the states vary depending on nucleotide concentration, template sequence and position of the polymerase along the template.

48. The method of claim **45**, wherein the nucleotides in the reaction mixture comprise one or more nucleotide analogues.

49. The method of claim **45**, wherein the rate transition rates between states include a conformational transition rate for the polymerase during use of the nucleotide analogues

by the polymerase, the rate set to be equal to a conformational transition rate for a natural nucleotide.

50. The method of claim **45**, wherein the multidimensional probability matrix is automatically generated based upon the template sequence, a standardized matrix of probability states, and the nucleotides in the reaction mixture.

51. The method of claim **45**, wherein the probability matrix is simplified by assuming that all possible Watson-Crick base pairings are equivalent in all state transitions.

52. The method of claim **45**, wherein a second reagent concentration matrix is generated to account for reagent concentration changes that result from position of the polymerase along a template, based on an output of the probability matrix.

53. The method of claim **45**, comprising vectorizing the probability matrix for multiple templates and multiplying the resulting vectorized probability matrix by the multidimensional probability matrix to provide a state distribution matrix.

54. The method of claim **45**, comprising defining an exponential time factor for the probability matrix to account for repeated sequences within the template sequence.

55. The method of claim **45**, comprising defining a polymerase nucleotide mismatch fraction using either a continuum model or a counting model.

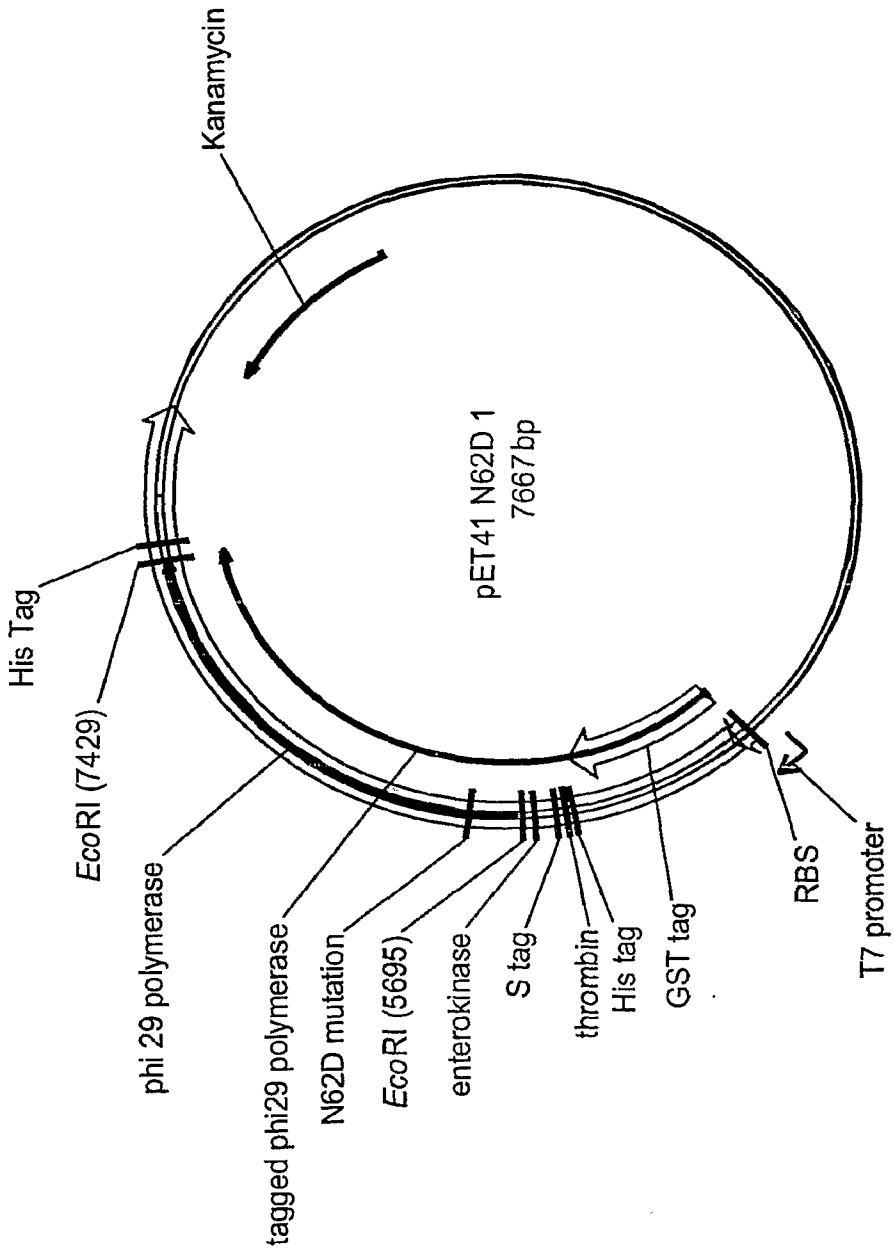

Kanamycin
pET41 N62D 1
7667bp
His Tag
EcoRI (7429)
phi 29 polymerase
tagged phi29 polymerase
N62D mutation
EcoRI (5695)
enterokinase
S tag
thrombin
His tag
GST tag
RBS
T7 promoter

Fig. 1

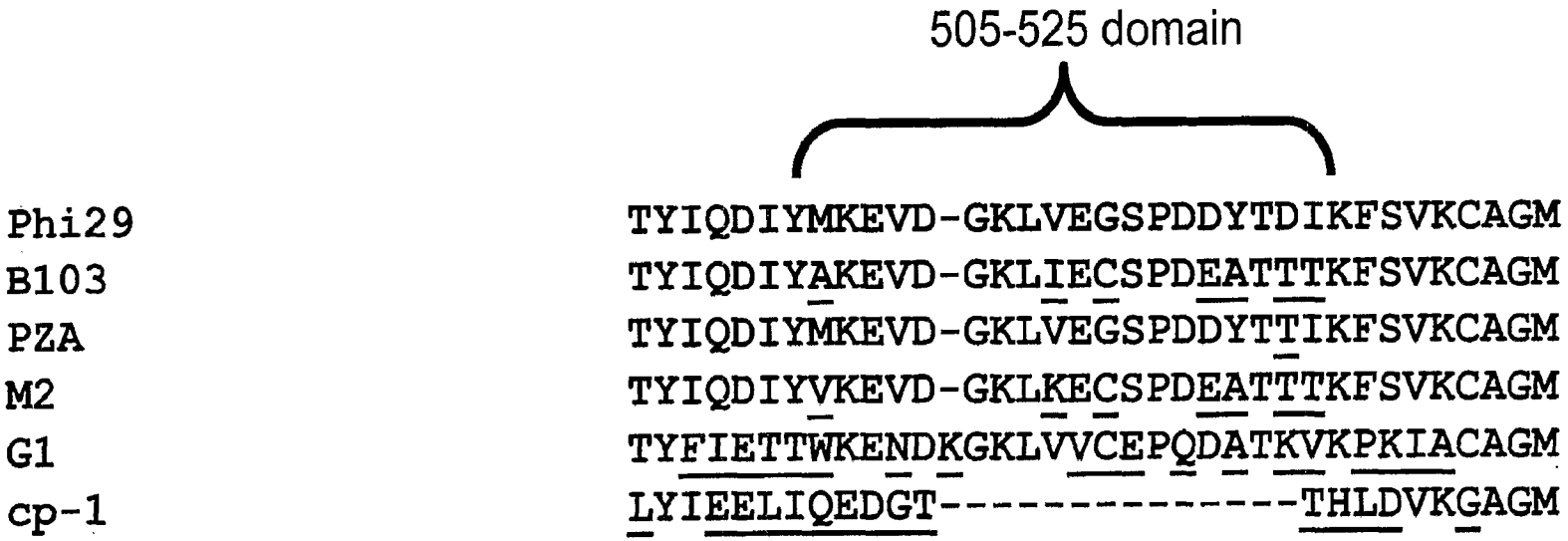

505-525 domain
Phi29 TYIQDIYMKEVD-GKLVEGSPDDYTDIKFSVKCAGM
B103 TYIQDIYAKEVD-GKLIECSPDEATTTKFSVKCAGM
PZA TYIQDIYMKEVD-GKLVEGSPDDYTTIKFSVKCAGM
M2 TYIQDIYVKEVD-GKLKECSPDEATTTKFSVKCAGM
G1 TYFIETTWKENDKGKLVVCEPQDATKVKPKIACAGM
cp-1 LYIEELIQEDGT----------THLDVKGAGM

Fig. 2A

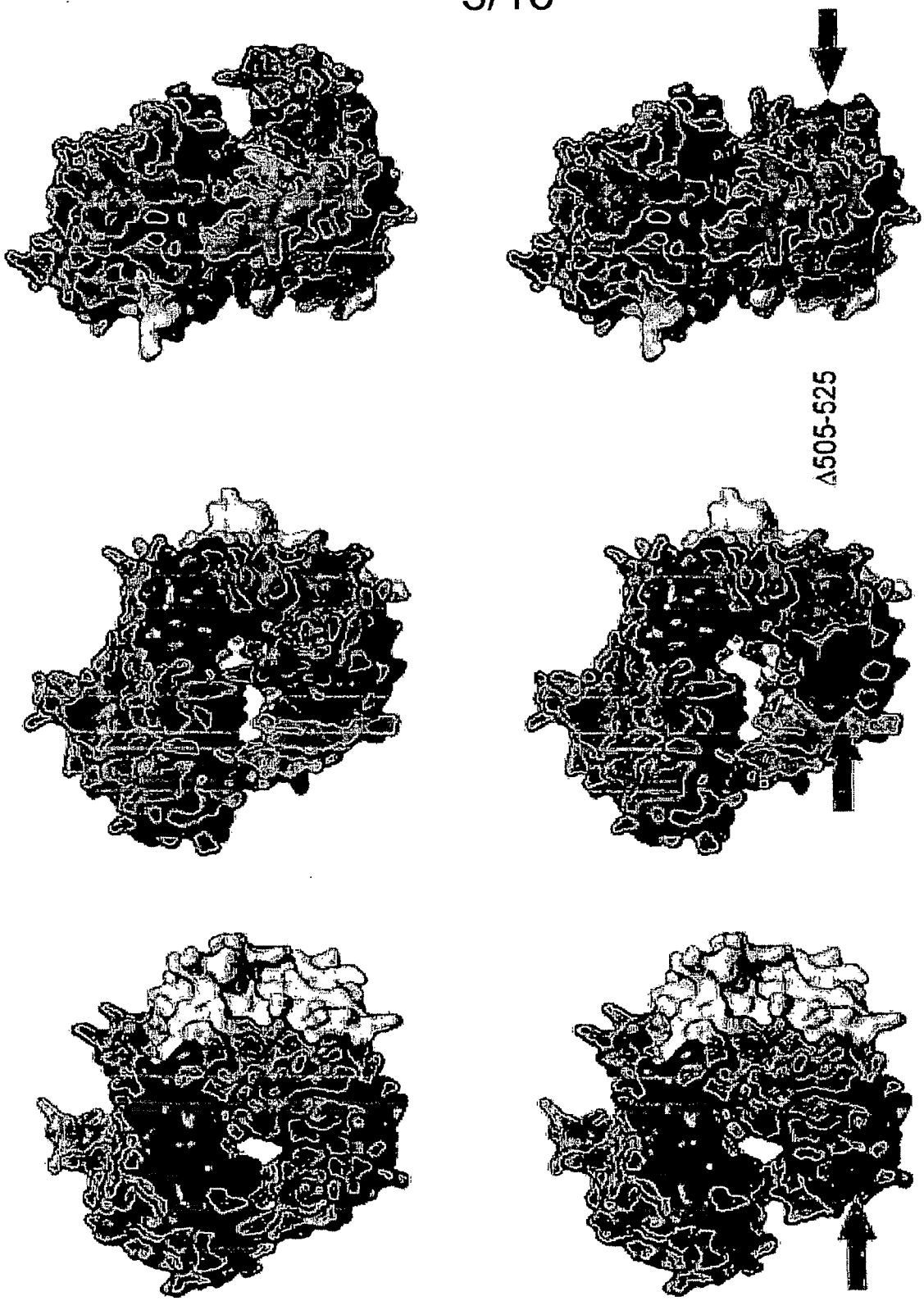


Fig. 2B

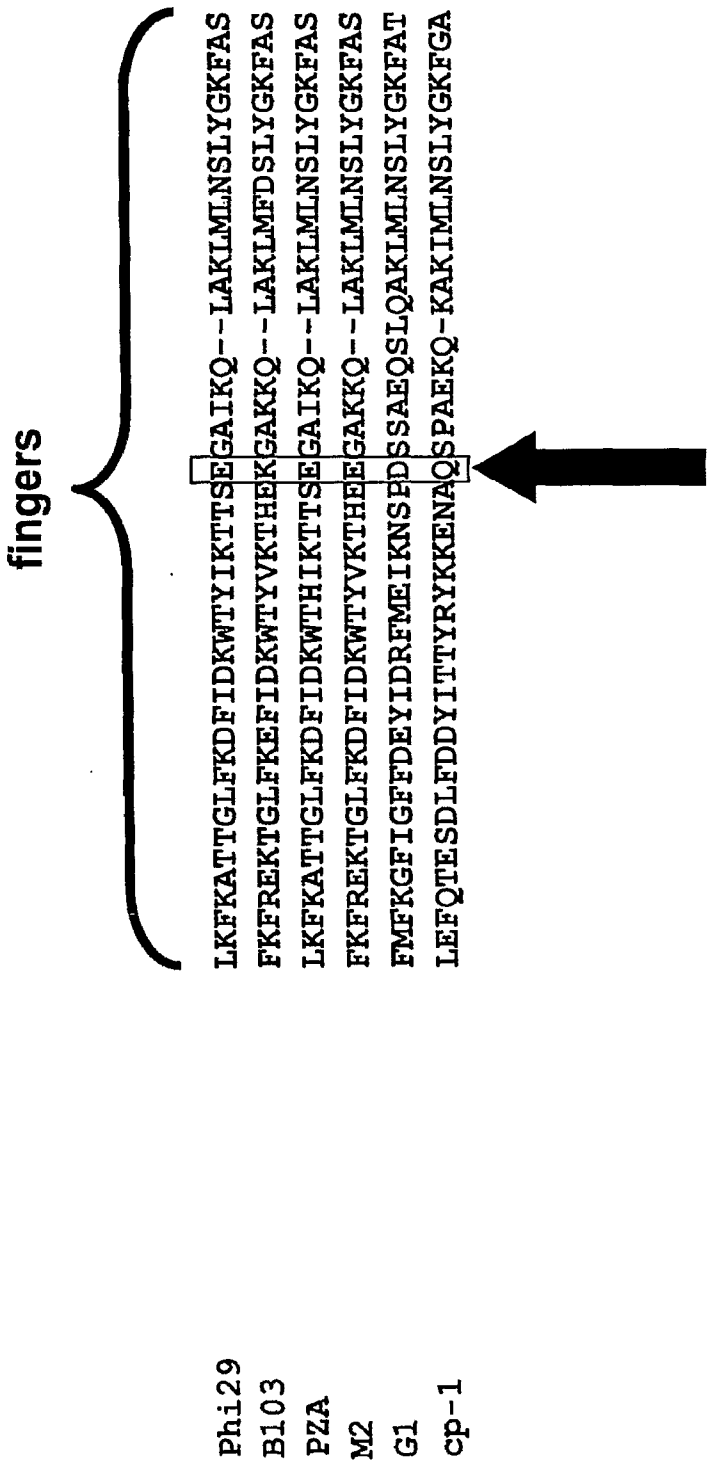

fingers
Phi29 LKFKATTGLFKDFIDKWTYIKTTSEGAIKQ--LAKLMLNSLYGKFAS
B103 FKFREKTGLFKEFIDKWTYVKTHEKGAKKQ--LAKLMFDSLYGKFAS
PZA LKFKATTGLFKDFIDKWTHIKTTSEGAIKQ--LAKLMLNSLYGKFAS
M2 FKFREKTGLFKDFIDKWTYVKTHEEGAKKQ--LAKLMLNSLYGKFAS
G1 FMFKGFIGFFDEYIDRFMEIKNSPDSSAEQSLQAKLMLNSLYGKFAT
cp-1 LEFQTESDLFDDYITTYRYKKENAQSPAEKQ-KAKIMLNSLYGKFGA
E375

Fig. 3A

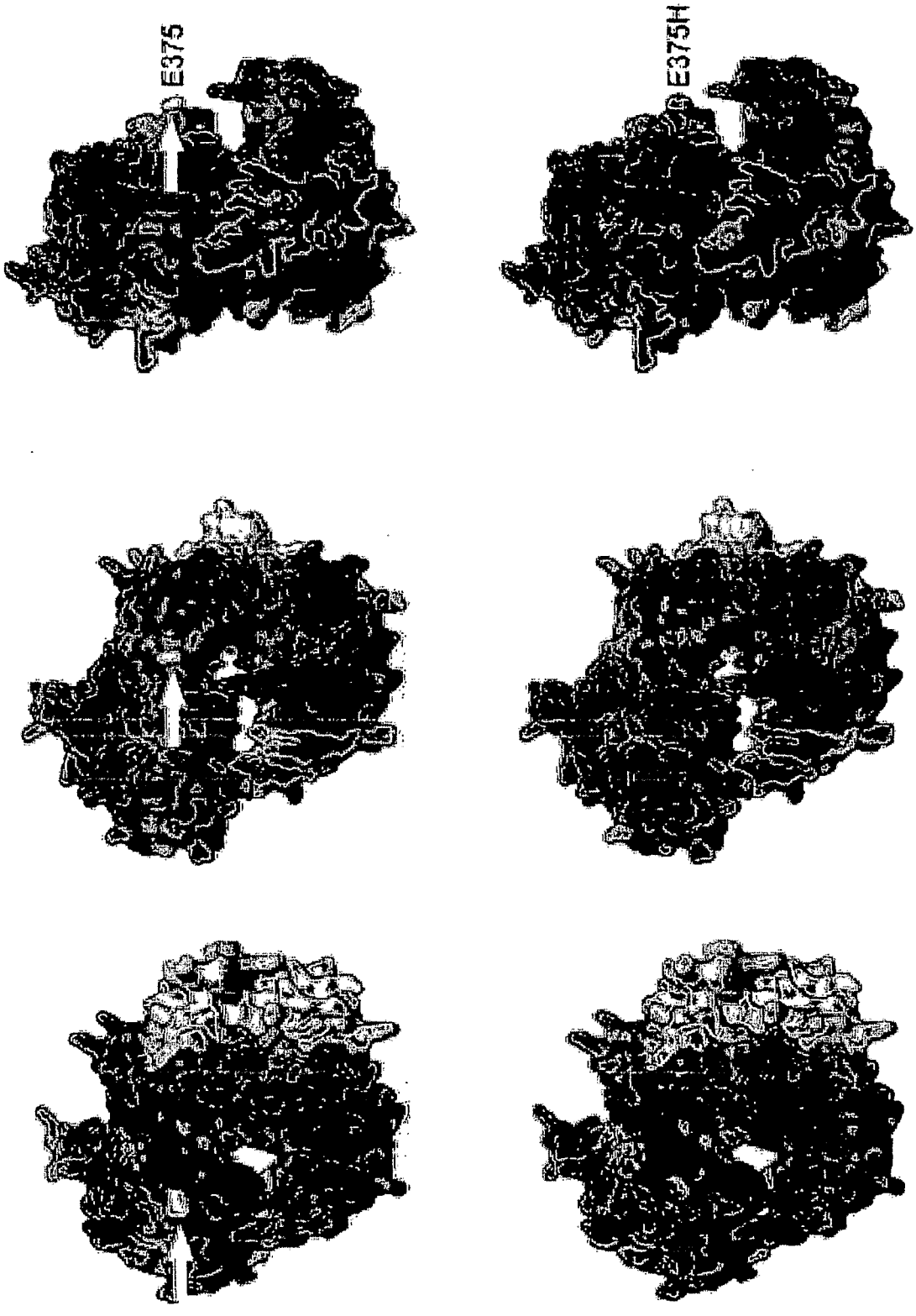


Fig. 3B

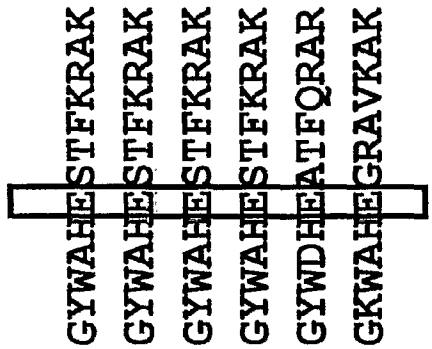

Phi29 GYWAHESTFKRAK
B103 GYWAHESTFKRAK
PZA GYWAHESTFKRAK
M2 GYWAHESTFKRAK
G1 GYWDHEATFQRAR
cp-1 GKWAHEGRAVKAK

Fig. 4A

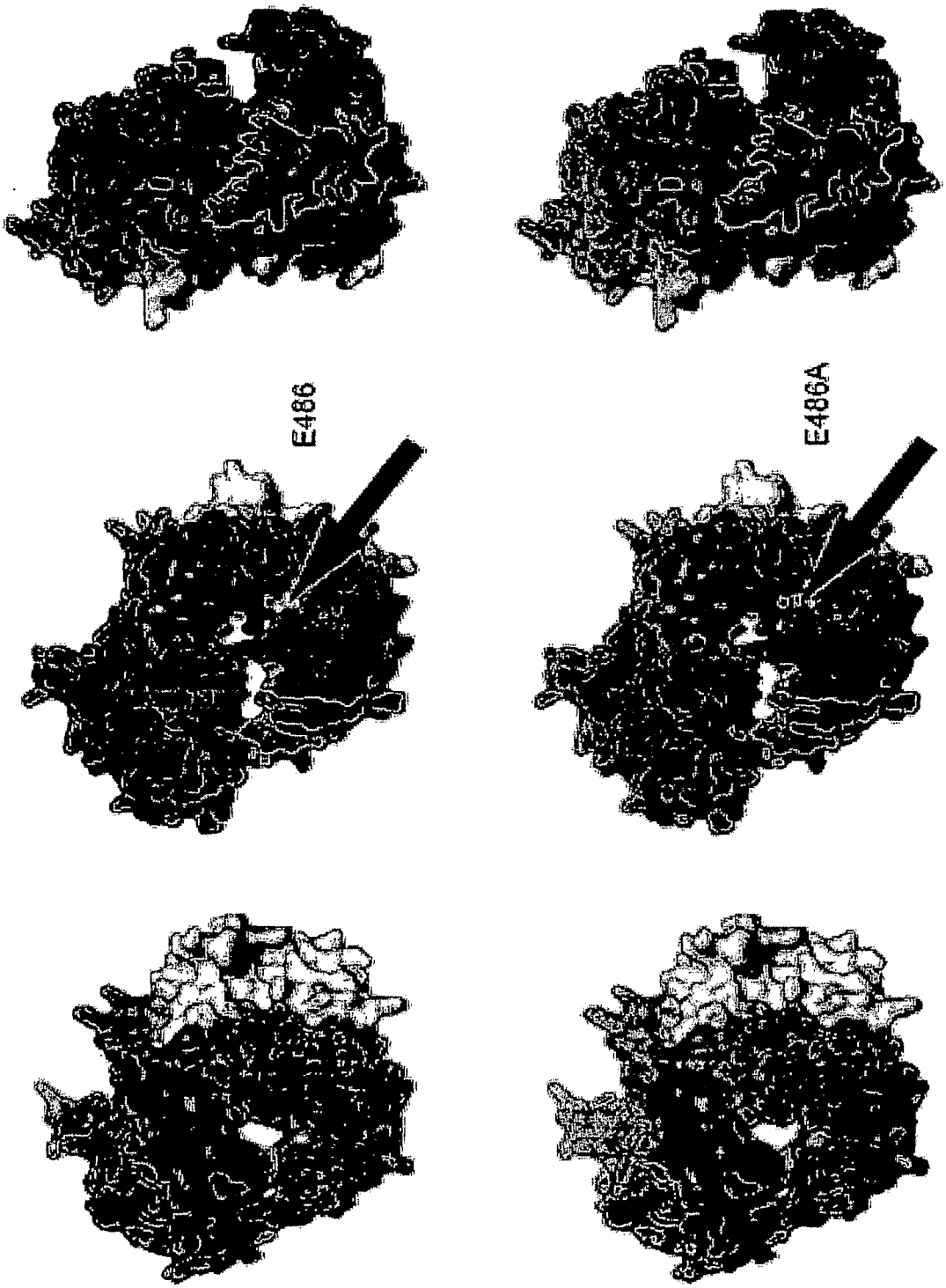


Fig. 4B

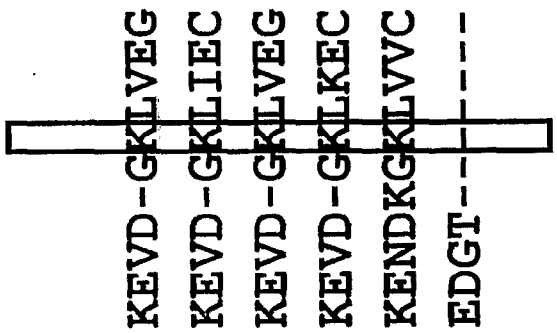

KEVD-GKLVEG
KEVD-GKLIEC
KEVD-GKLVEG
KEVD-GKLKEC
KENDKGKLVVC
EDGT-------
Phi29
B103
PZA
M2
G1
cp-1

Fig. 5A

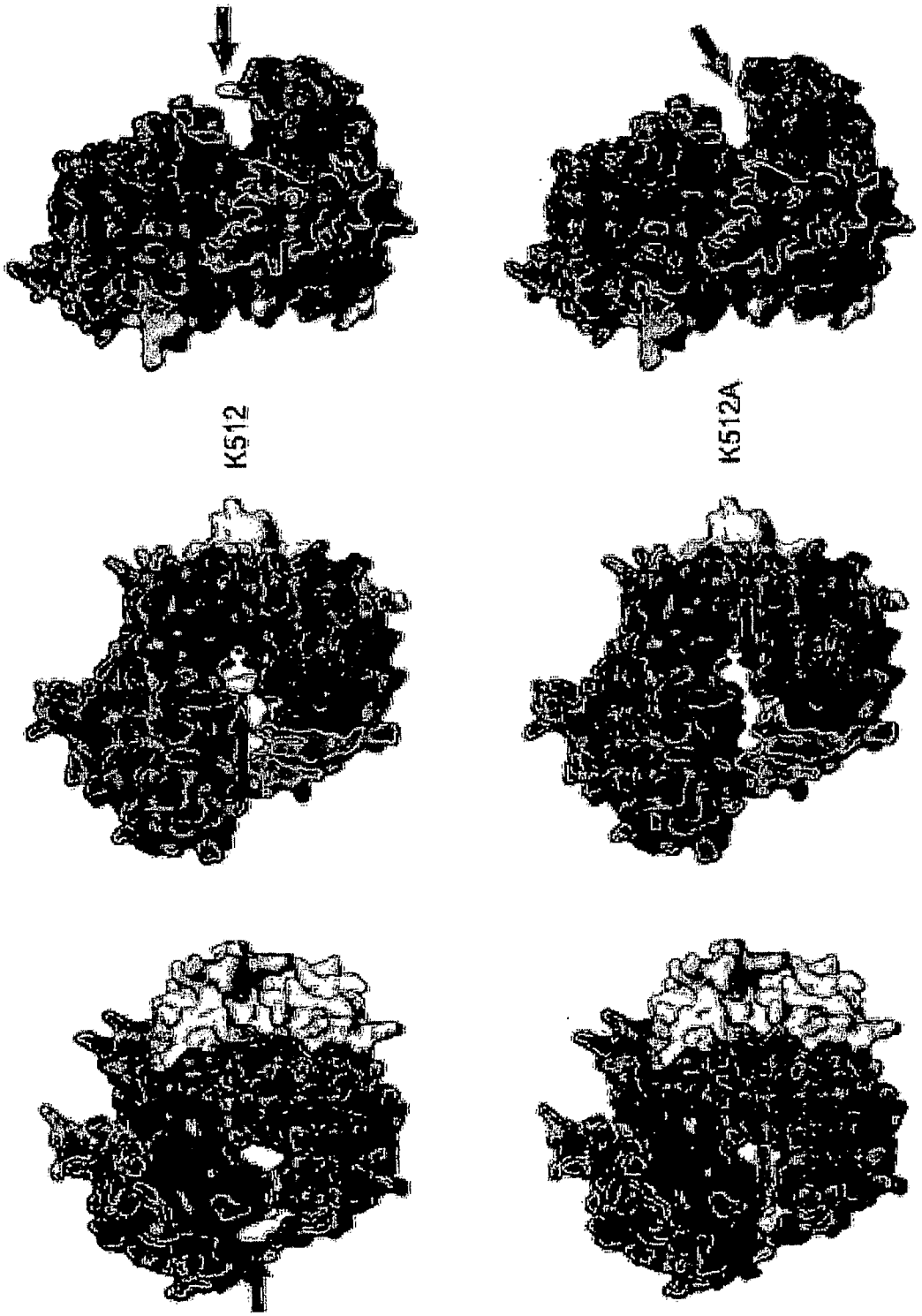


Fig. 5B

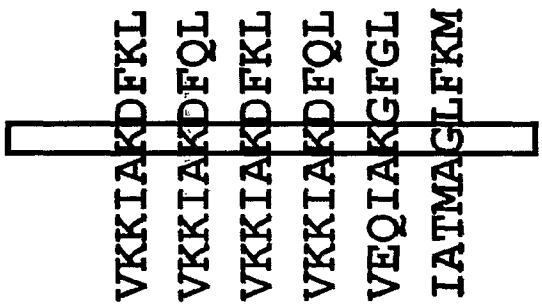


Fig. 6A

Phi29
B103
PZA
M2
G1
cp-1

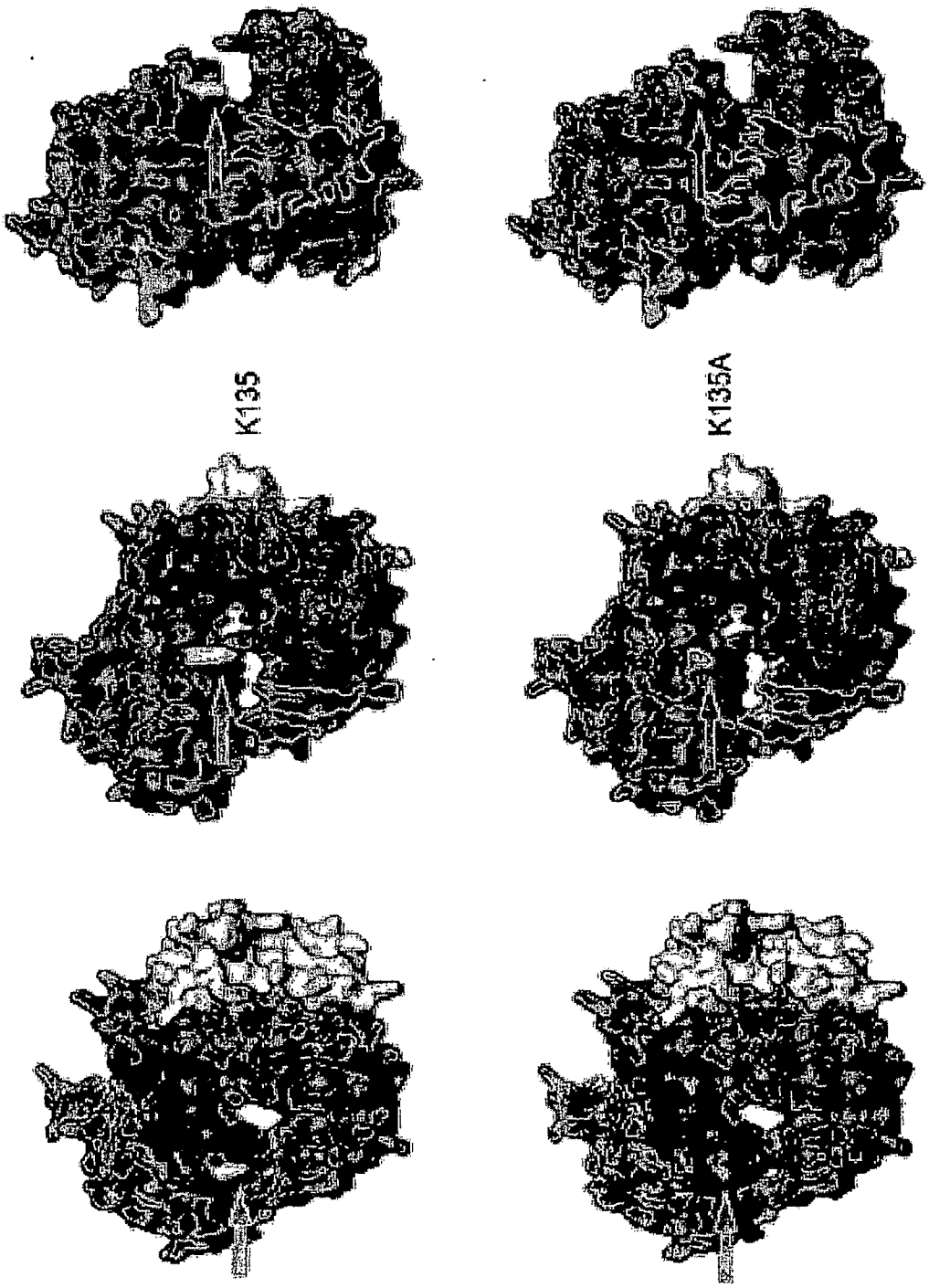


Fig. 6B

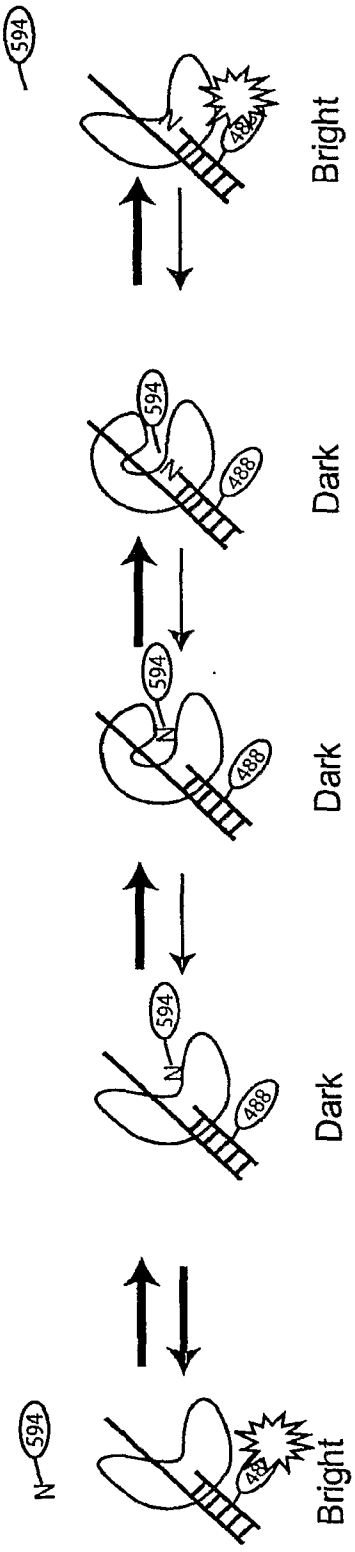

594
488
Bright
594
N
488
Dark
594
N
488
Dark
594
N
488
Dark
N
594
488
Bright

Fig. 7A

Fig 7B

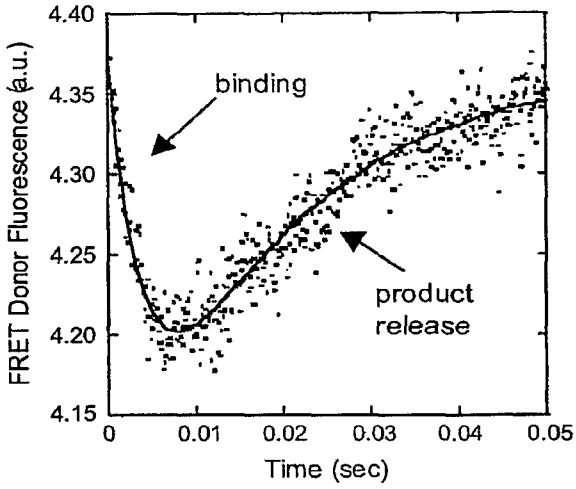


Fig 7C

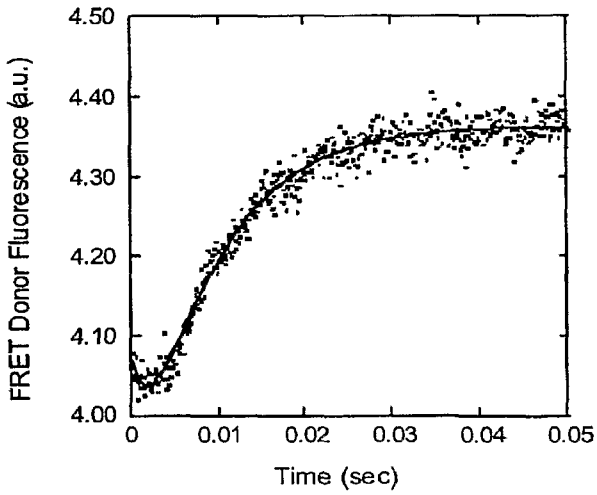
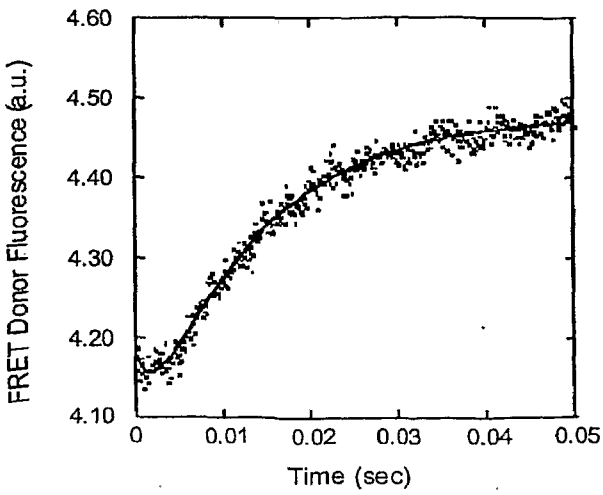


Fig 7D



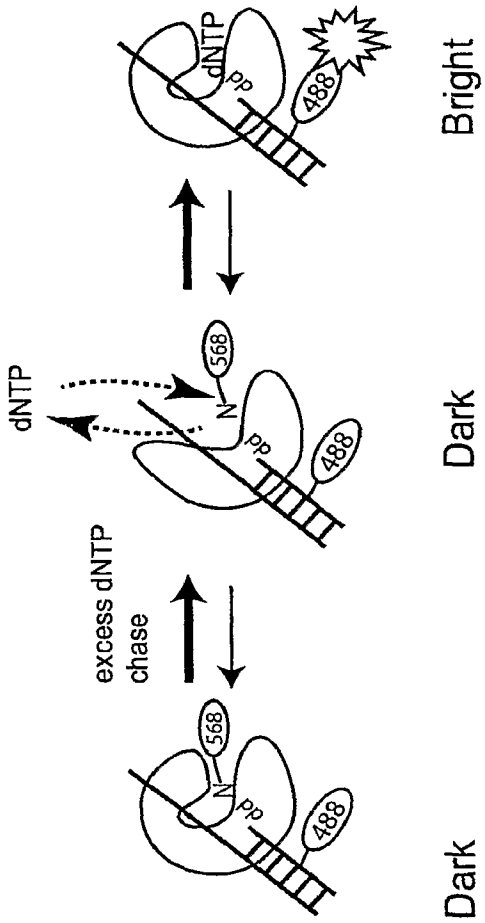

dNTP
pp
488
Bright
568
N
dNTP
pp
488
Dark
excess dNTP
chase
568
N
pp
488
Dark

Fig. 8A

Fig 8B

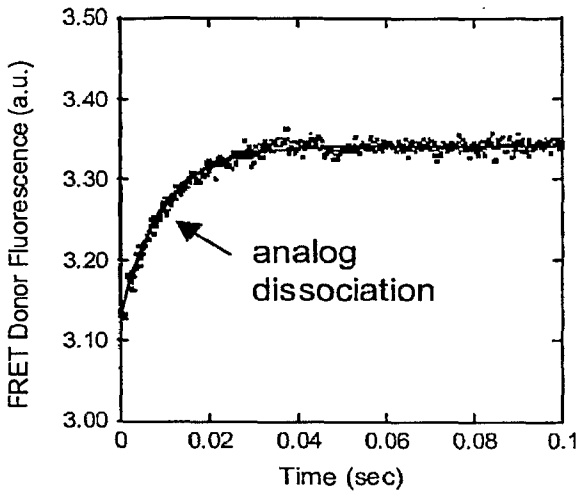


Fig 8C

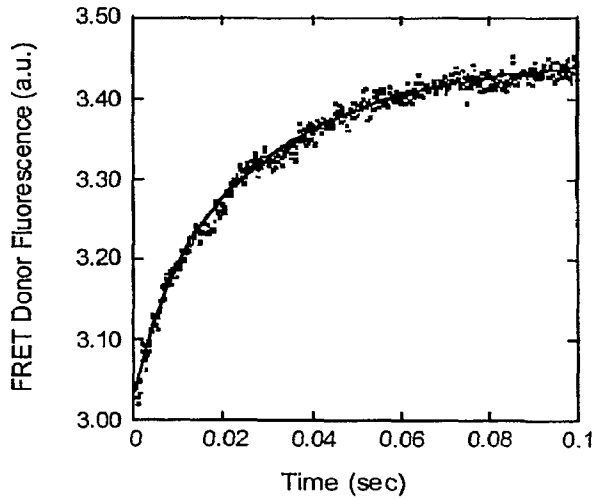
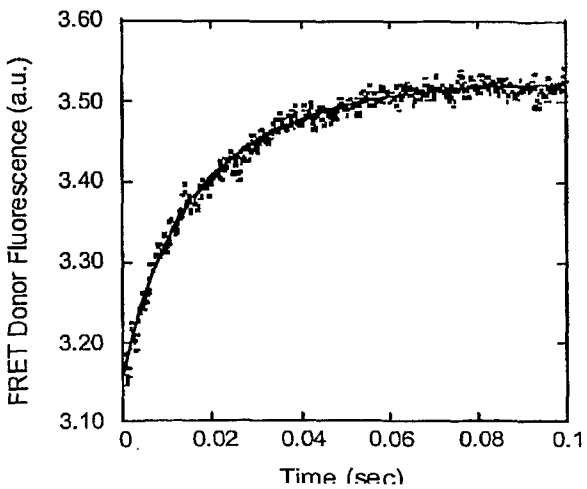


Fig 8D



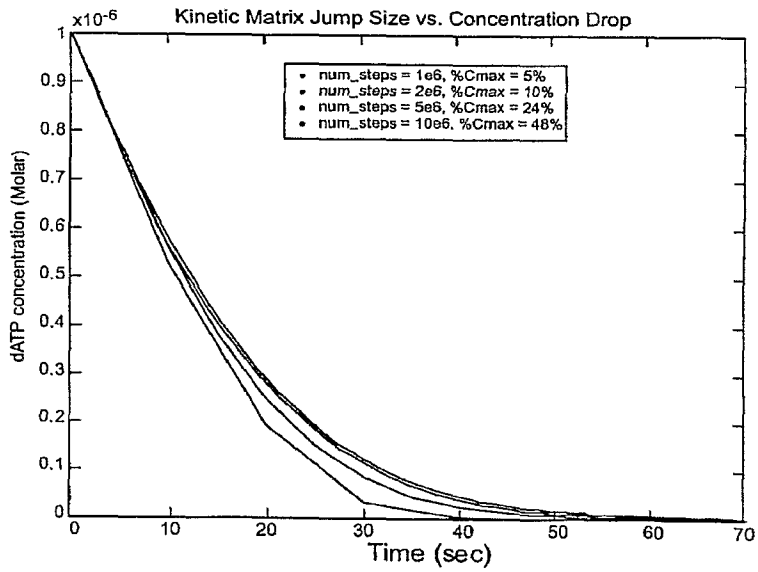

x10-6
Kinetic Matrix Jump Size vs. Concentration Drop
num_steps = 1e6, %Cmax = 5%
num_steps = 2e6, %Cmax = 10%
num_steps = 5e6, %Cmax = 24%
num_steps = 10e6, %Cmax = 48%
dATP concentration (Molar)
Time (sec)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70

Fig 9