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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 February 2008 (07.02.2008)

PCT

(10) International Publication Number
WO 2008/016906 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2007/074828

(22) International Filing Date: 31 July 2007 (31.07.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/496,274 31 July 2006 (31.07.2006) US

(71) Applicant *(for all designated States except US)*: **HELICOS BIOSCIENCES CORPORATION** [US/US]; One Kendall Square, Building 700, Cambridge, Massachusetts 02139 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SIDDIQI, Suhaib, M.** [US/US]; 37 University Avenue, Burlington, Massachusetts 01803 (US). **KRZYMANSKA-OLEJNIK, Edyta** [PL/US]; 48 Beaconsfield Road, Brookline, Massachusetts 02445 (US). **ORGUEIRA, Herman, Antonio** [AR/US]; 60 Ellery Street, Apt. 9, Cambridge, Massachusetts 02138 (US). **BAI, Xiaopeng** [CN/US]; 190 Waterman Street, Apt. 1, Providence, Rhode Island 02906 (US).

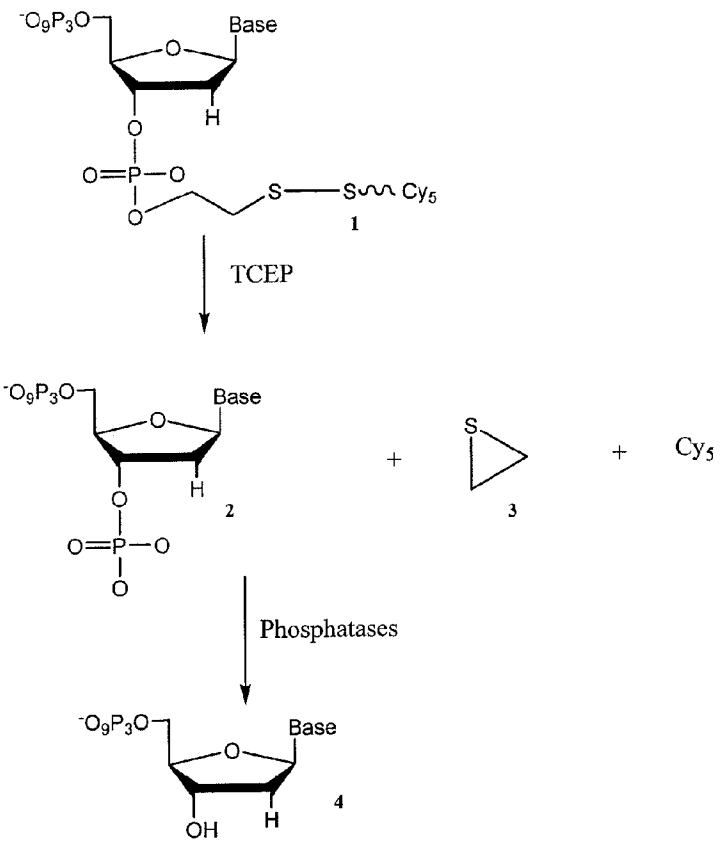
(74) Agent: **LINNIK, Konstantin, M.**; Cooley Godward Kronish Llp, ATTN: Patent Group, 11951 Freedom Drive, Reston, Virginia 20190 (US).

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

[Continued on next page]

(54) Title: NUCLEOTIDE ANALOGS



(57) Abstract: The invention provides nucleotide analogs for use in sequencing nucleic acid molecules.

WO 2008/016906 A2

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, MT, 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).

Declaration under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *without international search report and to be republished upon receipt of that report*

Nucleotide Analogs

Field of the Invention

[0001] The invention relates to nucleotide analogs and methods for sequencing a nucleic acid using the nucleotide analogs.

Background

[0002] New sequencing technologies, based on single-molecule measurements, have been proposed. These proposals include sequencing strategies based on the observation of an interaction of particular proteins with DNA, or by using ultra high resolution scanned probe microscopy. See, *e.g.*, Rigler, *et al.*, J. Biotechnol., 86(3):161 (2001); Goodwin, P.M., *et al.*, Nucleosides & Nucleotides, 16(5-6):543-550 (1997); Howorka, S., *et al.*, Nature Biotechnol., 19(7):636-639 (2001); Meller, A., *et al.*, Proc. Nat'l. Acad. Sci., 97(3):1079-1084 (2000); Driscoll, R.J., *et al.*, Nature, 346(6281):294-296 (1990).

[0003] Sequencing-by-synthesis methodology that results in sequence determination, but without consecutive base incorporation, has also been proposed. See, Braslavsky, *et al.*, Proc. Nat'l Acad. Sci., 100: 3960-3964 (2003). Bulky fluorophores that impede sequential base incorporation can be an impediment to base-over-base sequencing. Even when the label is removed, some fluorescently-labeled nucleotides hinder subsequent base incorporation, possibly due to the residue of the linker that is left behind after label removal.

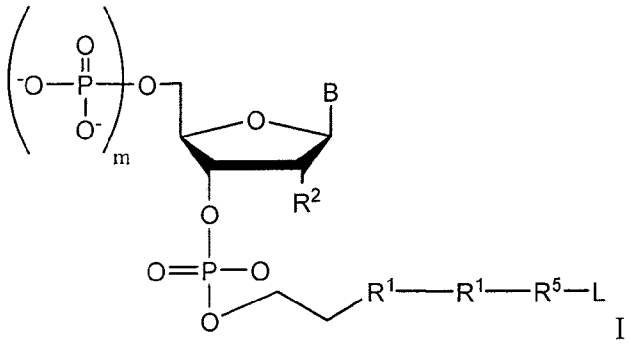
[0004] A need therefore exists for nucleotide analogs that promote accurate base-over-base incorporation in sequencing-by-synthesis reactions, resulting in greater read-lengths.

Summary of the Invention

[0005] The present invention provides nucleotide analogs and methods of using nucleotide analogs in sequencing. A nucleotide analog of the invention comprises a removable detectable moiety that is attached to a nucleotide analog, and that upon removal of the detectable moiety, leaves no or substantially no residue or "scar" on the incorporated base or nucleotide and therefore does not substantially hinder subsequent nucleotide (or nucleotide analog) incorporation, thereby permitting multiple base over base template-directed incorporation and longer runs of sequence determination. Before removal of a detectable moiety, analogs of the

invention may allow only limited base addition in any given cycle of template-dependent nucleotide incorporation.

[0006] Nucleotide analogs of the present invention include those depicted by Formula I:



I

wherein,

[0007] B is selected from the group consisting of a purine, a pyrimidine, and analogs thereof,

[0008] R^1 at each occurrence, independently is selected from the group consisting of S, NR^3 and O,

[0009] R^2 is selected from the group consisting of H and OH,

[0010] R^3 is selected from the group consisting of H and alkyl,

[0011] R^5 is an aliphatic moiety,

[0012] L is a label, and

[0013] m, at each occurrence, independently is an integer from 1 to 3.

[0014] B may selected from the group consisting of cytosine, uracil, thymine, adenine, guanine, and analogs thereof, such as for example, inosine.

[0015] In certain embodiments, R^1 for each occurrence is S.

[0016] L may be an optically detectable label, such as a fluorescent label. An optically detectable label may be selected from the group consisting of cyanine, rhodamine, fluoroscein, coumarin, BODIPY, alexa and conjugated multi-dyes. In some embodiments, the optically detectable label is Cy3 or Cy5.

[0017] In general, methods of sequencing a nucleic acid template provided herein comprise exposing a nucleic acid template hybridized to a primer having a free 3' hydroxyl group (end) to a polymerase and to nucleotide analogs disclosed herein under conditions to permit the analogs to be added to the primer (or extended primer). Incorporated nucleotide analogs are detected and the labels subsequently removed. The template sequence is determined by repeating these steps one or more times. In some embodiments, the nucleotide analog resulting from removal of the label is substantially identical to a native nucleotide. As used herein, the term "primer" includes sequences hybridized to the templates that have been previously extended, e.g., using the methods disclosed herein.

[0018] In preferred embodiments, the primer, template, or both is/are immobilized to a solid support. In a highly preferred embodiment, the primer is immobilized. In other embodiments, a duplex is immobilized so as to be individually optically resolvable.

[0019] The label and any linker attaching the label to the nucleotide analog may be chemically removed from the nucleotide analogs. In a preferred embodiment, a label is attached via a disulfide linkage and removed by exposure to a reducing agent such as dithiothreitol, tris(2-carboxyethyl)phosphine and tris(2-chloropropyl)phosphate. This serves to remove all moieties from the 3' position of the analog, leaving in its place an OH group ready for further extension by the polymerase in subsequent cycles.

[0020] While the invention is exemplified herein with fluorescent labels, the invention is not so limited and can be practiced using nucleotides labeled with any detectable label, preferably an optically detectable label, such as chemiluminescent labels, luminescent labels, phosphorescent labels, fluorescence polarization labels, as well as charge labels.

[0021] A detailed description of the certain embodiments of the invention is provided below. Other embodiments of the invention are apparent upon review of the detailed description that follows.

Brief Description of the Drawing

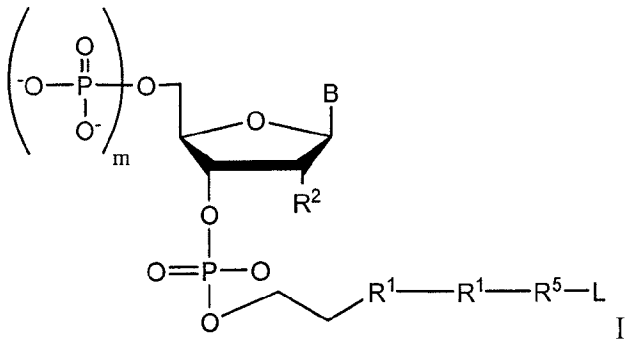
[0022] Figure 1 depicts an nucleotide analog disclosed herein having a label attached to the 3' position of the nucleotide, and a synthetic route for removal of the label yielding a nucleotide with a 3' OH group.

Detailed Description of the Invention

[0023] The invention relates generally to nucleotide analogs that, when used in sequencing reactions, allow extended base-over-base incorporation into a primer in a template-dependent sequencing reaction. Nucleotide analogs of the invention include nucleoside 5' triphosphates having a linker between a pentose of the nucleotide and a detectable label, wherein the linker is cleavable to produce an un-labeled residue that is substantially identical to the native (*i.e.*, unlabeled) nucleotide. Such an analog permits polymerase to recognize the analog as a nucleotide and add bases, and does not affect subsequent base pairing. Analogs of the invention are thus useful in sequencing-by-synthesis reactions in which consecutive bases are added to a primer in a template-dependent manner.

Nucleotide Analogs

[0024] Nucleotide analogs of the invention have the generalized structure:



I

[0025] The base B can be, for example, a purine or a pyrimidine. For example, B can be an adenine, cytosine, guanine, thymine, uracil, or hypoxanthine. The base B also can be, for example, naturally-occurring and synthetic derivatives of a base, including pyrazolo[3,4-d]pyrimidines, 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo (*e.g.*, 8-bromo), 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-

azaadenine, deazaguanine, 7-deazaguanine, 3-deazaguanine, deazaadenine, 7-deazaadenine, 3-deazaadenine, pyrazolo[3,4-d]pyrimidine, imidazo[1,5-a]1,3,5 triazinones, 9-deazapurines, imidazo[4,5-d]pyrazines, thiazolo[4,5-d]pyrimidines, pyrazin-2-ones, 1,2,4-triazine, pyridazine; and 1,3,5 triazine. Bases useful according to the invention may permit a nucleotide, that includes the base, to be incorporated into a polynucleotide chain by a polymerase and may form base pairs with a base on an antiparallel nucleic acid strand. The term base pair encompasses not only the standard AT, AU or GC base pairs, but also base pairs formed between nucleotides and/or nucleotide analogs comprising non-standard or modified bases, wherein the arrangement of hydrogen bond donors and hydrogen bond acceptors permits hydrogen bonding between a non-standard base and a standard base or between two complementary non-standard base structures. One example of such non-standard base pairing is the base pairing between the nucleotide analog inosine and adenine, cytosine or uracil, where the two hydrogen bonds are formed.

[0026] Label L may be any moiety that can be attached to or associated with an oligonucleotide and that functions to provide a detectable signal, and/or to interact with a second label to modify the detectable signal provided by the first or second label, e.g. fluorescence resonance energy transfer (FRET). The label preferably is an optically-detectable label. In one embodiment, the label is an optically-detectable label such as a fluorescent, chemiluminescence, or electrochemically luminescent label. Examples of fluorescent labels include, but are not limited to, 4-acetamido-4'-isothiocyanatostilbene-2,2'disulfonic acid; acridine and derivatives: acridine, acridine isothiocyanate; 5-(*2'*-aminoethyl)aminonaphthalene-1 -sulfonic acid (EDANS); 4-amino-N-[3-vinylsulfonyl)phenyl]naphthalimide-3,5 disulfonate; N-(4-anilino-l-naphthyl)maleimide; anthranilamide; BODIPY; Brilliant Yellow; coumarin and derivatives; coumarin, 7-amino-4-methylcoumarin (AMC, *C*oumarin 120), 7-amino-4-trifluoromethylcouluarin (Coumaran 151); cyanine dyes; cyanosine; 4',6-diaminidino-2-phenylindole (DAPI); 5'5"-dibromopyrogallol-sulfonaphthalein (Bromopyrogallol Red); 7-diethylamino-3-(4'-isothiocyanatophenyl)-4-methylcoumarin; diethylenetriamine pentaacetate; 4,4'-diisothiocyanatodihydro-stilbene-2,2'-disulfonic acid; 4,4'-diisothiocyanatostilbene-2,2'-disulfonic acid; 5-[dimethylamino]naphthalene- 1-sulfonyl chloride (DNS, dansylchloride); 4-dimethylaminophenylazophenyl-4'-isothiocyanate (DABITC); eosin and derivatives; eosin, eosin isothiocyanate, erythrosin and derivatives; erythrosin B, erythrosin, isothiocyanate; ethidium;

fluorescein and derivatives; 5-carboxyfluorescein (FAM), 5-(4,6-dichlorotriazin-2-yl)aminofluorescein (DTAF), 2',7'-dimethoxy-4'5'-dichloro-6-carboxyfluorescein, fluorescein, fluorescein isothiocyanate, QFITC, (XRITC); fluorescamine; IR144; IR1446; Malachite Green isothiocyanate; 4-methylumbelliferoneortho cresolphthalein; nitrotyrosine; pararosaniline; Phenol Red; B-phycoerythrin; o-phthaldialdehyde; pyrene and derivatives: pyrene, pyrene butyrate, succinimidyl 1-pyrene; butyrate quantum dots; Reactive Red 4 (CibacronTM Brilliant Red 3B-A) rhodamine and derivatives: 6-carboxy-X-rhodamine (ROX), 6-carboxyrhodamine (R6G), lissamine rhodamine B sulfonyl chloride rhodarnine (Rhod), rhodamine B, rhodamine 123, rhodamine X isothiocyanate, sulforhodamine B, sulforhodamine 101, sulfonyl chloride derivative of sulforhodamine 101 (Texas Red); N,N,N',N'tetramethyl-6-carboxyrhodamine (TAMRA); tetramethyl rhodamine; tetramethyl rhodamine isothiocyanate (TRITC); riboflavin; rosolic acid; terbium chelate derivatives; Cy3; Cy5; Cy5.5; Cy7; IRD 700; IRD 800; La Jolta Blue; phthalo cyanine; and naphthalo cyanine. Preferred fluorescent labels are cyanine-3 and cyanine-5. Labels other than fluorescent labels are contemplated by the invention, including other optically-detectable labels. Any appropriate detectable label can be used according to the invention, and numerous other labels are known to those skilled in the art.

[0027] R^1 at each occurrence may be independently selected from the group consisting of S, NR^3 and O, where R^3 may be selected from the group consisting of H and alkyl.

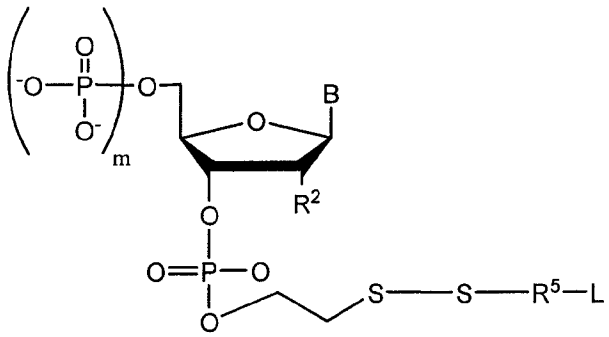
[0028] Alkyl moieties include saturated aliphatic groups, including straight-chain alkyl groups, branched-chain alkyl groups, cycloalkyl (alicyclic) groups, alkyl substituted cycloalkyl groups, and cycloalkyl substituted alkyl groups. In certain embodiments, a straight chain or branched chain alkyl has about 30 or fewer carbon atoms in its backbone (e.g., C_1-C_{30} for straight chain, C_3-C_{30} for branched chain), and alternatively, about 20 or fewer. Likewise, cycloalkyls have from about 3 to about 10 carbon atoms in their ring structure, and alternatively about 5, 6 or 7 carbons in the ring structure. The term "alkyl" also includes halosubstituted alkyls. Moreover, the term "alkyl" (or "lower alkyl") includes "substituted alkyls", which refers to alkyl moieties having substituents replacing a hydrogen on one or more carbons of the hydrocarbon backbone.

[0029] In order to prevent or reduce degradation of the primer containing the nucleotide analog or degradation of the nucleotide analogs, the nucleotide analog can further comprise a non-bridging sulfur on the α phosphate group of the nucleotide.

[0030] R^2 may be selected from H and OH. R^5 may be an aliphatic linker, such as a divalent linear, branched, cyclic alkane, alkene, or alkyne. In certain embodiments, aliphatic groups may be linear or branched and have from 1 to about 20 carbon atoms.

[0031] The integer m, at each occurrence, independently may be an integer from 1 to 3. In some embodiments, m is 1.

[0032] In certain embodiments, a nucleotide analog of the invention can be represented by:



where B, L, R^2, and R^5 are defined above.

Nucleic Acid Sequencing

[0033] The invention also includes methods for nucleic acid sequence determination using the nucleotide analogs described herein. The nucleotide analogs of the present invention are particularly suitable for use in single molecule sequencing techniques. Such techniques are described for example in U.S. Patent Application 10/831,214 filed April 2004; 10/852,028 filed May 24, 2004; 10/866,388 filed June 10, 2005; 10/099,459 filed March 12, 2002; and U.S. Published Application 2003/013880 published July 24, 2003, the teachings of which are incorporated herein in their entireties. In general, methods for nucleic acid sequence determination comprise exposing a target nucleic acid (also referred to herein as template nucleic acid or template) to a primer that is complementary to at least a portion of the target nucleic acid,

under conditions suitable for hybridizing the primer to the target nucleic acid, forming a template/primer duplex.

[0034] Target nucleic acids include deoxyribonucleic acid (DNA) and/or ribonucleic acid (RNA). Target nucleic acid molecules can be obtained from any cellular material obtained from an animal, plant, bacterium, virus, fungus, or any other cellular organism, or may be synthetic DNA. Target nucleic acids may be obtained directly from an organism or from a biological sample obtained from an organism, *e.g.*, from blood, urine, cerebrospinal fluid, seminal fluid, saliva, sputum, stool and tissue. Any tissue or body fluid specimen may be used as a source for nucleic acid for use in the invention. Nucleic acid molecules may also be isolated from cultured cells, such as a primary cell culture or a cell line. The cells from which target nucleic acids are obtained can be infected with a virus or other intracellular pathogen. Nucleic acid molecules may also include those of animal (including human), wild type or engineered prokaryotic or eukaryotic cells, viruses or completely or partially synthetic RNAs or DNAs. A sample can also be total RNA extracted from a biological specimen, a cDNA library, or genomic DNA.

[0035] Nucleic acid typically is fragmented to produce suitable fragments for analysis. In one embodiment, nucleic acid from a biological sample is fragmented by sonication. Test samples can be obtained as described in U.S. Patent Application 2002/0190663 A1, published October 9, 2003, the teachings of which are incorporated herein in their entirety. Generally, nucleic acid can be extracted from a biological sample by a variety of techniques such as those described by Maniatis, *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, N.Y., pp. 280-281 (1982). Generally, target nucleic acid molecules can be from about 5 bases to about 20 kb, about 30 kb, or even about 40 kb or more. Nucleic acid molecules may be single-stranded, double-stranded, or double-stranded with single-stranded regions (for example, stem- and loop-structures).

[0036] Single molecule sequencing includes a template nucleic acid molecule/primer duplex that is immobilized on a surface such that the duplex and/or the nucleotides (or nucleotide analogs) added to the immobilized primer are individually optically resolvable. The primer, template and/or nucleotide analogs are detectably labeled such that the position of an individual duplex molecule is individually optically resolvable. Either the primer or the template is

immobilized to a solid support. The primer and template can be hybridized to each other and optionally covalently cross-linked prior to or after attachment of either the template or the primer to the solid support.

[0037] In general, methods for facilitating the incorporation of a nucleotide analog as an extension of a primer include exposing a target nucleic acid/primer duplex to one or more nucleotide analogs disclosed herein and a polymerase under conditions suitable to extend the primer in a template dependent manner. Generally, the primer is sufficiently complementary to at least a portion of the target nucleic acid to hybridize to the target nucleic acid and allow template-dependent nucleotide polymerization. The primer extension process can be repeated to identify additional nucleotide analogs in the template. The sequence of the template is determined by compiling the detected nucleotides, thereby determining the complementary sequence of the target nucleic acid molecule.

[0038] Any polymerase and/or polymerizing enzyme may be employed. A preferred polymerase is Klenow with reduced exonuclease activity. Nucleic acid polymerases generally useful in the invention include DNA polymerases, RNA polymerases, reverse transcriptases, and mutant or altered forms of any of the foregoing. DNA polymerases and their properties are described in detail in, among other places, DNA Replication 2nd edition, Komberg and Baker, W. H. Freeman, New York, N.Y. (1991). Known conventional DNA polymerases useful in the invention include, but are not limited to, Pyrococcus furiosus (Pfu) DNA polymerase (Lundberg *et al.*, 1991, Gene, 108: 1, Stratagene), Pyrococcus woesei (Pwo) DNA polymerase (Hinnisdaels *et al.*, 1996, Biotechniques, 20:186-8, Boehringer Mannheim), Thermus thermophilus (Tth) DNA polymerase (Myers and Gelfand 1991, Biochemistry 30:7661), Bacillus stearothermophilus DNA polymerase (Stenesh and McGowan, 1977, Biochim Biophys Acta 475:32), Thermococcus litoralis (Tli) DNA polymerase (also referred to as Vent™ DNA polymerase, Cariello *et al.*, 1991, Polynucleotides Res, 19: 4193, New England Biolabs), 9°Nm™ DNA polymerase (New England Biolabs), Stoffel fragment, ThermoSequenase® (Amersham Pharmacia Biotech UK), Therminator™ (New England Biolabs), Thermotoga maritima (Tma) DNA polymerase (Diaz and Sabino, 1998 Braz J Med. Res, 31:1239), Thermus aquaticus (Taq) DNA polymerase (Chien *et al.*, 1976, J. Bacteoriol, 127: 1550), DNA polymerase, Pyrococcus kodakaraensis KOD DNA polymerase (Takagi *et al.*, 1997, Appl. Environ. Microbiol. 63:4504), JDF-3 DNA polymerase

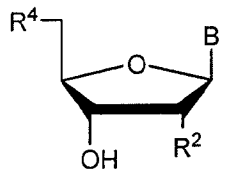
(from thermococcus sp. JDF-3, Patent application WO 0132887), Pyrococcus GB-D (PGB-D) DNA polymerase (also referred as Deep Vent™ DNA polymerase, Juncosa-Ginesta *et al.*, 1994, Biotechniques, 16:820, New England Biolabs), UlTma DNA polymerase (from thermophile Thermotoga maritima; Diaz and Sabino, 1998 Braz J. Med. Res, 31:1239; PE Applied Biosystems), Tgo DNA polymerase (from thermococcus gorgonarius, Roche Molecular Biochemicals), E. coli DNA polymerase I (Lecomte and Doubleday, 1983, Polynucleotides Res. 11:7505), T7 DNA polymerase (Nordstrom *et al.*, 1981, J Biol. Chem. 256:3112), and archaeal DP1I/DP2 DNA polymerase II (Cann *et al.*, 1998, Proc Natl Acad. Sci. USA 95:14250-->5).

[0039] Other DNA polymerases include, but are not limited to, ThermoSequenase®, 9°Nm™, Therminator™, Taq, Tne, Tma, Pfu, Tfl, Tth, Tli, Stoffel fragment, Vent™ and Deep Vent™ DNA polymerase, KOD DNA polymerase, Tgo, JDF-3, and mutants, variants and derivatives thereof. Reverse transcriptases useful in the invention include, but are not limited to, reverse transcriptases from HIV, HTLV-1, HTLV-II, FeLV, FIV, SIV, AMV, MMTV, MoMuLV and other retroviruses (see Levin, Cell 88:5-8 (1997); Verma, Biochim Biophys Acta. 473:1-38 (1977); Wu *et al.*, CRC Crit Rev Biochem. 3:289-347(1975)).

[0040] Unincorporated nucleotide analog molecules may be removed prior to or after detecting. Unincorporated nucleotide analog molecules may be removed by washing.

[0041] A template/primer duplex is treated to remove the label. The steps of exposing template/primer duplex to one or more nucleotide analogs and polymerase, detecting incorporated nucleotides, and then treating to remove the label. These steps can be repeated, thereby identifying additional bases in the template nucleic acid, the identified bases can be compiled, thereby determining the sequence of the target nucleic acid. All portions of the label and the linkage from the label to the nucleotide analog are removed.

[0042] In some embodiments, a nucleotide analog, after removal of the label and portions of the molecular chain connecting the label to the nucleotide can be represented by:



II

where B can be any base, and can be for example selected from the group consisting of a purine, a pyrimidine, and analogs thereof. R^2 may be selected from the group consisting of H and OH. R^4 can be a phosphodiester linkage connecting the nucleotide analog to a sugar of an adjacent nucleotide in the nucleic acid, or a phosphoryl group.

[0043] One embodiment of a method for sequencing a nucleic acid template includes exposing a nucleic acid template to a primer capable of hybridizing to the template to a polymerase capable of catalyzing nucleotide addition to the primer and a labeled nucleotide analog disclosed herein under conditions to permit the polymerase to add the nucleotide analog to the primer. A method for sequencing may further include identifying or detecting the incorporated labeled nucleotide. A cleavable bond may then be cleaved, removing at least the label from the nucleotide analog. The exposing, detecting, and removing steps are repeated at least once. In certain embodiments, the exposing, detecting, and removing steps are repeated at least three, five, ten or even more times. The sequence of the template can be determined based upon the order of incorporation of the labeled nucleotides.

[0044] In another embodiment, a method for sequencing a nucleic acid template includes exposing a nucleic acid template to a primer capable of hybridizing to the template and a polymerase capable of catalyzing nucleotide addition to the primer. The polymerase is, for example, Klenow with reduced exonuclease activity. The polymerase adds a labeled nucleotide analog disclosed herein. The method may include identifying the incorporated labeled nucleotide. Once the labeled nucleotide is identified, the label is removed and resulting nucleotide analog has a hydroxyl group or a phosphate group at the 3' position. The exposing, incorporating, identifying, and removing steps are repeated at least once, preferably multiple

times. The sequence of the template is determined based upon the order of incorporation of the labeled nucleotides.

[0045] Removal of a label from a disclosed labeled nucleotide analog and/or cleavage of the molecular chain linking a disclosed nucleotide to a label may include contacting or exposing the labeled nucleotide with a reducing agent. Such reducing agents include, for example, dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP), tris(3-hydroxy-propyl) phosphine, tris(2-chloropropyl) phosphate (TCPP), 2-mercaptoethanol, 2-mercaptoethylamine, cystein and ethylmaleimide. Such contacting or exposing the reducing agent to a labeled nucleotide analog may occur at a range of pH, for example at a pH of about 5 to about 10, or about 7 to about 9.

[0046] In an embodiment, a nucleotide resulting from a label removal may be contacted with an enzyme, e.g. phophatase, that may hydrolysis aphosphate group at the 3' position.

[0047] Any 3' phosphate moiety can be removed enzymatically from a nucleotide resulting from a label removal. In one embodiment, an optional phosphate can be removed using alkaline phosphatase or T_4 polynucleotide kinase. Suitable enzymes for removing optional phosphate include, any phosphatase, for example, alkaline phosphatase such as shrimp alkaline phosphatase, bacterial alkaline phosphatase, or calf intestinal alkaline phosphatase.

[0048] Reference to the following figure illustrating exemplary reaction schemes and nucleotide analogs is intended in no way to limit the scope of this invention but are provided to illustrate how to prepare and use the compounds of the present invention. Many other embodiments of this invention will be apparent to one skilled in the art.

[0049] Figure 1 depicts an exemplary labeled nucleotide analog of this disclosure. The labeled nucleotide of compound **1** is prepared using standard chemistry. Upon exposure to TCEP, the label of **1** is removed and the molecular chain linking the label to the phosphate is removed as heterocyclic compound **2**; resulting in nucleotide analog **4**, which is identical to a native nucleotide. Upon exposure to a reducing agent, the label from **1** is removed resulting in analog **3**.

Detection

[0050] Any detection method may be used to identify an incorporated nucleotide analog that is suitable for the type of label employed. Thus, exemplary detection methods include radioactive detection, optical absorbance detection, e.g., UV-visible absorbance detection, optical emission detection, e.g., fluorescence or chemiluminescence. Single-molecule fluorescence can be made using a conventional microscope equipped with total internal reflection (TIR) objective. The detectable moiety associated with the extended primers can be detected on a substrate by scanning all or portions of each substrate simultaneously or serially, depending on the scanning method used. For fluorescence labeling, selected regions on a substrate may be serially scanned one-by-one or row-by-row using a fluorescence microscope apparatus, such as described in Fodor (U.S. Patent No. 5,445,934) and Mathies et al. (U.S. Patent No. 5,091,652). Devices capable of sensing fluorescence from a single molecule include scanning tunneling microscope (siM) and the atomic force microscope (AFM). Hybridization patterns may also be scanned using a CCD camera (e.g., Model TE/CCD512SF, Princeton Instruments, Trenton, N.J.) with suitable optics (Ploem, in Fluorescent and Luminescent Probes for Biological Activity Mason, T.G. Ed., Academic Press, Landon, pp. 1-11 (1993), such as described in Yershov et al., Proc. Natl. Aca. Sci. 93:4913 (1996), or may be imaged by TV monitoring. For radioactive signals, a phosphorimager device can be used (Johnston et al., Electrophoresis, 13:566, 1990; Drmanac et al., Electrophoresis, 13:566, 1992; 1993). Other commercial suppliers of imaging instruments include General Scanning Inc., (Watertown, Mass. on the World Wide Web at genscan.com), Genix Technologies (Waterloo, Ontario, Canada; on the World Wide Web at confocal.com), and Applied Precision Inc. Such detection methods are particularly useful to achieve simultaneous scanning of multiple attached target nucleic acids.

[0051] The present invention provides for detection of molecules from a single nucleotide to a single target nucleic acid molecule. A number of methods are available for this purpose. Methods for visualizing single molecules within nucleic acids labeled with an intercalating dye include, for example, fluorescence microscopy. For example, the fluorescent spectrum and lifetime of a single molecule excited-state can be measured. Standard detectors such as a photomultiplier tube or avalanche photodiode can be used. Full field imaging with a

two-stage image intensified CCD camera also can be used. Additionally, low noise cooled CCD can also be used to detect single fluorescent molecules.

[0052] The detection system for the signal may depend upon the labeling moiety used. For optical signals, a combination of an optical fiber or charged couple device (CCD) can be used in the detection step. In those circumstances where the substrate is itself transparent to the radiation used, it is possible to have an incident light beam pass through the substrate with the detector located opposite the substrate from the target nucleic acid. For electromagnetic labeling moieties, various forms of spectroscopy systems can be used. Various physical orientations for the detection system are available and discussion of important design parameters is provided in the art.

[0053] A number of approaches can be used to detect incorporation of fluorescently-labeled nucleotides into a single nucleic acid molecule. Optical setups include near-field scanning microscopy, far-field confocal microscopy, wide-field epi-illumination, light scattering, dark field microscopy, photoconversion, single and/or multiphoton excitation, spectral wavelength discrimination, fluorophore identification, evanescent wave illumination, and total internal reflection fluorescence (TIRF) microscopy. In general, certain methods involve detection of laser-activated fluorescence using a microscope equipped with a camera. Suitable photon detection systems include, but are not limited to, photodiodes and intensified CCD cameras. For example, an intensified charge couple device (ICCD) camera can be used. The use of an ICCD camera to image individual fluorescent dye molecules in a fluid near a surface provides numerous advantages. For example, with an ICCD optical setup, it is possible to acquire a sequence of images (movies) of fluorophores.

[0054] Some embodiments of the present invention use TIRF microscopy for two-dimensional imaging. TIRF microscopy uses totally internally reflected excitation light and is well known in the art. See, e g., the World Wide Web at nikon-instruments.jp/eng/page/products/tirf.aspx. In certain embodiments, detection is carried out using evanescent wave illumination and total internal reflection fluorescence microscopy. An evanescent light field can be set up at the surface, for example, to image fluorescently-labeled nucleic acid molecules. When a laser beam is totally reflected at the interface between a liquid

and a solid substrate (e.g., a glass), the excitation light beam penetrates only a short distance into the liquid. The optical field does not end abruptly at the reflective interface, but its intensity falls off exponentially with distance. This surface electromagnetic field, called the "evanescent wave", can selectively excite fluorescent molecules in the liquid near the interface. The thin evanescent optical field at the interface provides low background and facilitates the detection of single molecules with high signal-to-noise ratio at visible wavelengths.

[0055] The evanescent field also can image fluorescently-labeled nucleotides upon their incorporation into the attached target nucleic acid target molecule/primer complex in the presence of a polymerase. Total internal reflectance fluorescence microscopy is then used to visualize the attached target nucleic acid target molecule/primer complex and/or the incorporated nucleotides with single molecule resolution.

[0056] Fluorescence resonance energy transfer (FRET) can be used as a detection scheme. FRET in the context of sequencing is described generally in Braslavasky, et al., Proc. Nat'l Acad. Sci., 100: 3960-3964 (2003), incorporated by reference herein. In an embodiment, a donor fluorophore is attached to the primer, polymerase, or template. Nucleotides added for incorporation into the primer comprise an acceptor fluorophore that is activated by the donor when the two are in proximity.

[0057] Measured signals can be analyzed manually or preferably by appropriate computer methods to tabulate results. Preferably, the signals of millions of analogs are read in parallel and then deconvoluted to ascertain a sequence. The substrates and reaction conditions can include appropriate controls for verifying the integrity of hybridization and extension conditions, and for providing standard curves for quantification, if desired. For example, a control nucleic acid can be added to the sample. The absence of the expected extension product is an indication that there is a defect with the sample or assay components requiring correction.

[0058] **Example**

[0059] The 7249 nucleotide genome of the bacteriophage M13mp18 is sequenced using nucleotide analogs of the invention.

[0060] Purified, single-stranded viral M13mp18 genomic DNA is obtained from New England Biolabs. Approximately 25 ug of M13 DNA is digested to an average fragment size of 40 bp with 0.1 U Dnase I (New England Biolabs) for 10 minutes at 37° C. Digested DNA fragment sizes are estimated by running an aliquot of the digestion mixture on a precast denaturing (TBE-Urea) 10% polyacrylamide gel (Novagen) and staining with SYBR Gold (Invitrogen/Molecular Probes). The DNase I-digested genomic DNA is filtered through a YM10 ultrafiltration spin column (Millipore) to remove small digestion products less than about 30 nt. Approximately 20 pmol of the filtered DNase I digest was then polyadenylated with terminal transferase according to known methods (Roychoudhury, R and Wu, R.1980, Terminal transferase-catalyzed addition of nucleotides to the 3' termini of DNA. Methods Enzymol. 65(1):43-62.). The average dA tail length is about 50+/-5 nucleotides. Terminal transferase is then used to label the fragments with Cy3-dUTP. Fragments are then terminated with dideoxyTTP (also added using terminal transferase). The resulting fragments are again filtered with a YM10 ultrafiltration spin column to remove free nucleotides and stored in ddH2O at –20° C.

[0061] Epoxide-coated glass slides are prepared for oligo attachment. Epoxide-functionalized 40mm diameter #1.5 glass cover slips (slides) are obtained from Erie Scientific (Salem, NH). The slides are preconditioned by soaking in 3xSSC for 15 minutes at 37° C. Next, a 500 pM aliquot of 5' aminated polydT(50) (polythymidine of 50 bp in length with a 5' terminal amine) is incubated with each slide for 30 minutes at room temperature in a volume of 80 ml. The resulting slides have poly(dT50) primer attached by direct amine linker to the epoxide. The slides are then treated with phosphate (1 M) for 4 hours at room temperature in order to passivate the surface. Slides are then stored in polymerase rinse buffer (20 mM Tris, 100 mM NaCl, 0.001% Triton® X-100 (polyoxyethylene octyl phenyl ether), pH 8.0) until used for sequencing.

[0062] For sequencing, the slides are placed in a modified FCS2 flow cell (Bioptechs, Butler, PA) using a 50 um thick gasket. The flow cell is placed on a movable stage that is part of a high-efficiency fluorescence imaging system built around a Nikon TE-2000 inverted microscope equipped with a total internal reflection (TIR) objective. The slide is then rinsed with HEPES buffer with 100 mM NaCl and equilibrated to a temperature of 50° C. An aliquot of the M13 template fragments described above is diluted in 3xSSC to a final concentration of 1.2

nM. A 100 ul aliquot is placed in the flow cell and incubated on the slide for 15 minutes. After incubation, the flow cell is rinsed with 1xSSC/HEPES/0.1% SDS followed by HEPES/NaCl. A passive vacuum apparatus is used to pull fluid across the flow cell. The resulting slide contains M13 template/oligo(dT) primer duplex. The temperature of the flow cell is then reduced to 37° C for sequencing and the objective is brought into contact with the flow cell.

[0063] For sequencing, cytosine triphosphate analog, guanidine triphosphate analog, adenine triphosphate analog, and uracil triphosphate analog, each having a fluorescent label, such as a Cy5, attached to a nucleotide, such as the labeled nucleotide analogs disclosed herein. The analogs are stored separately in buffer containing 20 mM Tris-HCl, pH 8.8, 10 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$, 10 mM HCl, and 0.1% Triton® X-100 (polyoxyethylene octyl phenyl ether), and 100U Klenow exo$^-$ polymerase (NEN). Sequencing proceeds as follows.

[0064] First, initial imaging is used to determine the positions of duplex on the epoxide surface. The Cy3 label attached to the M13 templates is imaged by excitation using a laser tuned to 532 nm radiation (Verdi V-2 Laser, Coherent, Inc., Santa Clara, CA) in order to establish duplex position. For each slide only single fluorescent molecules imaged in this step are counted. Imaging of incorporated nucleotides as described below is accomplished by excitation of a cyanine-5 dye using a 635 nm radiation laser (Coherent). 5 uM of a Cy5-labeled CTP analog as described above is placed into the flow cell and exposed to the slide for 2 minutes. After incubation, the slide is rinsed in 1xSSC/15 mM HEPES/0.1% SDS/pH 7.0 ("SSC/HEPES/SDS") (15 times in 60 ul volumes each, followed by 150 mM HEPES/150 mM NaCl/pH 7.0 ("HEPES/NaCl") (10 times at 60 ul volumes)). An oxygen scavenger containing 30% acetonitrile and scavenger buffer (134 ul HEPES/NaCl, 24 ul 100 mM Trolox in MES, pH 6.1, 10 ul DABCO in MES, pH 6.1, 8 ul 2M glucose, 20 ul NaI (50 mM stock in water), and 4 ul glucose oxidase) is next added. The slide is then imaged (500 frames) for 0.2 seconds using an Inova301K laser (Coherent) at 647nm, followed by green imaging with a Verdi V-2 laser (Coherent) at 532 nm for 2 seconds to confirm duplex position. The positions having detectable fluorescence are recorded. After imaging, the flow cell is rinsed 5 times each with SSC/HEPES/SDS (60 ul) and HEPES/NaCl (60 ul).

[0065] Next, the fluorescent label (*e.g.*, the cyanine-5) is removed or cleaved off of the incorporated CTP analogs. The Cy5 label is removed by introduction into the flow cell of 50 mM TCEP for 5 minutes, after which the flow cell was rinsed 5 times each with SSC/HEPES/SDS (60 ul) and HEPES/NaCl (60 ul), and the remaining nucleotide is capped with 50 mM iodoacetamide for 5 minutes followed by rinsing 5 times each with SSC/HEPES/SDS (60 ul) and HEPES/NaCl (60 ul). The scavenger is applied again in the manner described above, and the slide is again imaged to determine the effectiveness of the cleave/cap steps and to identify non-incorporated fluorescent objects.

[0066] The procedure described above is then conducted 100 nM Cy5dATP analog, followed by 100 nM Cy5dGTP analog, and finally 500 nM Cy5dUTP, each as described above. The procedure (expose to nucleotide, polymerase, rinse, scavenger, image, rinse, cleave, rinse, cap, rinse, scavenger, final image, removal of optional phosphate group) is repeated exactly as described for ATP, GTP, and UTP except that Cy5dUTP is incubated for 5 minutes instead of 2 minutes. Uridine is used instead of thymidine due to the fact that the Cy5 label is incorporated at the position normally occupied by the methyl group in thymidine triphosphate, thus turning the dTTP into dUTP. In all 64 cycles (C, A, G, U) are conducted as described in this and the preceding paragraph.

[0067] Once 64 cycles are completed, the image stack data (*i.e.*, the single molecule sequences obtained from the various surface-bound duplex) is aligned to the M13 reference sequence.

[0068] The alignment algorithm matches sequences obtained as described above with the actual M13 linear sequence. Placement of obtained sequence on M13 is based upon the best match between the obtained sequence and a portion of M13 of the same length, taking into consideration 0, 1, or 2 possible errors. All obtained 9-mers with 0 errors (meaning that they exactly matched a 9-mer in the M13 reference sequence) are first aligned with M13. Then 10-, 11-, and 12-mers with 0 or 1 error are aligned. Finally, all 13-mers or greater with 0, 1, or 2 errors are aligned.

[0069] All publications, patents, and patent applications cited herein are hereby expressly incorporated by reference in their entirety and for all purposes to the same extent as if each was

so individually denoted. The patent applications entitled "Nucleotide Analogs" filed on even date herewith (Attorney Docket Numbers: HEL-040; HEL-039) are each expressly incorporated by reference.

Equivalents

[0070] While specific embodiments of the subject invention have been discussed, the above specification is illustrative and not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. Contemplated equivalents of the nucleotide analogs disclosed here include compounds which otherwise correspond thereto, and which have the same general properties thereof, wherein one or more simple variations of substituents or components are made which do not adversely affect the characteristics of the nucleotide analogs of interest. In general, the components of the nucleotide analogs disclosed herein may be prepared by the methods illustrated in the general reaction schema as described herein or by modifications thereof, using readily available starting materials, reagents, and conventional synthesis procedures. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

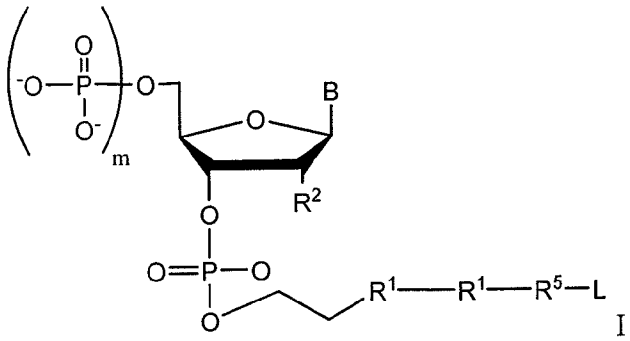
[0071] Unless otherwise indicated, all numbers expressing quantities of ingredients, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about." Accordingly, unless indicated to the contrary, the numerical parameters set forth in this specification and attached claims are approximations that may vary depending upon the desired properties sought to be obtained by the present invention.

[0072] The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in all respects illustrative rather than limiting on the invention described herein. Scope of the invention is thus indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

Claims:

What is claimed is:

1. A labeled nucleotide analog of Formula I:



I

wherein,

R^1 at each occurrence, independently is selected from the group consisting of S, NR^3 and O,

R^2 is selected from the group consisting of H and OH,

R^3 is selected from the group consisting of H and alkyl,

R^5 is an aliphatic moiety,

B is selected from the group consisting of a purine, a pyrimidine, and analogs thereof;

L is a label, and

m is an integer from 1 to 3.

2. The labeled nucleotide of claim 1, wherein, in each occurrence, R^1 is S.

3. The labeled nucleotide of claim 1 or 2, wherein B is selected from the group consisting of cytosine, uracil, thymine, adenine, guanine, and analogs thereof.

4. The labeled nucleotide analog of claim 1, wherein L is an optically detectable label.

5. The labeled nucleotide analog of claim 4, wherein the optically detectable label is a fluorescent label.

6. The labeled nucleotide analog of claim 4, wherein the optically detectable label is selected from the group consisting of cyanine, rhodamine, fluoroscein, coumarin, BODIPY, alexa and conjugated multi-dyes.

7. The labeled nucleotide analog of claim 4, wherein the optically detectable label is Cy3 or Cy5.

8. A method of removing a label or protecting group from a nucleotide, the method comprising the steps of:

(a) providing a nucleotide comprising a sugar and a label or protecting group linked via a phosphoryl moiety to a 3' position of the sugar; and

(b) exposing the nucleotide to a reducing agent in an amount and under conditions to remove the label or protecting group.

9. The method of claim 8, wherein after step (b), the nucleotide comprises a phosphoryl group.

10. The method of claim 9, further comprising the step of exposing the nucleotide to a phosphatase to remove the phosphoryl moiety and produce a hydroxyl group.

11. The method of claim 8, wherein in step (b), the reducing agent is tris(2-chloroethyl)phosphate.

12. A method of sequencing a nucleic acid template, the method comprising the steps of:

(a) exposing a nucleic acid template hybridized to a primer having a 3' end to (i) a polymerase capable of catalyzing nucleotide additions to the primer, and (ii) the nucleotide analog of claim 1 under conditions to permit the polymerase to add the nucleotide analog to the 3' end of the primer;

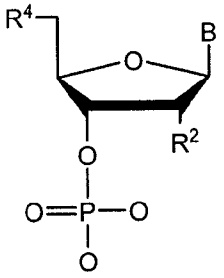
(b) detecting the nucleotide analog added to the primer in step (a); and

(c) removing the label from the nucleotide analog.

13. The method of claim 12, further comprising repeating steps (a), (b) and (c) thereby to determine the sequence of the template.

14. The method of claim 12, wherein, after step (c), the nucleotide analog has a hydroxyl group or the phosphoryl group.

15. The method of claim 14, wherein after step (c), the nucleotide analog is represented by formula II:



II

wherein,

R^2 is selected from the group consisting of H or OH,

R^4 is a phosphodiester linkage connecting the nucleotide analog to the primer, and

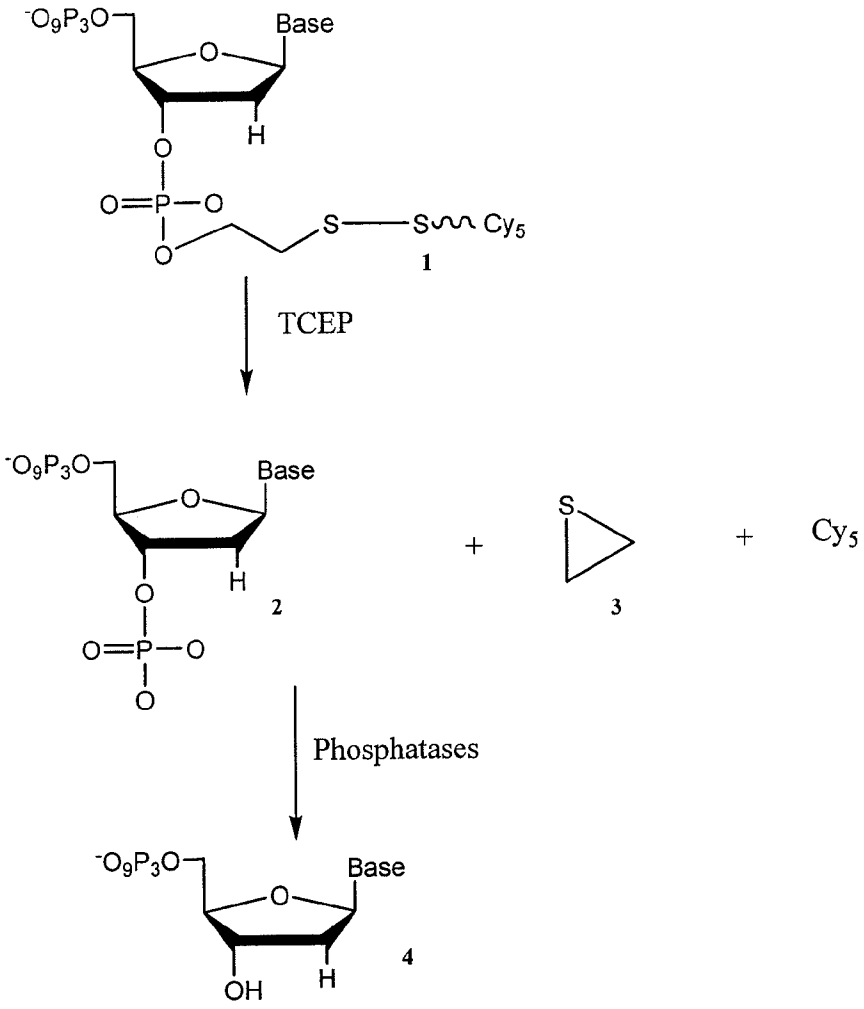
B is selected from the group consisting of a purine, a pyrimidine, and analogs thereof.

16. The method of claim 15, wherein, at step (c), the label is removed by exposure to a reducing agent.

17. The method of claim 16, where the reducing agent is tris(2-carboxyethyl) phosphine.

18. The method of claim 16, further comprising contacting the nucleotide analog with a phosphatase.

FIGURE 1


-O9P3O
Base
O
H
O
O=P-O
O
S
S
Cy5
1
TCEP
-O9P3O
Base
O
H
O
O=P-O
O
2
+
S
3
+
Cy5
Phosphatases
-O9P3O
Base
O
OH
H
4