

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 December 2008 (11.12.2008)

PCT

(10) International Publication Number
WO 2008/151127 A1

(51) International Patent Classification:
C07H 21/00 (2006.01) *G01N 33/00* (2006.01)

(74) Agents: IWANICKI, John, P. et al.; Banner & Witcoff, Ltd., 28 State Street, 28th Floor, Boston, MA 02109 (US).

(21) International Application Number:
PCT/US2008/065523

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

(22) International Filing Date: 2 June 2008 (02.06.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/941,805 4 June 2007 (04.06.2007) US

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

(71) Applicant (*for all designated States except US*): PRES-IDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

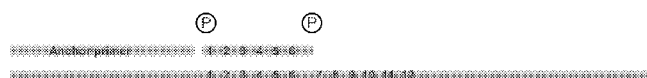
(72) Inventors; and

(75) Inventors/Applicants (*for US only*): CHURCH, George, M. [US/US]; 218 Kent Street, Brookline, MA 02446 (US). SISMOUR, Michael [US/US]; 85 Hancock Street, Apt. 7, Cambridge, MA 02139 (US).

Published:
— with international search report

(54) Title: COMPOUNDS AND METHODS FOR CHEMICAL LIGATION

5A



5B

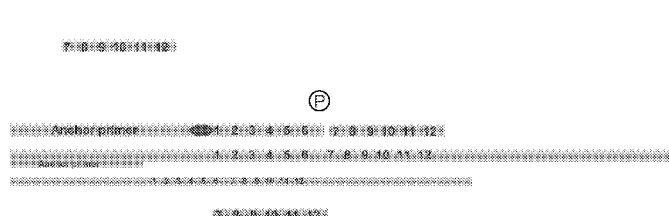


Figure 5

(57) Abstract: Compositions and methods for chemical ligation are provided. Methods for nucleic acid sequencing, nucleic acid assembly and nucleic acid synthesis are also provided.

WO 2008/151127 A1

COMPOUNDS AND METHODS FOR CHEMICAL LIGATION

STATEMENT OF GOVERNMENT INTERESTS

- [001] This application was funded in part by grant number DE-FG02-02ER63445, fund 148780, from the United States Department of Energy. The Government has certain rights in the invention.

RELATED APPLICATION

- [002] This application claims priority from U.S. provisional patent application number 60/941,805, filed June 4, 2007, which is hereby incorporated herein by reference in its entirety for all purposes.

FIELD OF THE INVENTION

- [003] The present invention relates to compounds and methods for chemically ligating nucleic acids.

BACKGROUND OF THE INVENTION

- [004] Known methods for 'next generation' (e.g., polony) sequencing use degenerate oligonucleotides and protein ligases to discriminate among oligonucleotides with perfect match at one or more positions relative to mismatched oligonucleotides (See, e.g., Shendure et al. (2005) *Science* 309:1728; Applied Biosystems Solid; and Complete Genomics). Once ligated to an anchor oligonucleotide bound to an immobilized template, the solid phase is extensively washed to remove free fluor-oligonucleotides and then visualized with a digital camera. The oligonucleotides are typically labeled with fluorophores such that a given emission wavelength corresponds with given base(s) (e.g., A, C, G, T) at specific locations in the oligonucleotide bound to the template (e.g., unknown) DNA.
- [005] Methods employing degenerate oligonucleotides and protein ligases suffer from many disadvantages. In one approach, the coupling to each monomer goes to 50% completion in 0.15 - 2.5 hours and is, thus, too slow to be useful, perhaps in part because there is too little hybridization energy from a single base-pair (Griesang et al.

(2006) *Angew. Chem. Int. Ed. Engl.* 45:6144). Another disadvantage of these art-known methods is that photocleavable protection chemistry requires intense irradiation with UV light, which can damage the DNA template to be sequenced after just a few deprotections.

SUMMARY OF THE INVENTION

- [006] The present invention is based in part on the discovery of novel reagents and methods for performing chemical ligation. Although protein ligases can, in principle, be engineered to accommodate a variety of template, oligonucleotide and anchor substrates, chemical ligation is far more versatile than protein ligases in accepting non-standard bases, backbone, and chemical bonds at the ligation position as well as throughout the probe oligonucleotide sequence. The chemical ligation reagents and methods described herein provide many advantages over traditional methods used for sequencing reactions, DNA assembly and the like.
- [007] In contrast with traditional ligase-based methods that query a single base at a time, the identity of multiple base pairs can be resolved in sequencing methods employing the chemical ligation reagents and methods described further herein. Further, the chemical ligation reagents and/or methods can utilize a no deprotection method for multiple (e.g., two rounds of ligation) and/or chemical deprotection method(s) (e.g., non-photochemical methods) for more than two rounds, thus eliminating the risk of UV damage to the DNA template.
- [008] In certain exemplary embodiments, a method of determining a nucleotide sequence of a reference oligonucleotide is provided. The method includes the steps of providing a reference oligonucleotide having an anchor primer bound thereto, providing a first probe oligonucleotide having a preactivated chemical ligation compound attached thereto, allowing the first probe oligonucleotide to hybridize to the reference oligonucleotide, and allowing the preactivated chemical ligation compound to mediate ligation between the hybridized first probe oligonucleotide and the anchor primer. In certain aspects, the preactivated chemical ligation compound is selected from the group consisting of a cyanogen bromide-based compound (e.g., a phosphocyanate such as N-Morpholino)Ethane Sulfonic Acid-cyanogen), a carbodiimide-based

compound (e.g., an O-phospho-isocarbamide compound, such as phospho-1-Ethyl-3-[3-Dimethylaminopropyl]Carbodiimide), an amidate-based compound (e.g., a phosphoramidate that may optionally include an R group selected from the group consisting of pyridine, aniline, imidazole, ethanolamine, azaoxybenzotriazolidine, N-oxy succinimide and succinimide), and a phosphodiester mimetic (e.g., an imine). In certain aspects, the first probe oligonucleotide further includes one or more S²T molecules and/or one or more diaminopurines. In other aspects, one or more thymines of the first probe oligonucleotide are replaced by one or more S²T molecules. In yet other aspects, one or more adenines of the first probe oligonucleotide are replaced by one or more diaminopurines. In still other aspects, the method further includes the steps of providing a second probe oligonucleotide, allowing the second probe oligonucleotide to hybridize to the reference oligonucleotide, and allowing the preactivated chemical ligation compound to mediate ligation between the hybridized second probe oligonucleotide and the hybridized first probe oligonucleotide. The second probe oligonucleotide may optionally have a preactivated chemical ligation compound attached thereto.

[009] In other exemplary embodiments, a method of assembling a polynucleotide sequence is provided. The method includes the steps of providing a plurality of overlapping oligonucleotides, at least a portion of which have a preactivated chemical ligation compound attached thereto, allowing hybridization of at least a portion of the overlapping oligonucleotides to each other, and allowing the preactivated chemical ligation compounds to mediate ligation between adjacent 5' and 3' ends of hybridized oligonucleotides to assemble a polynucleotide sequence. In certain aspects, the polynucleotide sequence is DNA such as, for example, one or more of a vector, a gene, a gene fragment, an exon, an intron or an intergenic DNA sequence.

[010] In other exemplary embodiments, a method of synthesizing a polynucleotide sequence is provided. The method includes the steps of providing a first oligonucleotide having a preactivated chemical ligation compound attached thereto, providing a second oligonucleotide, and allowing the preactivated chemical ligation compound to mediate ligation between the first oligonucleotide and the second oligonucleotide. In certain aspects, the method further includes the step of allowing the first oligonucleotide to

bind a reference oligonucleotide. In other aspects, the reference oligonucleotide is attached to a substrate.

BRIEF DESCRIPTION OF THE DRAWINGS

- [011] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee. The foregoing and other features and advantages of the present invention will be more fully understood from the following detailed description of illustrative embodiments taken in conjunction with the accompanying drawings in which:
- [012] **Figures 1A-1C** depict certain nitrogen compounds useful in chemical ligation reactions. A) cyanogen bromide (CNBr). B) MES-CN (MES = 2-(N-Morpholino)Ethane Sulfonic Acid). C) a 3' phosphocyanate compound.
- [013] **Figures 2A-2B** depict carbodiimide compounds. A) EDC (1-Ethyl-3-[3-Dimethylaminopropyl]Carbodiimide). B) a 3' phospho-EDC compound.
- [014] **Figures 3A-3D** depict phosphoramidate compounds. A) a 3' phosphoramidate compound. B) a 5' phosphoramidate compound. C) a 3' phosphoramidate compound. D) a variety of R groups useful with phosphoramidate compounds such as, for example, the compounds depicted in A) – C). Left to right: pyridine, aniline, imidazole, ethanolamine, azaoxybenzotriazolide, N-oxy succinimide and succinimide.
- [015] **Figures 4A-4B** depict cyanogen bromide and EDC chemical ligation reactions. These reactions were very fast (a few seconds) and were limited only by the hybridization kinetics in a complex mixture of 6-mer to 9-mer oligonucleotides.
- [016] **Figures 5A-5B** schematically depict chemical ligation using pre-activated phosphates. A) first round hybridization. Template, lower strand 1-12; Probe with phosphates at both ends, upper right, 1-6. B) second round hybridization. As part A) after addition of the first probe depicting efficient reaction with the remaining phosphate (upper) and failure of a second probe to react in the first position (even in the absence of pair mismatches) (lower).

- [017] **Figures 6A-6B** depict reaction schemes for the ligation, capping of non-reacted primers, and deprotection of 3'-azido probes for subsequent ligations. The reaction scheme (A) includes, from left to right, the ligation of a 3'-amino primer oligonucleotide with an oligonucleotide probe, followed by capping any non-reacted 3'-amino primer oligonucleotides by addition of excess aldehyde for an imine that is subsequently reduced to an amide using NaCNBH_3 . This was followed by deprotection of the 3'-azido probe oligonucleotide using NaBH_4 in the presence of CoCl_2 (A, B). This capping/deprotection scheme can be repeated for subsequent ligations of 3'-azido probe oligonucleotides.
- [018] **Figures 7A-7D** depict various phosphodiester mimetic schemes. Aldehyde functional groups can be formed at the 3' terminus of an oligonucleotide using an oligonucleotide containing a (A) 3' alkyl vicinal diol or (B) an oligonucleotide terminated at the 3' end in a nucleotide (as opposed to a deoxynucleotide) by oxidation with periodate. Such aldehydes can form (C) imines with oligonucleotides containing a 5' amine. In addition to such imine phosphodiester mimetics, (D) the vicinal diol of oligonucleotides terminated at the 3' end in a nucleotide (as opposed to a deoxynucleotide) can react with cyanogen bromide to form a hydroxy-isocyanate (left most molecule) that can further react with a 5'-OH on an oligonucleotide to form the hydroxy-isocyanate phosphodiester mimetic.
- [019] **Figures 8A-8B** depict various nucleotide substitutions that increase T_m (A) and increase specificity (B).

DETAILED DESCRIPTION

- [020] The principles of the present invention may be applied with particular advantage to determine the identity of oligonucleotide sequences. The principles of the present invention may further be applied with particular advantage to assemble and/or synthesize polynucleotide sequences.
- [021] DNA ligation is the process of joining together two DNA molecule ends (either from the same or different molecules). Specifically, it typically involves creating a phosphodiester bond between the 3' hydroxyl of one nucleotide and the 5' phosphate

of another, and is usually catalyzed by a DNA ligase. As used herein, the term “chemical ligation compound” refers to a variety of chemical compounds that can be used, alone or in combination, to mediate one or more chemical ligation reactions. Useful chemical ligation compounds include, but are not limited to: 1) cyanogen bromide-based compounds (e.g., phosphocyanates such as MES-CN (MES = 2-(N-Morpholino)Ethane Sulfonic Acid) and the like); 2) carbodiimide-based compounds (e.g., O-phospho-isocarbamide compounds such as phospho-EDC; 3) amidate-based compounds (e.g., phosphoramidates); and 4) phosphodiester mimetics (e.g., aldehydes generated by periodate, Schiff’s base intermediates and the like). Chemical ligation compounds and reactions including one or more of them are described further herein.

- [022] In certain exemplary embodiments, an oligonucleotide sequence having one or more preactivated phosphates is provided. As used herein, the term “preactivated phosphate” refers to a nucleotide or an oligonucleotide having a phosphate group (e.g., at the 5' and/or 3' end of the nucleotide or oligonucleotide) available for covalently binding a nucleotide and/or oligonucleotide sequence having an available non-phosphate reactive group, e.g., an hydroxyl, a sulfhydryl, an amine or the like.
- [023] As used herein, the terms “nucleic acid molecule,” “nucleic acid sequence,” “nucleic acid fragment,” “oligonucleotide” and “polynucleotide” are used interchangeably and are intended to include, but not limited to, a polymeric form of nucleotides that may have various lengths, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Different polynucleotides may have different three-dimensional structures, and may perform various functions, known or unknown. Non-limiting examples of polynucleotides include a gene, a gene fragment, an exon, an intron, intergenic DNA (including, without limitation, heterochromatic DNA), messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, small interfering RNA (siRNA), cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of a sequence, isolated RNA of a sequence, nucleic acid probes, and primers. Oligonucleotides useful in the methods described herein may comprise natural nucleic acid sequences and variants thereof, artificial nucleic acid sequences, or a combination of such sequences.

- [024] A polynucleotide is typically composed of a specific sequence of four nucleotide bases: adenine (A); cytosine (C); guanine (G); and thymine (T) (uracil (U) for thymine (T) when the polynucleotide is RNA). Thus, the term “polynucleotide sequence” is the alphabetical representation of a polynucleotide molecule; alternatively, the term may be applied to the polynucleotide molecule itself. This alphabetical representation can be input into databases in a computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching. Polynucleotides may optionally include one or more non-standard nucleotide(s), nucleotide analog(s) and/or modified nucleotides.
- [025] Examples of modified nucleotides include, but are not limited to diaminopurine, S²T, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl)uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-D46-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, 2,6-diaminopurine and the like. Nucleic acid molecules may also be modified at the base moiety (e.g., at one or more atoms that typically are available to form a hydrogen bond with a complementary nucleotide and/or at one or more atoms that are not typically capable of forming a hydrogen bond with a complementary nucleotide), sugar moiety or phosphate backbone.
- [026] In accordance with certain examples, methods of sequencing nucleic acid sequences by hybridization using one or more of the chemical ligation compounds and/or chemical ligation methods described herein are provided. General sequencing methods known in the art, such as sequencing by extension with reversible

terminators, fluorescent *in situ* sequencing (FISSEQ), pyrosequencing, massively parallel signature sequencing (MPSS), solid sequencing (Applied Biosystems, Foster city, CA), and the like (described in Shendure et al. (2004) *Nat. Rev.* 5:335; reversible termination methods: U.S. Patent Nos. 5,750,341 and 6,306,597; FISSEQ: Mitra et al. (2003) *Anal. Biochem.* 320:55; Pyrosequencing: Ronaghi et al. (1998) *Science* 281:363; MPSS: Brenner et al. (2000) *Nat. Biotech.* 18:630), are suitable for use with the chemical ligation compounds and assays described herein.

- [027] In exemplary embodiments, a plurality of overlapping oligonucleotide sequences will be used to assemble a larger polynucleotide sequence, such as DNA. The plurality of overlapping oligonucleotides can be exposed to conditions that allow hybridization to form duplexes. For example, oligonucleotide sequences may be exposed to hybridization conditions followed by ligation using one or more chemical ligation reagents and/or chemical ligation methods described further herein. Methods of ligating preformed duplexes using DNA ligases are well-known in the art (*See* Scarpulla et al. (1982) *Anal. Biochem.* 121:356, and Gupta et al. (1968) *Proc. Natl. Acad. Sci. USA* 60:1338).
- [028] Oligonucleotide sequences may be isolated from natural sources or purchased from commercial sources. In certain exemplary embodiments, oligonucleotide sequences may be prepared using one or more of the chemical ligation compounds and/or the chemical ligation methods described herein. Oligonucleotide sequences may also be prepared by any suitable method, e.g., the phosphoramidite method described by Beaucage and Carruthers ((1981) *Tetrahedron Lett.* 22: 1859) or the triester method according to Matteucci et al. (1981) *J. Am. Chem. Soc.* 103:3185), or by other chemical methods using either a commercial automated oligonucleotide synthesizer or high-throughput, high-density array methods known in the art (see U.S. Patent Nos. 5,602,244, 5,574,146, 5,554,744, 5,428,148, 5,264,566, 5,141,813, 5,959,463, 4,861,571 and 4,659,774, incorporated herein by reference in its entirety for all purposes). Pre-synthesized oligonucleotides may also be obtained commercially from a variety of vendors.

- [029] In certain exemplary embodiments, oligonucleotide sequences may be prepared using a variety of microarray technologies known in the art. Pre-synthesized oligonucleotide and/or polynucleotide sequences may be attached to a support or synthesized *in situ* using light-directed methods, flow channel and spotting methods, inkjet methods, pin-based methods and bead-based methods set forth in the following references: McGall et al. (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93:13555; *Synthetic DNA Arrays* In Genetic Engineering, Vol. 20:111, Plenum Press (1998); Duggan et al. (1999) *Nat. Genet.* S21:10; *Microarrays: Making Them and Using Them* In Microarray Bioinformatics, Cambridge University Press, 2003; U.S. Patent Application Publication Nos. 2003/0068633 and 2002/0081582; U.S. Patent Nos. 6,833,450, 6,830,890, 6,824,866, 6,800,439, 6,375,903 and 5,700,637; and PCT Application Nos. WO 04/031399, WO 04/031351, WO 04/029586, WO 03/100012, WO 03/066212, WO 03/065038, WO 03/064699, WO 03/064027, WO 03/064026, WO 03/046223, WO 03/040410 and WO 02/24597.
- [030] In certain exemplary embodiments, one or more oligonucleotide sequences described herein are immobilized on a support. Suitable supports include, but are not limited to, slides, beads, chips, particles, strands, gels, sheets, tubing, spheres, containers, capillaries, pads, slices, films, plates and the like. In various embodiments, the solid supports may be biological, nonbiological, organic, inorganic, or combinations thereof. When using supports that are substantially planar, the support may be physically separated into regions, for example, with trenches, grooves, wells, or chemical barriers (e.g., hydrophobic coatings, etc.).
- [031] In certain exemplary embodiments, a support is a microarray. As used herein, the term “microarray” refers in one embodiment to a type of assay that comprises a solid phase support having a substantially planar surface on which there is an array of spatially defined non-overlapping regions or sites that each contain an immobilized hybridization probe. “Substantially planar” means that features or objects of interest, such as probe sites, on a surface may occupy a volume that extends above or below a surface and whose dimensions are small relative to the dimensions of the surface. For example, beads disposed on the face of a fiber optic bundle create a substantially planar surface of probe sites, or oligonucleotides disposed or synthesized on a porous

planar substrate creates a substantially planar surface. Spatially defined sites may additionally be “addressable” in that its location and the identity of the immobilized probe at that location are known or determinable.

- [032] Probes immobilized on microarrays include nucleic acids that are generated in or from an assay reaction. Typically, the oligonucleotides or polynucleotides on microarrays are single stranded and are covalently attached to the solid phase support, usually by a 5'-end or a 3'-end. The density of non-overlapping regions containing nucleic acids in a microarray is typically greater than 100 per cm², and more typically, greater than 1000 per cm². Microarray technology relating to nucleic acid probes is reviewed in the following exemplary references: Schena, Editor, *Microarrays: A Practical Approach* (IRL Press, Oxford, 2000); Southern, *Current Opin. Chem. Biol.*, 2: 404-410 (1998); *Nature Genetics* Supplement, 21:1-60 (1999); and Fodor et al, U.S. Pat. Nos. 5,424,186; 5,445,934; and 5,744,305.
- [033] Methods of immobilizing oligonucleotides to a support are known in the art (beads: Dressman et al. (2003) *Proc. Natl. Acad. Sci. USA* 100:8817, Brenner et al. (2000) *Nat. Biotech.* 18:630, Albretsen et al. (1990) *Anal. Biochem.* 189:40, and Lang et al. *Nucleic Acids Res.* (1988) 16:10861; nitrocellulose: Ranki et al. (1983) *Gene* 21:77; cellulose: Goldkorn (1986) *Nucleic Acids Res.* 14:9171; polystyrene: Ruth et al. (1987) Conference of Therapeutic and Diagnostic Applications of Synthetic Nucleic Acids, Cambridge U.K.; teflon-acrylamide: Duncan et al. (1988) *Anal. Biochem.* 169:104; polypropylene: Polsky-Cynkin et al. (1985) *Clin. Chem.* 31:1438; nylon: Van Ness et al. (1991) *Nucleic Acids Res.* 19:3345; agarose: Polsky-Cynkin et al., *Clin. Chem.* (1985) 31:1438; and sephacryl: Langdale et al. (1985) *Gene* 36:201; latex: Wolf et al. (1987) *Nucleic Acids Res.* 15:2911).
- [034] As used herein, the term “attach” refers to both covalent interactions and noncovalent interactions. A covalent interaction is a chemical linkage between two atoms or radicals formed by the sharing of a pair of electrons (i.e., a single bond), two pairs of electrons (i.e., a double bond) or three pairs of electrons (i.e., a triple bond). Covalent interactions are also known in the art as electron pair interactions or electron pair bonds. Noncovalent interactions include, but are not limited to, van der Waals

interactions, hydrogen bonds, weak chemical bonds (i.e., via short-range noncovalent forces), hydrophobic interactions, ionic bonds and the like. A review of noncovalent interactions can be found in Alberts et al., in *Molecular Biology of the Cell*, 3d edition, Garland Publishing, 1994.

- [035] In certain exemplary embodiments, a detectable label can be used to detect one or more nucleotides and/or oligonucleotides described herein. Examples of detectable markers include various radioactive moieties, enzymes, prosthetic groups, fluorescent markers, luminescent markers, bioluminescent markers, metal particles, protein-protein binding pairs, protein-antibody binding pairs and the like. Examples of fluorescent proteins include, but are not limited to, yellow fluorescent protein (YFP), green fluorescence protein (GFP), cyan fluorescence protein (CFP), umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride, phycoerythrin and the like. Examples of bioluminescent markers include, but are not limited to, luciferase (e.g., bacterial, firefly, click beetle and the like), luciferin, aequorin and the like. Examples of enzyme systems having visually detectable signals include, but are not limited to, galactosidases, glucuronidases, phosphatases, peroxidases, cholinesterases and the like. Identifiable markers also include radioactive compounds such as ^{125}I , ^{35}S , ^{14}C , or ^3H . Identifiable markers are commercially available from a variety of sources.
- [036] Fluorescent labels and their attachment to nucleotides and/or oligonucleotides are described in many reviews, including Haugland, *Handbook of Fluorescent Probes and Research Chemicals*, Ninth Edition (Molecular Probes, Inc., Eugene, 2002); Keller and Manak, *DNA Probes*, 2nd Edition (Stockton Press, New York, 1993); Eckstein, editor, *Oligonucleotides and Analogues: A Practical Approach* (IRL Press, Oxford, 1991); and Wetmur, *Critical Reviews in Biochemistry and Molecular Biology*, 26:227-259 (1991). Particular methodologies applicable to the invention are disclosed in the following sample of references: Fung et al., U.S. Patent No. 4,757,141; Hobbs, Jr., et al. U.S. Patent No. 5,151,507; Cruickshank, U.S. Patent No. 5,091,519. In one aspect, one or more fluorescent dyes are used as labels for labeled target sequences, e.g., as disclosed by Menchen et al., U.S. Patent No. 5,188,934 (4,7-dichlorofluorescein dyes); Begot et al., U.S. Patent No. 5,366,860 (spectrally

resolvable rhodamine dyes); Lee et al., U.S. Patent No. 5,847,162 (4,7-dichlororhodamine dyes); Khanna et al., U.S. Patent No. 4,318,846 (ether-substituted fluorescein dyes); Lee et al., U.S. Patent No. 5,800,996 (energy transfer dyes); Lee et al., U.S. Patent No. 5,066,580 (xanthine dyes); Mathies et al., U.S. Patent No. 5,688,648 (energy transfer dyes); and the like. Labelling can also be carried out with quantum dots, as disclosed in the following patents and patent publications: U.S. Patent Nos. 6,322,901; 6,576,291; 6,423,551; 6,251,303; 6,319,426; 6,426,513; 6,444,143; 5,990,479; 6,207,392; 2002/0045045; and 2003/0017264. As used herein, the term “fluorescent label” includes a signaling moiety that conveys information through the fluorescent absorption and/or emission properties of one or more molecules. Such fluorescent properties include fluorescence intensity, fluorescence lifetime, emission spectrum characteristics, energy transfer, and the like.

[037] Commercially available fluorescent nucleotide analogues readily incorporated into nucleotide and/or oligonucleotide sequences include, but are not limited to, Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy5-dUTP (Amersham Biosciences, Piscataway, NJ), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, TEXAS REDTM-5-dUTP, CASCADE BLUETM-7-dUTP, BODIPY TMFL-14-dUTP, BODIPY TMR-14-dUTP, BODIPY TMTR-14-dUTP, RHODAMINE GREENTM-5-dUTP, OREGON GREENTM 488-5-dUTP, TEXAS REDTM-12-dUTP, BODIPY TM 630/650-14-dUTP, BODIPY TM 650/665-14-dUTP, ALEXA FLUORTM 488-5-dUTP, ALEXA FLUORTM 532-5-dUTP, ALEXA FLUORTM 568-5-dUTP, ALEXA FLUORTM 594-5-dUTP, ALEXA FLUORTM 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, TEXAS REDTM-5-UTP, mCherry, CASCADE BLUETM-7-UTP, BODIPY TM FL-14-UTP, BODIPY TMR-14-UTP, BODIPY TM TR-14-UTP, RHODAMINE GREENTM-5-UTP, ALEXA FLUORTM 488-5-UTP, ALEXA FLUORTM 546-14-UTP (Molecular Probes, Inc. Eugene, OR) and the like. Protocols are known in the art for custom synthesis of nucleotides having other fluorophores (*See*, Henegariu et al. (2000) *Nature Biotechnol.* 18:345).

[038] Other fluorophores available for post-synthetic attachment include, but are not limited to, ALEXA FLUORTM 350, ALEXA FLUORTM 532, ALEXA FLUORTM 546, ALEXA FLUORTM 568, ALEXA FLUORTM 594, ALEXA FLUORTM 647, BODIPY

493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethyl rhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR), Cy2, Cy3.5, Cy5.5, Cy7 (Amersham Biosciences, Piscataway, NJ) and the like. FRET tandem fluorophores may also be used, including, but not limited to, PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, APC-Cy7, PE-Alexa dyes (610, 647, 680), APC-Alexa dyes and the like.

- [039] Metallic silver or gold particles may be used to enhance signal from fluorescently labeled nucleotide and/or oligonucleotide sequences (Lakowicz et al. (2003) *BioTechniques* 34:62).
- [040] Biotin, or a derivative thereof, may also be used as a label on a nucleotide and/or an oligonucleotide sequence, and subsequently bound by a detectably labeled avidin/streptavidin derivative (e.g. phycoerythrin-conjugated streptavidin), or a detectably labeled anti-biotin antibody. Digoxigenin may be incorporated as a label and subsequently bound by a detectably labeled anti-digoxigenin antibody (e.g. fluoresceinated anti-digoxigenin). An aminoallyl-dUTP residue may be incorporated into an oligonucleotide sequence and subsequently coupled to an N-hydroxy succinimide (NHS) derivatized fluorescent dye. In general, any member of a conjugate pair may be incorporated into a detection oligonucleotide provided that a detectably labeled conjugate partner can be bound to permit detection. As used herein, the term antibody refers to an antibody molecule of any class, or any sub-fragment thereof, such as an Fab.
- [041] Other suitable labels for an oligonucleotide sequence may include fluorescein (FAM), digoxigenin, dinitrophenol (DNP), dansyl, biotin, bromodeoxyuridine (BrdU), hexahistidine (6xHis), phosphor-amino acids (e.g. P-tyr, P-ser, P-thr) and the like. In one embodiment the following hapten/antibody pairs are used for detection, in which each of the antibodies is derivatized with a detectable label: biotin/ α -biotin,

digoxigenin/ α -digoxigenin, dinitrophenol (DNP)/ α -DNP, 5-Carboxyfluorescein (FAM)/ α -FAM.

- [042] In certain exemplary embodiments, a nucleotide and/or an oligonucleotide sequence can be indirectly labeled, especially with a hapten that is then bound by a capture agent, e.g., as disclosed in Holtke et al., U.S. Patent Nos. 5,344,757; 5,702,888; and 5,354,657; Huber et al., U.S. Patent No. 5,198,537; Miyoshi, U.S. Patent No. 4,849,336; Misiura and Gait, PCT publication WO 91/17160; and the like. Many different hapten-capture agent pairs are available for use. Exemplary haptens include, but are not limited to, biotin, des-biotin and other derivatives, dinitrophenol, dansyl, fluorescein, CY5, digoxigenin and the like. For biotin, a capture agent may be avidin, streptavidin, or antibodies. Antibodies may be used as capture agents for the other haptens (many dye-antibody pairs being commercially available, e.g., Molecular Probes, Eugene, OR).
- [043] In certain exemplary embodiments, a first (e.g., probe) oligonucleotide sequence is annealed to a second (e.g., reference) oligonucleotide sequence. The terms “annealing” and “hybridization,” as used herein, are used interchangeably to mean the formation of a stable duplex. In one aspect, stable duplex means that a duplex structure is not destroyed by a stringent wash, e.g., conditions including temperature of about 5 °C less than the T_m of a strand of the duplex and low monovalent salt concentration, e.g., less than 0.2 M, or less than 0.1 M. The term “perfectly matched,” when used in reference to a duplex means that the polynucleotide and/or oligonucleotide strands making up the duplex form a double stranded structure with one another such that every nucleotide in each strand undergoes Watson-Crick base pairing with a nucleotide in the other strand. The term “duplex” includes, but is not limited to, the pairing of nucleoside analogs, such as deoxyinosine, nucleosides with 2-aminopurine bases, PNAs, and the like, that may be employed. A “mismatch” in a duplex between two oligonucleotides means that a pair of nucleotides in the duplex fails to undergo Watson-Crick bonding.
- [044] The term “hybridization,” as used herein, refers to the process in which two single-stranded oligonucleotides bind non-covalently to form a stable double-stranded

oligonucleotide. The term “hybridization” may also refer to triple-stranded hybridization. The resulting (usually) double-stranded polynucleotide is a “hybrid” or “duplex.”

- [045] As used herein, the term “hybridization conditions,” will typically include salt concentrations of less than about 1 M, more usually less than about 500 mM and even more usually less than about 200 mM. Hybridization temperatures can be as low as 5 °C, but are typically greater than 22 °C, more typically greater than about 30 °C, and often in excess of about 37 °C. Hybridizations are usually performed under stringent conditions, i.e., conditions under which a probe will specifically hybridize to its target subsequence. Stringent conditions are sequence-dependent and are different in different circumstances. Longer fragments may require higher hybridization temperatures for specific hybridization. As other factors may affect the stringency of hybridization, including base composition and length of the complementary strands, presence of organic solvents and extent of base mismatching, the combination of parameters is more important than the absolute measure of any one alone.
- [046] Generally, stringent conditions are selected to be about 5 °C lower than the T_m for the specific sequence at a defined ionic strength and pH. Exemplary stringent conditions include salt concentration of at least 0.01 M to no more than 1 M Na ion concentration (or other salts) at a pH 7.0 to 8.3 and a temperature of at least 25 °C. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM Na phosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30 °C are suitable for allele-specific probe hybridizations. For stringent conditions, see for example, Sambrook, Fritsche and Maniatis, *Molecular Cloning A Laboratory Manual*, 2nd Ed. Cold Spring Harbor Press (1989) and Anderson *Nucleic Acid Hybridization*, 1st Ed., BIOS Scientific Publishers Limited (1999). As used herein, the terms “hybridizing specifically to” or “specifically hybridizing to” or similar terms refer to the binding, duplexing, or hybridizing of a molecule substantially to a particular nucleotide sequence or sequences under stringent conditions.
- [047] As used herein, the term “hybridization-based assay” is intended to refer to an assay that relies on the formation of a stable complex as the result of a specific binding

event. In one aspect, a hybridization-based assay means any assay that relies on the formation of a stable duplex or triplex between a probe and a target nucleotide sequence for detecting or measuring such a sequence. A “probe” in reference to a hybridization-based assay refers to an oligonucleotide sequence that has a sequence that is capable of forming a stable hybrid (or triplex) with its complement in a target nucleic acid and that is capable of being detected, either directly or indirectly.

[048] The following examples are set forth as being representative of the present invention. These examples are not to be construed as limiting the scope of the invention as these and other equivalent embodiments will be apparent in view of the present disclosure, figures, tables, and accompanying claims. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference in their entirety for all purposes.

EXAMPLE I

Four Classes of Fast, Accurate Chemical Reactions

[049] Four classes of fast and accurate chemical reactions have been identified: 1) cyanogen bromide-based reactions; 2) carbodiimide-based reactions; 3) amidate-based reactions; and 4) phosphodiester mimetic-based reactions. These classes are suitable for use in a variety of methods that can utilize ligations including, but not limited to, sequencing by chemical ligation, synthesis of oligonucleotide sequences and assembly of oligonucleotide sequences.

[050] Cyanogen bromide-based reactions (Figure 1) and carbodiimide-based reactions (Figure 2) were used to perform sequencing by ligation reactions. The reactions proceeded in a very fast manner (i.e., on the order of a few seconds). The reaction rate was limited only by hybridization kinetics in a complex mixture of 6-mer to 9-mer oligonucleotide sequences. Cyanogen bromide-based and carbodiimide-based chemical ligation reactions are depicted in Figure 4. Amidate-based reactions and compounds useful in such reactions are described in Figure 3. Reaction schemes are depicted in Figure 6.

EXAMPLE II

Phosphodiester Mimetics

- [051] Phosphodiester mimetics, e.g. aldehydes generated by periodate, Schiff's base intermediates formed on reaction with primary amines and the like, are depicted in Figure 7. The chemical ligation reactions have a broader range of substrates so that the variation in binding as a function of template, anchor and probe can be altered, for example, as depicted in Figure 8. Certain mimetics may also be suitable for protein-ligase or polymerase reactions.
- [052] The probe can substitute one or more T to S²T, which binds better to A (higher T_m) on the template strand. Similarly, one or more A's in the probe can be replaced by diaminopurine (DAP) which binds better to T on the template strand. Due to its bulk and lowered hydrogen-bonding ability, the S²T is also better at discriminating against G or DAP in the probe molecules or G in the template molecules (James et al. (1997) *Chem. Biol.* 4:595).

EXAMPLE III

Preactivated Phosphates

- [053] In certain exemplary embodiments, preactivated phosphates are used to accelerate ligation reactions. Preactivated phosphates can be present on, for example, one or more of the 3' and/or 5' ends of an oligonucleotide sequence. If the oligonucleotide sequence has preactivated phosphates at both ends, this permits a second round of ligation (e.g., sequencing by chemical ligation) without need for a potentially inefficient deprotection step. The second round probe would have a non-phosphate OH, SH or NH₂ moiety at its 5' end if the original anchor is on the 5' end, and/or a non-phosphate OH, SH or NH₂ moiety at its 3' end of the second probe if the original anchor is to the 3' end of the template region to be sequenced. Without intending to be bound by scientific theory, an oligonucleotide sequence having two non-phosphate ends could serve this purpose as well. Figures 5A-5B schematically depict chemical ligation using preactivated phosphates. Ligation using oligonucleotide sequences having one or more preactivated phosphates is applicable for use with any combination of the four classes of fast and accurate chemical reactions described

herein: 1) cyanogen bromide-based reactions; 2) carbodiimide-based reactions; 3) amidate-based reactions; and 4) phosphodiester mimetic-based reactions.

[054] It is to be understood that the embodiments of the present invention which have been described are merely illustrative of some of the applications of the principles of the present invention. Numerous modifications may be made by those skilled in the art based upon the teachings presented herein without departing from the true spirit and scope of the invention.

What is claimed is:

1. A method of determining a nucleotide sequence of a reference oligonucleotide comprising the steps of:
providing a reference oligonucleotide having an anchor primer bound thereto;
providing a first probe oligonucleotide having a preactivated chemical ligation compound attached thereto;
allowing the first probe oligonucleotide to hybridize to the reference oligonucleotide;
and
allowing the preactivated chemical ligation compound to mediate ligation between the hybridized first probe oligonucleotide and the anchor primer.
2. The method of claim 1, wherein the preactivated chemical ligation compound is selected from the group consisting of a cyanogen bromide-based compound, a carbodiimide-based compound, an amidate-based compound, and a phosphodiester mimetic.
3. The method of claim 2, wherein the cyanogen bromide-based compound is a phosphocyanate.
4. The method of claim 3, wherein the phosphocyanate is 2-(N-Morpholino)Ethane Sulfonic Acid-cyanogen.
5. The method of claim 2, wherein the carbodiimide-based compound is an O-phospho-isocarbamide compound.
6. The method of claim 5, wherein the O-phospho-isocarbamide compound is phospho-1-Ethyl-3-[3-Dimethylaminopropyl]Carbodiimide.
7. The method of claim 2, wherein the amidate-based compound is a phosphoramidate.

8. The method of claim 7, wherein the phosphoramidate includes an R group selected from the group consisting of pyridine, aniline, imidazole, ethanolamine, azaoxybenzotriazolidine, N-oxysuccinimide and succinimide.

9. The method of claim 2, wherein the phosphodiester mimetic is an imine.

10. The method of claim 1, wherein the first probe oligonucleotide further comprises one or more S²T molecules.

11. The method of claim 1, wherein one or more thymines of the first probe oligonucleotide are replaced by one or more S²T molecules.

12. The method of claim 1, wherein the first probe oligonucleotide further comprises one or more diaminopurines.

13. The method of claim 1, wherein one or more adenines of the first probe oligonucleotide are replaced by one or more diaminopurines.

14. The method of claim 1, wherein the first probe oligonucleotide has two preactivated chemical ligation compounds attached thereto.

15. The method of claim 1, further comprising the steps of:
providing a second probe oligonucleotide;
allowing the second probe oligonucleotide to hybridize to the reference oligonucleotide; and
allowing the preactivated chemical ligation compound to mediate ligation between the hybridized second probe oligonucleotide and the hybridized first probe oligonucleotide.

16. The method of claim 15, wherein the second probe oligonucleotide has a preactivated chemical ligation compound attached thereto.

17. A method of assembling a polynucleotide sequence comprising the steps of:

providing a plurality of overlapping oligonucleotides, at least a portion of which have a preactivated chemical ligation compound attached thereto;

allowing hybridization of at least a portion of the overlapping oligonucleotides to each other; and

allowing the preactivated chemical ligation compounds to mediate ligation between adjacent 5' and 3' ends of hybridized oligonucleotides to assemble a polynucleotide sequence.

18. The method of claim 17, wherein the polynucleotide sequence is DNA.

19. The method of claim 18, wherein the DNA is selected from the group consisting of a vector, a gene, a gene fragment, an exon, an intron and an intergenic DNA sequence.

20. A method of synthesizing a polynucleotide sequence comprising the steps of:
providing a first oligonucleotide having a preactivated chemical ligation compound attached thereto;

providing a second oligonucleotide; and

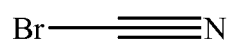
allowing the preactivated chemical ligation compound to mediate ligation between the first oligonucleotide and the second oligonucleotide.

21. The method of claim 20, further comprising the step of allowing the first oligonucleotide to bind a reference oligonucleotide.

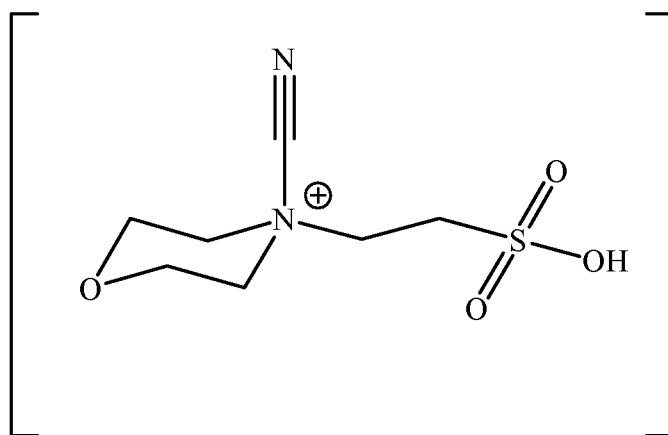
22. The method of claim 21, wherein the reference oligonucleotide is attached to a substrate.

1/8

1A



1B



1C

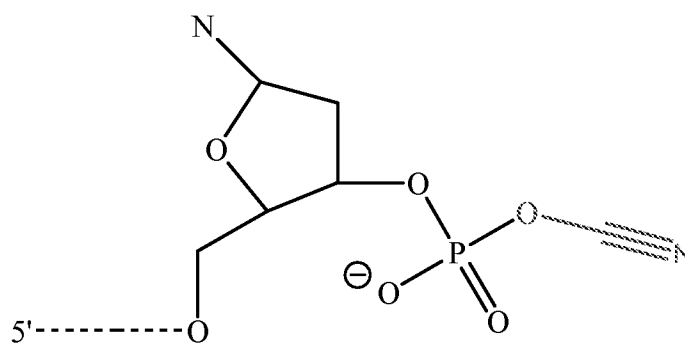
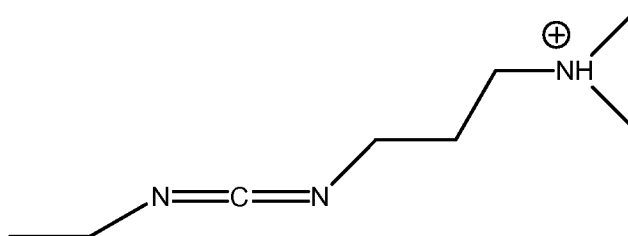


Figure 1

2/8

2A



2B

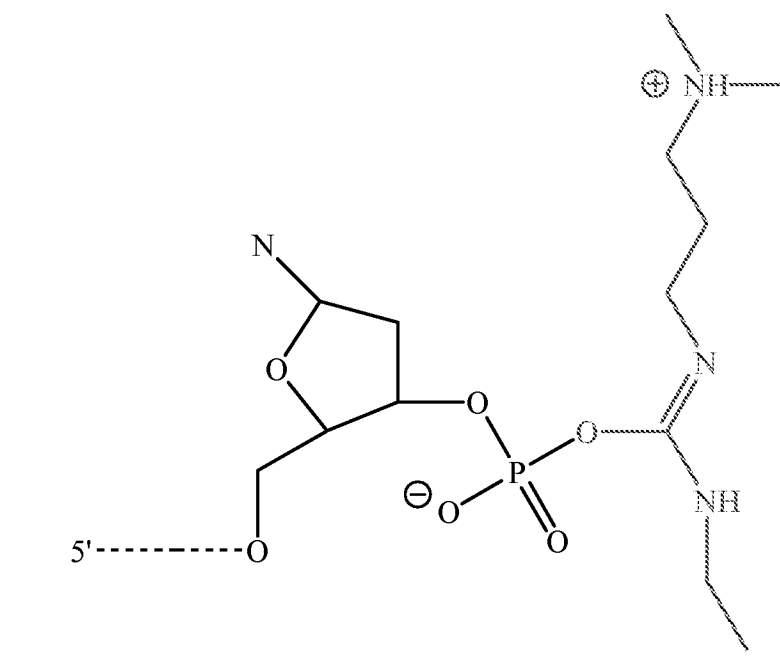
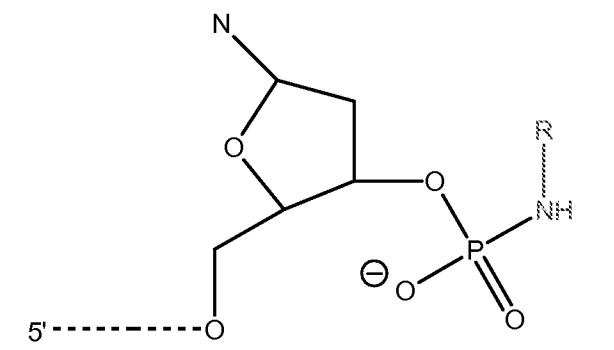


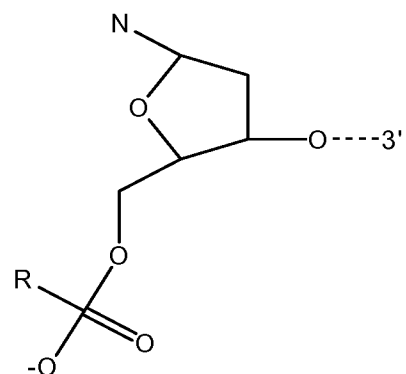
Figure 2

3/8

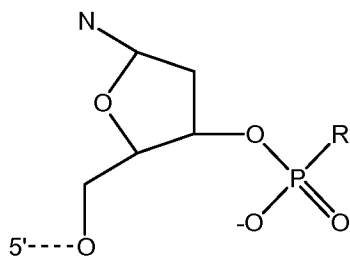
3A



3B



3C



3D

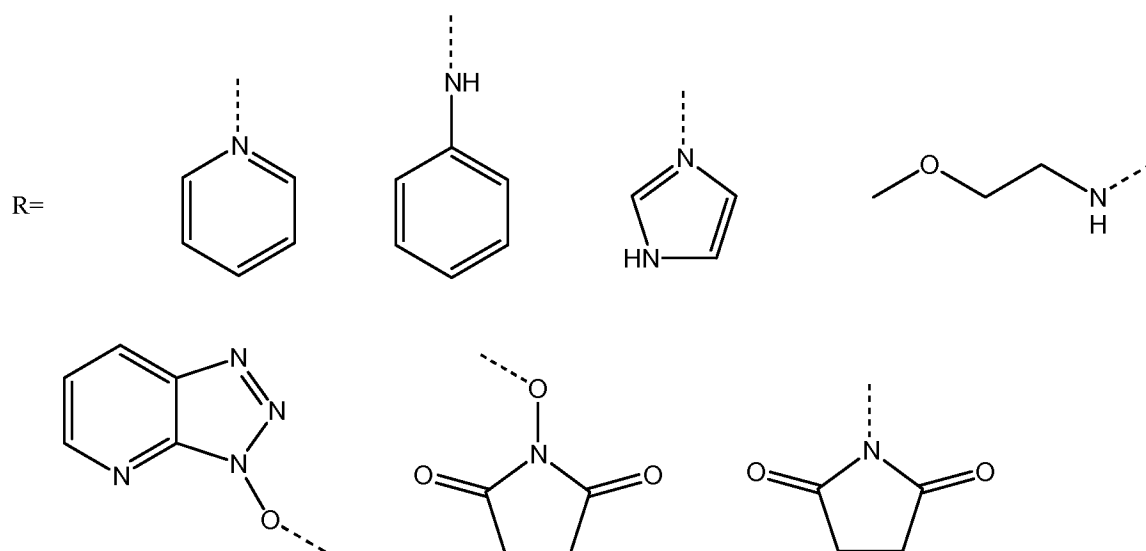
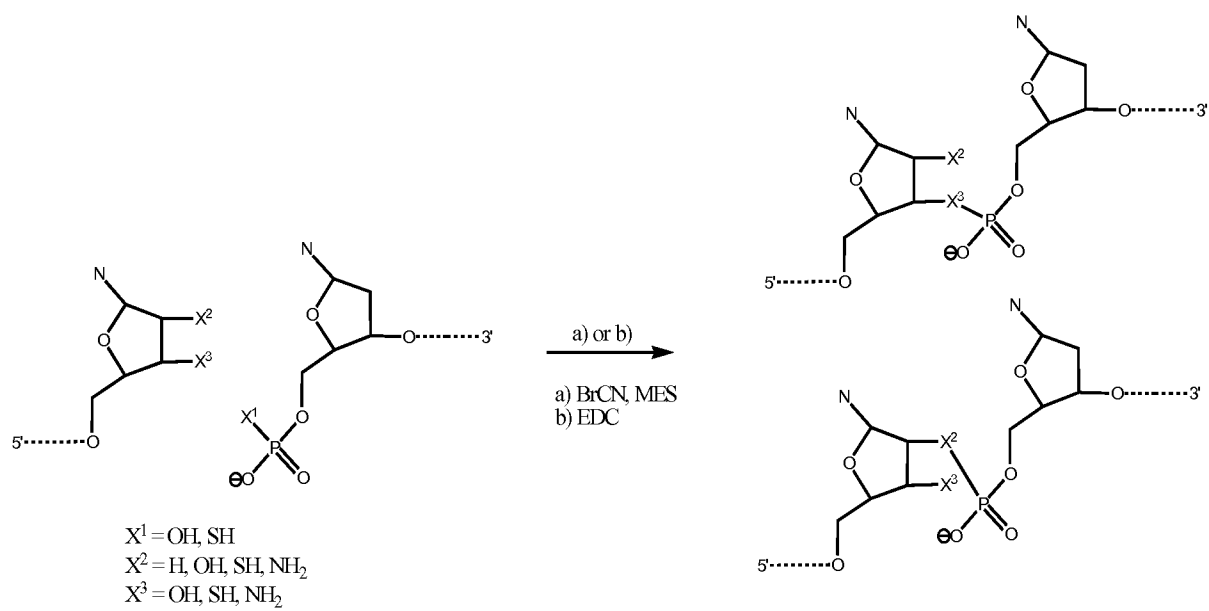


Figure 3

4/8

4A



4B

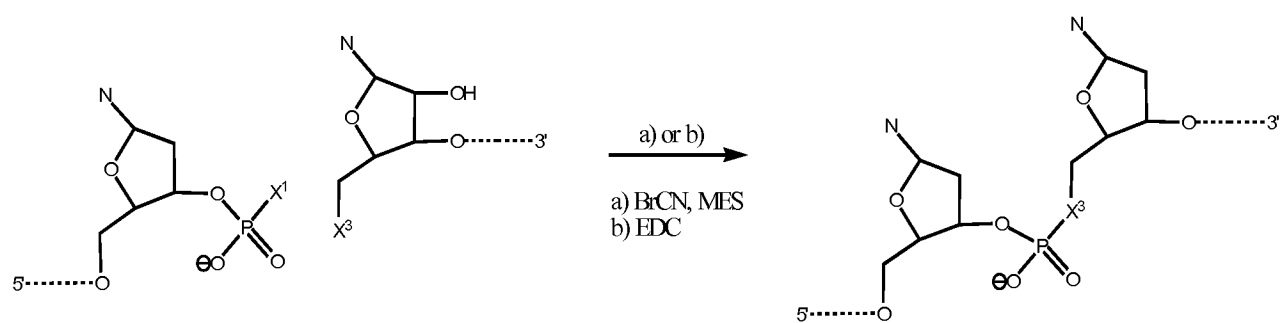
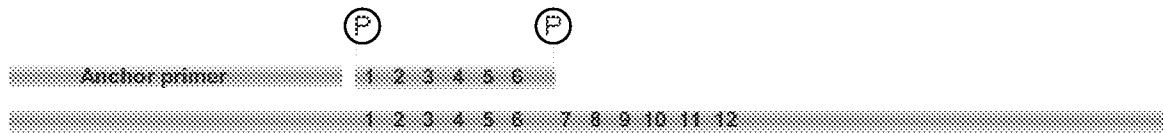


Figure 4

5/8

5A



5B

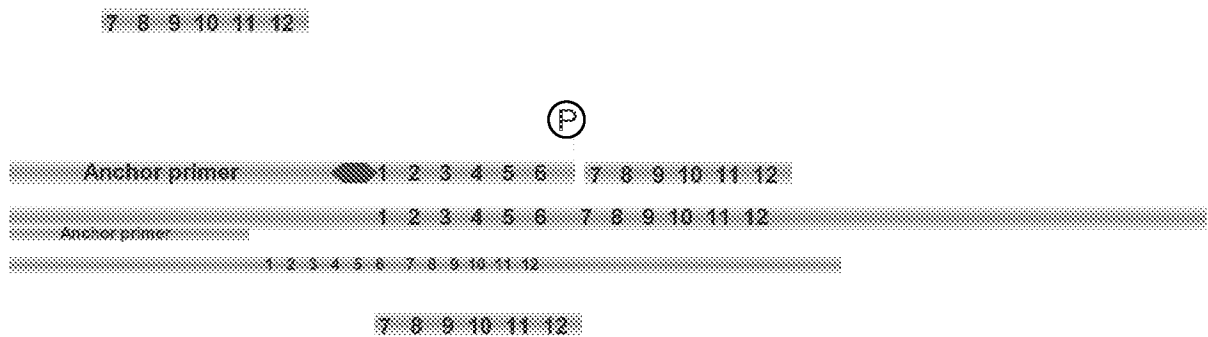
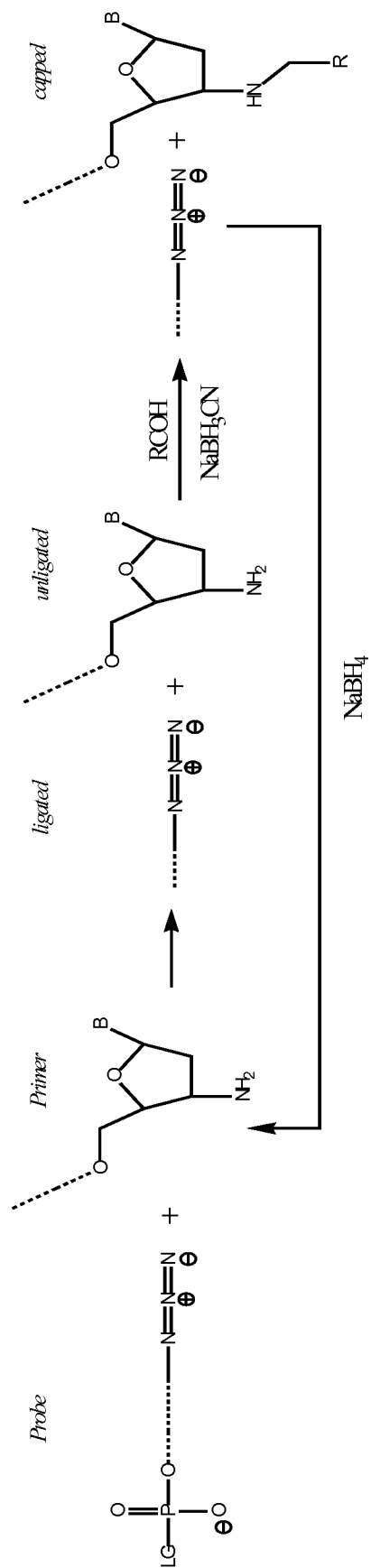


Figure 5

6/8

6A



6B

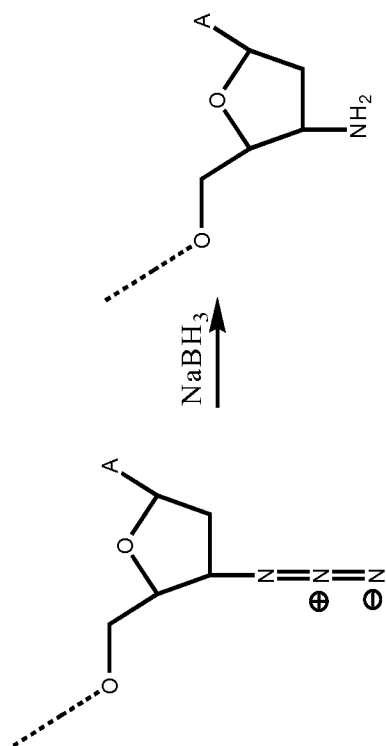
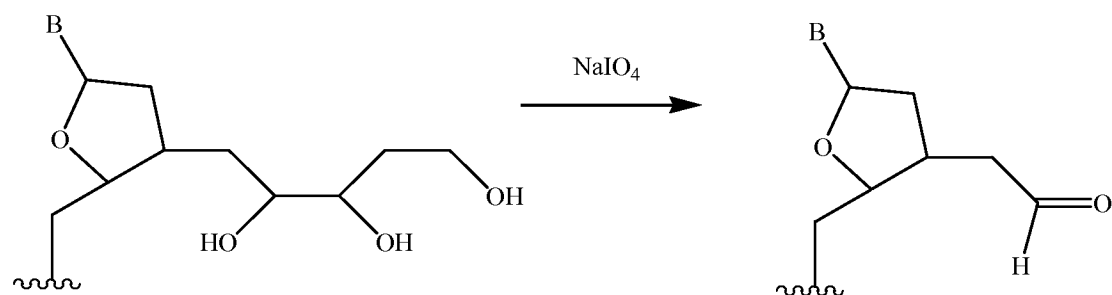


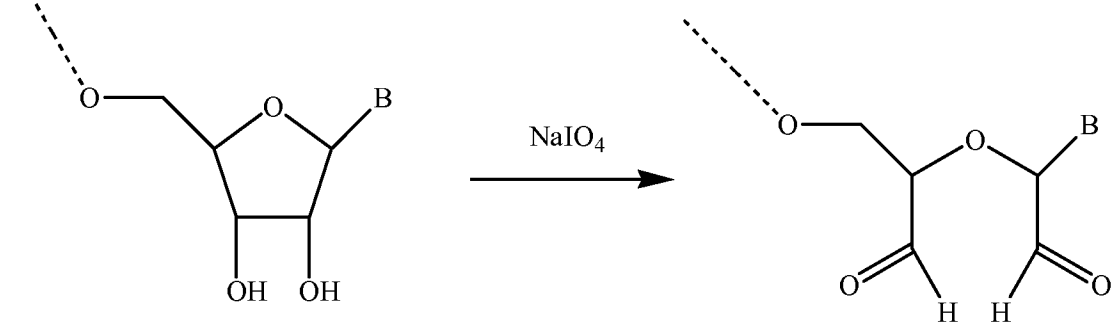
Figure 6

7/8

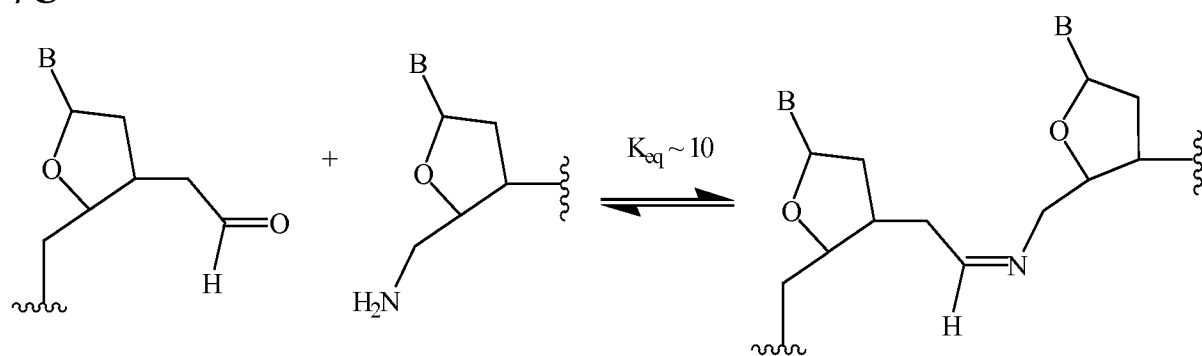
7A



7B



7C



7D

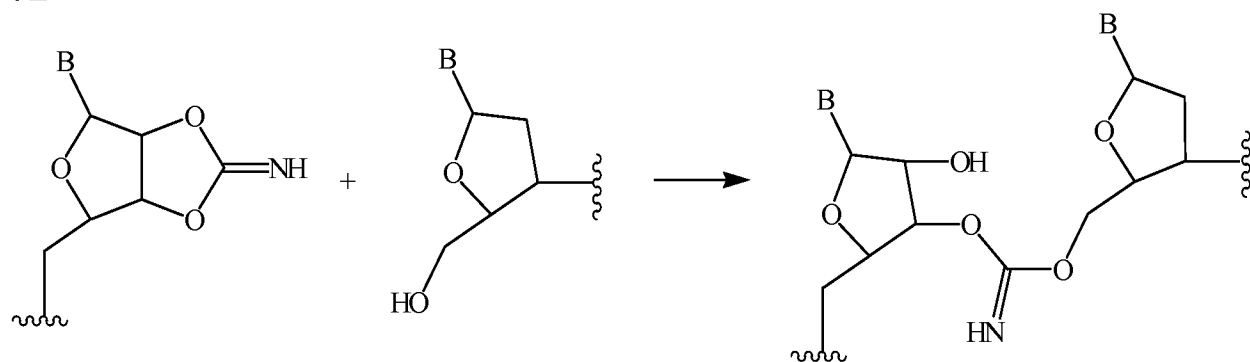
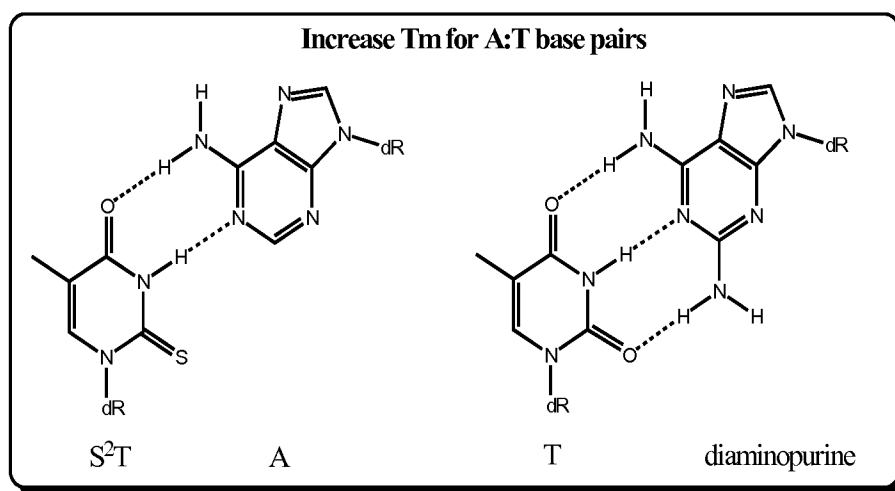


Figure 7

8A



8B

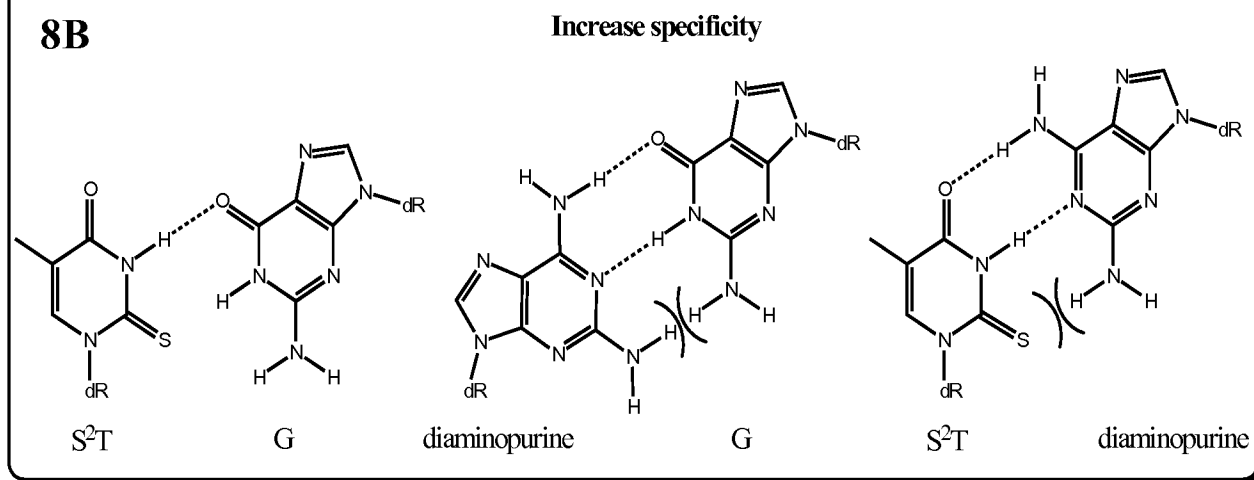


Figure 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 08/65523

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C07H 21/00; G01N 33/00 (2008.04)

USPC - 436/94, 536/23.1, 24.3

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

USPC: 436/94, 536/23.1, 24.3

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

USPC: 436/94, 536/23.1, 24.3

(text search)

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

PubWEST(PGPB,USPT,USOC,EPAB,JPAB); Google; PubMed

Search terms: oligonucleotide, chemical ligation, sequencing, MES, carbodiimide, phosphoraidate, CNBr, imine, crosslinker

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SHENDURE, J. et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 9 September 2005 (09.09.2005), Vol. 309, No. 5741, pages 1728-1732; pg 1729, para 5; pg 1730, para 4; pg 1731, para 5; Fig. S7	1-22
Y	SIDDELL, S.G. et al. RNA hybridization to DNA coupled with cyanogen-bromide-activated sephadex. The purification of polyoma messenger RNA. Eur. J. Biochem. December 1978 (12.1978), Vol. 92, No. 2, pages 621-629; pg 622, para 5	1-22
Y	PALL, G.S. et al. Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. Nucleic Acids. Res. 2 April 2007 (02.04.2007), Vol. 35, No. 8, article e60, pages 1-9; pg 2, para 3	5-6
Y	BARRE, F.X. et al. Unambiguous demonstration of triple-helix-directed gene modification. Proc. Natl. Acad. Sci. USA. 28 March 2000 (28.03.2000), Vol. 97, No. 7, pages 3084-3088; abstract; pg 3084, para 4	7-8
Y	US 4,996,142 A (AL-HAKIM, A.H. et al.) 26 February 1991 (26.02.1991) col 1, ln 59-65; col 2, ln 21-24	9
Y	US 2003/0207271 A1 (HOLWITT et al.) 6 November 2003 (06.11.2003) ttitle; para [0097]; Table 1	10-13



Further documents are listed in the continuation of Box C.



* Special categories of cited documents:

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

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

4 August 2008 (04.08.2008)

Date of mailing of the international search report

08 AUG 2008

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774