(19) DANMARK

(10) DK/EP 2298787 T3

(12) Oversættelse af europæisk patentskrift

Patent- og Varemærkestyrelsen

(51) Int.Cl.: *C 07 H 21/04 (2006.01)* *C 12 P 19/34 (2006.01)* *C 12 Q 1/68 (2006.01)*

(45) Oversættelsen bekendtgjort den: **2016-01-11**

(80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2015-10-28**

(86) Europæisk ansøgning nr.: **10186062.5**

(86) Europæisk indleveringsdag: **1996-04-16**

(87) Den europæiske ansøgnings publiceringsdag: **2011-03-23**

(30) Prioritet: **1995-04-17 US 424663**

(62) Stamansøgningsnr: **96912800.8**

(84) Designerede stater: **BE CH DE DK FR GB IT LI NL SE**

(73) Patenthaver: **Illumina, Inc., 5200 Ilumina Way, San Diego, CA 92122, USA**

(72) Opfinder: **Macevicz, Stephen, 21890 Rucker Drive, Cupertino, CA 95014, USA**

(74) Fuldmægtig i Danmark: **Holme Patent A/S, Valbygårdsvej 33, 2500 Valby, Danmark**

(54) Benævnelse: **DNA-sekventering ved hjælp af parallelle oligonucleotidforlængelser**

(56) Fremdragne publikationer:
WO-A-93/05183
WO-A-94/01582
WO-A-94/11530
WO-A-95/04160
US-A- 5 011 769
US-A- 5 380 833
US-A- 5 403 708

DESCRIPTION

Field of the Invention

[0001] The invention relates generally to methods for determining the nucleotide sequence of a polynucleotide, and more particularly, to a method of identifying nucleotides in a template by stepwise extension of one or more primers by successive ligations of oligonucleotide blocks.

Background

[0002] Analysis of polynucleotides with currently available techniques provides a spectrum of information ranging from the confirmation that a test polynucleotide is the same or different than a standard or an isolated fragment to the express identification and ordering of each nucleoside of the test polynucleotide. Not only are such techniques crucial for understanding the function and control of genes and for applying many of the basic techniques of molecular biology, but they have also become increasingly important as tools in genomic analysis and a great many non-research applications, such as genetic identification, forensic analysis, genetic counselling, medical diagnostics, and the like. In these latter applications both techniques providing partial sequence information, such as fingerprinting and sequence comparisons, and techniques providing full sequence determination have been employed, e.g. Gibbs et al, Proc. Natl. Acad. Sci., 86: 1919-1923 (1989); Gyllensten et al, Proc. Natl. Acad. Sci, 85: 7652-7656 (1988); Carrano et al, Genomics, 4:129-136 (1989); Caetano-Anolles et al, Mol. Gen. Genet., 235: 157-165 (1992); Brenner and Livak, Proc. Natl. Acad. Sci., 86: 8902-8906 (1989); Green et al, PCR Methods and Applications, 1: 77-90 (1991); and Versalovic et al, Nucleic Acids Research, 19: 6823-6831(1991).

[0003] Native DNA consists of two linear polymers, or strands of nucleotides. Each strand is a chain of nucleosides linked by phosphodiester bonds. The two strands are held together in an antiparallel orientation by hydrogen bonds between complementary bases of the nucleotides of the two strands: deoxyadenosine (A) pairs with thymidine (T) and deoxyguanosine (G) pairs with deoxycytidine (C).

[0004] Presently there are two basic approaches to DNA sequence determination: the dideoxy chain termination method, e.g. Sanger et al, Proc. Natl. Acad. Sci., 74: 5463-5467 (1977); and the chemical degradation method, e.g. Maxam et al, Proc. Natl. Acad. Sci., 74: 560-564 (1977). The chain termination method has been improved in several ways, and serves as the basis for all currently available automated DNA sequencing machines, e.g. Sanger et al, J. Mol. Biol., 143:161-178 (1980); Schreier et al, J. Mol. Biol., 129: 169-172 (1979); Smith et al, Nucleic Acids Research, 13: 2399-2412 (1985); Smith et al, Nature, 321: 674-679 (1987); Prober et al, Science, 238: 336-341 (1987); Section II, Meth. Enzymol., 155: 51-334 (1987); Church et al, Science, 240: 185-188 (1988); Hunkapiller et al, Science, 254: 59-67 (1991); Bevan et al, PCR Methods and Applications, 1: 222-228 (1992).

[0005] Both the chain termination and chemical degradation methods require the generation of one or more sets of labeled DNA fragments, each having a common origin and each terminating with a known base. The set or sets of fragments must then be separated by size to obtain sequence information. In both methods, the DNA fragments are separated by high resolution gel electrophoresis, which must have the capacity of distinguishing very large fragments differing in size by no more than a single nucleotide. Unfortunately, this step severely limits the size of the DNA chain that can be sequenced at one time. Sequencing using these techniques can reliably accommodate a DNA chain of up to about 400-450 nucleotides, Bankier et al, Meth. Enzymol., 155: 51-93 (1987); and Hawkins et al, Electrophoresis, 13: 552-559 (1992).

[0006] Several significant technical problems have seriously impeded the application of such techniques to the sequencing of long target polynucleotides, e.g. in excess of 500-600 nucleotides, or to the sequencing of high volumes of many target polynucleotides. Such problems include i) the gel electrophoretic separation step which is labor intensive, is difficult to automate, and introduces an extra degree of variability in the analysis of data, e.g. band broadening due to temperature effects, compressions due to secondary structure in the DNA sequencing fragments, inhomogeneities in the separation gel, and the like; ii) nucleic acid polymerases whose properties, such as processivity, fidelity, rate of polymerization, rate of incorporation of chain terminators, and the like, are often sequence dependent; iii) detection and analysis of DNA sequencing fragments which are typically present in fmol quantities in spacially overlapping bands in a gel; iv) lower signals because the labelling moiety is distributed over the many hundred spacially separated bands rather than being concentrated in a single homogeneous phase, and v) in the case of single-lane fluorescence detection, the availability of dyes with suitable emission and absorption properties, quantum yield, and spectral resolvability, e.g. Trainor, Anal. Biochem., 62: 418-426 (1990); Connell et al, Biotechniques, 5: 342-348 (1987); Karger et al, Nucleic Acids Research, 19: 4955-4962 (1991); Fung et al, U.S. patent 4,855,225; and Nishikawa et al,

Electrophoresis, 12: 623-631 (1991) WO0305183 discloses a method of sequencing by nucleotide extension.

[0007] Another problem exists with current technology in the area of diagnostic sequencing. An ever widening array of disorders, susceptibilities to disorders, prognoses of disease conditions, and the like, have been correlated with the presence of particular DNA sequences, or the degree of variation (or mutation) in DNA sequences, at one or more genetic loci. Examples of such phenomena include human leukocyte antigen (HLA) typing, cystic fibrosis, tumor progression and heterogeneity, p53 proto-oncogene mutations, ras proto-oncogene mutations, and the like, e.g. Gyllensten et al, PCR Methods and Applications, 1: 91-98 (1991); Santamaria et al, International application WO1992015711; Tsui et al, International application WO1991002796. A difficulty in determining DNA sequences associated with such conditions to obtain diagnostic or prognostic information is the frequent presence of multiple subpopulations of DNA, e.g. allelic variants, multiple mutant forms, and the like. Distinguishing the presence and identity of multiple sequences with current sequencing technology is virtually impossible, without additional work to isolate and perhaps clone the separate species of DNA.

[0008] A major advance in sequencing technology could be made if an alternative approach was available for sequencing DNA that did not required high resolution electrophoretic separations of DNA fragments, that generated signals more amenable to analysis, and that provided a means for readily analyzing DNA from heterozygous genetic loci.

[0009] An objective of the invention is to provide such an alternative approach to presently available DNA sequencing technologies.

Summary of the Invention

[0010] The invention provides a method of nucleic acid sequence analysis based on duplex extension along a single stranded template. Preferably, such extension starts from a duplex formed between an initializing oligonucleotide and the template. The initializing oligonucleotide is extended in an initial extension cycle by ligating an oligonucleotide probe to its end to form an extended duplex. The identity of one or more nucleotides in the template is determined by a fluorescent label on, or associated with, a successfully ligated oligonucleotide probe. Preferably, the oligonucleotide probe has a blocking moiety, e.g. a chain-terminating nucleotide, in a terminal position so that only a single extension of the extended duplex takes place in a single cycle. The blocking can optionally be remove to regenerate an extendable end.

[0011] In one aspect of the invention, a plurality of different initializing oligonucleotides is provided for separate samples of the template. Each initializing oligonucleotide forms a duplex with the template such that the end undergoing extension is one or more nucleotides out of register, or phase, with that of every other initializing oligonucleotide of the plurality. In other words, the starting nucleotide for extension is different by one or more nucleotides for each of the different initializing oligonucleotides. In this manner, after each cycle of extension with oligonucleotide probes of the same length, the same relative phase exists between the ends of the initializing oligonucleotides on the different templates. Thus, in a preferred embodiment, where, for example, i) the initializing oligonucleotides are out of phase by one nucleotide, ii) 9-mer oligonucleotide probes are used in the extension step, and iii) nine different initializing oligonucleotides are employed, nine template nucleotides will be identified simultaneously in each extension cycle.

Brief Description of the Drawings

[0012]

Figure 1 diagrammatically illustrates parallel extensions of multiple templates in accordance with the invention.

Figure 2 diagrammatically illustrates an embodiment of the invention employing acid-labile linkages.

Figure 3A diagrammatically illustrates an embodiment of the invention employing RNase H labile oligonucleotides with 3'-> 5' extensions.

Figure 3B diagrammatically illustrates an embodiment of the invention employing RNase H labile oligonucleotides with 5'-> 3' extensions.

Figure 4 diagrammatically illustrates an embodiment of the invention employing ligation followed by polymerase extension and cleavage.

Definitions

[0013] As used herein "sequence determination," "determining a nucleotide sequence," "sequencing," and like terms, in reference to polynucleotides includes determination of partial as well as full sequence information of the polynucleotide. That is, the term includes sequence comparisons, fingerprinting, and like levels of information about a target polynucleotide, as well as the express identification and ordering of each nucleoside of the test polynucleotide.

[0014] "Perfectly matched duplex" in reference to the protruding strands of probes and target polynucleotides means that the protruding strand from one forms a double stranded structure with the other such that each nucleotide in the double stranded structure undergoes Watson-Crick basepairing with a nucleotide on the opposite strand. The term also comprehends the pairing of nucleoside analogs, such as deoxyinosine, nucleosides with 2-aminopurine bases, and the like, that may be employed to reduce the degeneracy of the probes.

[0015] The term "oligonucleotide" as used herein includes linear oligomers of nucleosides or analogs thereof, including deoxyribonucleosides, ribonucleosides, and the like. Usually oligonucleotides range in size from a few monomeric units, e.g. 3-4, to several hundreds of monomeric units. Whenever an oligonucleotide is represented by a sequence of letters, such as "ATGCCTG," it will be understood that the nucleotides are in 5'- > 3' order from left to right and that "A" denotes deoxyadenosine, "C" denotes deoxycytidine, "G" denotes deoxyguanosine, and "T" denotes thymidine, unless otherwise noted.

[0016] As used herein, "nucleoside" includes the natural nucleosides, including 2'-deoxy and 2'-hydroxyl forms, e.g. as described in Kornberg and Baker, DNA Replication, 2nd Ed. (Freeman, San Francisco, 1992). "Analogs" in reference to nucleosides includes synthetic nucleosides having modified base moieties and/or modified sugar moieties, e.g. described generally by Scheit, Nucleotide Analogs (John Wiley, New York, 1980). Such analogs include synthetic nucleosides designed to enhance binding properties, reduce degeneracy, or increase specificity.

[0017] As used herein, "ligation" means to form a covalent bond or linkage between the termini of two or more nucleic acids, e.g. oligonucleotides and/or polynucleotides, in a template-driven reaction. The nature of the bond or linkage may vary widely and the ligation may be carried out enzymatically or chemically.

Detailed Description of the Invention

[0018] The invention provides a method of sequencing nucleic acids which obviates electrophoretic separation of similarly sized DNA fragments, and which eliminates the difficulties associated with the detection and analysis of spacially overlapping bands of DNA fragments in a gel or like medium. The invention also obviates the need to generate DNA fragments from long single stranded templates with a DNA polymerase.

[0019] The general scheme of one aspect of the invention is shown diagrammatically in Figure 1. As described more fully below, the invention is not meant to be limited by the particular features of this embodiment. Template (20) comprising a polynucleotide (50) of unknown sequence and binding region (40) is attached to solid phase support (10).

[0020] Preferably, for embodiments employing N-mer probes, the template is divided into N aliquots, and for each aliquot a different initializing oligonucleotide i_K is provided that forms a perfectly matched duplex at a location in binding region (40) different from that of the other initializing oligonucleotides. That is, the initializing oligonucleotides i_1-i_N form a set of duplexes with the template in the binding region (40), such that the ends of the duplexes proximal to the unknown sequence are from 0 to N-1 nucleotides from the start the unknown sequence. Thus, in the first cycle of ligations with N-mer probes, a terminal nucleotide (16) of probe (30) ligated to i_1, in Figure 1 will be complementary to the N-1 nucleotide of binding region (40). Likewise, a terminal nucleotide (17) of probe (30) ligated to i_2 in Figure 1 will be complementary to the N-2 nucleotide of binding region (40); a terminal nucleotide (18) of probe (30) ligated to i_3 in Figure 1 will be complementary to the N-3 nucleotide of binding region (40), and so on. Finally, a terminal nucleotide (15) of probe (30) ligated to i_n in will be complementary to the first nucleotide of unknown sequence (50). In the second cycle of ligations, a terminal nucleotide (19) of probe (31) will be complementary to the second nucleotide (19) of unknown sequence (50) in duplexes starting with initializing oligonucleotide i_1. Likewise, terminal nucleotides of probes ligated to duplexes starting with initializing oligonucleotides i_2, i_3, i_4, and so on, will be complementary to the third, fourth,

and fifth nucleotides of unknown sequence (50).

[0021] In the above embodiment, the oligonucleotide probes are labeled with a fluorescent label so that the identity of the nucleotide abutting the extended duplex can be determined from the label.

[0022] Binding region (40) has a known sequence, but can vary greatly in length and composition. It must be sufficiently long to accommodate the hybridization of an initializing oligonucleotide. Different binding regions can be employed with either identical or different initializing oligonucleotides, but for convenience of preparation, it is preferable to provide identical binding regions and different initializing oligonucleotides. Thus, all the templates are prepared identically and then separated into aliquots for use with different initializing oligonucleotides. Preferably, the binding region should be long enough to accommodate a set of different initializing oligonucleotides, each hybridridizing to the template to produce a different starting point for subsequent ligations. Most preferably, the binding region is between about 20 to 50 nucleotides in length.

[0023] Initializing oligonucleotides are selected to form highly stable duplexes with the binding region that remain intact during any washing steps of the extension cycles. This is conveniently achieved by selecting the length(s) of the initializing oligonucleotides to be considerably longer than that, or those, of the oligonucleotide probes and/or by selecting them to be GC-rich. Initializing oligonucleotides may also be cross-linked to the template strand by a variety of techniques, e.g. Summerton et al, U.S. patent 4,123,610; or they may be comprised of nucleotide analogs that form duplexes of greater stability than their natural counterparts, e.g. peptide nucleic acids, Science, 254:1497-1500 (1991); Hanvey et al, Science, 258: 1481-1485 (1992); and PCT applications WO1992020702 and WO1992020703.

[0024] Preferably, the length of the initializing oligonucleotide is from about 20 to 30 nucleotides and its composition comprises a sufficient percentage of G's and C's to provide a duplex melting temperature that exceeds those of the oligonucleotide probes being employed by about 10-50°C. More preferably, the duplex melting temperature of the initializing oligonucleotide exceeds those of the oligonucleotide probes by about 20-50°C. The number, N, of distinct initializing oligonucleotides employed in a sequencing operation can vary from one, in the case where a single nucleotide is identified at each cycle, to a plurality whose size is limited only by the size of oligonucleotide probe that can be practically employed. Factors limiting the size of the oligonucleotide probe include the difficulty in preparing mixtures having sufficiently high concentrations of individual probes to drive hybridization reactions at a reasonable rate, the susceptibility of longer probes to forming secondary structures, reduction in sensitivity to single base mismatches. Preferably, N is in the range of from 1 to 16; more preferably, N is in the range of from 1 to 12; and most preferably, N is in the range of from 1 to 8.

[0025] A wide variety of oligonucleotide probes can be used with the invention. Generally, the oligonucleotide probes should be capable of being ligated to an initializing oligonucleotide or extended duplex to generate the extended duplex of the next extension cycle; the ligation should be template-driven in that the probe should form a duplex with the template prior to ligation; the probe should possess a blocking moiety to prevent multiple probe ligations on the same template in a single extension cycle, the probe should be capable of being treated or modified to regenerate an extendable end after ligation, and the probe should possess a signaling moiety that permits the acquisition of sequence information relating to the template after a successful ligation. As described more fully below, depending on the embodiment, the extended duplex or initializing oligonucleotide may be extended in either the 5'- > 3' direction or the 3'- > 5' direction by oligonucleotide probes. Generally, the oligonucleotide probe need not form a perfectly matched duplex with the template, although such binding is usually preferred. In preferred embodiments in which a single nucleotide in the template is identified in each extension cycle, perfect base pairing is only required for identifying that particular nucleotide. For example, in embodiments where the oligonucleotide probe is enzymatically ligated to an extended duplex, perfect base pairing--i.e. proper Watson-Crick base pairing-is required between the terminal nucleotide of the probe which is ligated and its complement in the template. Generally, in such embodiments, the rest of the nucleotides of the probe serve as "spacers" that ensure the next ligation will take place at a predetermined site, or number of bases, along the template. That is, their pairing, or lack thereof, does not provide further sequence information. Likewise, in embodiments that rely on polymerase extension for base identification, the probe primarily serves as a spacer, so specific hybridization to the template is not critical, although it is desirable.

[0026] Preferably, the oligonucleotide probes are applied to templates as mixtures comprising oligonucleotides of all possible sequences of a predetermined length. The complexity of such mixtures can be reduced by a number of methods, including using so-called degeneracy-reducing analogs, such as deoxyinosine, e.g. as taught by Kong Thoo Lin et al, Nucleic Acids Research, 20: 5149-5152; U.S. patent 5,002,867; Nichols et al, Nature, 369: 492-493 (1994); or by separately applying multiple mixtures of oligonucleotide probes, e.g. four mixtures comprising four disjoint subsets of oligonucleotide sequences that taken together would comprise all possible sequences of the predetermined length.

[0027] Initializing oligonucleotides and oligonucleotide probes of the invention are conveniently synthesized on an automated

DNA synthesizer, e.g. an Applied Biosystems, Inc. (Foster City, California) model 392 or 394 DNA/RNA Synthesizer, using standard chemistries, such as phosphoramidite chemistry, e.g. disclosed in the following references: Beaucage and Iyer, Tetrahedron, 48: 2223-2311 (1992); Molko et al, U.S. patent 4,980,460; Koster et al, U.S. patent 4,725,677; Caruthers et al, U.S. patents 4,415,732; 4,458,066; and 4,973,679. Alternative chemistries, e.g. resulting in non-natural backbone groups, such as phosphorothioate, and phosphoramidate, may also be employed provided that the resulting oligonucleotides are compatible with the ligation and other reagents of a particular embodiment. Mixtures of oligonucleotide probes are readily synthesized using well known techniques, e.g. as disclosed in Telenius et al, Genomics, 13: 718-725 (1992); Welsh et al, Nucleic Acids Research, 19: 5275-5279 (1991); Grothues et al, Nucleic Acids Research, 21: 1321-1322 (1993); Hartley, European patent application 90304496.4. Generally, these techniques simply call for the application of mixtures of the activated monomers to the growing oligonucleotide during the coupling steps where one desires to introduce the degeneracy.

[0028] When conventional ligases are employed in the invention, as described more fully below, the 5' end of the probe may be phosphorylated in some embodiments. A 5' monophosphate can be attached to an oligonucleotide either chemically or enzymatically with a kinase, e.g. Sambrook et al, Molecular Cloning: A Laboratory Manual, 2nd Edition (Cold Spring Harbor Laboratory, New York, 1989). Chemical phosphorylation is described by Horn and Urdea, Tetrahedron Lett., 27: 4705 (1986), and reagents for carrying out the disclosed protocols are commercially available, e.g. 5' Phosphate-ON™ from Clontech Laboratories (Palo Alto, California). Preferably, when required, oligonucleotide probes are chemically phosphorylated.

[0029] The probes of the invention can be labeled with a fluorescent moeity, including the direct or indirect attachment of a fluorescent moiety. Many comprehensive reviews of methodologies for labeling DNA and constructing DNA probes provide guidance applicable to constructing probes of the present invention. Such reviews include Matthews et al, Anal. Biochem., Vol 169, pgs. 1-25 (1988); Haugland, Handbook of Fluorescent Probes and Research Chemicals (Molecular Probes, Inc., Eugene, 1992); Keller and Manak, DNA Probes, 2nd Edition (Stockton Press, New York, 1993); and Eckstein, editor, Oligonucleotides and Analogues: A Practical Approach (IRL Press, Oxford, 1991). Many more particular methodologies applicable to the invention are disclosed in the following sample of references: Fung et al, U.S. patent 4,757,141; Hobbs, Jr., et al U.S. patent 5,151,507; Cruickshank, U.S. patent 5,091,519; (synthesis of functionalized oligonucleotides for attachment of reporter groups); Jablonski et al, Nucleic Acids Research, 14: 6115-6128 (1986)(enzyme-oligonucleotide conjugates); and

[0030] Urdea et al, U.S. patent 5,124,246 (branched DNA).

[0031] Preferably, the probes are labeled with one or more fluorescent dyes, e.g. as disclosed by Menchen et al, U.S. patent 5,188,934; Begot et al PCT application WO1991005060.

[0032] Guidance in selecting hybridization conditions for the application of oligonucleotide probes to templates can be found in numerous references, e.g. Wetmur, Critical Reviews in Biochemistry and Molecular Biology, 26: 227-259 (1991); Dove and Davidson, J. Mol. Biol. 5: 467-478 (1962); Hutton, Nucleic Acids Research, 10: 3537-3555 (1977); Breslauer et al, Proc. Natl. Acad. Sci. 83: 3746-3750 (1986); Innis et al, editors, PCR Protocols (Academic Press, New York, 1990).

[0033] Generally, when an oligonucleotide probe anneals to a template in juxtaposition to an end of the extended duplex, the duplex and probe are ligated, i.e. are caused to be covalently linked to one another. Ligation can be accomplished either enzymatically or chemically. Chemical ligation methods are well known in the art, e.g. Ferris et al, Nucleosides & Nucleotides, 8: 407-414 (1989); Shabarova et al, Nucleic Acids Research, 19: 4247-4251 (1991). Preferably, enzymatic ligation is carried out using a ligase in a standard protocol. Many ligases are known and are suitable for use in the invention, e.g. Lehman, Science, 186: 790-797 (1974); Engler et al, DNA Ligases, pages 3-30 in Boyer, editor, The Enzymes, Vol. 15B (Academic Press, New York, 1982). Preferred ligases include T4 DNA ligase, T7 DNA ligase, E. coli DNA ligase, Taq ligase, Pfu ligase, and Tth ligase. Protocols for their use are well known, e.g. Sambrook et al (cited above); Barany, PCR Methods and Applications, 1: 5-16 (1991); Marsh et al, Strategies, 5: 73-76 (1992). Generally, ligases require that a 5' phosphate group be present for ligation to the 3' hydroxyl of an abutting strand.

Preparing Target Polynucleotides

[0034] Preferably, a target polynucleotide is conjugated to a binding region to form a template, and the template is attached to a solid phase support, such as a magnetic particle, polymeric microsphere, or filter material, which permits the sequential application of reagents without complicated and time-consuming purification steps. The length of the target polynucleotide can vary widely; however, for convenience of preparation, lengths employed in conventional sequencing are preferred. For example, lengths in the range of a few hundred basepairs, 200-300, to 1 to 2 kilobase pairs are preferred.

[0035] The target polynucleotides can be prepared by various conventional methods. For example, target polynucleotides can be prepared as inserts of any of the conventional cloning vectors, including those used in conventional DNA sequencing. Extensive guidance for selecting and using appropriate cloning vectors is found in Sambrook et al, Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor Laboratory, New York, 1989). Sambrook et al and Innis et al, editors, PCR Protocols (Academic Press, New York, 1990) also provide guidance for using polymerase chain reactions to prepare target polynucleotides.

[0036] Preferably, cloned or PCR-amplified target polynucleotides are prepared which permit attachment to magnetic beads, or other solid supports, for ease of separating the target polynucleotide from other reagents used in the method. Protocols for such preparative techniques are described fully in Wahlberg et al, Electrophoresis, 13: 547-551 (1992); Tong et al, Anal. Chem., 64: 2672-2677 (1992); Hultman et al, Nucleic Acids Research, 17: 4937-4946 (1989); Hultman et al, Biotechniques, 10: 84-93 (1991); Syvanen et al, Nucleic Acids Research, 16: 11327-11338 (1988); Dattagupta et al, U.S. patent 4,734,363; Uhlen, PCT application WO1989009282 Kits are also commercially available for practicing such methods, e.g. Dynabeads™ template preparation kit from Dynal AS. (Oslo, Norway). Generally, the size and shape of a microparticle or beads employed in the method of the invention is not critical; however, microparticles in the size range of a few, e.g. 1-2, to several hundred, e.g. 200-1000 m diameter are preferable, as they minimize reagent and sample usage while permitting the generation of readily detectable signals, e.g. from fluorescently labeled probes.

Schemes for Lighting, Capping, and Regenerating Extendable Termini

[0037] In one aspect, the invention calls for repeated steps of ligating and identifying of oligonucleotide probes. However, since the ligation of multiple probes to the same extended duplex in the same step would usually introduce identification problems, it is useful to prevent multiple extensions and to regenerate extendable termini. Moreover, if the ligation step is not 100% efficient, it would be desirable to cap extended duplexes that fail to undergo ligation so that they do not participate in any further ligation steps. That is, a capping step preferably occurs after a ligation step, by analogy with other synthetic chemical processes, such as polynucleotide synthesis, e.g. Andrus et al. US patent 4.816,571. This would remove a potentially significant source of noise from signals generated in subsequent identification steps.

[0038] Below, several exemplary schemes for carrying out ligation, capping, regeneration, and identification steps in accordance with the invention are described. They are presented for purposes of guidance and are not meant to be limiting.

[0039] A scheme for extending an initializing oligonucleotide or an extended duplex in the 3'->5' direction is illustrated in Figure 2. Template (20) is attached to solid phase support (10) by its 5' end. This can be conveniently accomplished via a biotin, or like linking moiety, using conventional techniques. Initialising oligonucleotide (200) having a 5' phosphate group is annealed to template (20) as described above prior to the initial cycle of ligation and identification. An oligonucleotide probe (202) of the following form is employed:

HO-(3')BBB ... BBB(5')-OP(=O)(O-)NH-B_t*

where BBB ... BBB represents the sequence of nucleotides of oligonucleotide probe (202) and B_t* is a fluorescently labeled chain-terminating moiety linked to the 5' carbon of the oligonucleotide via a phosphoramidate group, or other labile linkage, such as a photocleavable linkage. The nature of B_t* may vary widely. It can be a fluorescently labeled nucleoside (e.g. coupled via a 5'P3'N phosphoramidate) or other moiety, so long as it prevents successive ligations. It may simply be a fluorescent label connected by a linker, such as described in Agrawal and Tang, International application WO1992008728. An important feature of the oligonucleotide probe is that after annealing and ligation (204), the fluorescent label may be removed and the extendable end regenerated by treating the phosphoramidate linkage with acid, e.g. as taught by Letsinger et al, J. Am. Chem. Soc., 94: 292-293 (1971); Letsinger et al, Biochem., 15: 2810-2816 (1976); Gryaznov et al, Nucleic Acid Research, 20: 3403-3409 (1992). By way of example, hydrolysis of the phosphoramidate may be accomplished by treatment with 0.8% trifluoroacetic acid in dichloromethane for 40 minutes at room temperature. Thus, after annealing, ligating, and identifying the ligated probe via the fluorescent label on B_t*, the chain-terminating moiety is cleaved by acid hydrolysis (206) thereby breaking the phosphorus linkage and leaving a 5' monophosphate on the ligated oligonucleotide. The steps can be repeated (208) in successive cycles. In one aspect of this embodiment, a single initializing oligonucleotide may be employed such that only one nucleotide is identified in each sequencing cycle. For such an embodiment, the above probe preferably has the following form:

HO-(3')B(5')-OP(= O)(O^-)NHBB ... BBB-B_t*

[0040] Thus, after each ligation and acid cleavage steps the duplex will be extended by one nucleotide.

[0041] A capping step may be introduced prior to hydrolysis. For example, probe (202) may have the form:

HO-(3')BB ... Bp^B ... BB(5')-OP(=O)(O)NH-B_t*

where "p^" is a exonuclease resistant linkage, such as phosphorothioate or methylphosphonate. In such an embodiment, capping can be achieved by treating the extended duplexes with an exonuclease, such as λ exonuclease, which will cleave the unligated extended duplexes back to the exonuclease resistant linkage. The presence of this linkage at the 5' end of the extended duplex will then prevent it from participating in subsequent ligations. Clearly, many other capping methodologies may be employed, e.g. acylation or ligation of an inert oligonucleotide. When free 3' hydroxyls are involved, capping may be accomplished by extending the duplex with a DNA polymerase in the presence of chain-terminating nucleoside triphosphates, e.g. dideoxynucleoside triphosphates. The phosphoramidate linkage described above is an example of a general class of internucleosidic linkages referred to herein as "chemically scissile internucleosidic linkages." These are internucleosidic linkages that may be cleaved by treating them with characteristic chemical or physical conditions, such as an oxidizing environment, a reducing environment, or light of a characteristic wavelength (for photolabile linkages). Other examples of chemically scissile internucleosidic linkages which may be used in accordance with the invention are described in Urdea 5,380,833; Gryaznov et al, Nucleic Acids Research, 21: 1403-1408 (1993)(disulfide); Gryaznov et al, Nucleic Acids Research, 22: 2366-2369 (1994) (bromoacetyl); Urdea et al, International application WO1992002528 (photolabile)

[0042] Further chemically scissile linkages that may be employed with the invention include chain-terminating nucleotides that may be chemically converted into an extendable nucleoside. Examples of such compounds are described in the following references: Canard et al, International application WO1994023064; Ansorge, German patent application No. DE 4141178 A1; Metzker et al, Nucleic Acids Research, 22: 4259-4267 (1994); Cheeseman, U.S. patent 5,302,509; Ross et al, International application WO1991006678.

[0043] A scheme for extending an initializing oligonucleotide or an extended duplex in the 5'- > 3' direction is illustrated in Figure 3A. Template (20) is attached to solid phase support (10) by its 3' end. As above, this can be conveniently accomplished via a biotin, or like linking moiety, using conventional techniques. Initializing oligonucleotide (300) having a 3' hydroxyl group is annealed to template (20) as described above prior to the initial cycle of ligation and identification. An oligonucleotide probe (302) of the following form is employed:

OP(=O)(O-)O-(5')BBB... BBBRRRRB$_t$*

where BBB ... BBBRRRR represents the sequence of 2'-deoxynucleotides of oligonucleotide probe (302), "RRRR" represent a sequence of four ribonucleotides of probe (302), and B_t* is a fluorescently labeled chain-terminating moiety, as described above. Such mixed RNA-DNA oligonucleotides are readily synthesized using conventional automated DNA synthesizers, e.g. Duck et al, U.S. patent 5,011,769. RNase H will cleave the probe specifically in the center of the four ribonucleotide segment, Hogrefe et al, J. Biol. Chem., 265: 5561-5566 (1990), leaving a 3' hydroxyl (312) on the extended duplex, which may participate in subsequent ligation steps. Thus, a cycle in the present embodiment proceeds by annealing probe (302) to template (20) and ligating (304) to form extended duplex (306). After identification via B_t*, the extended duplex is treated with RNase H to cleave the fluorescent label and regenerate an extendable end. The cycle is then repeated (314). Capping (310) can be carried out prior to RNase H treatment by extending the unligated ends with a DNA polymerase in the presence of the four dideoxynucleoside triphosphates, ddATP, ddCTP, ddGTP, and ddTTP.

[0044] As illustrated in Figure 3B, a similar scheme can be employed for 3'5' extensions. In such an embodiment, initiating oligonucleotide or extended duplex (330) has a 5' monophosphate and the oligonucleotide probe (332) has the form:

HO-(3')BBB ... BBBRRRRB .. BB$_t$*

[0045] As above, after annealing, ligating (334), and identifying (338), extended duplex (336) is cleaved by RNase H which in this case leaves a 5' monophosphate (342) at the terminus of the extended duplex. With the regenerated extendable end, the cycle can be repeated (344). A capping step can be included prior to RNase H hydrolysis by either ligating an unlabeled non-RNA-containing probe, or by removing any remaining 5' monophosphates by treatment with a phosphatase.

[0046] Identification of nucleotides can be accomplished by polymerase extension following ligation. As exemplified in Figure 4, for this embodiment, template (20) is attached to solid phase support (10) as described above and initializing oligonucleotide (400) having a 3' hydroxyl is annealed to the template prior to the initial cycle. Oligonucleotide probes (402) of the form:

$OP(=O)(O^-)O$-(5')BBB ... BBBRRRRB ... B(3')$OP(=O)(O^-)O$

are annealed to template (20) and ligated (404) to form extended duplex (406). The 3' monophosphate, which prevents successive ligations of probes in the same cycle, is removed with phosphatase (408) to expose a free 3' hydroxyl (410). Clearly, alternative blocking approaches may also be used. Extended duplex (406) is further extended by a nucleic acid polymerase in the presence of fluorescently labeled dideoxynucleoside triphosphates (412), thereby permitting the identification of a nucleotide of template (20) by the fluorescent label of the incorporated dideoxynucleotide. The fluorescently labeled dideoxynucleotide and a portion of probe (402) are then cleaved (414), for example, by RNase H treatment, to regenerate an extendable end on extended duplex (406). The cycle is then repeated (416).

[0047] In order to reduce the number of separate annealing reactions that must be carried out, the oligonucleotide probes may be grouped into mixtures, or subsets, of probes whose perfectly matched duplexes with complementary sequences have similar stability or free energy of binding. Such subsets of oligonucleotide probes having similar duplex stability are referred to herein as "stringency classes" of oligonucleotide probes. The mixtures, or stringency classes, of oligonucleotide probes are then separately combined with the target polynucleotide under conditions such that substantially only oligonucleotide probes complementary to the target polynucleotide form duplexes. That is, the stringency of the hybridization reaction is selected so that substantially only perfectly complementary oligonucleotide probes form duplexes. These perfectly matched duplexes are then ligated to form extended duplexes. For a given oligonucleotide probe length, the number of oligonucleotide probes within each stringency class can vary widely. Selection of oligonucleotide probe length and stringency class size depends on several factors, such as length of target sequence and how it is prepared, the extent to which the hybridization reactions can be automated, the degree to which the stringency of the hybridization reaction can be controlled, and the presence or absence of oligonucleotide probes with complementary sequences. Guidance in selecting an appropriate size of stringency class for a particular embodiment can be found in the general literature on nucleic acid hybridization and polymerase chain reaction methodology, e.g. Gotoh, Adv. Biophys. 16: 1-52 (1983); Wetmer, Critical Reviews in Biochemistry and Molecular Biology 26: 227-259 (1991); Breslauer et al, Proc. Natl. Acad. Sci. 83: 3746-3750 (1986); Wolf et al, Nucleic Acids Research, 15: 2911-2926 (1987); Innis et al, editors, PCR Protocols (Academic Press, New York, 1990); McGraw et al, Biotechniques, 8: 674-678 (1990). Stringency can be controlled by several varying several parameters, including temperature, salt concentration, concentration of certain organic solvents, such as formamide.

[0048] Preferably, temperature is used to define the stringency classes because the activity of the various polymerases or ligases employed limits the degree to which salt concentration or organic solvent concentration can be varied for ensuring specific annealing of the oligonucleotide probes.

[0049] Generally, the larger the stringency class the greater the complexity of the hybridizing mixture and the lower the concentration of any particular oligonucleotide probe in the mixture. A lower concentration of a oligonucleotide probe having a complementary site on a target polynucleotide reduces the relative likelihood of the oligonucleotide probe hybridizing and being ligated. This, in turn, leads to reduced sensitivity. Larger stringency classes also have a greater variance in the stabilities of the duplexes that form between a oligonucleotide probe and a complementary sequence. On the other hand, smaller stringency classes require a larger number of hybridization reactions to ensure that all oligonucleotide probes of a set are hybridized to a target polynucleotide.

[0050] For example, when 8-mer oligonucleotide probes are employed stringency classes may include between about 50 to about 500 oligonucleotide probes each. Thus, several hundred to several thousand hybrization/ligation reactions are required. For larger sized oligonucleotide probes, much larger stringency classes are required to make the number of hybridization/extension reactions practical, e.g. 10^4-10^5, or more.

[0051] Oligonucleotide probes of the same stringency class can be synthesized simultaneously, in a manner similar to which fully random oligonucleotide probes are synthesized, e.g. as disclosed in Telenius et al, Genomics, 13: 718-725 (1992); Welsh et al, Nucleic Acids Research, 19: 5275-5279 (1991); Grothues et al, Nucleic Acids Research, 21: 1321-1322 (1993); Hartley, European patent application 90304496.4. The difference is that at each cycle different mixtures of monomers are applied to the growing oligonucleotide probe chain, wherein the proportion of each monomer in the mixture is dictated by the proportion of each nucleoside at the position of the oligonucleotide probe in the stringency class. Stringency classes are readily formed by computing the free energy of duplex formation by available algorithms, e.g. Breslauer et al, Proc. Natl. Acad. Sci., 83: 3746-3750 (1986); Lowe et al, Nucleic Acids Research, 18: 1757-1761 (1990). The oligonucleotide probes can be ordered according to the

free energy of binding to their complement under standard reaction conditions, e.g. with a standard bubble sort, Baase,

[0052] Computer Algorithms (Addison-Wesley, Menlo Park, 1978). For example the following is the list of ten 6-mers with the greatest stability (from top to bottom) in terms of free energy of duplex formation under standard hybridization conditions and the least stability in terms of free energy of duplex formation (the free energies being computed via Breslauer (cited above)):

Oligonucleotide Probe Ranking Sequence (5'- > 3')
1 GCGCGC
2 CGCGCG
3 CCCGCG
4 CGCCCG
5 CGCGCC
6 CGCGGC
7 CGGCGC
8 GCCGCG
9 GCGCCG
10 GCGCGG
4087 TCATAT
4088 TGATAT
4089 CATATA
4090 TATATG
4091 ATCATG
4092 ATGATG
4093 CATCAT
4094 CATGAT
4095 CATATG
4096 ATATAT

[0053] Thus, if a stringency class consisted of the first ten 6-mers the mixture monomers for the first (3'-most) position would be 0:4:6:0 (A:C:G:T), for the second position it would be 0:6:4:0, and so on. If a stringency class consisted of the last ten 6-mers the mixture of monomers for the first position would be 1:0:4:5, for the second position it would be 5:0:0:5, and so on. The resulting mixtures may then be further enriched for sequences of the desired stringency class by thermal elution, e.g. Miyazawa et al, J. Mol. Biol., 11: 223-237 (1965).

[0054] More conveniently, stringency classes containing several hundred to several thousands of oligonucleotides may be synthesized directly by a variety of parallel synthesis approaches, e.g. Frank et al, U.S. patent 4,689,405; Matson et al, Anal. Biochem., 224: 110-116 (1995); Fodor et al, International application WO1993022684; Pease et al, Proc. Natl. Acad. Sci., 91: 5022-5026 (1994); Southern et al, J. Biotechnology, 35: 217-227 (1994), Brennan, International application WO199427719.

[0055] In some cases it may be desirable to form additional stringency classes of oligonucleotide probes by placing in a separate subset oligonucleotide probes having complementary sequences to other oligonucleotide probes in a subset or oligonucleotide probe that are susceptible of forming oligonucleotide probe-dimers.

[0056] Clearly, one of ordinary skill in the art could combine features of the embodiments set forth above to design still further embodiments in accordance with the invention, but not expressly set forth above.

[0057] There are described systems and apparatus for carrying out the methods of the invention automatically. Such systems and apparatus can take a variety of forms depending on several design constraints, including i) the nature of the solid phase support used to anchor the target polynucleotide, ii) the degree of parallel operation desired, iii) the detection scheme employed; and iv) whether reagents are re-used or discarded. Generally, the apparatus comprises a series of reagent reservoirs, one or more reaction vessels containing target polynucleotide, preferably attached to a solid phase support, e.g. magnetic beads, one or more detection stations, and a computer controlled means for transferring in a predetermined manner reagents from the reagent

reservoirs to and from the reaction vessels and the detection stations. The computer controlled means for transferring reagents and controlling temperature can be implemented by a variety of general purpose laboratory robots, such as that disclosed by Harrison et al, Biotechniques, 14: 88-97 (1993);
Fujita et al, Biotechniques, 9: 584-591(1990); Wada et al, Rev. Sci. Instrum., 54: 1569-1572 (1983). Such laboratory robots are also available commercially, e.g. Applied Biosystems model 800 Catalyst (Foster City, CA).

[0058] A variety of kits may be provided for carrying out different embodiments of the invention. Generally, kits of the invention include oligonucleotide probes, initializing oligonucleotides, and a detection system. Kits further include ligation reagents and instructions for practicing the particular embodiment of the invention. In embodiments employing protein ligases, RNase H, nucleic acid polymerases, or other enzymes, their respective buffers may be included. In some cases, these buffers may be identical. Preferably, kits also include a solid phase support, e.g. magnetic beads, for anchoring templates. In one preferred kit, fluorescently labeled oligonucleotide probes are provided such that probes corresponding to different terminal nucleotides of the target polynucleotide carry distinct spectrally resolvable fluorescent dyes. As used herein, "spectrally resolvable" means that the dyes may be distinguished on basis of their spectral characteristics, particularly fluorescence emission wavelength, under conditions of operation. Thus, the identity of the one or more terminal nucleotides would be correlated to a distinct color, or perhaps ratio of intensities at different wavelengths. More preferably, four such probes are provided that allow a one-to-one correspondence between each of four spectrally resolvable fluorescent dyes and the four possible terminal nucleotides on a target polynucleotide. Sets of spectrally resolvable dyes are disclosed in U.S. patents 4,855,225 and 5,188,934; International application WO1991005060; and Lee et al, Nucleic Acids Research, 20: 2471-2483 (1992).

Example 1

Sequencing a Target Polynucleotide Amplified from pUC19 with Four Initializing Oligonucleotides

[0059] In this example, a template comprising a binding region and a portion of the pUC19 plasmid is amplified by PCR and attached to magnetic beads. Four initializing oligonucleotides are employed in separate reactions as indicated below. 8-mer oligonucleotide probes are employed having 4 central ribonucleotides and both 5' and 3' monophosphates, as shown in the following formula:

OP(= O)(O^-)O-(5')BBRRRRBB(3')-OP(= O)(O^-)O

[0060] After annealing, probes are enzymatically ligated to the initializing oligonucleotides and the magnetic bead supports are washed. The 3' phosphates of the ligated probes are removed with phosphatase, after which the probes are extended with DNA polymerase in the presence of the four labeled dideoxynucleoside triphosphate chain terminators. After washing and identification of the extended nucleotide, the ligated probes are cleaved at the ribonucleotide moiety with RNAse H to remove the label and to regenerate an extendable end.

[0061] The following double stranded fragment comprising a 36-mer binding region is ligated into a Sac I/Xma I-digested pUC19:
CCTCTCCCTTCCCTCTCCTCCCTCTCCCCTCTCCCTC
TCGAGGAGAGGGAAGGGAGAGGAGGGAGAGGGGAGAGGGAGGGCC

[0062] After isolation and amplification, a 402 basepair fragment of the modified pUC19 is amplified by PCR for use as a template. The fragment spans a region of pUC19 from position 41 to the binding region inserted adjacent to the Sac I site in the polylinker region (position 413 of the unmodified pUC19), Yanisch-Perron et al, Gene, 33: 103-119 (1985).

[0063] Two 18-mer oligonucleotide probes are employed having sequences 5' CCCTCTCCCCTCTCCCTCx-3' and 5'-GCAGCTCCCGGAGACGGT-3', where "x" is a 3' biotin moiety is attached during synthesis using a commercially available reagent with manufacturer's protocol, e.g. 3' Biotin-ON CPG (Clontech Laboratories, Palo Alto, California). The amplified template is isolated and attached to streptavidin-coated magnetic beads (Dynabeads) using manufacturer's protocol, Dynabeads Template Preparation Kit, with M280-streptavidin (Dynal, Inc., Great Neck, New York). A sufficient quantity of the biotinylated 313 basepair fragment is provided to load about 300 g of Dynabeads M280 Streptavidin.

[0064] The binding region sequence is chosen so that the duplexes formed with the initiating oligonucleotides have compositions

of about 66% GC to enhance duplex stability. The sequence is also chosen to prevent secondary structure formation and fortuitous hybridization of an initializing oligonucleotide to more than one location within the binding region. Any shifting of position of a given initializing oligonucleotide within the binding region results in a significant number of mis-matched bases.

[0065] After loading, the non-biotinylated strand of template is removed by heat denaturation, after which the magnetic beads are washed and separated into four aliquots. The template attached to the magnetic beads has the following sequence:
(Magnetic bead)-(linker)-(3')-CTCCCTCTCCCCTCTCCCTCCTC-

TCCCTTCCTCTCCTCGAGCTTAAGT ... CTCGACG-(5')

[0066] The following four oligonucleotides are employed as initializing oligonucleotides in each of the separate aliquots of template:

5'-GAGGAGAGGGAAGGAGAGGAG

5'-GGAGGAGAGGGAAGGAGAGGA

5'-GGGAGGAGAGGGAAGGAGAGG

5'-AGGGAGGAGAGGGAAGGAGAG

[0067] Reactions and washes below are generally carried out in 50 L volumes of manufacturer's (New England Biolabs) recommended buffers for the enzymes employed, unless otherwise indicated. Standard buffers are also described in Sambrook et al, Molecular Cloning, 2nd Edition (Cold Spring Harbor Laboratory Press, 1989).

[0068] 96 stringency classes of 684 or 682 oligonucleotide probes each (2 subsets for each of 48 different annealing temperatures) are formed which together contain all 8-mer probes for each of the four aliquots. The probes of each of the 96 classes are separately annealed to the target polynucleotide in reaction mixtures having the same components, with the exception that extensions and ligations carried out with Sequenase and T4 DNA ligase at temperatures less than 37°C and extensions and ligations carried out with Taq Stoffel fragment and a thermostable ligase otherwise.

[0069] The 48 stringency conditions are defined by annealing temperatures which range from 22°C to 70°C, such that each grouping of subsets at the same temperature differ in annealing temperature by 1°C from that of the subset groupings containing the next highest and next lowest stringency classes. The range of annealing temperatures (22-70°C) is roughly bounded by the temperatures 5-10 degrees below the temperatures at which the least stable and most stable 8-mers, respectively, are expected to have about fifty percent maximum annealing in a standard PCR buffer solution.

[0070] After 5-10 minutes incubation at 80°C, the reaction mixtures are brought down to their respective annealing temperatures over a period of 20-30 minutes. After ligation, washing and treatment with phosphatase, 2 units of polymerase and labeled dideoxynucleotide triphosphates (.08 mM final reaction concentration and labeled with TAMRA (tetramethylrhodamine), FAM (fluorescein), ROX (rhodamine X), and JOE (2',7' dimethoxy-4',5'-dichlorofluorescein)) are added. After 15 minutes, the beads are washed with H_2O and the identity of the extended nucleotide is determined by illuminating each reaction mixture with standard wavelengths, e.g Users Manual, model 373 DNA Sequencer (Applied Biosystems, Foster City, CA).

[0071] After identification, the reaction mixtures are treated with RNase H using the manufacturer's suggested protocol and washed. The RNase H treated extended duplexes have regenerated 3' hydroxyls and are ready for the next cycle of ligation/extension/ cleavage. The cycles are carried out until all the nucleotides of the test sequence are identified.

Example 2

Sequencing a Target Polynucleotide Amplified from pUC19 with One Initializing Oligonucleotide

[0072] In this example, a template is prepared in accordance with Example 1, except that since extension is in the 5'- > 3' direction in this example, the biotin moiety is attached to the 5' end of the primer hybridizing to the CT-rich strand of the binding region. Thus, in this example, the binding region of the single stranded template will be a GA-rich segment (essentially the

complement of the binding region of Example 1). Two 18-mer oligonucleotide probes are employed having sequences 5'-xGAGGGAGAGGGGAGAGGG-3' and 5'-ACCGTCTCCGGGAGCTGC-3', where "x" is a 5' biotin moiety is attached during synthesis using commercially available reagents with manufacturers' protocols, e.g. the Aminolink aminoalkylphosphoramidite linking agent (Applied Biosystems, Foster City, California) and Biotin-X-NHS Ester available form Clontech Laboratories (Palo Alto, California).

[0073] A single 21-mer initializing oligonucleotide is employed with the following sequence:

5'-OP(= O)(O^-)O-CCTCTCCCTTCCCTCTCCTCC-3'

6-mer oligonucleotide probes are employed that have an acid labile phosphoramidate linkage between the 3'-most nucleoside and 3'-penultimate nucleoside of the probe, as shown in the following formula:

HO-(3')B(5')-OP(= O)(O^-)NH-(3')BBBBB$_t$*

where B,* is a JOE-, FAM-, TAMRA-, or ROX-labeled dideoxynucleoside, such that the label corresponds to the identity of the 3'-most nucleotide (so 16 different labeled dideoxynucleosides are used in the synthesis of the probes).

[0074] As above, the 6-mer probes are prepared in 96 stringency classes of 42 or 43 probes each (2 subsets for each of 48 different annealing temperatures). Hybridizations and ligations are carried out as described above. After ligation and washing, a nucleoside in the target polynucleotide is identified by the fluorescent signal of the oligonucleotide probe. Acid cleavage is then carried out by treating the extended duplex with 0.8% trifluoroacetic acid in dichloromethane for 40 minutes at room temperature to regenerate an extendable end on the extended duplex. The process continues until the sequence of the target polynucleotide is determined.

SEQUENCE LISTING

[0075]

<110> Solexa, Inc.

<120> DNA SEQUENCING BY PARALLEL OLIGONUCLEOTIDE EXTENSIONS

<130> NLW/P87919EP02

<160> 8

<170> PatentIn version 3.5

<210> 1
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 1
cctctccctt ccctctcctc cctctcccct ctccctc 37

<210> 2
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonuceotide

<400> 2
gaggagaggg aaggagagga g 21

<210> 3
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 3
ggaggagagg gaaggagagg a 21

<210> 4
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 4
gggaggagag ggaaggagag g 21

<210> 5
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 5
agggaggaga gggaaggaga g 21

<210> 6
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucelotide

<400> 6
gagggagagg ggagaggg 18

<210> 7
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 7
accgtctccg ggagctgc 18

<210> 8
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 8
cctctccctt ccctctcctc c 21

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US4855225A [0006] [0058]
- WO0305183A [0006]
- WO1992015711A [0007]
- WO1991002796A [0007]
- US4123610A [0023]
- WO1992020702A [0023]
- WO1992020703A [0023]
- US5002867A [0026]
- US4980460A [0027]
- US4725677A [0027]
- US4415732A [0027]
- US4458066A [0027]
- US4973679A [0027]
- EP90304496A [0027] [0051]
- US4757141A [0029]
- US5151507A [0029]
- US5091519A [0029]
- US5124246A [0030]
- US5188934A [0031] [0058]
- WO1991005060A [0031] [0058]
- US4734363A [0036]
- WO1989009262A [0036]
- US4816571A [0037]
- WO1992008728A [0039]
- WO5380833A [0041]
- WO1992002528A [0041]
- WO1994023064A [0042]
- DE4141178A1 [0042]
- US5302509A [0042]
- WO1991006678A [0042]
- US5011769A [0043]
- US4689405A [0054]
- WO1993022684A [0054]
- WO199427719A [0054]

Non-patent literature cited in the description

- **GIBBS et al.**Proc. Natl. Acad. Sci., 1989, vol. 86, 1919-1923 [0002]
- **GYLLENSTEN et al.**Proc. Natl. Acad. Sci, 1988, vol. 85, 7652-7656 [0002]
- **CARRANO et al.**Genomics, 1989, vol. 4, 129-136 [0002]
- **CAETANO-ANOLLES et al.**Mol. Gen. Genet., 1992, vol. 235, 157-165 [0002]
- **BRENNERLIVAK**Proc. Natl. Acad. Sci., 1989, vol. 86, 8902-8906 [0002]
- **GREEN et al.**PCR Methods and Applications, 1991, vol. 1, 77-90 [0002]
- **VERSALOVIC et al.**Nucleic Acids Research, 1991, vol. 19, 6823-6831 [0002]
- **SANGER et al.**Proc. Natl. Acad. Sci., 1977, vol. 74, 5463-5467 [0004]
- **MAXAM et al.**Proc. Natl. Acad. Sci., 1977, vol. 74, 560-564 [0004]
- **SANGER et al.**J. Mol. Biol., 1980, vol. 143, 161-178 [0004]
- **SCHREIER et al.**J. Mol. Biol., 1979, vol. 129, 169-172 [0004]
- **SMITH et al.**Nucleic Acids Research, 1985, vol. 13, 2399-2412 [0004]
- **SMITH et al.**Nature, 1987, vol. 321, 674-679 [0004]
- **PROBER et al.**Science, 1987, vol. 238, 336-341 [0004]
- Section II, Meth. Enzymol., 1987, vol. 155, 51-334 [0004]
- **CHURCH et al.**Science, 1988, vol. 240, 185-188 [0004]
- **HUNKAPILLER et al.**Science, 1991, vol. 254, 59-67 [0004]
- **BEVAN et al.**PCR Methods and Applications, 1992, vol. 1, 222-228 [0004]
- **BANKIER et al.**Meth. Enzymol., 1987, vol. 155, 51-93 [0005]
- **HAWKINS et al.**Electrophoresis, 1992, vol. 13, 552-559 [0005]
- **TRAINOR**Anal. Biochem., 1990, vol. 62, 418-426 [0006]
- **CONNELL et al.**Biotechniques, 1987, vol. 5, 342-348 [0006]
- **KARGER et al.**Nucleic Acids Research, 1991, vol. 19, 4955-4962 [0006]
- **NISHIKAWA et al.**Electrophoresis, 1991, vol. 12, 623-631 [0006]
- **GYLLENSTEN et al.**PCR Methods and Applications, 1991, vol. 1, 91-98 [0007]
- **KORNBERGBAKER**DNA ReplicationFreeman, San Francisco19920000 [0016]
- **SCHEIT**Nucleotide AnalogsJohn Wiley19800000 [0016]
- Science, 1991, vol. 254, 1497-1500 [0023]
- **HANVEY et al.**Science, 1992, vol. 258, 1481-1485 [0023]
- **KONG THOO LIN et al.**Nucleic Acids Research, vol. 20, 5149-5152 [0026]
- **NICHOLS et al.**Nature, 1994, vol. 369, 492-493 [0026]
- **BEAUCAGEIYER**Tetrahedron, 1992, vol. 48, 2223-2311 [0027]
- **TELENIUS et al.**Genomics, 1992, vol. 13, 718-725 [0027] [0051]
- **WELSH et al.**Nucleic Acids Research, 1991, vol. 19, 5275-5279 [0027] [0051]
- **GROTHUES et al.**Nucleic Acids Research, 1993, vol. 21, 1321-1322 [0027] [0051]
- **SAMBROOK et al.**Molecular Cloning: A Laboratory ManualCold Spring Harbor Laboratory19890000 [0028] [0035]
- **HORNURDEA**Tetrahedron Lett., 1986, vol. 27, 4705- [0028]
- **MATTHEWS et al.**Anal. Biochem., 1988, vol. 169, 1-25 [0029]
- **HAUGLAND**Handbook of Fluorescent Probes and Research ChemicalsMolecular Probes, Inc.19920000 [0029]
- **KELLERMANAK**DNA ProbesStockton Press19930000 [0029]
- Oligonucleotides and Analogues: A Practical ApproachIRL Press19910000 [0029]
- **JABLONSKI et al.**Nucleic Acids Research, 1986, vol. 14, 6115-6128 [0029]
- **WETMUR**Critical Reviews in Biochemistry and Molecular Biology, 1991, vol. 26, 227-259 [0032]
- **DOVEDAVIDSON**J. Mol. Biol., 1962, vol. 5, 467-478 [0032]
- **HUTTON**Nucleic Acids Research, 1977, vol. 10, 3537-3555 [0032]
- **BRESLAUER et al.**Proc. Natl. Acad. Sci.19860000vol. 83, 3746-3750 [0032]
- PCR ProtocolsAcademic Press19900000 [0032] [0035] [0047]
- **FERRIS et al.**Nucleosides & Nucleotides, 1989, vol. 8, 407-414 [0033]
- **SHABAROVA et al.**Nucleic Acids Research, 1991, vol. 19, 4247-4251 [0033]
- **LEHMAN**Science, 1974, vol. 186, 790-797 [0033]
- DNA Ligases**ENGLER et al.**The EnzymesAcademic Press19820000vol. 15B, 3-30 [0033]
- **BARANY**PCR Methods and Applications, 1991, vol. 1, 5-16 [0033]
- **MARSH et al.**Strategies, 1992, vol. 5, 73-76 [0033]
- **WAHLBERG et al.**Electrophoresis, 1992, vol. 13, 547-551 [0036]
- **TONG et al.**Anal. Chem., 1992, vol. 64, 2672-2677 [0036]
- **HULTMAN et al.**Nucleic Acids Research, 1989, vol. 17, 4937-4946 [0036]
- **HULTMAN et al.**Biotechniques, 1991, vol. 10, 84-93 [0036]
- **SYVANEN et al.**Nucleic Acids Research, 1988, vol. 16, 11327-11338 [0036]

- **LETSINGER et al.**J. Am. Chem. Soc., 1971, vol. 94, 292-293 [0039]
- **LETSINGER et al.**Biochem., 1976, vol. 15, 2810-2816 [0039]
- **GRYAZNOV et al.**Nucleic Acid Research, 1992, vol. 20, 3403-3409 [0039]
- **GRYAZNOV et al.**Nucleic Acids Research, 1993, vol. 21, 1403-1408 [0041]
- **GRYAZNOV et al.**Nucleic Acids Research, 1994, vol. 22, 2366-2369 [0041]
- **METZKER et al.**Nucleic Acids Research, 1994, vol. 22, 4259-4267 [0042]
- **HOGREFE et al.**J. Biol. Chem., 1990, vol. 265, 5561-5566 [0043]
- **GOTOH**Adv. Biophys., 1983, vol. 16, 1-52 [0047]
- **WETMER**Critical Reviews in Biochemistry and Molecular Biology, 1991, vol. 26, 227-259 [0047]
- **BRESLAUER et al.**Proc. Natl. Acad. Sci., 1986, vol. 83, 3746-3750 [0047] [0051]
- **WOLF et al.**Nucleic Acids Research, 1987, vol. 15, 2911-2926 [0047]
- **MCGRAW et al.**Biotechniques, 1990, vol. 8, 674-678 [0047]
- **LOWE et al.**Nucleic Acids Research, 1990, vol. 18, 1757-1761 [0051]
- Computer AlgorithmsAddison-Wesley [0052]
- **MIYAZAWA et al.**J. Mol. Biol., 1965, vol. 11, 223-237 [0053]
- **MATSON et al.**Anal. Biochem., 1995, vol. 224, 110-116 [0054]
- **PEASE et al.**Proc. Natl. Acad. Sci., vol. 91, 5022-5026 [0054]
- **SOUTHERN et al.**J. Biotechnology, 1994, vol. 35, 217-227 [0054]
- **HARRISON et al.**Biotechniques, 1993, vol. 14, 88-97 [0057]
- **FUJITA et al.**Biotechniques, 1990, vol. 9, 584-591 [0057]
- **WADA et al.**Rev. Sci. Instrum., 1983, vol. 54, 1569-1572 [0057]
- **LEE et al.**Nucleic Acids Research, 1992, vol. 20, 2471-2483 [0058]
- **YANISCH-PERRON et al.**Gene, 1985, vol. 33, 103-119 [0062]
- **SAMBROOK et al.**Molecular CloningCold Spring Harbor Laboratory Press19890000 [0067]
- Users Manual, model 373 DNA SequencerApplied Biosystems [0070]

Krav

1. Fremgangsmåde til at identificere et eller flere nucleotider i et polynucleotid, fremgangsmåden omfatter at
   (a) forlænge et initieringsoligonucletid langs polynucleotidet ved at ligere en oligonucleotidprobe med en blokeringsdel dertil for at danne en forlænget duplex, som ikke er i stand til at ligere med en yderligere oligonucletidprobe; og
   (b) identificere et eller flere af polynucleotidets nucleotider ved hjælp af et fluorescerende mærke på den ligerede oligonucleotidprobe.

2. Fremgangsmåde ifølge krav 1, hvori proben er i stand til at blive behandlet for at regenerere en forlængbar ende efter ligering.

3. Fremgangsmåde ifølge krav 1, hvori nævnte identifikationstrin består af at identificere et enkelt nucleotid i nævnte polynucleotid.

4. Fremgangsmåde ifølge ethvert af kravene 1 – 3, hvori mindst nogle af de tilvejebragte prober omfatter mindst en deoxyinosinbase.

5. Fremgangsmåde ifølge ethvert af kravene 1 – 3, hvori de tilvejebragte prober har en forudbestemt længde valgt blandt 8 nucleotider, 9 nucleotider og 12 nucleotider.

6. Fremgangsmåde ifølge ethvert af kravene 1 – 5, hvori mærket er fastgjort til den ligerede probes endenucleotid modsat den ende af den ligerede probe, som bliver ligeret.

7. Fremgangsmåde ifølge ethvert af kravene 1 – 6, yderligere omfattende at amplificere det første polynucleotid.

8. Fremgangsmåde ifølge krav 7, hvori det amplificerede polynucleotid omfatter multiple kopier af det tilvejebragte polynucleotid båret på en perle.

9. Fremgangsmåde ifølge ethvert af kravene 1 – 8, yderligere omfattende at bestemme en konsekutiv basesekvens af forskellige ukendte sekvenser af forskellige polynucleotider

baseret på gennemførelsen af forlængelses- og identifikationstrinnene af multiple forskellige parallelle polynucleotider.

10. Fremgangsmåde ifølge krav 9, hvori de multiple forskellige polynucleotider har en kendt sekvens til fælles med hinanden som binder til initieringsoligonucletidet.

11. Fremgangsmåde ifølge krav 9 eller krav 10, hvori de forskellige polynucleotider er fastgjorte til et fast underlag.

12. Fremgangsmåde ifølge ethvert af kravene 1 – 10, yderligere omfattende at bestemme rækkefølgen af hvert nucleosid i polynucleotidet, som bliver sekventeret.

13. Fremgangsmåde ifølge ethvert af kravene 1 – 11, hvori initieringsoligonucletiderne er mindst 20 baser lange.

14. Fremgangsmåde ifølge ethvert af kravene 1 – 13, hvori multiple initieringsoligonucletiderhar den samme sekvens.

15. Fremgangsmåde ifølge ethvert af kravene 1 – 14, hvori initieringsoligonucletidet kan forlænges i enten 3’ til 5’-retningen eller i 5’ til 3’-retningen.

DRAWINGS

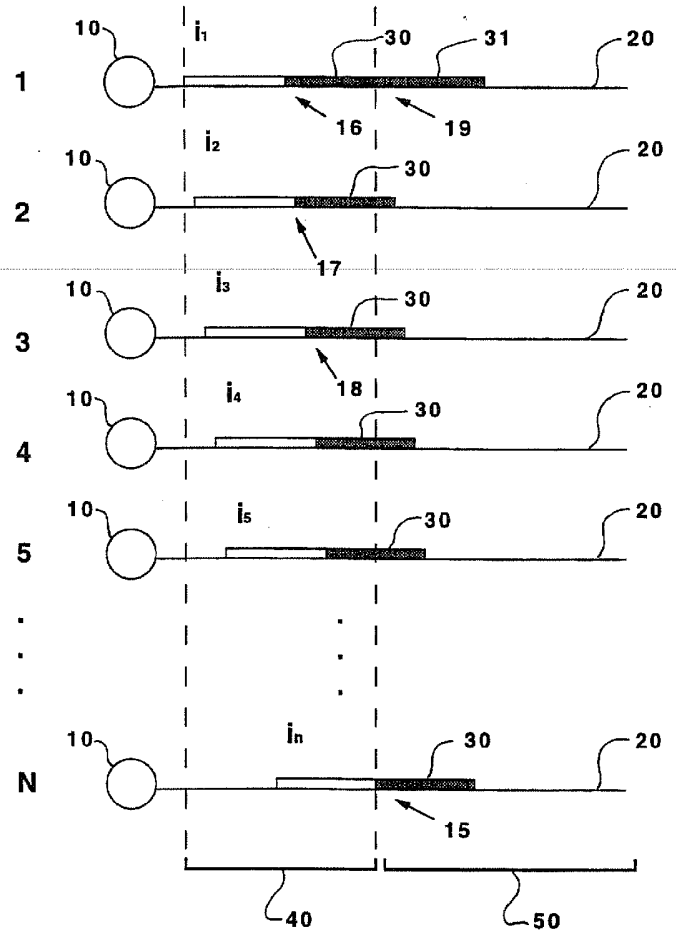

Fig. 1

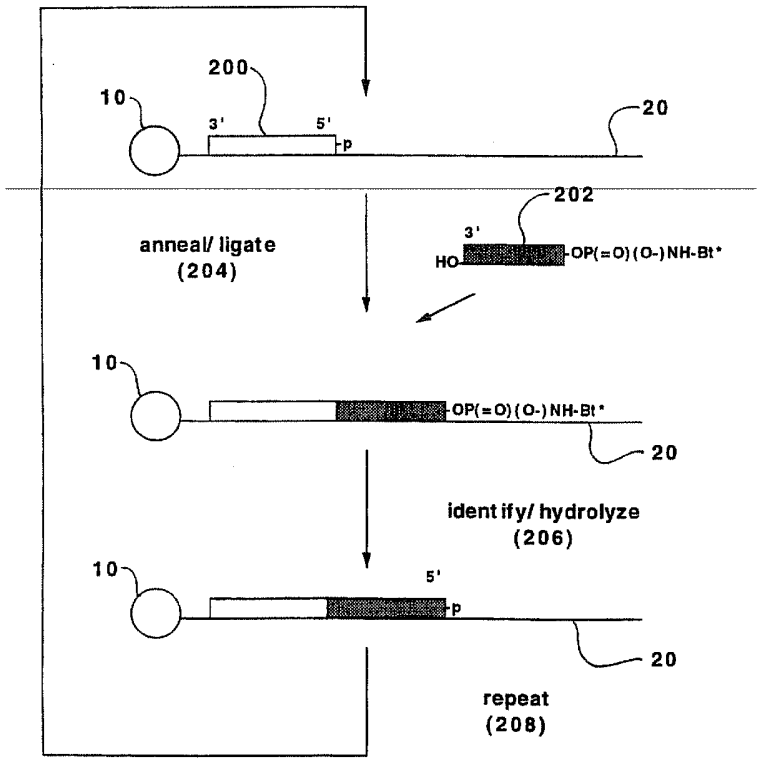
200
10
3'
5'
p
20
202
3'
HO
OP(=O)(O-)NH-Bt*
anneal/ ligate
(204)
10
OP(=O)(O-)NH-Bt*
20
identify/ hydrolyze
(206)
10
5'
p
20
repeat
(208)

Fig. 2

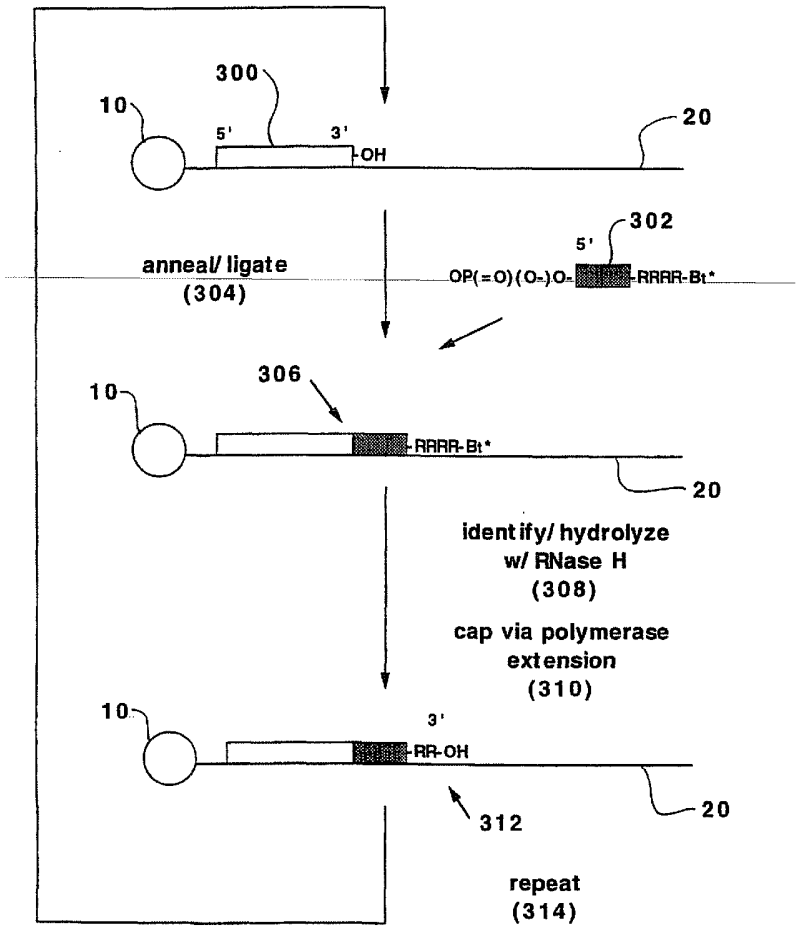
300
10
5'
3'
OH
20
302
5'
anneal/ ligate
(304)
OP(=O)(O-)O-
RRRR-Bt*
306
10
RRRR-Bt*
20
identify/ hydrolyze
w/ RNase H
(308)
cap via polymerase
extension
(310)
10
3'
RR-OH
20
312
repeat
(314)

Fig. 3a

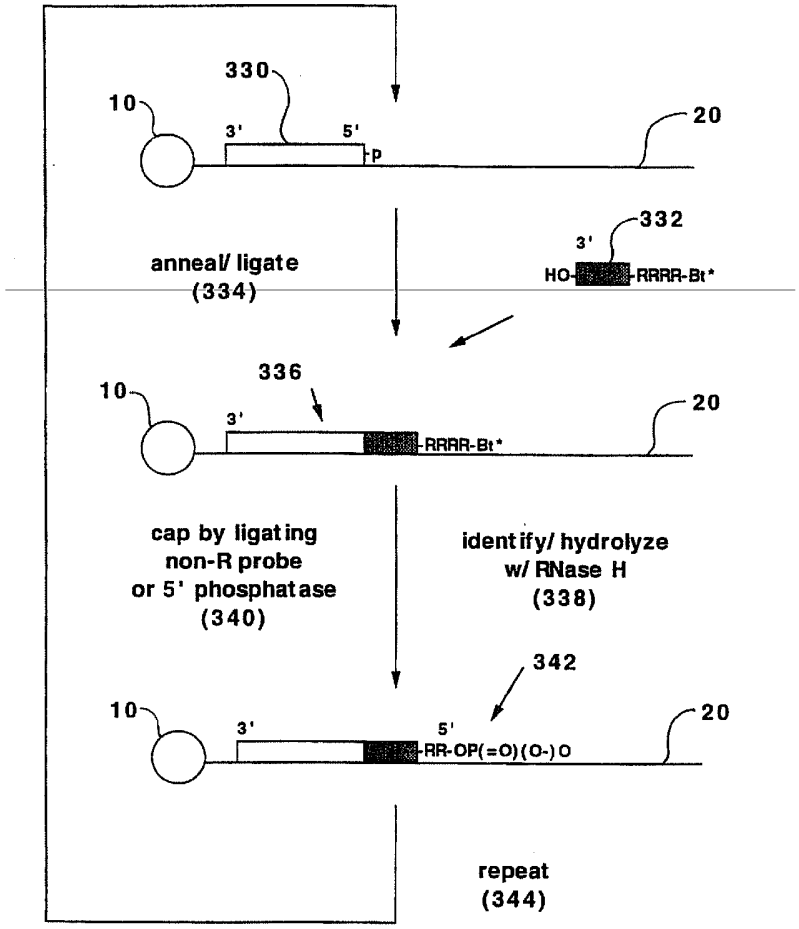
330
10
3'
5'
P
20
anneal/ ligate
(334)
332
3'
HO-
-RRRR-Bt*
336
10
3'
-RRRR-Bt*
20
cap by ligating
non-R probe
or 5' phosphatase
(340)
identify/ hydrolyze
w/ RNase H
(338)
342
10
3'
5'
-RR-OP(=O)(O-)O
20
repeat
(344)

Fig. 3b

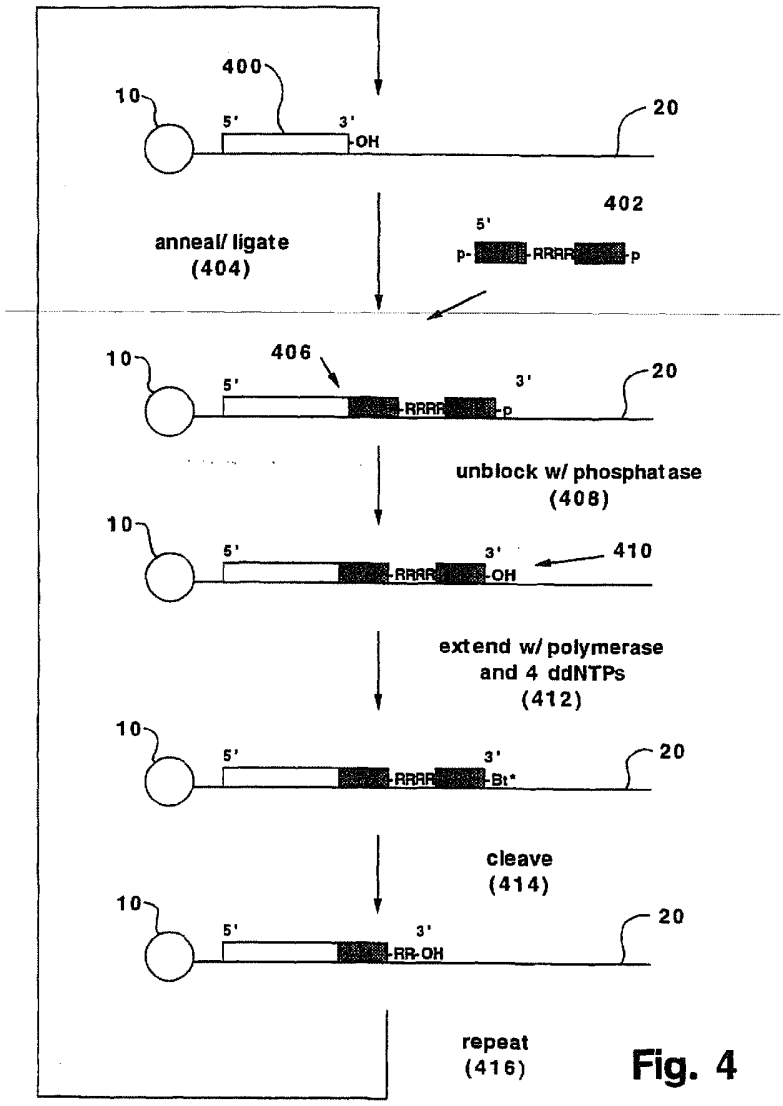
400
10
5'
3'
OH
20
anneal/ ligate
(404)
402
5'
p-
RRRR
p
406
10
5'
3'
RRRR
p
20
unblock w/ phosphatase
(408)
10
5'
3'
410
RRRR
OH
extend w/ polymerase
and 4 ddNTPs
(412)
10
5'
3'
20
RRRR
Bt*
cleave
(414)
10
5'
3'
20
RR-OH
repeat
(416)

Fig. 4