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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 November 2007 (29.11.2007)

PCT

(10) International Publication Number
WO 2007/136736 A2

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

(21) **International Application Number:** PCT/US2007/011886

(22) **International Filing Date:** 18 May 2007 (18.05.2007)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
60/802,197 19 May 2006 (19.05.2006) US

(71) **Applicant** *(for all designated States except US)*: **CODON DEVICES, INC.** [US/US]; One Kendall Square, Building 300, Cambridge, MA 02139 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **CHURCH, George** [US/US]; 218 Kent Street, Brookline, MA 02446 (US). **CERRINA, Francisco** [US/US]; 15 Windswept Way, Madison, WI 53719 (US). **BAYNES, Brian, M.** [US/US]; 19 Orrin Street, Cambridge, MA 02138 (US). **PITCHER, Edmund, R.** [US/US]; 2 Smith Farm Lane, Lexington, MA 02421 (US).

(74) **Agent: WALLER, Patrick, R., H.**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).

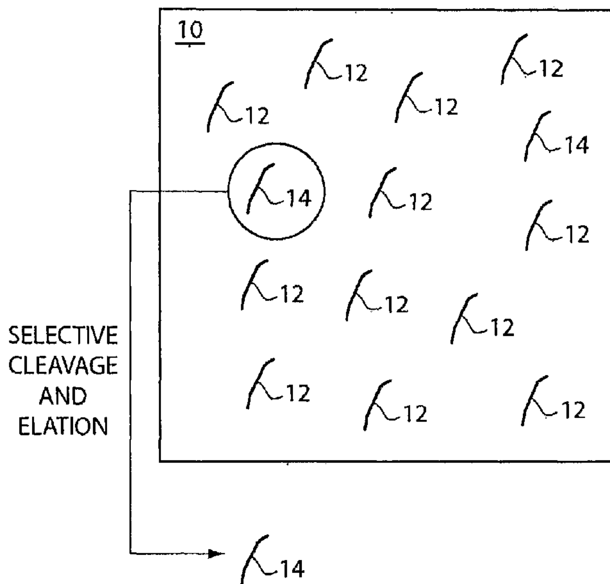
(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN,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

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** METHODS FOR NUCLEIC ACID SORTING AND SYNTHESIS



(57) **Abstract:** Methods of sorting oligonucleotide sequences are provided. The methods include determining the sequences of oligonucleotides and selectively isolating one or more that are determined to have the desired sequence. Methods of nucleic acid preparation are also disclosed.

WO 2007/136736 A2

METHODS FOR NUCLEIC ACID SORTING AND SYNTHESIS

RELATED APPLICATIONS

This application claims the benefit under 35 U.S.C. §119(e) from U.S. provisional application serial number 60/802,197, filed May 19, 2006, the entire contents
5 of which are herein incorporated by reference.

FIELD OF THE INVENTION

The invention relates to methods of synthesis and assembly of nucleic acids.

BACKGROUND

The ability to synthesize stepwise and later assemble oligonucleotides has
10 permitted the de novo synthesis of small and large nucleic acids, including entire viral genomes. However, chemically synthesized oligonucleotides may have an error rate (i.e., a rate of omitting a correct nucleotide at a particular position and/or of inserting an incorrect nucleotide at that position and/or of switching one nucleotide for another) exceeding the error rate obtainable through enzymatic means of replicating an existing
15 nucleic acid (e.g., PCR). If an error resulting from chemical synthesis is introduced into an assembled nucleic acid, either the error must be corrected or the assembled nucleic acid discarded. In either event, the cost of assembling a correct nucleic acid is thereby increased. Massively parallel oligonucleotide syntheses on chips have been performed in an effort to reduce total costs associated with nucleic acid synthesis. The error rate in
20 these massively parallel syntheses, however, are generally greater than for more conventional chemical synthesis methods, further complicating efforts to assemble nucleic acids of correct sequence.

SUMMARY OF THE INVENTION

The present invention permits the selective isolation of individual oligonucleotide
25 molecules with more accurate sequences from a population of oligonucleotide molecules with less accurate sequences. Once isolated, the individual oligonucleotide molecules

can be used for any purpose that benefits from increased accuracy, such as the synthesis of longer nucleic acid molecules.

Aspects of the invention relate to determining the sequence of one or more oligonucleotide molecules and isolating oligonucleotides having correct sequences of interest. In some embodiments, oligonucleotide sequences may be determined using non-destructive sequencing techniques after synthesis. Accordingly, error-free oligonucleotides may be identified and isolated. In some embodiments, oligonucleotide sequences may be monitored (e.g., determined using non-destructive techniques) during synthesis and error-free oligonucleotides may be isolated upon completion of synthesis without any further sequencing step. Error-free oligonucleotides may be isolated by separating oligonucleotides containing one or more synthesis errors from oligonucleotides having no sequence errors. The separation may be performed by selectively isolating (e.g., releasing from a solid support) oligonucleotide(s) having a correct sequence, and/or by selectively removing (e.g., degrading, or damaging) oligonucleotide(s) containing one or more sequence errors. In one embodiment, the invention facilitates the sorting of oligonucleotides by consecutively interrogating the identity of each of a plurality of contiguous nucleobases for each of several of the oligonucleotides. This can include, for example, a sequencing by synthesis reaction, where the identity of each template oligonucleotide is consecutively interrogated over the course of synthesis of its complement and, conversely, the identity of each oligonucleotide of the complementary oligonucleotide is consecutively interrogated as it is synthesized. Thus, the sequences of several oligonucleotides are determined. It is understood that the determination need not be perfect, but need merely be indicative of accuracy. In other words, it is sufficient if an oligonucleotide determined to have the desired sequence is indeed more likely to have the desired sequence than other oligonucleotides are. The invention permits the selective isolation of one or more oligonucleotides determined to have the desired sequence from one or more oligonucleotides determined not to have the desired sequence. "Selective isolation," as used herein, can involve physical isolation of a desired oligonucleotide from others, as by selective physical movement of the desired oligonucleotide; selective release of the desired oligonucleotide from a solid support or other molecule; selective inactivation, destruction, release, or removal of other oligonucleotide molecules; and /or selective

amplification of the desired oligonucleotide, in which case each oligonucleotide can optionally incorporate a pair of primer binding sites to facilitate amplification.

The sorting of individual oligonucleotide molecules can be facilitated by the use of one or more solid supports to which the oligonucleotides are attached. The solid support can be, for example, a bead, a substantially planar surface such as a slide or coverslip, or another insoluble, optionally polymeric material facilitating the separation of an attached oligonucleotide from soluble components of a contacting solution by, for example, filtration, centrifugation, or manual manipulation. Accordingly, in one embodiment, the sequence of each of several oligonucleotides attached to one or more solid supports is determined. Oligonucleotides determined to have the desired sequence are selectively released or selectively copied. The attachment of the oligonucleotides to one or more solid supports can involve, for example, attachment at random locations on one or more solid supports; attachment at fixed locations on one or more solid supports; or amplifying each of a plurality of oligonucleotides by PCR and attaching individual amplicons to individual solid supports, thereby generating "PCR colonies" or "polonies": colocalized populations of oligonucleotides amplified from a common ancestor. Sequencing of any member of an amplicon, or of a sampling of the amplicon, provides sequence information indicative of the sequences of other members of the amplicon. More generally, determinations of the sequences of the attached oligonucleotides can involve, for example, sequencing by hybridization, or a method involving the consecutive interrogation of the identity of each of a plurality of contiguous nucleobases, such as sequencing by synthesis. A nondestructive method of determining the sequence should be used if the intention is thereafter to isolate the individual oligonucleotide molecule sequenced, although other methods of sequence determination are acceptable if the intention is to determine the sequence of one or more members of an amplicon and to isolate a different member of the same amplicon.

If the oligonucleotides to be sorted are attached to a solid support, the introduction of a selectively cleavable linker between the oligonucleotide and the solid support can facilitate selective release of a desired oligonucleotide. Useful selectively cleavable linkers include, for example, those that are selectively cleaved upon exposure to electromagnetic radiation, acidic pH, basic pH, enzymes, or particular chemical

agents, or upon exposure to a combination of triggers, whether sequentially or simultaneously. For example, if the selectively cleavable linkers are sensitive to electromagnetic radiation, selectively irradiating the linker of an oligonucleotide determined to have the desired sequence can selectively release the oligonucleotide from the solid support.

Alternatively, or in addition, oligonucleotides determined not to have the desired sequence can be selectively released from a solid support or selectively inactivated. Thus, they could be selectively released by selectively triggering cleavage of linkers attaching the undesired oligonucleotides to their solid support(s). Similarly, selectively irradiating the undesired oligonucleotides with a destructive radiation can selectively inactivate them, leaving an enriched population of accurate oligonucleotides for subsequent copying or release.

The invention also provides a method of preparing nucleic acids with increased accuracy. The method includes synthesizing a multiplicity (e.g., more than one hundred, more than one thousand, or more than one million) of individual oligonucleotide molecules, each intended to have the same nucleotide sequence. The synthesis can be, for example, a conventional chemical synthesis or a chip-based chemical synthesis. The synthesized nucleic acids are optionally attached to solid supports and can be amplified by polymerase colony amplification prior to attachment to individual solid supports. The synthesis is followed by the consecutive interrogation of the identity of each of a plurality of contiguous nucleobases of various individual oligonucleotide molecules, their complements, or copies thereof (e.g., other members of a polony). This process identifies one or more desirable nucleic acids, regardless of whether the desirable nucleic acid is one of the synthesized nucleic acids or a complement or copy thereof. One or more of the desirable nucleic acids is then selectively isolated from other nucleic acids present. The selective isolation can involve separating an individual solid support attached to desirable nucleic acids from at least one individual solid support attached to other nucleic acids and can involve selectively releasing desirable nucleic acids from one or more solid supports.

The present invention can be practiced as a high-throughput method of nucleic acid preparation. The method includes providing one or more solid supports attached to

a multiplicity of populations of oligonucleotides, each member of a population having the same intended sequence and each population having a different intended sequence. The oligonucleotides provided can be synthesized on a surface, for example, and can optionally be released from the surface and amplified by polymerase colony amplification, followed by attaching each amplicon to a solid support. The sequences of the oligonucleotides on the one or more solid supports are determined in parallel (e.g., by sequencing by synthesis, sequencing by hybridization, or other means) to identify at least one member of each population determined to have its intended sequence. Population members determined to have their intended sequences are then selectively released or selectively copied.

Selectively released or selectively copied members are then available for any desired purpose. For example, they can be exposed to conditions promoting hybridization of a member of one population to a member of another population to permit assembly of a double stranded nucleic acid including members of a plurality of populations. The selective release or selective copying can be simultaneous, leading to a single pool of oligonucleotides determined to have their desired sequences from all populations; sequential, leading to a number of substantially pure pools, each derived from a single population; or a combination thereof, leading to a plurality of pools derived from different sets of populations. The ability to isolate desired oligonucleotides in more than one pool, each from a defined set of populations, facilitates their subsequent use by isolating from each other any oligonucleotides that could interfere with each other, such as by undesired annealing. Thus, for example, the invention permits assembling a nucleic acid incorporating oligonucleotides from one or more pools and subsequently assembling a longer nucleic acid that further incorporates an oligonucleotide from a different pool, which can simplify the assembly process.

Other features, objects, and advantages of the present invention are apparent in the drawings and detailed description that follow. It should be understood, however, that these descriptions of various embodiments of the present invention are merely illustrative and do not limit the scope of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic of an exemplary base-labile safety-catch linker with a photocleavable protecting group.

Figure 2 depicts an exemplary class of safety-catch linkers optionally used to
5 attach an oligonucleotide to a solid support. The linkers include aromatic ring R2 connected to photocleavable protecting group PCPG via acid-labile ketal R1.

Figure 3 depicts a specific exemplary safety-catch linker.

Figure 4 schematically depicts a method of selectively isolating oligonucleotides having a desired sequence by selectively removing or inactivating other oligonucleotides.

10 Figure 5 schematically depicts a method of selectively isolating an oligonucleotide having a desired sequence by selectively releasing the oligonucleotide from a solid support.

Figure 6 schematically depicts one embodiment of a plurality of oligonucleotides that may be assembled in a polymerase -based multiplex oligonucleotide assembly
15 reaction.

Figure 7 schematically depicts an embodiment of a plurality of oligonucleotides that may be assembled in a directional polymerase -based multiplex oligonucleotide assembly reaction.

Figure 8 schematically depicts an embodiment of a plurality of oligonucleotides
20 that may be assembled in a ligase reaction.

Figure 9 schematically depicts an embodiment of a ligase-based assembly where one or more of the plurality of oligonucleotides is bound to a support.

Figure 10 schematically depicts an exemplary process of oligonucleotide synthesis, polony amplification, sequencing, selective isolation of desired nucleic acids,
25 error filtration, cleavage and removal of PCR primer binding sites, and hierarchical assembly.

Figure 11 schematically depicts an exemplary optical system incorporating sequence detection and electromagnetic radiation delivery subsystems.

DETAILED DESCRIPTION

The present invention permits the sorting of individual oligonucleotide molecules determined to be more likely to have a desired sequence from individual oligonucleotide molecules determined to be less likely to have that sequence. The desired sequence can correspond to an error-free sequence the sequence intended when the oligonucleotide molecule was synthesized. The desired oligonucleotide molecule is sortable from other molecules, such as those containing one or more synthesis errors, because the invention permits the selective isolation of individual oligonucleotide molecules of individually determined sequences, whether by sequencing of that individual oligonucleotide molecule or by sequencing of a copy thereof. As used herein, one molecule is a "copy" of a second molecule if either is directly or indirectly copied from the other in a template-dependent process, such as PCR, or if both are copied from a common parental molecule. Thus, all molecules in a single PCR colony are "copies," as they each result from the amplification of a common parental molecule. Once sorted from other molecules, the individual oligonucleotide molecules determined to be more likely to have the desired sequence can be used for any purpose that benefits from increased accuracy, such as assembly into longer nucleic acids.

Oligonucleotides

The invention can be used to sort individual oligonucleotide molecules synthesized by any available means, such as chemical synthesis. A synthetic oligonucleotide can be of any suitable size, for example between 10 and 200 nucleotides long. The individual oligonucleotide molecules can be prepared by conventional phosphoramidite synthesis, which involves de-blocking a base bearing a protecting group; activating the base to be added and combining the two to form a phosphite linkage; capping any remaining de-blocked base; oxidizing the phosphite linkage to a phosphate linkage; and repeating until the oligonucleotide has reached the desired length. Phosphoramidite synthesis can achieve an error rate as low as one in one hundred for deletions and one in four hundred for mismatches and insertions and can be carried out

using a commercially available machine as is available, for example, from Integrated DNA Technologies. The resulting oligonucleotides can be desalted and can be purified, such as by polyacrylamide gel electrophoresis or by hybridization (Tian et al., (2004) Nature 432:1050-1054).

5 The individual oligonucleotide molecules can also be prepared by chip-based chemical syntheses. For example, oligonucleotides can be synthesized using a photoprogrammable microfluidic microarray (Zhou et al., (2004) Nucleic Acids Res. 32(18):5409-5417). As described by Zhou et al., the microarray can include thousands of individual reaction sites, each capable of synthesizing a different population of
10 oligonucleotides using standard phosphoramidite chemistry, except for the use of photogenerated acid in the deprotection step. For the deprotection step, digital photolithographic projection (e.g., using a Texas Instruments Digital Light Processor) is used to project a predetermined pattern of light on the microarray (e.g., 405 nm at 7 mW/cm2) in the presence of a photogenerated acid precursor solution.
15 Photolithographic syntheses of oligonucleotides have also been described elsewhere in the literature: see, for example, Richmond et al., ((2004) Nucleic Acids Res. 32(17): 5011-5018) and U.S. Patent Application Publication No. 2004-0126757 the complete contents of which are incorporated herein by reference. Richmond et al., used a 1024 x 768 array of micromirrors, each 16 µm wide, to project a UV image (365 nm from a
20 1000 W mercury arc lamp) of a virtual mask using a mono-hydroxysilane slide preferably coated with a base-labile linker as a solid support for oligonucleotide synthesis. The slide was located in a flow-cell reaction chamber connected to a DNA synthesizer.

An optical system for use in photo-programmable chip based syntheses
25 preferably incorporates micromirrors and/or an image locking subsystem, for example, as described in U.S. Patent Application Publication No. 2005-0249396 (Now U.S. Patent No. 7,072,500), the complete contents of which are incorporated herein by reference. An image locking subsystem that includes the use of detection or reference marks provides a feedback system to stabilize or lock the image with respect to an image capture device,
30 such as a camera and/or microscope. When a shift in image position is detected, a correction signal is sent to one of two mirrors, moving the image to correct for the shift

in image position. Using a feedback loop can permit optical stability with a drift of less than 0.5 μm over a period of hours, improving oligonucleotide synthesis accuracy and permitting the synthesis of longer oligonucleotides (e.g., up to 90 bases in length).

Sequencing

5 The invention permits the selective use of oligonucleotide molecules whose individual sequence has been queried. Sequence determinations can be made by any available method permitting the querying of the sequence of an individual molecule ("single molecule sequencing"), whether directly or through the querying of an amplified population of nucleic acids derived from a single molecule ("polony sequencing").
10 Generally, the method of sequence determination should be non-destructive, to the extent that the objective of the sequence determination is the identification of a subsequently useful oligonucleotide.

Methods of polymerase amplification and sequencing are described, for example, in U.S. Patent Application Nos. 2005-0079510 and 2006-0040297; in Mitra et al., (2003)
15 Analytical Biochemistry 320: 55-65; Shendure et al., (2005) Science 309:1728-1732; and in Margulies et al., (2005) Nature 437:376-380, the complete disclosures of each of which are herein incorporated by reference. Although some of these references discuss sequencing protocols in the context of genomic sequencing, the inventors have appreciated that the methods are also useful in oligonucleotide analysis. As discussed in
20 Shendure et al., (2005) Science 309:1729, polony amplification can involve, for example, in situ polonies, in situ rolling circle amplification, bridge PCR, picotiter PCR, or emulsion PCR. Generally, an oligonucleotide to be amplified is prepared to include primer binding sites, whether as part of its sequence when initially synthesized or by subsequent ligation to adaptor molecules bearing the primer binding sites.

25 In emulsion PCR, the oligonucleotides bearing primer binding sites are generally captured by beads bearing one of the PCR primers under conditions favoring binding at a stoichiometry of one oligonucleotide per bead. The beads are subsequently captured in a water-in-oil emulsion, the aqueous phase of which contains the reagents for a PCR reaction. In the method of Shendure et al., the aqueous:oil ratio is preferably about 1:6;
30 the nucleotide concentration is preferably about 3.5 mM; the MgC12 concentration is

preferably about 18.8 mM; Taq polymerase is preferably used at about 270 U; the extension time is preferably about 75 seconds; and PCR preferably continues for about 120 cycles. PCR amplification occurs within each water droplet, leading to the immobilization of many, substantially identical copies of the oligonucleotide on each bead. For beads that began with only one parental oligonucleotide sequence, the sequences of the nucleic acids on the beads at the end of the procedure differ only by the rate of PCR sequence infidelity, which is generally a rate substantially lower than the error rate from phosphoramidite synthesis. The beads are optionally paramagnetic, facilitating their purification from the emulsion. After PCR, the beads are optionally subjected to hybridization with larger, less dense beads bearing a sequence complementary to the amplicon to enrich the sample for successful amplification events.

Prior to sequencing, the oligonucleotides are generally separated from each other by, for example, placement in distinct wells; immobilization at distinct locations (e.g., predetermined, addressable locations or random locations) on a solid support; or immobilization on distinct solid supports. In the Genome Sequencer 20 System from 454 Life Sciences, for example, beads from polony amplification are deposited into wells of a fiber-optic slide. In the method of Shendure et al., beads from polony amplification are poured in a 5% acrylamide gel onto a glass coverslip manipulated to form a circular gel approximately 30 microns thick, giving a disordered monolayer.

If the oligonucleotides have been immobilized on a solid support, they can then be sequenced by any non-destructive method such as, for example, sequencing by hybridization. Other methods, such as sequencing by synthesis, permit the iterative interrogation of nucleobases of an oligonucleotide, which is advantageous when iterative interrogation provides a higher accuracy determination of sequence identity. For example, Margulies et al., describe a sequencing by synthesis technique in which the polony beads are sequenced in picoliter-sized wells using a pyrosequencing protocol. As another approach, Shendure et al., describe a four color sequencing by ligation method in which the identity of nucleobases is iteratively determined by ligation of anchor primers to second primers. The second primers are labeled with fluorescent dyes, the color of which identifies the nucleobase at one position in the primer; other positions are degenerate. Because ligation occurs only when the anchor primer and the second primer

are properly annealed, the color of the second primer identifies the nucleobase in the oligonucleotide at the position corresponding to the non-degenerate position of the second primer. By stripping the complexes and repeating the process with different populations of second primers in which the non-degenerate position varies, nucleobases
5 in the oligonucleotide can be iteratively identified. In yet another approach, Mitra et al., describe methods for sequencing polonies in parallel by fluorescent in situ sequential quantitation, "FISSEQ"; i.e., by performing repeated cycles of primer extension with reversibly-labeled fluorescent deoxynucleotides (for example, cycling sequentially through dATP, dCTP, dGTP, and dUTP or dTTP,). Incorporation of labeled dNTPs is
10 monitored using a scanning fluorescence microscope and software for automated image alignment and sequence calling. If a polony has incorporated a base, it will fluoresce, thereby identifying the template base immediately 3' of the primer. Once the incorporated base is identified, the dye linker is cleaved by a reducing agent (for example, by thiol reduction), or exposure to near UV light. Cleaved dye is washed away
15 and the cycle is repeated by adding a different dye-labeled base, washing away unincorporated dNTP, and scanning the gel. The sequence of the template nucleic acid is compiled as the primed template is interrogated at each cycle for incorporated nucleotide. The invention embraces any method of non-destructive sequencing. Non-limiting examples of non-destructive sequencing include pyrosequencing, as originally
20 described by Hyman et al., (1988, Anal Biochem 74: 324-436) and bead-based sequencing, described for instance by Leamon et al., (2004, Electrophoressi 24: 3769-3777). Non-destructive sequencing also includes methods using cleavable labeled oligonucleotides, as the above described Mitra et al., (2003, Anal Biochem 320:55-62) and photocleavable linkers (Seo et al., 2005, PNAS 102: 5926-5933). Methods using
25 reversible terminators are also embraced by the invention (Metzker et al,. 1994, NAR 22: 4259-4267). Further methods for non-destructive sequencing (including single molecule sequencing) are described in US patents 7,133,782 and US 7,169,560 which are hereby incorporated by reference.

Linkers

30 If oligonucleotides are synthesized or sequenced while attached to a solid support, the attachment is generally through a chemical linker. Although most standard

linkers are generally acceptable, cleavable linkers are preferred if an oligonucleotide is to be separated from the solid support (rather than, for example, being amplified without removal from the solid support). A cleavable linker permits the removal of an oligonucleotide from the substrate without destruction of the oligonucleotide itself. A cleavable linker may be cleaved or rendered susceptible to cleavage by any agent or stimulus of interest such as, for example, acidic conditions, basic conditions, nucleophiles, electromagnetic radiation, or any combination of the above. In some embodiments, a cleavable linker may be photo-cleavable. A cleavable linker can incorporate more than one cleavage site to permit, for example, a first cleavage of an oligonucleotide from a solid support used during synthesis of the oligonucleotide; attachment to a second solid support used during sequencing of the oligonucleotide; and subsequent cleavage of the oligonucleotide from the second solid support. An exemplary base-labile linker is depicted in FIG. 1 and is described in further detail in Richmond et al.

One useful class of linkers are referred to as "safety -catch" linkers, and incorporate a protecting group that is removed before the linker becomes labile. Thus, as one example, a safety -catch linker can incorporate a photocleavable protecting group that is removed before the linker becomes cleavable by treatment with a base. After removal of the protecting group by illumination at the appropriate wavelength, treatment with ammonium hydroxide or another base cleaves the linker. As another example, a safety-catch linker can incorporate an acid-cleavable protecting group that is removed before the linker becomes photocleavable. One group of safety-catch linkers are represented by the formula shown in FIG. 2. In the linker of FIG. 2, aromatic ring R2 is connected to a surface via an extender, which can be, for example, an unbranched alkyl group. The aromatic ring is connected via acid labile ketal R1 to a photocleavable protecting group (PCPG). A specific example of such a safety-catch linker, with an NVOC group as the photocleavable protecting group and a dimethoxy ketal linkage, is shown in FIG. 3. Following treatment of the linker with trifluoroacetic acid, converting the dimethoxy ketal linkage to a ketone linkage, the NVOC group is rendered photocleavable.

Selective isolation

Analysis of the sequence information from the oligonucleotides permits the identification of those oligonucleotides that appear to have desirable sequences and those that do not. Oligonucleotides of interest can then be selectively isolated from the rest of the population. If the oligonucleotides are in separate wells of a substrate, oligonucleotides can be taken selectively from the wells identified as containing oligonucleotides with desirable sequences. For example, in the apparatus of Margulies et al., polony beads are located in individual wells of a fiber-optic slide. Physical extraction of the bead from the appropriate well of the apparatus permits the subsequent amplification or purification of the desirable oligonucleotides free of other contaminating oligonucleotides. Alternatively, if the oligonucleotides are attached to the beads using a selectively cleavable linker, cleavage of the linker (e.g., by increasing the pH in the well to cleave a base-labile linker) followed by extraction of the solvent in the well can be used to selectively isolate the oligonucleotides without physical manipulation of the bead. Likewise, if the method of Shendure et al., is used, physical extraction of the beads or of the portions of the gel containing the oligonucleotides of interest can be used to selectively isolate desired oligonucleotides.

Certain other methods of selective isolation involve the targeting of oligonucleotide molecules without a requirement for physical manipulation of a solid support. Some such methods incorporate the use of an optical system to specifically target radiation to individual oligonucleotide molecules. One method based on the selective use of destructive radiation is depicted schematically in FIG. 4. Referring to FIG. 4, oligonucleotides of differing sequences are attached to one or more solid supports 10. The oligonucleotides include desired oligonucleotides 14 determined to contain a sequence of interest; the oligonucleotides also include other undesired oligonucleotides 12. Destructive radiation is selectively targeted against undesired oligonucleotides 12 (e.g., using micromirror technology) to destroy or disable them, leaving a population of oligonucleotides enriched for desired oligonucleotides 14. This enriched population can then be released from solid support 10 and/or amplified, e.g., by PCR.

Another method selectively releases desired or undesired oligonucleotides from one or more solid supports. An example of such a method is depicted schematically in FIG. 5. In FIG. 5, as in FIG. 4, desired oligonucleotides 14 and undesired oligonucleotides 12 are attached to one or more solid supports 10 via selectively cleavable linkers. In FIG. 5, agents triggering cleavage or lability of the selectively cleavable linkers are selectively delivered to desired oligonucleotides 14 determined to have a desirable sequence. Desired oligonucleotides 14 are selectively released from the one or more solid supports 10, leaving undesired oligonucleotides 12 bound to the solid support. Generally speaking, the agent can be delivered by any mechanism that permits preferential targeting to desired oligonucleotides 14. For example, if the agent is a chemical agent, such as an acid, a base, or a nucleophile, it can be delivered by inkjet technology. If the agent is radiation, it can be delivered by any mechanism that permits patterning of the radiation, whether assisted by a galvanometer, for example, or by micromirrors. Where the agent is electromagnetic radiation, one useful optical system is the optical system of Richmond et al.; the wavelength of the emitted light can of course be modified to correspond to the preferred wavelength for cleavage of the protecting group. In some embodiments, an optical system for use in selectively releasing one or more oligonucleotides (e.g., from a solid support such as a chip) may incorporate micromirrors and/or an image locking subsystem, for example, as described in U.S. Patent Application Publication No. 2005-0249396 (Now U.S. Patent No. 7,072,500), the entire disclosure of which is incorporated by reference herein. It should be appreciated that such optical systems also may be used to selectively destroy or damage one or more oligonucleotides (e.g., isolated on a solid support) using appropriately targeted electromagnetic radiation sufficient to damage or destroy the targeted oligonucleotide (e.g., alone or in the presence of an additional damaging or sensitizing agent). In some embodiments, a single molecule of an oligonucleotide of interest may be released and/or amplified (e.g., from each location on a solid support where an oligonucleotide having a correct or probably correct sequence is identified). In certain embodiments, a plurality of copies of an oligonucleotide of interest may be released and/or amplified.

Selective release permits the optional collection of a plurality of pools of oligonucleotides. The invention permits the synthesis, sequencing and sorting of many (e.g., hundreds, or thousands, or more) oligonucleotides in parallel through parallel chip-

based synthesis and parallel high-throughput sequencing. To maximize efficiency and minimize cost, many different oligonucleotide pools can be prepared at once. These different pools can be for totally unrelated purposes or, for example, can be intended for use in hierarchical assembly of a longer nucleic acid. In methods where the selective isolation involves physical isolation of a solid support or portion thereof, each isolate can be considered its own pool. In methods where selective isolation involves selective release from a solid support, the oligonucleotide to be released is preferably attached to the solid support by a selectively cleavable linker, such as a linker selectively cleavable following exposure to light, exposure to electricity, or other addressable stimulus. As one example, oligonucleotides can be coupled to a substrate using the base-labile linker with photocleavable protecting group of FIG. 1. Illumination of a first subset of oligonucleotides followed by base cleavage permits the elution of a first pool of oligonucleotides. Subsequent illumination of a second subset of oligonucleotides followed by base cleavage permits the elution of a second pool of oligonucleotides. This process can be repeated to generate any number of pools of oligonucleotides of interest.

Nucleic acid assembly

One particular use for oligonucleotides sorted and purified by the present methods is in the assembly of larger nucleic acids. A nucleic acid assembly process can incorporate one or more of the multiplex nucleic acid assembly procedures described below.

Multiplex Nucleic Acid Assembly

In aspects of the invention, multiplex nucleic acid assembly relates to the assembly of a plurality of nucleic acids to generate a longer nucleic acid product. In one aspect, multiplex oligonucleotide assembly relates to the assembly of a plurality of oligonucleotides to generate a longer nucleic acid molecule. However, it should be appreciated that other nucleic acids (e.g., single or double-stranded nucleic acid degradation products, restriction fragments, amplification products, naturally occurring small nucleic acids, other polynucleotides, etc.) may be assembled or included in a multiplex assembly reaction (e.g., along with one or more oligonucleotides) in order to generate an assembled nucleic acid molecule that is longer than any of the single starting

nucleic acids (e.g., oligonucleotides) that were added to the assembly reaction. In certain embodiments, one or more nucleic acid fragments that each were assembled in separate multiplex assembly reactions (e.g., separate multiplex oligonucleotide assembly reactions) may be combined and assembled to form a further nucleic acid that is longer than any of the input nucleic acid fragments. In certain embodiments, one or more nucleic acid fragments that each were assembled in separate multiplex assembly reactions (e.g., separate multiplex oligonucleotide assembly reactions) may be combined with one or more additional nucleic acids (e.g., single or double-stranded nucleic acid degradation products, restriction fragments, amplification products, naturally occurring small nucleic acids, other polynucleotides, etc.) and assembled to form a further nucleic acid that is longer than any of the input nucleic acids.

In aspects of the invention, one or more multiplex assembly reactions may be used to generate target nucleic acids having predetermined sequences. In one aspect, a target nucleic acid may have a sequence of a naturally occurring gene and/or other naturally occurring nucleic acid (e.g., a naturally occurring coding sequence, regulatory sequence, non-coding sequence, chromosomal structural sequence such as a telomere or centromere sequence, etc., any fragment thereof or any combination of two or more thereof). In another aspect, a target nucleic acid may have a sequence that is not naturally-occurring. In one embodiment, a target nucleic acid may be designed to have a sequence that differs from a natural sequence at one or more positions. In other embodiments, a target nucleic acid may be designed to have an entirely novel sequence. However, it should be appreciated that target nucleic acids may include one or more naturally occurring sequences, non-naturally occurring sequences, or combinations thereof.

In one aspect of the invention, multiplex assembly may be used to generate libraries of nucleic acids having different sequences. In some embodiments, a library may contain nucleic acids having random sequences. In certain embodiments, a predetermined target nucleic acid may be designed and assembled to include one or more random sequences at one or more predetermined positions.

In certain embodiments, a target nucleic acid may include a functional sequence (e.g., a protein binding sequence, a regulatory sequence, a sequence encoding a

functional protein, etc., or any combination thereof). However, some embodiments of a target nucleic acid may lack a specific functional sequence (e.g., a target nucleic acid may include only nonfunctional fragments or variants of a protein binding sequence, regulatory sequence, or protein encoding sequence, or any other non-functional naturally-occurring or synthetic sequence, or any non-functional combination thereof). Certain target nucleic acids may include both functional and non-functional sequences. These and other aspects of target nucleic acids and their uses are described in more detail herein.

A target nucleic acid may be assembled in a single multiplex assembly reaction (e.g., a single oligonucleotide assembly reaction). However, a target nucleic acid also may be assembled from a plurality of nucleic acid fragments, each of which may have been generated in a separate multiplex oligonucleotide assembly reaction. It should be appreciated that one or more nucleic acid fragments generated via multiplex oligonucleotide assembly also may be combined with one or more nucleic acid molecules obtained from another source (e.g., a restriction fragment, a nucleic acid amplification product, etc.) to form a target nucleic acid. In some embodiments, a target nucleic acid that is assembled in a first reaction may be used as an input nucleic acid fragment for a subsequent assembly reaction to produce a larger target nucleic acid.

Accordingly, different strategies may be used to produce a target nucleic acid having a predetermined sequence. For example, different starting nucleic acids (e.g., different sets of predetermined nucleic acids) may be assembled to produce the same predetermined target nucleic acid sequence. Also, predetermined nucleic acid fragments may be assembled using one or more different in vitro and/or in vivo techniques. For example, nucleic acids (e.g., overlapping nucleic acid fragments) may be assembled in an in vitro reaction using an enzyme (e.g., a ligase and/or a polymerase) or a chemical reaction (e.g., a chemical ligation) or in vivo (e.g., assembled in a host cell after transfection into the host cell), or a combination thereof. Similarly, each nucleic acid fragment that is used to make a target nucleic acid may be assembled from different sets of oligonucleotides. Also, a nucleic acid fragment may be assembled using an in vitro or an in vivo technique (e.g., an in vitro or in vivo polymerase, recombinase, and/or ligase based assembly process). In addition, different in vitro assembly reactions may be used

to produce a nucleic acid fragment. For example, an in vitro oligonucleotide assembly reaction may involve one or more polymerases, ligases, other suitable enzymes, chemical reactions, or any combination thereof.

Multiplex oligonucleotide assembly

5 A predetermined nucleic acid fragment may be assembled from a plurality of different starting nucleic acids (e.g., oligonucleotides) in a multiplex assembly reaction (e.g., a multiplex enzyme-mediated reaction, a multiplex chemical assembly reaction, or a combination thereof). Certain aspects of multiplex nucleic acid assembly reactions are illustrated by the following description of certain embodiments of multiplex 10 oligonucleotide assembly reactions. It should be appreciated that the description of the assembly reactions in the context of oligonucleotides is not intended to be limiting. The assembly reactions described herein may be performed using starting nucleic acids obtained from one or more different sources (e.g., synthetic or natural polynucleotides, nucleic acid amplification products, nucleic acid degradation products, oligonucleotides, 15 etc.). The starting nucleic acids may be referred to as assembly nucleic acids (e.g., assembly oligonucleotides). As used herein, an assembly nucleic acid has a sequence that is designed to be incorporated into the nucleic acid product generated during the assembly process. However, it should be appreciated that the description of the assembly reactions in the context of single-stranded nucleic acids is not intended to 20 be limiting. In some embodiments, one or more of the starting nucleic acids illustrated in the figures and described herein may be provided as double stranded nucleic acids. Accordingly, it should be appreciated that where the figures and description illustrate the assembly of single-stranded nucleic acids, the presence of one or more complementary nucleic acids is contemplated. Accordingly, one or more double -stranded 25 complementary nucleic acids may be included in a reaction that is described herein in the context of a single-stranded assembly nucleic acid. However, in some embodiments the presence of one or more complementary nucleic acids may interfere with an assembly reaction by competing for hybridization with one of the input assembly nucleic acids. Accordingly, in some embodiments an assembly reaction may involve only single-30 stranded assembly nucleic acids (i.e., the assembly nucleic acids may be provided in a single-stranded form without their complementary strand) as described or illustrated

herein. However, in certain embodiments the presence of one or more complementary nucleic acids may have no or little effect on the assembly reaction. In some embodiments, complementary nucleic acid(s) may be incorporated during one or more steps of an assembly. In yet further embodiments, assembly nucleic acids and their complementary strands may be assembled under the same assembly conditions via parallel assembly reactions in the same reaction mixture. In certain embodiments, a nucleic acid product resulting from the assembly of a plurality of starting nucleic acids may be identical to the nucleic acid product that results from the assembly of nucleic acids that are complementary to the starting nucleic acids (e.g., in some embodiments where the assembly steps result in the production of a double-stranded nucleic acid product). As used herein, an oligonucleotide may be a nucleic acid molecule comprising at least two covalently bonded nucleotide residues. In some embodiments, an oligonucleotide may be between 10 and 1,000 nucleotides long. For example, an oligonucleotide may be between 10 and 500 nucleotides long, or between 500 and 1,000 nucleotides long. In some embodiments, an oligonucleotide may be between about 20 and about 100 nucleotides long (e.g., from about 30 to 90, 40 to 85, 50 to 80, 60 to 75, or about 65 or about 70 nucleotides long), between about 100 and about 200, between about 200 and about 300 nucleotides, between about 300 and about 400, or between about 400 and about 500 nucleotides long. However, shorter or longer oligonucleotides may be used. An oligonucleotide may be a single-stranded nucleic acid. However, in some embodiments a double-stranded oligonucleotide may be used as described herein. In certain embodiments, an oligonucleotide may be chemically synthesized as described in more detail below.

In some embodiments, an input nucleic acid (e.g., oligonucleotide) may be amplified before use. The resulting product may be double-stranded. In some embodiments, one of the strands of a double-stranded nucleic acid may be removed before use so that only a predetermined single strand is added to an assembly reaction. In certain embodiments, each oligonucleotide may be designed to have a sequence that is identical to a different portion of the sequence of a predetermined target nucleic acid that is to be assembled. Accordingly, in some embodiments each oligonucleotide may have a sequence that is identical to a portion of one of the two strands of a double-stranded target nucleic acid. For clarity, the two complementary strands of a double stranded

nucleic acid are referred to herein as the positive (P) and negative (N) strands. This designation is not intended to imply that the strands are sense and anti-sense strands of a coding sequence. They refer only to the two complementary strands of a nucleic acid (e.g., a target nucleic acid, an intermediate nucleic acid fragment, etc.) regardless of the
5 sequence or function of the nucleic acid. Accordingly, in some embodiments a P strand may be a sense strand of a coding sequence, whereas in other embodiments a P strand may be an antisense strand of a coding sequence. According to the invention, a target nucleic acid may be either the P strand, the N strand, or a double-stranded nucleic acid comprising both the P and N strands. It should be appreciated that the reference to
10 complementary nucleic acids or complementary nucleic acid regions herein refers to nucleic acids or regions thereof that have sequences which are reverse complements of each other so that they can hybridize in an antiparallel fashion typical of natural DNA.

It should be appreciated that different oligonucleotides may be designed to have different lengths. In some embodiments, one or more different oligonucleotides may
15 have overlapping sequence regions (e.g., overlapping 5' regions or overlapping 3' regions). Overlapping sequence regions may be identical (i.e., corresponding to the same strand of the nucleic acid fragment) or complementary (i.e., corresponding to complementary strands of the nucleic acid fragment). The plurality of oligonucleotides may include one or more oligonucleotide pairs with overlapping identical sequence
20 regions, one or more oligonucleotide pairs with overlapping complementary sequence regions, or a combination thereof. Overlapping sequences may be of any suitable length. For example, overlapping sequences may encompass the entire length of one or more nucleic acids used in an assembly reaction. Overlapping sequences may be between about 5 and about 500 oligonucleotides long (e.g., between about 10 and 100, between
25 about 10 and 75, between about 10 and 50, about 20, about 25, about 30, about 35, about 45, about 50, etc.). However, shorter, longer, or intermediate overlapping lengths may be used. It should be appreciated that overlaps between different input nucleic acids used in an assembly reaction may have different lengths.

In a multiplex oligonucleotide assembly reaction designed to generate a
30 predetermined nucleic acid fragment, the combined sequences of the different oligonucleotides in the reaction may span the sequence of the entire nucleic acid

fragment on either the positive strand, the negative strand, both strands, or a combination of portions of the positive strand and portions of the negative strand. The plurality of different oligonucleotides may provide either positive sequences, negative sequences, or a combination of both positive and negative sequences corresponding to the entire sequence of the nucleic acid fragment to be assembled. In some embodiments, the plurality of oligonucleotides may include one or more oligonucleotides having sequences identical to one or more portions of the positive sequence, and one or more oligonucleotides having sequences that are identical to one or more portions of the negative sequence of the nucleic acid fragment. One or more pairs of different oligonucleotides may include sequences that are identical to overlapping portions of the predetermined nucleic acid fragment sequence as described herein (e.g., overlapping sequence portions from the same or from complementary strands of the nucleic acid fragment). In some embodiments, the plurality of oligonucleotides includes a set of oligonucleotides having sequences that combine to span the entire positive sequence and a set oligonucleotides having sequences that combine to span the entire negative sequence of the predetermined nucleic acid fragment. However, in certain embodiments, the plurality of oligonucleotides may include one or more oligonucleotides with sequences that are identical to sequence portions on one strand (either the positive or negative strand) of the nucleic acid fragment, but no oligonucleotides with sequences that are complementary to those sequence portions. In one embodiment, a plurality of oligonucleotides includes only oligonucleotides having sequences identical to portions of the positive sequence of the predetermined nucleic acid fragment. In one embodiment, a plurality of oligonucleotides includes only oligonucleotides having sequences identical to portions of the negative sequence of the predetermined nucleic acid fragment. These oligonucleotides may be assembled by sequential ligation or in an extension-based reaction (e.g., if an oligonucleotide having a 3' region that is complementary to one of the plurality of oligonucleotides is added to the reaction).

In one aspect, a nucleic acid fragment may be assembled in a polymerase mediated assembly reaction from a plurality of oligonucleotides that are combined and extended in one or more rounds of polymerase-mediated extensions. In another aspect, a nucleic acid fragment may be assembled in a ligase-mediated reaction from a plurality of oligonucleotides that are combined and ligated in one or more rounds of ligase-mediated

ligations. In another aspect, a nucleic acid fragment may be assembled in a non-enzymatic reaction (e.g., a chemical reaction) from a plurality of oligonucleotides that are combined and assembled in one or more rounds of non-enzymatic reactions. In some embodiments, a nucleic acid fragment may be assembled using a combination of polymerase, ligase, and/or non-enzymatic reactions. For example, both polymerase(s) and ligase(s) may be included in an assembly reaction mixture. Accordingly, a nucleic acid may be assembled via coupled amplification and ligation or ligation during amplification. The resulting nucleic acid fragment from each assembly technique may have a sequence that includes the sequences of each of the plurality of assembly oligonucleotides that were used as described herein. These assembly reactions may be referred to as primerless assemblies, since the target nucleic acid is generated by assembling the input oligonucleotides rather than being generated in an amplification reaction where the oligonucleotides act as amplification primers to amplify a pre-existing template nucleic acid molecule corresponding to the target nucleic acid.

Polymerase-based assembly techniques may involve one or more suitable polymerase enzymes that can catalyze a template-based extension of a nucleic acid in a 5' to 3' direction in the presence of suitable nucleotides and an annealed template. A polymerase may be thermostable. A polymerase may be obtained from recombinant or natural sources. In some embodiments, a thermostable polymerase from a thermophilic organism may be used. In some embodiments, a polymerase may include a 3'--* 5' exonuclease/proofreading activity. In some embodiments, a polymerase may have no, or little, proofreading activity (e.g., a polymerase may be a recombinant variant of a natural polymerase that has been modified to reduce its proofreading activity). Examples of thermostable DNA polymerases include, but are not limited to: Taq (a heat-stable DNA polymerase from the bacterium Thermus aquaticus); Pfu (a thermophilic DNA polymerase with a 3'-). 5' exonuclease/proofreading activity from Pyrococcus furiosus, available from for example Promega); VentR® DNA Polymerase and VentRO (exo-) DNA Polymerase (thermophilic DNA polymerases with or without a 3'--+ 5' exonuclease/proofreading activity from Thermococcus litoralis; also known as Th polymerase); Deep VentR® DNA Polymerase and Deep VentR® (exo-) DNA Polymerase (thermophilic DNA polymerases with or without a 3'-> 5' exonuclease/proofreading activity from Pyrococcus species GB-D; available from New

England Biolabs); KOD HiFi (a recombinant Thermococcus kodakaraensis KODI DNA polymerase with a 3'--* 5'exonuclease/proofreading activity, available from Novagen,); BIO-X-ACT (a mix of polymerases that possesses 5'-3' DNA polymerase activity and 3'- + 5' proofreading activity); Klenow Fragment (an N-terminal truncation of E. coli DNA Polymerase I which retains polymerase activity, but has lost the 5'-* 3' exonuclease activity, available from, for example, Promega and NEB); SequenaseTM (T7 DNA polymerase deficient in T-5' exonuclease activity); Phi29 (bacteriophage 29 DNA polymerase, may be used for rolling circle amplification, for example, in a TempliPhiTM DNA Sequencing Template Amplification Kit, available from Amersham Biosciences); TopoTaq (a hybrid polymerase that combines hyperstable DNA binding domains and the DNA unlinking activity of Methanopyrus topoisomerase, with no exonuclease activity, available from Fidelity Systems); TopoTaq HiFi which incorporates a proofreading domain with exonuclease activity; PhusionTM (a Pyrococcus-like enzyme with a processivity-enhancing domain, available from New England Biolabs); any other suitable DNA polymerase, or any combination of two or more thereof.

Ligase-based assembly techniques may involve one or more suitable ligase enzymes that can catalyze the covalent linking of adjacent 3' and 5' nucleic acid termini (e.g., a 5' phosphate and a 3' hydroxyl of nucleic acid(s) annealed on a complementary template nucleic acid such that the 3' terminus is immediately adjacent to the 5' terminus). Accordingly, a ligase may catalyze a ligation reaction between the 5' phosphate of a first nucleic acid to the 3' hydroxyl of a second nucleic acid if the first and second nucleic acids are annealed next to each other on a template nucleic acid). A ligase may be obtained from recombinant or natural sources. A ligase may be a heat-stable ligase. In some embodiments, a thermostable ligase from a thermophilic organism may be used. Examples of thermostable DNA ligases include, but are not limited to: Tth DNA ligase (from Thermus thermophilus, available from, for example, Eurogentec and GeneCraft); Pfu DNA ligase (a hyperthermophilic ligase from Pyrococcus furiosus); Taq ligase (from Thermus aquaticus), any other suitable heat-stable ligase, or any combination thereof. In some embodiments, one or more lower temperature ligases may be used (e.g., T4 DNA ligase). A lower temperature ligase may be useful for shorter overhangs (e.g., about 3, about 4, about 5, or about 6 base overhangs) that may not be stable at higher temperatures.

Non-enzymatic techniques can be used to ligate nucleic acids. For example, a 5'-end (e.g., the 5' phosphate group) and a 3'-end (e.g., the 3' hydroxyl) of one or more nucleic acids may be covalently linked together without using enzymes (e.g., without using a ligase). In some embodiments, non-enzymatic techniques may offer certain advantages over enzyme-based ligations. For example, non-enzymatic techniques may have a high tolerance of non-natural nucleotide analogues in nucleic acid substrates, may be used to ligate short nucleic acid substrates, may be used to ligate RNA substrates, and/or may be cheaper and/or more suited to certain automated (e.g., high throughput) applications.

Non-enzymatic ligation may involve a chemical ligation. In some embodiments, nucleic acid termini of two or more different nucleic acids may be chemically ligated. In some embodiments, nucleic acid termini of a single nucleic acid may be chemically ligated (e.g., to circularize the nucleic acid). It should be appreciated that both strands at a first double-stranded nucleic acid terminus may be chemically ligated to both strands at a second double-stranded nucleic acid terminus. However, in some embodiments only one strand of a first nucleic acid terminus may be chemically ligated to a single strand of a second nucleic acid terminus. For example, the 5' end of one strand of a first nucleic acid terminus may be ligated to the 3' end of one strand of a second nucleic acid terminus without the ends of the complementary strands being chemically ligated.

Accordingly, a chemical ligation may be used to form a covalent linkage between a 5' terminus of a first nucleic acid end and a 3' terminus of a second nucleic acid end, wherein the first and second nucleic acid ends may be ends of a single nucleic acid or ends of separate nucleic acids. In one aspect, chemical ligation may involve at least one nucleic acid substrate having a modified end (e.g., a modified 5' and/or 3' terminus) including one or more chemically reactive moieties that facilitate or promote linkage formation. In some embodiments, chemical ligation occurs when one or more nucleic acid termini are brought together in close proximity (e.g., when the termini are brought together due to annealing between complementary nucleic acid sequences). Accordingly, annealing between complementary 3' or 5' overhangs (e.g., overhangs generated by restriction enzyme cleavage of a double-stranded nucleic acid) or between any combination of complementary nucleic acids that results in a 3' terminus being

brought into close proximity with a 5' terminus (e.g., the 3' and 5' termini are adjacent to each other when the nucleic acids are annealed to a complementary template nucleic acid) may promote a template-directed chemical ligation. Examples of chemical reactions may include, but are not limited to, condensation, reduction, and/or photo-chemical ligation reactions. It should be appreciated that in some embodiments chemical ligation can be used to produce naturally occurring phosphodiester internucleotide linkages, non-naturally-occurring phosphamide pyrophosphate internucleotide linkages, and/or other non-naturally-occurring internucleotide linkages.

In some embodiments, the process of chemical ligation may involve one or more coupling agents to catalyze the ligation reaction. A coupling agent may promote a ligation reaction between reactive groups in adjacent nucleic acids (e.g., between a 5'-reactive moiety and a 3'-reactive moiety at adjacent sites along a complementary template). In some embodiments, a coupling agent may be a reducing reagent (e.g., ferricyanide), a condensing reagent such (e.g., cyanoimidazole, cyanogen bromide, carbodiimide, etc.), or irradiation (e.g., UV irradiation for photo-ligation).

In some embodiments, a chemical ligation may be an autoligation reaction that does not involve a separate coupling agent. In autoligation, the presence of a reactive group on one or more nucleic acids may be sufficient to catalyze a chemical ligation between nucleic acid termini without the addition of a coupling agent (see, for example, Xu et al., (1997) Tetrahedron Lett. 38:5595-8). Non-limiting examples of these reagent-free ligation reactions may involve nucleophilic displacements of sulfur on bromoacetyl, tosyl, or iodo-nucleoside groups (see, for example, Xu et al., (2001) Nat. Biotech. 19:148-52). Nucleic acids containing reactive groups suitable for autoligation can be prepared directly on automated synthesizers (see, for example, Xu et al., (1999) Nuc. Acids Res. 27:875-81). In some embodiments, a phosphorothioate at a 3' terminus may react with a leaving group (such as tosylate or iodide) on a thymidine at an adjacent 5' terminus. In some embodiments, two nucleic acid strands bound at adjacent sites on a complementary target strand may undergo auto-ligation by displacement of a 5'-end iodide moiety (or tosylate) with a 3'-end sulfur moiety. Accordingly, in some embodiments the product of an autoligation may include a non-naturally-occurring

internucleotide linkage (e.g., a single oxygen atom may be replaced with a sulfur atom in the ligated product).

In some embodiments, a synthetic nucleic acid duplex can be assembled via chemical ligation in a one step reaction involving simultaneous chemical ligation of nucleic acids on both strands of the duplex. For example, a mixture of 5'-phosphorylated oligonucleotides corresponding to both strands of a target nucleic acid may be chemically ligated by a) exposure to heat (e.g., to 97 °C) and slow cooling to form a complex of annealed oligonucleotides, and b) exposure to cyanogen bromide or any other suitable coupling agent under conditions sufficient to chemically ligate adjacent 3' and 5' ends in the nucleic acid complex.

In some embodiments, a synthetic nucleic acid duplex can be assembled via chemical ligation in a two step reaction involving separate chemical ligations for the complementary strands of the duplex. For example, each strand of a target nucleic acid may be ligated in a separate reaction containing phosphorylated oligonucleotides corresponding to the strand that is to be ligated and non-phosphorylated oligonucleotides corresponding to the complementary strand. The non-phosphorylated oligonucleotides may serve as a template for the phosphorylated oligonucleotides during a chemical ligation (e.g., using cyanogen bromide). The resulting single-stranded ligated nucleic acid may be purified and annealed to a complementary ligated single-stranded nucleic acid to form the target duplex nucleic acid (see, for example, Shabarova et al., (1991) Nucl. Acids Res. 19:4247-51).

Aspects of the invention may be used to enhance different types of nucleic acid assembly reactions (e.g., multiplex nucleic acid assembly reactions). Aspects of the invention may be used in combination with one or more assembly reactions described in, for example, Carr et al., (2004) Nucl. Acids Res., 32(20), e162 (9 pages); Richmond et al., (2004) Nucl. Acids Res., 32(17):5011-5018; Caruthers et al., (1972) J. Mol. Biol. 72:475-492; Hecker et al., (1998) Biotechniques 24:256-260; Kodumal et al., (2004) PNAS 101(44):15573-15578; Tian et al., (2004) Nature 432:1050-1054; and US Patent Nos. 6,008,031 and 5,922,539, the disclosures of which are incorporated herein by reference. Certain embodiments of multiplex nucleic acid assembly reactions for generating a predetermined nucleic acid fragment are illustrated with reference to FIGS.

6-9. It should be appreciated that synthesis and assembly methods described herein (including, for example, oligonucleotide synthesis, multiplex nucleic acid assembly, or any combination thereof) may be performed in any suitable format, including in a reaction tube, in a multi-well plate, on a surface, on a column, in a microfluidic device (e.g., a microfluidic tube), a capillary tube, etc.

FIG. 6 shows one embodiment of a plurality of oligonucleotides that may be assembled in a polymerase-based multiplex oligonucleotide assembly reaction. FIG. 6A shows two groups of oligonucleotides (Group P and Group N) that have sequences of portions of the two complementary strands of a nucleic acid fragment to be assembled. Group P includes oligonucleotides with positive strand sequences (P1, P2, ... Pn-1, Pn, Pn+1, ...PT, shown from 5'43' on the positive strand). Group N includes oligonucleotides with negative strand sequences (NT, ..., Nn+l, Nn, Nn-1, ..., N2, N1, shown from 543' on the negative strand). In this example, none of the P group oligonucleotides overlap with each other and none of the N group oligonucleotides overlap with each other. However, in some embodiments, one or more of the oligonucleotides within the S or N group may overlap. Furthermore, FIG. 6A shows gaps between consecutive oligonucleotides in Group P and gaps between consecutive oligonucleotides in Group N. However, each P group oligonucleotide (except for P 1) and each N group oligonucleotide (except for NT) overlaps with complementary regions of two oligonucleotides from the complementary group of oligonucleotides. P 1 and NT overlap with a complementary region of only one oligonucleotide from the other group (the complementary 3'-most oligonucleotides N1 and PT, respectively). FIG. 6B shows a structure of an embodiment of a Group P or Group N oligonucleotide represented in FIG. 6A. This oligonucleotide includes a 5' region that is complementary to a 5' region of a first oligonucleotide from the other group, a 3' region that is complementary to a 3' region of a second oligonucleotide from the other group, and a core or central region that is not complementary to any oligonucleotide sequence from the other group (or its own group). This central region is illustrated as the B region in FIG. 6B. The sequence of the B region may be different for each different oligonucleotide. As defined herein, the B region of an oligonucleotide in one group corresponds to a gap between two consecutive oligonucleotides in the complementary group of oligonucleotides. It should be noted that the 5'-most oligonucleotide in each group (P1 in Group P and NT in Group N) does not

have a 5' region that is complementary to the 5' region of any other oligonucleotide in either group. Accordingly, the 5'-most oligonucleotides (P1 and NT) that are illustrated in FIG. 6A each have a 3' complementary region and a 5' non-complementary region (the B region of FIG. 6B), but no 5' complementary region. However, it should be appreciated that any one or more of the oligonucleotides in Group P and/or Group N (including all of the oligonucleotides in Group P and/or Group N) can be designed to have no B region. In the absence of a B region, a 5'-most oligonucleotide has only the 3' complementary region (meaning that the entire oligonucleotide is complementary to the 3' region of the 3'-most oligonucleotide from the other group (e.g., the 3' region of Nl or PT shown in FIG. 6A). In the absence of a B region, one of the other oligonucleotides in either Group P or Group N has only a 5' complementary region and a 3' complementary region (meaning that the entire oligonucleotide is complementary to the 5' and 3' sequence regions of the two overlapping oligonucleotides from the complementary group). In some embodiments, only a subset of oligonucleotides in an assembly reaction may include B regions. It should be appreciated that the length of the 5', 3', and B regions may be different for each oligonucleotide. However, for each oligonucleotide the length of the 5' region is the same as the length of the complementary 5' region in the 5' overlapping oligonucleotide from the other group. Similarly, the length of the 3' region is the same as the length of the complementary 3' region in the 3' overlapping oligonucleotide from the other group. However, in certain embodiments a 3'-most oligonucleotide may be designed with a 3' region that extends beyond the 5' region of the 5'-most oligonucleotide. In this embodiment, an assembled product may include the 5' end of the 5'-most oligonucleotide, but not the 3' end of the 3'-most oligonucleotide that extends beyond it.

FIG. 6C illustrates a subset of the oligonucleotides from FIG. 6A, each oligonucleotide having a 5', a 3', and an optional B region. Oligonucleotide Pn is shown with a 5' region that is complementary to (and can anneal to) the 5' region of oligonucleotide Nn-1. Oligonucleotide Pn also has a 3' region that is complementary to (and can anneal to) the 3' region of oligonucleotide Nn. Nn is also shown with a 5' region that is complementary (and can anneal to) the 5' region of oligonucleotide Pn+l. This pattern could be repeated for all of oligonucleotides P2 to PT and N1 to NT-1 (with the 5'-most oligonucleotides only having 3' complementary regions as discussed herein).

If all of the oligonucleotides from Group P and Group N are mixed together under appropriate hybridization conditions, they may anneal to form a long chain such as the oligonucleotide complex illustrated in FIG. 6A. However, subsets of the oligonucleotides may form shorter chains and even oligonucleotide dimers with annealed 5' or 3' regions. It should be appreciated that many copies of each oligonucleotide are included in a typical reaction mixture. Accordingly, the resulting hybridized reaction mixture may contain a distribution of different oligonucleotide dimers and complexes. Polymerase-mediated extension of the hybridized oligonucleotides results in a template-based extension of the 3' ends of oligonucleotides that have annealed 3' regions. Accordingly, polymerase-mediated extension of the oligonucleotides shown in FIG. 6C would result in extension of the 3' ends only of oligonucleotides Pn and Nn generating extended oligonucleotides containing sequences that are complementary to all the regions of Nn and Pn, respectively. Extended oligonucleotide products with sequences complementary to all of Nn-1 and Pn+1 would not be generated unless oligonucleotides Pn-1 and Nn+1 were included in the reaction mixture. Accordingly, if all of the oligonucleotide sequences in a plurality of oligonucleotides are to be incorporated into an assembled nucleic acid fragment using a polymerase, the plurality of oligonucleotides should include 5'-most oligonucleotides that are at least complementary to the entire 3' regions of the 3'-most oligonucleotides. In some embodiments, the 5 '-most oligonucleotides also may have 5' regions that extend beyond the 3' ends of the 3'-most oligonucleotides as illustrated in FIG. 6A. In some embodiments, a ligase also maybe added to ligate adjacent 5' and 3' ends that may be formed upon 3' extension of annealed oligonucleotides in an oligonucleotide complex such as the one illustrated in FIG. 6A.

When assembling a nucleic acid fragment using a polymerase, a single cycle of polymerase extension extends oligonucleotide pairs with annealed 3' regions. Accordingly, if a plurality of oligonucleotides were annealed to form an annealed complex such as the one illustrated in FIG. 6A, a single cycle of polymerase extension would result in the extension of the 3' ends of the P1/Nl, P2/N2, ..., Pn-1/Nn-1, Pn/Nn, Pn+l/Nn+l, ..., PT/NT oligonucleotide pairs. In one embodiment, a single molecule could be generated by ligating the extended oligonucleotide dimers. In one embodiment, a single molecule incorporating all of the oligonucleotide sequences may be generated by performing several polymerase extension cycles.

In one embodiment, FIG. 6D illustrates two cycles of polymerase extension (separated by a denaturing step and an annealing step) and the resulting nucleic acid products. It should be appreciated that several cycles of polymerase extension may be required to assemble a single nucleic acid fragment containing all the sequences of an initial plurality of oligonucleotides. In one embodiment, a minimal number of extension cycles for assembling a nucleic acid may be calculated as log2n, where n is the number of oligonucleotides being assembled. In some embodiments, progressive assembly of the nucleic acid may be achieved without using temperature cycles. For example, an enzyme capable of rolling circle amplification may be used (e.g., phi 29 polymerase) when a circularized nucleic acid (e.g., oligonucleotide) complex is used as a template to produce a large amount of circular product for subsequent processing using MutS or a MutS homolog as described herein. In step 1 of FIG. 6D, annealed oligonucleotide pairs Pn/Nn and Pn+ l/Nn+l are extended to form oligonucleotide dimer products incorporating the sequences covered by the respective oligonucleotide pairs. For example, Pn is extended to incorporate sequences that are complementary to the B and 5' regions of Nn (indicated as N'n in FIG. 6D). Similarly, Nn+l is extended to incorporate sequences that are complementary to the 5' and B regions of Pn+ 1 (indicated as P'n+1 in FIG. 6D). These dimer products may be denatured and reannealed to form the starting material of step 2 where the 3' end of the extended Pn oligonucleotide is annealed to the 3' end of the extended Nn+1 oligonucleotide. This product may be extended in a polymerase-mediated reaction to form a product that incorporates the sequences of the four oligonucleotides (Pn, Nn, Pn +1, Nn+l). One strand of this extended product has a sequence that includes (in 5' to 3' order) the 5', B, and 3' regions of P,,, the complement of the B region of Nn, the 5', B, and 3' regions of Pn+l, and the complements of the B and 5' regions of Nn+1. The other strand of this extended product has the complementary sequence. It should be appreciated that the 3' regions of Pn and Nn are complementary, the 5' regions of Nn and Pn+1 are complementary, and the 3' regions of Pn+1 and Nn+1 are complementary. It also should be appreciated that the reaction products shown in FIG. 6D are a subset of the reaction products that would be obtained using all of the oligonucleotides of Group P and Group N. A first polymerase extension reaction using all of the oligonucleotides would result in a plurality of overlapping oligonucleotide dimers from P1/N1 to PT/NT. Each of these may be denatured and at

least one of the strands could then anneal to an overlapping complementary strand from an adjacent (either 3' or 5') oligonucleotide dimer and be extended in a second cycle of polymerase extension as shown in FIG. 6D. Subsequent cycles of denaturing, annealing, and extension produce progressively larger products including a nucleic acid fragment that includes the sequences of all of the initial oligonucleotides. It should be appreciated that these subsequent rounds of extension also produce many nucleic acid products of intermediate length. The reaction product may be complex since not all of the 3' regions may be extended in each cycle. Accordingly, unextended oligonucleotides may be available in each cycle to anneal to other unextended oligonucleotides or to previously extended oligonucleotides. Similarly, extended products of different sizes may anneal to each other in each cycle. Accordingly, a mixture of extended products of different sizes covering different regions of the sequence may be generated along with the nucleic acid fragment covering the entire sequence. This mixture also may contain any remaining unextended oligonucleotides.

FIG. 7 shows an embodiment of a plurality of oligonucleotides that may be assembled in a directional polymerase-based multiplex oligonucleotide assembly reaction. In this embodiment, only the 5'-most oligonucleotide of Group P may be provided. In contrast to the example shown in FIG. 7, the remainder of the sequence of the predetermined nucleic acid fragment is provided by oligonucleotides of Group N. The 3'-most oligonucleotide of Group N (N1) has a 3' region that is complementary to the 3' region of P1 as shown in FIG. 7B. However, the remainder of the oligonucleotides in Group N have overlapping (but noncomplementary) 3' and 5' regions as illustrated in FIG. 7B for oligonucleotides Nl-N3. Each Group N oligonucleotide (e.g., Nn) overlaps with two adjacent oligonucleotides: one overlaps with the 3' region (Nn-1) and one with the 5' region (Nn+1), except for N1 that overlaps with the 3' regions of P1 (complementary overlap) and N2 (non-complementary overlap), and NT that overlaps only with NT-1. It should be appreciated that all of the overlaps shown in FIG. 7A between adjacent oligonucleotides N2 to NT-1 are non-complementary overlaps between the 5' region of one oligonucleotide and the 3' region of the adjacent oligonucleotide illustrated in a 3' to 5' direction on the N strand of the predetermined nucleic acid fragment. It also should be appreciated that each oligonucleotide may have 3', B, and 5'regions of different lengths (including no B region in some embodiments). In some

embodiments, none of the oligonucleotides may have B regions, meaning that the entire sequence of each oligonucleotide may overlap with the combined 5' and 3' region sequences of its two adjacent oligonucleotides.

Assembly of a predetermined nucleic acid fragment from the plurality of oligonucleotides shown in FIG. 7A may involve multiple cycles of polymerase-mediated extension. Each extension cycle may be separated by a denaturing and an annealing step. FIG. 7C illustrates the first two steps in this assembly process. In step 1, annealed oligonucleotides P1 and N1 are extended to form an oligonucleotide dimer. P1 is shown with a 5' region that is noncomplementary to the 3' region of N1 and extends beyond the 3' region of N1 when the oligonucleotides are annealed. However, in some embodiments, P1 may lack the 5' noncomplementary region and include only sequences that overlap with the 3' region of N1. The product of P1 extension is shown after step 1 containing an extended region that is complementary to the 5' end of N1. The single strand illustrated in FIG. 7C may be obtained by denaturing the oligonucleotide dimer that results from the extension of P1/N1 in step 1. The product of P 1 extension is shown annealed to the 3' region of N2. This annealed complex may be extended in step 2 to generate an extended product that now includes sequences complementary to the B and 5' regions of N2. Again, the single strand illustrated in FIG. 7C may be obtained by denaturing the oligonucleotide dimer that results from the extension reaction of step 2. Additional cycles of extension may be performed to further assemble a predetermined nucleic acid fragment. In each cycle, extension results in the addition of sequences complementary to the B and 5' regions of the next Group N oligonucleotide. Each cycle may include a denaturing and annealing step. However, the extension may occur under the annealing conditions. Accordingly, in one embodiment, cycles of extension may be obtained by alternating between denaturing conditions (e.g., a denaturing temperature) and annealing/extension conditions (e.g., an annealing/extension temperature). In one embodiment, T (the number of group N oligonucleotides) may determine the minimal number of temperature cycles used to assemble the oligonucleotides. However, in some embodiments, progressive extension may be achieved without temperature cycling. For example, an enzyme capable promoting rolling circle amplification may be used (e.g., TempliPhi). It should be appreciated that a reaction mixture containing an assembled predetermined nucleic acid fragment also may contain a distribution of shorter extension

products that may result from incomplete extension during one or more of the cycles or may be the result of an P1/NI extension that was initiated after the first cycle.

FIG. 7D illustrates an example of a sequential extension reaction where the 5'-most Pl oligonucleotide is bound to a support and the Group N oligonucleotides are unbound. The reaction steps are similar to those described for FIG. 7C. However, an extended predetermined nucleic acid fragment will be bound to the support via the 5'-most P1 oligonucleotide. Accordingly, the complementary strand (the negative strand) may readily be obtained by denaturing the bound fragment and releasing the negative strand. In some embodiments, the attachment to the support may be labile or readily reversed (e.g., using light, a chemical reagent, a pH change, etc.) and the positive strand also may be released. Accordingly, either the positive strand, the negative strand, or the double-stranded product may be obtained. FIG. 7E illustrates an example of a sequential reaction where P 1 is unbound and the Group N oligonucleotides are bound to a support. The reaction steps are similar to those described for FIG. 7C. However, an extended predetermined nucleic acid fragment will be bound to the support via the 5'-most NT oligonucleotide. Accordingly, the complementary strand (the positive strand) may readily be obtained by denaturing the bound fragment and releasing the positive strand. In some embodiments, the attachment to the support may be labile or readily reversed (e.g., using light, a chemical reagent, a pH change, etc.) and the negative strand also may be released. Accordingly, either the positive strand, the negative strand, or the double stranded product may be obtained.

FIG. 8 shows an embodiment of a plurality of oligonucleotides that may be assembled in a ligase reaction. FIG. 8A illustrates the alignment of the oligonucleotides showing that they do not contain gaps (i.e., no B region as described herein). Accordingly, the oligonucleotides may anneal to form a complex with no nucleotide gaps between the 3' and 5' ends of the annealed oligonucleotides in either Group P or Group N. These oligonucleotides provide a suitable template for assembly using a ligase under appropriate reaction conditions. However, it should be appreciated that these oligonucleotides also may be assembled using a polymerase-based assembly reaction as described herein. FIG. 8B shows two individual ligation reactions. These reactions are illustrated in two steps. However, it should be appreciated that these ligation reactions

may occur simultaneously or sequentially in any order and may occur as such in a reaction maintained under constant reaction conditions (e.g., with no temperature cycling) or in a reaction exposed to several temperature cycles. For example, the reaction illustrated in step 2 may occur before the reaction illustrated in step 1. In each ligation reaction illustrated in FIG. 8B, a Group N oligonucleotide is annealed to two adjacent Group P oligonucleotides (due to the complementary 5' and 3' regions between the P and N oligonucleotides), providing a template for ligation of the adjacent P oligonucleotides. Although not illustrated, ligation of the N group oligonucleotides also may proceed in similar manner to assemble adjacent N oligonucleotides that are annealed to their complementary P oligonucleotide. Assembly of the predetermined nucleic acid fragment may be obtained through ligation of all of the oligonucleotides to generate a double stranded product. However, in some embodiments, a single stranded product of either the positive or negative strand may be obtained. In certain embodiments, a plurality of oligonucleotides may be designed to generate only single stranded reaction products in a ligation reaction. For example, a first group of oligonucleotides (of either Group P or Group N) may be provided to cover the entire sequence on one strand of the predetermined nucleic acid fragment (on either the positive or negative strand). In contrast, a second group of oligonucleotides (from the complementary group to the first group) may be designed to be long enough to anneal to complementary regions in the first group but not long enough to provide adjacent 5' and 3' ends between oligonucleotides in the second group. This provides substrates that are suitable for ligation of oligonucleotides from the first group but not the second group. The result is a single-stranded product having a sequence corresponding to the oligonucleotides in the first group. Again, as with other assembly reactions described herein, a ligase reaction mixture that contains an assembled predetermined nucleic acid fragment also may contain a distribution of smaller fragments resulting from the assembly of a subset of the oligonucleotides.

FIG. 9 shows an embodiment of a ligase-based assembly where one or more of the plurality of oligonucleotides is bound to a support. In FIG. 9A, the 5' most oligonucleotide of the P group oligonucleotides is bound to a support. Ligation of adjacent oligonucleotides in the 5' to 3' direction results in the assembly of a predetermined nucleic acid fragment. FIG. 9A illustrates an example where adjacent

oligonucleotides P2 and P3 are added sequentially. However, the ligation of any two adjacent oligonucleotides from Group P may occur independently and in any order in a ligation reaction mixture. For example, when P1 is ligated to the 5' end of N2, N2 may be in the form of a single oligonucleotide or it already may be ligated to one or more downstream oligonucleotides (N3, N4, etc.). It should be appreciated that for a ligation assembly bound to a support, either the 5'-most (e.g., P1 for Group P, or NT for Group N) or the 3'-most (e.g., PT for Group P, or N1 for Group N) oligonucleotide may be bound to a support since the reaction can proceed in any direction. In some embodiments, a predetermined nucleic acid fragment may be assembled with a central oligonucleotide (i.e., neither the 5'-most or the 3'-most) that is bound to a support provided that the attachment to the support does not interfere with ligation.

FIG. 9B illustrates an example where a plurality of N group oligonucleotides are bound to a support and a predetermined nucleic acid fragment is assembled from P group oligonucleotides that anneal to their complementary support-bound N group oligonucleotides. Again, FIG. 9B illustrates a sequential addition. However, adjacent P group oligonucleotides may be ligated in any order. Also, the bound oligonucleotides may be attached at their 5' end, 3' end, or at any other position provided that the attachment does not interfere with their ability to bind to complementary 5' and 3' regions on the oligonucleotides that are being assembled. This reaction may involve one or more reaction condition changes (e.g., temperature cycles) so that ligated oligonucleotides bound to one immobilized N group oligonucleotide can be dissociated from the support and bind to a different immobilized N group oligonucleotide to provide a substrate for ligation to another P group oligonucleotide.

As with other assembly reactions described herein, support-bound ligase reactions (e.g., those illustrated in FIG. 9B) that generate a full length predetermined nucleic acid fragment also may generate a distribution of smaller fragments resulting from the assembly of subsets of the oligonucleotides. A support used in any of the assembly reactions described herein (e.g., polymerase-based, ligase-based, or other assembly reaction) may include any suitable support medium. A support may be solid, porous, a matrix, a gel, beads, beads in a gel, etc. A support may be of any suitable size.

A solid support may be provided in any suitable configuration or shape (e.g., a chip, a bead, a gel, a microfluidic channel, a planar surface, a spherical shape, a column, etc.).

As illustrated herein, different oligonucleotide assembly reactions may be used to assemble a plurality of overlapping oligonucleotides (with overlaps that are either 5'15', 3'/3', 5'/3', complementary, non-complementary, or a combination thereof). Many of these reactions include at least one pair of oligonucleotides (the pair including one oligonucleotide from a first group or P group of oligonucleotides and one oligonucleotide from a second group or N group of oligonucleotides) have overlapping complementary 3' regions. However, in some embodiments, a predetermined nucleic acid may be assembled from non-overlapping oligonucleotides using blunt-ended ligation reactions. In some embodiments, the order of assembly of the nonoverlapping oligonucleotides may be biased by selective phosphorylation of different 5' ends. In some embodiments, size purification may be used to select for the correct order of assembly. In some embodiments, the correct order of assembly may be promoted by sequentially adding appropriate oligonucleotide substrates into the reaction (e.g., the ligation reaction).

In order to obtain a full-length nucleic acid fragment from a multiplex oligonucleotide assembly reaction, a purification step may be used to remove starting oligonucleotides and/or incompletely assembled fragments. In some embodiments, a purification step may involve chromatography, electrophoresis, or other physical size separation technique. In certain embodiments, a purification step may involve amplifying the full length product. For example, a pair of amplification primers (e.g., PCR primers) that correspond to the predetermined 5' and 3' ends of the nucleic acid fragment being assembled will preferentially amplify full length product in an exponential fashion. It should be appreciated that smaller assembled products may be amplified if they contain the predetermined 5' and 3' ends. However, such smaller-than-expected products containing the predetermined 5' and 3' ends should only be generated if an error occurred during assembly (e.g., resulting in the deletion or omission of one or more regions of the target nucleic acid) and may be removed by size fractionation of the amplified product. Accordingly, a preparation containing a relatively high amount of full length product may be obtained directly by amplifying the product of an assembly

reaction using primers that correspond to the predetermined 5' and 3' ends. In some embodiments, additional purification (e.g., size selection) techniques may be used to obtain a more purified preparation of amplified full-length nucleic acid fragment.

When designing a plurality of oligonucleotides to assemble a predetermined nucleic acid fragment, the sequence of the predetermined fragment will be provided by the oligonucleotides as described herein. However, the oligonucleotides may contain additional sequence information that may be removed during assembly or may be provided to assist in subsequent manipulations of the assembled nucleic acid fragment. Examples of additional sequences include, but are not limited to, primer recognition sequences for amplification (e.g., PCR primer recognition sequences), restriction enzyme recognition sequences, recombination sequences, other binding or recognition sequences, labeled sequences, etc. In some embodiments, one or more of the 5'-most oligonucleotides, one or more of the 3'-most oligonucleotides, or any combination thereof, may contain one or more additional sequences. In some embodiments, the additional sequence information may be contained in two or more adjacent oligonucleotides on either strand of the predetermined nucleic acid sequence. Accordingly, an assembled nucleic acid fragment may contain additional sequences that may be used to connect the assembled fragment to one or more additional nucleic acid fragments (e.g., one or more other assembled fragments, fragments obtained from other sources, vectors, etc.) via ligation, recombination, polymerase-mediated assembly, etc. In some embodiments, purification may involve cloning one or more assembled nucleic acid fragments. The cloned product may be screened (e.g., sequenced, analyzed for an insert of the expected size, etc.).

In some embodiments, a nucleic acid fragment assembled from a plurality of oligonucleotides may be combined with one or more additional nucleic acid fragments using a polymerase-based and/or a ligase-based extension reaction similar to those described herein for oligonucleotide assembly. Accordingly, one or more overlapping nucleic acid fragments may be combined and assembled to produce a larger nucleic acid fragment as described herein. In certain embodiments, double-stranded overlapping oligonucleotide fragments may be combined. However, single-stranded fragments, or combinations of single-stranded and double-stranded fragments may be combined as

described herein. A nucleic acid fragment assembled from a plurality of oligonucleotides may be of any length depending on the number and length of the oligonucleotides used in the assembly reaction. For example, a nucleic acid fragment (either single-stranded or double-stranded) assembled from a plurality of
5 oligonucleotides may be between 50 and 1,000 nucleotides long (for example, about 70 nucleotides long, between 100 and 500 nucleotides long, between 200 and 400 nucleotides long, about 200 nucleotides long, about 300 nucleotides long, about 400 nucleotides long, etc.). One or more such nucleic acid fragments (e.g., with overlapping 3' and/or 5' ends) may be assembled to form a larger nucleic acid fragment (single -
10 stranded or double-stranded) as described herein.

A full length product assembled from smaller nucleic acid fragments also may be isolated or purified as described herein (e.g., using a size selection, cloning, selective binding or other suitable purification procedure). In addition, any assembled nucleic acid fragment (e.g., full-length nucleic acid fragment) described herein may be amplified
15 (prior to, as part of, or after, a purification procedure) using appropriate 5' and 3' amplification primers.

It should be appreciated that the terms P Group and N Group oligonucleotides are used herein for clarity purposes only, and to illustrate several embodiments of multiplex oligonucleotide assembly. The Group P and Group N oligonucleotides described herein
20 are interchangeable, and may be referred to as first and second groups of oligonucleotides corresponding to sequences on complementary strands of a target nucleic acid fragment.

Applications:

Aspects of the invention may be useful for a range of applications involving the
25 production and/or use of synthetic nucleic acids. As described herein, the invention provides methods for assembling synthetic nucleic acids with increased efficiency. The resulting assembled nucleic acids may be amplified in vitro (e.g., using PCR, LCR, or any suitable amplification technique), amplified in vivo (e.g., via cloning into a suitable vector), isolated and/or purified. An assembled nucleic acid (alone or cloned into a
30 vector) may be transformed into a host cell (e.g., a prokaryotic, eukaryotic, insect,

mammalian, or other host cell). In some embodiments, the host cell may be used to propagate the nucleic acid. In certain embodiments, the nucleic acid may be integrated into the genome of the host cell. In some embodiments, the nucleic acid may replace a corresponding nucleic acid region on the genome of the cell (e.g., via homologous
5 recombination). Accordingly, nucleic acids may be used to produce recombinant organisms. In some embodiments, a target nucleic acid may be an entire genome or large fragments of a genome that are used to replace all or part of the genome of a host organism. Recombinant organisms also may be used for a variety of research, industrial, agricultural, and/or medical applications.

10 Many of the techniques described herein can be used together, applying enrichment steps at one or more points to produce long nucleic acid molecules. Correct sequence enrichment techniques of the invention can be applied to double-stranded nucleic acids of any size. For example, sorting techniques may be used with oligonucleotide duplexes and nucleic acid fragments of less than 100 to more than
15 10,000 base pairs in length (e.g., 100 mers to 500 mers, 500 mers to 1,000 mers, 1,000 mers to 5,000 mers, 5, 000 mers to 10,000 mers, 25,000 mess, 50,000 mers, 75,000 mers, 100,000 mers, etc.). In an exemplary embodiment, methods described herein may be used to during the assembly of an entire genome (or a large fragment thereof, e.g., about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or more) of an organism (e.g., of a
20 viral, bacterial, yeast, or other prokaryotic or eukaryotic organism) optionally incorporating specific modifications into the sequence at one or more desired locations.

Any of the nucleic acid products (e.g., including nucleic acids that are amplified, cloned, purified, isolated, etc.) may be packaged in any suitable format (e.g., in a stable buffer, lyophilized, etc.) for storage and/or shipping (e.g., for shipping to a distribution
25 center or to a customer). Similarly, any of the host cells (e.g., cells transformed with a vector or having a modified genome) may be prepared in a suitable buffer for storage and or transport (e.g., for distribution to a customer). In some embodiments, cells may be frozen. However, other stable cell preparations also may be used.

Host cells may be grown and expanded in culture. Host cells may be used for
30 expressing one or more RNAs or polypeptides of interest (e.g., therapeutic, industrial, agricultural, and/or medical proteins). The expressed polypeptides may be natural

polypeptides or non-natural polypeptides. The polypeptides may be isolated or purified for subsequent use.

Accordingly, nucleic acid molecules generated using methods of the invention can be incorporated into a vector. The vector may be a cloning vector or an expression vector. In some embodiments, the vector may be a viral vector. A viral vector may comprise nucleic acid sequences capable of infecting target cells. Similarly, in some embodiments, a prokaryotic expression vector operably linked to an appropriate promoter system can be used to transform target cells. In other embodiments, a eukaryotic vector operably linked to an appropriate promoter system can be used to transfect target cells or tissues.

Transcription and/or translation of the constructs described herein may be carried out in vitro (i.e., using cell-free systems) or in vivo (i.e., expressed in cells). In some embodiments, cell lysates may be prepared. In certain embodiments, expressed RNAs or polypeptides may be isolated or purified. Nucleic acids of the invention also may be used to add detection and/or purification tags to expressed polypeptides or fragments thereof. Examples of polypeptide-based fusion/tag include, but are not limited to, hexa-histidine (His6) Myc and HA, and other polypeptides with utility, such as GFP, GST, MBP, chitin and the like. In some embodiments, polypeptides may comprise one or more unnatural amino acid residue(s).

In some embodiments, antibodies can be made against polypeptides or fragment(s) thereof encoded by one or more synthetic nucleic acids.

In certain embodiments, synthetic nucleic acids may be provided as libraries for screening in research and development (e.g., to identify potential therapeutic proteins or peptides, to identify potential protein targets for drug development, etc.)

In some embodiments, a synthetic nucleic acid may be used as a therapeutic (e.g., for gene therapy, or for gene regulation). For example, a synthetic nucleic acid may be administered to a patient in an amount sufficient to express a therapeutic amount of a protein. In other embodiments, a synthetic nucleic acid may be administered to a patient in an amount sufficient to regulate (e.g., down-regulate) the expression of a gene.

It should be appreciated that different acts or embodiments described herein may be performed independently and may be performed at different locations in the United States or outside the United States. For example, each of the acts of receiving an order for a target nucleic acid, analyzing a target nucleic acid sequence, designing one or more starting nucleic acids (e.g., oligonucleotides), synthesizing starting nucleic acid(s), purifying starting nucleic acid(s), assembling starting nucleic acid(s), isolating assembled nucleic acid(s), confirming the sequence of assembled nucleic acid(s), manipulating assembled nucleic acid(s) (e.g., amplifying, cloning, inserting into a host genome, etc.), and any other acts or any parts of these acts may be performed independently either at one location or at different sites within the United States or outside the United States. In some embodiments, an assembly procedure may involve a combination of acts that are performed at one site (in the United States or outside the United States) and acts that are performed at one or more remote sites (within the United States or outside the United States).

Automated applications:

Aspects of the invention may include automating one or more acts described herein. For example, a sequence analysis may be automated in order to generate a synthesis strategy automatically. The synthesis strategy may include i) the design of the starting nucleic acids that are to be assembled into the target nucleic acid, ii) the choice of the assembly technique(s) to be used, iii) the number of rounds of assembly and error screening or sequencing steps to include, and/or decisions relating to subsequent processing of an assembled target nucleic acid. Similarly, one or more steps of an assembly reaction may be automated using one or more automated sample handling devices (e.g., one or more automated liquid or fluid handling devices). For example, the synthesis and optional selection of starting nucleic acids (e.g., oligonucleotides) may be automated using an nucleic acid synthesizer and automated procedures. Automated devices and procedures may be used to mix reaction reagents, including one or more of the following: starting nucleic acids, buffers, enzymes (e.g., one or more ligases and/or polymerases), nucleotides, nucleic acid binding proteins or recombinases, salts, and any other suitable agents such as stabilizing agents. Automated devices and procedures also may be used to control the reaction conditions. For example, an automated thermal

cycler may be used to control reaction temperatures and any temperature cycles that may be used. In some embodiments, a thermal cycler may be automated to provide one or more reaction temperatures or temperature cycles suitable for incubating nucleic acid fragments prior to transformation. Similarly, subsequent purification and analysis of
5 assembled nucleic acid products may be automated. For example, fidelity optimization steps (e.g., a MutS error screening procedure) may be automated using appropriate sample processing devices and associated protocols. Sequencing also may be automated using a sequencing device and automated sequencing protocols. Additional steps (e.g., amplification, cloning, etc.) also may be automated using one or more appropriate
10 devices and related protocols. It should be appreciated that one or more of the device or device components described herein may be combined in a system (e.g., a robotic system) or in a micro-environment (e.g., a micro-fluidic reaction chamber). Assembly reaction mixtures (e.g., liquid reaction samples) may be transferred from one component of the system to another using automated devices and procedures (e.g., robotic
15 manipulation and/or transfer of samples and/or sample containers, including automated pipetting devices, micro-systems, etc.). The system and any components thereof may be controlled by a control system.

Accordingly, acts of the invention may be automated using, for example, a computer system (e.g., a computer controlled system). A computer system on which
20 aspects of the invention can be implemented may include a computer for any type of processing (e.g., sequence analysis and/or automated device control as described herein). However, it should be appreciated that certain processing steps may be provided by one or more of the automated devices that are part of the assembly system. In some embodiments, a computer system may include two or more computers. For example,
25 one computer may be coupled, via a network, to a second computer. One computer may perform sequence analysis. The second computer may control one or more of the automated synthesis and assembly devices in the system. In other aspects, additional computers may be included in the network to control one or more of the analysis or processing acts. Each computer may include a memory and processor. The computers
30 can take any form, as the aspects of the present invention are not limited to being implemented on any particular computer platform. Similarly, the network can take any form, including a private network or a public network (e.g., the Internet). Display

devices can be associated with one or more of the devices and computers. Alternatively, or in addition, a display device may be located at a remote site and connected for displaying the output of an analysis in accordance with the invention. Connections between the different components of the system may be via wire, optical fiber, wireless transmission, satellite transmission, any other suitable transmission, or any combination of two or more of the above.

In accordance with one embodiment of the present invention for use on a computer system it is contemplated that sequence information (e.g., a target sequence, a processed analysis of the target sequence, etc.) can be obtained and then sent over a private or public network, such as the Internet, to a remote location to be processed by computer to produce any of the various types of outputs discussed herein (e.g., in connection with oligonucleotide design). However, it should be appreciated that the aspects of the present invention described herein are not limited in that respect, and that numerous other configurations are possible. For example, all of the analysis and processing described herein can alternatively be implemented on a computer that is attached locally to a device, an assembly system, or one or more components of an assembly system. As a further alternative, as opposed to transmitting sequence information (e.g., a target sequence, a processed analysis of the target sequence, etc.) over a communication medium (e.g., the network), the information can be loaded onto a computer readable medium that can then be physically transported to another computer for processing in the manners described herein. In another embodiment, a combination of two or more transmission/delivery techniques may be used. It also should be appreciated that computer implementable programs for performing a sequence analysis or controlling one or more of the devices, systems, or system components described herein also may be transmitted via a network or loaded onto a computer readable medium as described herein. Accordingly, aspects of the invention may involve performing one or more steps within the United States and additional steps outside the United States. In some embodiments, sequence information (e.g., a customer order) may be received at one location (e.g., in one country) and sent to a remote location for processing (e.g., in the same country or in a different country (e.g., for sequence analysis to determine a synthesis strategy and/or design oligonucleotides). In certain embodiments, a portion of the sequence analysis may be performed at one site (e.g., in

one country) and another portion at another site (e.g., in the same country or in another country). In some embodiments, different steps in the sequence analysis may be performed at multiple sites (e.g., all in one country or in several different countries). The results of a sequence analysis then may be sent to a further site for synthesis. However, in some embodiments, different synthesis and quality control steps may be performed at more than one site (e.g., within one county or in two or more countries). An assembled nucleic acid then may be shipped to a further site (e.g., either to a central shipping center or directly to a client).

Each of the different aspects, embodiments, or acts of the present invention described herein can be independently automated and implemented in any of numerous ways. For example, each aspect, embodiment, or act can be independently implemented using hardware, software or a combination thereof. When implemented in software, the software code can be executed on any suitable processor or collection of processors, whether provided in a single computer or distributed among multiple computers. It should be appreciated that any component or collection of components that perform the functions described above can be generically considered as one or more controllers that control the above-discussed functions. The one or more controllers can be implemented in numerous ways, such as with dedicated hardware, or with general purpose hardware (e.g., one or more processors) that is programmed using microcode or software to perform the functions recited above.

In this respect, it should be appreciated that one implementation of the embodiments of the present invention comprises at least one computer-readable medium (e.g., a computer memory, a floppy disk, a compact disk, a tape, etc.) encoded with a computer program (i.e., a plurality of instructions), which, when executed on a processor, performs one or more of the above-discussed functions of the present invention. The computer-readable medium can be transportable such that the program stored thereon can be loaded onto any computer system resource to implement one or more functions of the present invention discussed herein. In addition, it should be appreciated that the reference to a computer program which, when executed, performs the above-discussed functions, is not limited to an application program running on a host computer. Rather, the term computer program is used herein in a generic sense to

reference any type of computer code (e.g., software or microcode) that can be employed to program a processor to implement the above-discussed aspects of the present invention.

It should be appreciated that in accordance with several embodiments of the present invention wherein processes are stored in a computer-readable medium, the computer implemented processes may, during the course of their execution, receive input manually (e.g., from a user).

Accordingly, overall system-level control of the assembly devices or components described herein may be performed by a system controller which may provide control signals to the associated nucleic acid synthesizers, liquid handling devices, thermal cyclers, sequencing devices, associated robotic components, as well as other suitable systems for performing the desired input/output or other control functions. Thus, the system controller along with any device controllers together form a controller that controls the operation of a nucleic acid assembly system. The controller may include a general purpose data processing system, which can be a general purpose computer, or network of general purpose computers, and other associated devices, including communications devices, modems, and/or other circuitry or components necessary to perform the desired input/output or other functions. The controller can also be implemented, at least in part, as a single special purpose integrated circuit (e.g., ASIC) or an array of ASICs, each having a main or central processor section for overall, system-level control, and separate sections dedicated to performing various different specific computations, functions and other processes under the control of the central processor section. The controller can also be implemented using a plurality of separate dedicated programmable integrated or other electronic circuits or devices, e.g., hard wired electronic or logic circuits such as discrete element circuits or programmable logic devices. The controller can also include any other components or devices, such as user input/output devices (monitors, displays, printers, a keyboard, a user pointing device, touch screen, or other user interface, etc.), data storage devices, drive motors, linkages, valve controllers, robotic devices, vacuum and other pumps, pressure sensors, detectors, power supplies, pulse sources, , communication devices or other electronic circuitry or components, and so on. The controller also may control operation of other portions of a

system, such as automated client order processing, quality control, packaging, shipping, billing, etc., to perform other suitable functions known in the art but not described in detail herein.

Business applications:

5 Aspects of the invention may be useful to streamline nucleic acid assembly reactions. Accordingly, aspects of the invention relate to marketing methods, compositions, kits, devices, and systems for increasing nucleic acid assembly throughput involving correct sequence enrichment using sorting techniques described herein.

Aspects of the invention may be useful for reducing the time and/or cost of
10 production, commercialization, and/or development of synthetic nucleic acids, and/or related compositions. Accordingly, aspects of the invention relate to business methods that involve collaboratively (e.g., with a partner) or independently marketing one or more methods, kits, compositions, devices, or systems for analyzing and/or assembling synthetic nucleic acids as described herein. For example, certain embodiments of the
15 invention may involve marketing a procedure and/or associated devices or systems involving correct sequence enrichment using sliding clamp techniques described herein.. In some embodiments, synthetic nucleic acids, libraries of synthetic nucleic acids, host cells containing synthetic nucleic acids, expressed polypeptides or proteins, etc., also may be marketed.

20 Marketing may involve providing information and/or samples relating to methods, kits, compositions, devices, and/or systems described herein. Potential customers or partners may be, for example, companies in the pharmaceutical, biotechnology and agricultural industries, as well as academic centers and government research organizations or institutes. Business applications also may involve generating
25 revenue through sales and/or licenses of methods, kits, compositions, devices, and/or systems of the invention.

It should be understood that the above-described embodiments and the following example(s) are given by way of illustration, not limitation. Various changes and modifications within the scope of the present invention will become apparent to those
30 skilled in the art from the present description.

EXAMPLES

Example 1. Synthesis, sorting and assembly

An exemplary proposed oligonucleotide synthesis and sorting followed by assembly into a longer nucleic acid is described below and depicted schematically in FIG. 10.

To synthesize a DNA sequence approximately ten thousand base pairs in length, 400 oligonucleotides are designed. Half of the oligonucleotides each include a sequence of approximately 50 bases corresponding to a portion of one strand of the desired DNA sequence, such that in combination the oligonucleotides contain the sequence of one entire strand. The other half of the oligonucleotides each include a sequence of approximately 50 bases corresponding to a portion of the other strand and collectively contain the entire sequence of the other strand. The sequences of approximately 50 bases will be referred to hereafter as "target sequences." Each target sequence is flanked by restriction enzyme sites and by PCR primer binding sites 10-20 bases in length.

All of the 400 oligonucleotides are synthesized in parallel on a glass slide as described by Richmond et al. Each oligonucleotide is synthesized in its own 13 μm^2 region of the surface of the slide, twenty-five of which are depicted at 600 in FIG. 10. The slide is covered with the base-labile linker with photocleavable protecting group depicted in FIG. 1. Specifically, 7 mg of the linker are mixed with 2 mg P-dimethylaminopyride (DMAP), 6 mg 2-(1Hbenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and 10 µl of dry di-isopropyethylamine in 300 µl of dry acetonitrile in a sealed vial. 50 µl of the reaction solution are transferred immediately to a dry, amine-derivatized glass slide under a coverslip for 20 minutes in the dark. The slide is then washed extensively with acetonitrile; dried with a nitrogen stream; and stored in a dessicator.

At the time of synthesis step 500, the glass slide is mounted in a flow-cell reaction chamber connected to a DNA synthesizer. All oligonucleotides are synthesized in parallel using photolithographic chemical synthesis as described in Richmond et al., using a micromirror array projecting a 365 nm virtual mask on the slide to direct the photodeprotection and growth of the oligonucleotides on the chip. At the end of

synthesis process 500, synthesis of each of the 400 different oligonucleotide sequences is complete. As each of the 400 different oligonucleotide sequences is synthesized on a region of the slide with many copies of the base-labile linker, many individual oligonucleotide molecules are synthesized for each of the 400 different sequences, each
5 individual oligonucleotide molecule being attached to a separate linker molecule in the same region of the slide. Alternatively, an acid-labile, photolabile, uracil-containing, or other cleavable linker can be substituted. At 610 of FIG. 10, for simplicity of illustration, only several oligonucleotide molecules are shown attached to each of the 25 representative regions of the slide.

10 The 400 oligonucleotides are eluted from the slide (depicted at 510 in FIG. 10) in a single pool 620 by treatment of the slide with ammonium hydroxide for one hour to cleave the base-labile linker. The eluate is dried in a speed vacuum centrifuge at a force of 240 g and the precipitate is resuspended in 5 μl of sterile water.

The oligonucleotides are loaded on beads and subjected to emulsion PCR (520)
15 essentially as described in Shendure et al., except that the biotinylated primers used to immobilize the PCR products on beads incorporates the photo-labile linker of FIG. 3. 100 μl of 1um Dynal MyOne paramagnetic streptavidin beads are mixed with 100 μl of wash buffer (1.0 M NaCl, 5 mM Tris-HCI, 0.5 mM EDTA, pH 7.5). The liquid is removed and the beads are washed twice again and finally resuspended in 198 μl of wash
20 buffer. 2 μl of 1mM of a dual-biotinylated PCR forward primer complementary to one of the two PCR primer binding sites on the 400 oligonucleotides is added to the beads for twenty minutes with occasional agitation. All liquid is removed and the beads are washed twice in 200 μl of wash buffer and once in 200 μl of TE and the beads are resuspended in 200 μl of TE. 60 μl of the beads are combined with the 5μl of the 400
25 resuspended oligonucleotides in a 960 μl solution including lx MgC12-PCR buffer (Invitrogen), 18.8 mM MgC12 (Invitrogen), 3.5 mM of each dNTP (Invitrogen), 25 μM reverse primer, 50 nM unmodified forward primer, and 270U Platinum Taq (Invitrogen). 12 aliquots of 75 μl of this aqueous solution are taken and each is added to a 2 ml Corning cryogenic vial containing 400 μl of an oil phase (54.5% light mineral oil
30 (Sigma), 45% light mineral oil containing 10% Span 80 (Sigma), 4% Tween 80 (Sigma), 0.5% Triton X-100 (Sigma)). The oil and water phases are mixed by stirring for 30

minutes to prepare the emulsion. 50 µl aliquots are taken from the cryogenic tubes and placed in 200 µl PCR tubes.

PCR proceeds according to the following program: step 1, 94C for 2 minutes; step 2, 94C for 15 seconds; step 3, 57C for 30 seconds, step 4, 70C for 75 seconds; step 5, return to step 2 for an additional 119 iterations; step 6, 72C for 2 minutes; hold at 4C. The contents of the tubes are pooled in groups of eight in 1.5 ml tubes and the emulsion is broken by adding 800 µl of breaking buffer (100 mM NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1% Triton X-100), vortexing 20 seconds, spinning for 1.5 minutes at 13.2K rpm, removing most liquid, adding 800 µl of breaking buffer, vortexing 20 seconds, spinning for 1.5 minutes at 9K rpm, removing most liquid, adding 700 µl of breaking buffer, vortexing 20 seconds, spinning for 1.5 minutes at 9K rpm, removing most liquid, adding 600 µl of breaking buffer, vortexing 20 seconds, spinning for 1.5 minutes at 9K rpm, removing most liquid, placing the tube against a magnet and removing all liquid. The beads are resuspended in 20 µl of TE and transferred to a new tube. All resuspended beads from all tubes are now pooled and washed twice in 250 µl TE with removal of all liquid with a magnet. Free reverse strands are denatured by adding 250 µl of 0.1M NaOH to the tube and incubating at room temperature for ten minutes; washing once in 0.1M NaOH and three times in TE. The beads are then resuspended in 6.5 µl TE. The resulting polony beads are depicted at 630 in FIG. 10.

The polonies are sequenced (530) essentially as described in Shendure et al. The polony beads are first immobilized in a 5% polyacrylamide gel. 40 mm round #1.5 glass coverslips are washed for twenty minutes in 1% Triton X-100; rinsed in water; incubated for one hour in a solution of 350 mL water, 1.3 mL Bind Silane (Amersham) and 73 µl acetic acid; washed three times in water and once in 95% ethanol; and stored in a dessicator. The polony beads are mixed with 1.25 µl of 40% acrylamide/bis (19/1, Roche), 0.5 µl Rinohide gel strengthener (Molecular Probes), 0.5 µl 5% TEMED, and 0.75 µl 0.5% ammonium persulfate. The mixture is pipetted on a Teflon-coated slide and the Bind Silane-treated coverslip is dropped on top and the slide and coverslip are inverted. Polymerization proceeds for one hour at 25C. The coverslip with attached gel

is removed, immersed in water, and immediately assembled into a flow cell for sequencing.

The polonies are sequenced by degenerate ligation. Each base is sequentially interrogated by 1) hybridizing an anchor primer complementary to a common library
5 sequence; 2) ligating a pool of labeled query primers specific to one tag-position; imaging to determine which primer pool ligated to each bead; 3) stripping the anchor/query primer complex; and 4) repeating. The query primers are degenerate at all positions except the query position, at which only one base was present for a given fluorophore. Hybridization of anchor primers at 100 Um primer in 6x SSPE proceeded
10 for 5 minutes at 56C; the flowcell was then cooled to 42C and held at that temperature for 2 minutes. Excess primer was washed out at room temperature for two minutes with wash lE solution (10 mM Tris-HCl pH 7.5, 2 mM EDTA pH 8.0, 0.01% Triton X-100). Query primers were provided as an 8 μM query primer mix (with 2 μM of each query primer) with 6000U T4 DNA ligase and lx T4 DNA ligase buffer (NEB) at 35C for 30
15 minutes, after which excess query primer was removed at room temperature using wash lE solution for five minutes. Four-color imaging is as described in Shendure et al. The anchor/query primer complex is then stripped with USER (NEB) as described in Shendure et al., and the process is repeated for the next set of primers. By varying the fixed position in the query primer nonamer, nine positions can be queried. By using
20 query primers on both sides of an anchor primer (one side at a time), eighteen positions can be queried. By using anchor primers at several different positions, the entire target sequence of each oligonucleotide can be determined. Importantly, by designing the anchor primers as a pool having members complementary to each of the 400 target sequences, sequence determinations can proceed in a massively parallel manner.

25 Thus, the sequence determinations lead to the identification of nucleic acids determined to have sequences of interest for each of the 400 oligonucleotides. In FIG. 10, representative nucleic acids determined to have sequences of interest are depicted schematically as polonies A, B, C, and D at 640.

Sequences of interest are then selectively isolated. The polony beads in the 5%
30 polyacrylamide gel are treated with trifluoroacetic acid, converting the dimethoxy ketal linkage of the linker to a ketone linkage, such that the NVOC group is rendered

photocleavable. Selected oligonucleotides determined to have sequences of interest (represented by A and B at 640) are selectively irradiated (540) with UV radiation using the micromirror array of Richmond et al., to cleave the linkers, releasing the selected oligonucleotides from the beads, which are then selectively eluted from the polyacrylamide gel in lx TE, dried, and resuspended in lx TE to form a first pool (650) of eluted nucleic acids. Subsequently, other oligonucleotides determined to have sequences of interest (represented by C and D at 640 are selectively irradiated (540') and eluted to form a separate pool (650') of oligonucleotides. The process is continued until oligonucleotides corresponding to each of the 400 of the target sequences have been isolated in one or more pools.

The pools are optionally subjected to error filtration (550 and 550') to preferentially remove remaining incorrect oligonucleotides (depicted in pool 650 as A' and in pool 650' as D'). Error filtration is performed, for example, using the mutS system of Carr et al., (2004) Nucl. Acids Res. 32(20): e162 doi: 10. 1 093/nar/gnhl 60, leaving purified pools, two of which are represented in FIG. 10 as 660 and 6601. The PCR primer binding sites are cleaved from the target sequences by restriction enzymes and are purified from the target sequences by polyacrylamide gel electrophoresis. The final, purified pools of target sequences (representative pools being shown in FIG. 10 at 660 and 660') are subjected to hierarchical assembly (560, 560') to form subassemblies (such as representative subassemblies 670 and 670' of FIG. 10) that are themselves assembled (560") to form the final target sequence approximately ten thousand base pairs in length (represented in FIG. 10 as molecule AC/BD at 680.

Example 2. Optical systems

The sequences of individual oligonucleotides can be determined by sequencing methods that involve the emission of electromagnetic radiation following the addition of each nucleotide to a growing strand. This electromagnetic radiation can be detected with an appropriate optical system. An optical system can also be used to deliver electromagnetic radiation to individual oligonucleotides to cleave photolabile groups and selectively remove individual oligonucleotides from a solid support, or to selectively damage individual oligonucleotides. Thus, two or more optical systems can be used while sequencing and sorting oligonucleotides. Generally, the use of reference marks or

other means of calibration is recommended where different optical systems are used for sequence detection and subsequent delivery of targeted electromagnetic radiation. Reference marks can permit a more accurate targeting of oligonucleotides using a second optical system to the oligonucleotide locations mapped using a first optical system.

5 The calibration of separate optical systems for sequence detection and for radiation delivery can be avoided by integrating both functions into a single optical system such as the exemplary optical system shown in FIG. 11. Referring to FIG. 11, optical system 700 includes microscope objective 710 positioned near stage 720 bearing reaction chamber 730 for housing oligonucleotides, immobilized with a photolabile 10 linker, and reagents for sequence determinations. Microscope objective 710 can be from an inverted microscope such as a Nikon TE-2000. During sequencing, light from reaction chamber 730 passes through microscope objective 710 to camera 740, which can be a CMOS camera or, more preferably, a CCD camera, which converts the detected light to an electrical signal for communication to computer 750. Using the sequencing 15 data, computer 750 determines which nucleotides in reaction chamber 730 appear to contain a correct or desirable sequence.

Targeted electromagnetic radiation, such as ultraviolet light, is delivered to reaction chamber 730 via the same microscope objective 710 used during sequence determination. The radiation is targeted using a suitable optical addressing system 760, 20 such as a micro-mirror array (e.g., a Digital Light Processing array, Texas Instruments). Thus, for each location identified for targeting by computer 750 , computer 750, directly or via one or more intervening computers 755, controls optical addressing system 760 to direct electromagnetic radiation through microscope objective 710 into reaction chamber 730, permitting the selective isolation of desired oligonucleotides.

25 In some embodiments, a system for targeted electromagnetic radiation, may incorporate micromirrors and/or an image locking subsystem, for example, as described in U.S. Patent Application Publication No. 2005-0249396 (Now U.S. Patent No. 7,072,500), the entire disclosure of which is incorporated by reference herein.

The foregoing description of the present invention provides illustration and 30 description, but is not intended to be exhaustive or to limit the invention to the precise

embodiments disclosed. Modifications and variations are possible consistent with the above teachings or may be acquired from practice of the invention. Thus, it is noted that the scope of the invention is defined by the claims and their equivalents. The entire contents of the scientific publications and patent documents cited above are incorporated
5 herein by reference. In the event of conflict, the present specification controls.

What is claimed is:

CLAIMS

1. A method of sorting oligonucleotides, the method comprising the steps of:

a) for each of a plurality of oligonucleotides, consecutively interrogating the identity of each of a plurality of contiguous nucleobases of the oligonucleotide, thereby to determine a sequence of each of a plurality of the oligonucleotides; and

b) selectively isolating one or more oligonucleotides determined to have the desired sequence from one or more oligonucleotides determined not to have the desired sequence.

2. The method of claim 1, wherein the oligonucleotides comprise hybridization sites flanking the sequence determined in step a), and step b) comprises selectively amplifying an oligonucleotide determined to have the desired sequence.

3. The method of claim 1, wherein step b) comprises selectively inactivating one or more oligonucleotides determined not to have the desired sequence.

4. A method of sorting oligonucleotides, the method comprising the steps of:

a) providing one or more solid supports attached to a plurality of oligonucleotides;

b) determining a sequence of each of a plurality of the oligonucleotides; and

c) selectively releasing or selectively copying one or more oligonucleotides determined to have the desired sequence.

5. The method of claim 4, wherein step a) comprises amplifying each of a plurality of oligonucleotides by polymerase colony amplification and attaching individual amplicons to individual solid supports.

6. The method of claim 4, wherein step b) comprises, for each of a plurality of oligonucleotides, consecutively interrogating the identity of each of a plurality of contiguous nucleobases.

7. The method of claim 4, wherein the oligonucleotides are attached to the one or more solid supports by a linker whose susceptibility to cleavage is increased by exposure to radiation.

8. The method of claim 7, wherein step c) comprises selectively irradiating
5 the linker of an oligonucleotide determined to have the desired sequence.

9. The method of claim 8, wherein an optical system is used to specifically target radiation to individual oligonucleotides to selectively release the oligonucleotides.

10. The method of claim 9, wherein micromirrors are used to target the radiation.

10 11. The method of claim 9, wherein targeting comprises the use of an image locking subsystem.

12. The method of claim 4, further comprising the step, prior to step c), of selectively releasing or selectively inactivating one or more oligonucleotides determined not to have the desired sequence.

15 13. The method of claim 12, wherein the step of selectively releasing or selectively inactivating one or more oligonucleotides determined not to have the desired sequence comprises selectively irradiating one or more oligonucleotides determined not to have the desired sequence.

14. A method of high fidelity nucleic acid preparation, the method comprising
20 the steps of:

a) synthesizing a multiplicity of individual oligonucleotide molecules, each with the same intended nucleotide sequence;

b) for each of a plurality of the individual oligonucleotide molecules, consecutively interrogating the identity of each of a plurality of contiguous nucleobases
25 of (i) the individual oligonucleotide molecule, (ii) its complement, or (iii) a copy thereof, thereby to identify one or more desirable nucleic acids determined to have the intended nucleotide sequence or the complement of the intended nucleotide sequence; and

c) selectively isolating one or more desirable nucleic acids from other nucleic acids.

15. The method of claim 14, wherein, in step b), the individual oligonucleotide molecules are attached to one or more solid supports.

5 16. The method of claim 15, further comprising, after step a), the step of amplifying each of a plurality of the individual oligonucleotide molecules by polymerase colony amplification and attaching individual amplicons to individual solid supports.

17. The method of claim 16, wherein step c) comprises manually separating an individual solid support attached to desirable nucleic acids from at least one
10 individual solid support attached to other nucleic acids.

18. The method of claim 15, wherein step c) comprises selectively releasing desirable nucleic acids from the one or more solid supports.

19. A high-throughput method of nucleic acid preparation, the method comprising the steps of:

15 a) providing one or more solid supports attached to a multiplicity of populations of oligonucleotides, each member of a population having the same intended sequence and each population having a different intended sequence;

b) determining sequences the oligonucleotides in parallel, thereby to identify at least one member of each population determined to have its intended
20 sequence; and

c) selectively releasing or selectively copying said at least one member of each population.

20. The method of claim 19, further comprising the step of assembling a double-stranded nucleic acid comprising members of a plurality of the populations, the
25 assembling step comprising exposing the selectively released or selectively copied members to conditions promoting hybridization of a member of one population to a member of another population.

21. The method of claim 19, wherein step a) comprises synthesizing the oligonucleotides on a surface.

22. The method of claim 21, wherein step a) further comprises releasing the oligonucleotides from the surface, amplifying them by polymerase colony amplification, 5 and attaching each amplicon to a solid support.

23. The method of claim 19, wherein step b) comprises sequencing by synthesis.

24. The method of claim 19, wherein step c) comprises (i) selectively releasing or selectively copying at least one member of one population determined to 10 have its intended sequence and, thereafter, (ii) selectively releasing or selectively copying at least one member of a second, different population determined to have its intended sequence, thereby to collect at least two distinct pools of oligonucleotides determined to have their respective intended sequences.

25. The method of claim 24, further comprising the step of:

15 d) assembling a nucleic acid incorporating oligonucleotides from the first pool and, thereafter,

e) assembling a longer nucleic acid incorporating oligonucleotides from the first and second pools.

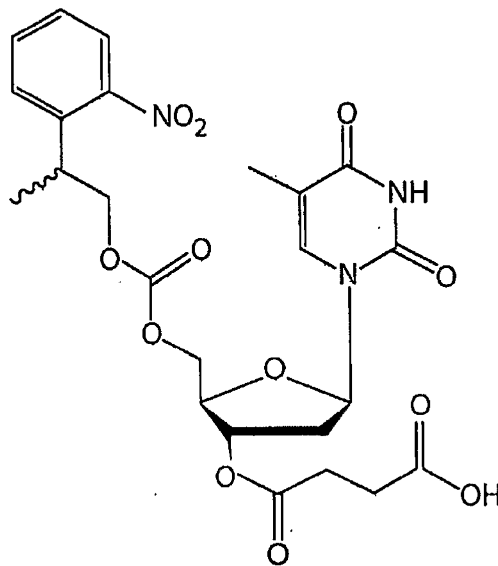


Fig. 1

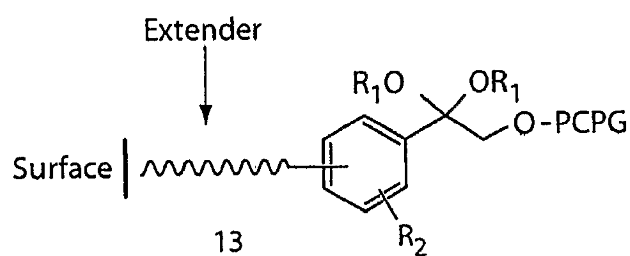


Fig. 2

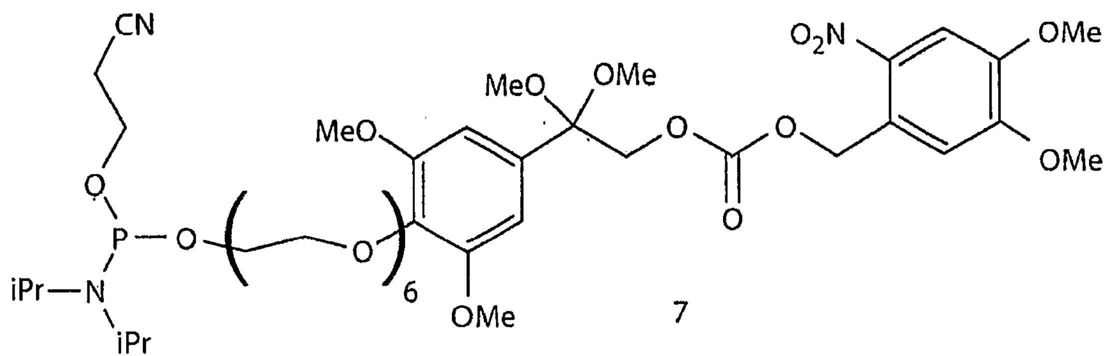


Fig. 3

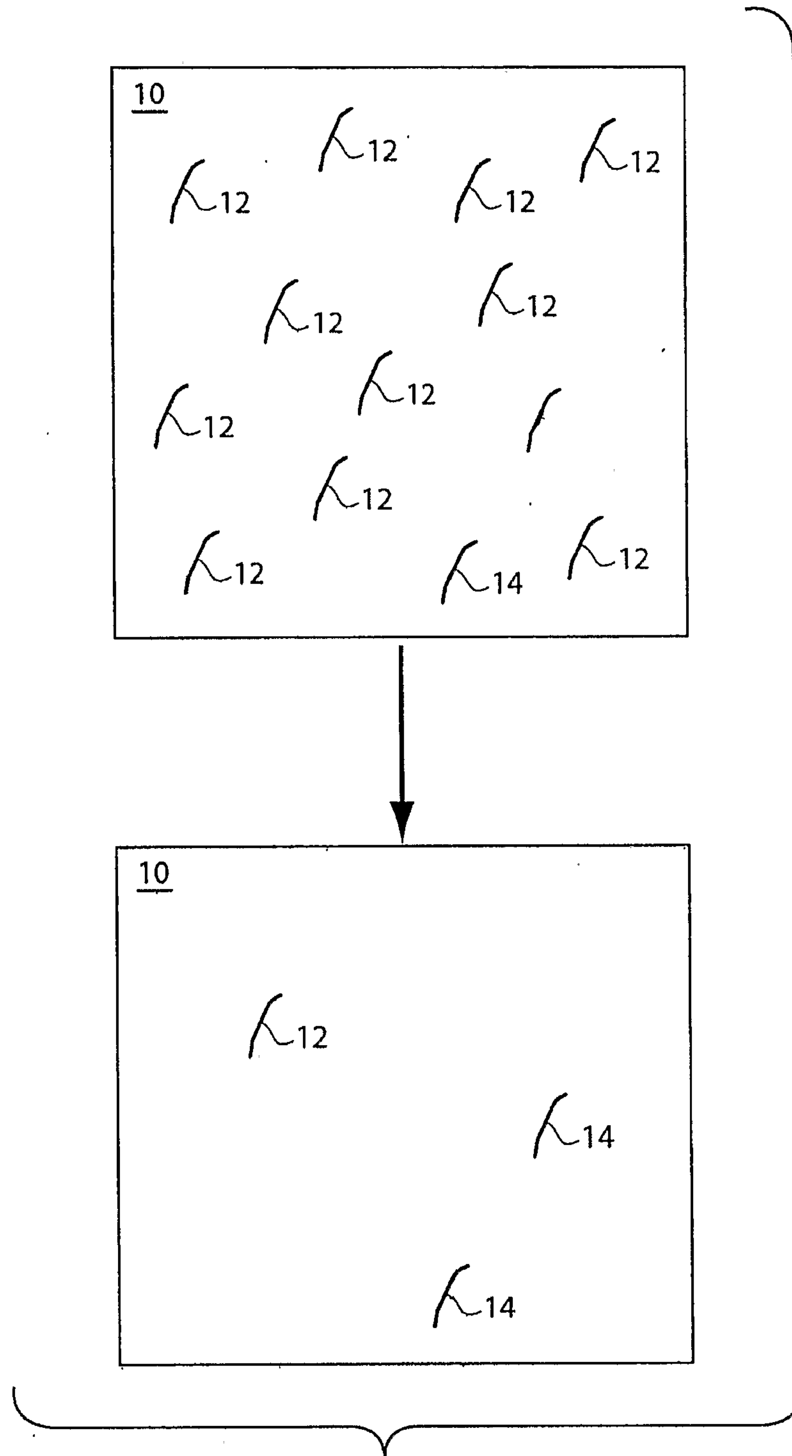


Fig. 4

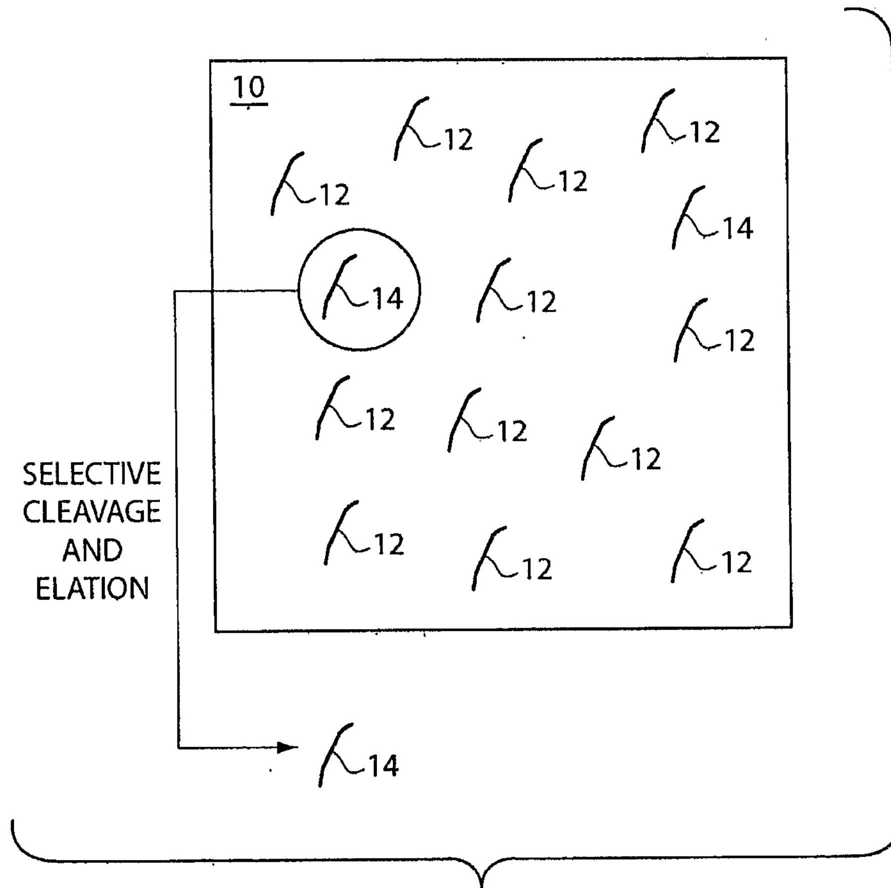


Fig. 5

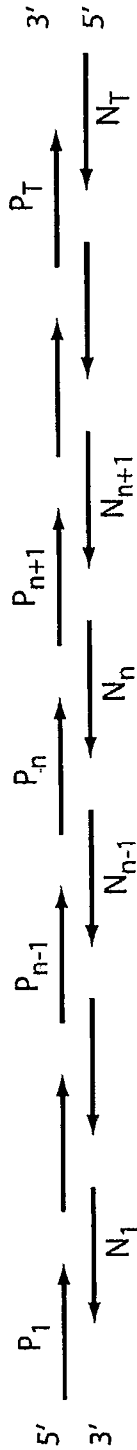


Fig. 6A

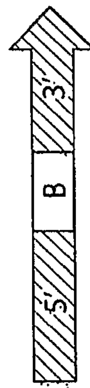


Fig. 6B

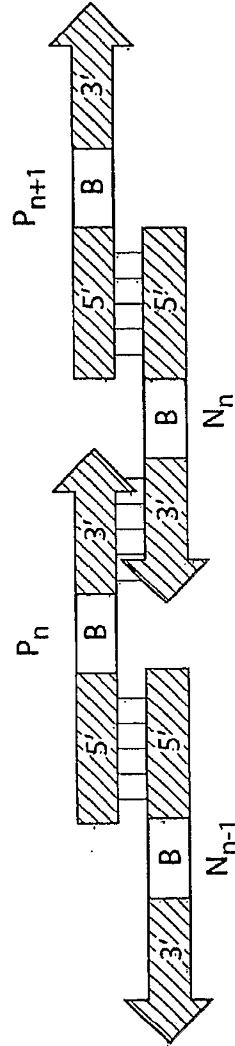


Fig. 6C

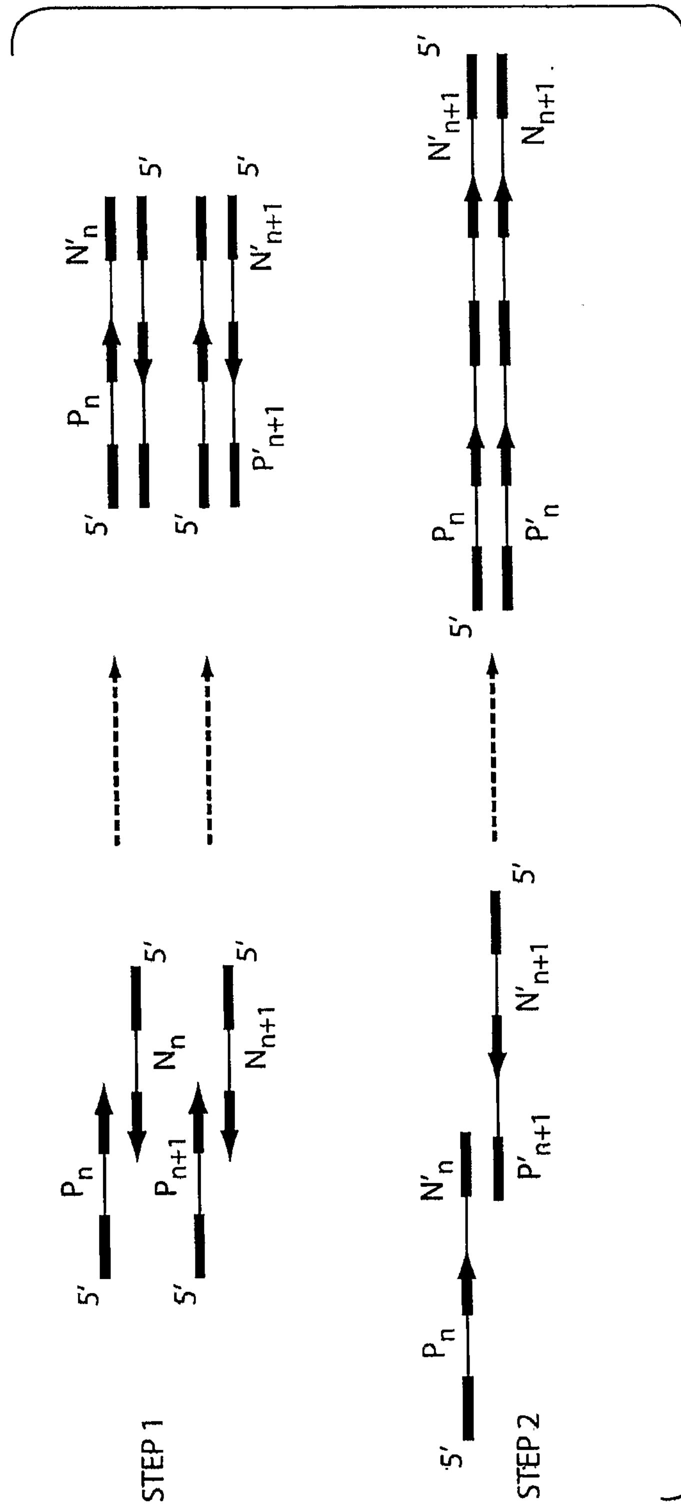


Fig. 6D

Fig. 7A

Fig. 7B

Fig. 7C

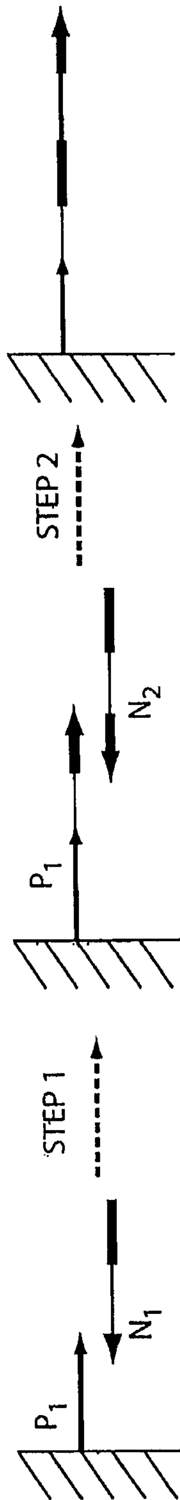


Fig. 7D

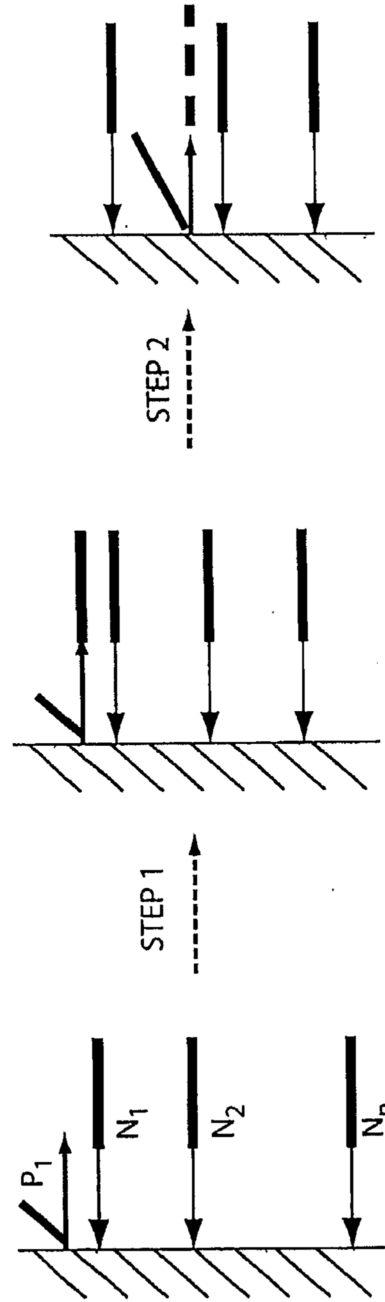


Fig. 7E

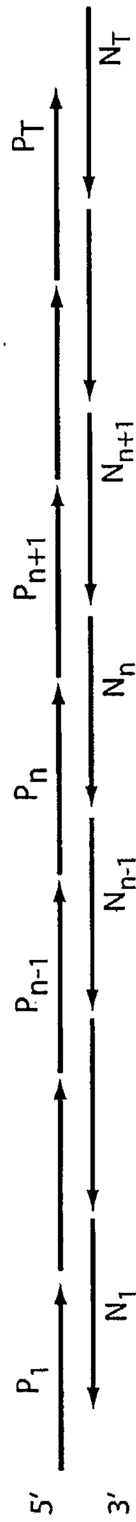


Fig. 8A

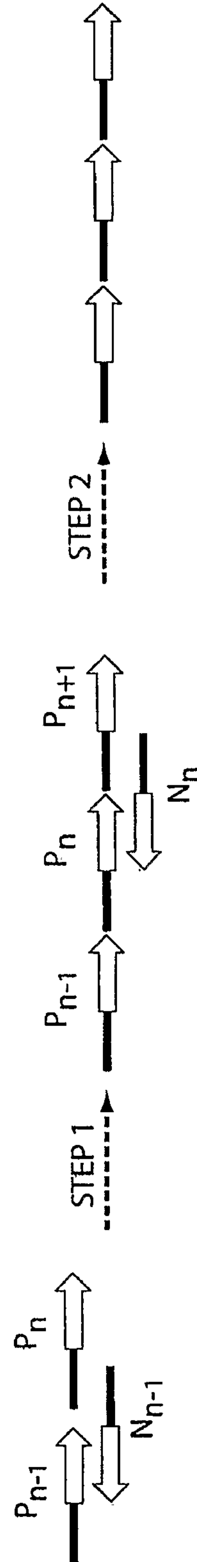


Fig. 8B

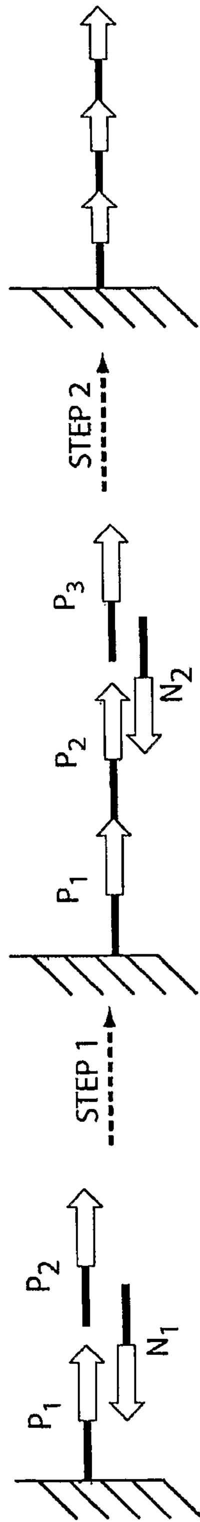


Fig. 9A

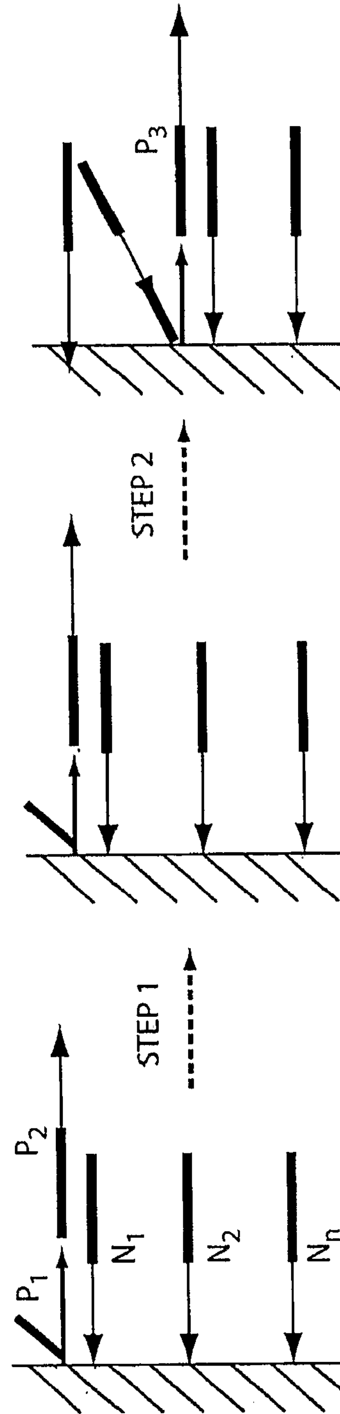


Fig. 9B

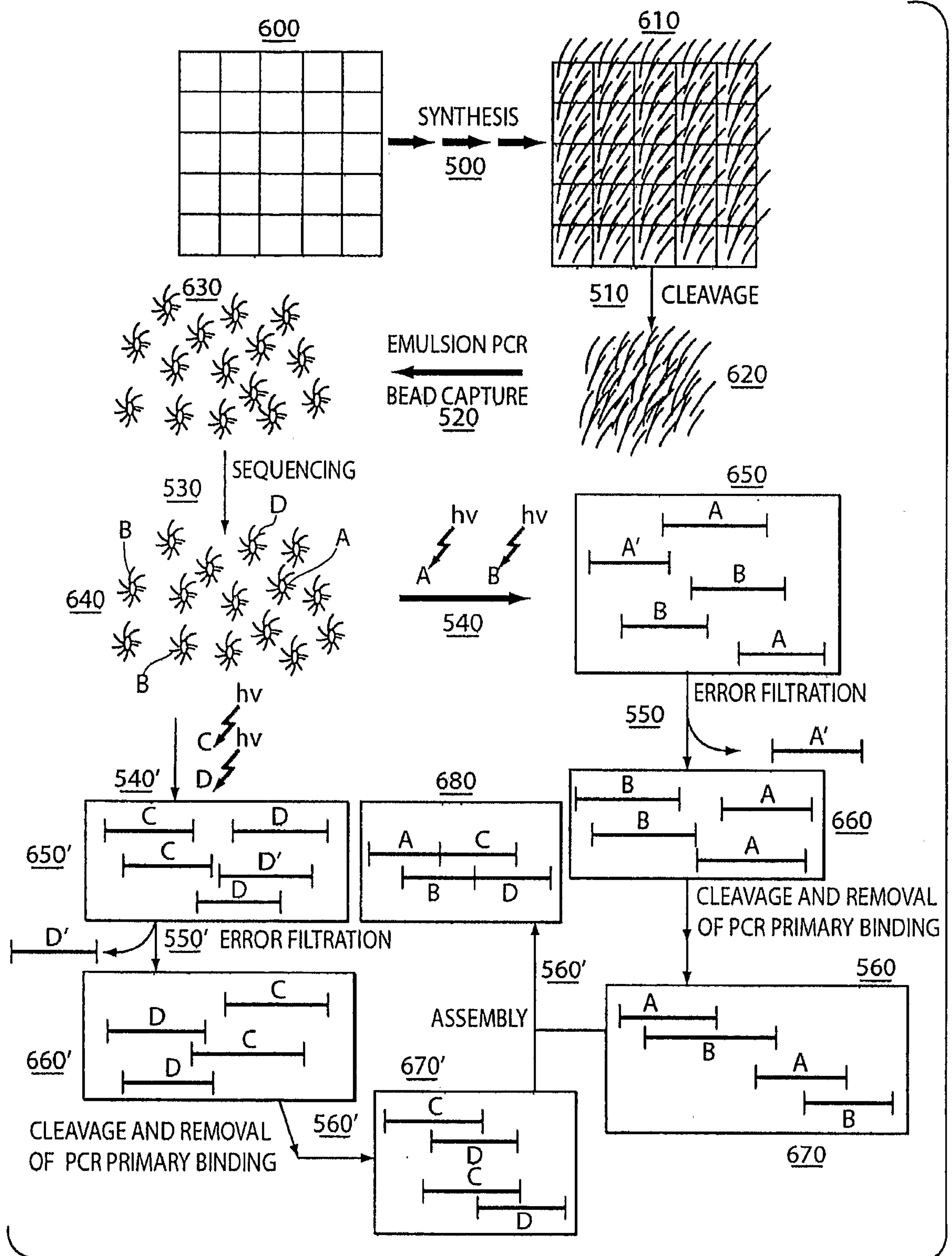


Fig. 10

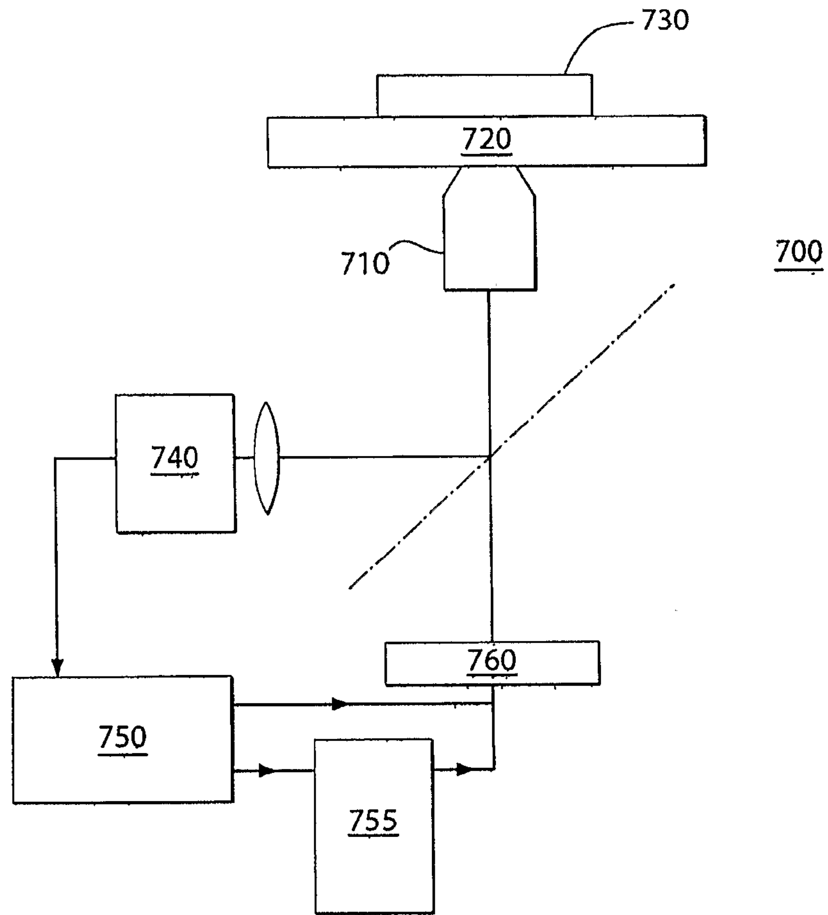


Fig. 11