

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
1 November 2007 (01.11.2007)

PCT

(10) International Publication Number
WO 2007/123742 A2

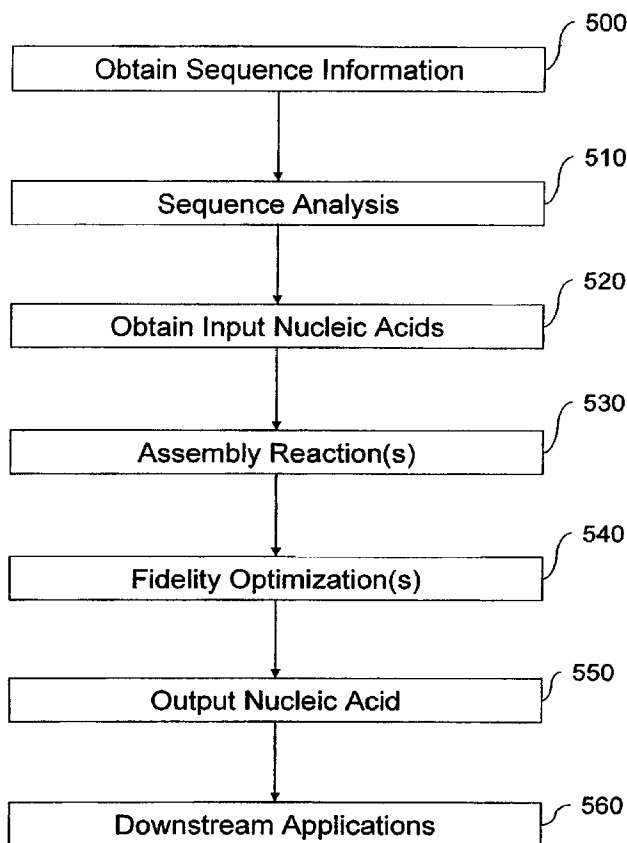
- (51) International Patent Classification:
C12Q 1/68 (2006.01)
- (21) International Application Number:
PCT/US2007/007988
- (22) International Filing Date: 30 March 2007 (30.03.2007)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/788,293 31 March 2006 (31.03.2006) US
60/788,440 31 March 2006 (31.03.2006) 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): **BAYNES, Brian, M.** [US/US]; 19 Orrin Street, Cambridge, MA 02138 (US). **NESMITH, Kenneth, G.** [US/US]; 125 Hampshire Street, Apt. 5, Cambridge, MA 02139 (US). **MARSISCHKY,**

Gerald, Thomas [US/US]; 32 Howard Street, Newton, MA 02458 (US). **CAMACHO, Jennifer, A.** [US/US]; 10 Bear Hill Road, Natick, MA 01760 (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, 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,

[Continued on next page]

(54) Title: METHODS AND COMPOSITIONS FOR INCREASING THE FIDELITY OF MULTIPLEX NUCLEIC ACID ASSEMBLY



(57) Abstract: Certain aspects of the present invention provide methods for increasing the fidelity of nucleic acid assembly reactions. Fidelity-optimized conditions may be used in certain embodiments. Nucleic acid binding proteins and/or recombinases (e.g., heat stable variants) may be used in some embodiments. A RecA protein (e.g., a heat stable RecA) may be included in some assembly reactions. Aspects of the invention also provide related kits, compositions, devices, and systems to increase the fidelity of multiplex nucleic acid assembly reaction.

WO 2007/123742 A2



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).

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.

Published:

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

METHODS AND COMPOSITIONS FOR INCREASING THE FIDELITY OF MULTIPLEX NUCLEIC ACID ASSEMBLY

RELATED APPLICATIONS

This application claims the benefit under 35 U.S.C. § 119(e) of United States provisional patent applications serial number 60/788,293, filed March 31, 2006, and serial number 60/788,440, filed March 31, 2006, the contents of both of which are incorporated herein by reference in their entirety.

FIELD OF THE INVENTION

Methods and compositions of the invention relate to nucleic acid assembly, and particularly to multiplex nucleic acid assembly reactions.

BACKGROUND

Recombinant and synthetic nucleic acids have many applications in research, industry, agriculture, and medicine. Recombinant and synthetic nucleic acids can be used to express and obtain large amounts of polypeptides, including enzymes, antibodies, growth factors, receptors, and other polypeptides that may be used for a variety of medical, industrial, or agricultural purposes. Recombinant and synthetic nucleic acids also can be used to produce genetically modified organisms including modified bacteria, yeast, mammals, plants, and other organisms. Genetically modified organisms may be used in research (e.g., as animal models of disease, as tools for understanding biological processes, etc.), in industry (e.g., as host organisms for protein expression, as bioreactors for generating industrial products, as tools for environmental remediation, for isolating or modifying natural compounds with industrial applications, etc.), in agriculture (e.g., modified crops with increased yield or increased resistance to disease or environmental stress, etc.), and for other applications. Recombinant and synthetic nucleic acids also may be used as therapeutic compositions (e.g., for modifying gene expression, for gene therapy, etc.) or as diagnostic tools (e.g., as probes for disease conditions, etc.).

Numerous techniques have been developed for modifying existing nucleic acids (e.g., naturally occurring nucleic acids) to generate recombinant nucleic acids. For example, combinations of nucleic acid amplification, mutagenesis, nuclease digestion, ligation, cloning and other techniques may be used to produce many different

recombinant nucleic acids. Chemically synthesized polynucleotides are often used as primers or adaptors for nucleic acid amplification, mutagenesis, and cloning.

Techniques also are being developed for *de novo* nucleic acid assembly whereby nucleic acids are made (e.g., chemically synthesized) and assembled to produce longer target nucleic acids of interest. For example, different multiplex assembly techniques are being developed for assembling oligonucleotides into larger synthetic nucleic acids that can be used in research, industry, agriculture, and/or medicine.

SUMMARY OF THE INVENTION

Aspects of the invention relate to multiplex nucleic acid assembly reactions. In some embodiments, methods, compositions, devices and systems of the invention are useful for enhancing the fidelity of nucleic acid assembly reactions. Some aspects of the invention relate to the use of assembly optimized conditions during the assembly of a plurality of starting nucleic acids to generate a target nucleic acid. Certain aspects relate to using one or more nucleic acid binding proteins and/or recombinases in assembly reactions.

In one aspect, the invention provides fidelity-optimized conditions for producing a target nucleic acid from a plurality of starting nucleic acids. In some embodiments, methods of assembling a target nucleic acid involve providing a plurality of starting nucleic acids each designed to have a sequence identical to a portion of a sequence of the target nucleic acid and assembling the plurality of starting nucleic acids in an assembly reaction involving one or more fidelity-optimized conditions. In certain embodiments, one or more recombinase proteins (e.g., RecA) and/or nucleic acid binding proteins are included in an assembly reaction in order to increase assembly fidelity. In some embodiments, at least two of the starting nucleic acids have complementary 5' regions. In some embodiments, at least two of the starting nucleic acids have complementary 3' regions. The assembly reaction may include a polymerase and/or a ligase. The polymerase and/or ligase may be heat stable. Similarly, a recombinase and/or nucleic acid binding protein may be heat stable (e.g., RecA). In some embodiments, a chemical assembly reaction may be used to connect two or more starting nucleic acids. For example, a ligation may be a chemical ligation. In some embodiments an assembly reaction may involve two or more cycles of denaturing, annealing, extension and/or ligation conditions. In some embodiments, the target nucleic acid may be amplified,

sequenced or cloned after it is made. In some embodiments, a host cell may be transformed with the assembled target nucleic acid. The target nucleic acid may be integrated into the genome of the host cell. In some embodiments, the target nucleic acid may encode a polypeptide. The polypeptide may be expressed (e.g., under the control of an inducible promoter). The polypeptide may be isolated or purified. A cell transformed with an assembled nucleic acid may be stored, shipped, and/or propagated (e.g., grown in culture).

Certain aspects of the invention are based, at least in part, on modified assembly reaction conditions that are not adapted for high product yield, but instead favor the assembly of correctly matched perfectly complementary sequences by selecting against the incorporation of hybridized nucleic acids (e.g., starting nucleic acids) that have even one or a few mismatches. It should be appreciated that conditions optimized for amplification yield (e.g., in a PCR reaction) focus on the efficiency and accuracy of nucleotide incorporation. Also, amplification annealing conditions are typically chosen to promote efficient primer hybridization in order to optimize yield. In contrast, in the context of a multiplex assembly reaction, the presence of errors in the starting nucleic acids can generate significant error frequencies in the end product. Accordingly, assembly conditions should be optimized to prevent incorporation of error-containing starting nucleic acids. Extension fidelity should be maintained, but assembly yield can be low since the final product can be amplified (e.g., in a PCR reaction).

Aspects of the invention provide methods of producing a target nucleic acid from a plurality of starting nucleic acids, each designed to comprise a sequence identical to a different portion of a sequence of the target nucleic acid, wherein at least two of the starting nucleic acids have complementary 5' or 3' regions. The starting nucleic acids may be assembled using one or more fidelity-optimized annealing conditions wherein starting nucleic acid duplexes comprising a mismatch are selectively destabilized relative to annealed duplexes comprising regions of complete complementarity.

Aspects of the invention provide methods of producing a target nucleic acid from a plurality of starting nucleic acids, each designed to comprise a sequence identical to a different portion of a sequence of a target nucleic acid, and wherein at least two starting nucleic acids have complementary 3' regions. The starting nucleic acids may be assembled in an assembly reaction including a high processivity polymerase under one or more fidelity-optimized conditions that cause the high processivity polymerase to

function as a low processivity polymerase. The starting nucleic acids may be assembled in an assembly reaction using a low processivity polymerase.

In some embodiments, fidelity optimized conditions may only allow less than about 50%, less than about 45%, less than about 40%, less than about 35%, less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, less than about 5%, less than about 1%, or lower or intermediate percentages of the perfectly complementary nucleic acid regions to anneal during an assembly reaction (e.g., during a first, second, third, etc., cycle of annealing of starting nucleic acid regions). According to the invention, an even lower percentage of error-containing nucleic acids would hybridize under these conditions.

In some embodiments, nucleic acid annealing may be performed at a temperature that is at least 2 °C above the T_m of the complementary regions of the starting nucleic acids (e.g., a temperature that is 3 °C, 4 °C, 5 °C, 6 °C, 7 °C, 8 °C, 9 °C, 10 °C, 11 °C, 12 °C, 13 °C, 14 °C, 15 °C, or higher above the T_m of the complementary regions of the starting nucleic acids. In some embodiments, the T_m of the complementary regions of the starting nucleic acids that is used for determining the reaction conditions is the average T_m of all of the complementary region pairs amongst the starting nucleic acids. In other embodiments, the lowest T_m amongst complementary regions of starting nucleic acid pairs is used as a reference for the reaction conditions. In further embodiments, the highest T_m amongst complementary regions of starting nucleic acid pairs is used as a reference point. In yet further embodiments, the T_m of complementary regions of any starting nucleic acid pair may be used as a reference point for determining the temperature of the extension reaction as described herein.

In any of the embodiments described herein, a low processivity polymerase may be used and/or conditions may be used to reduce polymerase processivity during annealing and/or extension. At least two cycles of denaturing, annealing, and extending conditions may be used for assembly (at least one of which may be a fidelity-optimized condition). In some embodiments, annealing and extending are performed under the same conditions. One or more fidelity-optimized conditions may destabilize annealing between nucleic acids that have a single mismatch (or 2, 3, 4, 5, 6, 7, 8, 9, 10, or more mismatches) in otherwise complementary 5' or 3' regions.

Accordingly, aspects of the invention can be used to produce a nucleic acid containing a low frequency of sequence errors (e.g., a lower frequency of sequence errors

than found in the plurality of starting nucleic acids as described herein, or lower than the frequency of errors found in a nucleic acid assembled from the same starting nucleic acids without using fidelity optimized conditions).

In some embodiments, reaction conditions including one or more of the temperature, the ionic strength, the presence of an organic solvent and/or an inorganic solvent and/or a nonionic detergent, or any combination thereof may be fidelity optimized. DMSO may be included (e.g., at a concentration near or above 10%). Formamide may be included (e.g., at a concentration near or above 10%). In some embodiments, a high stringency amount of DMSO, dimethyl formamide (DMF), formamide, glycerol, betaine, acetamide, a non-ionic detergent, Triton X-100, Nonidet P-40, Tween-20 or any combination of two or more thereof may be used. In some embodiments, reaction conditions (e.g., salt concentrations) may be substantially lower than an optimal amplification salt concentration, for example a salt concentration that is less than 50% of the optimal amplification salt concentration. In alternative embodiments, reaction conditions (e.g., salt concentrations) may be substantially higher than an optimal amplification salt concentration, for example a salt concentration that is at least 200% of the optimal amplification salt concentration. It should be understood that the optimal amplification conditions used above refer to standard amplification (e.g., PCR or LCR) conditions that are optimized for yield as opposed to fidelity.

In some embodiments, a fidelity-optimized conditions may include an extension time that is shorter than an extension time that is optimal for amplification (e.g., less than 30 seconds, less than 20 seconds, less than 10 seconds, etc.). In some embodiments, an annealing temperature may be at or about an estimated melting temperature (T_m) of the complementary 5' or 3' regions. In certain embodiments, each of the plurality of starting nucleic acids is a preparation of oligonucleotides designed to have a predetermined sequence comprising a 5' and/or 3' terminal sequence that overlaps with a terminal sequence found in only one other oligonucleotide preparation in an assembly reaction. An annealing temperature may be at or about the lowest estimated melting temperature (T_m) of overlapping complementary sequences between any two oligonucleotide preparations in the assembly reaction. An annealing temperature may be at or about the highest estimated melting temperature (T_m) of overlapping complementary sequences between any two oligonucleotide preparations in the assembly reaction. In either case,

the annealing temperature may be within $\pm 1^\circ\text{C}$ of the estimated T_m , or above the estimated T_m by about $1\text{-}2^\circ\text{C}$, about $2\text{-}5^\circ\text{C}$, about $5\text{-}10^\circ\text{C}$, about $10\text{-}15^\circ\text{C}$, or more.

In some embodiments, the starting nucleic acids are exposed to between about 3 and 30 cycles (e.g., about 5, about 10, about 15, about 20, about 25, or more cycles) of denaturing, annealing, and extending conditions. However, fewer or more cycles may be used. A low processivity polymerase (e.g., Pfu) may be used for nucleic acid extension.

In certain embodiments, an assembled target nucleic acid product may be amplified (*in vivo* and/or *in vitro*). A high processivity polymerase may be used for amplification. In some embodiments, an assembled target nucleic acid product may be inserted (e.g., cloned) into a nucleic acid vector. The vector may be transformed into a host cell. The assembled nucleic acid may be transformed directly into a host cell without a vector. In some embodiment, an assembled nucleic acid may be integrated into the genome of a host cell. Any transformed host cell may be propagated. A polypeptide product may be expressed from an assembled nucleic acid transformed into a host cell. An expressed polypeptide product may be isolated and/or purified.

Aspects of the invention relate to compositions comprising a plurality of nucleic acid complexes each comprising two different oligonucleotide molecules with overlapping complementary terminal sequences that are annealed to form a double-stranded annealed region, wherein the annealed regions comprise a nucleotide mismatch at a frequency equal to or less than $1/1,800$ nucleotides. Each nucleic acid complex may comprise two polymerase molecules associated with the annealed region.

In some embodiments, the starting nucleic acids are synthetic oligonucleotides. In certain embodiments, between about 5 and about 40 different synthetic oligonucleotides may be assembled. However, more than 40 different synthetic oligonucleotides may be assembled. In some embodiments, each synthetic oligonucleotide may be between about 20 and about 200 nucleotides long, for example between about 50 and 100 nucleotides long (e.g., about 50, about 60, about 70, about 75, about 80, about 90, or about 100 nucleotides long). However, oligonucleotides may be shorter, longer, or of intermediate length.

In some embodiments, an assembled target nucleic acid is evaluated by size analysis, restriction digestion, and/or sequencing.

Aspects of the invention relate to propagating a target nucleic acid that was assembled from a plurality of starting nucleic acids in an assembly reaction using a

fidelity-optimized reaction condition, and/or a recombinase, and/or a nucleic acid binding protein. The assembled target nucleic acid may be cloned, transformed into a host cell (which may be grown, e.g., in culture), amplified (e.g., *in vitro*, for example by PCR or LCR).

In one aspect, the invention provides a method of isolating a polypeptide by obtaining a host cell transformed with a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction including a fidelity-optimized reaction condition, and isolating, from the host cell, a polypeptide encoded by the target nucleic acid. In another aspect, the invention provides a method of isolating a polypeptide by obtaining a host cell transformed with a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction including a recombinase or nucleic acid binding protein, and isolating, from the host cell, a polypeptide encoded by the target nucleic acid.

In one aspect, the invention provides a method of isolating a polypeptide by obtaining a lysate of a host cell transformed with a target nucleic acid is assembled from a plurality of starting nucleic acids in an assembly reaction including a fidelity-optimized reaction condition, and isolating, from the lysate, a polypeptide encoded by the target nucleic acid. In another aspect, the invention provides a method of isolating a polypeptide by obtaining a lysate of a host cell transformed with a target nucleic acid is assembled from a plurality of starting nucleic acids in an assembly reaction including a recombinase and/or nucleic acid binding protein, and isolating, from the lysate, a polypeptide encoded by the target nucleic acid.

In one aspect, the invention provides a method of obtaining a target nucleic acid by sending sequence information for a target nucleic acid to a remote site, and sending a delivery address to the remote site, wherein the target nucleic acid is assembled at the remote site from a plurality of starting nucleic acids in an assembly reaction comprising a fidelity-optimized reaction condition, and wherein the target nucleic acid is delivered to the delivery address. In another aspect, the invention provides a method of obtaining a target nucleic acid by sending sequence information for a target nucleic acid to a remote site, and sending a delivery address to the remote site, wherein the target nucleic acid is assembled at the remote site from a plurality of starting nucleic acids in an assembly reaction comprising a recombinase and/or nucleic acid binding protein, and wherein the target nucleic acid is delivered to the delivery address.

Aspects of the invention also provide systems for designing and/or assembling target nucleic acids. A system may include a means for obtaining a sequence of a target nucleic acid. A system may include a means for analyzing the sequence to design a plurality of starting nucleic acids for assembly using a fidelity-optimized reaction condition and/or a recombinase and/or a nucleic acid binding protein. A system may include a means for assembling a target nucleic acid in an assembly reaction including one or more fidelity-optimized reaction conditions and/or a recombinase and/or nucleic acid binding protein. A system may be automated using a computer-implemented means. The recombinase protein may be a heat-stable recombinase protein. The recombinase may be a RecA protein.

Aspects of the invention relate to business methods that include collaboratively or independently marketing a system for designing and/or assembling a target nucleic acid using a fidelity-optimized reaction condition and/or a recombinase and/or nucleic acid binding protein. The recombinase protein may be a heat-stable recombinase protein. The recombinase may be a RecA protein.

Aspects of the invention relate to assembling a target nucleic acid from a plurality of starting nucleic acids, each designed to comprise a sequence identical to a different portion of a sequence of the target nucleic acid, wherein at least two starting nucleic acids have complementary 5' or 3' regions. The plurality of starting nucleic acids may be assembled in an assembly reaction including a recombinase protein, thereby assembling a nucleic acid product having a low frequency of sequence errors. In some embodiments, the recombinase protein is a heat-stable recombinase protein. The recombinase may be a RecA protein. The recombinase protein may be provided in a preparation including a polymerase protein. The recombinase protein may be added to an oligonucleotide preparation before adding the polymerase protein. The polymerase protein may be a low processivity polymerase protein. In some embodiments, a low concentration of polymerase protein may be used in the assembly reaction. For example, a polymerase concentration of between 0.01 U/ μ l and 0.1 U/ μ l may be used in the assembly reaction. However, higher, lower, or intermediate concentrations may be used. The recombinase protein may be used at a concentration of between 1-1000 ng/reaction. However, higher, lower, or intermediate concentrations may be used. In some embodiments, the recombinase protein may be provided in a preparation including a

ligase protein. The recombinase protein may be added to an oligonucleotide preparation before adding a ligase protein.

In another aspect, the invention provides methods of obtaining target nucleic acids by sending sequence information and delivery information to a remote site. The sequence may be analyzed at the remote site. The starting nucleic acids may be designed and/or produced at the remote site. The starting nucleic acids may be assembled in a reaction involving one or more fidelity-optimized conditions at the remote site. In some embodiments, one or more recombinases and/or nucleic acid binding proteins may be used at the remote site. In some embodiments, the starting nucleic acids, an intermediate product in the assembly reaction, and/or the assembled target nucleic acid may be shipped to the delivery address that was provided.

Other aspects of the invention provide systems for designing starting nucleic acids and/or for assembling the starting nucleic acids to make a target nucleic acid. Other aspects of the invention relate to methods and devices for automating a multiplex oligonucleotide assembly reactions that involve one or more techniques of the invention. Yet further aspects of the invention relate to business methods for marketing one or more techniques, systems, and/or automated procedures that involve a multiplex nucleic acid assembly reaction of the invention.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims. The claims provided below are hereby incorporated into this section by reference.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 illustrates non-limiting embodiments of polymerase-based multiplex oligonucleotide assembly reactions;

FIG. 2 illustrates non-limiting embodiments of sequential assembly of a plurality of oligonucleotides in polymerase-based multiplex assembly reactions;

FIG. 3 illustrates non-limiting embodiments of ligation-based multiplex oligonucleotide assembly reactions;

FIG. 4 illustrates non-limiting embodiments of ligation-based multiplex oligonucleotide assembly reactions on supports; and

FIG. 5 illustrates a non-limiting embodiment of a nucleic acid assembly procedure.

DETAILED DESCRIPTION OF THE INVENTION

Aspects of the invention relate to methods and compositions for enhancing nucleic acid assembly by increasing the fidelity of the reaction. In some embodiments, fidelity-optimized assembly reaction conditions may be used. In certain embodiments, nucleic acid binding proteins and/or recombinases may be included in an assembly reaction. Aspects of the invention may be used in one or more steps of a nucleic acid assembly procedure in order to increase the fidelity of any one or more assembly reactions (e.g., multiplex oligonucleotide assembly reactions). Increased fidelity may result from a reduced number or frequency of incorrectly assembled nucleic acids generated during the assembly procedure and/or a reduced number or frequency of misincorporation of nucleic acids containing sequence errors. Accordingly, aspects of the invention may be useful for increasing the frequency of assembled nucleic acids having a correct predetermined target sequence. As a result, fewer error correction, screening, and/or sequencing steps may be necessary when using aspects of the invention to assemble a predetermined target nucleic acid from a plurality of starting nucleic acids. In other aspects, increased fidelity of the assembly procedure provides for greater flexibility in the choice of starting nucleic acids. For example, starting nucleic acids (e.g., oligonucleotides) with higher sequence error rates may be tolerated more readily in an assembly reaction that has a higher fidelity. Increased fidelity of the assembly procedure also provides for greater flexibility in the design of the starting nucleic acids. For example, nucleic acids with shorter overlapping sequences or overlapping sequences containing one or more sequence repeats may be assembled more efficiently and/or more specifically. Therefore, aspects of the invention may be useful to increase the throughput rate of a nucleic acid assembly procedure and/or reduce the number of steps or amounts of reagent used to generate a correctly assembled nucleic acid. In certain embodiments, aspects of the invention may be useful in the context of automated nucleic acid assembly to reduce the assembly time, number of steps, amount of reagents, and other factors required for the assembly of each correct nucleic acid. Accordingly, these and other aspects of the invention may be useful to reduce the cost and time of one or more nucleic acid assembly procedures.

Aspects of the invention may be used in conjunction with *in vitro* and/or *in vivo* nucleic acid assembly reactions. Fidelity-optimized assembly conditions may be used to

increase the ratio of complementary to non-complementary sequence annealing during one or more steps of an assembly reaction. According to aspects of the invention, a nucleic acid assembly reaction involves one or more steps to assemble a plurality of nucleic acids (e.g., polynucleotides, oligonucleotides, etc.) to form a longer nucleic acid product. A nucleic acid binding protein and/or a recombinase may be used to promote annealing of complementary sequences during the assembly reaction. In certain embodiments, the plurality of nucleic acids being assembled may include at least one pair of nucleic acids having complementary sequences at one or more stages in the assembly procedure. The complementary sequences may be limited to certain portions of each nucleic acid in the pair. For example, a 3' end region of a first nucleic acid may be complementary to a 3' end region of the second nucleic acid in the pair (or the entire length of the second nucleic acid) so that the annealed pair of nucleic acids includes a double-stranded region of complementarity and at least one single-stranded 5' overhang. In some embodiments, regions of sequence complementarity may be generated during the assembly process (e.g., a sequence incorporated in one of the intermediate products of the assembly reaction may be complementary to a sequence of another nucleic acid in the assembly reaction). Methods and compositions of the invention may be used to reduce the assembly of error containing nucleic acid products using any of a variety of nucleic acid assembly procedures. Non-limiting examples of assembly reactions in with aspects of the invention may be used are described herein and illustrated in FIGS. 1-4.

According to aspects of the invention, a fidelity-optimized condition may increase the fidelity of an assembly reaction by increasing the annealing specificity (selectivity) between nucleic acids regions that have complementary sequences. Accordingly, fidelity-optimized conditions may increase the ratio of annealing of two complementary nucleic acid regions relative to the annealing of regions of similar size that are non-complementary. In some embodiments, this ratio may be increased by destabilizing the annealing of nucleic acid regions having non-complementary sequences. It should be appreciated that in some embodiments the annealing of complementary nucleic acids also may be destabilized. However, annealing specificity (selectivity) between complementary nucleic acids may be promoted if the annealing of non-complementary nucleic acids is destabilized to a greater extent than the annealing of the complementary nucleic acids. In some embodiments, a nucleic acid binding protein

or a recombinase preferentially promotes the annealing of two complementary nucleic acid regions relative to regions of similar size that are non-complementary.

Accordingly, fidelity of an assembly reaction may be increased if the annealing of complementary nucleic acid regions is promoted relative to nucleic acid regions. As used herein, complementary nucleic acid regions are regions that can anneal without any base mismatching, meaning that the sequences are completely complementary over the entire length of the complementary regions. Accordingly, nucleic acids having complementary regions can anneal to form a stretch of uninterrupted double-stranded base-pairing over the entire length of the complementary regions. In some embodiments, non-complementary sequences may be very similar to complementary sequences and may differ by as little as a single non-complementary base mismatching due to a sequence change (e.g., a base transition, transversion, deletion, or insertion) in one of two otherwise complementary nucleic acid regions. For example, the introduction of a mismatch will interrupt the length of complete complementarity between two otherwise complementary regions and result in two shorter regions of complementarity, one on either side of the mismatching. Accordingly, the original complementary regions are no longer complementary over their entire length (they may be considered non-complementary) and will form a heteroduplex if annealed. Therefore, the introduction of a single mismatch may produce two shorter complementary regions. Accordingly, the 5'-most or 3'-most complementary regions may be shortened by the presence of sequence mismatch. The presence of additional mismatches further reduces the degree of complementarity between two otherwise complementary regions and may further shorten the 5'-most or 3'-most complementary regions.

According to aspects of the invention, two major sources of errors in multiplex nucleic acid assembly reactions involve incorrect annealing of non-complementary sequences. As a result, errors may be incorporated into assembled nucleic acid products due to the extension and/or ligation of mismatched nucleic acids. In some embodiments, a mismatch may be caused by the presence of one or more sequence errors in the starting nucleic acids that are used for the assembly reactions. For example, oligonucleotides designed to have certain sequences may include sequence errors that were introduced during chemical synthesis. In other embodiments, a mismatch may occur when two regions that are not designed to be complementary to each other nonetheless anneal. This may result in nucleic acids being joined together in an incorrect order. In some

embodiments, both types of mismatches may occur. Aspects of the invention may be used to reduce the level of mismatches due to any combination of nucleic acid sequence errors and/or mismatches between nucleic acids that are not designed to anneal to each other.

According to aspects of the invention, fidelity-optimized assembly conditions may involve using one or more reagents, enzymes, reaction parameters, etc., or any combination thereof, to decrease the ratio of incorrect nucleic acid incorporation relative to correct nucleic acid incorporation during a multiplex nucleic acid assembly procedure, regardless of whether a subsequent error selection, error screening, error correction or other fidelity optimizing step is performed. In certain embodiments, fidelity-optimized assembly conditions will decrease the frequency of sequence errors (due to starting nucleic acids having sequence errors) in assembled nucleic acids relative to the frequency of sequence errors in the pool of starting nucleic acids prior to assembly. For example, the frequency of sequence errors in the assembled nucleic acids may be reduced by about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or more, relative to the frequency of sequence errors in the pool of starting nucleic acids prior to assembly. In some embodiments, one or more fidelity-optimized assembly conditions may be used to produce one or more assembled nucleic acids having fewer sequence errors (or a lower frequency of sequence errors) than the nucleic acid(s) assembled from the same pool of starting nucleic acids using assembly conditions that are not fidelity-optimized. For example, the fidelity-optimized assembly conditions may reduce the number or the frequency of sequence errors in one or more assembled nucleic acids by about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or more, relative to the number or frequency of sequence errors in nucleic acids assembled from the same pool of starting nucleic acids using assembly conditions that are not fidelity-optimized. In some embodiments, fidelity optimized assembly conditions may result in fewer sequence errors than the assembly conditions of Example 1. However, in some embodiments, the assembly conditions of Example 1 may be used. In some embodiments, an assembly reaction is optimized so that the error frequency in the assembled nucleic acids (prior to error removal, e.g., using a mismatch protein) is lower than the error frequency in the pool of starting nucleic acids used in the assembly reaction. For example, the error frequency in the assembled nucleic acids may be about 99%, about 99%-90%, about 90%-75%, about 75%-50%, about 50%-25%, about 25%-1%, about 1%-0.5%, about

0.5%-0.1%, about 0.1%-0.05%, about 0.05%-0.01%, about 0.01%-0.005%, about 0.005%-0.001%, or less of the error frequency of the pool of starting nucleic acids (e.g., of the starting oligonucleotides). In some embodiments, the error frequency may be less than about 1/500; 1/1,000; 1/1,500; 1/1,800; or even lower in a nucleic acid assembled using one or more fidelity optimized conditions. In some embodiments, the error frequency may be further reduced by performing an error screening technique (e.g., using MutS) after a nucleic acid is assembled.

In some embodiments, a polymerase-based assembly reaction may be fidelity-optimized by using a low (or relatively low) processivity polymerase in one or more of the assembly steps.

Generally, processivity is a measure of the number of deoxynucleotides that a polymerase incorporates before dissociating from the DNA template. Thus, it measures the number of deoxynucleotides incorporated per DNA-binding event (von Hippel, P. H., Fairfield, F. R. & Dolejsi, M. K. (1994) *Ann. N.Y. Acad. Sci.* **726**, 118–131; Bambara, R. A., Fay, P. J. & Mallaber, L. M. (1995) *Methods Enzymol.* **262**, 270–280).

Quantitatively, processivity, P_N , can be expressed as the probability that following each deoxynucleotide addition N , the polymerase will incorporate at least one more deoxynucleotide $N + 1$. Processivity can be thus calculated by a modification of the procedure described by von Hippel et al. (*Ann. N.Y. Acad. Sci.* **726**, 118–131; 1994). Briefly, gel-band intensities of each of the extended primers at the 120-s time point of the processivity assay were quantitated and used to calculate the percent of active polymerases that incorporated at least N deoxynucleotides by using the following equation:

$$\% \text{ active polymerases at } n = (I_N + I_{N+1} + \dots) \times 100\% / (I_1 + I_2 + \dots + I_N + \dots)$$

where I_1 is the intensity of the band resulting from one incorporation, I_N is the intensity of the band resulting from N incorporations, and so on. Processivity, P_N , is defined as the probability at deoxynucleotide incorporation N of the polymerase moving ahead to incorporate the $N + 1$ deoxynucleotide rather than dissociating from the DNA template. Taking into the account the ratio of active polymerases at N that are also active at $N + 1$, the processivity may be derived by calculating:

$$P_N = (I_{N+1} + I_{N+2} + \dots) / (I_N + I_{N+1} + I_{N+2} + \dots)$$

For further detail, see: M. Todd Washington, Robert E. Johnson, Satya Prakash, and Louise Prakash, 2000. Proc.Natl.Acad.Sci. 97: 3094–3099.

According to aspects of the invention, a lower processivity polymerase stabilizes a double-stranded nucleic acid template (e.g., a hybridized oligonucleotide duplex, an oligonucleotide hybridized to a longer nucleic acid template, etc.) to a lesser degree than a relatively higher processivity polymerase. Accordingly, a lower processivity polymerase does not stabilize a heteroduplex nucleic acid (e.g., double-stranded nucleic acid containing at least one mismatch) as much as a higher processivity polymerase. Therefore, a lower processivity polymerase may incorporate fewer error-containing nucleic acids than would be incorporated by a higher processivity polymerase during an assembly reaction.

In some embodiments, the processivity of a polymerase may be evaluated by the Gibbs free energy change associated with the polymerase binding to a nucleic acid duplex (e.g., a primer-template complex). In certain embodiments, a higher Gibbs free energy change is associated with a higher processivity polymerase. Accordingly, the relative processivity of two different polymerases may be evaluated by the relative change in Gibbs free energy associated with their binding to a nucleic acid duplex.

In some embodiments, the processivity of a polymerase may be associated with the binding affinity of the polymerase for a nucleic acid duplex. A polymerase that binds to an annealed nucleic acid substrate with higher affinity may generate a longer extended nucleic acid product than a polymerase that binds with relatively lower affinity. Accordingly, in some embodiments the processivity of a polymerase may be evaluated by the number of nucleotides that are added to a substrate template by the polymerase before it dissociates from the extended substrate template. A higher processivity polymerase may add more nucleotides than a lower processivity polymerase. In some embodiments, the processivity of a polymerase may be assayed using any appropriate assay, for example, one or more assays described in Example 3.

According to aspects of the invention, a polymerase may be a DNA polymerase, an RNA polymerase, a heat stable polymerase, a heat sensitive polymerase, or any other suitable form of nucleic acid polymerase.

In some embodiments, a low processivity polymerase may be used without a high processivity polymerase. However, a combination of low and high processivity polymerases may be used. In certain embodiments, a low (or relatively low) processivity polymerase may be used to extend annealed nucleic acids during one or more primerless assembly steps, and a high (or relatively higher) processivity polymerase may be used in one or more subsequent primer-based amplification steps used to amplify the assembled nucleic acid. In some embodiments, a purification step is used to remove a low processivity polymerase used during assembly before using a high processivity polymerase for amplification. In some embodiments, a purification step is used to remove a high processivity polymerase after amplification before using the amplified product in a subsequent assembly reaction (e.g., an assembly reaction using a low processivity polymerase or a ligase).

According to aspects of the invention, a low processivity polymerase may be an *E. coli* DNA polymerase (e.g., *E. coli* DNA polymerase I); a klenow fragment, Pfu, any other suitable low processivity polymerase; or any combination of two or more thereof. In some embodiments, a low processivity polymerase may incorporate less than 500 nucleotides (e.g., about 250 nucleotides) per extension reaction before falling off a template. Accordingly, a low processivity polymerase may be a polymerase that incorporates less than 500 nucleotides, about 250 nucleotides, or less than 250 nucleotides before dissociating from a nucleic acid template. In some embodiments, a low processivity polymerase incorporates less than about twice the number of nucleotides (e.g., about the same number or less) that are incorporated by Klenow or Pfu before they dissociate from a nucleic acid template.

A high processivity polymerase may be Taq DNA polymerase; Taq Full DNA polymerase; Vent DNA polymerase; Phusion; T7 DNA polymerase (or a modified version thereof such as Sequenase); Bio-X-Act DNA polymerase; HiFi polymerase; any other suitable high processivity polymerase; or any combination of two or more thereof. In some embodiments, a high processivity polymerase may incorporate more than 500 nucleotides (e.g., about 1,000 nucleotides) per extension reaction before falling off a template. Accordingly, a high processivity polymerase may be a polymerase that incorporates more than 500 nucleotides, about 1,000 nucleotides, or more than 1,000 nucleotides before dissociating from a nucleic acid template. In some embodiments, a high processivity polymerase may incorporate more than about 5,000, about 10,000,

about 20,000, about 30,000, or more nucleotides in an extension reaction. In some embodiments, a high processivity polymerase incorporates more than twice the number of nucleotides (e.g., 3 fold, 4 fold, 5 fold, 10 fold, 100 fold or more) that are incorporated by Klenow or Pfu before it dissociates from a nucleic acid template. In some embodiments, a high processivity polymerase may be a chimeric polymerase including a polymerase domain and a nucleic acid (e.g., DNA) binding domain, or a polymerase domain and a topoisomerase domain (e.g., topotaq). In some embodiments, a low processivity polymerase (e.g., Pfu) may become a high processivity polymerase if it is fused to a DNA binding domain.

According to aspects of the invention, many recombinant or modified enzymes that are adapted for nucleic acid sequencing are high processivity polymerases. In contrast, many natural polymerases are low processivity polymerases.

However, it should be appreciated that low and high processivity are relative properties. Accordingly, in some embodiments low and high processivity polymerases may be polymerases with different relative processivities, but that are all considered to be low or high according to other criteria described herein.

It should be appreciated that the processivity of a polymerase may be affected by certain conditions of the assembly reaction (e.g., salt, pH, buffer, temperature, solvent, etc., or any combination thereof). Accordingly, an otherwise high processivity polymerase may be used under certain conditions that promote low processivity in order to optimize the fidelity of the assembly reaction by reducing the stability of annealed starting nucleic acids containing mismatched regions relative to annealed starting nucleic acids with no mismatches. Accordingly, in some embodiments, one or more primerless assembly steps may be performed under conditions that reduce the processivity of a polymerase. In some embodiments, the conditions may be changed to increase the processivity of the polymerase during a subsequent primer-based amplification procedure. In certain embodiments, an assembled nucleic acid may be purified (at least partially) after one or more primerless assembly steps before being added to suitable reaction conditions for a high processivity amplification procedure. Similarly, in some embodiments an amplified nucleic acid may be purified (at least partially) after one or more primer-based amplification steps before being added to a subsequent primerless assembly reaction.

In some aspects of the invention, a fidelity-optimized condition may include one or more high stringency conditions that destabilize non-complementary annealing relative to annealing between nucleic acid regions that are completely complementary. It should be appreciated that fidelity-optimized conditions also may destabilize annealing between complementary nucleic acid regions to a certain degree. However, fidelity optimized conditions have a weaker destabilizing effect on annealing between regions of complete complementarity relative to their effect on annealing between non-complementary regions. For example, the destabilizing effect on non-complementary annealing may be greater than the effect on complementary annealing by about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or greater (e.g., about 3 fold, 4 fold, 5 fold, 10 fold, or more). In some embodiments, non-complementary sequences may differ from complementary sequences by as little as a single nucleotide mismatch between two regions designed to be complementary. In certain embodiments, two or more mismatches may be present. For example, potential mismatches due to incorrect nucleic acid sequences may be present at a frequency of between about 1/50 base pairs and 1/200 base pairs. However, potential mismatches also may be present in higher or lower frequencies. In certain embodiments, potential mismatching may involve annealing between regions that were not designed to be complementary (e.g., involving certain regions containing one or more direct sequence repeats or indirect sequence repeats).

In certain aspects, one or more different factors that affect the equilibrium specificity and/or kinetics of nucleic acid hybridization may be optimized. Parameters that may be varied may include, but are not limited to, reaction temperatures, reaction times, reaction pH, the presence and/or concentration of one or more co-solvents, helix destabilizers, helix stabilizers, salts, and/or buffers, and/or oligonucleotide concentration, and/or any combination of two or more thereof.

Accordingly, fidelity optimized reaction conditions may involve fidelity-optimized salt concentrations. In certain embodiments, low salt concentrations may be used. In some embodiments, high salt concentrations may be used. In certain embodiments, one or more water-excluding agents may be used. In some embodiments, organic solvents may be used. For example, one or more of the following co-solvents, additives, or helix destabilizers may be added: DMSO (e.g., 1-10%), formamide (e.g., 1-10%), glycerol (e.g., 5-10%), betaine (e.g., 0.5 to 2M, for example 1M), acetamide,

tetramethyl ammonium chloride, and/or ionic detergents such as SDS, and/or non-ionic detergents (e.g., 0.05-0.1%) such as Triton X-100, NP-40 and/or Tween 20.

In some embodiments, a reaction buffer may contain about 10-50 mM Tris, up to 50 mM KCl or NaCl, and 1-4 mM Mg²⁺. The pH of a reaction solution may be between 5 and 9, for example around 8 (e.g., ~pH 8.3). However, higher or lower concentrations and or pHs may be used. Different buffers, salts, and/or solvents also may be used.

It should be appreciated that a fidelity-optimized amount of a reagent (e.g., buffer, salt, solvent, etc.) may be one that reduces the processivity of a polymerase that is being used in an assembly reaction. In some embodiments, fidelity-optimized reaction conditions may cause a high processivity polymerase to have a low-processivity function. In certain embodiments, fidelity-optimized reaction conditions may be conditions that decrease the processivity of a polymerase (e.g., a high processivity polymerase or a low processivity polymerase) by at least 5% (e.g., by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or more) relative to its processivity in a standard buffer (e.g., a commercially available buffer) that is optimized for processivity for the polymerase.

In some embodiments, one or more temperatures may be fidelity-optimized. For example, an annealing temperature, a reaction temperature (e.g., an extension temperature and/or a ligation temperature), and/or a denaturing temperature may be fidelity optimized. Temperature optimization may involve changing a reaction temperature and/or a reaction time. In some embodiments, an annealing time may be less than 1 minute (for example, less than 45 seconds, less than 30 seconds, less than 25 seconds, less than 20 seconds, less than 15 seconds, or shorter). In certain embodiments, an extension time may be less than 2 minutes (for example, less than 1 minute, less than 45 seconds, less than 30 seconds, less than 25 seconds, less than 20 seconds, less than 15 seconds, or shorter). In some embodiments, exposure to a single temperature may be used both for annealing and reaction (e.g., extension and/or ligation). In some embodiments, this time may be less than 2 minutes (for example, less than 1 minute, less than 45 seconds, less than 30 seconds, less than 25 seconds, less than 20 seconds, less than 15 seconds, or shorter).

In some embodiments, an annealing temperature for an optimized assembly reaction may be set at or about the theoretical T_m for complementary nucleic acid sequences. Accordingly, a fidelity-optimized annealing temperature may be between

about 5 °C below a theoretical T_m and about 5 °C above a theoretical T_m. In some embodiments, the annealing temperature may be set at 5 °C below the T_m, 4 °C below the T_m, 3 °C below the T_m, 2 °C below the T_m, 1 °C below the T_m, at the T_m, 1 °C above the T_m, 2 °C above the T_m, 3 °C above the T_m, 4 °C above the T_m, 5 °C above the T_m, or higher (e.g., 5-10 °C above the T_m, 10-15 °C above the T_m, 15-20 °C above the T_m, or higher). As used herein, the T_m (melting temperature) is the temperature at a particular salt concentration and total strand concentration at which 50% of two complementary nucleic acid regions are annealed in a duplex form. The following equation may be used to calculate the theoretical T_m based on the Wallace rule:

$$T_m = 2^{\circ}\text{C}(\text{A}+\text{T}) + 4^{\circ}\text{C}(\text{G}+\text{C})$$

where A, G, C, and T are the number of occurrences of each nucleotide on one strand of the complementary region. This equation may be used for complementary regions of any size. However, in some embodiments this equation may be more accurate for complementary regions extending over about 10 to about 50 nucleotides in length (e.g., from about 10 to about 30 nucleotides in length, from about 10 to about 20 nucleotides in length, from about 15 to about 20 nucleotides in length, etc.).

It should be appreciated that in general the T_m of duplex nucleic acids is a function of a number of factors, including salt concentrations, molar concentrations of nucleic acid molecules to be assembled, as well as the activity of the polymerases used for catalyzing the reaction. The skilled artisan would know how to determine a theoretical value of T_m, based on a number of calculation methods that are known in the art. For example, melting temperature calculations may be based on the thermodynamic relationship between entropy, enthalpy, free energy and temperature, where

$$T_m = \Delta H / (\Delta S + R \ln[\text{nucleic acid}])$$

The change in entropy (order or a measure of the randomness of the oligonucleotide) and enthalpy (heat released or absorbed by the oligonucleotide) are directly calculated by summing the values for nucleotide pairs obtained by Sugimoto *et al.*, (1996) *Nucleic Acids Res* 24:4501-4505. Assuming that the concentration of sense strands and the concentration of the antisense strands are equal, the effective concentration is 0.25 the total concentration of oligonucleotide (Wetmur, J.G., (1991) *Crit Rev Biochem Mol Biol* 26:227-259. It has been determined empirically that there is a 5 kcal free energy change (3.4 by Sugimoto *et al.* and the value used by OligoCalc) during the transition from

single stranded to B-form DNA. This represents the helix initiation energy. The T_m value can be adjusted for salt concentrations by adding a component:

$$+ 16.6 \log_{10}([\text{Na}^+])$$

to the equation. Such an adjustment constant for salt concentration is not needed if the various parameters are determined at 1 Molar NaCl, and the log of 1 is zero. It should be understood that in addition to the variables that are accounted for by the T_m equation, other factors may also contribute to T_m , such as pH of a reaction and magnesium concentrations. Based on a theoretical (i.e., calculated) value of T_m under a given set of parameters as a starting point, an actual T_m may be experimentally determined as a part of routine optimization steps. Such procedures are well known in the art.

In some cases, optimization of fidelity involves lowering the nucleic acid (e.g., oligonucleotide) concentration in a reaction. Using lower concentrations of nucleic acids in the primerless PCR extension will result in a lower T_m and therefore may increase overall specificity. For example, oligonucleotide concentrations described in the present examples may be used. In some embodiments, oligonucleotide concentrations lower than those described in the examples may be used (e.g., 0.9-0.5, 0.5-0.1, 0.1-0.05, 0.05-0.01, 0.01-0.005 times the concentrations exemplified herein).

In certain embodiments, optimization of fidelity involves lowering the polymerase concentration. This may increase overall specificity of the primerless PCR extension reaction. For example, polymerase concentrations described in the present examples may be used. In some embodiments, polymerase concentrations lower than those described herein in the examples may be used (e.g., 0.9-0.5, 0.5-0.1, 0.1-0.05, 0.05-0.01, 0.01-0.005 times the concentrations exemplified herein).

In some embodiments, a fidelity-optimized assembly procedure may involve a plurality of different reaction conditions of different stringencies, one or more of which may be fidelity-optimized. It should be appreciated that certain reaction conditions (e.g., salts, reagent concentrations, buffers, solvents, etc.) that alter (e.g., lower) the T_m of nucleic acid annealing can be used to allow an annealing temperature that is several °C higher than the T_m (e.g., 1-5, 5-10, 10-15 °C, or more) without changing the actual annealing temperature (e.g., 50-60 or 60-70 °C) of the reaction.

In some cases, the reaction may employ so-called a “hot-start” strategy. This is based on the rationale that non-specific binding of primers frequently occurs and can be due to repeat sequences in the overlapping portion of nucleic acids, and non-specific annealing between overlapping nucleic acids, and incomplete primer binding. Manipulation of annealing temperature and/or magnesium ion (which stabilize DNA and RNA interactions) concentrations, and/or other reaction conditions, can generally increase specificity. Accordingly, “hot start” polymerase enzymes may be used to reduce non-specific annealing during assembly reaction preparation at lower temperatures. The active site of hot-start enzymes is blocked by an antibody or chemical that only dislodges once the reaction is sufficiently heated (e.g., to ~95°C) during the denaturation step of the first cycle, thereby minimizing non-specific assembly of nucleic acids. A number of such hot-start enzymes are known in the art.

In certain circumstances, it may be desirable to vary reaction temperatures during the course of a reaction. In some cases, a reaction is started at a temperature near or above a calculated T_m then incrementally increased or decreased. Touch down assembly amplification is a variant of assembly that reduces nonspecific primer annealing by gradually lowering the annealing temperature as cycling progresses. The annealing temperature at the initial cycles is usually a few degrees (e.g., 3, 4, 5, 6, 7, 8, 9, 10°C) above the T_m of the primers used, while at the later cycles, it is near the calculated T_m . The higher temperatures give greater specificity for assembly, and the lower temperatures permit more efficient amplification from the specific products formed during the initial cycles.

However, in some embodiments the annealing temperature may be increased in successive cycles of assembly during which both the average length of the assembled nucleic acids and the average length of complementary regions increases. Accordingly, an assembly reaction (e.g., a ligation based or a polymerase extension based reaction) may include a first annealing temperature adapted (e.g., fidelity optimized) for the length of the complementary regions of the starting nucleic acids. A subsequent annealing temperature may be increased to a second annealing temperature that is higher than the first annealing temperature (e.g., by 1, 2, 3, 4, 5 or more °C) to account for the increased length of the complementary regions due to extension (e.g., extension due to polymerization or due to ligation). The annealing temperature may be increased several times during an assembly reaction. In some embodiments, the annealing temperature is

increased in each cycle during an assembly reaction. However, it should be appreciated that the annealing temperature is not necessarily increased in each cycle but may be increased every 2, 3, 4, 5, etc. cycles.

In certain embodiments, an assembly procedure may involve two or more cycles of extension or ligation performed under different conditions. In some embodiments, two or more successive cycles may be performed under conditions of decreasing stringency. For example, the stringency of one or more of a denaturing, annealing, extension and/or ligation condition may be decreased in successive cycles of assembly. In some embodiments, the annealing temperature may be decreased in successive cycles of a polymerase-based assembly procedure. In some embodiments, the annealing temperature may be decreased in successive cycles of a ligation-based assembly procedure. For example, a ligation-based assembly may involve cycling between a high temperature (e.g., a denaturing temperature) and a low temperature (e.g., an annealing temperature). In some embodiments, the low temperature may be decreased as the cycles progress. For example, a low temperature of between about 70 °C and about 90 °C (e.g., between about 80 °C and about 90 °C, or about 85 °C) may be used in one or more early cycles. This temperature may be progressively decreased in subsequent cycles. For example, the low temperature may be decreased by between about 0.1 °C and about 5 °C (e.g., by about 1 °C) in subsequent cycles (e.g., in each cycle, every other cycle, etc.). The low temperature may be progressively decreased until it reaches about 60 °C, about 55 °C, about 50 °C, about 45 °C, about 40 °C, about 35 °C, about 30 °C, etc. According to aspects of the invention, a progressive decrease in an annealing temperature during a ligation-based assembly may increase assembly fidelity by reducing the number or percentage of incorrect ligations or ligations involving error-containing nucleic acids. For example, error-free input nucleic acids are preferentially ligated during initial cycles under high annealing temperatures. Error-free nucleic acids may be depleted during these initial cycles. As a result, error-containing nucleic acids may have reduced numbers of potential annealing partners as the annealing temperature is dropped in successive cycles. Accordingly, reduced numbers of error-containing nucleic acids may be incorporated into assembled products. It should be appreciated that a thermostable ligase may be used in such reactions. In some embodiments, input nucleic acids with relatively short overlapping complementary regions may be used to increase the sensitivity of the annealing process to the high annealing temperatures used

during the initial cycles of the assembly procedure. For example, overlapping complementary regions of less than about 20, less than about 15, less than about 10, between about 5 and about 10, about 6, about 7, about 8, about 9, or about 10 nucleotides in length may be used. Longer overlapping complementary regions may not discriminate against error-containing sequences to the same extent, because longer complementary regions may stabilize the annealing of nucleic acids containing one or more sequence errors thereby reducing the degree of preferential incorporation of error-free nucleic acids during the initial ligation cycles. Accordingly, in some embodiments, ligation-based assembly may be optimized by using input nucleic acids that are between about 15 and about 30 nucleotides long (e.g., between about 18 and about 22, or about 20 nucleotides long).

It should be appreciated that progressively decreasing annealing stringency may be obtained by changing one or more reaction conditions in addition to, or instead of, the progressively decreasing annealing temperature. For example, the reaction buffer, salt, solvent (e.g., DMSO), betaine amount, etc., may be progressively changed.

In some embodiments, a progressively decreasing annealing stringency may be used to increase the fidelity of a polymerase-based assembly reaction. For example, a polymerase-based assembly reaction may use an annealing temperature of about 72 °C during initial rounds of extension (e.g., for at least the first 1-5 cycles of extension). The annealing temperature may be decreased progressively during subsequent cycles of extension (e.g., by about 1-5 °C every 1-5 cycles) down to an annealing temperature of about 45 to 60 °C (e.g., about 50 °C). It should be appreciated that a progressively decreasing annealing stringency during a polymerase-based assembly also may be obtained by changing additional or alternative reaction conditions as discussed above.

In some embodiments, a progressively decreasing annealing stringency may be used to increase the fidelity of an enzyme-free assembly process (e.g., using a chemical assembly reaction).

In some embodiments, a nucleic acid that was assembled using one or more fidelity-optimized assembly steps described herein may be purified (at least partially) before a subsequent amplification step. Similarly, an amplified nucleic acid may be purified (at least partially) before being added to a subsequent fidelity-optimized assembly reaction. Purification may be useful to remove unincorporated nucleic acids

(e.g., oligonucleotides) or partially assembled nucleic acids that may interfere with subsequent assembly and/or amplification techniques.

It should be appreciated that one or more fidelity-optimized conditions may be used. Different fidelity-optimized conditions may be used in one or more steps of an assembly reaction. For example, if an assembly involves a plurality of cycles of denaturing, annealing, extension and/or ligation conditions, one or more different fidelity-optimized conditions may be used in different cycles.

In some embodiments, a recombinase (e.g., RecA) or nucleic acid binding protein may be used to increase the fidelity of one or more assembly reactions. In some embodiments, a heat stable RecA protein may be included in one or more reagents or steps of a multiplex nucleic acid assembly reaction. A heat stable RecA protein is disclosed, for example, in Shigemori et al., 2005, *Nucleic Acids Research*, Vol. 33, No. 14, e126. Heat stable RecA proteins may be from one or more thermophilic organisms (e.g., *Thermus thermophilus* or other thermophilic organisms). Heat stable RecA proteins also may be isolated as sequence variants of one or more heat sensitive RecA proteins.

A nucleic acid binding protein or recombinase may function via any mechanism provided that it promotes annealing of complementary sequences relative to non-complementary sequences. In some embodiments, a nucleic acid binding protein or recombinase may stabilize a perfectly annealed pair of nucleic acids (e.g., a region of perfect complementarity). For example, a nucleic acid binding protein may stabilize a region of complementarity by binding to it. In other embodiments, a nucleic acid binding protein or recombinase may destabilize an annealed pair of nucleic acids containing a mismatched region. For example, the nucleic acid binding protein or recombinase may recognize the mismatched region (e.g., a physical disruption or loop associated with a mismatching) and promote dissociation of the two nucleic acids. In other embodiments, a recombinase may promote annealing of complementary nucleic acids by promoting strand exchange or pairing between complementary nucleic acid regions. It should be appreciated that the extent of stabilization, destabilization, or strand exchange or pairing may be related to the length of the complementary regions, the number and/or extent of non-complementary mismatchings, or a combination thereof.

In some embodiments, a nucleic acid binding protein or recombinase may be heat stable. For example, it may be more active at high temperatures than at low temperatures

or it may retain a significant amount of its activity at high temperatures. High temperatures may be temperatures above 40 °C, above 50 °C, above 60 °C, above 70 °C, above 80 °C, or above 90 °C. Increased discrimination between complementary and non-complementary regions or between different lengths of complementarity may be achieved by using a heat stable nucleic acid binding protein or recombinase at a temperature above room temperature, for example at a temperature above 25 °C, above 30 °C, above 35 °C, above 40 °C, above 50 °C, above 60 °C, above 70 °C, above 80 °C, or above 90 °C. In some embodiments, the annealing of complementary nucleic acid regions relative to non-complementary nucleic acid regions may be promoted by exposing the nucleic acids to a nucleic acid binding protein or recombinase (e.g., a heat stable nucleic acid binding protein or recombinase) at a temperature that is close to the expected T_m of the complementary regions (e.g., at about 10 °C below the T_m , at about 5 °C below the T_m , at about the T_m , at about 5 °C above the T_m , or at about 10 °C above the T_m) or at a higher temperature.

In certain embodiments, a nucleic acid assembly reaction may comprise one or more endogenous and/or exogenous recombinase proteins. Recombinases are proteins that may provide a measurable increase in the recombination frequency and/or localization frequency between a targeting polynucleotide and a desired target sequence. The most common recombinase is a family of RecA-like recombination proteins all having essentially all or most of the same functions, particularly: (i) the recombinase protein's ability to properly bind to and position targeting polynucleotides on their homologous targets and (ii) the ability of recombinase protein/targeting polynucleotide complexes to efficiently find and bind to complementary endogenous sequences.

In certain embodiments, a recombinase may be a RecA protein or a homolog or ortholog thereof. Recombinases like the RecA protein of *E. coli* are proteins which promote strand pairing and exchange. The best characterized RecA protein is from *E. coli*. In *E. coli*, RecA is a component of the recombinational repair system for DNA. RecA promotes homologous recombination and is involved in several recombinational events in addition to DNA repair. RecA is involved in homology search and strand exchange reactions (see, Cox and Lehman (1987), *supra*). RecA is required for induction of the SOS repair response, DNA repair, and efficient genetic recombination in *E. coli*. RecA can catalyze homologous pairing of a linear duplex DNA and a homologous single strand DNA *in vitro*. In contrast to site-specific recombinases,

proteins like RecA which are involved in general recombination recognize and promote pairing of DNA structures on the basis of shared homology, as has been shown by several *in vitro* experiments (Hsieh and Camerini-Otero (1989) *J. Biol. Chem.* 264: 5089; Howard-Flanders et al. (1984) *Nature* 309: 215; Stasiak et al. (1984) *Cold Spring Harbor Symp. Quant. Biol.* 49: 561; Register et al. (1987) *J. Biol. Chem.* 262: 12812).

In addition to the wild-type protein, a number of mutant RecA-like proteins have been identified (e.g., recA803). RecA proteins have been identified in many bacterial systems. RecA orthologs also have been identified in eukaryotes. The Rad51 protein is an ortholog of RecA in certain eukaryotes (e.g., yeast). The Rad51 protein plays a central role in homologous recombination in yeast. In mammals, 7 RecA-like proteins have been identified: Rad51, Rad51L1/B, Rad51L2/C, Rad51L3/D, XrcC2, XrcC3, and Dmc1. Different active sequence variants of RecA also have been identified. In some embodiments, a heat stable RecA protein may be used. Furthermore, RecA-like recombinases with strand-transfer activities have been identified in many organisms (e.g., Fugisawa et al., (1985) *Nucl. Acids Res.* 13: 7473; Hsieh et al., (1986) *Cell* 44: 885; Hsieh et al., (1989) *J. Biol. Chem.* 264: 5089; Fishel et al., (1988) *Proc. Natl. Acad. Sci. USA* 85: 36-40; Cassuto et al., (1987) *Mol. Gen. Genet.* 208: 10; Ganea et al., (1987) *Mol. Cell Biol.* 7: 3124; Moore et al., (1990) *J. Biol. Chem.* 19: 11108; Keene et al., (1984) *Nucl. Acids Res.* 12: 3057; Kimeic, (1984) *Cold Spring Harbor Symp.* 48:675; Kimeic, (1986) *Cell* 44: 545; Kolodner et al., (1987) *Proc. Natl. Acad. Sci. USA* 84 :5560; Sugino et al., (1985) *Proc. Natl. Acad. Sci. USA* 85: 3683; Halbrook et al., (1989) *J. Biol. Chem.* 264: 21403; Eisen et al., (1988) *Proc. Natl. Acad. Sci. USA* 85: 7481; McCarthy et al., (1988) *Proc. Natl. Acad. Sci. USA* 85: 5854; Lowenhaupt et al., (1989) *J. Biol. Chem.* 264: 20568). Examples of such recombinase proteins include, for example, but are not limited to: RecA, RecA803, UvsX, and other RecA mutants and RecA-like recombinases (Roca, A. I. (1990) *Crit. Rev. Biochem. Molec. Biol.* 25: 415), sep1 (Kolodner et al. (1987) *Proc. Natl. Acad. Sci. (U.S.A.)* 84: 5560; Tishkoff et al. *Molec. Cell. Biol.* 11: 2593), RuvC (Dunderdale et al. (1991) *Nature* 354: 506), DST2, KEM1, XRN1 (Dykstra et al. (1991) *Molec. Cell. Biol.* 11: 2583), STP-alpha/DST1 (Clark et al. (1991) *Molec. Cell. Biol.* 11: 2576), HPP-1 (Moore et al. (1991) *Proc. Natl. Acad. Sci. (U.S.A.)* 88: 9067), other eukaryotic recombinases (Bishop et al. (1992) *Cell* 69: 439; Shinohara et al. (1992) *Cell* 69: 457). RecA may be purified from *E. coli* strains, such as *E. coli* strains JC12772 and JC15369 (available from A. J. Clark and M.

Madiraju, University of California-Berkeley). These strains contain the *recA* coding sequences on a "runaway" replicating plasmid vector present at a high copy numbers per cell. The RecA803 protein is a high-activity mutant of wild-type RecA. The art teaches several examples of recombinase proteins, for example, from *Drosophila*, yeast, plant, human, and non-human mammalian cells, including proteins with biological properties similar to RecA (i.e., RecA-like recombinases).

In some embodiments a nucleic acid binding protein may be a single-strand binding protein (e.g., SSBP), for example a single-strand DNA binding protein (e.g., gene 32 protein from phage T4, *E. coli* single-strand binding protein, etc.). In some embodiments, one or more nucleic acid binding proteins may be used along with one or more recombinases.

It should be appreciated that nucleic acid binding proteins and/or recombinases may be obtained from recombinant or natural sources.

In some aspects, methods and compositions of the invention may be useful to reduce the amount or frequency of errors in an assembly reaction. However, in other aspects, methods and compositions of the invention may be useful to modify the reaction composition or conditions. For example, increased fidelity may allow less starting material to be used (e.g., less starting nucleic acids, less polymerase, less ligase, less nucleotides, etc.).

In some embodiments, starting nucleic acids with higher error rates may be used since the presence of errors may be overcome, at least in part, by the use of one or more fidelity-optimized conditions. In some embodiments, starting nucleic acids with higher error rates may be used since the presence of errors may be overcome, at least in part, by the use of one or more nucleic acid binding protein and/or recombinase. This may result in certain cost savings since cheaper nucleic acid (e.g., cheaper oligonucleotides) may be used.

In certain embodiments, the starting nucleic acids may include sequences that have certain secondary structures or repeat sequences at their 5' and/or 3' ends, since any incorrect pairing between starting nucleic acids that may result from the presence of secondary structures or repeat sequences in the 5' or 3' ends may be reduced by the use of one or more fidelity-optimized conditions. In some embodiments, the use of nucleic acid binding protein and/or recombinase may be advantageous for sequences with high GC content. This allows for greater flexibility in the design of starting nucleic acids.

In some embodiments, starting nucleic acids may be designed with shorter overlapping sequences. Any nucleic acid mismatching due to decreased hybridization specificity associated with shorter overlaps may be reduced by the use of one or more fidelity-optimized conditions. Similarly, any nucleic acid mismatching due to decreased hybridization specificity associated with shorter overlaps may be reduced by the use of nucleic acid binding protein and/or recombinase. This also allows for greater flexibility in the design of the starting nucleic acids. This also may reduce the cost of the assembly reaction since shorter starting nucleic acids may be used to produce the same target nucleic acid.

It should be appreciated that fidelity-optimized conditions may be used at any stage during an assembly reaction that involves hybridization between complementary nucleic acid sequences. It should also be appreciated that nucleic acid binding proteins and/or recombinases may be included at any stage during the assembly where hybridization between complementary nucleic acid sequences. In some embodiments, reagents (e.g., buffers, salts, solvents, enzymes, etc.) that are useful for fidelity optimization may be provided in a combination preparation and/or in a kit. Similarly, a nucleic acid binding protein and/or a recombinase may be provided along with one or more of the other reaction components (e.g., the starting nucleic acids, nucleotides, buffers, ligase, polymerase, etc.). In some embodiments, a nucleic acid binding protein and/or recombinase (e.g., RecA) may be covalently linked to a ligase or a polymerase in order to promote assembly of correctly hybridized nucleic acids. In some embodiments, a nucleic acid binding protein and/or recombinase (e.g., RecA) may be provided attached to a support (e.g., a column, or a reaction vessel, etc.). In some embodiments a heat stable RecA (e.g., isolated from a thermophilic organism) may be attached to a ligase, a polymerase (e.g., a heat stable polymerase) or a support.

In some embodiments, a nucleic acid binding protein and/or a recombinase may be used in an assembly reaction-performed under high stringency (e.g., using high annealing temperatures, etc.).

In some embodiments, fidelity-optimized conditions may be used to increase the number of different starting nucleic acids than can be effectively assembled in a single reaction. For example, fidelity-optimized conditions may provide sufficient sequence discrimination to reduce incorrect cross-reactions between different starting nucleic acids thereby allowing a higher number of different starting nucleic acids to be combined in a

single multiplex assembly reaction. For example, fidelity-optimized conditions may be used to assemble over 100 different starting nucleic acids (e.g., 100 to 500; 500 to 1,000; 1,000 to 2,000; or more) different starting nucleic acids.

In some embodiments, a nucleic acid binding protein and/or a recombinase may be used to assemble a higher number of different starting nucleic acids in a single assembly reaction. For example, the presence of a nucleic acid binding protein and/or a recombinase may provide sufficient sequence discrimination between different starting nucleic acids to reduce problems associated with higher mismatch rates when a higher number of different starting nucleic acids are used in an assembly reaction. For example, a nucleic acid binding protein and/or a recombinase may be used to assemble over 100 different starting nucleic acids (e.g., 100 to 500; 500 to 1,000; 1,000 to 2,000; or more different starting nucleic acids) in a single multiplex reaction.

However, it should be appreciated that methods of the invention also may be used with less than 100 different starting nucleic acids (e.g., 10-50 or 50-100 different starting nucleic acids) in order to increase the fidelity of the assembly reaction. In some embodiments, a plurality (e.g., 2, 3, etc.) of related sequences (e.g., similar target sequences that are designed to have slightly different sequences) may be assembled in parallel in a single reaction mixture.

In some embodiments, certain fidelity-optimized conditions may be used to increase the nucleic acid fragment length that can be generated in a single multiplex assembly reaction thereby producing a long nucleic acid fragment. In some embodiments, a long nucleic acid fragment can be assembled from a high number of different starting nucleic acids in a single multiplex assembly reaction using a nucleic acid binding protein and/or a recombinase. A long nucleic acid fragment may be over 500 nucleotides long (e.g., 500 to 1,000; 1,000 to 2,000; 2,000 to 5,000; 5,000 to 10,000; or more nucleotides long). However, it should be appreciated that assembly reactions performed under reaction conditions of the invention also may be used to assemble nucleic acid fragments shorter than 500 nucleotides long (e.g., 100 to 500 nucleotides long or even shorter).

According to aspects of the invention, a nucleic acid binding protein and/or a recombinase protein may be provided in any suitable form at any stage in an assembly reaction. Different concentrations may be used depending on the stability and activity of the protein, the amount of nucleic acid being assembled, and the reaction conditions

being used. A useful amount of protein may be determined by performing parallel assembly reactions in the presence of different amounts of the protein and determining which concentrations promote assembly fidelity (e.g., reduce the number or frequency of sequence error incorporation). It should be appreciated that the cost of the protein also may be considered when determining the optimal amount of protein to be included in a reaction. Accordingly, a RecA protein may be added at any stage in an assembly reaction. As described herein, different types or combinations of assembly reactions may be used. A RecA protein may be provided in any one stage or at all stages.

In some embodiments, the amount of RecA (or other recombinase or nucleic acid binding protein) may be between about 1 ng and about 1,000 ng per reaction (e.g., between about 400 ng and about 500 ng per reaction). However, lower or higher amounts may be used. In some embodiments, RecA (or other recombinase or nucleic acid binding protein) may be in molar excess relative to the number of ends able for hybridization during assembly (e.g., sequences that can hybridize during annealing or extension steps of a polymerase-based assembly, during annealing of a ligase-based assembly, etc.). In some embodiments, a sufficient amount of RecA is added to form a multimeric filament on annealing nucleic acid regions. Accordingly, the amount of RecA (or other recombinase or nucleic acid binding protein) added to a reaction may vary as a function of the amount of nucleic acids that are added to the reaction for assembly.

In some embodiments, the presence of RecA (or other recombinase or nucleic acid binding protein) in an assembly reaction (e.g., in a polymerase-based, a ligase based, or a nonenzymatic assembly) may allow direct amplification of the product without involving any purification steps. In some embodiments, the presence of RecA (or other recombinase or nucleic acid binding protein) in an assembly or amplification reaction may reduce the amount of annealing of off-target sequences (e.g., including sequences that result from premature termination during a prior amplification reaction).

Aspects of the invention may be used in conjunction with any suitable multiplex nucleic acid assembly procedure involving at least two nucleic acids with complementary regions (e.g., at least one pair of nucleic acids that have complementary 3' end regions) designed to promote assembly of a target nucleic acid. For example, one or more fidelity-optimized conditions optionally employing nucleic acid binding proteins and or recombinases may be used in one or more steps of the multiplex nucleic acid

assembly procedures described below. In some embodiments, one or more nucleic acid binding proteins and/or recombinases may be used to increase the fidelity of a nucleic acid assembly reaction without using any fidelity optimized buffer or reagent conditions.

FIG. 5 illustrates a method for assembling a nucleic acid in accordance with one embodiment of the invention. Initially, in act 500, sequence information is obtained. The sequence information may be the sequence of a predetermined target nucleic acid that is to be assembled. In some embodiments, the sequence may be received in the form of an order from a customer. The order may be received electronically or on a paper copy. In some embodiments, the sequence may be received as a nucleic acid sequence (e.g., DNA or RNA). In some embodiments, the sequence may be received as a protein sequence. The sequence may be converted into a DNA sequence. For example, if the sequence obtained in act 500 is an RNA sequence, the Us may be replaced with Ts to obtain the corresponding DNA sequence. If the sequence obtained in act 500 is a protein sequence, it may be converted into a DNA sequence using appropriate codons for the amino acids. When choosing codons for each amino acid, consideration may be given to one or more of the following factors: i) using codons that correspond to the codon bias in the organism in which the target nucleic acid may be expressed, ii) avoiding excessively high or low GC or AT contents in the target nucleic acid (for example, above 60% or below 40%; e.g., greater than 65%, 70%, 75%, 80%, 85%, or 90%; or less than 35%, 30%, 25%, 20%, 15%, or 10%), and iii) avoiding sequence features that may interfere with the assembly procedure (e.g., the presence of repeat sequences or stem loop structures). However, these factors may be ignored in some embodiments as the invention is not limited in this respect. Also, aspects of the invention may be used to reduce errors caused by one or more of these factors. Accordingly, a DNA sequence determination (e.g., a sequence determination algorithm or an automated process for determining a target DNA sequence) may omit one or more steps relating to the analysis of the GC or AT content of the target nucleic acid sequence (e.g., the GC or AT content may be ignored in some embodiments) or one or more steps relating to the analysis of certain sequence features (e.g., sequence repeats, inverted repeats, etc.) that could interfere with an assembly reaction performed under standard conditions but may not to interfere with an assembly reaction including one or more fidelity-optimized conditions.

In act 510, the sequence information may be analyzed to determine an assembly strategy. This may involve determining whether the target nucleic acid will be

assembled as a single fragment or if several intermediate fragments will be assembled separately and then combined in one or more additional rounds of assembly to generate the target nucleic acid. Once the overall assembly strategy has been determined, input nucleic acids (e.g., oligonucleotides) for assembling the one or more nucleic acid fragments may be designed. The sizes and numbers of the input nucleic acids may be based in part on the type of assembly reaction (e.g., the type of polymerase-based assembly, ligase-based assembly, chemical assembly, or combination thereof) that is being used for each fragment. The input nucleic acids also may be designed to avoid 5' and/or 3' regions that may cross-react incorrectly and be assembled to produce undesired nucleic acid fragments. Other structural and/or sequence factors also may be considered when designing the input nucleic acids. In certain embodiments, some of the input nucleic acids may be designed to incorporate one or more specific sequences (e.g., primer binding sequences, restriction enzyme sites, etc.) at one or both ends of the assembled nucleic acid fragment.

In act 520, the input nucleic acids are obtained. These may be synthetic oligonucleotides that are synthesized on-site or obtained from a different site (e.g., from a commercial supplier). In some embodiments, one or more input nucleic acids may be amplification products (e.g., PCR products), restriction fragments, or other suitable nucleic acid molecules. Synthetic oligonucleotides may be synthesized using any appropriate technique as described in more detail herein. It should be appreciated that synthetic oligonucleotides often have sequence errors. Accordingly, oligonucleotide preparations may be selected or screened to remove error-containing molecules as described in more detail herein.

In act 530, an assembly reaction may be performed for each nucleic acid fragment. For each fragment, the input nucleic acids may be assembled using any appropriate assembly technique (e.g., a polymerase-based assembly, a ligase-based assembly, a chemical assembly, or any other multiplex nucleic acid assembly technique, or any combination thereof). An assembly reaction may result in the assembly of a number of different nucleic acid products in addition to the predetermined nucleic acid fragment. Accordingly, in some embodiments, an assembly reaction may be processed to remove incorrectly assembled nucleic acids (e.g., by size fractionation) and/or to enrich correctly assembled nucleic acids (e.g., by amplification, optionally followed by size fractionation). In some embodiments, correctly assembled nucleic acids may be

amplified (e.g., in a PCR reaction) using primers that bind to the ends of the predetermined nucleic acid fragment. It should be appreciated that act 530 may be repeated one or more times. For example, in a first round of assembly a first plurality of input nucleic acids (e.g., oligonucleotides) may be assembled to generate a first nucleic acid fragment. In a second round of assembly, the first nucleic acid fragment may be combined with one or more additional nucleic acid fragments and used as starting material for the assembly of a larger nucleic acid fragment. In a third round of assembly, this larger fragment may be combined with yet further nucleic acids and used as starting material for the assembly of yet a larger nucleic acid. This procedure may be repeated as many times as needed for the synthesis of a target nucleic acid. Accordingly, progressively larger nucleic acids may be assembled. At each stage, nucleic acids of different sizes may be combined. At each stage, the nucleic acids being combined may have been previously assembled in a multiplex assembly reaction. However, at each stage, one or more nucleic acids being combined may have been obtained from different sources (e.g., PCR amplification of genomic DNA or cDNA, restriction digestion of a plasmid or genomic DNA, or any other suitable source). It should be appreciated that nucleic acids generated in each cycle of assembly may contain sequence errors if they incorporated one or more input nucleic acids with sequence error(s). Accordingly, a fidelity optimization procedure may be performed after a cycle of assembly in order to remove or correct sequence errors. It should be appreciated that fidelity optimization may be performed after each assembly reaction when several successive cycles of assembly are performed. However, in certain embodiments fidelity optimization may be performed only after a subset (e.g., 2 or more) of successive assembly reactions are complete. In some embodiments, no fidelity optimization is performed.

Accordingly, act 540 is an optional fidelity optimization procedure. Act 540 may be used in some embodiments to remove nucleic acid fragments that seem to be correctly assembled (e.g., based on their size or restriction enzyme digestion pattern) but that may have incorporated input nucleic acids containing sequence errors as described herein. For example, since synthetic oligonucleotides may contain incorrect sequences due to errors introduced during oligonucleotide synthesis, it may be useful to remove nucleic acid fragments that have incorporated one or more error-containing oligonucleotides during assembly. In some embodiments, one or more assembled nucleic acid fragments may be sequenced to determine whether they contain the predetermined sequence or not.

This procedure allows fragments with the correct sequence to be identified. However, in some embodiments, other techniques may be used to remove error containing nucleic acid fragments. It should be appreciated that error containing-nucleic acids may be double-stranded homoduplexes having the error on both strands (i.e., incorrect complementary nucleotide(s), deletion(s), or addition(s) on both strands), because the assembly procedure may involve one or more rounds of polymerase extension (e.g., during assembly or after assembly to amplify the assembled product) during which an input nucleic acid containing an error may serve as a template thereby producing a complementary strand with the complementary error. In certain embodiments, a preparation of double-stranded nucleic acid fragments may be suspected to contain a mixture of nucleic acids that have the correct sequence and nucleic acids that incorporated one or more sequence errors during assembly. In some embodiments, sequence errors may be removed using a technique that involves denaturing and reannealing the double-stranded nucleic acids. In some embodiments, single strands of nucleic acids that contain complementary errors may be unlikely to reanneal together if nucleic acids containing each individual error are present in the nucleic acid preparation at a lower frequency than nucleic acids having the correct sequence at the same position. Rather, error containing single strands may reanneal with a complementary strand that contains no errors or that contains one or more different errors. As a result, error-containing strands may end up in the form of heteroduplex molecules in the reannealed reaction product. Nucleic acid strands that are error-free may reanneal with error-containing strands or with other error-free strands. Reannealed error-free strands form homoduplexes in the reannealed sample. Accordingly, by removing heteroduplex molecules from the reannealed preparation of nucleic acid fragments, the amount or frequency of error containing nucleic acids may be reduced. Any suitable method for removing heteroduplex molecules may be used, including chromatography, electrophoresis, selective binding of heteroduplex molecules, etc. In some embodiments, mismatch binding proteins that selectively (e.g., specifically) bind to heteroduplex nucleic acid molecules may be used. One example includes using MutS, a MutS homolog, or a combination thereof to bind to heteroduplex molecules. In *E. coli*, the MutS protein, which appears to function as a homodimer, serves as a mismatch recognition factor. In eukaryotes, at least three MutS Homolog (MSH) proteins have been identified; namely, MSH2, MSH3, and MSH6, and they form heterodimers. For

example in the yeast, *Saccharomyces cerevisiae*, the MSH2-MSH6 complex (also known as MutS α) recognizes base mismatches and single nucleotide insertion/deletion loops, while the MSH2-MSH3 complex (also known as MutS β) recognizes insertions/deletions of up to 12-16 nucleotides, although they exert substantially redundant functions. A mismatch binding protein may be obtained from recombinant or natural sources. A mismatch binding protein may be heat-stable. In some embodiments, a thermostable mismatch binding protein from a thermophilic organism may be used. Examples of thermostable DNA mismatch binding proteins include, but are not limited to: Tth MutS (from *Thermus thermophilus*); Taq MutS (from *Thermus aquaticus*); Apy MutS (from *Aquifex pyrophilus*); Tma MutS (from *Thermotoga maritima*); any other suitable MutS; or any combination of two or more thereof.

According to aspects of the invention, protein-bound heteroduplex molecules (e.g., heteroduplex molecules bound to one or more MutS proteins) may be removed from a sample using any suitable technique (binding to a column, a filter, a nitrocellulose filter, etc., or any combination thereof). It should be appreciated that this procedure may not be 100% efficient. Some errors may remain for at least one of the following reasons. Depending on the reaction conditions, not all of the double-stranded error-containing nucleic acids may be denatured. In addition, some of the denatured error-containing strands may reanneal with complementary error-containing strands to form an error containing homoduplex. Also, the MutS/heteroduplex interaction and the MutS/heteroduplex removal procedures may not be 100% efficient. Accordingly, in some embodiments the fidelity optimization act 540 may be repeated one or more times after each assembly reaction. For example, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more cycles of fidelity optimization may be performed after each assembly reaction. In some embodiments, the nucleic acid is amplified after each fidelity optimization procedure. It should be appreciated that each cycle of fidelity optimization will remove additional error-containing nucleic acid molecules. However, the proportion of correct sequences is expected to reach a saturation level after a few cycles of this procedure.

In some embodiments, the size of an assembled nucleic acid that is fidelity optimized (e.g., using MutS or a MutS homolog) may be determined by the expected number of sequence errors that are suspected to be incorporated into the nucleic acid during assembly. For example, an assembled nucleic acid product should include error free nucleic acids prior to fidelity optimization in order to be able to enrich for the error

free nucleic acids. Accordingly, error screening (e.g., using MutS or a MutS homolog) should be performed on shorter nucleic acid fragments when input nucleic acids have higher error rates. In some embodiments, one or more nucleic acid fragments of between about 200 and about 800 nucleotides (e.g., about 200, about 300, about 400, about 500, about 600, about 700 or about 800 nucleotides in length) are assembled prior to fidelity optimization. After assembly, the one or more fragments may be exposed to one or more rounds of fidelity optimization as described herein. In some embodiments, several assembled fragments may be ligated together (e.g., to produce a larger nucleic acid fragment of between about 1,000 and about 5,000 bases in length, or larger), and optionally cloned into a vector, prior to fidelity optimization as described herein.

At act 550, an output nucleic acid is obtained. As discussed herein, several rounds of act 530 and/or 540 may be performed to obtain the output nucleic acid, depending on the assembly strategy that is implemented. The output nucleic acid may be amplified, cloned, stored, etc., for subsequent uses at act 560. In some embodiments, an output nucleic acid may be cloned with one or more other nucleic acids (e.g., other output nucleic acids) for subsequent applications. Subsequent applications may include one or more research, diagnostic, medical, clinical, industrial, therapeutic, environmental, agricultural, or other uses.

Aspects of the invention may include automating one or more acts described herein. For example, an analysis may be automated in order to generate an output automatically. Acts of the invention may be automated using, for example, a computer system.

Aspects of the invention may be used in conjunction with any suitable multiplex nucleic acid assembly procedure. For example, fidelity-optimized conditions may be used in connection with one or more of the multiplex nucleic acid assembly procedures described below. In some embodiments, one or more recombinase and/or nucleic acid binding proteins may be used in connection with 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 non-functional 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

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 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, 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 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 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-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 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 anti-sense 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 different oligonucleotides may be designed to have different lengths. In some embodiments, one or more different oligonucleotides may 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 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 nucleotides long (e.g., between about 10 and 100, between about 10 and 75, between about 10 and 50, about 20, about 25, about 30, about 35, about 40, 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 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 VentR® (exo-) DNA Polymerase (thermophilic DNA polymerases with or without a 3' → 5' exonuclease/proofreading activity from *Thermococcus litoralis*; also known as Tli 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* KOD1 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); Sequenase™ (T7 DNA polymerase deficient in 3'-5' exonuclease activity); Phi29 (bacteriophage 29 DNA

polymerase, may be used for rolling circle amplification, for example, in a TempliPhi™ 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; Phusion™ (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 Y & Kool ET, 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 Y 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 Y & Kool ET, 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 ZA et al., 1991, *Nuc. 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, *Nucleic Acids Research*, Vol. 32, No 20, e162 (9 pages); Richmond et al., 2004, *Nucleic Acids Research*, Vol. 32, No 17, pp. 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* Vol. 101, No. 44, pp. 15573-15578; Tian et al., 2004, *Nature*, Vol. 432, pp. 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. 1-4. It should be appreciated that multiplex nucleic acid assembly reactions 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.

It should be appreciated that the reference to 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.

FIG. 1 shows one embodiment of a plurality of oligonucleotides that may be assembled in a polymerase-based multiplex oligonucleotide assembly reaction. Figure 1A 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 ($P_1, P_2, \dots, P_{n-1}, P_n, P_{n+1}, \dots, P_T$, shown from $5' \rightarrow 3'$ on the positive strand). Group N includes oligonucleotides with negative strand sequences ($N_T, \dots, N_{n+1}, N_n, N_{n-1}, \dots, N_2, N_1$, shown from $5' \rightarrow 3'$ 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. 1A 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 N_T) overlaps with complementary regions of two oligonucleotides from the complementary group of oligonucleotides. P_1 and N_T overlap with a complementary region of only one oligonucleotide from the other group (the complementary 3'-most oligonucleotides N_1 and P_T , respectively). FIG. 1B shows a structure of an embodiment of a Group P or Group N oligonucleotide represented in FIG. 1A. 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. 1B. 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 (P_1 in Group P and N_T 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 (P_1 and N_T) that are illustrated in FIG. 1A each have a 3' complementary region and a 5' non-complementary region (the B region of FIG. 1B), 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 N_1 or P_T shown in FIG. 1A). 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. 1C illustrates a subset of the oligonucleotides from FIG. 1A, each oligonucleotide having a 5', a 3', and an optional B region. Oligonucleotide P_n is shown with a 5' region that is complementary to (and can anneal to) the 5' region of oligonucleotide N_{n-1} . Oligonucleotide P_n also has a 3' region that is complementary to (and can anneal to) the 3' region of oligonucleotide N_n . N_n is also shown with a 5' region that is complementary (and can anneal to) the 5' region of oligonucleotide P_{n+1} . This pattern could be repeated for all of oligonucleotides P_2 to P_T and N_1 to N_{T-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. 1A. 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. 1C would result in extension of the 3' ends only of oligonucleotides P_n and N_n generating

extended oligonucleotides containing sequences that are complementary to all the regions of N_n and P_n , respectively. Extended oligonucleotide products with sequences complementary to all of N_{n-1} and P_{n+1} would not be generated unless oligonucleotides P_{n-1} and N_{n+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. 1A. In some embodiments, a ligase also may be 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. 1A.

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. 1A, a single cycle of polymerase extension would result in the extension of the 3' ends of the P_1/N_1 , P_2/N_2 , ..., P_{n-1}/N_{n-1} , P_n/N_n , P_{n+1}/N_{n+1} , ..., P_T/N_T 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. 1D 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 $\log_2 n$, 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. 1D, annealed oligonucleotide pairs

P_n/N_n and P_{n+1}/N_{n+1} are extended to form oligonucleotide dimer products incorporating the sequences covered by the respective oligonucleotide pairs. For example, P_n is extended to incorporate sequences that are complementary to the B and 5' regions of N_n (indicated as N'_n in FIG. 1D). Similarly, N_{n+1} is extended to incorporate sequences that are complementary to the 5' and B regions of P_{n+1} (indicated as P'_{n+1} in FIG. 1D). These dimer products may be denatured and reannealed to form the starting material of step 2 where the 3' end of the extended P_n oligonucleotide is annealed to the 3' end of the extended N_{n+1} oligonucleotide. This product may be extended in a polymerase-mediated reaction to form a product that incorporates the sequences of the four oligonucleotides ($P_n, N_n, P_{n+1}, N_{n+1}$). One strand of this extended product has a sequence that includes (in 5' to 3' order) the 5', B, and 3' regions of P_n , the complement of the B region of N_n , the 5', B, and 3' regions of P_{n+1} , and the complements of the B and 5' regions of N_{n+1} . The other strand of this extended product has the complementary sequence. It should be appreciated that the 3' regions of P_n and N_n are complementary, the 5' regions of N_n and P_{n+1} are complementary, and the 3' regions of P_{n+1} and N_{n+1} are complementary. It also should be appreciated that the reaction products shown in FIG. 1D 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 P_1/N_1 to P_T/N_T . 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. 1D. 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. 2 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. 1, 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 (N_1) has a 3' region that is complementary to the 3' region of P_1 as shown in FIG. 2B. However, the remainder of the oligonucleotides in Group N have overlapping (but non-complementary) 3' and 5' regions as illustrated in FIG. 2B for oligonucleotides N_1 - N_3 . Each Group N oligonucleotide (e.g., N_n) overlaps with two adjacent oligonucleotides: one overlaps with the 3' region (N_{n-1}) and one with the 5' region (N_{n+1}), except for N_1 that overlaps with the 3' regions of P_1 (complementary overlap) and N_2 (non-complementary overlap), and N_T that overlaps only with N_{T-1} . It should be appreciated that all of the overlaps shown in FIG. 2A between adjacent oligonucleotides N_2 to N_{T-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. 2A may involve multiple cycles of polymerase-mediated extension. Each extension cycle may be separated by a denaturing and an annealing step. FIG. 2C illustrates the first two steps in this assembly process. In step 1, annealed oligonucleotides P_1 and N_1 are extended to form an oligonucleotide dimer. P_1 is shown with a 5' region that is non-complementary to the 3' region of N_1 and extends beyond the 3' region of N_1 when the oligonucleotides are annealed. However, in some embodiments, P_1 may lack the 5' non-complementary region and include only sequences that overlap with the 3' region of N_1 . The product of P_1 extension is shown after step 1 containing an extended region that is complementary to the 5' end of N_1 . The single

strand illustrated in FIG. 2C may be obtained by denaturing the oligonucleotide dimer that results from the extension of P_1/N_1 in step 1. The product of P_1 extension is shown annealed to the 3' region of N_2 . 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 N_2 . Again, the single strand illustrated in FIG. 2C 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 P_1/N_1 extension that was initiated after the first cycle.

FIG. 2D illustrates an example of a sequential extension reaction where the 5'-most P_1 oligonucleotide is bound to a support and the Group N oligonucleotides are unbound. The reaction steps are similar to those described for FIG. 2C. However, an extended predetermined nucleic acid fragment will be bound to the support via the 5'-most P_1 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. 2E 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. 2C. However, an extended

predetermined nucleic acid fragment will be bound to the support via the 5'-most N_T 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.

It should be appreciated that other configurations of oligonucleotides may be used to assemble a nucleic acid via two or more cycles of polymerase-based extension. In many configurations, at least one pair of oligonucleotides have complementary 3' end regions. FIG. 2F illustrates an example where an oligonucleotide pair with complementary 3' end regions is flanked on either side by a series of oligonucleotides with overlapping non-complementary sequences. The oligonucleotides illustrated to the right of the complementary pair have overlapping 3' and 5' regions (with the 3' region of one oligonucleotide being identical to the 5' region of the adjacent oligonucleotide) that corresponding to a sequence of one strand of the target nucleic acid to be assembled. The oligonucleotides illustrated to the left of the complementary pair have overlapping 3' and 5' regions (with the 3' region of one oligonucleotide being identical to the 5' region of the adjacent oligonucleotide) that correspond to a sequence of the complementary strand of the target nucleic acid. These oligonucleotides may be assembled via sequential polymerase-based extension reactions as described herein (see also, for example, Xiong et al., 2004, *Nucleic Acids Research*, Vol. 32, No. 12, e98, 10 pages, the disclosure of which is incorporated by reference herein). It should be appreciated that different numbers and/or lengths of oligonucleotides may be used on either side of the complementary pair. Accordingly, the illustration of the complementary pair as the central pair in FIG. 2F is not intended to be limiting as other configuration of a complementary oligonucleotide pair flanked by a different number of non-complementary pairs on either side may be used according to methods of the invention.

FIG. 3 shows an embodiment of a plurality of oligonucleotides that may be assembled in a ligase reaction. FIG. 3A 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. 3B 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. 3B, 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. 4 shows an embodiment of a ligase-based assembly where one or more of the plurality of oligonucleotides is bound to a support. In FIG. 4A, 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. 4A 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 N_T for Group N) or the 3'-most (e.g., P_T 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. 4B 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. 4B 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. 4B) 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'/5', 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 non-overlapping 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 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-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 (prior to, as part of, or after, a purification procedure) using appropriate 5' and 3' amplification primers.

Synthetic Oligonucleotides:

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 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.

Oligonucleotides may be synthesized using any suitable technique. For example, oligonucleotides may be synthesized on a column or other support (e.g., a chip). Examples of chip-based synthesis techniques include techniques used in synthesis devices or methods available from Combimatrix, Agilent, Affymetrix, or other sources. A synthetic oligonucleotide may be of any suitable size, for example between 10 and 1,000 nucleotides long (e.g., between 10 and 200, 200 and 500, 500 and 1,000 nucleotides long, or any combination thereof). An assembly reaction may include a plurality of oligonucleotides, each of which independently may be between 10 and 200 nucleotides in length (e.g., between 20 and 150, between 30 and 100, 30 to 90, 30-80, 30-70, 30-60, 35-55, 40-50, or any intermediate number of nucleotides). However, one or more shorter or longer oligonucleotides may be used in certain embodiments.

Oligonucleotides may be provided as single stranded synthetic products. However, in some embodiments, oligonucleotides may be provided as double-stranded preparations including an annealed complementary strand. Oligonucleotides may be molecules of DNA, RNA, PNA, or any combination thereof. A double-stranded oligonucleotide may be produced by amplifying a single-stranded synthetic oligonucleotide or other suitable template (e.g., a sequence in a nucleic acid preparation such as a nucleic acid vector or genomic nucleic acid). Accordingly, a plurality of oligonucleotides designed to have the sequence features described herein may be provided as a plurality of single-stranded oligonucleotides having those feature, or also may be provided along with complementary oligonucleotides.

In some embodiments, an oligonucleotide may be amplified using an appropriate primer or primer pair with one primer corresponding to each end of the oligonucleotide (e.g., one that is complementary to the 3' end of the oligonucleotide and one that is identical to the 5' end of the oligonucleotide). In some embodiments, an oligonucleotide may be designed to contain a central assembly sequence (designed to be incorporated into the target nucleic acid) flanked by a 5' amplification sequence (e.g., a 5' universal sequence) and/or a 3' amplification sequence (e.g., a 3' universal sequence). Amplification primers (e.g., between 10 and 50 nucleotides long, between 15 and 45 nucleotides long, about 25 nucleotides long, etc.) corresponding to the flanking amplification sequences may be used to amplify the oligonucleotide (e.g., one primer may be complementary to the 3' amplification sequence and one primer may have the same sequence as the 5' amplification sequence). The amplification sequences then may be removed from the amplified oligonucleotide using any suitable technique to produce an oligonucleotide that contains only the assembly sequence. In some embodiments, a plurality of different oligonucleotides (e.g., about 5, 10, 50, 100, or more) with different central assembly sequences may have identical 5' amplification sequences and/or identical 3' amplification sequences. These oligonucleotides can all be amplified in the same reaction using the same amplification primers. Any oligonucleotide amplification with at least one primer may involve a polymerase and optionally a recombinase.

A preparation of an oligonucleotide designed to have a certain sequence may include oligonucleotide molecules having the designed sequence in addition to oligonucleotide molecules that contain errors (e.g., that differ from the designed sequence at least at one position). A sequence error may include one or more nucleotide

deletions, additions, substitutions (e.g., transversion or transition), inversions, duplications, or any combination of two or more thereof. Oligonucleotide errors may be generated during oligonucleotide synthesis. Different synthetic techniques may be prone to different error profiles and frequencies. In some embodiments, error rates may vary from 1/10 to 1/200 errors per base depending on the synthesis protocol that is used. However, in some embodiments lower error rates may be achieved. Also, the types of errors may depend on the synthetic techniques that are used. For example, in some embodiments chip-based oligonucleotide synthesis may result in relatively more deletions than column-based synthetic techniques.

In some embodiments, one or more oligonucleotide preparations may be processed to remove (or reduce the frequency of) error-containing oligonucleotides. In some embodiments, a hybridization technique may be used wherein an oligonucleotide preparation is hybridized under stringent conditions one or more times to an immobilized oligonucleotide preparation designed to have a complementary sequence. Oligonucleotides that do not bind may be removed in order to selectively or specifically remove oligonucleotides that contain errors that would destabilize hybridization under the conditions used. It should be appreciated that this processing may not remove all error-containing oligonucleotides since many have only one or two sequence errors and may still bind to the immobilized oligonucleotides with sufficient affinity for a fraction of them to remain bound through this selection processing procedure.

In some embodiments, a nucleic acid binding protein or recombinase (e.g., RecA) may be included in one or more of the oligonucleotide processing steps to improve the selection of error free oligonucleotides. For example, by preferentially promoting the hybridization of oligonucleotides that are completely complementary with the immobilized oligonucleotides, the amount of error containing oligonucleotides that are bound may be reduced. As a result, this oligonucleotide processing procedure may remove more error-containing oligonucleotides and generate an oligonucleotide preparation that has a lower error frequency (e.g., with an error rate of less than 1/50, less than 1/100, less than 1/200, less than 1/300, less than 1/400, less than 1/500, less than 1/1,000, or less than 1/2,000 errors per base).

A plurality of oligonucleotides used in an assembly reaction may contain preparations of synthetic oligonucleotides, single-stranded oligonucleotides, double-stranded oligonucleotides, amplification products, oligonucleotides that are processed to

remove (or reduce the frequency of) error-containing variants, etc., or any combination of two or more thereof.

In some aspects, a synthetic oligonucleotide may be amplified prior to use. Either strand of a double-stranded amplification product may be used as an assembly oligonucleotide and added to an assembly reaction as described herein. A synthetic oligonucleotide may be amplified using a pair of amplification primers (e.g., a first primer that hybridizes to the 3' region of the oligonucleotide and a second primer that hybridizes to the 3' region of the complement of the oligonucleotide). The oligonucleotide may be synthesized on a support such as a chip (e.g., using an ink-jet-based synthesis technology). In some embodiments, the oligonucleotide may be amplified while it is still attached to the support. In some embodiments, the oligonucleotide may be removed or cleaved from the support prior to amplification. The two strands of a double-stranded amplification product may be separated and isolated using any suitable technique. In some embodiments, the two strands may be differentially labeled (e.g., using one or more different molecular weight, affinity, fluorescent, electrostatic, magnetic, and/or other suitable tags). The different labels may be used to purify and/or isolate one or both strands. In some embodiments, biotin may be used as a purification tag. In some embodiments, the strand that is to be used for assembly may be directly purified (e.g., using an affinity or other suitable tag). In some embodiments, the complementary strand is removed (e.g., using an affinity or other suitable tag) and the remaining strand is used for assembly.

In some embodiments, a synthetic oligonucleotide may include a central assembly sequence flanked by 5' and 3' amplification sequences. The central assembly sequence is designed for incorporation into an assembled nucleic acid. The flanking sequences are designed for amplification and are not intended to be incorporated into the assembled nucleic acid. The flanking amplification sequences may be used as universal primer sequences to amplify a plurality of different assembly oligonucleotides that share the same amplification sequences but have different central assembly sequences. In some embodiments, the flanking sequences are removed after amplification to produce an oligonucleotide that contains only the assembly sequence.

In some embodiments, one of the two amplification primers may be biotinylated. The nucleic acid strand that incorporates this biotinylated primer during amplification can be affinity purified using streptavidin (e.g., bound to a bead, column, or other

surface). In some embodiments, the amplification primers also may be designed to include certain sequence features that can be used to remove the primer regions after amplification in order to produce a single-stranded assembly oligonucleotide that includes the assembly sequence without the flanking amplification sequences.

In some embodiments, the non-biotinylated strand may be used for assembly. The assembly oligonucleotide may be purified by removing the biotinylated complementary strand. In some embodiments, the amplification sequences may be removed if the non-biotinylated primer includes a dU at its 3' end, and if the amplification sequence recognized by (i.e., complementary to) the biotinylated primer includes at most three of the four nucleotides and the fourth nucleotide is present in the assembly sequence at (or adjacent to) the junction between the amplification sequence and the assembly sequence. After amplification, the double-stranded product is incubated with T4 DNA polymerase (or other polymerase having a suitable editing activity) in the presence of the fourth nucleotide (without any of the nucleotides that are present in the amplification sequence recognized by the biotinylated primer) under appropriate reaction conditions. Under these conditions, the 3' nucleotides are progressively removed through to the nucleotide that is not present in the amplification sequence (referred to as the fourth nucleotide above). As a result, the amplification sequence that is recognized by the biotinylated primer is removed. The biotinylated strand is then removed. The remaining non-biotinylated strand is then treated with uracil-DNA glycosylase (UDG) to remove the non-biotinylated primer sequence. This technique generates a single-stranded assembly oligonucleotide without the flanking amplification sequences. It should be appreciated that this technique may be used to process a single amplified oligonucleotide preparation or a plurality of different amplified oligonucleotides in a single reaction if they share the same amplification sequence features described above.

In some embodiments, the biotinylated strand may be used for assembly. The assembly oligonucleotide may be obtained directly by isolating the biotinylated strand. In some embodiments, the amplification sequences may be removed if the biotinylated primer includes a dU at its 3' end, and if the amplification sequence recognized by (i.e., complementary to) the non-biotinylated primer includes at most three of the four nucleotides and the fourth nucleotide is present in the assembly sequence at (or adjacent to) the junction between the amplification sequence and the assembly sequence. After

amplification, the double-stranded product is incubated with T4 DNA polymerase (or other polymerase having a suitable editing activity) in the presence of the fourth nucleotide (without any of the nucleotides that are present in the amplification sequence recognized by the non-biotinylated primer) under appropriate reaction conditions. Under these conditions, the 3' nucleotides are progressively removed through to the nucleotide that is not present in the amplification sequence (referred to as the fourth nucleotide above). As a result, the amplification sequence that is recognized by the non-biotinylated primer is removed. The biotinylated strand is then isolated (and the non-biotinylated strand is removed). The isolated biotinylated strand is then treated with UDG to remove the biotinylated primer sequence. This technique generates a single-stranded assembly oligonucleotide without the flanking amplification sequences. It should be appreciated that this technique may be used to process a single amplified oligonucleotide preparation or a plurality of different amplified oligonucleotides in a single reaction if they share the same amplification sequence features described above.

It should be appreciated that the biotinylated primer may be designed to anneal to either the synthetic oligonucleotide or to its complement for the amplification and purification reactions described above. Similarly, the non-biotinylated primer may be designed to anneal to either strand provided it anneals to the strand that is complementary to the strand recognized by the biotinylated primer.

In certain embodiments, it may be helpful to include one or more modified oligonucleotides in an assembly reaction. An oligonucleotide may be modified by incorporating a modified-base (e.g., a nucleotide analog) during synthesis, by modifying the oligonucleotide after synthesis, or any combination thereof. Examples of modifications include, but are not limited to, one or more of the following: universal bases such as nitroindoles, dP and dK, inosine, uracil; halogenated bases such as BrdU; fluorescent labeled bases; non-radioactive labels such as biotin (as a derivative of dT) and digoxigenin (DIG); 2,4-Dinitrophenyl (DNP); radioactive nucleotides; post-coupling modification such as dR-NH₂ (deoxyribose-NH₂); Acridine (6-chloro-2-methoxyacridine); and spacer phosphoramides which are used during synthesis to add a spacer 'arm' into the sequence, such as C3, C8 (octanediol), C9, C12, HEG (hexaethylene glycol) and C18.

It should be appreciated that one or more nucleic acid binding proteins or recombinases are preferably not included in a post-assembly fidelity optimization

technique (e.g., a screening technique using a MutS or MutS homolog), because the optimization procedure involves removing error-containing nucleic acids via the production and removal of heteroduplexes. Accordingly, any nucleic acid binding proteins or recombinases (e.g., RecA) that were included in the assembly steps is preferably removed (e.g., by inactivation, column purification or other suitable technique) after assembly and prior to fidelity optimization.

Applications:

Aspects of the invention may be useful for a range of applications involving the production and/or use of synthetic nucleic acids. As described herein, the invention provides methods for producing synthetic nucleic acids with increased fidelity and/or for reducing the cost and/or time of synthetic assembly reactions. 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 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 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.

Many of the techniques described herein can be used together, applying fidelity-optimized conditions at one or more stages in an assembly to produce long nucleic acid molecules. For example, fidelity-optimized assembly conditions may be used to assemble oligonucleotide duplexes and nucleic acid fragments of less than 100 to more than 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 mers, 50,000 mers, 75,000 mers, 100,000 mers, etc.). In an exemplary embodiment, methods described herein may be used 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 viral, bacterial, yeast, or other prokaryotic or eukaryotic organism), optionally incorporating specific modifications into the sequence at one or more desired locations. Similarly, the embodiments described herein may be used in conjunction with nucleic acid binding proteins and/or recombinases.

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 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 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, hexahistidine (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 a 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. In some embodiments, reaction reagents may include one or more fidelity-optimized reagents or reaction conditions. In some embodiments, one or more nucleic acid binding proteins and/or recombinases may be included. 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 fidelity-optimized reaction temperatures or temperature cycles. Similarly, subsequent purification and analysis of 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 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). 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 manipulation and/or transfer of samples and/or sample containers, including automated pipetting devices, 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 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, 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 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, 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 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 implemented 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:

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 fidelity-optimized nucleic acid assembly techniques described herein. In some embodiments, the invention involves marketing techniques, compositions, kits, devices and systems that include or are adapted for use with one or more nucleic acid binding proteins and/or recombinases (e.g., RecA), for example with one or more heat stable nucleic acid binding proteins and/or recombinases.

Aspects of the invention may be useful for reducing the time and/or cost of 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 invention may involve marketing a procedure and/or associated devices or systems involving correct sequence enrichment using fidelity-optimized assembly techniques described herein. In some embodiments, the invention involves marketing a procedure and/or associated devices or systems for nucleic acid assembly reactions including one or more nucleic acid binding proteins and/or recombinases. 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.

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 revenue through sales and/or licenses of methods, kits, compositions, devices, and/or systems of the invention.

EXAMPLES

Example 1. Nucleic acid fragment assembly.

Gene assembly via a 2-step PCR method: In step (1), a primerless assembly of oligonucleotides is performed and in step (2) an assembled nucleic acid fragment is amplified in a primer-based amplification.

A 993 base long promoter>EGFP construct was assembled from 50-mer abutting oligonucleotides using a 2-step PCR assembly.

Mixed oligonucleotide pools were prepared as follows: 36 overlapping 50-mer oligonucleotides and two 5' terminal 59-mers were separated into 4 pools, each corresponding to overlapping 200-300 nucleotide segments of the final construct. The total oligonucleotide concentration in each pool was 5 μ M.

A primerless PCR extension reaction was used to stitch (assemble) overlapping oligonucleotides in each pool. The PCR extension reaction mixture was as follows:

oligonucleotide pool (5 μ M total)	1.0 μ l (~ 25 nM final each)
dNTP (10 mM each)	0.5 μ l (250 μ M final each)
Pfu buffer (10x)	2.0 μ l
Pfu polymerase (2.5 U/ μ l)	0.5 μ l
dH ₂ O to 20 μ l	

Assembly was achieved by cycling this mixture through several rounds of denaturing, annealing, and extension reactions as follows:

start 2 min. 95°C
 30 cycles of 95°C 30 sec., 65°C 30 sec., 72°C 1 min.
 final 72°C 2 min. extension step

The resulting product was exposed to amplification conditions to amplify the desired nucleic acid fragments (sub-segments of 200-300 nucleotides). The following PCR mix was used:

primerless PCR product	1.0 μ l
primer 5' (1.2 μ M)	5 μ l (300 nM final)
primer 3' (1.2 μ M)	5 μ l (300 nM final)
dNTP (10 mM each)	0.5 μ l (250 μ M final each)
Pfu buffer (10x)	2.0 μ l
Pfu polymerase (2.5 U/ μ l)	0.5 μ l
dH ₂ O to 20 μ l	

The following PCR cycle conditions were used:

start 2 min. 95°C
 35 cycles of 95°C 30 sec., 65°C 30 sec., 72°C 1 min.
 final 72°C 2 min. extension step

The amplified sub-segments were assembled using another round of primerless PCR as follows. A diluted amplification product was prepared for each sub-segment by diluting each amplified sub-segment PCR product 1:10 (4 μ l mix + 36 μ l dH₂O). This diluted mix was used as follows:

diluted sub-segment mix	1.0 μ l
dNTP (10mM each)	0.5 μ l (250 μ M final each)
Pfu buffer (10x)	2.0 μ l
Pfu polymerase (2.5 U/ μ l)	0.5 μ l
dH ₂ O to 20 μ l	

The following PCR cycle conditions were used:

start 2 min. 95°C
 30 cycles of 95°C 30 sec., 65°C 30 sec., 72°C 1 min.
 final 72°C 2 min. extension step

The full-length 993 nucleotide long promoter>EGFP was amplified in the following PCR mix:

assembled sub-segments	1.0 μ l
primer 5' (1.2 μ M)	5 μ l (300 nM final)
primer 3' (1.2 μ M)	5 μ l (300 nM final)

dNTP (10 mM each)	0.5 μ l (250 μ M final each)
Pfu buffer (10x)	2.0 μ l
Pfu polymerase (2.5 U/ μ l)	0.5 μ l
dH ₂ O to 20 μ l	

The following PCR cycle conditions were used:

start 2 min. 95°C
35 cycles of 95°C 30 sec., 65°C 30 sec., 72°C 1 min.
final 72°C 2 min. extension step

Example 2. Fidelity-optimized Conditions.

A 2 step nucleic acid assembly process as described above (a primerless oligonucleotide assembly followed by a primer-based amplification) was performed under several different conditions to assemble an approximately 850 base pair nucleic acid fragment. Different annealing conditions were tested (e.g., 65°C and 70°C). Different extension times were tested (e.g., 30 seconds, and 1 minute). Different buffers were tested (e.g., with or without 3% DMSO).

The resulting assembled nucleic acids were cloned and sequenced. Error rates ranged from 1/500 (using an annealing temperature of 65°C) to 1/1,200 (using a 30 second extension at 72°C in the presence of DMSO).

Example 3. Polymerase Processivity Assays.

In some embodiments, the processivity of a polymerase (e.g., a DNA polymerase) may be assayed as follows. A polymerase in a reaction solution (e.g., 120 mM Tris-HCl buffer solution (pH 8.0), 1 mM magnesium chloride, 10 mM KCl, 6 mM (NH₄)₂SO₄, 0.1% TritonX-100, 10 μ g/ml BSA, 0.2 mM dNTP) may be mixed with substrate nucleic acid (e.g., DNA consisting of single-stranded M13mp18 DNA including annealed primers that may be labeled for example with ³²P at the 5'-end) at an appropriate reaction temperature (e.g., at 75 °C). The concentrations may such that the substrate is present in a several-fold molar excess relative to the polymerase. Such conditions may be used to determine the number of nucleotides synthesized without release of the polymerase from the substrate DNA (e.g., during a given time period). After a certain reaction time, the reaction may be terminated by adding a reaction

terminating solution (e.g., 50 mM sodium hydroxide, 10 mM EDTA, 5% Ficoll, 0.05% Bromophenol Blue) in a volume equal to the reaction mixture.

The nucleic acids synthesized in the above reaction may be fractionated by electrophoresis on an alkaline agarose gel, and the gel may be dried and subjected to autoradiography. Processivity can be determined by measuring the number of extended nucleotides with reference to a marker band (e.g., using several labeled reference nucleic acids of known size).

In some embodiments, the processivity of a polymerase may be assessed by assaying its ability to initiate primer extension using primers that have different degrees of similarity with a template nucleic acid. For example, an amount of substrate may be assayed for a polymerase using one or more primers that contain one or more mismatches relative to a template. In certain embodiments, the amount of substrate may be compared to the amount generated in an assay using a primer that is perfectly complementary to the template. In certain embodiments, the amount of substrate generated using different primers with different numbers of mismatches may be assayed and compared.

EQUIVALENTS

The present invention provides among other things methods for assembling nucleic acids using fidelity-optimized conditions, whereby large polynucleotide constructs and organisms having increased genomic stability may be generated. While specific embodiments of the subject invention have been discussed, the above specification is illustrative and not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

INCORPORATION BY REFERENCE

All publications, patents and sequence database entries mentioned herein, including those items listed below, are hereby incorporated by reference in their entirety as if each individual publication or patent was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.

CLAIMS

1. A method of producing a target nucleic acid, the method comprising:
providing a plurality of starting nucleic acids, wherein each starting nucleic acid is designed to comprise a sequence identical to a different portion of a sequence of a target nucleic acid, and wherein at least two starting nucleic acids have complementary 5' or 3' regions, and
assembling said plurality of starting nucleic acids in an assembly reaction using one or more fidelity-optimized annealing conditions wherein starting nucleic acid duplexes comprising a mismatch are selectively destabilized relative to annealed duplexes comprising regions of complete complementarity,
thereby producing a target nucleic acid having a low frequency of sequence errors.
2. The method of claim 1, wherein the assembly reaction comprises one or more fidelity-optimized conditions wherein nucleic acid annealing is performed at a temperature that is at least 2 °C above the T_m of the complementary regions of the starting nucleic acids.
3. The method of claim 2, wherein the annealing is performed at a temperature that is 3 °C, 4 °C, 5 °C, 6 °C, 7 °C, 8 °C, 9 °C, 10 °C, 11 °C, 12 °C, 13 °C, 14 °C or 15 °C above the T_m of the complementary regions of the starting nucleic acids.
4. A method of producing a target nucleic acid, the method comprising:
providing a plurality of starting nucleic acids, wherein each starting nucleic acid is designed to comprise a sequence identical to a different portion of a sequence of a target nucleic acid, and wherein at least two starting nucleic acids have complementary 3' regions, and
assembling said plurality of starting nucleic acids in an assembly reaction comprising a high processivity polymerase and one or more fidelity-optimized conditions under which the high processivity polymerase functions as a low processivity polymerase,

thereby producing a target nucleic acid having a low frequency of sequence errors.

5. The method of any one of claims 1-4, wherein the assembly reaction comprises a low processivity polymerase.

6. A method of producing a target nucleic acid, the method comprising:
providing a plurality of starting nucleic acids, wherein each starting nucleic acid is designed to comprise a sequence identical to a different portion of a sequence of a target nucleic acid, and wherein at least two starting nucleic acids have complementary 3' regions, and
assembling said plurality of starting nucleic acids in an assembly reaction using a low processivity polymerase,
thereby producing a target nucleic acid having a low frequency of sequence errors.

7. The method of any one of the preceding claims, wherein at least two starting nucleic acids are designed to have complementary 3' regions.

8. The method of any one of claims 1-3, wherein at least two starting nucleic acids are designed to have complementary 5' regions.

9. The method of any one of the preceding claims, wherein the assembly reaction comprises exposing the plurality of starting nucleic acids to at least two cycles of denaturing, annealing, and extending conditions, at least one of which is a fidelity-optimized condition.

10. The method of claim 9, wherein annealing and extending are performed under the same conditions.

11. The method of any one of the preceding claims, wherein the one or more fidelity-optimized conditions destabilize annealing between nucleic acids that have a single mismatch in otherwise complementary 5' or 3' regions.

12. The method of any one of the preceding claims, wherein the one or more fidelity-optimized conditions comprise an extending condition that reduces polymerase processivity.
13. The method of any one of the preceding claims, wherein the at least two starting nucleic acids anneal with a probability of less than 75% under at least one fidelity-optimized condition.
14. The method of any one of the preceding claims, wherein at least one fidelity-optimized condition is an annealing condition comprising a fidelity optimized temperature, an ionic strength, an organic or inorganic solvent, a nonionic detergent, or combination thereof.
15. The method of claim 14, wherein the fidelity-optimized condition comprises a dimethyl sulfoxide (DMSO) concentration near or above 10%.
16. The method of claim 14, wherein the fidelity-optimized condition comprises a formamide concentration near or above 10%.
17. The method of claim 14, wherein the fidelity-optimized condition comprises a high stringency amount of DMSO, dimethyl formamide (DMF), formamide, glycerol, betaine, acetamide, a non-ionic detergent, Triton X-100, Nonidet P-40, Tween-20 or any combination of two or more thereof.
18. The method of any one of claims 1-6, wherein the one or more fidelity-optimized conditions comprise a salt concentration substantially lower than an optimal amplification salt concentration.
19. The method of claim 18, wherein the salt concentration is less than 50% of the optimal amplification salt concentration.

20. The method of any one of claims 1-6, wherein the one or more fidelity-optimized conditions comprise a salt concentration substantially higher than an optimal amplification salt concentration.
21. The method of claim 20, wherein the salt concentration is at least 200% of the optimal amplification salt concentration.
22. The method of any one of claims 1-6, wherein the one or more fidelity-optimized conditions comprise an extension time that is shorter than an extension time that is optimal for amplification.
23. The method of claim 22, wherein the extension time is 30 seconds or less.
24. The method of any one of claims 1, 4, 5, and 6, wherein the fidelity-optimized condition comprises an annealing temperature at or about an estimated melting temperature (T_m) of the complementary 5' or 3' regions.
25. The method of any one of claims 1, 4, 5, and 6, wherein each of the plurality of starting nucleic acids is a preparation of oligonucleotides designed to have a predetermined sequence comprising a 5' and/or 3' terminal sequence that overlaps with a terminal sequence found in only one other oligonucleotide preparation.
26. The method of claim 25, wherein at least one fidelity-optimized condition comprises an annealing temperature at or about the lowest estimated melting temperature (T_m) of overlapping complementary sequences between any two oligonucleotide preparations in the assembly reaction.
27. The method of claim 19, wherein at least one fidelity-optimized condition comprises an annealing temperature at or about the highest estimated melting temperature (T_m) of overlapping complementary sequences between any two oligonucleotide preparations in the assembly reaction.

28. The method of any one of claims 23-26 wherein the annealing temperature is within ± 1 °C of the estimated T_m .
29. The method of any one of claims 23-26 wherein the annealing temperature is about 1-2 °C above the estimated T_m .
30. The method of any one of claims 23-26 wherein the annealing temperature is about 2-5 °C above the estimated T_m .
31. The method of any one of claims 1-6, wherein the starting nucleic acids are exposed to 25 cycles of the denaturing, annealing, and extending conditions.
32. The method of any one of claims 1-5, wherein a low processivity polymerase is used for nucleic acid extension.
33. The method of claim 32, wherein the low processivity polymerase is a Pfu polymerase.
34. The method of any one of claims 1-6 and 32, wherein a nucleic acid amplification reaction is performed to amplify the target nucleic acid.
35. The method of claim 34, wherein the target nucleic acid is assembled using a low processivity polymerase and amplified using a high processivity polymerase.
36. A method of preparing a target nucleic acid, the method comprising the step(s) of:
assembling a target nucleic acid molecule from a plurality of overlapping nucleic acids using a low processivity polymerase; and,
amplifying the target nucleic acid using a high processivity polymerase.
37. A composition comprising a plurality of nucleic acid complexes each comprising two different oligonucleotide molecules with overlapping complementary terminal sequences that are annealed to form a double-stranded annealed region, wherein the

annealed regions comprise a nucleotide mismatch at a frequency equal to or less than 1/1,800 nucleotides.

38. The composition of claim 37, wherein each nucleic acid complex comprises two polymerase molecules associated with the annealed region.

39. The method of any one of claims 1-6, further comprising the act of amplifying the target nucleic acid.

40. The method of any one of claims 1-6, further comprising inserting the target nucleic acid into a vector.

41. The method of claim 40, further comprising transforming a host cell with the vector.

42. The method of any one of claims 1-6, further comprising transforming a host cell with the target nucleic acid.

43. The method of claim 42, wherein the target nucleic acid is integrated into the genome of the host cell.

44. The method of claim 41 or 42, further comprising propagating the host cell after it is transformed.

45. The method of any one of claims 1-6, further comprising expressing a polypeptide product encoded by the target nucleic acid.

46. The method of claim 45, further comprising isolating the polypeptide product.

47. The method of claim 41 or 42, further comprising expressing in said host cell a polypeptide product encoded by the target nucleic acid.

48. The method of claim 47, further comprising isolating the polypeptide product.

49. The method of any one of claims 1-6, wherein the starting nucleic acids are synthetic oligonucleotides.
50. The method of claim 49, wherein between 5 and 40 different synthetic oligonucleotides are assembled in the assembly reaction.
51. The method of claim 49, wherein more than 40 different synthetic oligonucleotides are assembled in the assembly reaction.
52. The method of claim 49, wherein each synthetic oligonucleotide is between about 50 and about 100 nucleotides long.
53. The method of claim 52 wherein each synthetic oligonucleotide is about 75 nucleotides long.
54. The method of any one of claims 1-6, further comprising sequencing the target nucleic acid.
55. A method of propagating a target nucleic acid, the method comprising:
obtaining a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a fidelity-optimized reaction condition,
and
cloning the target nucleic acid.
56. A method of propagating a target nucleic acid, the method comprising:
obtaining a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a fidelity-optimized reaction condition,
and
transforming a host cell with the target nucleic acid.
57. A method of propagating a target nucleic acid, the method comprising:

obtaining a host cell transformed with a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a fidelity-optimized reaction condition, and
growing the transformed host cell in culture.

58. A method of propagating a target nucleic acid, the method comprising:
obtaining a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a fidelity-optimized reaction condition, and
amplifying the target nucleic acid.
59. The method of claim 58, wherein the target nucleic acid is amplified in vitro.
60. A method of isolating a polypeptide, the method comprising:
obtaining a host cell transformed with a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a fidelity-optimized reaction condition, and
isolating, from the host cell, a polypeptide encoded by the target nucleic acid.
61. A method of isolating a polypeptide, the method comprising:
obtaining a lysate of a host cell transformed with a target nucleic acid is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a fidelity-optimized reaction condition, and
isolating, from the lysate, a polypeptide encoded by the target nucleic acid.
62. A method of obtaining a target nucleic acid, the method comprising:
sending sequence information for a target nucleic acid to a remote site, and
sending a delivery address to the remote site, wherein the target nucleic acid is assembled at the remote site from a plurality of starting nucleic acids in an assembly reaction comprising a fidelity-optimized reaction condition, and wherein the target nucleic acid is delivered to the delivery address.
63. A system for assembling a target nucleic acid, the system comprising:

a means for obtaining a plurality of starting nucleic acids, and
a means for assembling a target nucleic acid in an assembly reaction comprising the plurality of starting nucleic acids using a fidelity-optimized reaction condition.

64. A system for designing a plurality of starting nucleic acids to be assembled into a target nucleic acid, the system comprising:

a means for obtaining a sequence of a target nucleic acid, and
a means for analyzing the sequence to design a plurality of starting nucleic acids, wherein the plurality of starting nucleic acids are designed to be assembled in a reaction comprising a fidelity-optimized reaction condition.

65. The system of claim 63 or 64, wherein the system is automated using computer-implemented means.

66. A business method comprising:

providing a system for assembling a target nucleic acid using a fidelity-optimized reaction condition, and
collaboratively or independently marketing said system.

67. A business method comprising:

providing a system for designing starting nucleic acids to be assembled into a target nucleic acid in a fidelity-optimized reaction condition, and
collaboratively or independently marketing said system.

68. The method, composition or system of any one of the preceding claims, further comprising a recombinase protein.

69. A method of assembling a target nucleic acid, the method comprising:

providing a plurality of starting nucleic acids, wherein each starting nucleic acid is designed to comprise a sequence identical to a different portion of a sequence of a target nucleic acid, and wherein at least two starting nucleic acids have complementary 5' or 3' regions, and

assembling said plurality of starting nucleic acids in an assembly reaction comprising a recombinase protein,
thereby assembling a nucleic acid product having a low frequency of sequence errors.

70. The method or composition of claim 69, wherein the recombinase protein is a heat-stable recombinase protein.

71. The method or composition of claim 69 or 70, wherein the recombinase protein is a RecA protein.

72. The method of claim 69, wherein the recombinase protein is provided in a preparation including a polymerase protein.

73. The method of claim 69, wherein the recombinase protein is added to an oligonucleotide preparation before adding a polymerase protein.

74. The method of claim 72 or 73, wherein the polymerase protein is a low processivity polymerase protein.

75. The method of any one of claims 69-74, wherein a low concentration of polymerase protein is used in the assembly reaction.

76. The method of claim 75, wherein the polymerase protein is used at a concentration of between 0.01 U/ μ l and 0.1 U/ μ l in the assembly reaction.

77. The method of claim 69, wherein the recombinase protein is used at a concentration of between 1-1000 ng/reaction.

78. The method of claim 69, wherein the recombinase protein is provided in a preparation including a ligase protein.

79. The method of claim 69, wherein the recombinase protein is added to an oligonucleotide preparation before adding a ligase protein.
80. A method of propagating a target nucleic acid, the method comprising:
obtaining a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a recombinase protein, and
cloning the target nucleic acid.
81. A method of propagating a target nucleic acid, the method comprising:
obtaining a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a recombinase protein, and
transforming a host cell with the target nucleic acid.
82. A method of propagating a target nucleic acid, the method comprising:
obtaining a host cell transformed with a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a recombinase protein, and
growing the transformed host cell in culture.
83. A method of propagating a target nucleic acid, the method comprising:
obtaining a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a recombinase protein, and
amplifying the target nucleic acid.
84. The method of claim 83, wherein the target nucleic acid is amplified in vitro.
85. A method of isolating a polypeptide, the method comprising:
obtaining a host cell transformed with a target nucleic acid that is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a recombinase protein, and
isolating, from the host cell, a polypeptide encoded by the target nucleic acid.
86. A method of isolating a polypeptide, the method comprising:

obtaining a lysate of a host cell transformed with a target nucleic acid is assembled from a plurality of starting nucleic acids in an assembly reaction comprising a recombinase protein, and

isolating, from the lysate, a polypeptide encoded by the target nucleic acid.

87. A method of obtaining a target nucleic acid, the method comprising:
sending sequence information for a target nucleic acid to a remote site, and
sending a delivery address to the remote site, wherein the target nucleic acid is assembled at the remote site from a plurality of starting nucleic acids in an assembly reaction comprising a recombinase protein, and wherein the target nucleic acid is delivered to the delivery address.
88. A system for assembling a target nucleic acid, the system comprising:
a means for obtaining a plurality of starting nucleic acids, and
a means for assembling a target nucleic acid in an assembly reaction comprising the plurality of starting nucleic acids and a recombinase protein.
89. A system for designing a plurality of starting nucleic acids to be assembled into a target nucleic acid, the system comprising:
a means for obtaining a sequence of a target nucleic acid, and
a means for analyzing the sequence to design a plurality of starting nucleic acids, wherein the plurality of starting nucleic acids are designed to be assembled in a reaction comprising a recombinase protein.
90. The system of claim 88 or 89, wherein the system is automated using computer-implemented means.
91. A business method comprising:
providing a system for assembling a target nucleic acid using an assembly reaction comprising a recombinase protein, and
collaboratively or independently marketing said system.
92. A business method comprising:

providing a system for designing starting nucleic acids to be assembled into a target nucleic acid in an assembly reaction comprising a recombinase protein, and collaboratively or independently marketing said system.

93. The method of any one of claims 80-92, wherein the recombinase protein is a heat-stable recombinase protein.

94. The method of claim 93, wherein the recombinase protein is a RecA protein.

Figure 1

FIG. 1A

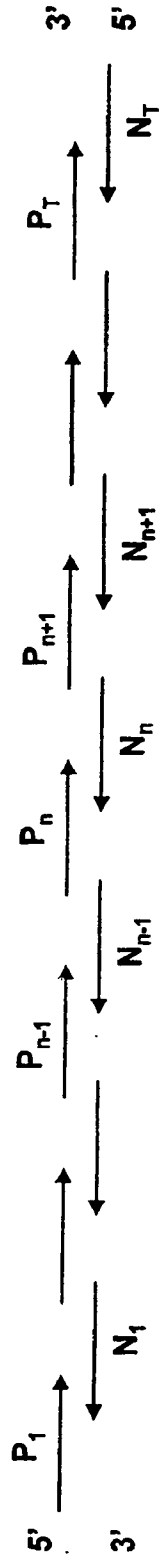


FIG. 1B

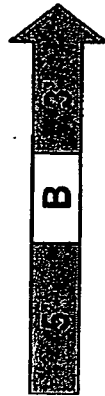


FIG. 1C

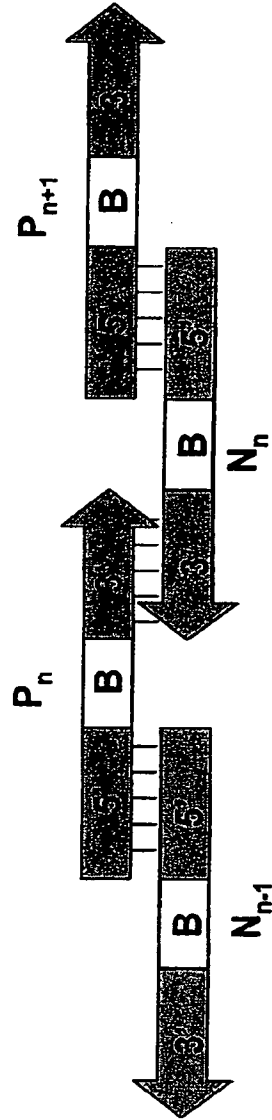


Figure 1 (ctd)

FIG. 1D

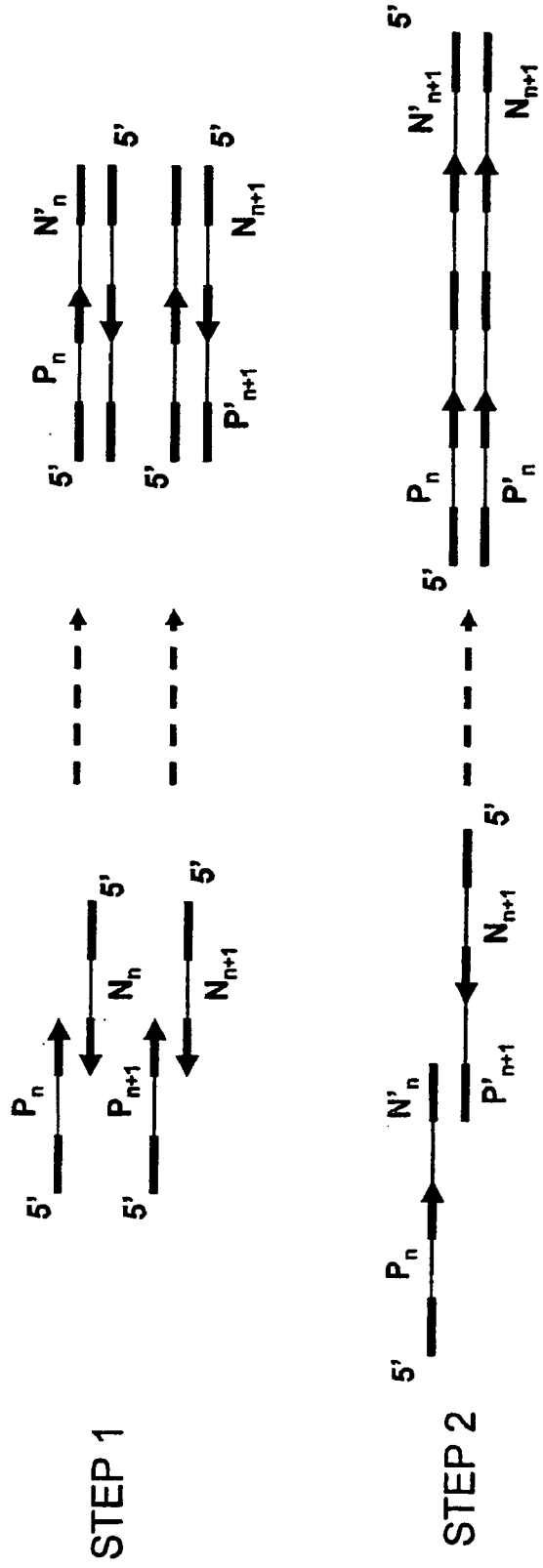


Figure 2

FIG. 2A

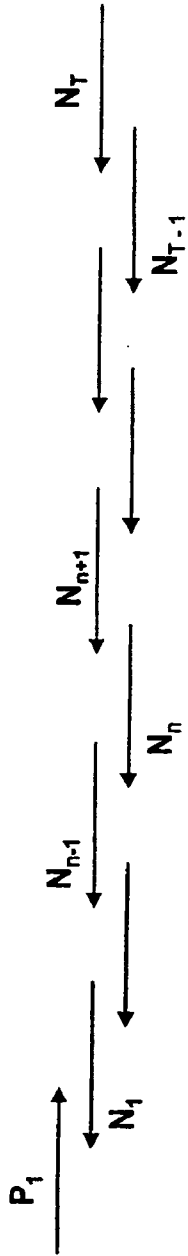


FIG. 2B

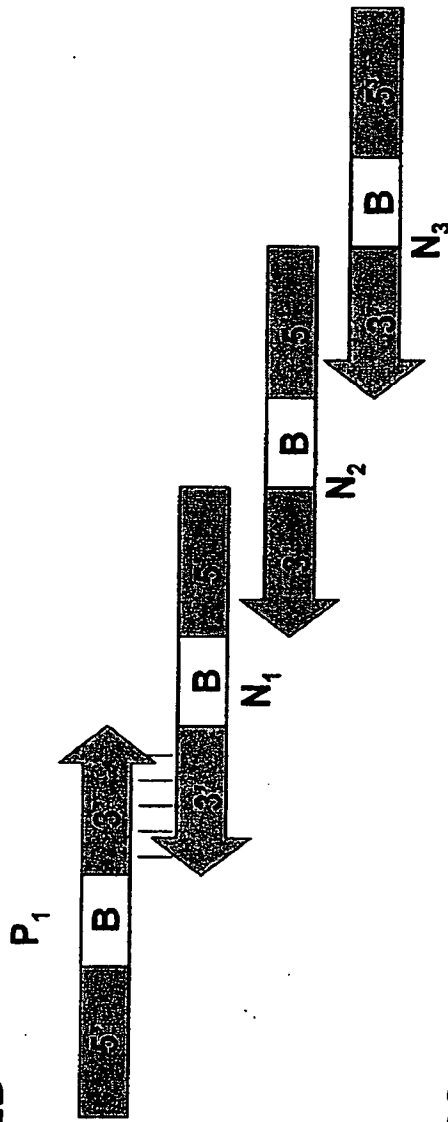


FIG. 2C

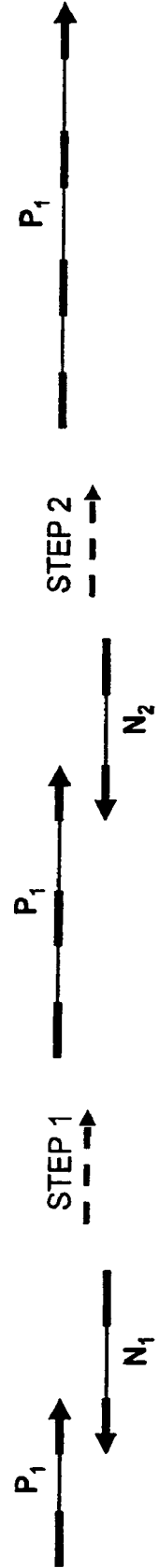


Figure 2 (ctd)

FIG. 2D

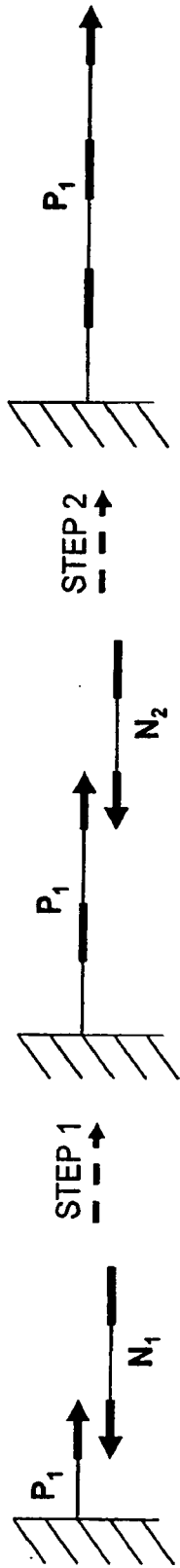


FIG. 2E

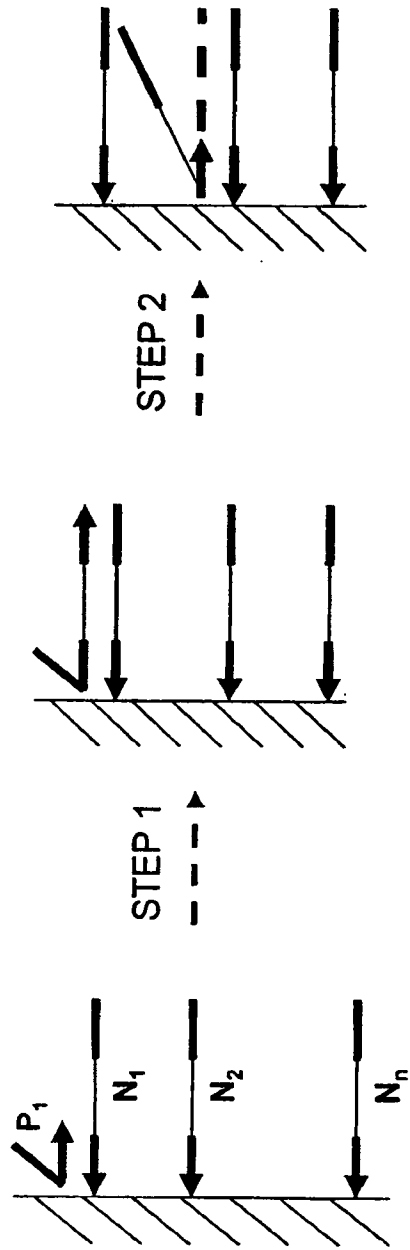


Figure 2 (ctd)

FIG. 2F

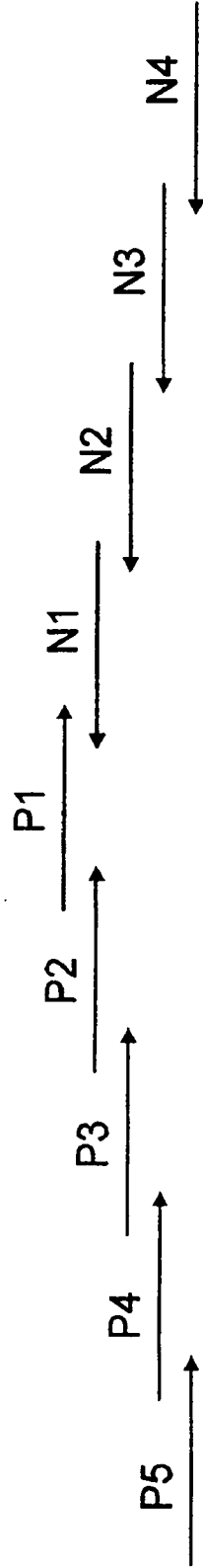


Figure 3

FIG. 3A

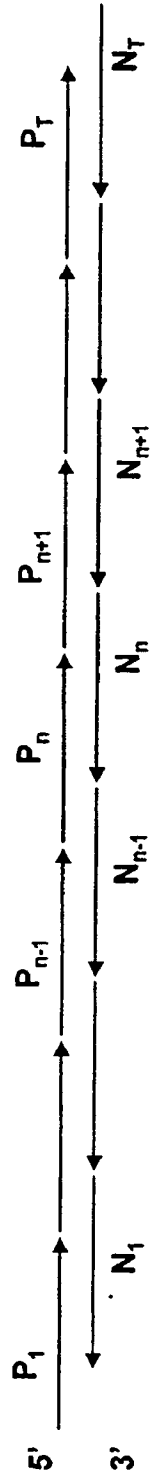


FIG. 3B

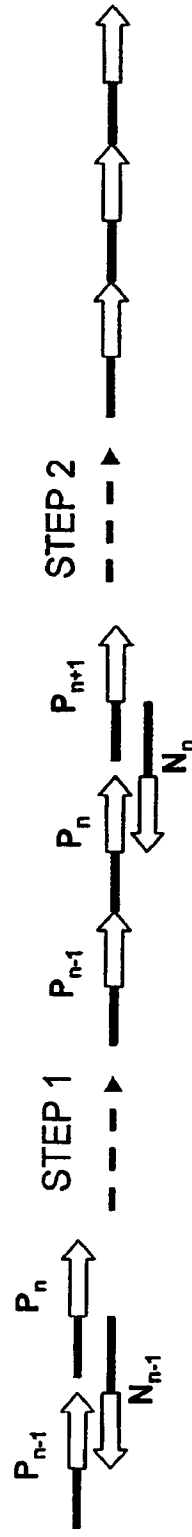


Figure 4

FIG. 4A

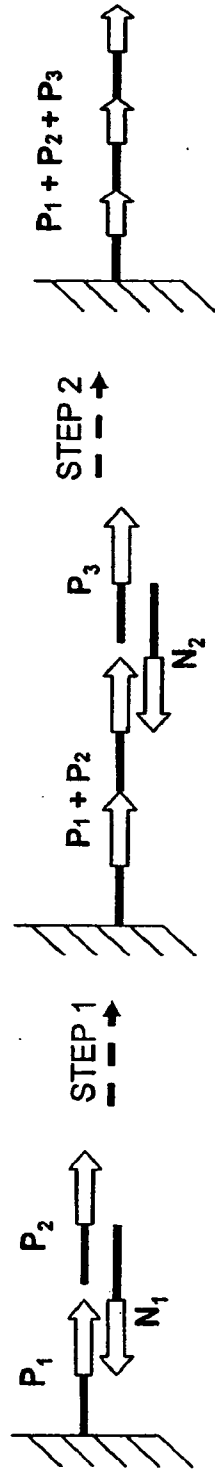
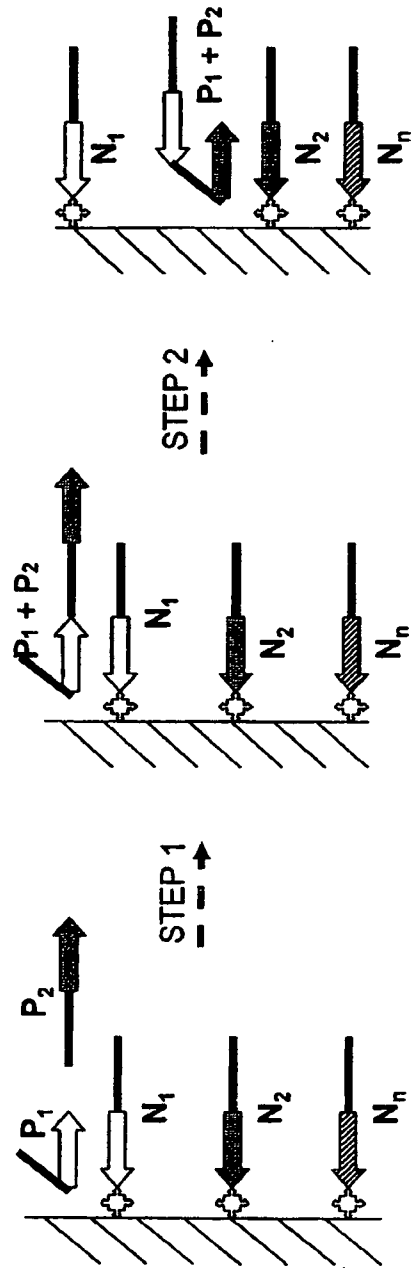


FIG. 4B



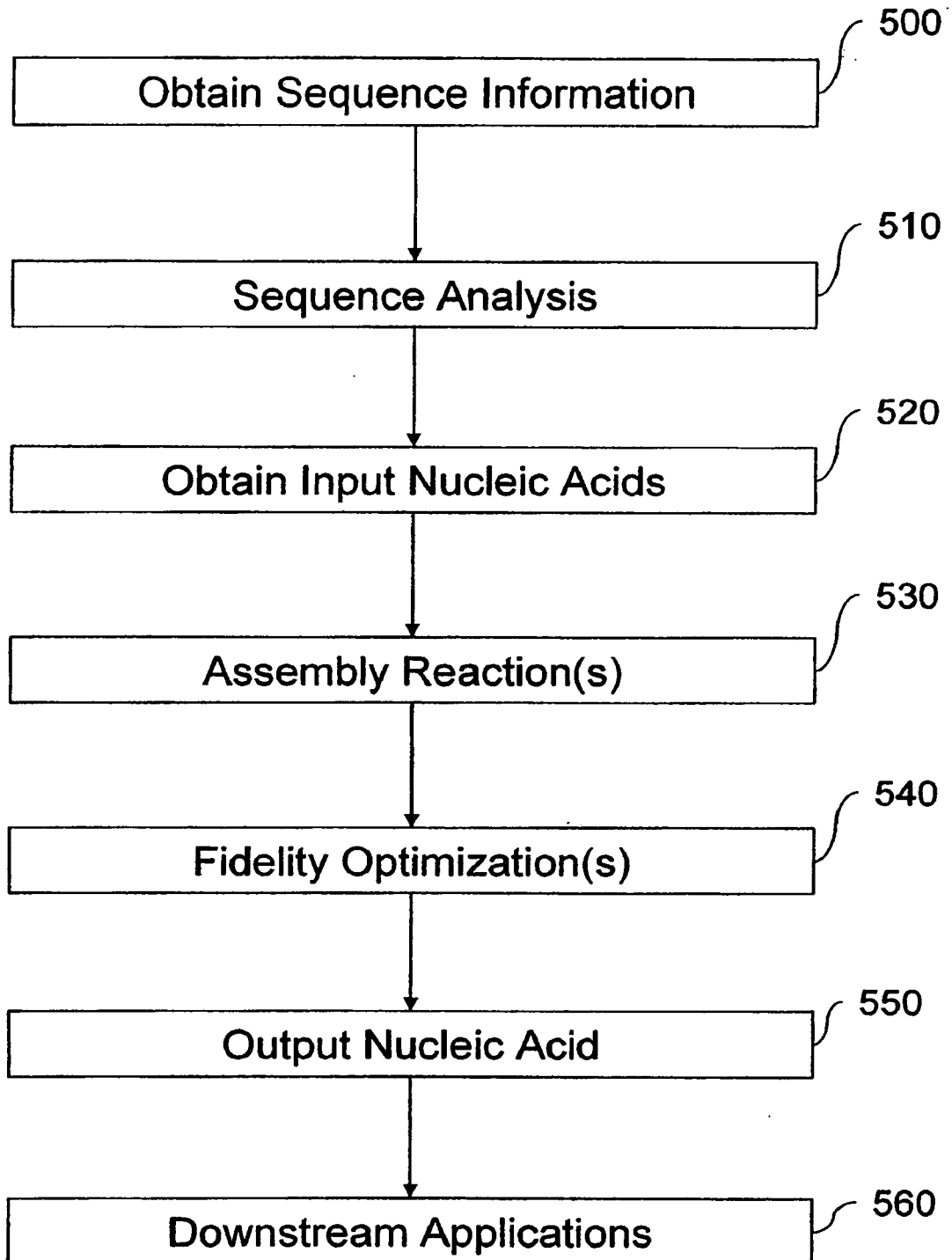


Figure 5