



- (51) **International Patent Classification:**
C12N 15/00 (2006.01) *C12N 15/64* (2006.01)
C12N 15/10 (2006.01)
- (21) **International Application Number:**
PCT/US2016/016636
- (22) **International Filing Date:**
4 February 2016 (04.02.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/112,022 4 February 2015 (04.02.2015) US
- (71) **Applicant: TWIST BIOSCIENCE CORPORATION**
[US/US]; 455 Mission Bay Boulevard South, Suite 545,
San Francisco, California 94158 (US).

- (72) **Inventors: TORO, Esteban;** 38562 McDole Terrace, Fremont, California 94536 (US). **TREUSCH, Sebastian;** 933 Woolsey Street, San Francisco, California 94134 (US). **CHEN, Siyuan;** 1302 Aberdeen Way, San Mateo, California 94402 (US). **WU, Cheng-Hsien;** 1129 Juanita Avenue, Burlingame, California 94010 (US).
- (74) **Agent: HARBURGER, David;** Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, California 94304 (US).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,

[Continued on nextpage]

(54) **Title:** COMPOSITIONS AND METHODS FOR SYNTHETIC GENE ASSEMBLY

(57) **Abstract:** Methods and compositions are provided for assembly of large nucleic acids where the assembled large nucleic acids lack internal sequence modifications made during the assembly process.

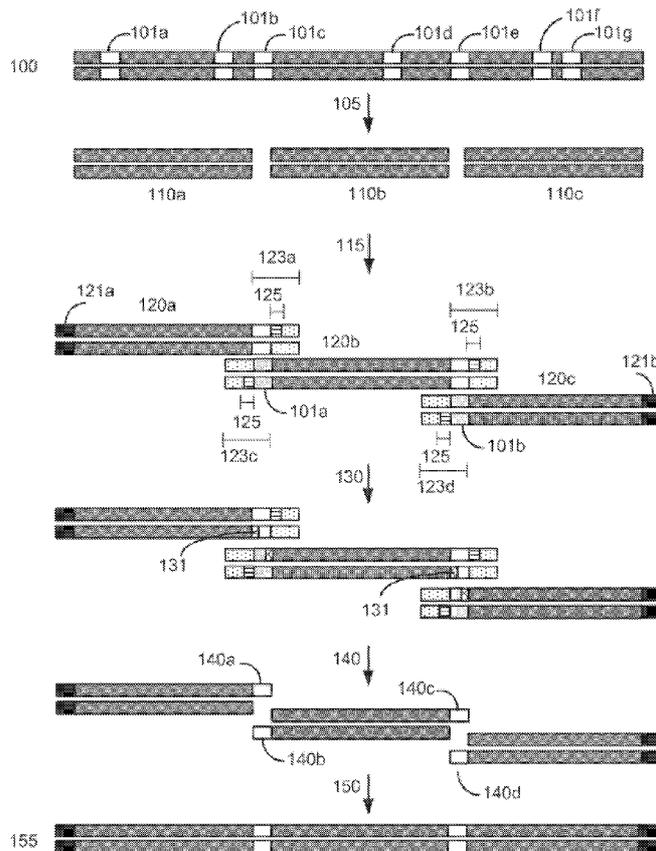


FIG. 1

WO 2016/126987 A1

MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,
SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ,
TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA,
ZM, ZW.

DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT,
LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE,
SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA,
GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,

Published:

- *with international search report (Art. 21(3))*
- *with sequence listing part of description (Rule 5.2(a))*

COMPOSITIONS AND METHODS FOR SYNTHETIC GENE ASSEMBLY

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 62/112,022 filed February 4, 2015, which is herein incorporated by reference in its entirety.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on January 25, 2016, is named 44854_709_601_SL.txt and is 41,005 bytes in size.

BACKGROUND

[0003] De novo nucleic acid synthesis is a powerful tool for basic biological research and biotechnology applications. While various methods are known for the synthesis of relatively short fragments of nucleic acids in a small scale, these techniques suffer from scalability, automation, speed, accuracy, and cost. In many cases, the assembly of nucleic acids from shorter segments is limited by the availability of non-degenerate overhangs that can be annealed to join the segments.

BRIEF SUMMARY

[0004] Provided herein are methods for nucleic acid assembly, comprising: providing a predetermined nucleic acid sequence; providing a plurality of precursor double-stranded nucleic acid fragments, each precursor double-stranded nucleic acid fragment having two strands, wherein each of the two strands comprises a sticky end sequence of 5'-A (N^x)T-3' (SEQ ID NO.: 1) or 5'-G (N^x)C-3' (SEQ ID NO.: 16), wherein N is a nucleotide, wherein x is the number of nucleotides between nucleotides A and T or between G and C, and wherein x is 1 to 10, and wherein no more than two precursor double-stranded nucleic acid fragments comprise the same sticky end sequence; providing primers comprising a nicking endonuclease recognition site and a sequence comprising (i) 5'-A (N^x)U-3' (SEQ ID NO.: 80) corresponding to each of the different sticky end sequences of 5'-A (N^x)T-3' (SEQ ID NO.: 1) or (ii) 5'-G (N^x)U-3' (SEQ ID NO.: 81) corresponding to each of the different sticky end sequences of 5'-G (N^x)C-3' (SEQ ID NO.: 16); and performing a polynucleotide extension reaction to form double-stranded nucleic acid fragments; subjecting the polynucleotide extension reaction product to nicking and cleavage reactions to form double-stranded nucleic acid fragments with 3' overhangs; and annealing the double-stranded nucleic acid fragments to form a nucleic acid encoding for the predetermined nucleic acid sequence that does not include the nicking endonuclease recognition site. Methods

are further provided wherein x is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. Methods are further provided wherein the predetermined nucleic acid sequence is 1 kb to 100 kb in length. Methods are further provided wherein the predetermined nucleic acid sequence is 1 kb to 25 kb in length. Methods are further provided wherein the predetermined nucleic acid sequence is 2 kb to 20 kb in length. Methods are further provided wherein the predetermined nucleic acid sequence is at least 2 kb in length. Methods are further provided wherein the plurality of single-stranded nucleic acid fragments are each at least 100 bases in length. Methods are further provided wherein the double-stranded nucleic acid fragments are each at least 500 bases in length. Methods are further provided wherein the double-stranded nucleic acid fragments are each at least 1 kb in length. Methods are further provided wherein the double-stranded nucleic acid fragments are each at least 20 kb in length. Methods are further provided wherein the sticky ends are at least 4 bases long. Methods are further provided wherein the sticky ends are 6 bases long. Methods are further provided wherein step c further comprises providing (i) a forward primer comprising, in order 5' to 3': a first outer adaptor region and nucleic acid sequence from a first terminal portion of predetermined nucleic acid sequence; and (ii) a reverse primer, comprising, in order 5' to 3': a second outer adaptor region and nucleic acid sequence from a second terminal portion of predetermined nucleic acid sequence. Methods are further provided wherein the annealed double-stranded nucleic acid fragments comprise the first outer adaptor region and the second outer adapter region. Methods are further provided wherein the nicking and cleavage reagents comprise a nicking endonuclease. Methods are further provided wherein the nicking endonuclease comprises endonuclease VIII. Methods are further provided wherein the nicking endonuclease is selected from the list consisting of Nb.BbvCI, Nb.BsmI, Nb.BsrDI, Nb.BtsI, Nt.AlwI, Nt.BbvCI, Nt.BsmAI, Nt.BspQI, Nt.BstNBI, and Nt.CviPII. Methods are further provided wherein the method further comprises ligating the annealed double-stranded nucleic acid fragments. Methods are further provided wherein annealing comprises thermocycling between a maximum and a minimum temperature, thereby generating a first overhang from a first double-stranded DNA fragment and a second overhang from a second double-stranded DNA fragment, wherein the first and the second overhangs are complimentary, hybridizing the first and second overhangs to each other; and ligating. Methods are further provided wherein a polymerase lacking 3' to 5' proofreading activity is added during the polynucleotide extension reaction. Methods are further provided wherein the polymerase is a Family A polymerase. Methods are further provided wherein the polymerase is a Family B high fidelity polymerase engineered to tolerate base pairs comprising uracil. Methods are further provided wherein the precursor double-stranded nucleic acid fragments comprise an adaptor sequence comprising the nicking endonuclease recognition site. Methods are further provided wherein one of the plurality

of precursor double-stranded nucleic acid fragments is a linear vector. In some aspects, provided herein is a nucleic acid library generated by any of the aforementioned methods.

[0005] Methods are provided herein for nucleic acid assembly, comprising: providing a predetermined nucleic acid sequence; synthesizing a plurality of precursor double-stranded nucleic acid fragments, each precursor double-stranded nucleic acid fragment having two strands, wherein each of the two strands comprises a sticky end sequence of 5'-A (N_x) T-3' (SEQ ID NO.: 1) or 5'-G (N_x) C-3' (SEQ ID NO.: 16), wherein N is a nucleotide, wherein x is the number of nucleotides between nucleotides A and T or between G and C, and wherein x is 1 to 10, and wherein no more than two precursor double-stranded nucleic acid fragments comprise the same sticky end sequence; providing primers comprising a nicking endonuclease recognition site and a sequence comprising (i) 5'-A (N_x) M-3' (SEQ ID NO.: 82) corresponding to each of the different sticky end sequences of 5'-A (N_x) T-3' (SEQ ID NO.: 1) or (ii) 5'-G (N_x) M-3' (SEQ ID NO.: 83) corresponding to each of the different sticky end sequences of 5'-G (N_x) C-3' (SEQ ID NO.: 16), wherein M is a non-canonical base, wherein the primers are each 7 to 70 bases in length; and performing a polynucleotide extension reaction to form double-stranded nucleic acid fragments; subjecting the polynucleotide extension reaction product to nicking and cleavage reactions to form double-stranded nucleic acid fragments with 3' overhangs; and annealing the double-stranded nucleic acid fragments to form a nucleic acid encoding for the predetermined nucleic acid sequence that does not include the nicking endonuclease recognition site. Methods are further provided wherein x is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. Methods are further provided wherein x is 4. Methods are further provided wherein the non-canonical base is uracil, inosine, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, acetylcytosine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N-6-isopentenyl adenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 1-methyladenine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, 5-ethylcytosine, N6-adenine, N6-methyladenine, N,N-dimethyladenine, 8-bromoadenine, 7-methylguanine, 8-bromoguanine, 8-chloroguanine, 8-aminoguanine, 8-methylguanine, 8-thioguanine, 5-ethyluracil, 5-propyluracil, 5-methylaminomethyluracil, methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid, pseudouracil, 1-methylpseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-hydroxymethyluracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-S-oxyacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, 5-(2-bromovinyl)uracil, 2-aminopurine, 6-hydroxyaminopurine, 6-thiopurine, or 2,6-diaminopurine. Methods are further

provided wherein the non-canonical base is incorporated into the double-stranded nucleic acid fragments by performing a nucleic acid extension reaction from a primer comprising the non-canonical nucleotide. Methods are further provided wherein the non-canonical base is a uracil. Methods are further provided wherein the uracil is in a deoxyuridine-deoxyadenosine base pair. Methods are further provided wherein the primers are 10 to 30 bases in length. Methods are further provided wherein one of the plurality of precursor double-stranded nucleic acid fragments comprises a portion of linear vector. Methods are further provided wherein no more than 2 N nucleotides of the sticky end sequence have the same identity. Methods are further provided wherein the precursor double-stranded nucleic acid fragments comprise an adaptor sequence comprising the nicking endonuclease recognition site. Methods are further provided wherein the predetermined nucleic acid sequence is 1 kb to 100 kb in length. Methods are further provided wherein the plurality of precursor nucleic acid fragments are each at least 100 bases in length. Methods are further provided wherein the sticky ends are at least 4 bases long in each precursor nucleic acid. In some aspects, provided herein is a nucleic acid library generated by any of the aforementioned methods.

[0006] Methods are provided herein for nucleic acid assembly, comprising: providing a predetermined nucleic acid sequence; synthesizing a plurality of single-stranded nucleic acid fragments, wherein each single-stranded nucleic acid fragment encodes for a portion of the predetermined nucleic acid sequence and comprises at least one sticky end motif, wherein the sticky end motif comprises a sequence of 5'-A(N^x)T-3' (SEQ ID NO.: 1) or 5'-G(N^x)C-3' (SEQ ID NO.: 16) in the predetermined nucleic acid sequence, wherein N is a nucleotide, wherein x is the number of nucleotides between nucleotides A and T or between G and C, and wherein x is 1 to 10, and wherein no more than two single-stranded nucleic acid fragments comprise the same sticky end sequence; amplifying the plurality of single-stranded nucleic acid fragments to generate a plurality of double-stranded nucleic acid fragments, wherein the plurality of double-stranded nucleic acid fragments are modified from the predetermined nucleic acid sequence to comprise (i) a non-canonical base located at a 3' end of the sticky end motif on a first strand and (ii) a first adaptor region located 5' of the non-canonical base on the first strand, wherein the first adaptor region comprises a nicking enzyme recognition site; creating sticky ends, wherein creating sticky ends comprises: treating the plurality of double-stranded fragments with a first nicking enzyme that nicks the non-canonical base on a first strand of each double-stranded fragment, and cleaving the nicked non-canonical base; and treating the plurality of double-stranded fragments with a second nicking enzyme, wherein the second nicking enzyme binds to the first strand at the nicking enzyme recognition site and cleaves a second strand of each double-stranded fragment, wherein a cleavage site for the nicking enzyme is located at a

junction between the sticky end motif a sequence reverse complementary to the first adaptor region of the first strand; and annealing the double-stranded nucleic acid fragments to form a nucleic acid encoding for the predetermined nucleic acid sequence that does not include the nicking endonuclease recognition site. Methods are further provided wherein x is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. Methods are further provided wherein the predetermined nucleic acid sequence is 1 kb to 100 kb in length. Methods are further provided wherein the predetermined nucleic acid sequence is 1 kb to 25 kb in length. Methods are further provided wherein the predetermined nucleic acid sequence is 2 kb to 20 kb in length. Methods are further provided wherein the predetermined nucleic acid sequence is at least 2 kb in length. Methods are further provided wherein the plurality of single-stranded nucleic acid fragments are each at least 100 bases in length. Methods are further provided wherein the plurality of single-stranded nucleic acid fragments are each at least 500 bases in length. Methods are further provided wherein the plurality of single-stranded nucleic acid fragments are each at least 1 kb in length. Methods are further provided wherein the plurality of single-stranded nucleic acid fragments are each at least 20 kb in length. Methods are further provided wherein the sticky ends are at least 4 bases long. Methods are further provided wherein the sticky ends are 6 bases long. Methods are further provided wherein the non-canonical base is uracil, inosine, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, acetylcytosine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N-6-isopentenyl adenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 1-methyladenine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, 5-ethylcytosine, N6-adenine, N6-methyladenine, N,N-dimethyladenine, 8-bromoadenine, 7-methylguanine, 8-bromoguanine, 8-chloroguanine, 8-aminoguanine, 8-methylguanine, 8-thioguanine, 5-ethyluracil, 5-propyluracil, 5-methylaminomethyluracil, methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid, pseudouracil, 1-methylpseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-hydroxymethyluracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-S-oxyacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, 5-(2-bromovinyl)uracil, 2-aminopurine, 6-hydroxyaminopurine, 6-thiopurine, or 2,6-diaminopurine. Methods are further provided wherein the non-canonical base is incorporated into the double-stranded nucleic acid by performing a nucleic acid extension reaction from a primer comprising the non-canonical nucleotide. Methods are further provided wherein the non-canonical base is a uracil. Methods are further provided wherein the uracil is in a deoxyuridine-deoxyadenosine base pair. Methods are further provided

wherein the nicking recognition site is a nicking endonuclease recognition site. Methods are further provided wherein the distance between the non-canonical base the nicking enzyme cleavage site is less than 12 base pairs. Methods are further provided wherein the distance between the non-canonical base the nicking enzyme cleavage site is at least 5 base pairs. Methods are further provided wherein the first nicking enzyme comprises a base excision activity. Methods are further provided wherein the first nicking enzyme comprises uracil-DNA glycosylase (UDG). Methods are further provided wherein the first nicking enzyme comprises an AP endonuclease. Methods are further provided wherein the first nicking enzyme comprises endonuclease VIII. Methods are further provided wherein the second nicking enzyme a nicking endonuclease. Methods are further provided wherein the nicking endonuclease is selected from the list consisting of Nb.BbvCI, Nb.BsmI, Nb.BsrDI, Nb.BtsI, Nt.AlwI, Nt.BbvCI, Nt.BsmAI, Nt.BspQI, Nt.BstNBI, and Nt.CviPIO. Methods are further provided wherein each of the plurality of double-stranded nucleic acid fragments further comprises a two sticky ends. Methods are further provided wherein each of the two sticky ends have a different sequence from each other. Methods are further provided wherein the sticky ends comprises a 3' overhang. Methods are further provided wherein the method further comprises ligating the annealed double-stranded nucleic acid fragments. Methods are further provided wherein annealing comprises: thermocycling between a maximum and a minimum temperature, thereby generating a first overhang from a first double-stranded DNA fragment and a second overhang from a second double-stranded DNA fragment, wherein the first and the second overhangs are complimentary; hybridizing the first and second overhangs to each other; and ligating. Methods are further provided wherein the annealed double-stranded nucleic acid fragments comprise a 5' outer adaptor region and a 3'outer adaptor region. Methods are further provided wherein at least two non-identical single-stranded nucleic acid fragments are synthesized. Methods are further provided wherein at least 5 non-identical single-stranded nucleic acid fragments are synthesized. Methods are further provided wherein at least 20 non-identical single-stranded nucleic acid fragments are synthesized. Methods are further provided wherein a polymerase lacking 3' to 5' proofreading activity is added during the amplification step. Methods are further provided wherein the polymerase is a Family A polymerase. Methods are further provided wherein the polymerase is a Family B high fidelity polymerase engineered to tolerate base pairs comprising uracil. Methods are further provided wherein the amplified plurality of single-stranded nucleic acid fragments are not naturally occurring. Provided herein are nucleic acid libraries generated by any of the aforementioned methods.

[0007] Provided herein are DNA libraries comprising n DNA fragments, each comprising a first strand and a second strand, each of the n DNA fragments comprising, in order 5' to 3': a first

nicking endonuclease recognition site, a first sticky end motif, a template region, a second sticky end motif, and a second nicking endonuclease recognition site, wherein the first sticky end motif comprises a sequence of 5'-A (N^x)_iiU-3' (SEQ ID NO.: 13) in the first strand; and wherein the second sticky end motif comprises a sequence of 5'-A (N^x)_{i,2}U-3' (SEQ ID NO.: 14) in the second strand; wherein N^x denotes x nucleosides, wherein (N^x)_{i,2} is reverse complementary to (N^x)_ii and different from every other N^x found in any sticky end motif sequence within the fragment library, wherein the first nicking endonuclease recognition site in each of the DNA fragments are positioned such that there is a corresponding cleavage site immediately 3' of the sticky end motif in the second strand, and wherein the second nicking endonuclease recognition sites are positioned such that there is a corresponding cleavage site immediately 3' of the second sticky end motif in the first strand. Libraries are further provided wherein the first nicking endonuclease recognition site, the first sticky end motif, the variable insert, the second sticky end motif site, and the second nicking endonuclease recognition site are ordered as recited. Libraries are further provided wherein the library further comprises a starter DNA fragment comprising a template region, a second sticky end motif, and a second nicking endonuclease recognition site; wherein the second sticky end motif comprises a sequence of 5'-A (N^x)_{s,2}U-3' (SEQ ID NO.: 20) and wherein (N^x)_{s,2} is reverse complementary to (N^x)_{i,1}. Libraries are further provided wherein the library further comprises a finishing DNA fragment comprising a first nicking endonuclease recognition site, a first sticky end motif, and a template region; wherein the first sticky end motif comprises a sequence of 5'-A (N^x)_fiU-3' (SEQ ID NO.: 21) and wherein (N^x)_fi is reverse complementary to (N^x)_{n,2}. Libraries are further provided wherein the first and second nicking endonuclease recognition sites are the same. Libraries are further provided wherein n is at least 2. Libraries are further provided wherein n is less than 10. Libraries are further provided wherein x is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. Libraries are further provided wherein x is 4. Libraries are further provided wherein the template region of each of the n DNA fragments encodes for a different nucleic acid sequencing from the template region of every other of the n DNA fragments. Libraries are further provided wherein the sequences of the n DNA fragments are not naturally occurring. Libraries are further provided wherein the first nicking endonuclease recognition site is not naturally adjacent to the first sticky end motif.

INCORPORATION BY REFERENCE

[0008] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] **Figure 1** depicts a workflow through which a nucleic acid product is assembled from 1kbp nucleic acid fragments.

[0010] **Figure 2** depicts the assembly of a longer nucleic acid fragment from the ligation of two oligonucleic acid fragments having complementary overhangs and discloses SEQ ID NOs.: 4, 6, 3, 5, 3, 6, 3 and 6, respectively, in order of appearance.

[0011] **Figure 3** depicts a uracil-containing universal primer pair, and discloses SEQ ID NOs.: 7, 2, 8 and 2, respectively, in order of appearance.

[0012] **Figure 4** depicts the assembly of a nucleic acid product from oligonucleic acid fragments having complementary overhangs.

[0013] **Figures 5A-5B** depict the assembly of a recombinatorial library from a library of nucleic acid fragments each having at least one unspecified base.

[0014] **Figure 6** depicts a diagram of steps demonstrating a process workflow for oligonucleic acid synthesis and assembly.

[0015] **Figure 7** illustrates an example of a computer system.

[0016] **Figure 8** is a block diagram illustrating an example architecture of a computer system.

[0017] **Figure 9** is a diagram demonstrating a network configured to incorporate a plurality of computer systems, a plurality of cell phones and personal data assistants, and Network Attached Storage (NAS).

[0018] **Figure 10** is a block diagram of a multiprocessor computer system using a shared virtual address memory space.

[0019] **Figure 11** shows an image of an electrophoresis gel resolving amplicons of a LacZ gene assembled in a plasmid using scar-free assembly methods described herein.

DETAILED DESCRIPTION

[0020] Disclosed herein are methods and compositions for the assembly of nucleic acid fragments into longer nucleic acid molecules of desired predetermined sequence and length without leaving inserted nucleic acid sequence at assembly points, aka "scar" sequence. In addition, amplification steps are provided during the synthesis of the fragments which provide a means for increasing the mass of a long nucleic acid sequence to be amplified by amplifying the shorter fragments and then rejoining them in a processive manner such that the long nucleic acid is assembled.

[0021] Definitions

[0022] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of any embodiment. As used herein, the singular forms "a,"

"an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms "comprises" and/or "comprising," when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. As used herein, the term "and/or" includes any and all combinations of one or more of the associated listed items. Unless specifically stated or obvious from context, as used herein, the term "about" in reference to a number or range of numbers is understood to mean the stated number and numbers +/- 10% thereof, or 10% below the lower listed limit and 10% above the higher listed limit for the values listed for a range. As used herein, the terms "preselected sequence", "predefined sequence" or "predetermined sequence" are used interchangeably. The terms mean that the sequence of the polymer is known and chosen before synthesis or assembly of the polymer. In particular, various aspects of the invention are described herein primarily with regard to the preparation of nucleic acids molecules, the sequence of the oligonucleotide or polynucleotide being known and chosen before the synthesis or assembly of the nucleic acid molecules.

[0023] The term "nucleic acid" as used herein refers broadly to any type of coding or non-coding, long polynucleotide or polynucleotide analog. As used herein, the term "complementary" refers to the capacity for precise pairing between two nucleotides. If a nucleotide at a given position of a nucleic acid is capable of hydrogen bonding with a nucleotide of another nucleic acid, then the two nucleic acids are considered to be complementary to one another (or, more specifically in some usage, "reverse complementary") at that position. Complementarity between two single-stranded nucleic acid molecules may be "partial," in which only some of the nucleotides bind, or it may be complete when total complementarity exists between the single-stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

[0024] "Hybridization" and "annealing" refer to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The term "hybridized" as applied to a polynucleotide is a polynucleotide in a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogsteen binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more

extensive process, such as the initiation of a PCR or other amplification reactions, or the enzymatic cleavage of a polynucleotide by a ribozyme. A first sequence that can be stabilized via hydrogen bonding with the bases of the nucleotide residues of a second sequence is said to be "hybridizable" to the second sequence. In such a case, the second sequence can also be said to be hybridizable to the first sequence. In many cases a sequence hybridized with a given sequence is the "complement" of the given sequence.

[0025] In general, a "target nucleic acid" is a desired molecule of predetermined sequence to be synthesized, and any fragment thereof.

[0026] The term "primer" refers to an oligonucleotide that is capable of hybridizing (also termed "annealing") with a nucleic acid and serving as an initiation site for nucleotide (RNA or DNA) polymerization under appropriate conditions (i.e. in the presence of four different nucleoside triphosphates and an agent for polymerization, such as DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer. In some instances, primers are least 7 nucleotides long. In some instances, primers range from 7 to 70 nucleotides, 10 to 30 nucleotides, or from 15 to 30 nucleotides in length. In some instances, primers are from 30 to 50 or 40 to 70 nucleotides long. Oligonucleotides of various lengths as further described herein are used as primers or precursor fragments for amplification and/or gene assembly reactions. In this context, "primer length" refers to the portion of an oligonucleotide or nucleic acid that hybridizes to a complementary "target" sequence and primes nucleotide synthesis. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. The term "primer site" or "primer binding site" refers to the segment of the target nucleic acid to which a primer hybridizes.

[0027] Scar-free Nucleic Acid Assembly

[0028] An exemplary workflow illustrating the generation of a target nucleic acid using a scar-free nucleic acid assembly method is shown in **FIG. 1**. In a first step, the predetermined sequence of a double-stranded target nucleic acid **100** is analyzed to find short sequences, such as sequences of 3, 4, 5, 6, 7, 8, 9, or 10 bases, to serve as sticky end motifs **IOla-IOlg**. Each sticky end motif **101a-IOlg** identified in the target nucleic acid need not comprise a sequence unique from another sequence in the target nucleic acid, but each sticky end sequence involved in target nucleic acid assembly is used only once, that is, at only one pair of precursor nucleic acid fragment ends. Sticky end motifs are generally used more than once, that is, at more than one pair of precursor nucleic acid fragment ends. A sticky end motif comprises the sequence $A(N^x)T$ (SEQ ID NO. : 1), wherein x indicates from about 1 to about 10, N deoxyribonucleic acid bases of

any sequence. For example, x is 4, 5 or 6 and each N may be the same or different from another N in the motif. In some cases, a sticky end motif comprises an ANNNNT (SEQ ID NO.: 2) sequence. After the target nucleic acid sequence **100** is analyzed to identify sticky end motifs **101a-101g** and fragment sequences **110a-110c** selected **105**, the fragments are synthesized **115** with the sticky end motifs from the target nucleic acid **100**, for example, by de novo synthesis.

[0029] In one example of the de novo synthesis process as illustrated in **FIG. 1**, synthesis **115** results in double-stranded precursor nucleic acid fragments **120a-120c**. Each double-stranded precursor nucleic acid fragments **120a-120c** includes an adaptor sequence positioned at either end of target fragment sequence. The outer terminal portions of the double-stranded precursor nucleic acid fragments each comprise an outer adaptor **121a-121b**. Each double-stranded precursor nucleic acid fragment **121a-120c** is synthesized **115** such that it overlaps with another region of another fragment sequence via sticky end motifs **101a-101g** in a processed order. As illustrated in **FIG. 1**, at the region of the synthesized double-stranded precursor nucleic acid fragment comprising a sticky end motif **101a-101b**, synthesis also results in including additional sequence in a connecting adaptor region **123a-123d**. The "sticky end motif" occurs at a desired frequency in the nucleic acid sequence. The connecting adaptor region **123a-123d** includes a sticky end motif **101a-101b** and a first nicking enzyme recognition site **125**.

[0030] Further processing of the double-stranded precursor nucleic acid fragments **120a-120c** is done via primers in an amplification reaction via primers in an amplification reaction **130** to insert a non-canonical base **131**. In an alternative method, connecting adaptor regions **123a-123d** and/or outer adaptors **120a-120b** are and/or are appended to either end of the fragments during a processing step, for example, via primers in an amplification reaction **130**.

[0031] To generate fragments capable of annealing, the double-stranded precursor nucleic acid fragments **120a-120c** as subjected to enzymatic processing **140**. Enzymatic processing **140** as illustrated in **FIG. 1**, entails cleaving portions of the connecting adaptor regions **123a-123d**. In a first enzymatic reaction, a first nicking enzyme binds at a first nicking enzyme recognition site **125**, and then cleaves the opposite strand. In a second enzymatic reaction, a second nicking enzyme cleaves the non-canonical base **131**. The enzymatic reaction results in fragments having sticky ends **140a-140d** wherein pairs of sticky ends are revers complementary and correspond to sticky end motifs **101a-101b** in the original sequence. Finally, the fragments are subjected to an annealing and an ligation reaction **150** to form a reaction product **155** comprising target sequence. The annealing and ligation reactions **150** can include rounds of annealing, ligating and melting under conditions such that only desired sticky ends **140a-140d** are able to anneal and ligate, while cleaved end fragments remain unligated. Ordered assembly of nucleic acid

fragments includes linear and circular assembly, for example, fragments are assembled with a vector into a plasmid.

[0032] In one example, each double-stranded fragment is flanked on a terminal side by a double-stranded connecting adaptor comprising: a double-stranded sticky end motif derived from the target nucleic acid sequence, a nicking enzyme cleavage site located only a first strand of the adaptor, and a double-stranded nicking enzyme recognition sequence, such that upon incubation with a first nicking enzyme specific for the nicking enzyme recognition sequence, a single-strand break is introduced at the nicking enzyme cleavage site in the first strand. In exemplary cases, the sticky end motif of the connecting adaptor is located directly at the 5' or 3' end of a fragment so that each sticky end motif-fragment or fragment-sticky end motif construct comprises sequence native to the predetermined target nucleic acid sequence. The target nucleic acid sequences **100** may be partitioned in sticky end motifs **IOla-IOlg** of about 200 bp or other lengths, such as less than or about 50 bp, about 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, or 1000 bp, or more bp.

[0033] In various aspects, described herein are double-stranded nucleic acids comprising a first strand having a first cleavage site and a second strand having a second cleavage site; wherein the cleavage sites are positioned one or more bases from one another in sequence. As a non-limiting example, provided are double-stranded nucleic acids comprising a first strand comprising a non-canonical base and a second strand comprising a nicking enzyme cleavage site; wherein the non-canonical base and nicking enzyme cleavage site are positioned one or more bases from one another in sequence. Through the combined action of nicking enzymes directed to act in tandem at adjacent or near adjacent positions on opposite strands of a double-stranded nucleic acid, one may impact the generation of a sticky end at or near the end of a first nucleic acid fragment, wherein the sticky end sequence is unique and complementary only to the sticky end of a second nucleic acid fragment sequentially adjacent thereto in a predetermined sequence of a full-length target nucleic acid to be assembled from the fragments.

[0034] An example workflow illustrating the generation of a nick at a non-canonical base in a nucleic acid is shown in **FIGS. 2A-2B**. As a preliminary step, as illustrated in **FIG. 1**, a predetermined sequence of a target nucleic acid is partitioned *in silico* into fragments, where the sequence of each fragment is separated from an adjacent fragment by an identified sticky end motif. The connecting adaptor regions **123a-123d** appended to an end of a fragment include a sticky end motif corresponding to the sticky end motif **IOla-IOlg** adjacent to the fragment such that each motif can processively be aligned during enzymatic processing. For example, the 3' end of a first fragment **201** is configured for connection to the 5' end of fragment 2 **202** via a sticky end motif X **211a**. Similarly, fragment 2 **201** is configured for connection to fragment 3

203 in the target sequence via sticky end motif **Y 211d** and fragment **3 203** is configured for connection to fragment **4 204** in the target sequence via sticky end motif **Z 211c**.

[0035] In some instances, a connecting adaptor comprises a first and a second nicking enzyme recognition site such that tandem nicks made to the connecting adaptor do not affect the sequence of the fragment to which the adaptor is connected. For example, a detailed view of precursor fragments **203** and **204** having such connecting adaptors is shown in **FIG. 2 (220 and 215, respectively)**. The 5' connecting adaptor of the fragment **4 204** comprises a first double-stranded nicking enzyme recognition site **225**, a first nicking enzyme cleavage site **227** located on a first single-strand **221**, and a double-stranded sticky end motif **Z (AAGTCT, SEQ ID NO.: 3)** modified with a uracil (**AAGTCU, SEQ ID NO.: 4**) on a second single-strand **223**. The 3' connecting adaptor of fragment **3 230** comprises the double-stranded sticky end motif **Z 211c (SEQ ID NO.: 3)** modified with a uracil (**AGACTU, SEQ ID NO.: 5**) on a first single-strand **229**, the first nicking enzyme cleavage site **227** on a second single-strand **231**, and the first double-stranded nicking enzyme recognition site **225**. Accordingly, each strand of the connecting adaptors comprise two nicking sites —a first nicking enzyme cleavage site and a uracil—located at different positions and strands in the adaptor sequence.

[0036] Continuing this exemplary workflow, nicking reactions **240** are next described. The first nicking enzyme cleavage site **227** is located at the backbone of a single-strand of each connecting adaptor, adjacent to a first nicking enzyme recognition sequence **225**. In some instances, the cleavage site is located at a position adjacent to a 5' or 3' end of a nicking enzyme recognition site by 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 bases. Fragments are treated with a first nicking enzyme, in this case, a strand-adjacent nicking enzyme, which cleaves a single-strand of the connecting adaptor at the first nicking enzyme cleavage site; and a second nicking enzyme which excises uracil and cleaves a single-strand of the connecting adaptor at the excised uracil site. Cleaved fragments **241, 242** comprise sticky end overhangs. Fragments comprising complementary sticky end overhangs are annealed and ligated **250**. The ligation product **260** comprises predetermined target nucleic acid sequence comprising adjacent fragments separated by a sticky end motif, without the introduction of extraneous scar sequence.

[0037] As used herein, a sticky end motif includes forward, reverse, and reverse complements of a sticky end sequence. For example, a first strand of sticky end motif **Z** comprises **SEQ ID NO.: 3** and a second strand of sticky end motif **Z** comprises the reverse complement of **SEQ ID NO.: 3**, **AGACTT (SEQ ID NO.: 6), FIG. 2**.

[0038] To prepare double-stranded precursor fragments with one or two nicking enzyme cleavage sites, precursor fragments are either synthesized with one or both sites, assembled from smaller nucleic acids comprising one or both sites, amplified with a primer comprising one or

both sites, or any combination of the methods described or known in the art. For example, a precursor fragment can comprise a sticky end sequence and a primer is synthesized comprising a sequence that is complementary to the sticky end sequence, yet comprises a non-canonical base substitution at the 3' end of the sticky end sequence. Amplification of precursor nucleic acid fragments comprising sticky end sequences with the primer may introduce the non-canonical base to the precursor fragment sequence so that the precursor fragment amplicons comprise a nicking enzyme cleavage site defined by the position of the non-canonical base. In one example, a double-stranded precursor fragment is prepared comprising, in 5' to 3' or 3' to 5' order: a first double-stranded nicking enzyme recognition sequence, a first nicking enzyme cleavage site on a first single-strand, a double-stranded sticky end motif, and a double-stranded fragment of predetermined target sequence; wherein amplification of the precursor fragment with a non-canonical base-containing primer as described introduces a second nicking enzyme cleavage site between the sticky end motif and fragment of predetermined target sequence on a second single-strand.

[0039] In some cases, a collection of precursor nucleic acid fragments is provided, each precursor nucleic acid fragment comprising a fragment sequence of a predetermined sequence of a target nucleic acid and a 5' and/or 3' connecting adaptor, wherein each connecting adaptor comprises a shared sequence among the precursor fragments and optionally one or more bases variable among the precursor fragments. Amplification of collective fragments comprising a shared sequence can be performed using a universal primer targeting shared sequence of the adaptors.

[0040] An exemplary universal primer is one that comprises a base or sequence of bases which differs from a shared adaptor sequence of precursor nucleic acid fragments. For example, a universal primer comprises a non-canonical base as an addition and/or base substitution to shared adaptor sequence, and amplification of precursor fragments comprising the shared adaptor sequence with the primer introduces the non-canonical base into each adaptor sequence. An illustration of an exemplary universal primer pair comprising a non-canonical base substitution is shown in **FIG. 3**. Each primer comprises, in 5' to 3' order: one or more adaptor bases **301a**, **301b**, a nicking enzyme recognition site **302a**, **302b**, and a sticky end motif comprising a T to U base substitution (sticky end motif in forward primer **305**: AATGCU, SEQ ID NO.: 7 **303a**; sticky end motif in reverse primer **310**: AGCATU, SEQ ID NO.: 8 **303b**). Amplification of a first precursor nucleic acid having an adaptor comprising sticky end motif AATGCT (SEQ ID NO.: 9) with the forward primer introduces a uracil to a single-strand of the adaptors in the resulting amplicons. Amplification of a second precursor nucleic acid having an adaptor comprising sticky end motif AGCATT (SEQ ID NO.: 10) with the reverse primer introduces a

uracil to a single-strand of the adaptors in the resulting amplicons. The amplification products, cleavage steps described herein, have compatible sticky ends are suitable for annealing and ligating. In some cases, a set of two or more universal primer pairs is used in a method disclosed herein, wherein each pair comprises a universal forward primer and a universal reverse primer, and wherein the forward primers in the set each comprise a shared forward sequence and a variable forward sequence and the reverse primers in the set each comprise a shared reverse sequence and a variable reverse sequence. A set of universal primers designed to amplify the collection of nucleic acids may comprises differences within each set of universal forward and reverse primers relating to one or more bases of the sticky end motif sequence.

[0041] Provided herein are methods where a universal primer pair incorporates a universal primer sequence 5' to a sticky end motif sequence in a nucleic acid. As a non-limiting example, a universal primer sequence comprises a universal nicking enzyme recognition sequence to be incorporated at the end of each fragment in a library of precursor nucleic acid fragments. For the universal primers shown in **FIG. 3**, as one example, a primer fusion site comprises four bases 3' to an adenine (A) and 5' to a uracil (U). The 5' - A (N⁴) U -3' (SEQ ID NO.: 11) primer fusion sequence is located at the very 3' end of the exemplary primers, which conclude with a 3' uracil. Alternatively, the primer fusion can be sequence is 5' - G (N⁴) U -3' (SEQ ID NO.: 12). For some assembly reactions with precursor nucleic acid fragments, a number of such primers with varying N⁴ sequences are used within a reaction mixture, each targeting a complementary fusion site on one end of one of the fragments that are to be assembled. N⁴ represents any configuration of 4 bases (N), where each base N has the same or different identity than another base N. In some cases, the number of N bases is greater than or less than 4. Without being bound by theory, since mismatched base pairs toward the 3' end of a primer significantly reduce the efficiency of a nucleic acid extension reaction, placement of variable regions that target different fusion sites increases the specificity between the primer fusion site sequences and fragment fusion site sequences.

[0042] A plurality of precursor nucleic acid fragments comprising shared and variable regions of sequence is shown in **FIGS. 4**. Each precursor fragment 401-404 comprises at least one connecting adaptor and optionally an outer adaptor at each end of a target fragment sequence, wherein each of the connecting and outer adaptors comprise a shared sequence. Following PCR amplification 405 with primers (designate by arrows in above and below "401" in **FIG. 4**), the precursor fragment 401-404 are modified to include non-canonical bases 410, subject to enzymatic digestion 415 to generate fragments with overhangs 420, and subject to annealing and ligation 430. The primers may be universal primers described herein. The nucleic acids comprising fragment 1 401 and fragment 2 402 are appended at their 3' or 5' ends, respectively,

with sticky end motif X, wherein the sequence: fragment 1-sticky end motif X-fragment 2 occurs in the predetermined target sequence. The nucleic acids comprising fragment 2 **402** and fragment 3 **403** are appended at their 3' or 5' ends, respectively, with sticky end motif Y, wherein the sequence fragment 2-sticky end motif Y-fragment 3 occurs in the predetermined target sequence. The nucleic acids comprising fragment 3 **403** and fragment 4 **404** are appended at their 3' or 5' ends, respectively, with sticky end motif Z, wherein the sequence fragment 3-sticky end motif Z-fragment 4 occurs in the predetermined target sequence. The ligation product is then amplified by PCR **440** using primers **445**, **446** complementary to outer adaptor regions. The resulting final product is a plurality of nucleic acids which lack adaptor regions **450**.

[0043] Connecting adaptors disclosed herein may comprise a Type II restriction endonuclease recognition sequence. In such instances, a sticky end motif shared between adjacent fragments in a predetermined sequence is a Type II restriction endonuclease recognition sequence. As a non-limiting example, sticky end motif X is a first Type II restriction endonuclease recognition sequence so that upon digesting with the appropriate Type II restriction enzyme, a sticky end is produced at the ends of nucleic acids **401** and **402**. As another example, sticky end motifs Y and Z are also two different Type II restriction endonuclease recognition sequences native to the predetermined target nucleic acid sequence. In such cases a target nucleic acid having no scar sites is assembled from the Type II-digested fragments. In some cases, fragments assembled using Type II restriction endonucleases are small, for example, less than about 500, 200, or 100 bases so to reduce the possibility of cleavage at a site within the fragment sequence. In some instances, a combination of tandem, single-strand breaks and Type II restriction endonuclease cleavage is used to prepare precursor fragments for assembly.

[0044] In some cases, tandem nicking of a double-stranded nucleic acid and/or double-stranded cleavage by a Type II restriction endonuclease, results in undesired sequences terminal to cleavage sites remaining in the cleavage reaction. These terminal bases are optionally removed to facilitate downstream ligation. Cleaved termini are removed, for example, through size-exclusion column purification. Alternately or in combination, terminal ends are tagged with an affinity tag such as biotin such that the cleaved ends are removed from the reaction using avidin or streptavidin, such as streptavidin coated on beads. Alternately, for tandem nicking reactions, cleaved ends of precursor fragments are retained throughout annealing of the fragments to a larger target nucleic acid.

[0045] Provided herein are methods where the precursor fragments comprise a first nicking enzyme cleavage site defined by a first nicking enzyme recognition sequence, and a non-canonical base. In these cases, precursor fragments are treated with a first enzyme activity that excises the non-canonical base and a second enzyme activity that cleaves single-stranded nucleic

acids at the abasic site and first nicking enzyme cleavage site. Some of the cleaved ends produced at the first nicking enzyme cleavage site are able to reanneal to cleaved sticky end overhangs, and may re-ligate. However, such re-ligation will also reconstitute the cleavage site, and will be re-cleaved if the single-strand nicking enzyme activity is included in the reaction. The opposite strand, from which the non-canonical base has been excised and the phosphodiester backbone cleaved at that site, is incapable of re-ligation to the cleaved end because of the gap created at the now abasic site. However, sticky ends of precursor nucleic acid fragments that are end pairs intended to assemble into a larger fragment are capable of annealing to one another and ligating. Upon ligation, the molecule formed thereby will not have the first nicking enzyme cleavage site, as the sequence that specifies cleavage is in the cleaved-off terminal fragment rather than in the adjacent fragment sequence. Subsequently, ligated ends will not be re-cleaved by strand-adjacent nicking enzyme. Additionally, as neither strand has a gap position corresponding to the excised non-canonical base position, sticky ends of precursor nucleic acid fragments that are end pairs intended to assemble into a larger target are capable of annealing to one another across both strands.

[0046] Following successive rounds of thermocycling through annealing, ligation and denaturing, optionally in the presence of a nicking enzyme, sticky ends that bind to their partner ends will be ligated and drawn out of the sticky end pool, while sticky ends that bind to cleaved terminator sequence will remain available for ligation in successive rounds. Through successive iterations of annealing, ligation and melting, cleaved ends remain unligated while junction binding events become ligated to one another.

[0047] Sticky ends of cleaved precursor nucleic acid fragments are allowed to anneal to one another under conditions promoting stringent hybridization, such that in some cases, only perfectly reverse complementary sticky ends anneal. In some cases, less stringent annealing is permitted. Annealed sticky ends are ligated to form either complete target nucleic acid molecules, or larger fragment target nucleic acid molecules. Larger fragment molecules are in turn subjected to one or more additional rounds of assembly, using either methods described herein and additional sticky end sites, or one or more assembly techniques known in the art.

[0048] Methods and compositions described herein allow assembly of large nucleic acid target molecules with a high degree of confidence as to sequence integrity. The target molecules are assembled from precursor nucleic acid fragments that are in many cases synthesized to a length that is within a target level of sequence confidence—that is, they are synthesized to a length for which the synthesis method provides a high degree of confidence in sequence integrity. In some cases, this length is about 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleic acid bases.

[0049] In some cases, the methods provided herein generate a specific target sequence for a recombinatorial library, e.g., a chimeric construct or a construct comprising at least one targeted variation for codon mutation. Positions to vary include, without limitation, codons at residues of interest in an encoded protein, codons of residues of unknown function in an encoded protein, and pairs or larger combinations of codons encoding residues known or suspected to work in concert to influence a characteristic of a protein such as enzymatic activity, thermostability, protein folding, antigenicity, protein-protein interactions, solubility or other characteristics.

[0050] A library of variants may be prepared by synthesizing target nucleic acids from fragments having at least one indeterminate or partially determinate position among members of the library. In some cases, target fragments are synthesized having combinations of variants. Upon assembly of a target nucleic acid library, multiple combinations of variations at a first position and variations at a second position may be present in the library. In some instances, all possible combinations of variants are represented in a library. The library may be constructed such that variant base positions are each found on different target fragments, or alternately, multiple variant base positions are found on the same target fragment library.

[0051] **FIGS. 5A-5B** illustrate an exemplary workflow for recombinatorial library synthesis of a target gene. The target gene is partitioned into fragments 1-4 by motifs X, Y, and Z **500**, each fragment comprising one or two indeterminate sites (**FIG. 5A**). In some instances, not all fragments of a target gene comprise an indeterminate site. Precursor fragments **501** comprise an outer adaptor, a variant of fragment 1 comprising one indeterminate site, and a connecting adaptor comprising motif X. Precursor fragments **502** comprise a connecting adaptor comprising motif X, a variant of fragment 2 comprising one indeterminate site, and a connecting adaptor comprising motif Y. Precursor fragments **503** comprise a connecting adaptor comprising motif Y, a variant of fragment 3 comprising two indeterminate sites, and a connecting adaptor comprising motif Z. Precursor fragments **504** comprise a connecting adaptor comprising motif Z, a variant of fragment 4 comprising one indeterminate sites, and a second outer adaptor. PCR is used to generate amplicons **510** of each precursor fragment, collectively, **500**. In some cases, using a universal primer pair(s) (**FIG. 5B**). Precursor nucleic acids are digested at their connecting adaptor sequence to generate sticky ends, complements of which are annealed and ligated together to form a series of target genes comprising: fragment 1 sequence comprising one indeterminate site, motif X, fragment 2 sequence comprising one indeterminate site, motif Y, fragment 3 sequence comprising two indeterminate sites, motif Z, and fragment 4 sequence comprising one indeterminate site **520**. The number of possible target gene variants is 4^5 or 1,024 different genes. **FIG. 5B**, part **530**, shows a conceptual depiction of some of these target gene variants after PCR amplification.

[0052] Methods described herein comprise assembling double-stranded DNA ("dsDNA") target nucleic acid from shorter target nucleic acid fragments that are building block precursors. Assembly may proceed by hybridizing uniquely complimentary pairs of overhangs. Such uniquely complimentary pairs may be formed by incorporating sticky ends from two precursor fragments that appear successively in the assembled nucleic acid. In some cases, the pair of overhangs does not involve complete complementarity, but rather sufficient partial complementarity that allows for selective hybridization of successive precursor fragments under designated reaction conditions.

[0053] Generation of an overhang on a double-stranded nucleic acid is generally performed with two cleavage agents. A cleavage agent includes any molecule with enzymatic activity for base excision and/or single-strand cleavage of a double-stranded nucleic acid. For example, a cleavage agent is a nicking enzyme or has nicking enzymatic activity. A cleavage agent recognizes a cleavage or nicking enzyme recognition sequence, mismatched base pair, atypical base, non-canonical or modified nucleoside or nucleobase to be directed to a specific cleavage site. In some cases, two cleavage agents have independent recognition sites and cleavage sites. In some cases, a cleavage agent generates a single-stranded cleavage, *e.g.*, a nick or a gap, involving removal of one or more nucleosides from a single-strand of a double-stranded nucleic acid. In some cases, a cleavage agent cleaves a phosphodiester bond of a single-strand in a double-stranded nucleic acid.

[0054] Provided herein are methods for creating a sticky end on a double-stranded nucleic acid comprising: (a) providing a linear double-stranded nucleic acid comprising in order an insert region, a first fusion site, and a first adaptor region; (b) creating a first nick on a first strand of the double-stranded nucleic acid with a first cleavage agent having a first recognition site and a first specific cleavage site; and (c) creating a second nick on a second strand of the double-stranded nucleic acid with a second cleavage agent having a second recognition site and a second specific cleavage site; wherein the method produces a sticky end at the first fusion site; wherein the first recognition site is in the first fusion site or the first adaptor region; and wherein the second recognition site is in the first fusion site or first adaptor region. In some cases, the first adaptor region or first fusion site comprises a sticky end motif. In some cases, the first adaptor region or first fusion site comprises a strand-adjacent nicking enzyme recognition sequence. In some cases, a precursor nucleic acid sequence comprises a fusion site and adaptor region that is not naturally adjacent to each other.

[0055] Provided herein are methods for creating sticky ends on double-stranded nucleic acid comprising: (a) providing a plurality of double-stranded nucleic acids each comprising in order an insert region, a fusion site, and an adaptor region, wherein each of the plurality of double-

stranded nucleic acids have a different fusion site; (b) creating a first nick on a first strand of each of the plurality of double-stranded nucleic acids with a first cleavage agent having a first recognition site and a first specific cleavage site; and (c) creating a second nick on a second strand of each of the plurality of double-stranded nucleic acids with a second cleavage agent having a second recognition site and a second specific cleavage site; wherein the method produces a sticky end at each fusion site of the plurality of double-stranded nucleic acids; wherein the first recognition site is in the fusion site or the adaptor region of the plurality of double-stranded nucleic acids; and wherein the second recognition site is in the fusion site or adaptor region of the plurality of double-stranded nucleic acids. In some cases, the first adaptor region or first fusion site comprises a sticky end motif. In some cases, the first adaptor region or first fusion site comprises a strand-adjacent nicking enzyme recognition sequence.

[0056] Provided herein are methods for assembling a polynucleotide comprising: (a) providing a reaction mixture comprising a first dsDNA fragment comprising a uracil base on its first strand; a second dsDNA fragment comprising a uracil base on its first strand; a first cleaving agent that cuts dsDNA on a single-strand at the site of a uracil; a second cleaving agent that cuts dsDNA on a single-strand, wherein the cleavage site of the second cleaving agent is within k bp of the uracil in an opposite strand and wherein k is between 2 and 10; and a ligase; and (b) thermocycling the reaction mixture between a maximum and a minimum temperature, thereby generating a first overhang from the first dsDNA fragment and a second overhang from the second dsDNA fragment, wherein the first and the second overhangs are complimentary, hybridizing the first and second overhangs to each other, and ligating.

[0057] Provided herein are methods for assembling a polynucleotide comprising: (a) providing a reaction mixture comprising n dsDNA fragments each comprising a first and a second strand, and a first nicking endonuclease recognition site, a first fusion site, a variable insert, a second fusion site, and a second nick enzyme recognition site, wherein the second fusion site comprises a uracil base on the first strand and the first fusion site comprises a uracil base on the second strand; a first cleaving agent that cuts dsDNA on a single-strand at the site of a uracil; a second cleaving agent that cuts dsDNA on a single-strand, wherein the cleavage site of the second cleaving agent is within k bp of the uracil in an opposite strand and wherein k is between 2 and 10; and a ligase; and (b) thermocycling the reaction mixture between a maximum and a minimum temperature, thereby generating a first overhang and a second overhang on each end of the n dsDNA fragments, wherein the second overhang on the ith of the n dsDNA fragments is reverse complementary to the first overhang on the i+1 st of the n dsDNA fragments, hybridizing the complementary overhangs to each other, and ligating.

[0058] Provided herein are fragment libraries comprising n DNA fragments, each comprising a first strand and a second strand, each i th DNA fragment comprising a first nicking endonuclease recognition site, a first fusion site, a variable insert, a second fusion site, and a second nicking endonuclease recognition site; wherein the first fusion site comprises a sequence of 5'-A (Nx) _{i} U-3' (SEQ ID NO.: 13) in the first strand; and wherein the second fusion site comprises a sequence of 5'-A (Nx) _{i} U-3' (SEQ ID NO.: 14) in the second strand; wherein Nx denotes x nucleosides; wherein (N χ) _{i} is reverse complementary to (N χ) _{$i+1$} and different from every other Nx found in any fusion site sequence within the fragment library; wherein the first nicking endonuclease recognition sites are positioned such that there is a corresponding cleavage site immediately 3' of the first fusion site in the second strand; and wherein the second nicking endonuclease recognition sites are positioned such that there is a corresponding cleavage site immediately 3' of the second fusion site in the first strand.

[0059] Provided herein are primer libraries comprising n primers, each comprising a nicking endonuclease recognition sequence and a fusion sequence comprising 5'-A (Nx); U-3' (SEQ ID NO.: 15), wherein the nicking endonuclease recognition sequence is positioned 5' of the fusion sequence. In some cases, the nicking endonuclease recognition sites are positioned such that the nicking endonuclease recognition site in a primer is capable of generating a corresponding cleavage site in a reverse complementary DNA strand 3' of a first fusion site in the reverse complementary DNA strand, if the primer were hybridized to the reverse complementary DNA strand such that the fusion sequence hybridizes to the first fusion site in the reverse complementary DNA strand. In some cases, x is selected from the list consisting of the integers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10. In some cases, n is at least 2. In some cases, n is less than 10. In some cases, the sequences of the n primers are not naturally occurring. In some cases, the primers are in a kit further comprising a nicking endonuclease, UDG, and an AP endonuclease.

[0060] A primer is said to anneal to another nucleic acid if the primer, or a portion thereof, hybridizes to a nucleotide sequence within the nucleic acid. The statement that a primer hybridizes to a particular nucleotide sequence is not intended to imply that the primer hybridizes either completely or exclusively to that nucleotide sequence.

Sticky Ends

[0061] Provided herein are methods for the creation of a sticky end on a nucleic acid using a combination of independently acting single-strand cleaving enzymes rather than a single restriction endonuclease. In some cases, a sticky end is an end of a double-stranded nucleic acid having a 5' or 3' overhang, wherein a first strand of the nucleic acid comprises one or more bases at its 5' or 3' end, respectively, which are collectively not involved in a base-pair with bases of the second strand of the double-stranded nucleic acid. An overhang is capable of annealing to a

complementary overhang under suitable reaction conditions. In some cases, "sticky end" and "overhang" are used interchangeably. Non-limiting examples of overhang lengths include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more bases. For example an overhang has 4 to 10 bases, 4 to 8 bases, or 4 to 6 bases.

[0062] Sticky end motifs are generally identified from a predetermined sequence of a target nucleic acid to be synthesized from fragments partitioned by selected identified sticky end motifs. In some cases, ANNNNT (SEQ ID NO.: 2) motifs are identified as sources of potential sticky ends in a target sequence. In some cases, GNNNNC (SEQ ID NO.: 17) motifs are identified as a source of potential sticky ends in a target sequence. Each N is independently any base. Selected sticky ends serve as fusion sites for annealing and ligating together two fragments via complementary sticky ends.

[0063] In some cases, a sticky end comprises a sequence of A (N^x)T (SEQ ID NO.: 1), wherein N^x is x number of N bases of any sequence. In some cases, a sticky end comprises a sequence of G (N^x)C (SEQ ID NO.: 16), wherein N^x is x number of N bases of any sequence. A sticky end motif is a sequence of double-stranded polynucleotides in a nucleic acid that when treated with an appropriate cleavage agent make up a sticky end. For reactions comprising a plurality of double-stranded nucleic acid fragments to be assembled, in some instances the N^x sequence or full sequence of a sticky end at the 3' end of a first nucleic acid fragment is completely or partially reverse complementary to the N^x sequence of a sticky end at the 5' end of a second nucleic acid fragment. In similar instances the 3' end of the second nucleic acid fragment has a sticky end that is completely or partially reverse complementary to the N^x sequence of sticky end at the 5' end of a third nucleic acid fragment, and so on. In some instances, the motif of the sticky end complementary between the first and second nucleic acids is the same as the motif of the sticky end complementary between the second and third nucleic acids. This sequence similarity between sticky end motifs includes motifs having identical base number and sequence identities. In some cases, sticky end motifs of a plurality of nucleic acids are the same, yet have variable identities. For example, each motif shares the sequence ANNNNT (SEQ ID NO.: 2), but two or more motifs differ in the identity of the sequence of 4, N bases. A plurality of nucleic acid fragments to be assembled may each comprise a sticky end motif of A (N^x)T (SEQ ID NO.: 1), wherein the sequence of a given motif is only shared among two of the fragments adjacent to one another in a target nucleic acid sequence. Thus, these nucleic acid fragments, under appropriate conditions, anneal to each other in a linear sequence without degeneracy in the pairing of overhangs and hence the nucleic acid order within the linear sequence.

[0064] The number of bases x in N^x in a sticky end motif described herein may be the same for all sticky end motifs for a number of nucleic acids within a plurality of nucleic acids. In some

instances, sticky end motifs belonging to a number of nucleic acids within a plurality of nucleic acids comprise sequences of A (N^x) T (SEQ ID NO.: 1), G (N^x) C (SEQ ID NO.: 16), or combinations thereof, wherein the number of bases x in N^x is the same or varies among the plurality of nucleic acids. The number of bases x in N^x may be more than or equal to about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more. In some cases, the number of bases x in N^x sticky end motifs of a plurality of nucleic acids is less than or equal to 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 bases. In some cases, the number of bases x in N^x in sticky end motifs is 2-10 bases, 3-9 bases, 4-8 bases, or 5-10 bases. In some case, a sequence of N bases in a sticky end motif described herein comprises no more than 4, 3, 2, or 1 of the same base. For example, in a sticky end motif comprising x=4 N bases, no more than 1, 2, 3 or 4 bases have the same identity. In some cases, no more than 2, 3 or 4 bases in a sticky end motif sequence have the same identity. In some cases, a sequence adjacent to a sticky end motif in a nucleic acid described herein does not comprise a G or C in the first two positions adjacent to the 3' end of the sticky end motif.

[0065] Referring to the figures, **FIG. 2** depicts the preparation and annealing of two sticky ends in a plurality of precursor nucleic acid fragments. In **FIG. 2**, a plurality of fragments spanning a predetermined target nucleic acid sequence is generated for which sticky end motif sequences have been selected (sticky end motifs X, Y, and Z) such that only two fragments will share a particular compatible sticky end. Each precursor fragment comprises target nucleic acid fragment sequence, flanked by sticky end motif sequence ANNNNT (SEQ ID NO.: 2), wherein NNNN are specific to an end pair, and having a U in place of the T at the 3' end of one strand. In alternate embodiments the sequence is GNNNNC (SEQ ID NO.: 17), herein NNNN are specific to an end pair, and having a U in place of the C at the 3' end of one strand.

[0066] Another non-limiting depiction of sticky end use is shown in the example workflow of **FIG. 4**, which generally depicts the assembly of target nucleic acids from precursor nucleic acid fragments via assembly of complementary sticky ends in the precursor fragments. Connecting adaptors of two or more fragments may be synthesized to be flanked by Type II restriction endonuclease sites that are unique to a fragment pair. Compatible ends are ligated and PCR is used to amplify the full length target nucleic acids.

[0067] *Position-specific sticky end generation*

[0068] In some cases, methods and compositions described herein use two independent cleavage events that have to occur within a distance that allow for separation of a cleaved end sequence under specified reaction conditions. For example, two different cleaving agents are used that both cut DNA only at a single-strand. In some cases, one or both of the cleaving agents cut outside of its recognition sequence (a "strand-adjacent nicking enzyme"). This allows

independency of the process from the actual sequence of the overhangs which are to be assembled at sticky end sites. In some cases, one or more of the cleavage agents recognizes or cleaves at non-canonical bases that are not part of the Watson-Crick base pairs or typical base pairs, including, but not limited to a uracil, a mismatch, and a modified base.

[0069] Further provided herein are methods for generation of a sticky end in a double-stranded nucleic acid having a sticky end motif comprises cleaving a first strand of the nucleic acid at a first position adjacent to one end of the sticky end motif and cleaving a second strand of the nucleic acid at a second position adjacent to the other end of the sticky end motif. In some cases, the first and/or second position are defined by their location next to a nicking enzyme recognition sequence. For example, a strand-adjacent nicking enzyme recognitions the nicking enzyme recognition sequence and cleaves a single-strand adjacent to the recognition sequence. In some cases, the first and/or second position are defined by the presence of a non-canonical base, wherein excision and cleavage at the non-canonical base site occurs via one or more nicking enzymes collectively having excision and endonuclease activities. In some cases, two nicks on opposite strands of a nucleic acid are within a short nick-to-nick distance from each other, *e.g.*, a distance equal to or less than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 base pairs. A nicking enzyme recognition sequence is positioned such that its cleavage site is at the desired nick-to-nick distance from the other cleavage activity that is used together to create an overhang.

[0070] A single-strand of a sticky end motif may be modified with or comprises a non-canonical base positioned directly adjacent to a target nucleic acid sequence. In some cases, a non-canonical base identifies a cleavage site. In an exemplary arrangement, an adaptor sequence comprising a sticky end motif further comprises a nicking enzyme recognition sequence adjacent to the terminal end of the sticky end motif. In this configuration, if the nicking enzyme recognition sequence defines a cleavage site adjacent to the recognition sequence and is located next to the sticky end motif, treatment with a strand-adjacent nicking enzyme introduces a nick on a single-strand between the nicking enzyme recognition sequence and sticky end motif.

Examples of non-canonical bases for inclusion in a modified sticky end motif are, without limitation, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, acetylcytosine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N-6-isopentenyl adenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 1-methyladenine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, 5-ethylcytosine, N6-adenine, N6-methyladenine, N,N-dimethyladenine, 8-bromoadenine, 7-methylguanine, 8-bromoguanine, 8-chloroguanine, 8-aminoguanine, 8-methylguanine, 8-thioguanine, 5-ethyluracil, 5-propyluracil, 5-

methylaminomethyluracil, methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid, pseudouracil, 1-methylpseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-hydroxymethyluracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-S-oxyacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, 5-(2-bromovinyl)uracil, 2-aminopurine, 6-hydroxyaminopurine, 6-thiopurine, and 2,6-diaminopurine.

[0071] In addition, the terms "nucleoside" and "nucleotide" include those moieties that contain not only conventional ribose and deoxyribose sugars, but other sugars as well. Modified nucleosides or nucleotides also include modifications on the sugar moiety, *e.g.*, wherein one or more of the hydroxyl groups are replaced with halogen atoms or aliphatic groups, or are functionalized as ethers, amines, or the like. Examples of modified sugar moieties which can be used to modify nucleosides or nucleotides at any position on their structures include, but are not limited to arabinose, 2-fluoroarabinose, xylose, and hexose, or a modified component of the phosphate backbone, such as phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, or a formacetal or analog thereof.

[0072] A nucleic acid described herein may be treated with a chemical agent, or synthesized using modified nucleotides, thereby creating a modified nucleic acid. In various embodiments, a modified nucleic acid may be cleaved, for example at the site of the modified base. For example, a nucleic acid may comprise alkylated bases, such as N³-methyladenine and N³-methylguanine, which may be recognized and cleaved by an alkyl purine DNA-glycosylase, such as DNA glycosylase I (*E. coli* TAG) or AlkA. Similarly, uracil residues may be introduced site specifically, for example by the use of a primer comprising uracil at a specific site. The modified nucleic acid may be cleaved at the site of the uracil residue, for example by a uracil N-glycosylase. Guanine in its oxidized form, 8-hydroxyguanine, may be cleaved by formamidopyrimidine DNA N-glycosylase. Examples of chemical cleavage processes include without limitation alkylation, (*e.g.*, alkylation of phosphorothioate-modified nucleic acid); cleavage of acid lability of P^{3'}-N^{5'}-phosphoroamidate-containing nucleic acid; and osmium tetroxide and piperidine treatment of nucleic acid.

Nucleic Acid Synthesis and Modification

[0073] Methods described herein provide for synthesis of a precursor nucleic acid sequence, or a target fragment sequence thereof, has a length of about or at least about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 10000, 20000, or 30000 bases. In some cases, a plurality of precursor nucleic acid fragments are prepared with sticky ends, and the

sticky ends are annealed and ligated to generate the predetermined target nucleic acid sequence having a base length of about, or at least about, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 15000, 20000, 30000, 50000, or 100000 bases. In some cases, a precursor nucleic acid sequence is assembled with another precursor nucleic acid sequence via annealing and ligation of complementary sticky ends, followed by additional rounds of sticky end generation and assembly with other precursor fragment(s) to generate a long target nucleic acid sequence. In some cases, 2, 3, 4, 5, 6, 7, 8, 9, or 10 rounds of sticky end generation and assembly are performed to generate a long target nucleic acid of predetermined sequence. The precursor nucleic acid fragment or a plurality of precursor nucleic acid fragments may span a predetermined sequence of a target gene, or portion thereof. The precursor nucleic acid fragment or a plurality of precursor nucleic acid fragments may span a vector and a plasmid sequence, or portion thereof. For example, a precursor nucleic acid fragment comprises a sequence of a cloning vector from a plasmid. In some such cases, a cloning vector is generated using de novo synthesis and an assembly method described herein, and is subsequently assembled with a precursor nucleic acid fragment or fragments of a target gene to generate an expression plasmid harboring the target gene. A vector may be a nucleic acid, optionally derived from a virus, a plasmid, or a cell, which comprises features for expression in a host cell, including, for example, an origin of replication, selectable marker, reporter gene, promoter, and/or ribosomal binding site. A host cell includes, without limitation, a bacterial cell, a viral cell, a yeast cell, and a mammalian cell. Cloning vectors useful as precursor nucleic acid fragments include, without limitation, those derived from plasmids, bacteriophages, cosmids, bacterial artificial chromosomes, yeast artificial chromosomes, and human artificial chromosomes.

[0074] Provided herein are methods for synthesis of target nucleic acid fragments having an error rate of less than 1/500, 1/1000, 1/10,000 or less compared to a predetermined sequence(s). In some cases, target fragment length is selected in light of the location of desired sticky ends, such that target fragment length varies among fragments in light of the occurrence of desired sticky ends among target fragments. In some cases, target nucleic acid fragments are synthesized to a size of at least 20 but less than 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 500, 1000, 5000, 10000, or 30000 bases. In some cases, target fragments are synthesized de novo, such as through nonenzymatic nucleic acid synthesis. In some cases, target nucleic acid fragments are synthesized from template nucleic acids, such as templates of nucleic acids that are to be assembled into a single target nucleic acid but which, in some cases, do not naturally occur adjacent to one another.

[0075] Through the synthesis of target nucleic acid fragments having at least one indeterminate position, followed by the ligation at sticky ends to adjacent target nucleic acid fragments also

having at least one indeterminate position, one can synthesize a target nucleic acid population that comprises a recombinant library of all possible combinations of the base identities at the varying positions. Alternately, at least one base position is partially indeterminate in some cases, such that two or three base alternatives are permitted. In some such cases, target nucleic acid fragments are selected such that only one base varies within a given target nucleic acid fragment, which in turn allows for each position to independently vary in the target nucleic acid library.

[0076] An example workflow of nucleic acid synthesis is shown in FIG. 6. Methods of synthesis using this workflow are, in some instances, performed to generate a plurality of target nucleic acid fragments, or oligonucleotides thereof, for assembly using sticky end methods described herein. In some cases, oligonucleotides are prepared and assembled into precursor fragments using the methods depicted in FIG. 6. The workflow is divided generally into the following processes: (1) de novo synthesis of a single stranded oligonucleic acid library, (2) joining oligonucleic acids to form larger fragments, (3) error correction, (4) quality control, and (5) shipment. Prior to de novo synthesis, an intended nucleic acid sequence or group of nucleic acid sequences is preselected. For example, a library of precursor nucleic acid fragments is preselected for generation.

[0077] In some instances, a structure comprising a surface layer 601 is provided. In the example, chemistry of the surface is functionalized in order to improve the oligonucleic acid synthesis process. Areas of low surface energy are generated to repel liquid while areas of high surface energy are generated to attract liquids. The surface itself may be in the form of a planar surface or contain variations in shape, such as protrusions or nanowells which increase surface area. In the workflow example, high surface energy molecules selected support oligonucleic acid attachment and synthesis.

[0078] In step 602 of the workflow example, a device, such as a material deposition device, is designed to release reagents in a step wise fashion such that multiple oligonucleic acids extend from an actively functionalized surface region, in parallel, one residue at a time to generate oligomers with a predetermined nucleic acid sequence. In some cases, oligonucleic acids are cleaved from the surface at this stage. Cleavage includes gas cleavage, *e.g.*, with ammonia or methylamine.

[0079] The generated oligonucleic acid libraries are placed in a reaction chamber. In some instances, the reaction chamber (also referred to as "nanoreactor") is a silicon coated well containing PCR reagents lowered onto the oligonucleic acid library 603. Prior to or after the sealing 604 of the oligonucleic acids, a reagent is added to release the oligonucleic acids from the surface. In the exemplary workflow, the oligonucleic acids are released subsequent to sealing of the nanoreactor 605. Once released, fragments of single-stranded oligonucleic acids hybridize in

order to span an entire long range sequence of DNA. Partial hybridization **605** is possible because each synthesized oligonucleic acid is designed to have a small portion overlapping with at least one other oligonucleic acid in the pool.

[**0080**] After hybridization, oligonucleic acids are assembled in a PCA reaction. During the polymerase cycles of the PCA reaction, the oligonucleic acids anneal to complementary fragments and gaps are filled in by a polymerase. Each cycle increases the length of various fragments randomly depending on which oligonucleic acids find each other. Complementarity amongst the fragments allows for forming a complete large span of double-stranded DNA **606**, in some instances, a fragment of DNA to be assembled into a target nucleic acid.

[**0081**] After PCA is complete, the nanoreactor is separated from the surface **607** and positioned for interaction with a surface having primers for PCR **608**. After sealing, the nanoreactor is subject to PCR **609** and the larger nucleic acids are amplified. After PCR **610**, the nanochamber is opened **611**, error correction reagents are added **612**, the chamber is sealed **613** and an error correction reaction occurs to remove mismatched base pairs and/or strands with poor complementarity from the double-stranded PCR amplification products **614**. The nanoreactor is opened and separated **615**. Error corrected product is next subject to additional processing steps, such as PCR, nucleic acid sorting, and/or molecular bar coding, and then packaged **622** for shipment **623**.

[**0082**] In some cases, quality control measures are taken. After error correction, quality control steps include, for example, interaction with a wafer having sequencing primers for amplification of the error corrected product **616**, sealing the wafer to a chamber containing error corrected amplification product **617**, and performing an additional round of amplification **618**. The nanoreactor is opened **619** and the products are pooled **620** and sequenced **621**. In some cases, nucleic acid sorting is performed prior to sequencing. After an acceptable quality control determination is made, the packaged product **622** is approved for shipment **623**. Alternatively, the product is a library of precursor nucleic acids to be assembled using scar-free assembly methods and compositions described herein.

[**0083**] Provided herein is library of nucleic acids each synthesized with an adaptor sequence comprising a shared primer binding sequence. In some cases, the primer binding sequence is a universal primer binding sequence shared among all primers in a reaction. In some cases, different set of primers are used for generating different final nucleic acids. In some cases, multiple populations of primers each have their own "universal" primer binding sequence that is directed to hybridize with universal primer binding sites on multiple nucleic acids in a library. In such a configuration, different nucleic acids within a population share a universal primer binding site, but differ in other sequence elements. Thus, multiple populations of nucleic acids may be

used as a template in primer extension reactions in parallel through the use of different universal primer binding sites. Universal primers may comprise a fusion site sequence that is partially or completely complementary to a sticky end motif of one of the nucleic acids. The combination of a primer binding sequence and the sticky end motif sequence is used to hybridize the primer to template nucleic acids. In some cases, primers and/or adaptor sequences further comprise a recognition sequence for a cleavage agent, such as a nicking enzyme. In some cases, primers and/or primer binding sequences in an adaptor sequence further comprise a recognition sequence for a cleavage agent, such as a nicking enzyme. In some cases, a nicking enzyme recognition sequence is introduced to extension products by a primer.

[0084] Primer extension may be used to introduce a sequence element other than a typical DNA or RNA Watson-Crick base pair, including, without being limited to, a uracil, a mismatch, a loop, or a modified nucleoside; and thus creates a non-canonical base pair in a double-stranded target nucleic acid or fragment thereof. Primers are designed to contain such sequences in a way that still allows efficient hybridization that leads to primer extension. Such non-Watson-Crick sequence elements may be used to create a nick on one strand of the resulting double-stranded nucleic acid amplicon. In some cases, a primer extension reaction is used to produce extension products incorporating uracil into a precursor nucleic acid fragment sequence. Such primer extension reactions may be performed linearly or exponentially. In some cases, a polymerase in a primer extension reaction is a 'Family A' polymerase lacking 3'-5' proofreading activity. In some cases, a polymerase in a primer extension reaction is a Family B high fidelity polymerase engineered to tolerate base pairs comprising uracil. In some cases, a polymerase in a primer extension reaction is a Kappa Uracil polymerase, a FusionU polymerase, or a Pfu turbo polymerase as commercially available.

[0085] *Nicking enzyme recognition sequences and cleavage sites*

[0086] The generation of an overhang described herein in a double-stranded nucleic acid comprises may create two independent single-stranded nicks at an end of the double-stranded nucleic acid. In some cases, the two independent single-stranded nicks are generated by two cleavage agents having cleavage activities independent from each other. In some cases, a nick is created by including a recognition site for a cleavage agent, for example in an adaptor region or fusion site. In some cases, a cleavage agent is a nicking endonuclease using a nicking endonuclease recognition sequence or any other agent that produces a site-specific single-stranded cut. For example, a mismatch repair agent that creates a gap at the site of a mismatched base-pair, or a base excision system that creates a gap at the site of a recognized nucleoside, such as a deoxy-uridine, is used to create a single-stranded cut. In some cases, a deoxy-uridine is a non-canonical base in a non-canonical base pair formed with a deoxy-adenine, a deoxy-guanine,

or a deoxy-thymine. In some cases, for example, when using a uracil containing primer in a nucleic acid extension reaction, a nucleic acid comprises a deoxy-uridine/deoxy-adenine base pair. For example a glycosylase, such as UDG, alone or in combination with an AP endonuclease, such as endonuclease VIII, is used to excise uracil and create a gap. In some cases, a second nick is created similarly using any suitable single-stranded site-specific cleavage agent; wherein the second nick is created at a site not directly across from the first nick in the double-stranded nucleic acid. Such pairs of staggered nicks, when in proximity to each other and under appropriate reaction conditions, cause a sticky end when parts of the original nucleic acid melt away from each other. In various embodiments, one or more of the cleavage sites are situated apart from the sequence of the fusion site.

[0087] Two nicks in a double-stranded nucleic acid may be created such that the resulting overhang is co-extensive with the span of a sticky end site. For example, a first nick is created at the juncture between sticky end site and adaptor region at one end of a nucleic acid; and a second nick is created at the other end of the sticky end site. Thus, only one strand along a sticky end site is kept at the end of a nucleic acid along the entire sticky end sequence, while the other is cut off. A mixture of enzymatic uracil excision activity and nicking endonuclease activity may be provided in a mixture of engineered fragments. In some cases, a strand-adjacent nicking enzyme is provided, such that sticky ends that reanneal to their cleaved terminal ends and are re-ligated across a single-strand will be re-subjected to single-strand nicking due to the reconstitution of the strand-adjacent nicking site.

[0088] Overhangs of various sizes are prepared by adjusting the distance between two nicks on opposite strands of the end of a double-stranded nucleic acid. In some cases, the distance or the length of an overhang is equal to or less than 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 bases. Overhangs may be 3' or 5' overhangs. In various embodiments, the cleavage site of a cleavage agent is a fixed distance away from its recognition site. In some cases, the fixed distance between a cleavage agent's cleavage site and recognition site is more than or equal to about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 bases or more. In some cases, the fixed distance between a cleavage agent's cleavage site and recognition site is 2-10 bases, 3-9 bases, or 4-8 bases. The cleavage site of a cleaving agent may be outside of its recognition site, for example, it is adjacent to its recognition site and the agent is a strand-adjacent nicking enzyme. In some case, the recognition site of a cleavage agent is not cleaved.

A double-stranded nucleic acid disclosed herein may be modified to comprise a non-canonical base. As a non-limiting example, a nucleic acid fragment having a sticky end motif such as A (N^x)T (SEQ ID NO. : 1) or G (N^x)C (SEQ ID NO. : 16) is prepared. In some cases, the fragment further comprises a recognition site for a single-strand cleavage agent, such as a nicking

endonuclease, having a cleavage site immediately adjacent to the last base in the sticky end motif sequence. Alternatively, the recognition site is introduced by a primer in a nucleic acid extension reaction using a strand of the fragment comprising the sticky end motif as a template. For example, the recognition site is appended to the end of the fragment in an adaptor region. In a non-limiting example, a nucleic acid extension reaction using the strand of the fragment comprising the sticky end motif, such as A (N^x) T (SEQ ID NO.: 1) or G (N^x) C (SEQ ID NO.: 16), as a template is primed with a primer comprising a sticky end sequence comprising a non-canonical base substitution. For a sticky end motif of A (N^x) T (SEQ ID NO.: 1) in a template, one such primer comprises the sequence A (N^x)' U (SEQ ID NO.: 18), wherein (N^x)' is partially or completely reverse complementary to (N^x). For a sticky end motif of A (N^x) T (SEQ ID NO.: 1) in a template, one such primer comprises the sequence A (N^x) U (SEQ ID NO.: 19). In some cases, the A (N^x)' U (SEQ ID NO.: 18) and/or A (N^x) U (SEQ ID NO.: 19) sequence on the primer is located at the very 3' end of the primer. A plurality of such primers each having a sequence of A (N^x)' U (SEQ ID NO.: 18) and/or A (N^x) U (SEQ ID NO.: 19) corresponding to a sequence of A (N^x) T in one strand of a fragment may be used to perform a nucleic acid extension reaction. The exemplary sequences described have a sticky end motif comprising a first A or G and a terminal T or C prior to non-canonical base incorporation. However, any sticky end motif sequence is useful with the methods described herein.

Libraries

[0089] Provided herein are fragment libraries comprising n double-stranded precursor nucleic acid fragments. In some cases, each double-stranded nucleic acid precursor fragment of the n double-stranded nucleic acid fragments comprises a first nicking endonuclease recognition site, a first fusion site, a variable insert of predetermined fragment sequence, a second fusion site, and a second nick enzyme recognition site, optionally in that order. In some cases, the first fusion site comprises or is a first sticky end motif and the second fusion site comprises or is a second sticky end motif. In some instances, the first fusion site has the sequence of 5'-A (N^x)_{i,1} U-3' (SEQ ID NO.: 13) in the first strand, wherein denotes N^x x bases or nucleosides and the subscript "i,1" in (N^x)_{i,1} denotes the first strand of the ith fragment. In some cases, the second fusion site has the sequence of 5'-A (N^x)_{i,2} U-3' (SEQ ID NO.: 14) in the second strand, wherein denotes N^x x bases or nucleosides and the subscript "i,2" in (N^x)_{i,2} denotes the second strand of the ith fragment. In some instances (N^x)_{i,2} is completely or partially reverse complementary to (N^x)_{i+1,1} in the first strand of the i+1 'th fragment. Each N^x found in the fusion site sequences are the same or different than the N^x in any other fusion site sequence found within the fragment library. In some cases, the first nicking endonuclease recognition site is positioned such that there is a corresponding cleavage site immediately 3' of the first fusion site in the second strand and the

second nicking endonuclease recognition site is positioned such that there is a corresponding cleavage site immediately 3' of the second fusion site in the first strand.

[0090] A fragment library may comprise a starter DNA fragment comprising a variable insert, a second fusion site, and a second nick enzyme recognition site. In some cases, the second fusion site of the starter DNA fragment comprises a sequence of 5'-A (N^x)_{s,2} U-3' (SEQ ID NO.: 20), wherein the subscript "s₂" in (N^x)_{s,2} denotes the second strand of the starter fragment and (N^x)_{s,2} is reverse complementary to (N^x)_{i,i} in one of the fusion sites of the first nucleic acid fragment in the library. Similarly, the fragment library may also comprise a finishing DNA fragment comprising a first nicking endonuclease recognition site, a first fusion site, and a variable insert. In some cases, the first fusion site comprises a sequence of 5'-A (N^x)_{f,i} U-3' (SEQ ID NO.: 21), wherein the subscript "f_i" in (N^x)_{f,i} denotes the first strand of the finishing fragment and (N^x)_{f,i} is reverse complementary to (N^x)_{n,2} in one of the fusion sites of the nth nucleic acid fragment in the library. In some cases, the first and/or the second nicking endonuclease recognition sites are the same in all the fragments in the fragment library. In various embodiments, the fragment library comprises about or at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 75, 100, 125, 150, 200, 250, 500, or more nucleic acid fragments. In some instances, the fragment library comprises 2-75 fragments, 4-125 fragments, or 5-10 fragments.

[0091] Further described herein is a primer library of n primers. Each primer within the library may comprise a recognition sequence such a nicking endonuclease recognition sequence, and a fusion sequence comprising a sticky end motif. For example, a sticky end motif having the sequence 5'-A (N^x)_iU-3' (SEQ ID NO.: 15). In some cases, the recognition sequence is positioned 5' of the fusion site sequence. In some cases, the recognition sequence is positioned such that the recognition site in a primer is capable of generating a corresponding cleavage site in a reverse complementary DNA strand 3' of a first fusion site in the reverse complementary DNA strand, if the primer were hybridized to the reverse complementary DNA strand such that the fusion sequence hybridizes to the first fusion site in the reverse complementary DNA strand. In various aspects, a primer library described herein comprises about or at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 75, 100, 125, 150, 200, 250, 500, or more primers.

Cleavage agents

[0092] Provided herein are methods where two or more independent cleaving agents are selected to generate single-stranded cleavage on opposite strands of a double-stranded nucleic acid. As used herein, "nick" generally refers to enzymatic cleavage of only one strand of a double-stranded nucleic acid at a particular region, while leaving the other strand intact, regardless of whether one or more bases are removed. In some cases, one or more bases are removed while in

other cases no bases are removed and only phosphodiester bonds are broken. In some instances, such cleavage events leave behind intact double-stranded regions lacking nicks that are a short distance apart from each other on the double-stranded nucleic acid, for example a distance of about or at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 bases or more. In some cases, the distance between the intact double-stranded regions is equal to or less than 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 bases. In some instances, the distance between the intact double-stranded regions is 2 to 10 bases, 3 to 9 bases, or 4 to 8 bases.

[0093] Cleavage agents used in methods described herein may be selected from nicking endonucleases, DNA glycosylases, or any single-stranded cleavage agents described in further detail elsewhere herein. Enzymes for cleavage of single-stranded DNA may be used for cleaving heteroduplexes in the vicinity of mismatched bases, D-loops, heteroduplexes formed between two strands of DNA which differ by a single base, an insertion or deletion. Mismatch recognition proteins that cleave one strand of the mismatched DNA in the vicinity of the mismatch site may be used as cleavage agents. Nonenzymatic cleaving may also be done through photodegradation of a linker introduced through a custom oligonucleotide used in a PCR reaction.

[0094] Provided herein are fragments designed and synthesized such that the inherent cleavage sites are utilized in the preparation of fragments for assembly. For instance, these inherent cleavage sites are supplemented with a cleavage site that is introduced, *e.g.*, by recognition sites in adaptor sequences, by a mismatch, by a uracil, and/or by an un-natural nucleoside. In various embodiments, described herein is a plurality of double stranded nucleic acids such as dsDNA, comprising an atypical DNA base pair comprising a non-canonical base in a fusion site and a recognition site for a single-strand cleaving agent. Compositions according to embodiments described herein, in many cases, comprise two or more cleaving agents. In some cases, a first cleaving agent has the atypical DNA base pair as its recognition site and the cleaving agent cleaves a single-strand at or a fixed distance away from the atypical DNA base pair. In some cases, a second cleaving agent has an independent single-strand cleaving and/or recognition activity from the first cleaving agent. In some cases, the nucleic acid molecules in the composition are such that the recognition site for the second single-strand cleaving agent is not naturally adjacent to the fusion site or the remainder of the nucleic acid in any of the plurality of double stranded nucleic acids in the composition. In some instances, the cleavage sites of two cleavage agents are located on opposite strands.

[0095] *Type II enzymes*

[0096] Provided herein are methods and compositions described herein use a Type II restriction endonuclease in as a cleavage agent. Type II enzymes cleave within or at short specific distances

from a recognition site. There are a variety of different type II enzymes known in the art, many of which differ in the sequence they recognize. Type II restriction endonucleases comprise many sub-types with varying activities. Exemplary Type II restriction endonucleases include, without limitation, Type IIP, Type IIF, Type IIB (*e.g.* Bcgl and BpII), Type HE (*e.g.* NaeI), and Type IIM (DpnI) restriction endonucleases. The most common Type II enzymes are those like HhaI, HindIII, and NotI that cleave DNA within their recognition sequences. Many recognize DNA sequences that are symmetric, because, without being bound by theory, they bind to DNA as homodimers, but a few, (*e.g.*, BbvCI: CCTCAGC (SEQ ID NO.: 22)) recognize asymmetric DNA sequences, because, without being bound by theory, they bind as heterodimers. Some enzymes recognize continuous sequences (*e.g.*, EcoRI: GAATTC (SEQ ID NO.: 23)) in which the two half-sites of the recognition sequence are adjacent, while others recognize discontinuous sequences (*e.g.*, BglI: GCCNNNNNGGC (SEQ ID NO.: 24)) in which the half-sites are separated. Using this type, a 3'-hydroxyl on one side of each cut and a 5'-phosphate on the other may be created upon cleavage.

[0097] The next most common Type II enzymes, usually referred to as "Type IIS" are those like FokI and AlwI that cleave outside of their recognition sequence to one side. Type IIS enzymes recognize sequences that are continuous and asymmetric. Type IIS restriction endonucleases (*e.g.* FokI) cleave DNA at a defined distance from their non-palindromic asymmetric recognition sites. These enzymes may function as dimers. Type IIS enzymes typically comprise two distinct domains, one for DNA binding, and the other for DNA cleavage. Type IIA restriction endonucleases recognize asymmetric sequences but can cleave symmetrically within the recognition sequences (*e.g.* **BbvCI cleaves 2 bases downstream of the 5'-end of each strand of CCTCAGC (SEQ ID NO.: 25)**). Similar to Type IIS restriction endonucleases, Type IIT restriction enzymes (*e.g.*, BpuIOI and BslI) are composed of two different subunits. Type IIG restriction enzymes, the third major kind of Type II enzyme, are large, combination restriction-and-modification enzymes, Type IIG restriction endonucleases (*e.g.* Eco57I) do have a single subunit, like classical Type II restriction enzymes. The two enzymatic activities typically reside in the same protein chain. These enzymes cleave outside of their recognition sequences and can be classified as those that recognize continuous sequences (*e.g.*, AclI: CTGAAG (SEQ ID NO.: 26)) and cleave on just one side; and those that recognize discontinuous sequences (*e.g.*, Bcgl: CGANNNNNNTGC (SEQ ID NO.: 27)) and cleave on both sides releasing a small fragment containing the recognition sequence. When these enzymes bind to their substrates, they may switch into either restriction mode to cleave the DNA, or modification mode to methylate it.

[0098] Type III enzymes are also large combination restriction-and-modification enzymes. They cleave outside of their recognition sequences and require two such sequences in opposite

orientations within the same DNA molecule to accomplish cleavage. Type IV enzymes recognize modified DNA, *e.g.* methylated, hydroxymethylated and glucosyl-hydroxymethylated DNA and are exemplified by the McrBC and Mrr systems of *E. coli*.

[0099] Some naturally occurring and recombinant endonucleases make single-strand breaks. These nicking endonucleases (NEases) typically recognize non-palindromes. They can be *bona fide* nicking enzymes, such as frequent cutter Nt.CviPII and Nt.CviQII, or rare-cutting homing endonucleases (HEases) I-BasI and I-HmuI, both of which recognize a degenerate 24-bp sequence. As well, isolated large subunits of heterodimeric Type IIS REases such as BtsI, BsrDI and BstNBI/BspD6I display nicking activity.

[0100] Thus, properties of restriction endonucleases that make double-strand cuts may be retained by engineering variants of these enzymes such that they make single-strand breaks. In various embodiments, recognition sequence-specific nicking endonucleases are used as cleavage agents that cleave only a single-strand of double-stranded DNA at a cleavage site. Nicking endonucleases useful in various embodiments of methods and compositions described herein include Nb.BbvCI, Nb.BsmI, Nb.BsrDI, Nb.BtsI, Nt.AlwI, Nt.BbvCI, Nt.BsmAI, Nt.BspQI, Nt.BstNBI, and Nt.CviPII, used either alone or in various combinations. In various embodiments, nicking endonucleases that cleave outside of their recognition sequence, *e.g.* Nb.BsrDI, Nb.BtsI, Nt.AlwI, Nt.BsmAI, Nt.BspQI, Nt.BstNBI, and Nt.CviPII, are used. In some instances, nicking endonucleases that cut within their recognition sequences, *e.g.* Nb.BbvCI, Nb.BsmI, or Nt.BbvCI are used. Recognition sites for the various specific cleavage agents used herein, such as the nicking endonucleases, comprise a specific nucleic acid sequence. The nickase Nb.BbvCI (New England Biolabs, Ipswich, Mass) nicks at the following cleavage site with respect to its recognition site (with "|" specifying the nicking (cleavage) site and "N" representing any nucleoside, *e.g.* one of C, A, G or T):

[0101] 5'... CCTCA GC ... 3' (SEQ ID NO.: 28)

[0102] 3'... GGAGT|CG ... 5' (SEQ ID NO.: 29)

[0103] The nickase Nb.BsmI (New England Biolabs, Ipswich, Mass) nicks at the following cleavage site with respect to its recognition site:

[0104] 5' ... GAATGCN ... 3' (SEQ ID NO.: 30)

[0105] 3' ... CTTAC|GN ... 5' (SEQ ID NO.: 31)

[0106] The nickase Nb.BsrDI (New England Biolabs, Ipswich, Mass) nicks at the following cleavage site with respect to its recognition site:

[0107] 5' ... GCAATGNN ... 3' (SEQ ID NO.: 32)

[0108] 3' ... CGTTAC|NN ... 5' (SEQ ID NO.: 33)

[0109] The nickase Nb.BtsI (New England Biolabs, Ipswich, Mass) nicks at the following cleavage site with respect to its recognition site:

[0110] 5' ... GCAGTGNN ... 3' (SEQ ID NO.: 34)

[0111] 3' ... CGTCAC|NN ... 5' (SEQ ID NO.: 35)

[0112] The nickase Nt.AlwI (New England Biolabs, Ipswich, Mass) nicks at the following cleavage site with respect to its recognition site:

[0113] 5' ... GGATCNNNN|N ... 3' (SEQ ID NO.: 36)

[0114] 3' ... CCTAGNNNNN ... 5' (SEQ ID NO.: 37)

[0115] The nickase Nt.BbvCI (New England Biolabs, Ipswich, Mass) nicks at the following cleavage site with respect to its recognition site:

[0116] 5' ... CC|TCAGC ... 3' (SEQ ID NO.: 38)

[0117] 3' ... GGAGTCG ... 5' (SEQ ID NO.: 39)

[0118] The nickase Nt.BsmAI (New England Biolabs, Ipswich, Mass) nicks at the following cleavage site with respect to its recognition site:

[0119] 5' ... GTCTCN|N ... 3' (SEQ ID NO.: 40)

[0120] 3' ... CAGAGNN ... 5' (SEQ ID NO.: 41)

[0121] The nickase Nt.BspQI (New England Biolabs, Ipswich, Mass) nicks at the following cleavage site with respect to its recognition site:

[0122] 5' ... GCTCTTCNI ... 3' (SEQ ID NO.: 42)

[0123] 3' ... CGAGAAGN ... 5' (SEQ ID NO.: 43)

[0124] The nickase Nt.BstNBI (New England Biolabs, Ipswich, Mass) nicks at the following cleavage site with respect to its recognition site:

[0125] 5' ... GAGTCNNNN|N ... 3' (SEQ ID NO.: 44)

[0126] 3' ... CTCAGNNNNN ... 5' (SEQ ID NO.: 45)

[0127] The nickase Nt.CviPII (New England Biolabs, Ipswich, Mass) nicks at the following cleavage site with respect to its recognition site (wherein D denotes A or G or T and wherein H denotes A or C or T):

[0128] 5' ... |CCD ... 3' (SEQ ID NO.: 46)

[0129] 3' ... GGH ... 5' (SEQ ID NO.: 47)

[0130] *Non-canonical base recognizing enzymes*

[0131] A non-canonical base and/or non-canonical base pair in a sticky end motif and/or adaptor sequence may be recognized by an enzyme for cleavage at its 5' or 3' end. In some instances, the non-canonical base and/or non-canonical base pair comprises a uracil base. In some cases, the enzyme is a DNA repair enzyme. In some cases, the base and/or non-canonical base pair is recognized by an enzyme that catalyzes a first step in base excision, for example, a

DNA glycosylase. A DNA glycosylase is useful for removing a base from a nucleic acid while leaving the backbone of the nucleic acid intact, generating an apurinic or apyrimidinic site, or AP site. This removal is accomplished by flipping the base out of a double-stranded nucleic acid followed by cleavage of the N-glycosidic bond.

[0132] The non-canonical base or non-canonical base pair may be recognized by a bifunctional glycosylase. In this case, the glycosylase removes a non-canonical base from a nucleic acid by N-glycosylase activity. The resulting apurinic/apyrimidinic (AP) site is then incised by the **AP lyase activity of bifunctional glycosylase via β -elimination of the 3' phosphodiester bond.**

[0133] The glycosylase and/or DNA repair enzyme may recognize a uracil or a non-canonical base pair comprising uracil, for example U:G and/or U:A. Nucleic acid base substrates recognized by a glycosylase include, without limitation, uracil, 3-meA (3-methyladenine), hypoxanthine, 8-oxoG, FapyG, FapyA, Tg (thymine glycol), hoU (hydroxyuracil), hmU (hydroxymethyluracil), fU (formyluracil), hoC (hydroxycytosine), fC (formylcytosine), oxidized base, alkylated base, deaminated base, methylated base, and any non-canonical nucleobase provided herein or known in the art. In some cases, the glycosylase and/or DNA repair enzyme recognizes oxidized bases such as 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG) and 8-oxoguanine (8-oxo). Glycosylases and/or DNA repair enzymes which recognize oxidized bases include, without limitation, OGG1 (8-oxoG DNA glycosylase 1) or *E. coli* Fpg (recognizes 8-oxoG:C pair), MYH (MutY homolog DNA glycosylase) or *E. coli* MutY (recognizes 8-oxoG:A), NEIL1, NEIL2 and NEIL3. In some cases, the glycosylase and/or DNA repair enzyme recognizes methylated bases such as 3-methyladenine. An example of a glycosylase that recognizes methylated bases is *E. coli* AlkA or 3-methyladenine DNA glycosylase II, Magi and MPG (methylpurine glycosylase). Additional non-limiting examples of glycosylases include SMUG1 (single-strand specific monofunctional uracil DNA glycosylase 1), TDG (thymine DNA glycosylase), MBD4 (methyl-binding domain glycosylase 4), and NTHL1 (endonuclease III-like 1). Exemplary DNA glycosylases include, without limitation, uracil DNA glycosylases (UDGs), helix-hairpin-helix (HhH) glycosylases, 3-methyl-purine glycosylase (MPG) and endonuclease VHI-like (NEIL) glycosylases. Helix-hairpin-helix (HhH) glycosylases include, without limitation, Nth (homologs of the *E. coli* EndoIII protein), Ogg1 (8-oxoG DNA glycosylase I), MutY/Mig (A/G-mismatch-specific adenine glycosylase), AlkA (alkyladenine-DNA glycosylase), MpgII (N-methylpurine-DNA glycosylase II), and OggII (8-oxoG DNA glycosylase II). Exemplary 3-methyl-purine glycosylases (MPGs) substances include, in non-limiting examples, alkylated bases including 3-meA, 7-meG, 3-meG and ethylated bases. Endonuclease VHI-like glycosylase substrates include, without limitation, oxidized pyrimidines (e.g., Tg, 5-hC, FaPyA, PaPyG), 5-hU and 8-oxoG.

[0134] Exemplary uracil DNA glycosylases (UDGs) include, without limitation, thermophilic uracil DNA glycosylases, uracil-N glycosylases (UNGs), mismatch-specific uracil DNA glycosylases (MUGs) and single-strand specific monofunctional uracil DNA glycosylases (SMUGs). In non-limiting examples, UNGs include UNG1 isoforms and UNG2 isoforms. In non-limiting examples, MUGs include thymidine DNA glycosylase (TDG). A UDG may be active against uracil in ssDNA and dsDNA.

[0135] The non-canonical base pair included in a fragment disclosed herein is a mismatch base pair, for example a homopurine pair or a heteropurine pair. In some cases, a primer described herein comprises one or more bases which form a mismatch base pair with a base of a target nucleic acid or with a base of an adaptor sequence connected to a target nucleic acid. In some cases, an endonuclease, exonuclease, glycosylase, DNA repair enzyme, or any combination thereof recognizes the mismatch pair for subsequent removal and cleavage. For example, the TDG enzyme is capable of excising thymine from G:T mismatches. In some cases, the non-canonical base is released from a dsDNA molecule by a DNA glycosylase resulting in an abasic site. This abasic site (AP site) is further processed by an endonuclease which cleaves the phosphate backbone at the abasic site. Endonucleases included in methods herein may be AP endonucleases. For example, the endonuclease is a class I or class II AP endonuclease which incises DNA at the phosphate groups 3' and 5' to the baseless site leaving 3' OH and 5' phosphate termini. The endonuclease may also be a class III or class IV AP endonuclease which cleaves DNA at the phosphate groups 3' and 5' to the baseless site to generate 3' phosphate and 5' OH. In some cases, an endonuclease cleaving a fragment disclosed herein is an AP endonuclease which is grouped in a family based on sequence similarity and structure, for example, AP endonuclease family 1 or AP endonuclease family 2. Examples of AP endonuclease family 1 members include, without limitation, *E. coli* exonuclease III, *S. pneumoniae* and *B. subtilis* exonuclease A, mammalian AP endonuclease 1 (API), *Drosophila* recombination repair protein 1, *Arabidopsis thaliana* apurinic endonuclease-redox protein, *Dictyostelium* DNA-(apurinic or apyrimidinic site) lyase, enzymes comprising one or more domains thereof, and enzymes having at least 75% sequence identity to one or more domains or regions thereof. Examples of AP endonuclease family 2 members include, without limitation, bacterial endonuclease IV, fungal and *Caenorhabditis elegans* apurinic endonuclease APN1, *Dictyostelium* endonuclease 4 homolog, Archaeal probable endonuclease 4 homologs, mimivirus putative endonuclease 4, enzymes comprising one or more domains thereof, and enzymes having at least 75% sequence identity to one or more domains or regions thereof. Exemplary, endonucleases include endonucleases derived from both Prokaryotes (e.g., endonuclease IV, RecBCD endonuclease, T7 endonuclease, endonuclease II) and Eukaryotes (e.g., *Neurospora*

endonuclease, SI endonuclease, PI endonuclease, Mung bean nuclease I, Ustilago nuclease). In some case, an endonuclease functions as both a glycosylase and an AP-lyase. The endonuclease may be endonuclease VIII. In some cases, the endonuclease is SI endonuclease. In some instances, the endonuclease is endonuclease III. The endonuclease may be an endonuclease IV. In some case, an endonuclease is a protein comprising an endonuclease domain having endonuclease activity, i.e., cleaves a phosphodiester bond.

[0136] Provided herein are methods where a non-canonical base is removed with a DNA excision repair enzyme and endonuclease or lyase, wherein the endonuclease or lyase activity is optionally from an excision repair enzyme or a region of the excision repair enzyme. Excision repair enzymes include, without limitation, Methyl Purine DNA Glycosylase (recognizes methylated bases), 8-Oxo-Guanine Glycosylase 1 (recognizes 8-oxoG:C pairs and has lyase activity), Endonuclease Three Homolog 1 (recognizes T-glycol, C-glycol, and formamidopyrimidine and has lyase activity), inosine, hypoxanthine-DNA glycosylase; 5-Methylcytosine, 5-Methylcytosine DNA glycosylase; Formamidopyrimidine-DNA-glycosylase (excision of oxidized residue from DNA: hydrolysis of the N-glycosidic bond (DNA glycosylase), and beta-elimination (AP-lyase reaction)). In some cases, the DNA excision repair enzyme is uracil DNA glycosylase. DNA excision repair enzymes also include, without limitation, Aag (catalyzes excision of 3-methyladenine, 3-methylguanine, 7-methylguanine, hypoxanthine, 1,N6-ethenoadenine), endonuclease III (catalyzes excision of cis- and trans- thymine glycol, 5,6-dihydrothymine, 5,6-dihydroxydihydrothymine, 5-hydroxy-5-methylhydantoin, 6-hydroxy-5,6-dihydropyrimidines, 5-hydroxycytosine and 5-hydroxyuracil, 5-hydroxy-6-hydrothymine, 5,6-dihydrouracil, 5-hydroxy-6-hydrouracil, AP sites, uracil glycol, methyltartronylurea, alloxan), endonuclease V (cleaves AP sites on dsDNA and ssDNA), Fpg (catalyzes excision of 8-oxoguanine, 5-hydroxycytosine, 5-hydroxyuracil, aflatoxin-bound imidazole ring-opened guanine, imidazole ring-opened N-2-aminofluorene-C8-guanine, open ring forms of 7-methylguanine), and Mug (catalyzes the removal of uracil in U:G mismatches in double-stranded oligonucleic acids, excision of 3,N4-ethenocytosine (eC) in eC:G mismatches in double-, or single-stranded oligonucleic acids). Non-limiting DNA excision repair enzymes are listed in Curr Protoc Mol Biol. 2008 Oct;Chapter 3:Unit3.9. DNA excision repair enzymes, such as endonucleases, may be selected to excise a specific non-canonical base. As an example, endonuclease V, *T. maritima* is a 3'-endonuclease which initiates the removal of deaminated bases such as uracil, hypoxanthine, and xanthine. In some cases, a DNA excision repair enzyme having endonuclease activity functions to remove a modified or non-canonical base from a strand of a dsDNA molecule without the use of an enzyme having glycosylase activity.

[0137] In some cases, a DNA excision repair enzyme ("DNA repair enzyme") comprises glycosylase activity, lyase activity, endonuclease activity, or any combination thereof. In some cases, one or more DNA excision repair enzymes are used in the methods described herein, for example one or more glycosylases or a combination of one or more glycosylases and one or more endonucleases. As an example, Fpg (formamidopyrimidine [fapy]-DNA glycosylase), also known as 8-oxoguanine DNA glycosylase, acts both as a N-glycosylase and an AP-lyase. The N-glycosylase activity releases a non-canonical base (*e.g.*, 8-oxoguanine, 8-oxoadenine, fapy-guanine, methy-fapy-guanine, fapy-adenine, aflatoxin **Bi**-fapy-guanine, 5-hydroxy-cytosine, 5-hydroxy-uracil) from dsDNA, generating an abasic site. The lyase activity then cleaves both 3' and 5' to the abasic site thereby removing the abasic site and leaving a 1 base gap or nick. Additional enzymes which comprise more than enzymatic activities include, without limitation, endonuclease III (Nth) protein from *E. coli* (N-glycosylase and AP-lyase) and *Tma* endonuclease III (N-glycosylase and AP-lyase). For a list of DNA repair enzymes having lyase activity, see the New England BioLabs® Inc. catalog.

[0138] Provided herein are methods where mismatch endonucleases are used to nick DNA in the region of mismatches or damaged DNA, including but not limited to T7 Endonuclease I, *E. coli* Endonuclease V, T4 Endonuclease VII, mung bean nuclease, Cel-1 endonuclease, *E. coli* Endonuclease IV and UVDE. Cel-1 endonuclease from celery and similar enzymes, typically plant enzymes, exhibit properties that detect a variety of errors in double-stranded nucleic acids. For example, such enzymes can detect polynucleotide loops and insertions, detect mismatches in base pairing, recognize sequence differences in polynucleotide strands between about 100 bp and 3 kb in length and recognize such mutations in a target polynucleotide sequence without substantial adverse effects of flanking DNA sequences.

[0139] Provided herein are methods where one or more non-canonical bases are excised from a dsDNA molecule which is subsequently treated with an enzyme comprising exonuclease activity. In some cases, the exonuclease comprises 3' DNA polymerase activity. Exonucleases include those enzymes in the following groups: exonuclease I, exonuclease II, exonuclease III, exonuclease IV, exonuclease V, exonuclease VI, exonuclease VII, and exonuclease VIII. In some cases, an exonuclease has AP endonuclease activity. In some cases, the exonuclease is any enzyme comprising one or more domains or amino acid regions suitable for cleaving nucleotides from either 5' or 3' end or both ends, of a nucleic acid chain. Exonucleases include wild-type exonucleases and derivatives, chimeras, and/or mutants thereof. Mutant exonucleases include enzymes comprising one or more mutations, insertions, deletions or any combination thereof within the amino acid or nucleic acid sequence of an exonuclease.

[0140] Provided herein are methods where a polymerase is provided to a reaction comprising an enzyme treated dsDNA molecule, wherein one or more non-canonical bases of the dsDNA molecule has been excised, for example, by treatment with one or more DNA repair enzymes. In some cases, the DNA product has been treated with a glycosylase and endonuclease to remove a non-canonical base. In some cases, one or more nucleotides (e.g., dNTPs) are provided to a reaction comprising the treated dsDNA molecule and the polymerase. In some instances, the DNA product has been treated with a UDG and endonuclease VIII to remove at least one uracil. In some cases, one or more nucleotides (e.g., dNTPs) are provided to a reaction comprising the treated dsDNA molecule and the polymerase.

[0141] *DNA repair enzymes*

[0142] Provided herein are methods where a site-specific base excision reagents comprising one or more enzymes are used as cleavage agents that cleave only a single-strand of double-stranded DNA at a cleavage site. A number of repair enzymes are suitable alone or in combination with other agents to generate such nicks. An exemplary list of repair enzymes is provided in Table 1. Homologs or non-natural variants of the repair enzymes, including those in Table 1, are also be used according to various embodiments. Any of the repair enzymes for use according to the methods and compositions described herein may be naturally occurring, recombinant or synthetic. In some instances, a DNA repair enzyme is a native or an in vitro-created chimeric protein with one or more activities. Cleavage agents, in various embodiments, comprise enzymatic activities, including enzyme mixtures, which include one or more of nicking endonucleases, AP endonucleases, glycosylases and lyases involved in base excision repair.

[0143] Without being bound by theory, a damaged base is removed by a DNA enzyme with glycosylase activity, which hydrolyses an N-glycosylic bond between the deoxyribose sugar moiety and the base. For example, an *E. coli* glycosylase and an UDG endonuclease act upon deaminated cytosine while two 3-mAde glycosylases from *E. coli* (TagI and TagII) act upon alkylated bases. The product of removal of a damaged base by a glycosylase is an AP site (apurinic/apyrimidinic site), also known as an abasic site, is a location in a nucleic acid that has neither a purine nor a pyrimidine base. DNA repair systems are often used to correctly replace the AP site. This is achieved in various instances by an AP endonuclease that nicks the sugar phosphate backbone adjacent to the AP site and the abasic sugar is removed. Some naturally occurring or synthetic repair systems include activities, such as the DIMA polymerase/DNA ligase activity, to insert a new nucleotide.

[0144] Repair enzymes are found in prokaryotic and eukaryotic cells. Some enzymes having applicability herein have glycosylase and AP endonuclease activity in one molecule. AP endonucleases are classified according to their sites of incision. Class I AP endonucleases and

class II AP endonucleases incise DNA at the phosphate groups 3' and 5' to the baseless site leaving 3'-OH and 5'-phosphate termini. Class III and class IV AP endonucleases also cleave DNA at the phosphate groups 3' and 5' to the baseless site, but they generate a 3'-phosphate and a 5'-OH.

[0145] In some cases, AP endonucleases remove moieties attached to the 3' OH that inhibit polynucleotide polymerization. For example a 3' phosphate is converted to a 3' OH by *E. coli* endonuclease IV. In some cases, AP endonucleases work in conjunction with glycosylases to engineer nucleic acids at a site of mismatch, a non-canonical nucleoside or a base that is not one of the major nucleosides for a nucleic acid, such as a uracil in a DNA strand.

[0146] Examples of glycosylase substrates include, without limitation, uracil, hypoxanthine, 3-methyladenine (3-mAde), formamidopyrimidine (FAPY), 7,8 dihydro-8-oxyguanine and hydroxymethyluracil. In some instances, glycosylase substrates incorporated into DNA site-specifically by nucleic acid extension from a primer comprising the substrate. In some instances, glycosylase substrates are introduced by chemical modification of a nucleoside, for example by deamination of cytosine, *e.g.* by bisulfate, nitrous acids, or spontaneous deamination, producing uracil, or by deamination of adenine by nitrous acids or spontaneous deamination, producing hypoxanthine. Other examples of chemical modification of nucleic acids include generating 3-mAde as a product of alkylating agents, FAPY (7-mGua) as product of methylating agents of DNA, 7,8-dihydro-8 oxoguanine as a mutagenic oxidation product of guanine, 4,6-diamino-5-FAPY produced by gamma radiation, and hydroxymethyluracil produced by ionizing radiation or oxidative damage to thymidine. Some enzymes comprise AP endonuclease and glycosylase activities that are coordinated either in a concerted manner or sequentially.

[0147] Examples of polynucleotide cleavage enzymes used to generate single-stranded nicks include the following types of enzymes derived from but not limited to any particular organism or virus or non-naturally occurring variants thereof: *E. coli* endonuclease IV, Tth endonuclease IV, human AP endonuclease, glycosylases, such as UDG, *E. coli* 3-methyladenine DNA glycoylase (AlkA) and human Aag, glycosylase/lyases, such as *E. coli* endonuclease III, *E. coli* endonuclease VIII, *E. coli* Fpg, human OGG1, and T4 PDG, and lyases. Exemplary additional DNA repair enzymes are listed in Table 1.

Table 1. DNA repair enzymes.

Gene Name	Activity	Accession Number
UNG	Uracil-DNA glycosylase	NM_080911
SMUG1	Uracil-DNA glycosylase	NM_014311
MBD4	Removes U or T opposite G at CpG sequences	NM_003925

TDG	Removes U, T or ethenoC opposite G	NM_003211
OGG1	Removes 8-oxoG opposite C	NM_016821
MUTYH (MYH)	Removes A opposite 8-oxoG	NM_012222
NTHL1 (NTH1)	Removes Ring-saturated or fragmented pyrimidines	NM_002528
MPG	Removes 3-meA, ethenoA, hypoxanthine	NM_002434
NEIL1	Removes thymine glycol	NM_024608
NEIL2	Removes oxidative products of pyrimidines	NM_145043
XPC	Binds damaged DNA as complex with RAD23B, CETN2	NM_004628
RAD23B (HR23B)	Binds damaged DNA as complex with XPC, CETN2	NM_002874
CETN2	Binds damaged DNA as complex with XPC, RAD23B	NM_004344
RAD23A (HR23A)	Substitutes for HR23B	NM_005053
XPA	Binds damaged DNA in preincision complex	NM_000380
RPA1	Binds DNA in preincision complex	NM_002945
RPA2	Binds DNA in preincision complex	NM_002946
RPA3	Binds DNA in preincision complex	NM_002947
ERCC5 (XPG)	3' incision	NM_000123
ERCC1	5' incision subunit	NM_001983
ERCC4 (XPF)	5' incision subunit	NM_005236
LIG1	DNA joining	NM_000234
CKNI(CSA)	Cockayne syndrome; Needed for transcription-coupled NER	NM_000082
ERCC6 (CSB)	Cockayne syndrome; Needed for transcription-coupled NER	NM_000124
XAB2 (HCNP)	Cockayne syndrome; Needed for transcription-coupled NER	NM_020196
DDB1	Complex defective in XP group E	NM_001923
DDB2	DDB1, DDB2	NM_000107
MMS19L (MMS19)	Transcription and NER	NM_022362
FEN1 (DNase IV)	Flap endonuclease	NM_004111
SPO11	endonuclease	NM_012444
FLJ35220 (ENDOV)	incision 3' of hypoxanthine and uracil	NM_173627
FANCA	Involved in tolerance or repair of DNA crosslinks	NM_000135
FANCB	Involved in tolerance or repair of DNA crosslinks	NM_152633

FANCC	Involved in tolerance or repair of DNA crosslinks	NM_000136
FANCD2	Involved in tolerance or repair of DNA crosslinks	NM_033084
FANCE	Involved in tolerance or repair of DNA crosslinks	NM_021922
FANCF	Involved in tolerance or repair of DNA crosslinks	NM_022725
FANCG (XRCC9)	Involved in tolerance or repair of DNA crosslinks	NM_004629
FANCL	Involved in tolerance or repair of DNA crosslinks	NM_018062
DCLRE1A (SNM1)	DNA crosslink repair	NM_014881
DCLRE1B (SNM1B)	Related to SNM1	NM_022836
NEIL3	Resembles NEIL1 and NEIL2	NM_018248
ATRIP (TREX1)	ATR-interacting protein 5' alternative ORF of the TREX1 /ATRIP gene	NMJ30384
NTH	Removes damaged pyrimidines	NP_416150.1
NEI	Removes damaged pyrimidines	NP_415242.1
NFI	Deoxyinosine 3' endonuclease	NP_418426.1
MUTM	Formamidopyrimidine DNA glycosylase	NP_418092.1
UNG	Uracil-DNA glycosylase	NP_417075.1
UVRA	DNA excision repair enzyme complex	NP_418482.1
UVRB	DNA excision repair enzyme complex	NP_415300.1
UVRC	DNA excision repair enzyme complex	NP_416423.3
DENY	Pyrimidine dimer glycosylase	NP_049733.1

[0148] Provided herein are methods where one or more enzymatic activities, such as those of repair enzymes, are used in combination to generate a site-specific single-strand nick. For example, USER (Uracil-Specific Excision Reagent; New England BioLabs) generates a single nucleoside gap at the location of a uracil. USER is a mixture of Uracil DNA glycosylase (UDG) and the DNA glycosylase-lyase Endonuclease VIII. UDG catalyzes the excision of a uracil base, forming an abasic (apyrimidinic) site while leaving the phosphodiester backbone intact. The **lyase activity of Endonuclease VIII is used to break the phosphodiester backbone at the 3' and 5' sides of the abasic site so that the base-free deoxyribose is released, creating a one nucleotide gap at the site of uracil nucleotide.**

[0149] Provided herein are methods where a nucleic acid fragment is treated prior to assembly into a target nucleic acid of predetermined sequence. In some instances, nucleic acid fragments are treated to create a sticky end, such as a sticky end with a 3' overhang or a 5' overhang. For example, uracil bases are incorporated into one or both strands of the target nucleic acids, which are chewed off upon treatment with Uracil DNA glycosylase (UDG) and Endonuclease VIII (EndoVIII). In some instances, uracil bases are incorporated near the 5' ends (or 3' ends), such as at least or at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 bases from the 5' end (or 3' end), of one or both strands. In some cases, uracil bases are incorporated near the 5' ends such as at most or at most about 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1 base from the 5' end, of one or both strands. In some cases, uracil bases are incorporated near the 5' end such as between 1-20, 2-19, 3-18, 4-17, 5-16, 6-15, 7-14, 8-13, 9-12, 10-13, 11-14 bases from the 5' end, of one or both strands. Those of skill in art will appreciate that the uracil bases may be incorporated near the 5' end such that the distance between the uracil bases and the 5' end of one or both strands may fall within a range bound by any of these values, for example from 7-19 bases.

Nucleic Acid Assembly

[0150] Provided herein are methods where two or more of the cleavage, annealing and ligation reactions are performed concurrently within the same mixture and the mixture comprises a ligase. In some cases, one or more of the various reactions is sped up and one or more of the various reactions is slowed down by adjusting the reaction conditions such as temperature. In some cases, the reaction is thermocycled between a maximum and minimum temperature to repeatedly enhance cleavage, melting, annealing, and/or ligation. In some cases, the temperature ranges from a high of 80 degrees Celsius. In some cases, the temperature ranges from a low to 4 degrees Celsius. In some cases, the temperature ranges from 4 degrees Celsius to 80 degrees Celsius. In some cases, the temperature ranges among intermediates in this range. In some cases, the temperature ranges from a high of 60 degrees Celsius. In some cases, the temperature ranges to a low of 16 degrees Celsius. In some cases, the temperature ranges from a high of 60 degrees Celsius to a low of 16 degrees Celsius. In some cases, the mixture is temperature cycled to allow for the removal of cleaved sticky ended distal fragments from precursor fragments at elevated temperatures and to allow for the annealing of the fragments with complementary sticky ends at a lower temperature. In some cases, alternative combinations or alternative temperatures are used. In yet more alternate cases the reactions occur at a single temperature. In some cases, palindromic sequences are excluded from overhangs. The number of fragment populations to anneal in a reaction varies across target nucleic acids. In some cases, a ligation reaction comprises 2, 3, 4, 5, 6, 7, 8, or more than 8 types of target fragments to be assembled. For a

given target nucleic acid, in some cases, portions of the entire nucleic acid are synthesized in separate reactions. In some cases, intermediate nucleic acids are used in a subsequent assembly round that uses the same or a different method to assemble larger intermediates or the final target nucleic acid. The same or different cleavage agents, recognition sites, and cleavage sites are used in subsequent rounds of assembly. In some instances, consecutive rounds of assembly, *e.g.* pooled or parallel assembly, are used to synthesize larger fragments in a hierarchical manner. In some cases, described herein are methods and compositions for the preparation of a target nucleic acid, wherein the target nucleic acid is a gene, using assembly of shorter fragments.

[0151] Polymerase chain reaction (PCR)-based and non-polymerase-cycling-assembly (PCA)-based strategies may be used for gene synthesis. In addition, non-PCA-based gene synthesis using different strategies and methods, including enzymatic gene synthesis, annealing and ligation reaction, simultaneous synthesis of two genes *via* a hybrid gene, shotgun ligation and co-ligation, insertion gene synthesis, gene synthesis *via* one strand of DNA, template-directed ligation, ligase chain reaction, microarray-mediated gene synthesis, Golden Gate Gene Assembly, Blue Heron solid support technology, Sloning building block technology, RNA-mediated gene assembly, the PCR-based thermodynamically balanced inside-out (TBIO) (Gao et al, 2003), two-step total gene synthesis method that combines dual asymmetrical PCR (DA-PCR) (Sandhu et al, 1992), overlap extension PCR (Young and Dong, 2004), PCR-based two-step DNA synthesis (PTDS) (Xiong et al, 2004b), successive PCR method (Xiong et al, 2005, 2006a), or any other suitable method known in the art can be used in connection with the methods and compositions described herein, for the assembly of longer polynucleotides from shorter oligonucleotides.

[0152] *Amplification*

[0153] Amplification reactions described herein can be performed by any means known in the art. In some cases, the nucleic acids are amplified by polymerase chain reaction (PCR). Other methods of nucleic acid amplification include, for example, ligase chain reaction, oligonucleotide ligations assay, and hybridization assay. DNA polymerases described herein include enzymes that have DNA polymerase activity even though it may have other activities. A single DNA polymerase or a plurality of DNA polymerases may be used throughout the repair and copying reactions. The same DNA polymerase or set of DNA polymerases may be used at different stages of the present methods or the DNA polymerases may be varied or additional polymerase added during various steps. Amplification may be achieved through any process by which the copy number of a target sequence is increased, *e.g.* PCR. Amplification can be performed at any point during a multi reaction procedure, *e.g.* before or after pooling of sequencing libraries from

independent reaction volumes and may be used to amplify any suitable target molecule described herein.

[0154] *Oligonucleic acid Synthesis*

[0155] Oligonucleic acids serving as target nucleic acids for assembly may be synthesized de novo in parallel. The oligonucleic acids may be assembled into precursor fragments which are then assembled into target nucleic acids. In some case, greater than about 100, 1000, 16,000, 50,000 or 250,000 or even greater than about 1,000,000 different oligonucleic acids are synthesized together. In some cases, these oligonucleic acids are synthesized in less than 20, 10, 5, 1, 0.1 cm², or smaller surface area. In some instances, oligonucleic acids are synthesized on a support, *e.g.* surfaces, such as microarrays, beads, miniwells, channels, or substantially planar devices. In some case, oligonucleic acids are synthesized using phosphoramidite chemistry. In order to host phosphoramidite chemistry, the surface of the oligonucleotide synthesis loci of a substrate in some instances is chemically modified to provide a proper site for the linkage of the growing nucleotide chain to the surface. Various types of surface modification chemistry exists which allow a nucleotide to attached to the substrate surface.

[0156] The DNA and RNA synthesized according to the methods described herein may be used to express proteins in vivo or in vitro. The nucleic acids may be used alone or in combination to express one or more proteins each having one or more protein activities. Such protein activities may be linked together to create a naturally occurring or non-naturally occurring metabolic/enzymatic pathway. Further, proteins with binding activity may be expressed using the nucleic acids synthesized according to the methods described herein. Such binding activity may be used to form scaffolds of varying sizes.

Computers and Software

[0157] The methods and systems described herein may comprise and/or are performed using a software program on a computer system. Accordingly, computerized control for the optimization of design algorithms described herein and the synthesis and assembly of nucleic acids are within the bounds of this disclosure. For example, supply of reagents and control of PCR reaction conditions are controlled with a computer. In some instances, a computer system is programmed to search for sticky end motifs in a user specified predetermined nucleic acid sequence, interface these motifs with a list of suitable nicking enzymes, and/or determine one or more assembly algorithms to assemble fragments defined by the sticky end motifs. In some instances, a computer system described herein accepts as an input one or more orders for one or more nucleic acids of predetermined sequence, devises an algorithm(s) for the synthesis and/or assembly of the one or more nucleic acid fragments, provides an output in the form of instructions to a peripheral device(s) for the synthesis and/or assembly of the one or more nucleic acid fragments, and/or

instructs for the production of the one or more nucleic acid fragments by the peripheral devices to form the desired nucleic acid of predetermined sequence. In some instances, a computer system operates without human intervention during one or more of steps for the production of a target nucleic acid of predetermined sequence or nucleic acid fragment thereof.

[0158] In some cases, a software system is used to identify sticky end motif sequence for use in a target sequence assembly reaction consistent with the disclosure herein. For example, in some cases, a software system is used to identify a sticky end motif using at least one, up to and including all, of the steps as follows. Given a final target sequence of length I, a desired target fragment of J, and a desired sticky end overhang length of K (for 5' ANNNNT 3'(SEQ ID NO.: 2), K=6) and a maximum desired similarity between sites of L, assembly parameters are in some instances calculated as follows. In some cases, J is about 200. In some cases, J is about 1000. In some cases, J is a number selected from about 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 100, or more than 1000. In some cases, J is a value in the range from 70-250. I/J is the number of fragments to be assembled (x). X-1 breakpoints are added along the target sequence, reflecting the number of junctions in the target sequence to be assembled. In some cases, junctions are selected at equal intervals or at approximately equal intervals throughout the target sequence.

[0159] For at least one breakpoint, the nearest breakpoint site candidate is identified, for example having ANNNNT (SEQ ID NO.: 2), or GNNNNC (SEQ ID NO.: 17). Consistent with the disclosure herein, the breakpoint has a 6 base sequence in some cases, while in other cases the junction sequence is 1, 2, 3, 4, or 5 bases, and in other cases the junction is 7, 8, 9, 10, or more than 10 bases. In some cases, the breakpoint site candidate comprises a purine at a first position, a number of bases ranging from 0 to 8 or greater, preferably 1 or greater in some cases, and a pyrimidine at a final position such that the first position purine and the final position pyrimidine are a complementary base pair (either AT or GC).

[0160] In some cases, breakpoint selection is continued for sites up to and In some cases, including each breakpoint or near each breakpoint. Site candidates are evaluated so as to reduce the presence of at least one of palindromic sequences, homopolymers, extreme GC content, and extreme AT content. Sites are assessed in light of at least one of these criteria, optionally in combination with or alternatively viewing additional criteria for site candidate evaluation. If a site is determined or calculated to have undesirable qualities, then the next site in a vicinity is subjected to a comparable evaluation. Site candidates are further evaluated for cross-site similarity, for example excluding sites that share more than L bases in common at common positions or in common sequence. In some cases, L is 2, such that the central NNNN of some selected sticky ends must not share similar bases at similar positions. In some cases, L is 2, such

that the central NNNN of some selected sticky ends must not share similar bases in similar patterns. In alternate cases, L is 3, 4, 5, 6, or greater than 6. Site candidates are evaluated individually or in combination, until a satisfactory sticky end system or group of distinct sticky ends is identified for a given assembly reaction. Alternate methods employ at least one of the steps recited above, alone or in combination with additional steps recited above or in combination with at least one step not recited above, or in combination with a plurality of steps recited above and at least one step not recited above.

[0161] A method described herein may be operably linked to a computer, either remotely or locally. In some cases, a method described herein is performed using a software program on a computer. In some cases, a system described herein comprises a software program for performing and/or analyzing a method or product of a method described herein. Accordingly, computerized control of a process step of any method described herein is envisioned.

[0162] The computer system **700** illustrated in **FIG. 7** depicts a logical apparatus that reads instructions from media **711** and/or a network port **705**, which is optionally be connected to server **709** having fixed media **712**. In some cases, a computer system, such as shown in **FIG. 7**, includes a CPU **701**, disk drive **703**, optional input devices such as keyboard **715** and/or mouse **716** and optional monitor **707**. Data communication can be achieved through the indicated communication medium to a server at a local or a remote location. Communication medium includes any means of transmitting and/or receiving data. As non-limiting examples, communication medium is a network connection, a wireless connection, and/or an internet connection. Such a connection can provide for communication over the World Wide Web. It is envisioned that data relating to the present disclosure is transmittable over such networks or connections for reception and/or review by a user **722**, as illustrated in **FIG. 7**.

[0163] A block diagram illustrating a first example architecture of a computer system **800** for use in connection with example embodiments of the disclosure is shown in **FIG. 8**. The example computer system of **FIG. 8** includes a processor **802** for processing instructions. Non-limiting examples of processors include: Intel Xeon™ processor, AMD Opteron™ processor, Samsung 32-bit RISC ARM 1176JZ(F)-S v1.0™ processor, ARM Cortex-A8 Samsung S5PC100™ processor, ARM Cortex-A8 Apple A4™ processor, Marvell PXA 930™ processor, and a functionally-equivalent processor. Multiple threads of execution can be used for parallel processing. In some instances, multiple processors or processors with multiple cores are used, whether in a single computer system, in a cluster, or distributed across systems over a network comprising a plurality of computers, cell phones, and/or personal data assistant devices.

[0164] In the computer system of **FIG. 8**, a high speed cache **804** is connected to, or incorporated in, the processor **802** to provide a high speed memory for instructions or data that

have been recently, or are frequently, used by processor **802**. The processor **802** is connected to a north bridge **806** by a processor bus **808**. The north bridge **806** is connected to random access memory (RAM) **810** by a memory bus **812** and manages access to the RAM **810** by the processor **802**. The north bridge **806** is also connected to a south bridge **814** by a chipset bus **816**. The south bridge **814** is, in turn, connected to a peripheral bus **818**. The peripheral bus is, for example, PCI, PCI-X, PCI Express, or another peripheral bus. The north bridge and south bridge are often referred to as a processor chipset and manage data transfer between the processor, RAM, and peripheral components on the peripheral bus **818**. In some alternative architectures, the functionality of the north bridge is incorporated into the processor instead of using a separate north bridge chip. In some instances, system **800** includes an accelerator card **822** attached to the peripheral bus **818**. The accelerator may include field programmable gate arrays (FPGAs) or other hardware for accelerating certain processing. For example, an accelerator is used for adaptive data restructuring or to evaluate algebraic expressions used in extended set processing.

[0165] Software and data are stored in external storage **824**, which can then be loaded into RAM **810** and/or cache **804** for use by the processor. System **800** includes an operating system for managing system resources. Non-limiting examples of operating systems include: Linux, Windows™, MACOS™, BlackBerry OSTM, iOS™, and other functionally-equivalent operating systems, as well as application software running on top of the operating system for managing data storage and optimization in accordance with example embodiments of the present disclosure. System **800** includes network interface cards (NICs) **820** and **821** connected to the peripheral bus for providing network interfaces to external storage, such as Network Attached Storage (NAS) and other computer systems that can be used for distributed parallel processing.

[0166] FIG. 9 is a diagram showing a network **900** with a plurality of computer systems **902a**, and **902b**, a plurality of cell phones and personal data assistants **902c**, and Network Attached Storage (NAS) **904a**, and **904b**. In some instances, systems **902a**, **902b**, and **902c** manage data storage and optimize data access for data stored in NAS **904a** and **904b**. A mathematical model can be used for the data and be evaluated using distributed parallel processing across computer systems **902a** and **902b**, and cell phone and personal data assistant system **902c**. Computer systems **902a** and **902b**, and cell phone and personal data assistant system **902c** can provide parallel processing for adaptive data restructuring of the data stored in NAS **904a** and **904b**.

FIG. 9 illustrates an example only, and a wide variety of other computer architectures and systems can be used in conjunction with the various embodiments of the present disclosure. For example, a blade server can be used to provide parallel processing. Processor blades can be connected through a back plane to provide parallel processing. Storage can also be connected to the back plane or as NAS through a separate network interface.

[0167] In some instances, processors maintain separate memory spaces and transmit data through network interfaces, back plane or other connectors for parallel processing by other processors. In some instances, some or all of the processors use a shared virtual address memory space.

[0168] FIG. 10 is a block diagram of a multiprocessor computer system 1000 using a shared virtual address memory space in accordance with an example embodiment. The system includes a plurality of processors 1002a-f that can access a shared memory subsystem 1004. The system incorporates a plurality of programmable hardware memory algorithm processors (MAPs) 1006a-f in the memory subsystem 1004. Each MAP 1006a-f can comprise a memory 1008a-f and one or more field programmable gate arrays (FPGAs) 1010a-f. The MAP provides a configurable functional unit and particular algorithms or portions of algorithms can be provided to the FPGAs 1010a-f for processing in close coordination with a respective processor. For example, the MAPs are used to evaluate algebraic expressions regarding a data model and to perform adaptive data restructuring in example embodiments. In this example, each MAP is globally accessible by all of the processors for these purposes. In one configuration, each MAP uses Direct Memory Access (DMA) to access an associated memory 1008a-f, allowing it to execute tasks independently of, and asynchronously from, the respective microprocessor 1002a-f. In this configuration, a MAP can feed results directly to another MAP for pipelining and parallel execution of algorithms.

[0169] The above computer architectures and systems are examples only, and a wide variety of other computer, cell phone, and personal data assistant architectures and systems can be used in connection with example embodiments, including systems using any combination of general processors, co-processors, FPGAs and other programmable logic devices, system on chips (SOCs), application specific integrated circuits (ASICs), and other processing and logic elements. In some instances, all or part of the computer system can be implemented in software or hardware. Any variety of data storage media can be used in connection with example embodiments, including random access memory, hard drives, flash memory, tape drives, disk arrays, Network Attached Storage (NAS) and other local or distributed data storage devices and systems.

[0170] The following examples are set forth to illustrate more clearly the principle and practice of embodiments disclosed herein to those skilled in the art and are not to be construed as limiting the scope of any claimed embodiments.

EXAMPLES

EXAMPLE 1: Restriction enzyme-free ligation of a gene fragment using sticky ends

Amplification with uracil-containing PCR primers

[0171] A gene of about 1kB (the "1kB Gene Construct") was selected to perform restriction enzyme-free ligation with a vector:

5'CAGCAGTTCCTCGCTCTTCTCACGACGAGTTCGACATCAACAAGCTGCGCTACCAC
AAGATCGTGCTGATGGCCGACGCCGATGTTGACGGCCAGCACATCGCAACGCTGCTG
CTCACCTGCTTTTCCGCTTCATGCCAGACCTCGTCGCCGAAGGCCACGTCTACTTGG
CACAGCCACCTTTGTACAAACTGAAGTGGCAGCGCGGAGAGCCAGGATTCGCATAC
TCCGATGAGGAGCGCGATGAGCAGCTCAACGAAGGCCTTGCCGCTGGACGCAAGAT
CAACAAGGACGACGGCATCCAGCGCTACAAGGGTCTCGGCGAGATGAACGCCAGCG
AGCTGTGGGAAACCACCATGGACCCAAGTTCGATTCTGCGCCGCGTGGACATCA
CCGATGCTCAGCGTGCTGATGAACTGTTCTCCATCTTGATGGGTGACGACGTTGTGG
CTCGCCGACGCTTCATACCCGAAATGCCAAGGATGTTTCGTTTCTCGATATCTAAA
GCGCCTTACTTAACCCGCCCTGGAATTCTGGGGGCGGGTTTTGTGATTTTTAGGGTC
AGCACTTTATAAATGCAGGCTTCTATGGCTTCAAGTTGGCCAATACGTGGGGTTGAT
TTTTTAAAACCAGACTGGCGTGCCCAAGAGCTGAACTTTCGCTAGTCATGGGCATTC
CTGGCCGGTTTTCTTGGCCTTCAAACCGGACAGGAATGCCCAAGTTAACGGAAAAACC
GAAAGAGGGGCACGCCAGTCTGGTTCTCCCAAACCTCAGGACAAATCCTGCCTCGGC
GCCTGCGAAAAGTGCCCTCTCCTAAATCGTTTCTAAGGGCTCGTCAGACCCCAGTTG
ATACAAACATACATTCTGAAAATTCAGTCGCTTAAATGGGCGCAGCGGGAAATGCTG
AAAACATACATTAATCACCGATACCCTAGGGCACGTGACCTTACTGAACCCACCACC
ACAGCCCATGTTCCACTACCTGATGGATCTTCCACTCCAGTCCAAATTTGGGCGTAC
ACTGCGAGTCCACTACGAT3' (SEQ ID NO.: 48)

[0172] The 1kB Gene Construct, which is an assembled gene fragment with heterogeneous sequence populations, was purchased as a single gBlock (Integrated DNA Technologies). The 1kB Gene Construct was amplified in a PCR reaction with uracil-containing primers. The PCR reaction components were prepared according to Table 2.

Table 2. PCR reaction mixture comprising uracil-containing primers.

10 μ L 5X HF buffer (ThermoFisher Scientific)
0.8 μ L 10mM dNTP (NEB)
1 ng template (1kB Gene Construct)
2.5 μ L forward primer (10 μ M) 5'CAGCAGT/ideoxyU/CCTCGCTCTTCT3' (SEQ ID NO.: 49; Integrated DNA Technologies)
2.5 μ L reverse primer (10 μ M) 5'ATCGTAG/ideoxyU/GGACTCGCAGTGTA3' (SEQ ID NO.: 50; Integrated DNA Technologies)
0.5 μ L Phusion-U hot start DNA polymerase (ThermoFisher Scientific, 2U/ μ E)

Water up to 50 μ L

[0173] The 1kB Gene Construct was amplified with the uracil-containing primers in a PCR reaction performed using the thermal cycling conditions described in Table 3.

Table 3. PCR reaction conditions for amplifying a gene with uracil-containing primers.

Step	Cycle
1	1 cycle: 98°C, 30sec
2	20 cycles: 98°C, 10sec; 68°C, 15sec; 72°C, 60sec
3	1 cycle: 72°C, 5min
4	Hold: 4°C

[0174] The uracil-containing PCR products were purified using Qiagen MinElute column, eluted in 10 μ L EB buffer, analyzed by electrophoresis (BioAnalyzer), and quantified on aNanoDrop to be 93 ng[^]L. The uracil-containing PCR products of the 1kB Gene Construct were incubated with a mixture of Uracil DNA glycosylase (UDG) and Endonuclease VIII to generate sticky ends. The incubation occurred at 37°C for 30 min in a reaction mixture as described in Table 4.

Table 4. Digestion reaction conditions for generating sticky ends in a uracil-containing gene.

<u>Reaction component</u>	<u>Quantity</u>
Uracil-containing PCR product	15 nM (final concentration)
10x CutSmart buffer (NEB)	10 μ L
UDG/EndoVIII (NEB or Enzymatics)	2 μ L of 1 U/ μ L
Water	Up to 94.7 μ L

Preparation of Artificial Vector

[0175] Two synthetic oligonucleotides having 3' overhangs when annealed together ("Artificial Vector") were hybridized and ligated to the digested uracil-containing 1kB Gene Construct ("Sticky-end Construct"). The first oligo ("Upper Oligo", SEQ ID NO.: 51) contains a 5' phosphate for ligation. The second oligonucleotide ("Lower Oligo", SEQ ID NO.: 52) lacks a base on the 5' end such that it leaves a nucleotide gap after hybridizing to the Sticky-end Construct with the Upper Oligo. Further, the Lower Oligo lacks a 5' phosphate to ensure that no ligation occurs at this juncture. The first six phosphate bonds on the Lower Oligo are phosphorothioated to prevent exonuclease digestion from the gap. Oligonucleic acid sequences of the Artificial Vector are shown in Table 5. An asterisk denotes a phosphorothioate bond.

Table 5. Sequence identities of an artificial vector for ligation to a sticky-end gene product.

Sequence ID	Sequence
SEQ ID NO.: 51	5'/5phos/TACGCTCTTCCTCAGCAGTGGTCATCGTAGT3'
SEQ ID NO.: 52	5'A*C*C*A*C*T*GCTGAGGAAGAGCGTACAGCAGTT3'
Artificial Vector SEQ ID NO.: 79	TACGCTCTTCCTCAGCA G T G G T CATCGTAGT TTGACGACATGCGAGAAGGAGTCGT*C*A*C*C*A*

[0176] The Sticky-end Construct was mixed with Upper Oligo and Lower Oligo (5 µM each) in 1x CutSmart buffer (NEB). The mixture was heated to 95 °C for 5 min, and then slowly cooled to anneal. The annealed product comprised a circularized gene construct comprising the lkb Gene Construct. This construct was generated without the remnants of any restriction enzyme cleavage sites and thus lacked any associated enzymatic "scars."

EXAMPLE 2: Assembly of LacZ gene into a plasmid

[0177] A LacZ gene was assembled into a 5 kb plasmid from three precursor LacZ fragments and 1 precursor plasmid fragment. Assembly was performed using 9 different reaction conditions.

[0178] Preparation of precursor plasmid fragments

[0179] A 5 kb plasmid was amplified with two different sets of primers for introducing a sticky end motif comprising a non-canonical base (SEQ ID NO.: 53): set A (SEQ ID NOs.: 54 and 55) and set B (SEQ ID NOs.: 56 and 57), shown in Table 6, to produce plasmid precursor fragments A and B, respectively.

Table 6. Sequence identities of plasmid primers.

Sequence identity	Primer name	Sequence
SEQ ID NO.: 54	plasmid-Fa	TGATCGGCAATGATATG/ideoxyU/CTGGAAAGAACATGTG
SEQ ID NO.: 55	plasmid-Ra	TGATCGGCAATGATGGC/ideoxyU/TATAATGCGACAAACAACAG
SEQ ID NO.: 56	plasmid-Fb	TGATCGGCAATGATATG/ideoxyU/CGCTGGAAAGAACATG
SEQ ID	plasmid-Ra	TGATCGGCAATGATGGC/ideoxyU/CGTATAATGCGACAAACA

NO.: 57		AC
---------	--	----

[0180] Each primer set comprises, in 5' to 3' order: 6 adaptor bases (TGATCG, SEQ ID NO.: 58), a first nicking enzyme recognition site (GCAATG, SEQ ID NO.: 59), a sticky end motif comprising a non-canonical base (ANNNU, SEQ ID NO.: 53), and plasmid sequence. The first two bases of the plasmid sequence in the forward and reverse primers of set B are a CG. These two bases are absent from the forward and reverse primers of set A. Two plasmid fragments, plasmid A and plasmid B, were amplified using primer set A and primer B, respectively. The composition of the amplification reaction is shown in Table 7. The amplification reaction conditions are shown in Table 8.

Table 7. PCR reaction mixture for amplification of a 5 kb plasmid.

PCR component	Quantity (μL)	Concentration in mixture
Phusion U (2 U/ μL)	1	1 U/ 50 μL
5x Phusion HF buffer	20	1x
10 mM dNTP	4	400 μM
Plasmid template (50 pg/ μL)	4	100 pg/50 μL
plasmid-Fa or plasmid-Fb (200 μM)	0.25	0.5 μM
plasmid-Ra or plasmid-Rb (200 μM)	0.25	0.5 μM
Water	70.5	

Table 8. PCR reaction conditions for amplification of a 5 kb plasmid.

Step	Cycle
1	1 cycle: 98°C, 30 sec
2	30 cycles: 98°C, 10 sec; 49°C, 15 sec; 72°C, 90 sec
3	1 cycle: 72°C, 5 min
4	Hold: 4°C, 15 – 30 sec per kb

[0181] The precursor plasmid fragment was treated with DpnI, denatured and purified.

[0182] *Preparation of precursor LacZ fragments*

[0183] The LacZ sequence was analyzed to identify two sticky end motifs which partition the sequence into roughly 3, 1 kb fragments: LacZ fragments 1-3. Sequence identities of the two sticky end motifs and the LacZ fragments are shown in Table 9. SEQ ID NO.: 60 shows the

complete LacZ gene, wherein motifs are italicized, fragment 1 is underlined with a single line, fragment 2 is underlined with a squiggly line, and fragment 3 is underlined with a double line.

Table 9. Sequence identities of LacZ fragments and sticky end motifs.

Sequence identity	Sequence name	Sequence
SEQ ID NO.: 61	fragment 1	<u>ATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACG</u> <u>TCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCC</u> <u>TTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAA</u> <u>GAGGCCCGCACCGATCGCCCTTCCAACAGTTGCGCAGCCT</u> <u>GAATGGCGAATGGCGCTTTCCTGGTTCGGCACCAGAAG</u> <u>CGGTGCCGAAAGCTGGCTGGAGTGCATCTTCTGAGGCC</u> <u>GATACTGTCGTCGTCCTCAAACCTGGCAGATGCACGGTTA</u> <u>CGATGCGCCATCTACACCAACGTGACCTATCCCATTACGG</u> <u>TCAATCCGCCGTTTGTTCACGGAGAATCCGACGGGTTGT</u> <u>TACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGA</u> <u>AGGCCAGACGCGAATTATTTTTGATGGCGTTAACTCGGCGT</u> <u>TTCATCTGTGGTGCAACGGGCGCTGGGTTCGGTACGGCCAG</u> <u>GACAGTCGTTTGCCGTCTGAATTTGACCTGAGCGCATTTTT</u> <u>ACGCGCCGGAGAAAACCGCCTCGCGGTGATGGTGCTGCGC</u> <u>TGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGC</u> <u>GGATGAGCGGCATTTCCGTGACGTCCTGTTGCTGCATAAA</u> <u>CCGACTACACAAATCAGCGATTTCCATGTTGCCACTCGCTT</u> <u>TAATGATGATTTACGCCGCGCTGTACTGGAGGCTGAAGTTC</u> <u>AGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTT</u> <u>TCTTTATGGCAGGGTCAAACGCAGGTGCGCCAGCGGCACCG</u> <u>CGCCTTTCGGCGGTGAAATTATCGATGAGCGTGGTGGTTAT</u> <u>GCCGATCGCGTCACACTACGTCTGAACGTGAAAACCCGA</u> <u>AACTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTG</u> <u>GTTGAACTGCACACCGCCGACGGCACGCTGATTGAAGCAG</u> <u>AAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAA</u>
SEQ ID NO.: 62	fragment 2	<u>CTGCTGCTGCTGAACGGCAAGCCGTTGCTGATTCGAGGCGT</u> <u>TAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCATGG</u> <u>ATGAGCAGACGATGGTGCAGGATATCCTGCTGATGAAGCA</u> <u>GAACAACTTTAACGCCGTGCGCTGTTTCGCATTATCCGAACC</u> <u>ATCCGCTGTGGTACACGCTGTGCGACCCTACGGCCTGTAT</u> <u>GTGGTGGATGAAGCCAAATTGAAACCCACGGCATGGTGC</u> <u>CAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCG</u> <u>ATGAGCGAACGCGTAACGCGAATGGTGCAGCGCGATCGTA</u> <u>ATCACCCGAGTGTGATCATCTGGTCGCTGGGGAATGAATCA</u> <u>GGCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAA</u> <u>ATCTGTCGATCCTTCCCGCCCGGTGCAGTATGAAGGCGGCG</u> <u>GAGCCGACACCACGGCCACCGATATTATTTGCCGATGTAC</u> <u>GCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTGTGCCGA</u> <u>AATGGTCCATCAAAAAATGGCTTTCGCTACCTGGAGAGACG</u> <u>CGCCCGCTGATCCTTTGCGAATACGCCACGCGATGGGTAA</u> <u>CAGTCTTGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTC</u> <u>AGTATCCCGTTTACAGGGCGGCTTCGTCTGGGACTGGGTG</u> <u>GATCAGTCGCTGATTAATATGATGAAAACGGCAACCCGT</u> <u>GGTCGGCTTACGGCGGTGATTTGGCGATACGCCGAACGAT</u> <u>CGCCAGTTCTGTATGAACGGTCTGGTCTTTGCCGACCGCAC</u>

		<p>GCCGCATCCAGCGCTGACGGAAGCAAAACACCAGCAGCAG TTTTCAGrTCGTATATCCGGGCAAACCATCGAAGTGAC CAGCGAATACCTGTTC^ ^CATAGCGATAACGAGCTCCTGC ACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAGCGG TAA TGCCT T GATGT GCTCCACAAGGTAA</p>
SEQ ID NO.: 63	fragment 3	<p>GATTGAACTGCCTGAACTACCGCAGCCGGAGAGCGCCGGG CAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCGA CCGCATGGTCAGAAGCCGGGCACATCAGCGCCTGGCAGCA GTGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCCCCGCCG CGTCCCACGCCATCCCGCATCTGACCACCAGCGAAATGGAT TTTTGCATCGAGCTGGGTAATAAGCGTTGGCAATTTAACCG CCAGTCAGGCTTTCTTTCACAGATGTGGATTGGCGATAAAA ACAA CTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCA CCGCTGGATAACGACATTGGCGTAAGTGAAGCGACCCGCA TTGACCCTAACGCCTGGGTCGAACGCTGGAAGGCGGGCGGG CCATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCACGGCA GATACTTGCTGATGCGGTGCTGATTACGACCCTCACGC GTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAA ACCTACCGGATTGATGGTAGTGGTCAAATGGCGATTACCGT TGATGTTGAAGTGGCGAGCGATACACCGCATCCGGCGCGG ATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGG TAAACTGGCTCGGATTAGGGCCGCAAGAAA ACTATCCCGA CCGCCTTACTGCCGCTGTTTTGACCGCTGGGATCTGCCATT GTCAGACATGTATAACCCCGTACGTCTTCCCGAGCGAAAACG GTCTGCGCTGCGGGACGCGCAATTGAATTATGGCCACAC CAGTGGCGCGGCGACTTCCAGTTCAACATCAGCCGCTACAG TCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGC ACGCGGAAGAAGGCACATGGCTGAATATCGACGGTTTCCA TATGGGGATTGGTGGCGACGACTCCTGGAGCCCGTCAGTAT CGGCGGAATTCCAGCTGAGCGCCGGTCGCTACCATTACCAG TTGGTCTGGTGTCAAAAATAA</p>
SEQ ID NO.: 64	motif 1	AATGGT
SEQ ID NO.: 65	motif 2	ACAGTT
SEQ ID NO.: 60	LacZ	<p>AGCCA7ATGACCATGATTACGGATTCACTGGCCGTCGTTTTA CAACGTCGTGACTGGGAAAACCTGGCGTTACCCA ACTTAA TCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATA GCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGC AGCCTGAATGGCGAATGGCGCTTTCCTGGTTTCCGGCACC AGAAGCGGTGCCGGAAGCTGGCTGGAGTGCATCTTCCCT GAGGCCGATACTGTCGTCGTCCTTCAA ACTGGCAGATGCA CGGTTACGATGCGCCATCTACACCAACGTGACCTATCCCA TTACGGTCAATCCGCCGTTTGTTCACGGAGAATCCGACG GGTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCT ACAGGAAGGCCAGACGCGAATTATTTTTGATGGCGTTAACT CGGCGTTTCATCTGTGGTGAACGGGCGCTGGGTGCGTTAC GGCCAGGACAGTCGTTTGCCGTCTGAATTTGACCTGAGCGC ATTTTTACGCGCCGGAGAAAACCGCCTCGCGGTGATGGTGC TCGCTGGAGTGACGGCAGTTATCTGGAAGATCAGGATAT</p>

	<p>GTGGCGGATGAGCGGCATTTTCCGTGACGTCTCGTTGCTGC ATAAACCGACTACACAAATCAGCGATTTCCATGTTGCCACT CGCTTTAATGATGATTTTCCAGCCGCGCTGTACTGGAGGCTGA AGTTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAA CAGTTTCTTTATGGCAGGGTGAAACGCAGGTCGCCAGCGGC ACCGCGCCTTTCGGCGGTGAAATTATCGATGAGCGTGGTGG TTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACC CGAAACTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCG GTGGTTGAACTGCACACCGCCGACGGCACGCTGATTGAAG CAGAAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAA AArGGrCTCCTCCTCAACGG CAAGCCGTTGCTGATTCG AGGCGTTAACCGTCACGAGCATCATCCTCTGCATGGTCAGG ICATCGATCAGCAGACGA^ }GTGCAGGATATCCTGCTGATG AAGCA^AACAACIIXAACGCCCTOCGC^ }TTCGATTATCC GAACCATCOGCT^GGXACAOGC^^ GCGACCGCTACGGCC TGTATGTGGTGGATGAAGCCAATATTGAAACCCACGGCATG GTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACC GGCGATGAGCGAACGCGTAACGCGAATGGTGCAGCGCGAT CGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGAATGA ATCAGGCCACGGCGCTAATCACGACGCGCTGTATCGCTGGA T AAAT T T GAT CTCCCGCCCGGTGCAGTATGAAGGC GGCGGAGCCGACA£CAOggccaccga^ }TATTTGCCCGAT GTACGCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTGTGC CGAAATGGTCCATCAAAAAATGGCTTTCGCTACCTGGAGAG ACGCGC^CGCTGATCCTT^ }CGAATACGCCACGCGATGGG IAACAGTCTTGGCGGTTTCGCTAAATACTGGCAGGCGTTTC GT AGTAT CCCGTTTACAGGGCGGCTTCGTCTGGGACTGG T GGAT AGTCGT ATTAATATGATGAAAACGGCAACC CGTGGTCCGGCTTACGGCGGTGATTTTGGCGATACGCCGAAC GATCGCCAGTTCTGTATGAACGGTCTGGTCTTTGCCGACCG CACGCCGCATCCAGCGCTGACGGAAGCAAAACACCAGCAG CAGTTTTTCCAGTTCCGTTTATCCGGGCAAACCATCGAAGT GACCA GCGAATA CCTGIT CCGTC ATAGCGATAACGAGCTCC TGCACTCGATWrGGOGCTGGA^ }GTAAGCCGCTGGCAAG CGGTGAAGTCCCTCTGGAT^GI^}CTCCACAAGGTAAACAGT 7OATTGAACTGCCTGAACTACCGCAGCCGGAGAGCGCCGG GCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCG ACCGCATGGTCAGAAGCCGGGCACATCAGCGCCTGGCAGC AGTGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCCCGCC GCGTCCCACGCCATCCCGCATCTGACCACCAGCGAAATGGA TTTTTGCATCGAGCTGGGTAAATAAGCGTTGGCAATTTAACC GCCAGTCAGGCTTTCTTTTACAGATGTGGATTGGCGATAAA AAACAACCTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGC ACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACCCGC ATTGACCCTAACCGCTGGGTGCAACGCTGGAAGGCGGCGG GCCATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCACGGC AGATACACTTGCTGATGCGGTGCTGATTACGACCCTCACG CGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAA AACCTACCGGATTGATGGTAGTGGTCAAATGGCGATTACCG TTGATGTTGAAGTGGCGAGCGATACACCGCATCCGGCGCG GATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGG GTAAACTGGCTCGGATTAGGGCCGCAAGAAAACCTATCCCG</p>
--	---

		<p><u>ACCGCCTTACTGCCGCTGTTTTGACCGCTGGGATCTGCCA</u> <u>TTGTCAGACATGTATACCCCGTACGTCTTCCCGAGCGAAAA</u> <u>CGGTCTGCGCTGCGGGACGCGGAATTGAATTATGGCCAC</u> <u>ACCAGTGGCGGGCGACTTCCAGTTCAACATCAGCCGCTAC</u> <u>AGTCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCT</u> <u>GCACGCGGAAGAAGGCACATGGCTGAATATCGACGGTTTC</u> <u>CATATGGGGATTGGTGGCGACGACTCCTGGAGCCCGTCAGT</u> <u>ATCGGCGGAATTCCAGCTGAGCGCCGGTCGCTACCATTACC</u> <u>AGTTGGTCTGGTGTCAAATAAA4 TA TGT</u></p>
--	--	---

[0184] LacZ fragments 1-3 were assembled from smaller, synthesized oligonucleic acids. During fragment preparation, the 5' and/or 3' of each fragment end was appended with a connecting adaptor to generated adaptor-modified fragments 1-3. To prepare LacZ for assembly with the precursor plasmid fragments, the 5' end of fragment 1 and the 3' end of fragment 3 were appended with a first outer adaptor comprising outer adaptor motif 1 (AGCCAT, SEQ ID NO.: 66) and a second outer adaptor comprising outer adaptor motif 2 (TTATGT, SEQ ID NO.: 67), respectively. The sequences of modified fragments 1-3 are shown in Table 10. Each modified fragment comprises a first adaptor sequence (GTATGCTGACTGCT, SEQ ID NO. : 68) at the first end and second adaptor sequence (TTGCCCTACGGTCT, SEQ ID NO.: 69) at the second end, indicated by a dashed underline. Each modified fragment comprises a nicking enzyme recognition site (GCAATG, SEQ ID NO.: 59), indicated by a dotted underline. Each modified fragment comprises an ANNNNT motif (SEQ ID NO. : 2), indicated by italics.

Table 10. Sequence identities of modified LacZ fragments.

Sequence identity	Sequence name	Sequence
SEQ ID NO.: 70	modified fragment 1	<p><u>GGACTTGTATGCTGACTGCTGCAATGAGCCATATGACCATGA</u> <u>TTACGGATTCAGTGGCCGTCGTTTTACAACGTCGTGACTGGG</u> <u>AAAACCCTGGCGTTACCCAACCTTAATCGCCTTGCAGCACATC</u> <u>CCCCTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCG</u> <u>ATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGC</u> <u>GCTTTGCCTGGTTTCCGGCACAGAAAGCGGTGCCGGAAGCT</u> <u>GGCTGGAGTGCGATCTTCCTGAGGCCGATACTGTCGTCGTCC</u> <u>CCTCAAACCTGGCAGATGCACGGTTACGATGCGCCCATCTACA</u> <u>CCAACGTGACCTATCCCATTACGGTCAATCCGCCGTTTGTTC</u> <u>CACGGAGAATCCGACGGGTTGTTACTCGCTCACATTTAATGT</u> <u>TGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTT</u> <u>TGATGGCGTTAACTCGGCGTTTCATCTGTGGTGAACGGGCG</u> <u>CTGGGTCGGTTACGGCCAGGACAGTCGTTTGCCGTCTGAATT</u> <u>TGACCTGAGCGCATTTTACGCGCCGGAGAAAACCGCCTCGC</u> <u>GGTGATGGTGCTGCGCTGGAGTGACGGCAGTTATCTGGAAGA</u> <u>TCAGGATATGTGGCGGATGAGCGGCATTTTCCGTGACGTCTC</u> <u>GTTGCTGCATAAACCGACTACACAAATCAGCGATTTCCATGT</u> <u>TGCCACTCGCTTTAATGATGATTTTACGCCGCGCTGACTGGA</u> <u>GGCTGAAGTTCAGATGTGCGGCGAGTTGCGTGACTACCTACG</u></p>

		<p>GGTAACAGTTTCTTTATGGCAGGGTCAAACGCAGGTCGCCAG CGGCACCGCGCCTTTCGGCGGTGAAATTATCGATGAGCGTGG TGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAA CCCGAAACTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGC GGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTGAAGC AGAAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAA4 rGGrCATTGCAGACCGTAGGGCAATGATTC</p>
<p>SEQ ID NO.: 71</p>	<p>modified fragment 2</p>	<p>GGACTTGTATGCTGACTGCTGCAATGWGGrCTGCTGCTGCT GAACGGCAAGCCGTTGCTGATTCGAGGCGTTAACCGTCACGA GCATCATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGAT GGTGCAGGATATCCTGCTGATGAAGCAGAACAACCTTAAACGC CGTGCGCTGTTCCGATTATCCGAACCATCCGCTGTGGTACAC GCTGTGCGACCGCTACGGCCTGTATGTGGTGGATGAAGCCAA TATTGAAACCCACGGCATGGTGCCAATGAATCGTCTGACCGA TGATCCGCGCTGGCTACCGGCGATGAGCGAACGCGTAACGC GAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGATCATCT GGTCGCTGGGGAATGAATCAGGCCACGGCGCTAATCACGAC GCGCTGTATCGCTGGATCAAATCTGTGATCCTTCCCGCCCG GTGCAGTATGAAGGCGGCGGAGCCGACACCACGGCCACCGA TATTATTTGCCCGATGTACGCGCGCGTGGATGAAGACCAGCC CTTCCCGGCTGTGCCGAAATGGTCCATCAAAAATGGCTTTC GCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGC CCACGCGATGGGTAACAGTCTTGGCGGTTTCGCTAAATACTG GCAGGCGTTTCGTCAGTATCCCCGTTTACAGGGGCGGCTTCGT CTGGGACTGGGTGGATCAGTCGCTGATTAATATGATGAAAA CGGCAACCCGTGGTCCGCTTACGGCGGTGATTTTGGCGATAC GCCGAACGATCGCCAGTTCTGTATGAACGGTCTGGTCTTTGC CGACCGCACGCCGATCCAGCGCTGACGGAAGCAAAACACC AGCAGCAGTTTTTCCAGTTCGTTTATCCGGGCAAACCATCG AAGTGACCAGCGAATACTGTTCCGTCATAGCGATAACGAGC TCCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAA GCGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAA^C^Gr TCATTGCAGACCGTAGGGCAATGATTC</p>
<p>SEQ ID NO.: 72</p>	<p>modified fragment 3</p>	<p>GGACTTGTATGCTGACTGCTGCAATGUC4G7TGATTGAACTGC CTGAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTC ACAGTACGCGTAGTGCAACCGAACGCGACCGCATGGTCAGA AGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGG AAAACCTCAGTGTGACGCTCCCCGCGCGTCCCACGCCATCC CGCATCTGACCACCAGCGAAATGGATTTTGCATCGAGCTGG GTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTT CACAGATGTGGATTGGCGATAAAAAACAACCTGCTGACGCCG CTGCGCGATCAGTTCACCCGTGCACCGCTGGATAACGACATT GGCGTAAGTGAAGCGACCCGCATTGACCCTAACGCCTGGGTC GAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAAGCAGC GTTGTTCAGTGCACGGCAGATACTTGTGATGCGGTGCT GATTACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCTT ATTTATCAGCCGGAACCTACCGGATTGATGGTAGTGGTCA AATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATAACCC GCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGGT AGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAA ACTATCCCGACCGCCTTACTGCCGCTGTTTTGACCGCTGGG ATCTGCCATTGTCAGACATGTATACCCCGTACGTCTTCCCGA</p>

		GCGAAAACGGTCTGCGCTGCGGGACGCGCGAATTGAATTAT GGCCCACACCAGTGGCGCGGCGACTTCCAGTTCAACATCAGC CGCTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCAT CTGCTGCACGCGGAAGAAGGCACATGGCTGAATATCGACGG TTTCATATGGGGATTGGTGGC GACGACTCCTGGAGCCCCTC AGTATCGGCGGAATTCCAGCTGAGCGCCGGTCGCTACCATTA CCAGTTGGTCTGGTGTCAAATAA^ TA TGTCATTGCAGACCG <u>TAGGGCAATGATTC</u>
--	--	--

[0185] To generate a second nicking enzyme recognition site, a non-canonical base uracil, each modified fragment was amplified using the universal primers shown in Table 11. An asterisk indicates a phosphorothioated bond.

Table 11. Uracil-containing universal primers for amplification of modified LacZ fragments.

Sequence identity	Sequence name	Sequence
SEQ ID NO.: 73	modfrag1F	GTATGCTGACTGCTGCAATGAGCCA*/3deoxyU/
SEQ ID NO.: 74	modfrag1R	TTGCCCTACGGTCTGCAATGACCAT*/3deoxyU/
SEQ ID NO.: 75	modfrag2F	GTATGCTGACTGCTGCAATGAATGG*/3deoxyU/
SEQ ID NO.: 76	modfrag2R	TTGCCCTACGGTCTGCAATGAACTG*/3deoxyU/
SEQ ID NO.: 77	modfrag3F	GTATGCTGACTGCTGCAATGACAGT*/3deoxyU/
SEQ ID NO.: 78	modfrag3R	TTGCCCTACGGTCTGCAATGACATA*/3deoxyU/

[0186] Each primer set comprises, in 5' to 3' order: adaptor sequence, a first nicking enzyme recognition site (GCAATG, SEQ ID NO.: 59), and a sticky end motif comprising a non-canonical base (ANNNU, SEQ ID NO.: 53). Modified fragments 1-3 were amplified using their corresponding primers modfrag1F/modfrag1R, modfrag2F/modfrag2R and modfrag3F/modfrag3R, respectively. The composition of the amplification reaction is shown in Table 12. The amplification reaction conditions are shown in Table 13.

Table 12. PCR reaction mixture for amplification of modified LacZ fragments.

PCR component	Quantity (μ L)	Concentration in mixture
Phusion U (2 U/ μ L)	1	1 U/ 50 μ L
5x Phusion HF buffer	20	1x
10 mM dNTP	2	200 μ M
Plasmid template (50 pg/ μ L)	2	100 pg/100 μ L
Forward primer (200 μ M)	0.25	0.5 μ M
Reverse primer (200 μ M)	0.25	0.5 μ M
Water	70.5	

Table 13. PCR reaction conditions for amplification of modified LacZ fragments

Step	Cycle
1	1 cycle: 98°C, 30 sec
2	20 cycles: 98°C, 10 sec; 72°C, 30 sec
3	10 cycles: 98°C, 10 sec; 72°C, 45 sec
4	1 cycle: 72°C, 5 min
5	Hold: 4°C, 15 – 30 sec per kb

[0187] *Assembly of LacZ precursor fragments*

[0188] LacZ precursor fragments were annealed and ligated with the plasmid fragment according to reactions 1 and 2 under conditions A-I shown in Table 14. The nicking enzyme Nb.BsrDI was used to generate a nick adjacent to the nicking recognition site (GCAATG, SEQ ID NO.: 59) on one strand during reaction 1. USER (UDG and endonuclease VIII) was used to generate a nick at uracil in a second strand during reaction 2. Reaction 2 comprised three steps: cleavage of uracil, ligation, and enzymatic inactivation. Assembled fragments comprise LacZ inserted into the 5 kb plasmid. To determine efficiency of assembly into the plasmid, PCR of colonies resulting from the transformation of assembled plasmids into *E. coli* were amplified using plasmid-specific primers. Amplification products from 10 colonies of conditions A-I were amplified by colony PCR. The number amplicons with the correct size insert (about 3 kb), as identified by gel electrophoresis, are shown in Table 14. **FIG. 11** shows an image of a gel electrophoresis of LacZ amplified inserts generated from assembly conditions A-I.

Table 14. LacZ fragment efficiency assembly analysis.

Condition	Precursor fragments	Reaction 1	Reaction 2	Predicted insert size confirmed by electrophoresis
A	LacZ precursor fragments 1-3; plasmid precursor fragment A	Incubate fragments with nicking enzyme Nb.BsrDI and buffer at 65°C for 60 min	Incubate reaction 1 with USER, ATP, T7 ligase, and buffer at 37°C for 30 min, 16°C for 60 min, and 80°C for 20 min	4/10
B	LacZ precursor fragments 1-3; plasmid precursor fragment A	Incubate fragments with nicking enzyme Nb.BsrDI and buffer at 65°C for 60 min	Incubate reaction 1 with USER, ATP, T7 ligase, and buffer at 37°C for 30 min, 16°C for 60 min, and 80°C for 20 min	9/10
C	LacZ precursor fragments 1-3; plasmid precursor fragment A	Incubate fragments with nicking enzyme Nb.BsrDI and buffer at 65°C for 60 min	Incubate reaction 1 with USER, ATP, T7 ligase, and buffer at 37°C for 30 min, 16°C for 60 min, and 80°C for 20 min	8/10
D	LacZ precursor	Incubate fragments with	Incubate reaction 1 with	10/10

	fragments 1-3; plasmid precursor fragment A	nicking enzyme Nb.BsrDI and buffer at 65°C for 60 min	USER, ATP, T7 ligase, and buffer at 60°C for 30 min, 20 cycles of 37°C for 1 min and 16°C for 3 min, 80°C for 20 min, 4°C hold	
E	LacZ precursor fragments 1-3; plasmid precursor fragment A	Incubate fragments with nicking enzyme Nb.BsrDI and buffer at 65°C for 60 min	Incubate reaction 1 with USER, ATP, T7 ligase, and buffer at 60°C for 30 min, 20 cycles of 37°C for 1 min and 16°C for 3 min, 80°C for 20 min, 4°C hold	7/10
F	LacZ precursor fragments 1-3; plasmid precursor fragment A	Incubate fragments with nicking enzyme Nb.BsrDI and buffer at 65°C for 60 min	Incubate reaction 1 with USER, ATP, T7 ligase, and buffer at 60°C for 30 min, 20 cycles of 37°C for 1 min and 16°C for 3 min, 80°C for 20 min, 4°C hold	9/10
G	LacZ precursor fragments 1-3; plasmid precursor fragment B	Incubate fragments with nicking enzyme Nb.BsrDI, ATP and buffer at 65°C for 60 min	Incubate reaction 1 with USER, T7 ligase, and buffer at 37°C for 60 min, 16°C for 60 min, and 80°C for 20 min	0/10
H	plasmid precursor fragment A	Incubate fragments with nicking enzyme Nb.BsrDI and buffer at 65°C for 60 min	Incubate reaction 1 with USER, ATP, T7 ligase, and buffer at 37°C for 60 min, 16°C for 60 min, and 80°C for 20 min	0/4
I	plasmid precursor fragment B	Incubate fragments with nicking enzyme Nb.BsrDI and buffer at 65°C for 60 min	Incubate reaction 1 with USER, ATP, T7 ligase, and buffer at 37°C for 60 min, 16°C for 60 min, and 80°C for 20 min	0/4

EXAMPLE 3: Recombinatorial target nucleic acid library

[0189] An enzyme of interest having an activity to be improved is selected. Specific amino acid residues relevant to enzyme activity and stability are identified. The nucleic acid sequence encoding the enzyme is obtained. Bases corresponding to the specific amino acid residues are identified, and the nucleic acid is partitioned into fragments such that each fragment spans a single base position corresponding to a specific amino acid residue.

[0190] Target nucleic acid fragments are synthesized such that identified bases corresponding to the specific amino acid residues are indeterminate. Target nucleic acid fragments are amplified using a uridine primer and treated with a sequence adjacent nick enzyme and a uridine-specific nick enzyme. Cleaved end sequence is removed and target nucleic acid fragments are assembled to generate a target nucleic acid library. Aliquots of the library are sequenced to confirm success of the assembly, and aliquoted molecules of the library are individually cloned and transformed into a host cell for expression. Expressed enzymes are isolated and assayed for activity and stability.

[0191] Enzymes having increased stability due to single point mutations are identified. Enzymes having increased activity due to single point mutations are identified. Also identified are

enzymes having increased stability and/or activity due to combinations of point mutations, each of which individually is detrimental to enzyme activity or stability, and which would be unlikely to be pursued by more traditional, 'one mutation at a time' approaches.

EXAMPLE 4: De novo generation of a target nucleic acid

[0192] A 3 kb double-stranded target gene of predetermined sequence is prepared using a de novo synthesis and assembly method described herein. The predetermined gene sequence is first analyzed to identify fragments which will be synthesized and assembled into the final gene product.

[0193] *Determination of gene/fragment sequences*

[0194] The target nucleic acid sequence is analyzed to identify sticky end motifs having an ANNNNT sequence (SEQ ID NO. : 2). Two of the identified motifs are selected according to their position in the sequence, so that the first identified motif is located at roughly 1 kb and the second identified motif is located at roughly 2 kb. The two selected motifs thus partition the target sequence into three, approximately 1 kb precursor fragments, denoted fragments 1, 2 and 3.

[0195] *De novo synthesis of precursor/fragments*

[0196] Fragments 1, 2 and 3 are prepared by de novo synthesis and PCA assembly of oligonucleic acids. During this process, outer adaptor sequences are added to the 5' end of fragment 1 and the 3' end of fragment 3, and connecting adaptor sequences are added to the 3' end of fragment 1, the 5' and 3' ends of fragment 2, and the 5' end of fragment 3. The connecting adaptor sequences located at the 3' end of fragment 1 and the 5' end of fragment 2 comprise the sequence of the first identified ANNNNT motif (SEQ ID NO.: 2). The connecting adaptor sequences located at the 3' end of fragment 2 and the 5' end of fragment 3 comprise the sequence of the second identified ANNNNT motif (SEQ ID NO.: 2). Each connecting adaptor comprises, in order: a sequence of 1-10 bases (adaptor bases), a first nicking enzyme recognition site comprising a first nicking enzyme cleavage site on one strand, and a sticky end motif. The adaptor bases and first nicking enzyme cleavage site comprise the same bases for each connecting adaptor.

[0197] Fragment 1 prepared with adaptor sequence comprises, in 5' to 3' order: a first outer adaptor sequence; fragment 1 sequence; and a first connecting adaptor sequence comprising, in 5' to 3' order, the first ANNNNT motif (SEQ ID NO.: 2), the first nicking enzyme recognition site comprising the first nicking enzyme cleavage site on a first strand, and the sequence of adaptor bases. Fragment 2 prepared with adaptor sequence comprises, in 5' to 3' order: the first connecting adaptor sequence comprising, in 5' to 3' order, the sequence of adaptor bases, the first nicking enzyme recognition site comprising the first nicking enzyme cleavage site on a second

strand, and the first ANNNNT motif (SEQ ID NO.: 2); fragment 2 sequence; and a second connecting adaptor sequence comprising, in 5' to 3' order, the second ANNNNT motif (SEQ ID NO.: 2), the first nicking enzyme recognition site comprising the first nicking enzyme cleavage site on a first strand, and the sequence of adaptor bases. Fragment 3 prepared with adaptor sequence comprises, in 5' to 3' order: the second connecting adaptor sequence comprising, in 5' to 3' order, the sequence of adaptor bases, the first nicking enzyme recognition site comprising the first nicking enzyme cleavage site on a second strand, the second ANNNNT motif (SEQ ID NO.: 2); fragment 3 sequence; and a second outer adaptor sequence.

[0198] *Generation of fragments with two nicking enzyme cleavage sites*

[0199] Each of the prepared fragments are amplified to incorporate a second nicking enzyme cleavage site on a single-strand of each fragment such that the second nicking enzyme cleavage site is located from 1 to 10 bases away from the first nicking enzyme cleavage site of each fragment and on a different strand from the first nicking enzyme cleavage site. The second nicking enzyme cleavage site comprises a non-canonical base. The non-canonical base is added to each fragment during PCR via a primer comprising the sequence of adaptor bases, the first nicking enzyme recognition site, a sticky end motif ANNNNT (SEQ ID NO.: 2), and the non-canonical base.

[0200] Fragment 1 comprises, in 5' to 3' order: the first outer adaptor sequence, fragment 1 sequence, the non-canonical base on the second strand, the first ANNNNT motif (SEQ ID NO.: 2), the first nicking enzyme recognition site comprising the first nicking enzyme cleavage site on the first strand, and the sequence of adaptor bases. Fragment 2 comprises, in 5' to 3' order: the sequence of adaptor bases, the first nicking enzyme recognition site comprising the first nicking enzyme cleavage site on the second strand, the first ANNNNT motif (SEQ ID NO.: 2), the non-canonical base on the first strand, fragment 2 sequence, the non-canonical base on the second strand, the ANNNNT motif (SEQ ID NO.: 2), the first nicking enzyme recognition site comprising the first nicking enzyme cleavage site on the first strand, and the sequence of adaptor bases. Fragment 3 comprises, in 5' to 3' order: the sequence of adaptor bases, the first nicking enzyme recognition site comprising the first nicking enzyme cleavage site on the second strand, the second ANNNNT motif (SEQ ID NO.: 2), the non-canonical base on the first strand, fragment 3 sequence, and a second outer adaptor sequence.

[0201] *Cleavage of fragments with two nicking enzymes*

[0202] Each of the three fragments comprising two nicking enzyme cleavage sites are treated with a first nicking enzyme and a second nicking enzyme. The first nicking enzyme creates a nick at the first nicking enzyme cleavage site by cleaving a single-strand of the fragment. The second nicking enzyme creates a nick by removing the non-canonical base from the fragment.

The enzyme-treated fragments have an overhang comprising a sticky end motif ANNNNT (SEQ ID NO.: 2).

[0203] Enzyme-treated fragment 1 comprises, in 5' to 3' order: the first outer adaptor, fragment 1 sequence, and on the first strand, the first sticky end motif ANNNNT (SEQ ID NO.: 2).

Enzyme-treated fragment 2 comprises, in 5' to 3' order: on the second strand, the first sticky end motif ANNNNT (SEQ ID NO.: 2); fragment 2 sequence; and on the first strand, the second sticky end motif ANNNNT (SEQ ID NO.: 2). Enzyme-treated fragment 3 comprises, in 5' to 3' order: on the second strand, the second sticky end motif ANNNNT (SEQ ID NO.: 2); fragment 3 sequence; and the second outer adaptor.

[0204] *Assembly of cleaved fragments*

[0205] The first sticky ends of fragments 1 and 2 are annealed and the second sticky ends of fragments 2 and 3 are annealed, generating a gene comprising, in 5' to 3' order: the first outer adaptor, fragment 1 sequence, the first sticky end motif, fragment 2 sequence, the second sticky end motif, fragment 3 sequence, and the second outer adaptor. The assembled product comprises the predetermined sequence of the target gene without any scar sites. The assembled product is amplified using primers to the outer adaptors to generate desired quantities of the target gene.

EXAMPLE 5: Generation of precursor nucleic acid fragments using uracil as a non-canonical base

[0206] A double-stranded target gene of predetermined sequence is prepared using a de novo synthesis and assembly method described herein. The predetermined gene sequence is first analyzed to identify fragments which will be synthesized and assembled into the final gene product.

[0207] *Determination of gene fragment sequences*

[0208] The target nucleic acid sequence is analyzed to identify sticky end motifs. Three of the identified motifs are selected according to their position in the sequence, so that the motifs partition the predetermined sequence in four fragments having roughly similar sequence lengths. The sticky end motifs are designated sticky end motif x, sticky end motif y, and sticky end motif z. The precursor fragments are designated fragment 1, fragment 2, fragment 3, and fragment 4. Accordingly, the predetermined sequence comprises, in order: fragment 1 sequence, sticky end motif x, fragment 2 sequence, sticky end motif y, fragment 3 sequence, sticky end motif z, and fragment 4 sequence.

[0209] *De novo synthesis of precursor fragments*

[0210] Fragments 1-4 are prepared by de novo synthesis and PCA assembly of oligonucleic acids. During this process connecting adaptor sequences are added to the 3' end of fragment 1, the 5' and 3' ends of fragments 2 and 3, and the 5' end of fragment 4. The connecting adaptor

sequences located at the 3' end of fragment 1 and the 5' end of fragment 2 comprise sticky end motif x. The connecting adaptor sequences located at the 3' end of fragment 2 and the 5' end of fragment 3 comprise sticky end motif y. The connecting adaptor sequences located at the 3' end of fragment 3 and the 5' end of fragment 4 comprise sticky end motif z. Each connecting adaptor comprises, in order: a sequence of 1-10 bases (adaptor bases), a first nicking enzyme recognition site comprising a first nicking enzyme cleavage site on a first strand, a sticky end motif comprising a second nicking enzyme cleavage site on the 3' base of the second strand. The second nicking enzyme cleavage site comprises the non-canonical base uracil. The connecting adaptor sequences are positioned at the 5' and/or 3' end of a fragment such that the 3' uracil of the connecting adaptor is positioned directed next to the 5' and/or 3' end of the fragment. The adaptor bases and first nicking enzyme cleavage site comprise the same bases for each connecting adaptor.

[0211] Precursor fragment 1 comprises fragment 1 sequence and a first connecting adaptor comprising sticky end motif x. Precursor fragment 2 comprises the first connecting adaptor comprising sticky end motif x, fragment 2 sequence, and a second connecting adaptor comprising sticky end motif y. Precursor fragment 3 comprises the second connecting adaptor comprising sticky end motif y, fragment 3 sequence, and a third connecting adaptor comprising sticky end motif z. Precursor fragment 4 comprises the third connecting adaptor comprising sticky end motif z and fragment 4 sequence.

[0212] *Cleavage of fragments with two nicking enzymes*

[0213] Each of the four precursor fragments comprise one or two connecting adaptors, each connecting adaptor comprising: a first nicking enzyme recognition site comprising a first nicking enzyme cleavage site on a first strand, and uracil base on the second strand. The precursor fragments are treated with a first nicking enzyme which recognizes the first nicking enzyme recognition site to generate a nick at the first nicking enzyme cleavage site. The precursor fragments are treated with a second nicking enzyme, USER, which excises the uracil from the second strand, generating a nick where the uracil used to reside. USER comprises Uracil DNA glycosylase (UDG) and DNA glycosylase-lyase Endonuclease VIII (EndoVIII). Each precursor fragment now comprises an overhang consisting of a sticky end motif.

[0214] Precursor fragment 1 now comprises fragment 1 sequence and a 5' overhang consisting of sequence motif x. Precursor fragment 2 now comprises a 3' overhang consisting of sequence motif x, fragment 2 sequence, and a 5' overhang consisting of sequence motif y. Precursor fragment 3 now comprises a 3' overhang consisting of sequence motif y, fragment 3 sequence, and a 5' overhang consisting of sequence motif z. Precursor fragment 4 now comprises a 3' overhang consisting of sequence motif z and fragment 4 sequence.

[0215] *Assembly of cleaved fragments*

[0216] The sticky end motif x overhangs of precursor fragments 1 and 2 are annealed, the sticky end motif y overhangs of precursor fragments 2 and 3 are annealed, and the sticky end motif z overhangs of precursor fragments 3 and 4 are annealed, generating a gene comprising, in 5' to 3' order: fragment 1 sequence, sticky end motif x, fragment 2 sequence, sticky end motif y, fragment 3 sequence, sticky end motif z and fragment 4 sequence.

[0217] The product to be assembled comprises the predetermined sequence of the target gene without any scar sites. The assembled product is optionally amplified to generate desired quantities of the target gene. Alternatively, precursor fragments are generated at sufficient quantities such that amplification of the final gene is unnecessary. Such instances allow for the generation of large genes which are unable to be amplified using traditional amplification methods.

EXAMPLE 6: Universal primers to introduce a non-canonical base into a precursor nucleic acid fragment

[0218] A population of precursor nucleic acid fragments are amplified using a set of universal primer pairs, wherein each universal primer introduces a non-canonical base uracil to a single-strand of a precursor nucleic acid.

[0219] *Design of universal primers*

[0220] A predetermined sequence of a target gene is analyzed to select sticky end motifs that partition the gene into precursor fragments of desired size. The sticky end motifs have the sequence ANNNNT (SEQ ID NO.: 2), where each selected sticky end motif has a different NNNN sequence. The NNNN sequence for each selected sticky end motif is noted.

[0221] Universal forward primers are synthesized to comprise, in 5' to 3' order: 1-20 forward adaptor bases, a nicking enzyme recognition site, and a sticky end motif comprising ANNNNU (SEQ ID NO.: 53). A subpopulation of forward primers is generated so that each subpopulation comprises a NNNN sequence of a different sticky end motif selected from the target gene.

[0222] Universal reverse primers are synthesized to comprise, in 5' to 3' order: 1-20 reverse adaptor bases, a nicking enzyme recognition site, and a sticky end motif comprising ANNNNU (SEQ ID NO.: 53). A subpopulation of reverse primers is generated so that each subpopulation comprises the reverse complement of a NNNN sequence of a different sticky end motif selected from the target gene.

[0223] The nicking enzyme recognition site sequence in the universal primers is designed such that when the universal primers are incorporated into precursor fragments during an amplification reaction, the reverse complement sequence of the nicking enzyme recognition site sequence in the universal primer comprises a nicking enzyme cleavage site. Accordingly, upon treating with

a nicking enzyme specific for the nicking enzyme cleavage site, a nick is generated on a strand of the fragment not comprising the uracil base.

[0224] *Amplification of precursor nucleic acid fragments with universal primers*

[0225] Precursor fragments partitioned by the selected sticky end motifs are assembled from smaller, synthesized nucleic acids. The precursor fragments are amplified using the set of universal primers comprising the sticky end motif ANNNNT (SEQ ID NO.: 2), wherein the T is mutated with the non-canonical base uracil. The precursor fragments each comprise a nicking enzyme recognition site comprising a nicking enzyme cleavage site on one strand and a uracil base on the other strand.

[0226] *Enzymatic digestion of precursor fragments amplified with universal primers*

[0227] Precursor fragments amplified with universal primers are treated with a first nicking enzyme to create a nick at the nicking enzyme cleavage site and a second nicking enzyme comprising UDG and Endonuclease VIII activity to generate a nick at the uracil base site. The precursor fragments comprise overhangs with the sticky end motif ANNNNT (SEQ ID NO.: 2).

[0228] *Assembly of cleaved fragments*

[0229] Fragments comprising complementary overhangs are annealed to generate the target gene. The target gene comprises the predetermined sequence, with no extraneous scar sites.

EXAMPLE 7: Assembly of a target gene using Type II restriction endonucleases

[0230] A double-stranded target gene of predetermined sequence is prepared using a de novo synthesis and assembly method described herein. The predetermined gene sequence is first analyzed to identify fragments which will be synthesized and assembled into the final gene product.

[0231] *Determination of gene fragment sequences*

[0232] The target nucleic acid sequence is analyzed to identify sticky end motifs having a Type II restriction endonuclease recognition sequence. Three of the identified motifs are selected according to their position in the sequence, so that the motifs partition the predetermined sequence in four fragments having roughly similar sequence lengths of about 200 kb. The sticky end motifs are designated sticky end motif x, sticky end motif y, and sticky end motif z. The precursor fragments are designated fragment 1, fragment 2, fragment 3, and fragment 4. Accordingly, the predetermined sequence comprises, in order: fragment 1 sequence, sticky end motif x, fragment 2 sequence, sticky end motif y, fragment 3 sequence, sticky end motif z, and fragment 4 sequence.

[0233] *De novo synthesis of precursor fragments*

[0234] Precursor fragments 1-4 are prepared by de novo synthesis and PCA assembly of oligonucleic acids. During this process connecting adaptor sequences are added to the 3' end of

fragment 1, the 5' and 3' ends of fragments 2 and 3, and the 5' end of fragment 4. The connecting adaptor sequences located at the 3' end of fragment 1 and the 5' end of fragment 2 comprise sticky end motif x. The connecting adaptor sequences located at the 3' end of fragment 2 and the 5' end of fragment 3 comprise sticky end motif y. The connecting adaptor sequences located at the 3' end of fragment 3 and the 5' end of fragment 4 comprise sticky end motif z. Each connecting adaptor comprises a sequence of 1-10 adaptor bases and sticky end motif comprising a Type II restriction endonuclease recognition sequence. Also during preparation of precursor fragments 1-4, outer adaptors comprising 1-10 adaptor bases are added to the 5' and 3' ends of fragments 1 and 4, respectively. The adaptor bases comprise the same bases for each connecting adaptor and outer adaptor.

[0235] Precursor fragment 1 comprises outer adaptor sequence 1, fragment 1 sequence and a first connecting adaptor comprising sticky end motif x. Precursor fragment 2 comprises the first connecting adaptor comprising sticky end motif x, fragment 2 sequence, and a second connecting adaptor comprising sticky end motif y. Precursor fragment 3 comprises the second connecting adaptor comprising sticky end motif y, fragment 3 sequence, and a third connecting adaptor comprising sticky end motif z. Precursor fragment 4 comprises the third connecting adaptor comprising sticky end motif z, fragment 4 sequence, and outer adaptor sequence 2.

[0236] *Cleavage of fragments with Type II restriction enzymes*

[0237] Each of the four precursor fragments comprise one or two connecting adaptors, each connecting adaptor having a sticky end motif comprising a Type II restriction endonuclease recognition sequence. The precursor fragments are treated with three Type II restriction enzymes, each enzyme specific for a Type II recognition sequence in sticky end motifs X-Z, to generate four precursor fragments with sticky ends.

[0238] *Assembly of cleaved fragments*

[0239] The sticky end motif x overhangs of precursor fragments 1 and 2 are annealed, the sticky end motif y overhangs of precursor fragments 2 and 3 are annealed, and the sticky end motif z overhangs of precursor fragments 3 and 4 are annealed, generating a gene comprising, in 5' to 3' order: fragment 1 sequence, sticky end motif x, fragment 2 sequence, sticky end motif y, fragment 3 sequence, sticky end motif z and fragment 4 sequence. The product to be assembled comprises the predetermined sequence of the target gene without any scar sites.

[0240] While preferred embodiments of the present disclosure have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the disclosure. It should be understood that various alternatives to the embodiments described herein may be employed.

CLAIMS

WHAT IS CLAIMED IS:

1. A method for nucleic acid assembly, the method comprising:
 - a) providing a predetermined nucleic acid sequence;
 - b) providing a plurality of precursor double-stranded nucleic acid fragments, each precursor double-stranded nucleic acid fragment having two strands, wherein each of the two strands comprises a sticky end sequence of 5'-A (N^x)T-3' (SEQ ID NO.: 1) or 5'-G (N^x)C-3' (SEQ ID NO.: 16), wherein N is a nucleotide, wherein x is the number of nucleotides between nucleotides A and T or between G and C, and wherein x is 1 to 10, and wherein no more than two precursor double-stranded nucleic acid fragments comprise the same sticky end sequence;
 - c) providing primers comprising a nicking endonuclease recognition site and a sequence comprising (i) 5'-A (N^x)U-3' (SEQ ID NO.: 80) corresponding to each of the different sticky end sequences of 5'-A (N^x)T-3'(SEQ ID NO.: 1) or (ii) 5'-G (N^x)U-3' (SEQ ID NO.: 81) corresponding to each of the different sticky end sequences of 5'-G (N^x)C-3' (SEQ ID NO.: 16);
 - d) performing a polynucleotide extension reaction to form double-stranded nucleic acid fragments;
 - e) subjecting the polynucleotide extension reaction product to nicking and cleavage reactions to form double-stranded nucleic acid fragments with 3' overhangs; and
 - f) annealing the double-stranded nucleic acid fragments to form a nucleic acid encoding for the predetermined nucleic acid sequence that does not include the nicking endonuclease recognition site.
2. The method of claim 1, wherein x is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.
3. The method of any one of claims 1 or 2, wherein the predetermined nucleic acid sequence is 1 kb to 100 kb in length.
4. The method of any one of claims 1 to 3, wherein the predetermined nucleic acid sequence is 1 kb to 25 kb in length.
5. The method of any one of claims 1 to 4, wherein the predetermined nucleic acid sequence is 2 kb to 20 kb in length.
6. The method of any one of claims 1 or 2, wherein the predetermined nucleic acid sequence is at least 2 kb in length.
7. The method of any one of claims 1 to 6, wherein the plurality of single-stranded nucleic acid fragments are each at least 100 bases in length.

8. The method of any one of claims 1 to 7, wherein the double-stranded nucleic acid fragments are each at least 500 bases in length.
9. The method of any one of claims 1 to 8, wherein the double-stranded nucleic acid fragments are each at least 1 kb in length.
10. The method of claim 9, wherein the double-stranded nucleic acid fragments are each at least 20 kb in length.
11. The method of any one of claims 1 to 10, wherein the sticky ends are at least 4 bases long.
12. The method of any one of claims 1 to 11, wherein the sticky ends are 6 bases long.
13. The method of any one of claims 1 to 12, wherein step c further comprises providing (i) a forward primer comprising, in order 5' to 3': a first outer adaptor region and nucleic acid sequence from a first terminal portion of predetermined nucleic acid sequence; and (ii) a reverse primer, comprising, in order 5' to 3': a second outer adaptor region and nucleic acid sequence from a second terminal portion of predetermined nucleic acid sequence.
14. The method of claim 13, wherein the annealed double-stranded nucleic acid fragments comprise the first outer adaptor region and the second outer adapter region.
15. The method of any one of claims 1 to 14, wherein the nicking and cleavage reagents comprise a nicking endonuclease.
16. The method of claim 15, wherein the nicking endonuclease comprises endonuclease VIII.
17. The method of claim 15, wherein the nicking endonuclease is selected from the list consisting of Nb.BbvCI, Nb.BsmI, Nb.BsrDI, Nb.BtsI, Nt.AlwI, Nt.BbvCI, Nt.BsmAI, Nt.BspQI, Nt.BstNBI, and Nt.CviPII,
18. The method of any one of claims 1 to 17, further comprising ligating the annealed double-stranded nucleic acid fragments.
19. The method of any one of claims 1 to 17, wherein annealing comprises thermocycling between a maximum and a minimum temperature, thereby generating a first overhang from a first double-stranded DNA fragment and a second overhang from a second double-stranded DNA fragment, wherein the first and the second overhangs are complimentary, hybridizing the first and second overhangs to each other; and ligating.
20. The method of any one of claims 1 to 17, where a polymerase lacking 3' to 5' proofreading activity is added during the polynucleotide extension reaction.
21. The method of claim 18, wherein the polymerase is a Family A polymerase.
22. The method of claim 18, wherein the polymerase is a Family B high fidelity polymerase engineered to tolerate base pairs comprising uracil.

23. The method of any one of claims 1 to 22, wherein the precursor double-stranded nucleic acid fragments comprise an adaptor sequence comprising the nicking endonuclease recognition site.
24. The method of any one of claims 1 to 23, wherein one of the plurality of precursor double-stranded nucleic acid fragments is a linear vector.
25. A method for nucleic acid assembly, the method comprising:
- a) providing a predetermined nucleic acid sequence;
 - b) synthesizing a plurality of precursor double-stranded nucleic acid fragments, each precursor double-stranded nucleic acid fragment having two strands, wherein each of the two strands comprises a sticky end sequence of 5'-A (Nx) T-3' (SEQ ID NO.: 1) or 5'-G (Nx) C-3' (SEQ ID NO.: 16), wherein N is a nucleotide, wherein x is the number of nucleotides between nucleotides A and T or between G and C, and wherein x is 1 to 10, and wherein no more than two precursor double-stranded nucleic acid fragments comprise the same sticky end sequence;
 - c) providing primers comprising a nicking endonuclease recognition site and a sequence comprising (i) 5'-A (Nx) M-3' (SEQ ID NO.: 82) corresponding to each of the different sticky end sequences of 5'-A (Nx) T-3' (SEQ ID NO.: 1) or (ii) 5'-G (Nx) M-3' (SEQ ID NO.: 83) corresponding to each of the different sticky end sequences of 5'-G (Nx) C-3' (SEQ ID NO.: 16), wherein M is a non-canonical base, wherein the primers are each 7 to 70 bases in length;
 - d) performing a polynucleotide extension reaction to form double-stranded nucleic acid fragments;
 - e) subjecting the polynucleotide extension reaction product to nicking and cleavage reactions to form double-stranded nucleic acid fragments with 3' overhangs; and
 - f) annealing the double-stranded nucleic acid fragments to form a nucleic acid encoding for the predetermined nucleic acid sequence that does not include the nicking endonuclease recognition site.
26. The method of claim 25, wherein x is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.
27. The method of claim 26, wherein x is 4.
28. The method of any one of claims 25 to 27, wherein the non-canonical base is uracil, inosine, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, acetylcytosine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N-6-isopentenyl adenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 1-methyladenine, 2-methyladenine, 2-methylguanine, 3-

methylcytosine, 5-methylcytosine, 5-ethylcytosine, N6-adenine, N6-methyladenine, N,N-dimethyladenine, 8-bromoadenine, 7-methylguanine, 8-bromoguanine, 8-chloroguanine, 8-aminoguanine, 8-methylguanine, 8-thioguanine, 5-ethyluracil, 5-propyluracil, 5-methylaminomethyluracil, methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid, pseudouracil, 1-methylpseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-hydroxymethyluracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-S-oxyacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, 5-(2-bromovinyl)uracil, 2-aminopurine, 6-hydroxyaminopurine, 6-thiopurine, or 2,6-diaminopurine.

29. The method of any one of claims 25 to 28, wherein the non-canonical base is incorporated into the double-stranded nucleic acid fragments by performing a nucleic acid extension reaction from a primer comprising the non-canonical nucleotide.

30. The method of any one of claims 25 to 27, wherein the non-canonical base is a uracil.

31. The method of claim 30, wherein the uracil is in a deoxyuridine-deoxyadenosine base pair.

32. The method of any one of claims 25 to 31, wherein the primers are 10 to 30 bases in length.

33. The method of any one of claims 25 to 32, wherein one of the plurality of precursor double-stranded nucleic acid fragments comprises a portion of linear vector.

34. The method of any one of claims 25 to 33, wherein no more than 2 N nucleotides of the sticky end sequence have the same identity.

35. The method of any one of claims 25 to 34, wherein the precursor double-stranded nucleic acid fragments comprise an adaptor sequence comprising the nicking endonuclease recognition site.

36. The method of any one of claims 25 to 35, wherein the predetermined nucleic acid sequence is 1 kb to 100 kb in length.

37. The method of any one of claims 25 to 36, wherein the plurality of precursor nucleic acid fragments are each at least 100 bases in length.

38. The method of any one of claims 25 to 37, wherein the sticky ends are at least 4 bases long in each precursor nucleic acid.

39. A method for nucleic acid assembly, the method comprising:

- a) providing a predetermined nucleic acid sequence;

b) synthesizing a plurality of single-stranded nucleic acid fragments, wherein each single-stranded nucleic acid fragment encodes for a portion of the predetermined nucleic acid sequence and comprises at least one sticky end motif, wherein the sticky end motif comprises a sequence of 5'-A(N^x)T-3' (SEQ ID NO. : 1) or 5'-G(N^x)C-3' (SEQ ID NO. : 16) in the predetermined nucleic acid sequence, wherein N is a nucleotide, wherein x is the number of nucleotides between nucleotides A and T or between G and C, and wherein x is 1 to 10, and wherein no more than two single-stranded nucleic acid fragments comprise the same sticky end sequence;

c) amplifying the plurality of single-stranded nucleic acid fragments to generate a plurality of double-stranded nucleic acid fragments, wherein the plurality of double-stranded nucleic acid fragments are modified from the predetermined nucleic acid sequence to comprise (i) a non-canonical base located at a 3' end of the sticky end motif on a first strand and (ii) a first adaptor region located 5' of the non-canonical base on the first strand, wherein the first adaptor region comprises a nicking enzyme recognition site;

d) creating sticky ends, wherein creating sticky ends comprises:

i) treating the plurality of double-stranded fragments with a first nicking enzyme that nicks the non-canonical base on a first strand of each double-stranded fragment, and cleaving the nicked non-canonical base; and

ii) treating the plurality of double-stranded fragments with a second nicking enzyme, wherein the second nicking enzyme binds to the first strand at the nicking enzyme recognition site and cleaves a second strand of each double-stranded fragment, wherein a cleavage site for the nicking enzyme is located at a junction between the sticky end motif a sequence reverse complementary to the first adaptor region of the first strand; and

e) annealing the double-stranded nucleic acid fragments to form a nucleic acid encoding for the predetermined nucleic acid sequence that does not include the nicking endonuclease recognition site.

40. The method of claim 39, wherein x is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

41. The method of any one of claims 39 or 40, wherein the predetermined nucleic acid sequence is 1 kb to 100 kb in length.

42. The method of any one of claims 39 to 41, wherein the predetermined nucleic acid sequence is 1 kb to 25 kb in length.

43. The method of any one of claims 39 to 42, wherein the predetermined nucleic acid sequence is 2 kb to 20 kb in length.

44. The method of any one of claims 39 or 40, wherein the predetermined nucleic acid sequence is at least 2 kb in length.

45. The method of any one of claims 39 to 44, wherein the plurality of single-stranded nucleic acid fragments are each at least 100 bases in length.

46. The method of any one of claims 39 to 45, wherein the plurality of single-stranded nucleic acid fragments are each at least 500 bases in length.

47. The method of any one of claims 39 to 46, wherein the plurality of single-stranded nucleic acid fragments are each at least 1 kb in length.

48. The method of claim 45, wherein the plurality of single-stranded nucleic acid fragments are each at least 20 kb in length.

49. The method of any one of claims 39 to 48, wherein the sticky ends are at least 4 bases long.

50. The method of any one of claims 39 to 49, wherein the sticky ends are 6 bases long.

51. The method of any one of claims 39 to 50, wherein the non-canonical base is uracil, inosine, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, acetylcytosine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N-6-isopentenyl adenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 1-methyladenine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, 5-ethylcytosine, N6-adenine, N6-methyladenine, N,N-dimethyladenine, 8-bromoadenine, 7-methylguanine, 8-bromoguanine, 8-chloroguanine, 8-aminoguanine, 8-methylguanine, 8-thioguanine, 5-ethyluracil, 5-propyluracil, 5-methylaminomethyluracil, methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid, pseudouracil, 1-methylpseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-hydroxymethyluracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-S-oxyacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, 5-(2-bromovinyl)uracil, 2-aminopurine, 6-hydroxyaminopurine, 6-thiopurine, or 2,6-diaminopurine.

52. The method of claim 51, wherein the non-canonical base is incorporated into the double-stranded nucleic acid by performing a nucleic acid extension reaction from a primer comprising the non-canonical nucleotide.

53. The method of claim 51, wherein the non-canonical base is a uracil.
54. The method of claim 53, wherein the uracil is in a deoxyuridine-deoxyadenosine base pair.
55. The method of any one of claims 39 to 55, wherein the nicking recognition site is a nicking endonuclease recognition site.
56. The method of any one of claims 39 to 45, wherein the distance between the non-canonical base the nicking enzyme cleavage site is less than 12 base pairs.
57. The method of any one of claims 39 to 45, wherein the distance between the non-canonical base the nicking enzyme cleavage site is at least 5 base pairs.
58. The method of any one of claims 39 to 57, wherein the first nicking enzyme comprises a base excision activity.
59. The method of claim 58, wherein the first nicking enzyme comprises uracil-DNA glycosylase (UDG).
60. The method of claim 59, wherein the first nicking enzyme comprises an AP endonuclease.
61. The method of claim 60, wherein the first nicking enzyme comprises endonuclease VIII.
62. The method of any one of claims 39 to 61, wherein the second nicking enzyme a nicking endonuclease.
63. The method of claim 62, wherein the nicking endonuclease is selected from the list consisting of Nb.BbvCI, Nb.BsmI, Nb.BsrDI, Nb.BtsI, Nt.AlwI, Nt.BbvCI, Nt.BsmAI, Nt.BspQI, Nt.BstNBI, and Nt.CviPII,
64. The method of any one of claims 39 to 63, wherein each of the plurality of double-stranded nucleic acid fragments further comprises a two sticky ends.
65. The method of claim 64, wherein each of the two sticky ends have a different sequence from each other.
66. The method of any one of claims 39 to 65, wherein the sticky ends comprises a 3' overhang.
67. The method of any one of claims 39 to 66, further comprising ligating the annealed double-stranded nucleic acid fragments.
68. The method of any one of claims 39 to 66, wherein annealing comprises: thermocycling between a maximum and a minimum temperature, thereby generating a first overhang from a first double-stranded DNA fragment and a second overhang from a second double-stranded DNA fragment, wherein the first and the second overhangs are complimentary; hybridizing the first and second overhangs to each other; and

ligating.

69. The method of any one of claims 39 to 68, wherein the annealed double-stranded nucleic acid fragments comprise a 5' outer adaptor region and a 3' outer adaptor region.

70. The method of any one of claims 39 to 69, wherein at least two non-identical single-stranded nucleic acid fragments are synthesized.

71. The method of claim 70, wherein at least 5 non-identical single-stranded nucleic acid fragments are synthesized.

72. The method of claim 71, wherein at least 20 non-identical single-stranded nucleic acid fragments are synthesized.

73. The method of any one of claims 39 to 72, where a polymerase lacking 3' to 5' proofreading activity is added during the amplification step.

74. The method of claim 73, wherein the polymerase is a Family A polymerase.

75. The method of claim 73, wherein the polymerase is a Family B high fidelity polymerase engineered to tolerate base pairs comprising uracil.

76. The method of claim 39, wherein the amplified plurality of single-stranded nucleic acid fragments are not naturally occurring.

77. A nucleic acid library generated by any one of claims 1 to 76.

78. A DNA library comprising:
n DNA fragments, each comprising a first strand and a second strand, each of the n DNA fragments comprising, in order 5' to 3': a first nicking endonuclease recognition site, a first sticky end motif, a template region, a second sticky end motif, and a second nicking endonuclease recognition site, wherein the first sticky end motif comprises a sequence of 5'-A (N^x)_iiU-3' (SEQ ID NO.: 13) in the first strand; and wherein the second sticky end motif comprises a sequence of 5'-A (N^x)_{i,2}U-3'(SEQ ID NO. : 14) in the second strand; wherein N^x denotes x nucleosides; wherein (N^x)_{i,2} is reverse complementary to (N^x)_{i,1} and different from every other N^x found in any sticky end motif sequence within the fragment library, wherein the first nicking endonuclease recognition site in each of the DNA fragments are positioned such that there is a corresponding cleavage site immediately 3' of the sticky end motif in the second strand, and wherein the second nicking endonuclease recognition sites are positioned such that there is a corresponding cleavage site immediately 3' of the second sticky end motif in the first strand.

79. The fragment library of claim 78, wherein the first nicking endonuclease recognition site, the first sticky end motif, the variable insert, the second sticky end motif site, and the second nicking endonuclease recognition site are ordered as recited.

80. The fragment library of claim 78 further comprising a starter DNA fragment comprising a template region, a second sticky end motif, and a second nicking endonuclease recognition site; wherein the second sticky end motif comprises a sequence of 5'-A (N^x)_{s,2} U-3' (SEQ ID NO.: 20) and wherein (N^x)_{s,2} is reverse complementary to (N^x)_{1,1}.

81. The fragment library of claim 78, further comprising a finishing DNA fragment comprising a first nicking endonuclease recognition site, a first sticky end motif, and a template region; wherein the first sticky end motif comprises a sequence of 5'-A (N^x)_{f,i} U-3' (SEQ ID NO.: 21) and wherein (N^x)_{f,i} is reverse complementary to (N^x)_{n,2}.

82. The fragment library of claim 78, wherein the first and second nicking endonuclease recognition sites are the same.

83. The fragment library of claim 78, wherein n is at least 2.

84. The fragment library of claim 78, wherein n is less than 10.

85. The fragment library of claim 78, wherein x is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

86. The fragment library of claim 85, wherein x is 4.

87. The fragment library of claim 78, wherein the template region of each of the n DNA fragments encodes for a different nucleic acid sequencing from the template region of every other of the n DNA fragments.

88. The fragment library of claim 78, wherein the sequences of the n DNA fragments are not naturally occurring.

89. The fragment library of claim 78, wherein the first nicking endonuclease recognition site is not naturally adjacent to the first sticky end motif.

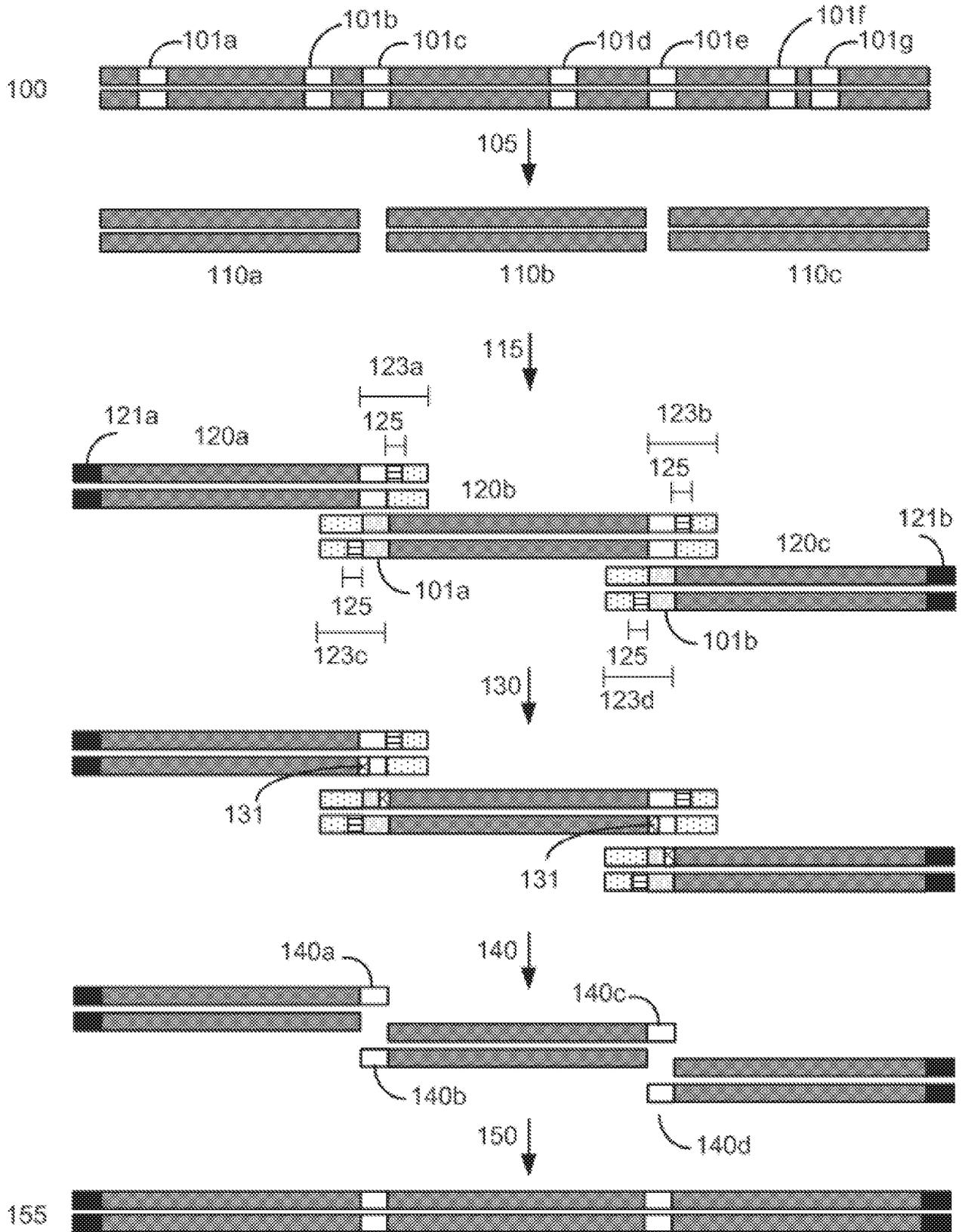


FIG. 1

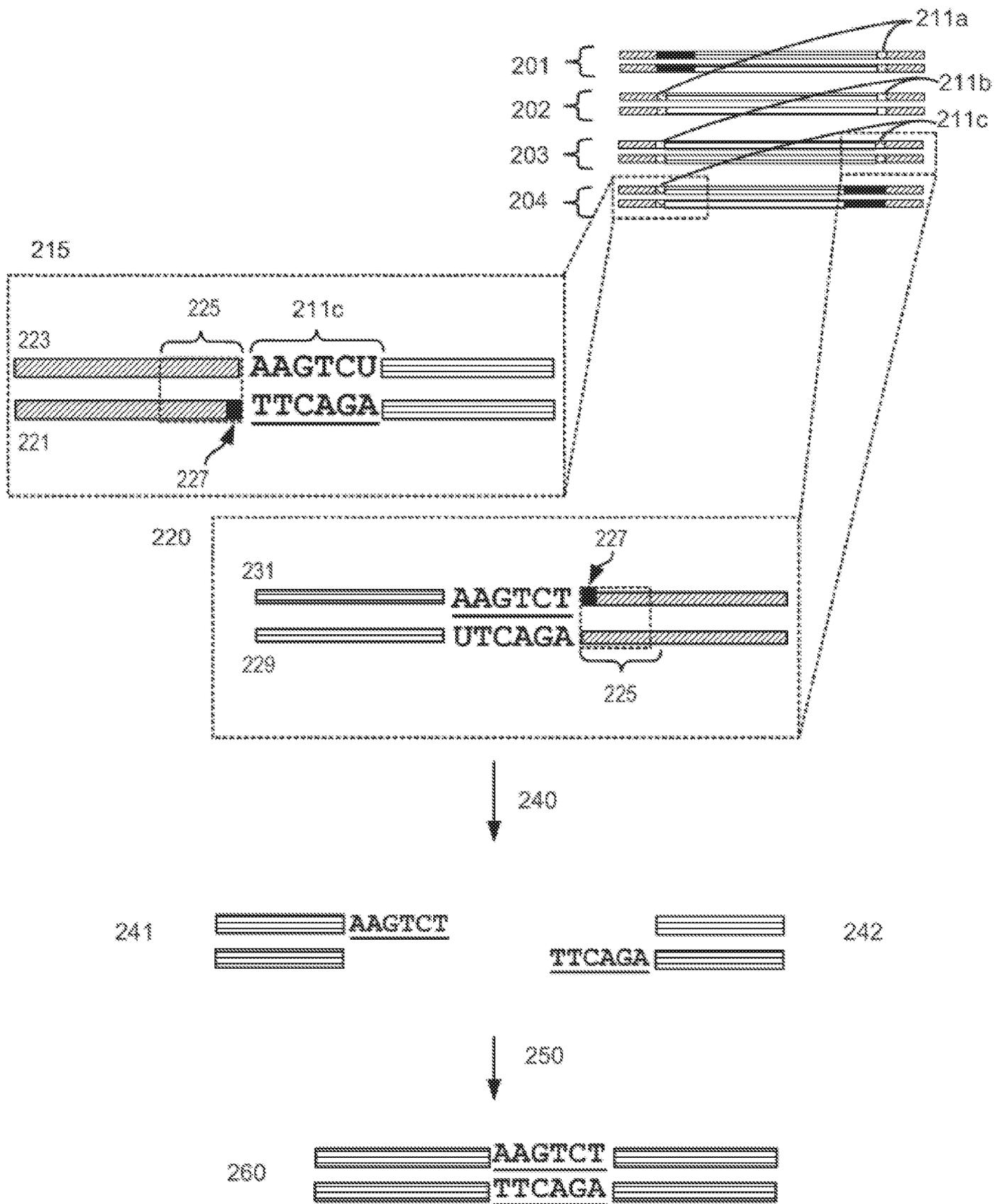


FIG. 2

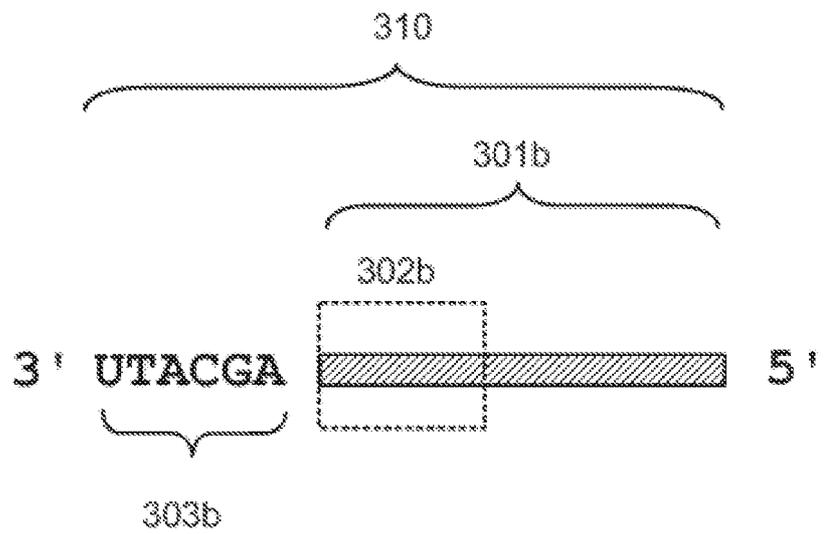
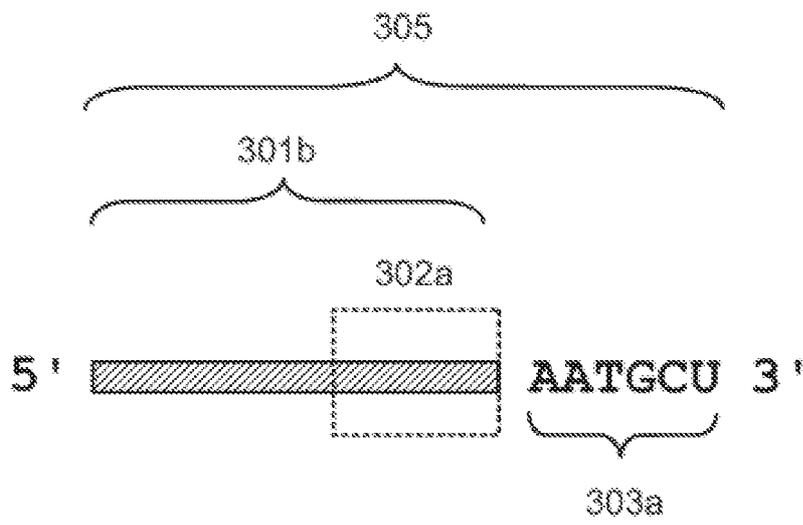


FIG. 3

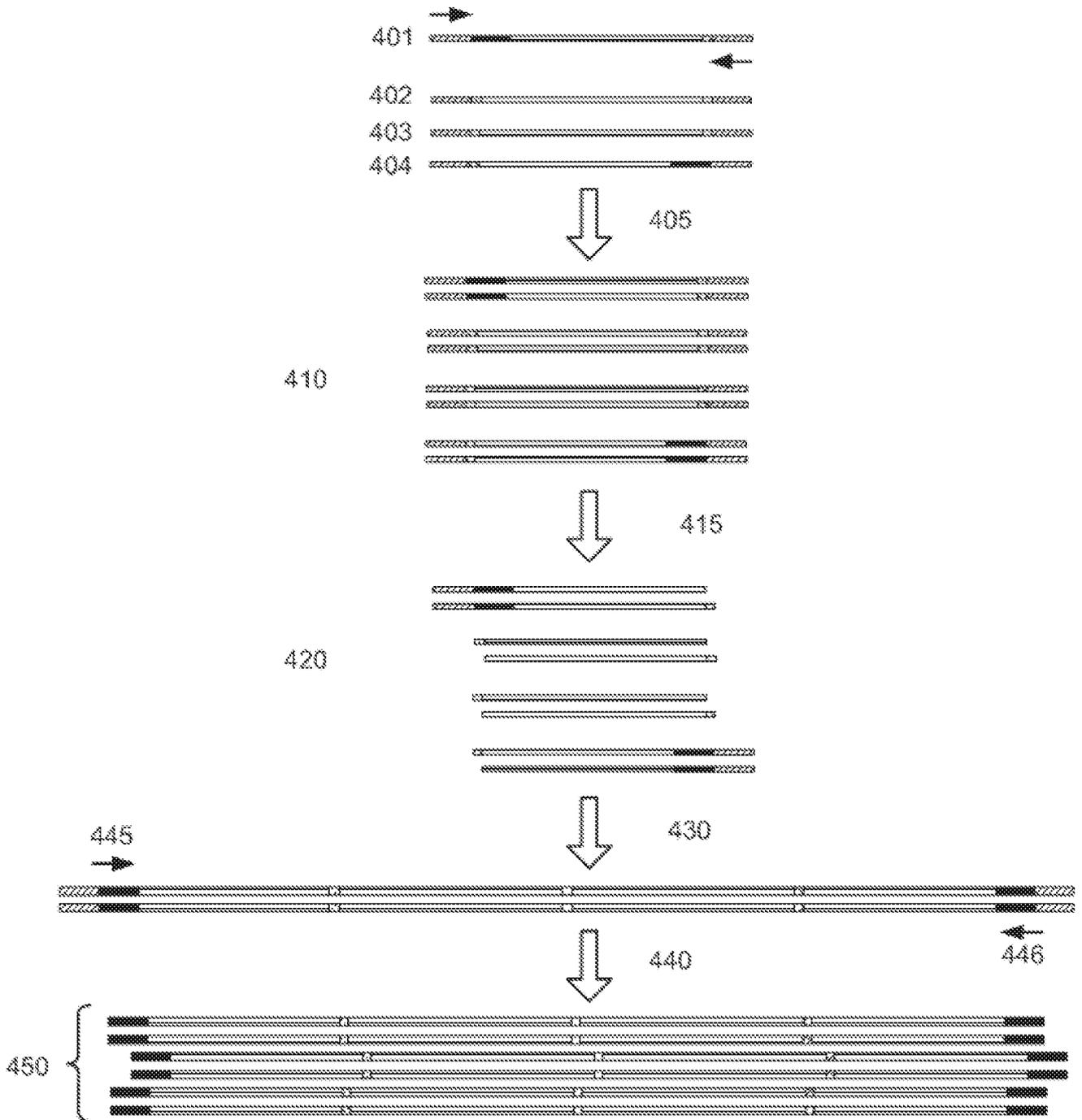


FIG. 4

500

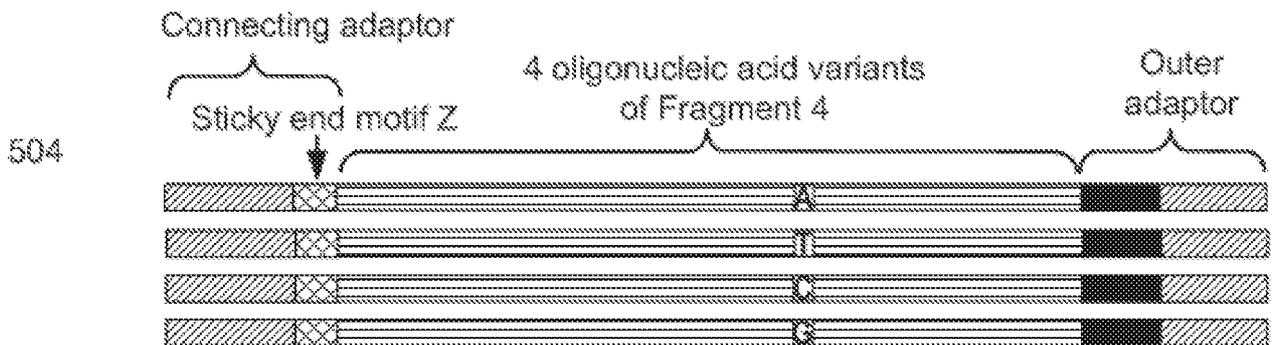
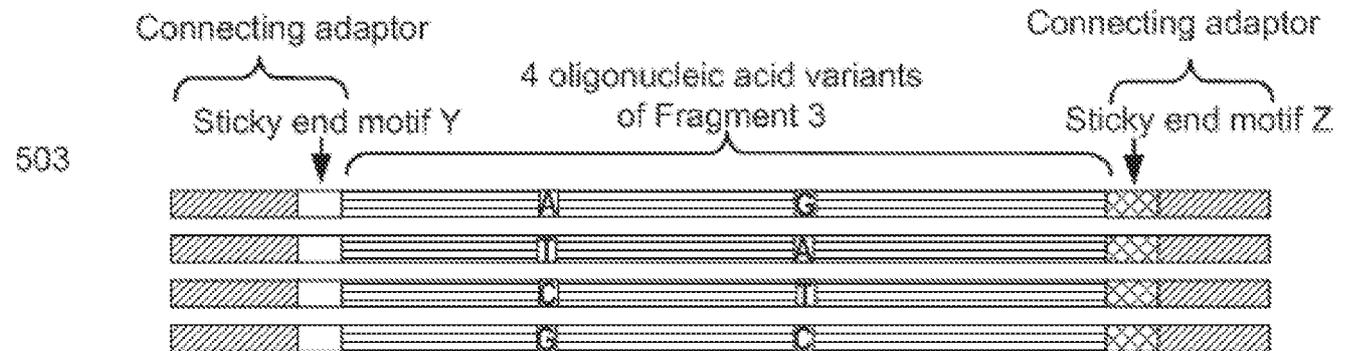
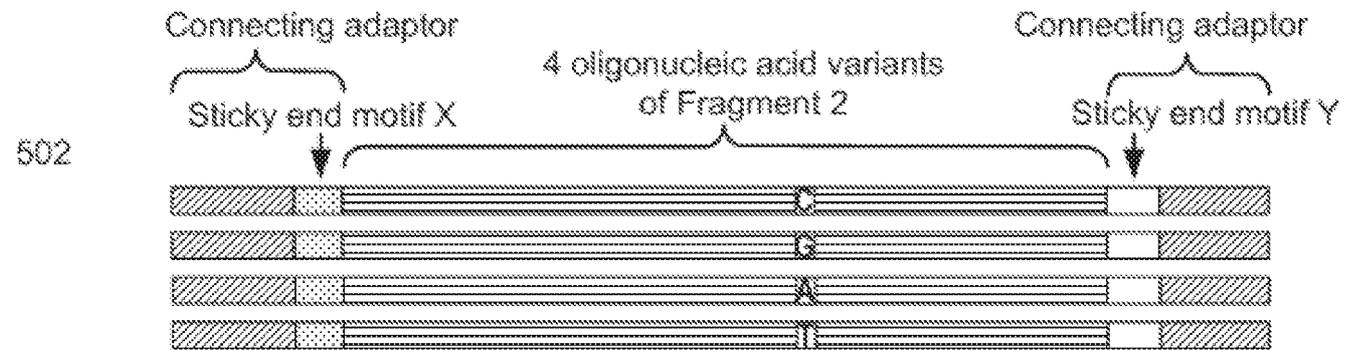
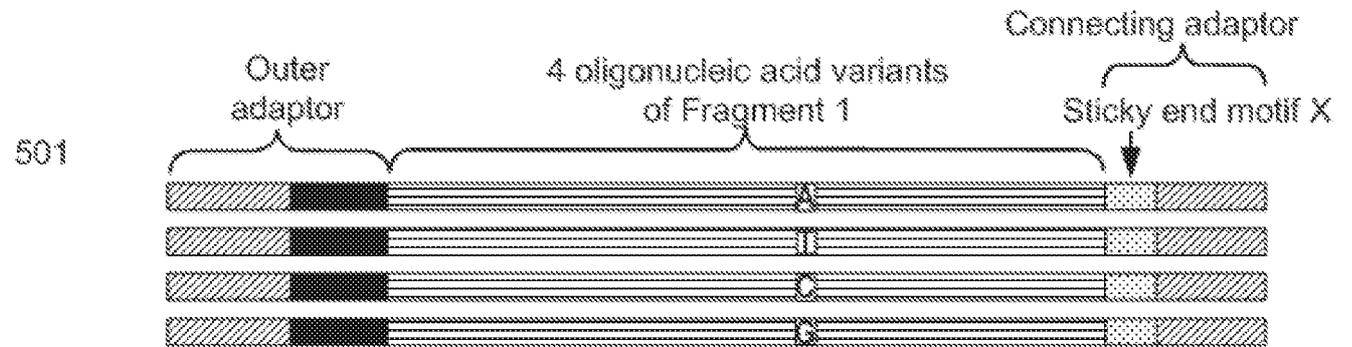


FIG. 5A

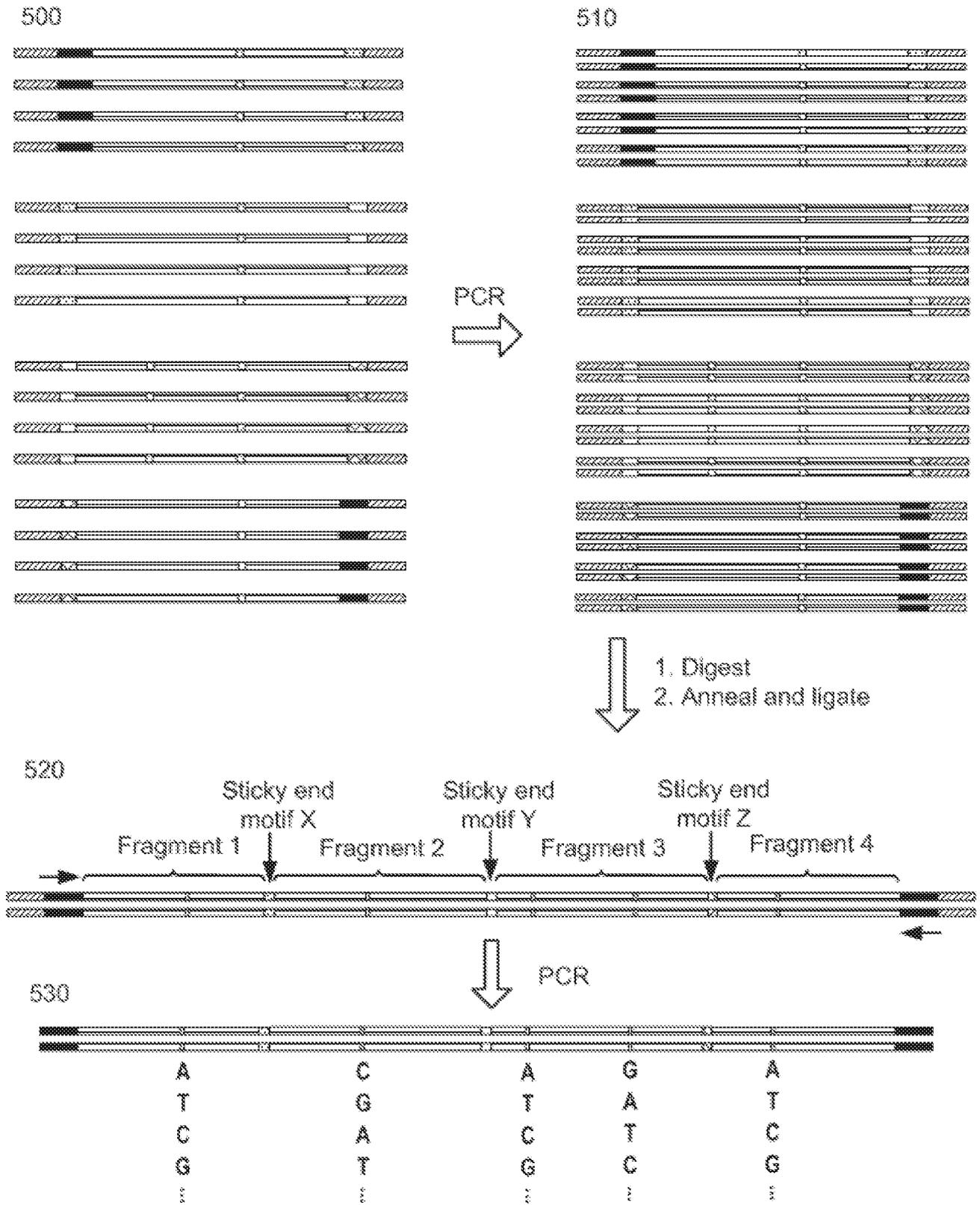


FIG. 5B

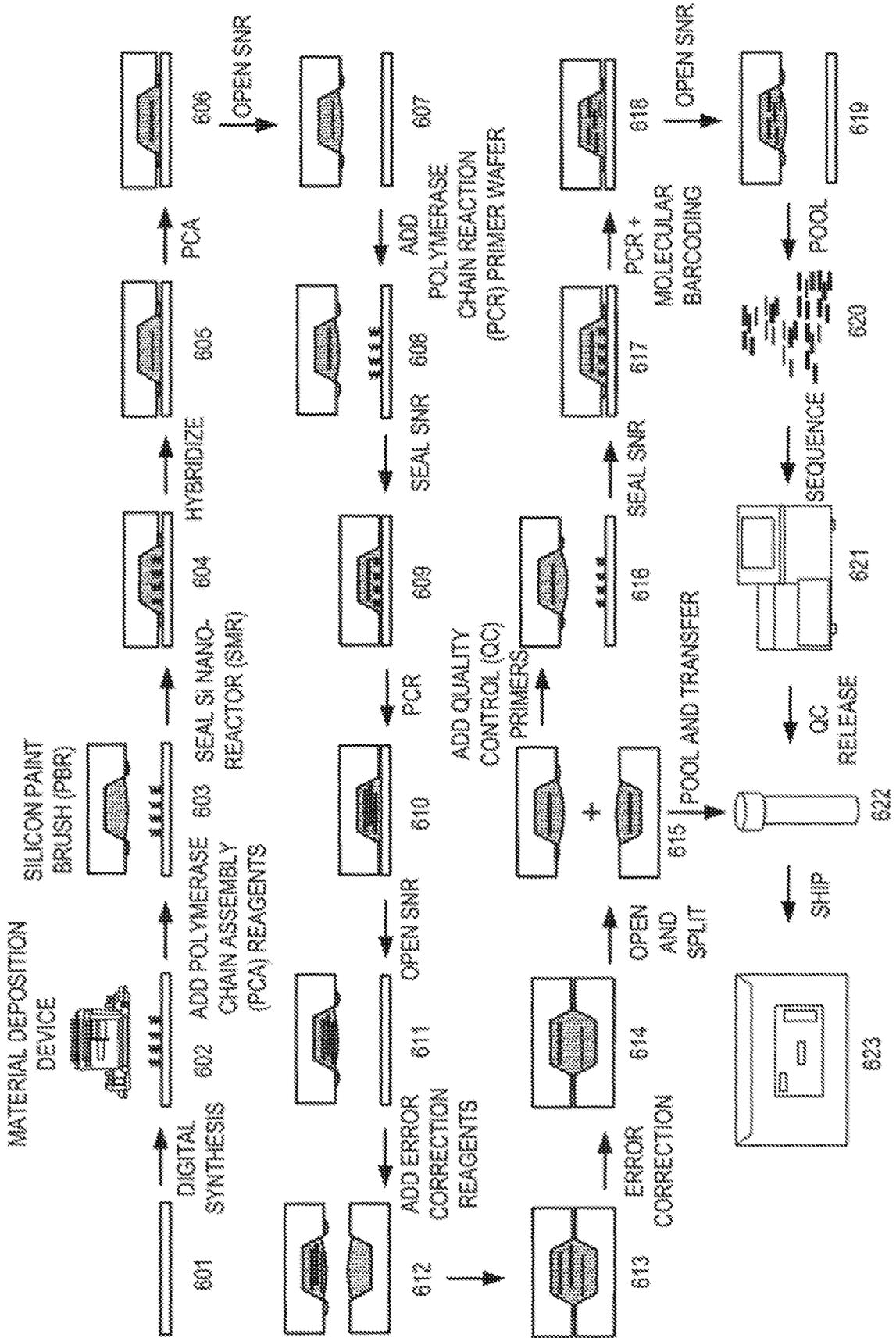


FIG. 6

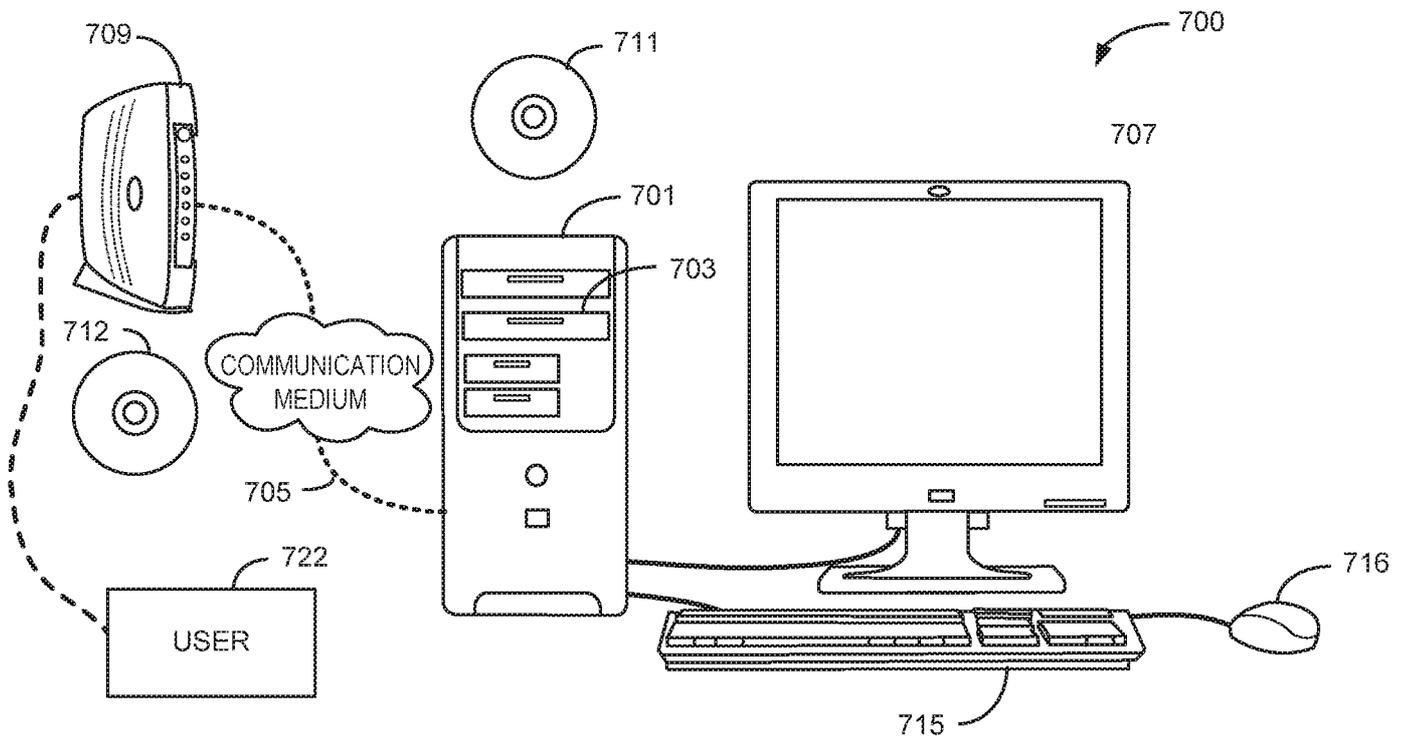


FIG. 7

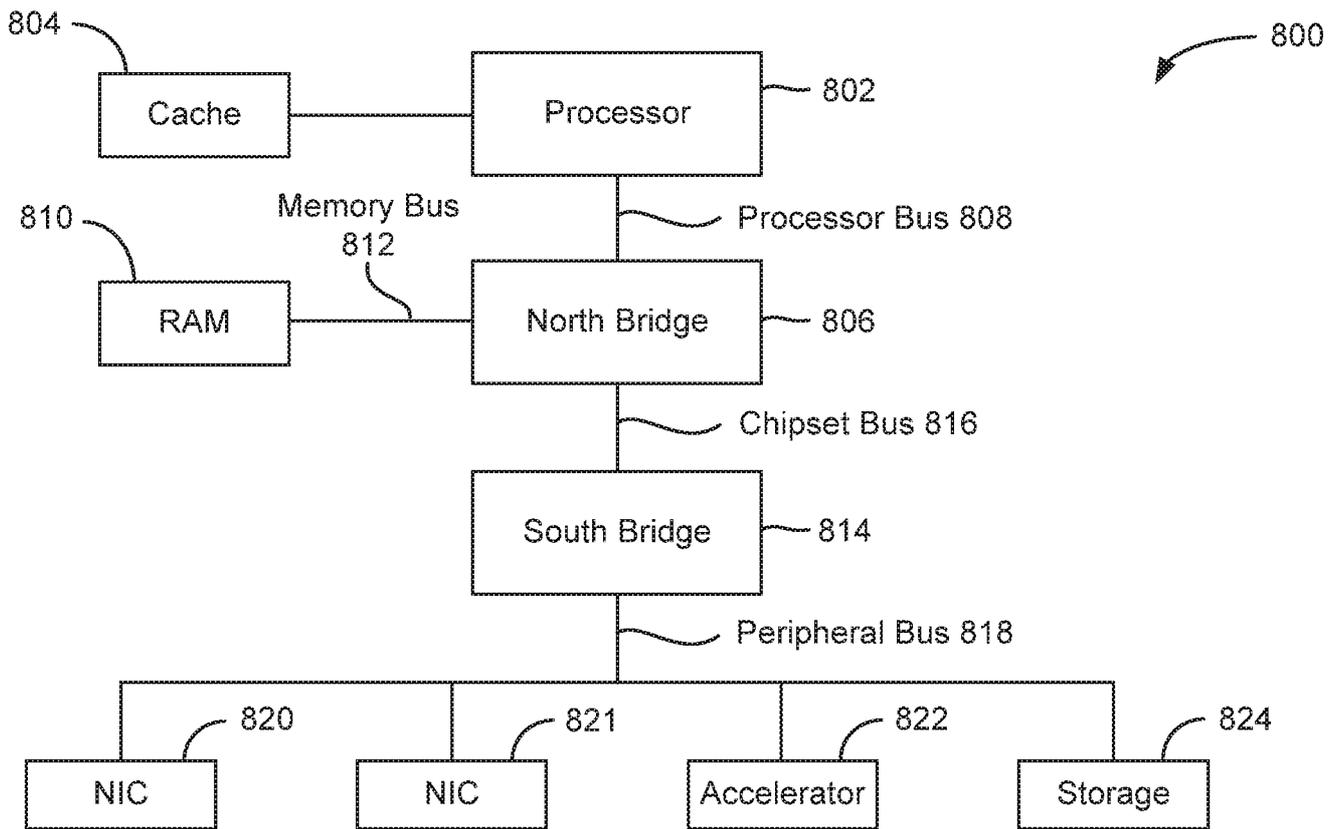


FIG. 8

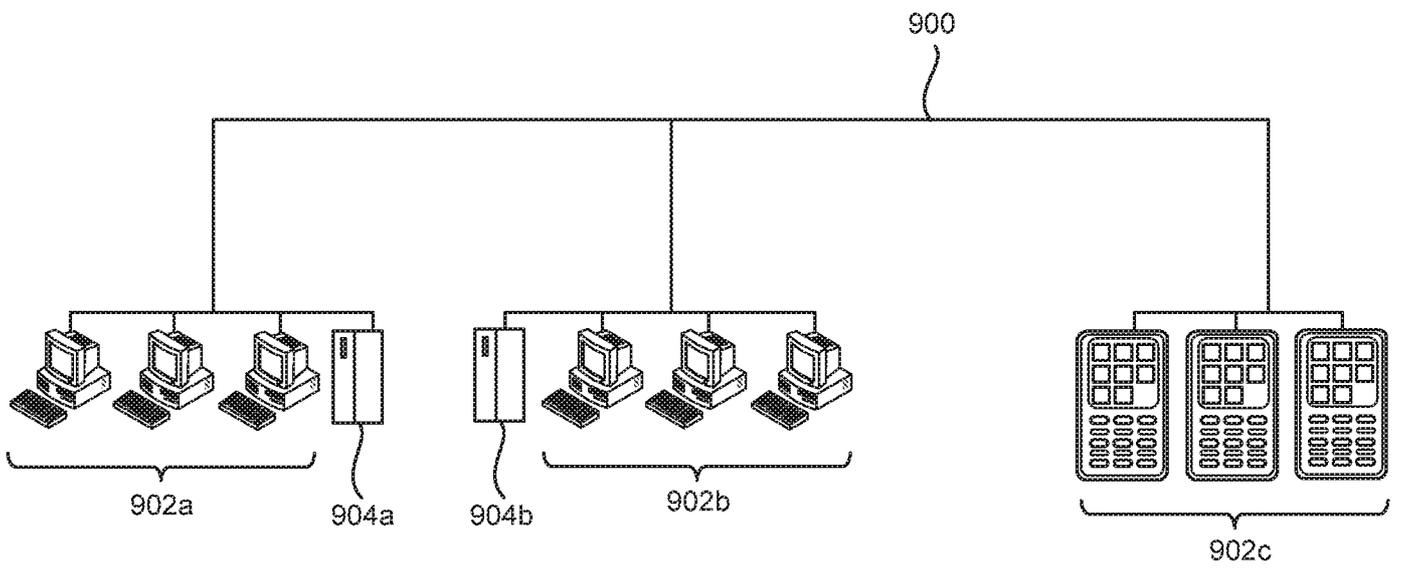


FIG. 9

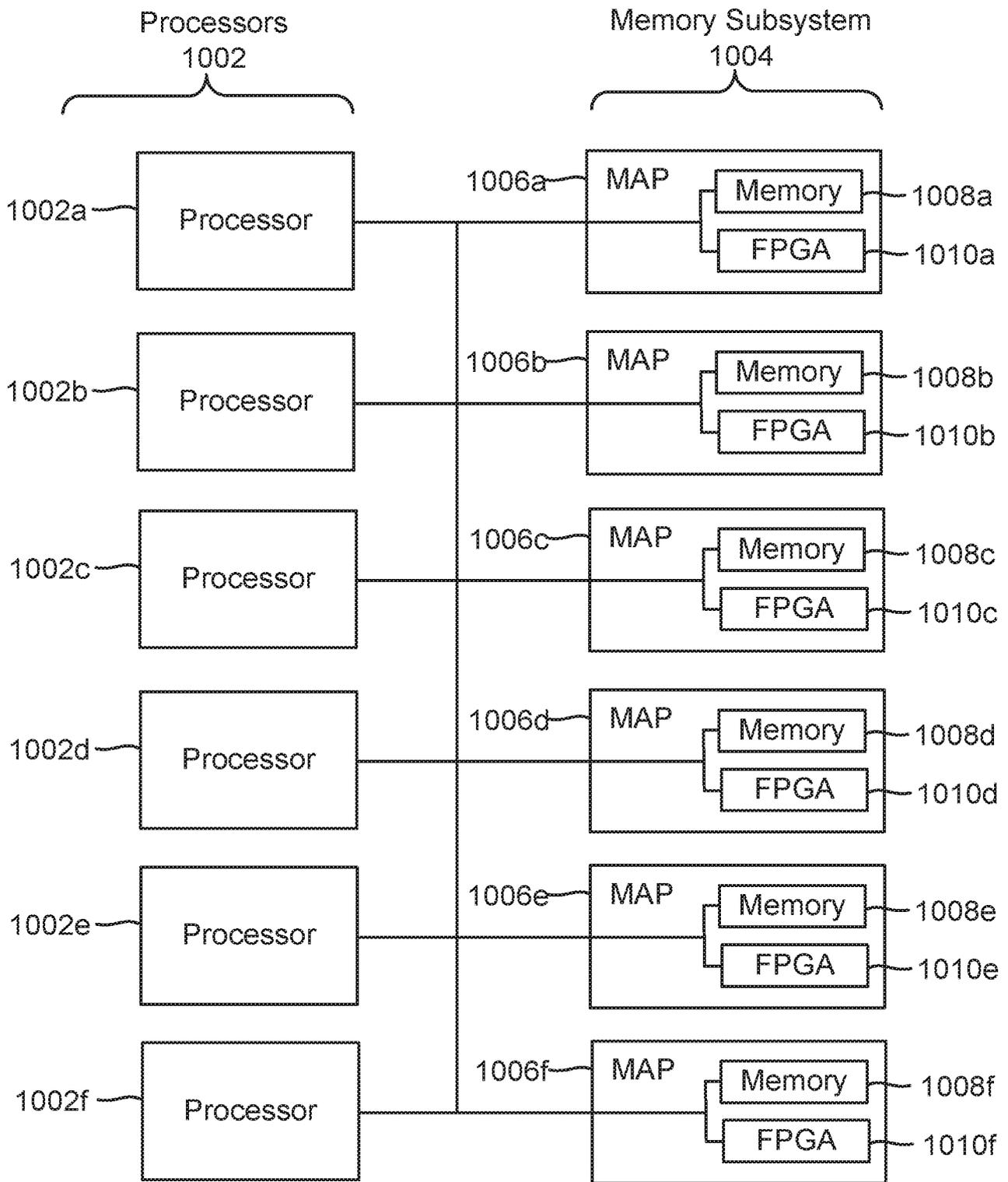


FIG. 10

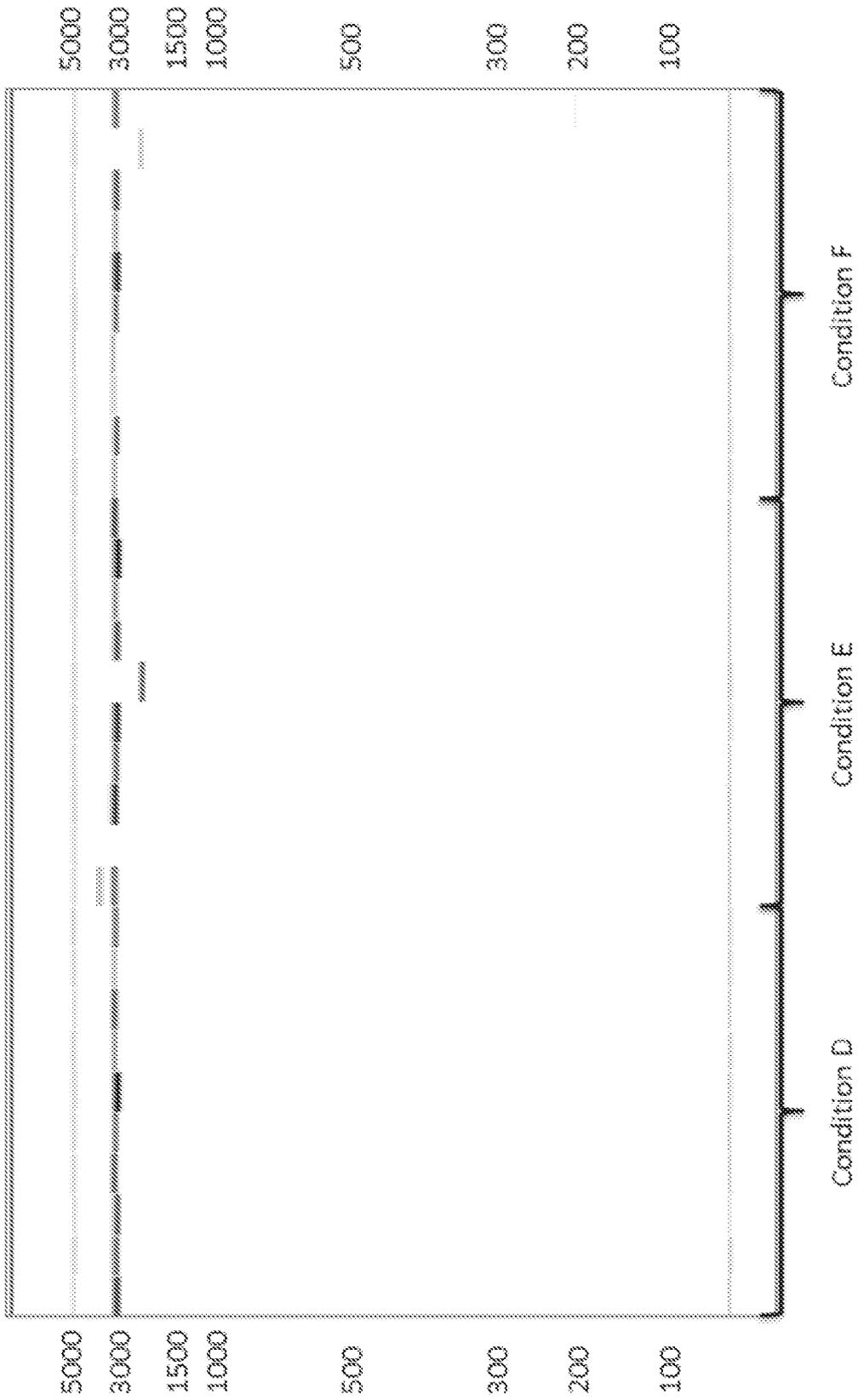


FIG. 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 16/1 6636

<p>A. CLASSIFICATION OF SUBJECT MATTER IPC(8) - C12N 15/00, 15/10, 15/64 (2016.01) CPC - C12N 15/00, 15/10, 15/64 According to International Patent Classification (IPC) or to both national classification and IPC</p>																
<p>B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC(8): C12N 15/00, 15/10, 15/64 (2016.01) CPC: C12N 15/00, 15/10, 15/64, 15/101 , 15/102 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) PatSccr (UO, EP, WO, JP, DE, tib, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); EBSCO Discovery; PubMed; Google; Google Scholar; Google Patents; KEYWORDS: nucleic acid, double strand, fragment, primer, nick, endonuclease recognition site, stick end</p>																
<p>C. DOCUMENTS CONSIDERED TO BE RELEVANT</p> <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>US 2003/0022317 A 1 (JACK, W et al.) January 30, 2003; abstract; figure 7 paragraphs [0013], [0017], [0024], [0025], [0028], [0049]; claims 1, 11</td> <td>1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31/30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89</td> </tr> <tr> <td>A</td> <td>US 5,712,124 A (WALKER) January 27, 1998; abstract; claim 1</td> <td>1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31/30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89</td> </tr> <tr> <td>A</td> <td>US 2011/0009607 A 1 (KOMIYAMA, M et al.) January 13, 2011; abstract; paragraph [0004]</td> <td>1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31/30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89</td> </tr> <tr> <td>A</td> <td>US 2012/0288476 A 1 (HARTMANN, G et al.) November 15, 2012; abstract; paragraph [0226]</td> <td>1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31(30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89</td> </tr> </tbody> </table>		Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	A	US 2003/0022317 A 1 (JACK, W et al.) January 30, 2003; abstract; figure 7 paragraphs [0013], [0017], [0024], [0025], [0028], [0049]; claims 1, 11	1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31/30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89	A	US 5,712,124 A (WALKER) January 27, 1998; abstract; claim 1	1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31/30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89	A	US 2011/0009607 A 1 (KOMIYAMA, M et al.) January 13, 2011; abstract; paragraph [0004]	1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31/30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89	A	US 2012/0288476 A 1 (HARTMANN, G et al.) November 15, 2012; abstract; paragraph [0226]	1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31(30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.														
A	US 2003/0022317 A 1 (JACK, W et al.) January 30, 2003; abstract; figure 7 paragraphs [0013], [0017], [0024], [0025], [0028], [0049]; claims 1, 11	1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31/30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89														
A	US 5,712,124 A (WALKER) January 27, 1998; abstract; claim 1	1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31/30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89														
A	US 2011/0009607 A 1 (KOMIYAMA, M et al.) January 13, 2011; abstract; paragraph [0004]	1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31/30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89														
A	US 2012/0288476 A 1 (HARTMANN, G et al.) November 15, 2012; abstract; paragraph [0226]	1-2, 3/1-2, 6/1-2, 25-27, 28/25-27, 30/25-27, 31(30/25-27, 39-40, 41/39-40, 44/39-40, 76, 78-89														
<p><input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.</p>																
<p>* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family</p>																
<p>Date of the actual completion of the international search:</p> <p>21 April 2016 (21.04.2016)</p>	<p>Date of mailing of the international search report</p> <p>02 MAY 2016</p>															
<p>Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300</p>	<p>Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774</p>															

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US16/16636

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2012/061832 A 1 (ILLUMINA, INC.) May 10, 2012; page 2, lines 7-10, 27-30; page 9, lines 6-11; page 44, lines 5-12; claim 1	78-89
A	US 2006/0024721 A1 (PEDERSEN, ML) February 02, 2006; paragraphs [0060], [0291], [0292]	78-89

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US16/16636

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
- a. forming part of the international application as filed:
 in the form of an Annex C/ST.2₅ text file.
 on paper or in the form of an image file.
- b. furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.2₅ text file.
- c. furnished subsequent to the international filing date for the purposes of international search only:
 in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).
 on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).
2. In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US16/16636

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

- 1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

- 2. Claims Nos.;;
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

- 3. Claims Nos.: 4, 5, 7-24, 29, 32-38, 42, 43, 45-75, 77
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

- 1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
- 2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
- 3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

- 4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.