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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 December 2015 (10.12.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/187786 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C07H 21/04* (2006.01)

(21) International Application Number:
PCT/US2015/033915

(22) International Filing Date:
3 June 2015 (03.06.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/007,235 3 June 2014 (03.06.2014) US

(71) Applicant: **HELIX NANOTECHNOLOGIES, INC.** [US/US]; 415 Hudson Street, Apt 3, Oakland, CA 94618 (US).

(72) Inventors: **NAMIH, Carina**; Flat G5 59, Chilton Street, London E2 6EA (GB). **WENDELSTADT, Katharina**; Rua Madre Teodora 159, Jardim Paulista, São Paulo, CEP 01428-010 (BR). **SIWO, Geoffrey**; 2 Canfield Avenue, Apt 605, White Plains, NY 10601 (US). **RAJANIEMI, Hannu**; 415 Hudson Street, Apt. 3, Oakland, CA 94618 (US). **LOOS, Remco**; 4 Celtic House, Hooper Street, Cambridge, Cambridgeshire CB1 2PB (GB). **MARTIN, Cristina, Cotobal**; 165 Victoria Road, Cambridge CB4 3BU (GB). **SMALLIE, Tim**; 7 White Hart Lane, London SW130PX (GB).

(74) Agent: **FOSTER, David C.**; Roberts Foster LLP, 1400 Civic Place, Suite 221, Southlake, Texas 76092 US (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, 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.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

[Continued on next page]

(54) Title: PROGRAMMABLE NUCLEOTIDE ENZYMES

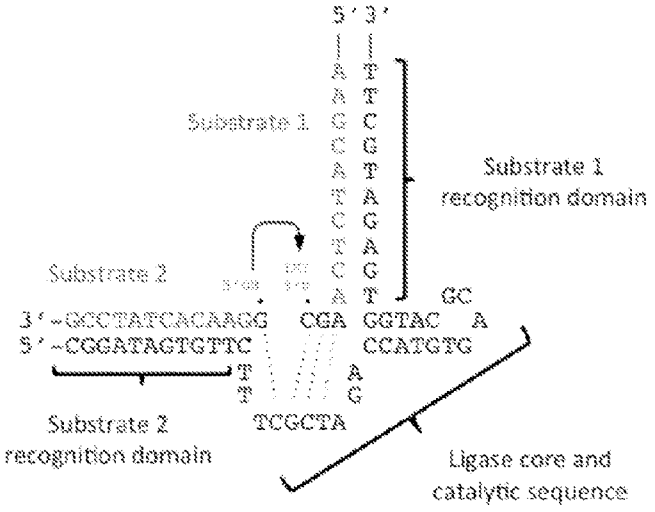

Figure 1

(57) Abstract: The present invention is directed to compositions and methods related to nucleotide enzymes, such as DNA ligases.

WO 2015/187786 A2

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, 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).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

PATENT APPLICATION

FOR

PROGRAMMABLE NUCLEOTIDE ENZYMES

The present application claims the benefit of U.S. Provisional Application Serial Number 62/007,235, entitled "Programmable Nucleotide Enzymes", filed June 3, 2014, which is incorporated herein by reference.

STATEMENT REGARDING FEDERALLY FUNDED RESEARCH

[001] NONE.

BACKGROUND

[002] DNA ligation is at the core of many recombinant DNA technologies including gene synthesis, cloning, DNA barcoding and tagging. DNA ligase, the enzyme that catalyzes formation of phosphodiester bonds between DNA fragments suffers from limitations, some of which are common to protein enzymes. These include sensitivity to temperature and instability. In addition DNA ligase requires the presence of overlapping complimentary regions between the fragments to be joined. The enzyme also suffers in its substrate flexibility. That is, its activity cannot be easily controlled or switched so as to perform ligations conditionally.

SUMMARY

[003] In one embodiment, principles of the present disclosure provide a composition comprising at least one or two or more nucleotide enzyme(s), wherein each of the nucleotide enzymes comprises at least a first and second target binding sequence and a catalytic sequence, wherein the first and second target binding sequences of the first and second nucleotide enzyme are substantially complementary to polynucleotide target sequences and wherein upon hybridization of polynucleotide target sequences with first and second target binding sequences of any of the nucleotide enzymes, the nucleotide enzyme is capable of catalyzing ligation of the first and second polynucleotide sequences.

[004] In one embodiment, principles of the present disclosure provide a method of assembling a plurality of polynucleotides comprising combining in a reaction vessel a plurality of polynucleotide target sequences with at least first and second nucleotide enzymes, wherein each of the nucleotide enzymes comprises at least a first and second target binding sequences and a catalytic sequence, wherein the first and second target binding sequences of the first nucleotide enzyme hybridize with substantially complementary sequences in first and second polynucleotide target sequences and wherein upon said hybridization the first and second polynucleotide target sequences are ligated with each other to form a first ligated polynucleotide.

[005] In one embodiment, principles of the present disclosure provide a method of assembling a plurality of polynucleotides comprising combining in a reaction vessel a plurality of polynucleotide target sequences with at least one, two or more nucleotide enzyme(s), wherein each of the nucleotide enzymes comprises at least a first and second target binding sequences, wherein the first target binding sequences hybridize with the 5' terminus of the first polynucleotide target sequence and the second target binding sequences hybridize with the 3' terminus of the second polynucleotide target sequence and a catalytic sequence, wherein the first and second target binding sequences of the first nucleotide enzyme hybridize with substantially complementary sequences in first and second polynucleotide target sequences and wherein upon hybridization the first and second polynucleotide target sequences are ligated with each other to form a first ligated polynucleotide.

[006] In one embodiment, principles of the present disclosure provide a method of assembling a plurality of polynucleotides comprising activating the 3' terminus of a plurality of target polynucleotides, combining in a reaction vessel a plurality of activated polynucleotide target sequences with at least one, two or more nucleotide enzyme(s), wherein each of the nucleotide enzymes comprises at least a first and second target binding sequences, wherein said first target binding sequences hybridize with the 5' terminus of the first polynucleotide target sequence and the second target binding sequences hybridize with the 3' terminus of the second polynucleotide target sequence; and a catalytic sequence, wherein the first and second target binding sequences of the first nucleotide enzyme hybridize with substantially complementary sequences in first and second polynucleotide target sequences and wherein upon said hybridization the first and second polynucleotide target sequences are ligated with each other to form a first ligated polynucleotide.

[007] In one embodiment, principles of the present disclosure provide a method for identifying the catalytic and substrate binding domains of a nucleotide enzyme involving aligning a library of variants of the nucleotide enzymes and identifying the substrate recognition domain by identifying the regions of the nucleotide enzymes that are complementary to the target polynucleotide.

[008] In one embodiment, principles of the present disclosure provide a method for designing a catalytic nucleic acid for ligating at least two nucleic acids comprising modularizing the catalytic nucleic acid into one or more catalytic domains and at least two substrate recognition domains, defining the catalytic domain as the region of the catalytic nucleic acid between at least two substrate recognition domains, defining the substrate recognition domains as the sequences that form complementary base pairs to substrate sequences to be joined, programming the catalytic nucleic acid to ligate two nucleic acids by designing at least one of the substrate recognition domains to be complementary to the terminus of one of the target nucleic acids and the second substrate recognition domain to be complementary to the terminus of the other target nucleic acid based on Watson-Crick base pairing.

[009] In addition, the method can include defining the substrate recognition domains by determining the optimal length of at least one of the substrate binding domains by a method comprising calculating the thermodynamic stability of Watson-Crick

pairs between the substrate recognition domains of a given length and the terminal regions of the target nucleic acid sequences, calculating the probability of secondary structures forming on the substrate recognition sequences, calculating the probability of intra-molecular interactions between the substrate recognition sequences and the catalytic domains, calculating the expected catalytic efficiency of the catalytic nucleic acid sequence given the sequence of the substrate recognition domain, calculating the error rate of the nucleic acid synthesizer for making the catalytic nucleic acid and/or calculating the cost of synthesizing the nucleic acid of a given length on the selected nucleic acid synthesizer

[010] It is contemplated that any embodiment of a method or composition described herein can be implemented with respect to any other method or composition described herein.

[011] The use of the word “a” or “an” when used in conjunction with the term “comprising” in the claims and/or the specification may mean “one,” but it is also consistent with the meaning of “one or more,” “at least one,” and “one or more than one.”

[012] The use of the term “or” in the claims is used to mean “and/or” unless explicitly indicated to refer to alternatives only or the alternative are mutually exclusive, although the disclosure supports a definition that refers to only alternatives and “and/or.”

[013] Throughout this application, the term "about" is used to indicate that a value includes the standard deviation of error for the device or method being employed to determine the value.

[014] As used in this specification and claim(s), the words "comprising" (and any form of comprising, such as "comprise" and "comprises"), "having" (and any form of having, such as "have" and "has"), "including" (and any form of including, such as "includes" and "include") or "containing" (and any form of containing, such as "contains" and "contain") are inclusive or open-ended and do not exclude additional, unrecited elements or method steps.

[015] Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating specific embodiments of the invention, are given by way of illustration only, since

various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

DESCRIPTION OF THE DRAWINGS

[016] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of the specification embodiments presented herein.

[017] FIG. 1 Diagram of the domains of a nucleotide ligase as described herein.

[018] FIG 2 Ligation reaction using two arbitrary DNA sequences, catalyzed by an E47-derived DNAzyme. One DNA fragment was labeled with FAM while the other was attached to biotin to allow tracking of the ligation reaction on a flow cytometer.

[019] FIG 3 Tracking the ligation rate of the E47-derived ligase programmed to join two pieces of DNA sequences encoding English sentences. Sequences of the re-programmed E47 ligase and substrates S1 and S2. S1 (Hello) was biotinylated (BiotinTEG) to allow attachment to streptavidin coated beads while S2 (World) was fluorescently labeled with 6-FAM.

[020] FIG 4 Ligation reactions were then monitored by injecting the reaction mixture at 30 minute intervals onto a flow cytometer. Only S1 beads onto which S2 (fluorescent) sequences have been ligated are detectable as fluorescent beads and can therefore be analyzed through gating. (B) Assessment of fluorescence on a FACS machine.

[021] FIG 5 is a flowchart and diagram showing parallel assembly of long DNA strands from short DNA sequences.

[022] FIG 6 is a flowchart and diagram showing parallel combinatorial and ordered assembly of DNA sequences.

[023] FIG 7 is a flowchart and diagram showing parallel assembly on DNA coated solid surfaces.

[024] FIG 8 is a flowchart and diagram showing a system for assessing DNAzyme performance, programmability and orthogonality using microarrays.

[025] FIG 9 is a diagram showing purification of ligation products.

[026] FIG 10 is a diagram showing parallel monitoring of nucleotide enzyme variants by sequencing.

[027] FIG. 11 shows a target sequence divided into smaller fragments, which are assembled by target sequence specific ligases to produce a desired target sequence.

[028] FIG 12 shows assembly of circular nucleotide products by the ligases disclosed herein.

[029] FIG 13 A. shows examples of nucleotide fragments of different length. B. shows an example of length of differentially assembled polynucleotides using different ligases. C. is a photo of an agarose gel showing differentially assembled polynucleotide products using different ligases.

[030] FIG 14 A. shows a cartoon of ligase mediated assembly of long DNA products. B. is a photo of an agarase gel showing assembly of DNA products of increasing size.

[031] FIG 15 A. shows a flow diagram of the method of the example. B. shows an image of GFP. C. An image of an agarose gel showing the showing the results of the example.

[032] FIG 16 A. Ligation of ssDNA fragments using phosporothioate-modified backbone ligases. Two 50 nucleotides long substrates (no modified, end modified or fully modified reactions 1, 2 and 3) were ligated into a product of 100 base pairs length by different modified backbone ligases. The product of theses ligations were compared to the product of a non modified ligase (Reactions 1,2 and 3 on the left). DNA fragments were resolved using TBE-UREA Polyacrylamide Gel Electrophoresis (PAGE). B. Ligation of ssDNA fragments using modified backbone ligases detected by real-time PCR. Two 50 nucleotides long substrates (no modified, end modified or fully modified reactions, were ligated into a product of 100 base pairs length by different modified backbone ligases. The product of these ligations were detected by real-time PCR that allows the quantification of the ligation product.

[033] FIG 17 Ligation of ssDNA and RNA using DNAzyme ligase or its RNA variant. Two 50 nucleotides long substrates, one of DNA and the other of RNA sequence,

were ligated to form a hybrid DNA-RNA product of 100 base pairs length by different modified backbone ligases. Reactions number 1 and 2 correspond to ligation of two DNA substrates by a DNA ligase and a RNA ligase respectively. Reactions 3 and 4 show DNA-RNA ligation products by DNA ligase and RNA ligase.

DESCRIPTION

[034] In view of the limitations of current DNA ligase enzymes, there exists a need for improved ligases. Accordingly, principles of the present disclosure provide novel nucleotide enzymes, such as novel nucleotide ligases and methods of their use.

[035] While DNA enzymes, such as the E47 DNA ligase have been described previously, development of related technologies has not advanced. For instance, previously disclosed nucleotide enzyme ligases were limited in their substrate specificity and developments in altering substrate specificity have languished. However, the present disclosure provides nucleotide enzymes with diverse substrate specificity that can be tailored or customized to ligate a target of choice. As such, nucleotides enzymes described herein may be described as "programmable" or "customizable" or "tailored" nucleotide enzymes. That is, nucleotide ligases as described herein bind to or hybridize with and ligate nucleotide target sequences different from E47. In view of the present disclosure it is appreciated that nucleotide ligases can be customized to hybridize with and ligate substrates of choice.

[036] By nucleotide ligase or DNA ligase or DNAzyme, as described herein is meant a nucleotide sequence having a first and second target binding sequence flanking a catalytic domain (Figure 1). The DNA ligase may be comprised of deoxyribonucleotides or ribonucleotides or combinations thereof, for instance a DNA-RNA hybrid, but also may contain or be made of modified nucleotides, such as, but not limited to phosphorothioate-modified backbones. The target binding sequences hybridize with target sequences that are complementary or substantially complementary to the target binding sequences. Target sequences are found in target molecules, generally separate target molecules, and upon hybridization with the first and second target binding sequences, the two target molecules are ligated together. Notably, this ligation requires only a single stranded target molecule and requires no overlapping regions of complementarity between the two target sequences. As such, nucleotide enzymes can be designed to ligate any two target

molecules of interest allowing for the first time protein-free, multiplex ligation of a plurality of molecules in a single reaction vessel. Moreover, ligation reactions are not limited by the size of the target molecule. Thus, very large substrates can be targeted for ligation and larger products generated. Thus, the present disclosure provides methods for sequential ligation of a plurality of target sequences to generate products larger than could be achieved by prior nucleotide enzymes.

[037] Another notable feature of nucleotide ligases is that the product formed can have a 2'-5'-phosphodiester bonds instead of 3'-5'- phosphodiesters. This provides numerous advantages, such as being resistant to nucleases while still being recognized by DNA polymerase.

[038] Accordingly, in one embodiment, the nucleotide ligases described herein may ligate any of a variety of nucleotide molecules. For instance, the disclosure need not be limited to ligation of DNA molecules. Rather, other nucleotides, such as RNA, LNA, PNA ssDNA, RNA-DNA hybrids, modified or capped RNA, and the like may be ligated. Other nucleotides that may be ligated include, but are not limited to XNAs where where X can be one of different types of sugars:1,5-anhydrohexitol nucleic acids (HNAs), cyclohexenyl nucleic acids (CeNAs), 2'-O,4'-C-methylene-beta-D-ribonucleic acids [locked nucleic acids (LNAs)], arabinonucleic acids (ANAs), 2'-fluoro-arabinonucleic acids (FANAs), alpha-L-threofurnaosyl Nucleic Acids (TNAs).

[039] As noted above, the nucleotide enzymes of the present disclosure comprises first and second target binding sequences. These regions hybridize with complementary or substantially complementary target sequences in target molecules. A benefit of the nucleotide enzymes described herein is that the target binding sequences are designed and may be of different lengths. That is, in some embodiments each of the target binding sequences may have as few as 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 nucleotides. In one embodiment, the target binding sequences of the nucleotide enzyme are up to 20 or 30 or 40 or 50 or more nucleotides in length. In some embodiments the target binding sequences are from 3 to 50, 4 to 40, 5 to 30, 6 to 25, or 8 to 10 nucleotides in length.

[040] The nucleotide ligases described herein also comprise a catalytic domain between the two target binding sequences.

[041] Additionally, the catalytic nucleotide ligases are generally metalloenzymes. That is, they require metal co-factors for enzymatic activity. Other metals include Zn^{2+} and/or Cu^{2+}, Ag(I), Pb(II), Hg(II), As(III), Fe(III), Zn(II), Cd(II), Cu(II), Sr(II), Sa(II), Ni(II), Co(II), As(V), U(VI), Cr(VI), Ca(II), Mg(II).

[042] As described above substrates can be RNA, DNA, hybrids of RNA-DNA, modified nucleotides and the like. To facilitate ligation of two target nucleotides, the substrates should be activated. Specifically, the 3' phosphate of at least one of the substrates should be activated for the ligation reaction to proceed. In one embodiment the substrates may be activated by phosphocyanate (cyanogen bromide), azide-alkyne (triazole linkage), phosphorimidazolide, including 3'-phosphoramidazolide, 3'-phosphorothioate, 5'-iodide, 5'-bromoacetyl, 5'-tosyl, phosphoramidate, boranophosphate linkages, peptide nucleic acid linkages (PNAs), locked nucleic acid linkages (LNAs), xeno nucleic acid linkage (XNAs), and morpholino linkage. Reaction classes that are applicable in activating nucleic acid templates or can be catalyzed by the nucleotide ligase include, but are not limited to reductive amination, amine acylation, acyl transfer, S_N2 reaction, conjugation addition, Henry reaction, Nitro-Michael reaction, Wittig reaction, 1,3-nitrone cycloaddition, Huisgen cycloaddition, Oxazolidine formation, Heck coupling, Cross coupling, Me-salen formation, Aldol Reaction (e.g. using DNA-linked glyceraldehyde and glycoladehyde)

[043] As such, the ligation reaction includes at least one, two or more nucleotide ligase(s), metal co-factors if necessary, and activated substrates. Once combined, the reaction continues at for at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 60, 120, 240, 360, 480 minutes or more. In one embodiment, the reaction occurs at room temperature. In other embodiments, the reaction occurs at 37°C. In some embodiments the nucleotide fragments may serve as indicators of the presence of an organism or other nucleotides. In some embodiments the two target nucleotides are ligated to assemble a functional ligation product, such as a chimeric polynucleotide encoding a protein or RNA molecule. In addition, the two target polynucleotides may assemble nucleotide regulatory regions, such as promoters/enhancers with coding sequences to be expressed or under the control of the promoter/enhancer.

[044] A benefit of using nucleotide ligases as described herein is the ability to perform simultaneous, multiplex sequence-specific ligation in a single reaction vessel.

First and second different nucleotide ligases are included in a single reaction vessel and perform at least two sequence specific ligations yielding two distinct products. Further, the reaction may include additional nucleotide ligases to produce a plurality of ligation products in a single reaction. In this embodiment, at least 5, 10, 15, 25, 50, 75, 100, 150, or more different nucleotide ligases are included in a single reaction to produce the plurality of ligation products. That is, up to 1000, 2000 or more ligation reactions can occur in a single reaction vessel. Moreover, following a first ligation reaction, substrates can be activated resulting in a second ligation reaction using the first set of nucleotide ligases. The cycle may continue until the appropriate number of ligations has occurred. By using the ligases described herein products of at least 1kb or 10kb or 100 kb or more can be assembled.

[045] The stability and flexibility provides for numerous uses for nucleotide ligases. For instance the ligases can be used in a variety of assays to determine if two target nucleotides are present in a sample such as a biological sample. By "biological sample" is meant any bodily fluids (including, but not limited to, blood, urine, serum, lymph, saliva, anal and vaginal secretions, perspiration and semen, of virtually any organism, with mammalian samples being preferred and human samples being particularly preferred); biopsy/tissue material; environmental samples (including, but not limited to, air, agricultural, water and soil samples); biological warfare agent samples; research samples; purified samples, such as purified genomic DNA, RNA, proteins, etc.; raw samples (bacteria, virus, genomic DNA, etc.). As will be appreciated by those in the art, virtually any experimental manipulation may have been done on the sample.

[046] In this embodiment, nucleotide enzymes are designed to ligate two nucleotide target sequences anticipated to be in the sample. Following the reaction, the longer, ligated product is detected by methods known in the art. Production of the ligation product is evidence of the presence of the two target nucleotide sequences.

[047] In some embodiments, such as diagnostic methods, one member of a target sequence is immobilized on a solid support while the other may be in solution, for instance in a biological sample. In this embodiment, the nucleotide ligase contains one target binding sequence complementary to the immobilized target and one target binding sequence complementary to the sequence to be detected in the

sample. The presence of the target sequence in the sample is then confirmed upon ligation of the two target nucleotide sequences by methods known in the art.

[048] In some embodiments the immobilized target nucleotide sequences can be the same yet in other embodiments they can be different. When the immobilized target are the same, the nucleotide ligases contain a first target binding sequence that also is the same, while the second target binding sequence of different nucleotide enzymes will be different. This difference allows the nucleotide enzymes to bind to different nucleotides to be detected.

[049] Detection of ligation products may be facilitated by incorporation of a label in to the product. This may be done in a variety of ways known in the art. By "detection label" or "detectable label" herein is meant a moiety that allows detection. This may be a primary label or a secondary label. Accordingly, detection labels may be primary labels (i.e. directly detectable) or secondary labels (indirectly detectable).

[050] In one embodiment, the detection label is a primary label. A primary label is one that can be directly detected, such as a fluorophore. In general, labels fall into three classes: a) isotopic labels, which may be radioactive or heavy isotopes; b) magnetic, electrical, thermal labels; and c) colored or luminescent dyes. Labels can also include enzymes (horseradish peroxidase, etc.) and magnetic particles. Other labels include chromophores or phosphors but are preferably fluorescent dyes. Suitable dyes for use in the invention include, but are not limited to, fluorescent lanthanide complexes, including those of Europium and Terbium, fluorescein, rhodamine, tetramethylrhodamine, eosin, erythrosin, coumarin, methyl-coumarins, quantum dots (also referred to as "nanocrystals"), pyrene, Malacite green, stilbene, Lucifer Yellow, Cascade Blue™, Texas Red, Cy dyes (Cy3, Cy5, etc.), alexa dyes, phycoerythin, bodipy, and others described in the 6th Edition of the Molecular Probes Handbook by Richard P. Haugland, hereby expressly incorporated by reference. Carbon nanotubes also find use as detection agents as outlined in Wang *et al. Analyst*, 2012, 137, 3667. Or fluorescent metal indicators such as calcein Tomita et al Nature Protocols 2008 (3) 877-882 or colorimetric metal indicators such as Hydroxy Napthol Blue (HNB) Goto et al Biotechniques 2009 (46) 167-172, both of which are incorporated herein by reference, find use in embodiments described herein.

[051] In one embodiment, a secondary detectable label is used. A secondary label is one that is indirectly detected; for example, a secondary label can bind or react with a

primary label for detection, can act on an additional product to generate a primary label (e.g. enzymes), or may allow the separation of the compound comprising the secondary label from unlabeled materials, etc. Secondary labels find particular use in systems requiring separation of labeled and unlabeled probes, such as SBE, OLA, invasive cleavage reactions, etc; in addition, these techniques may be used with many of the other techniques described herein. Secondary labels include, but are not limited to, one of a binding partner pair; chemically modifiable moieties; nuclease inhibitors, enzymes such as horseradish peroxidase, alkaline phosphatases, luciferases, streptavidin (when primary label is biotin), etc.

[052] In one embodiment, the secondary label is a binding partner pair. For example, the label may be a hapten or antigen, which will bind its binding partner. In one embodiment, the binding partner can be attached to a solid support to allow separation of extended and non-extended primers. For example, suitable binding partner pairs include, but are not limited to: antigens (such as proteins (including peptides)) and antibodies (including fragments thereof (FAbs, etc.)); proteins and small molecules, including biotin/streptavidin; enzymes and substrates or inhibitors; other protein-protein interacting pairs; receptor-ligands; and carbohydrates and their binding partners. Nucleic acid-nucleic acid binding proteins pairs are also useful. In general, the smaller of the pair is attached to the NTP for incorporation into the primer. Binding partner pairs include, but are not limited to, biotin (or imino-biotin) and streptavidin, digeoxinin and Abs, and Prolinx^{TM} reagents (see www.prolinxinc.com/ie4/home.hmtl).

[053] In one embodiment, the binding partner pair comprises biotin or imino-biotin and streptavidin. Imino-biotin is useful as imino-biotin disassociates from streptavidin in pH 4.0 buffer while biotin requires harsh denaturants (e.g. 6 M guanidinium HCl, pH 1.5 or 90% formamide at 95^{o}C).

[054] In one embodiment, the binding partner pair comprises a primary detection label (for example, attached to the NTP and therefore to the extended primer) and an antibody that will specifically bind to the primary detection label. By "specifically bind" herein is meant that the partners bind with specificity sufficient to differentiate between the pair and other components or contaminants of the system. The binding should be sufficient to remain bound under the conditions of the assay, including wash steps to remove non-specific binding. In some embodiments, the dissociation constants of the pair will be less than about 10^{-4}-10^{-}

^{6} M^{-1}, with less than about 10^{-5} to 10^{-9} M^{-1} being preferred and less than about 10^{-7}-10^{-9} M^{-1} being particularly preferred.

[055] In one embodiment, the secondary label is a chemically modifiable moiety. In this embodiment, labels comprising reactive functional groups are incorporated into the nucleic acid. The functional group can then be subsequently labeled with a primary label. Suitable functional groups include, but are not limited to, amino groups, carboxy groups, maleimide groups, oxo groups and thiol groups, with amino groups and thiol groups being particularly preferred. For example, primary labels containing amino groups can be attached to secondary labels comprising amino groups, for example using linkers as are known in the art; for example, homo- or hetero-bifunctional linkers as are well known (see 1994 Pierce Chemical Company catalog, technical section on cross-linkers, pages 155-200, incorporated herein by reference).

[056] In some embodiments, the reactions can occur in Eppendorf tubes, microtiter plates, ordered arrays, random bead arrays or other substrates as are known in the art. Detection may occur by any of a number of methods, including but not limited to placement on an ordered or random array, analysis of a multi-well plate, FACS, electrophoretic, spectrophotometric, colorimetric analysis, and the like.

[057] By "nucleic acid" or "oligonucleotide" or grammatical equivalents herein means at least two nucleotides covalently linked together. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, as outlined below, nucleic acid analogs are included that may have alternate backbones, comprising, for example, phosphoramide (Beaucage et al., Tetrahedron 49(10):1925 (1993) and references therein; Letsinger, J. Org. Chem. 35:3800 (1970); Sprinzl et al., Eur. J. Biochem. 81:579 (1977); Letsinger et al., Nucl. Acids Res. 14:3487 (1986); Sawai et al, Chem. Lett. 805 (1984), Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); and Pauwels et al., Chemica Scripta 26:141 91986)), phosphorothioate (Mag et al., Nucleic Acids Res. 19:1437 (1991); and U.S. Pat. No. 5,644,048), phosphorodithioate (Briu et al., J. Am. Chem. Soc. 111:2321 (1989), O-methylphophoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm, J. Am. Chem. Soc. 114:1895 (1992); Meier et al., Chem. Int. Ed. Engl. 31:1008 (1992); Nielsen, Nature, 365:566 (1993); Carlsson et al., Nature 380:207 (1996), all of which are

incorporated by reference). Other analog nucleic acids include those with positive backbones (Denpcy et al., Proc. Natl. Acad. Sci. USA 92:6097 (1995); non-ionic backbones (U.S. Pat. Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; Kiedrowshi et al., Angew. Chem. Intl. Ed. English 30:423 (1991); Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); Letsinger et al., Nucleoside & Nucleotide 13:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook; Mesmaeker et al., Bioorganic & Medicinal Chem. Lett. 4:395 (1994); Jeffs et al., J. Biomolecular NMR 34:17 (1994); Tetrahedron Lett. 37:743 (1996)) and non-ribose backbones, including those described in U.S. Pat. Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within the definition of nucleic acids (see Jenkins et al., Chem. Soc. Rev. (1995) pp 169-176). Nucleic acids capped with protective residues (Yisraeli, J.K., and Melton, D.A. (1989). Synthesis of long, capped transcripts in vitro by SP6 and T7 RNA polymerases. Methods Enzymol. 180; Warren et al, Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA, Cell Stem Cell, Volume 7, Issue 5, 5 November 2010, Pages 618-630). Several nucleic acid analogs are described in Rawls, C & E News Jun. 2, 1997 page 35. All of these references are hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone may be done to facilitate the addition of labels, or to increase the stability and half-life of such molecules in physiological environments.

[058] As will be appreciated by those in the art, all of these nucleic acid analogs may find use in the present invention. In addition, mixtures of naturally occurring nucleic acids and analogs can be made. Alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made.

[059] Peptide nucleic acids (PNA), which include peptide nucleic acid analogs also find use in the present disclosure. These backbones are substantially non-ionic under neutral conditions, in contrast to the highly charged phosphodiester backbone of naturally occurring nucleic acids. This results in two advantages. First, the PNA backbone exhibits improved hybridization kinetics. PNAs have larger changes in the melting temperature (Tm) for mismatched versus perfectly matched basepairs.

DNA and RNA typically exhibit a 2-4°C drop in Tm for an internal mismatch. With the non-ionic PNA backbone, the drop is closer to 7-9°C. This allows for better detection of mismatches. Similarly, due to their non-ionic nature, hybridization of the bases attached to these backbones is relatively insensitive to salt concentration.

[060] The nucleic acids may be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid contains any combination of deoxyribo- and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xathanine hypoxathanine, isocytosine, isoguanine, etc. One embodiment utilizes isocytosine and isoguanine in nucleic acids designed to be complementary to other probes, rather than target sequences, as this reduces non-specific hybridization, as is generally described in U.S. Pat. No. 5,681,702. As used herein, the term "nucleoside" includes nucleotides as well as nucleoside and nucleotide analogs, and modified nucleosides such as amino modified nucleosides. In addition, "nucleoside" includes non-naturally occurring analog structures. Thus for example the individual units of a peptide nucleic acid, each containing a base, are referred to herein as a nucleoside. In embodiments when target nucleic acids are found in double stranded nucleic acids the samples may be treated to disrupt the double strand so that the single stranded polynucleotides may serve as templates for nucleotide ligation.

[061] In some embodiments the disclosure provides methods for identifying catalytic and/or substrate binding domains from a plurality of different sequences. In this embodiment, a plurality of sequences are aligned with known DNAzymes and conserved catalytic regions are identified. In addition, upon identification of the catalytic regions, the location and identity of first and second substrate binding domains can be identified based on their proximity to the catalytic domain.

[062] In some embodiments, DNAzymes or nucleotide enzymes are specifically designed or programmed. In this embodiment, substrate-binding domains of a nucleotide enzyme are designed to be complementary to substrates with which they hybridize. Factors influencing hybridization include degree of complementarity, length of complementarity region, nucleotide composition of the hybridizing regions, secondary structure and other factors as known in the art.

Accordingly, the present disclosure also provides methods of designing an library of nucleotide enzymes or DNAzymes that can be used to ligate a plurality of target molecules.

[063] In some embodiments designing the DNAzymes or nucleotide enzymes includes modularizing the catalytic nucleic acid into one or more catalytic domains and at least two substrate recognition domains. Catalytic domains can be identified by their characteristic sequence of 5'-CCTGTTTCATGAGACCATGTGACGCATGGCCCG-3' (SEQ ID NO: 1). Variants, however, can be identified by methods disclosed herein. Substrate binding (also referred to as substrate recognition) domains found 5' and 3' of the catalytic domain. As described above, the substrate binding domains can be designed to be complementary to any target of choice and thus provide the ability for a DNAzyme to ligate any two nucleotides together. As such, the method provides programming the catalytic nucleic acid to ligate two nucleic acids by designing at least one of the substrate recognition domains to be complementary to the terminus of one of the target nucleic acids and the second substrate recognition domain to be complementary to the terminus of the other target nucleic acid, for instance, based on Watson-Crick base pairing.

[064] In some embodiments, the substrate binding domains are of the same or different length within a DNAzyme or within a set of DNAzymes. In one embodiment, designing DNAzymes, therefore, includes defining the optimal length of the substrate binding domain. Factors to consider and to calculate when designing optimal lengths include calculating the thermodynamic stability of Watson-Crick pairs between the substrate recognition domains of a given length and the terminal regions of the target nucleic acid sequences, calculating the probability of secondary structures forming on the substrate recognition sequences, calculating the probability of intra-molecular interactions between the substrate recognition sequences and the catalytic domains, calculating the expected catalytic efficiency of the catalytic nucleic acid sequence given the sequence of the substrate recognition domain, calculating the error rate of the nucleic acid synthesizer for making the catalytic nucleic acid; and/or calculating the cost of synthesizing the nucleic acid of a given length on the selected nucleic acid synthesizer.

[065] In one embodiment, the method provides method for designing a set of orthogonal catalytic nucleic acids to perform parallel ligation of more than two pre-specified

nucleic acid fragments. This method includes selecting or designing a set of catalytic nucleic acids as described above. In some embodiments the catalytic nucleic acids are analyzed and designed based on Smith-Waterman or Needleman-Wunsch alignments to avoid cross-hybridizations.

[066] In some embodiments, principles of the present disclosure provide a method for determining or designing orthogonal sets of catalytic nucleic acids by parallel monitoring of the catalytic efficiency of a plurality of catalytic nucleic acids. In some embodiments, the number of catalytic nucleic acids to be monitored can be at least 10,000, 100,000, 1,000,000, 10^7, 10^8, 10^9, 10^{10}, 10^{11}, 10^{12}, 10^{13}, 10^{14}, 10^{15} or more. According to this method, each catalytic nucleic acid represents substrate recognition or catalytic domain variants of various lengths and nucleotide content, the catalytic nucleic acids are synthesized by adding activated nucleotides to a linker sequence on a solid surface such as microarray, in some embodiments, one of the 2 substrate recognition domains in each catalytic nucleic acid has a self-complimentary region in the 5'-end (Figure 8), the substrate nucleic acid sequence complimentary to the 3'-substrate recognition domain is labeled, for instance, by a fluorescent dye such as, but not limited to Cy-5 or Cy-3 by methods known in the art. The catalytic nucleic acids are monitored by first activating the substrate nucleic acid sequences as described herein, followed by incubation on a microarray chip of for a desired period of time after which the microarray is washed under denaturing conditions, scanned and processed by image segmentation. See Figure 8 and Example 6.

[067] In one embodiment, principles of the present disclosure provide a method for determining catalytic activity of multiple catalytic nucleic acids by deep DNA sequencing. In this embodiment one-pot reactions in which multiple variants of catalytic nucleic acids and activated template substrates as described herein are mixed together and incubated. The reaction mixture is then sequenced after incubation at high depth, for example, 10X, 20X, 30X, 50X, 100X, 1000X or more. The read count or coverage depth of each ligated sequence species is then used as a measure of the catalytic activity of the corresponding nucleotide ligase variant containing complementary substrate recognition sequences. See Example 8 and Figure 10. Example 7 and FIG 9 provide examples of methods to purify reaction products. This approach is also applicable in profiling the activity of other nucleic acid based enzymes such as programmable nucleases (CRISPR/Cas9)

[068] Once the products are sequenced, statistical and machine learning methods are used for decoding the DNA sequence information into catalytic activity of individual catalytic nucleic acid. In one embodiment, methods for statistical and machine learning include obtaining and analyzing nucleic acid sequencing information including but not limited to sequence coverage, sequence read depth, consensus and error rate.

[069] In one embodiment, principles of the present disclosure provide a method for selecting an optimal set of catalytic nucleic acids from an undirected graph or network in which nodes are catalytic nucleic acids of varying substrate recognition sequences and edges between nodes represent a lack of complementarity between the recognition sequences of the nodes and optimal catalytic efficiency at a given temperature range. In one embodiment, the method involves finding the shortest set of paths passing through nodes whose sequence is complementary to that of the set of termini to be ligated in parallel; in other embodiments, the method involves finding clusters or communities or cliques of nodes in the graph representing orthogonal sets of catalytic nucleic acids.

[070] Utilization of a graph-theoretic depiction of orthogonal relationships between DNAzymes in determining the optimal set of DNAzymes to ligate a given set of templates. In the example given, the goal is to find the set of DNAzymes that can operate in parallel to ligate templates 1, 2, 3 and 4. To do this, the termini of pairs of consecutive templates are examined for sequences that are complementary to orthogonal DNAzymes using the graph shown in Figure 8. Each circle (node) in the graph represents a DNAzyme that is complementary to and ligates a specific pair of DNA sequences. Two DNAzymes are connected in the graph if minimally they are not complementary to each other and can operate at the same temperature. The graph shown is only representative since there are multiple variants of DNAzymes that can exist. In the basic implementation, the graph has 65,000 nodes (DNAzyme variants) whose orthogonal relationships have been determined by methods described herein. For example, templates 1 and 2 (green) can be joined by the green DNAzyme (green circle) while templates 2 and 4 can be joined by the orange DNAzyme (orange circle) since these 2 DNAzymes are connected in the graph by a short path (direct connection). In some embodiments, the graph consists of weighted edges (connections) between nodes (DNAzymes) in which a high weight reflects high orthogonality and a low weight corresponds to low orthogonality between the DNAzymes. When the connections are

weighted, the shortest path is defined as one which connects a set of DNAzymes through edges whose combined weights are the largest of all possible paths connecting DNAzymes complementary to the target sequences.

[071] Given a set of nucleic acid templates to be ligated, a computational algorithm for finding an optimal set of DNAzymes can be implemented as follows: For each of those fragments, the algorithm examines their junctions and determines the activity of ligase variants complimentary to those junctions using the ligase activity data generated from methods of described herein. The algorithm then selects a subset of fragments whose junctions have the highest activity relative to those of other fragments. It then evaluates the orthogonality of catalytic nucleic acids complementary to the termini of each of the fragments using multiple criteria: i) sequence alignments- i.e no DNAzyme is the selected set should be complementary to another DNAzyme in the same set, ii) the ligase activity of the selected set should be at the highest level at the same temperature, iii) the set of DNAzymes should have similar melting profiles.

[072] Given a target assembly problem, all fragments of the sequences (that can be synthesized accurately) can be generated *in silico* and mapped onto this graph. The shortest path connecting all nodes onto which the fragments map can then be selected as the set of orthogonal DNAzymes for the given assembly problem. The short path in the graph is one in which all the set of DNAzyme that can potentially ligate the desired fragments are directly connected to each other.

[073] In one embodiment, principles of the present disclosure provide a method for *in silico* decomposition of a long nucleic acid sequence into an optimal set of fragments that can be synthesized accurately and ligated by methods described herein using nucleotide enzymes. In this embodiment, considerations include the length of each nucleic acid fragment, which is a function of the synthetic capability and error profile of a given nucleic acid synthesizer, the length and number of fragments, which are a function of the cost and time of their synthesis on a given nucleic acid synthesizer , the fact that the junctions to be joined have minimal secondary structure, the termini of the fragments are unique based on Smith-Waterman alignments and the pair of any contiguous termini are unique.

[074] In one embodiment, principles of the present disclosure provide a method for designing a set of orthogonal catalytic nucleic acids for ligating modular nucleic acid sequences including but not limited to antibody or T-cell receptor VDJ

segments, TALEN repeat domains, ZFN domains, promoter elements, ribosome binding sites, signal peptides, protein sorting signals, exons, protein domains and libraries of standard biological parts such as BioBricks.

[075] In one embodiment, principles of the present disclosure provide a method for determining the biological effects of unnatural ligation junctions using single molecule approaches such as single molecule real-time sequencing inter-pulse duration to estimate the speed and processivity of nucleic acid polymerases on encountering such junctions.

[076] In one embodiment, principles of the present disclosure provide a method for predicting the location of an unnatural backbone in a given nucleic acid using a signature of nucleic acid poylmerase kinetics obtained from single molecule real-time sequencing of nucleic acids with unnatural backbones. The method is applicable in sequencing the nucleic acid backbone- the process of determining the identity of nucleic acid backbone and distinct from the conventional nucleic acid sequencing procedures which determine identify of nucleic acid bases.

[077] In one embodiment, principles of the present disclosure provide a method for minimizing the likelihood of nucleic acid polymerase stalling at ligation junctions due to unnatural 2',5'-phosphodiester linkage or other alternative nucleic acid backbones formed by some catalytic nucleic acids. The method involves the incorporation of measured polymerase kinetics using single molecule real time sequencing and specifically avoids the formation of ligation junctions at nucleotide sequences associated with polymerase stalling or reduced speed or polymerase processivity.

[078] In one embodiment, principles of the present disclosure provide a method for minimizing the biological effects of unnatural phosphodiester linkages that might be introduced by some catalytic nucleic acids by avoiding or only allowing for the lowest possible number of such linkages within biologically active nucleic acid motifs such as transcription factor binding sites, transcription start sites (TSS), ribosome binding sites (RBS), catalytic domains of catalytic nucleic acids, ligand recognition domains in aptamers.

[079] In one embodiment, principles of the present disclosure provide a method for 'healing' or correcting the unnatural linkages that might be introduced into the phosphodiester backbone by some catalytic nucleic acids involving copying the

nucleic acid using nucleic acid copying enzymes such as polymerases, reverse transcriptases or replicases or using DNA repair enzymes such as nucleotide excision repair enzymes or by the replication of the nucleic acid in cells.

[080] In one embodiment, principles of the present disclosure provide a method for altering the nuclease resistance of a nucleic acid sequence by using a catalytic nucleic acid and a combination of templates whose ligation results into unnatural phosphodiester linkages at sequence motifs that are the targets of nucleases such as but not limited to restriction endonuclease, Zinc Finger Nuclease (ZFN), TALEN, CRISPR, DNAse, RNAse and nickase.

[081] In one embodiment, principles of the present disclosure provide a method for altering the hybridization affinity or specificity or thermal stability or radiation resistance or in vivo half life or alkaline/ acidic resistance or ionic state or electrophoretic mobility or electric field effect or molecular weight or volume or charge of catalytic and non-catalytic nucleic acids by replacing their nucleotides with unnatural backbones such as but not limited to 2',5'-phosphodiester linkage, PNAs, LNAs, XNAs, TNAs, HNAs, FANAs and CeNAS.

[082] In one embodiment, principles of the present disclosure provide a method for altering the protein binding affinity of a nucleic acid by using catalytic nucleic acids to introduce unnatural linkages through ligation of templates within junctions of the nucleic acid domains that mediate protein binding or within motifs that bind sequence specific nucleic acid binding proteins such as but not limited to transcription factors, ribosomes, RNA binding proteins, nucleosomes, mRNA splicing machinery.

[083] In one embodiment, principles of the present disclosure provide a method for introducing additional chemical groups (fluorophores, biotin and chemical groups for attachment to solid surfaces) to specific junctions in the nucleic acid backbone using catalytic nucleic acids. The method enables programmable modifications of the nucleic acid backbone that can produce a molecular barcode of a nucleic acid sequence.

[084] In one embodiment, principles of the present disclosure provide a method for ligating nucleic acid templates containing unnatural backbones, or modified side-chains that cannot be tolerated by natural protein based DNA ligases or

chemically synthesized under the same reaction conditions. The method involves using catalytic nucleic acid with ligase activity to join such templates.

[085] In one embodiment, principles of the present disclosure provide a method for the purification or recovery or recycling of catalytic nucleic acids involving conjugation of the nucleic acids to biotin- for streptavidin separation or magnetic beads or gold nanoparticles or affinity purification using complementary sequences to the catalytic nucleic acids.

[086] In one embodiment, principles of the present disclosure provide a method for enhancing nucleic acid templated reactions between two molecules in which one of the molecules is conjugated to a nucleic acid sequence complementary to the 5' substrate recognition sequence of the catalytic nucleic acid, the second molecule is conjugated to a nucleic acid complementary to the 3' substrate recognition sequence of the catalytic nucleic acid, the two molecules to be joined are brought together by the formation of complementary interactions between the substrate recognition domains of the catalytic nucleic acid in the presence of a metal ion and the nucleic acids to which the molecules are conjugated.

[087] In one embodiment, principles of the present disclosure provide a method for improving or fine-tuning the catalytic activity of catalytic nucleic acids using data generated from parallel monitoring of catalytic nucleic acid variants and sequencing methods as described herein. In one instance of the method, catalytic nucleic acid variants are generated by random nucleic acid substitutions, insertions or deletions in the catalytic domain sequence of catalytic nucleic acids known in the art. In another instance, catalytic nucleic acid variants are generated by random nucleic acid substitutions, insertions or deletions in the substrate recognition domains of catalytic nucleic acids.

[088] In one embodiment, principles of the present disclosure provide a method of constructing or assembling functional DNA constructs. For instance, the method provides for construction of plasmids, antisense oligos, transcription factor decoys, siRNA and/or aptamers. In one embodiment, the sequence of the desired construct is known and can be assembled by dividing the sequence into smaller fragments. Fragments may range from 10-500 nucleotides or 50-250 nucleotides or 100-200 nucleotides in length. The fragments are then assembled by ligases designed to ligate particular target sequences which results in assembly of the desired functional construct. (Figure 11). In some embodiments the assembled

product is a circular product (Figure 12). This embodiment finds particular use in assembling plasmid constructs from polynucleotides having different functions, such as transcription factor binding sites, promoters, inserts, antibiotic resistance and the like. Particular benefits include that no special equipment is necessary, there is no need to remove viral or bacterial sequences and the assembly can occur at ambient temperature.

[089] While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and/or methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the present invention.

EXAMPLES

[090] The following examples are included to demonstrate certain embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLE 1
NUCLEOTIDE LIGASE CATALYZED ASSEMBLY OF TARGET DNA

[091] E47 ligase:

[092] 5-CGGATAGTGTTCTTTCGCTAGACCATGTGACGCATGGTGAGATGCTT-3 (SEQ ID NO:2)

[093] Substrate 1 (S1): 5-AAGCATCTCAAGC-3 (SEQ ID NO:3)

[094] Substrate 2 (S2): 5-GGAACACTATCCG-3 (SEQ ID NO:4)

IMIDAZOLE ACTIVATION OF S1

[095] Substrate 1 was first prephosphorylated at the 3' end (Integrated DNA Technologies). To add the imidazole group to the activated 3'phosphate group, 20 uL of 100 uM Substrate 1 with 2.5 uL of 1M imidazole (pH 6.0) and 4.5 uL of 1M EDC.HCL was incubated at room temperature for 1 hour. The mixture was then purified by a desalting column (PD-10, Amersham Biosciences) and the first fraction collected. Figure 1 depicts the E47 nucleotide ligase hybridized to target DNA.

E-47 CATALYZED LIGATION

[096] The activated Substrate 1 was mixed with the E-47 ligase (2 uM), Cu(N03)2 (20 uM) and S2 (3 uM) in 25 mM HEPES (pH 7.0) containing 300 mM NaCl. The solution is incubated for 4 hours at room temperature.

ASSESSMENT OF LIGATION

[097] The success of the ligation reaction was observed using FACS. Fig. 2 depicts results demonstrating increased ligation of two nucleotides. Fluorescent measurement of ligated DNA molecules was obtained after exposure of a DNA sequence tagged with a polystyrene bead (biotin labeled) and another sequence labeled with fluorescent dye FAM. This method allowed detection of only ligated DNA molecules because they have both the polystyrene bead and FAM dye which was detected by fluorescent activated cell sorter (FACS).

EXAMPLE 2: ASSESSMENT OF LIGATION

[098] A careful study of the E47 DNAzyme structure and sequence variants was performed. This revealed high modularity: the catalytic domain includes two sub-domains, non-conserved sequences that are complementary to the substrate, referred to as substrate recognition domains. Their existence strongly suggests that the ligase can be programmed. Software was written that takes the target DNA sequences as input, and redesigns a new catalytic nucleic acid ligase to join them. This was used to generate a ligase for DNA sequences encoding two short English sentences (denoted "Hello" and "World" for convenience, each approximately 30 nucleotides long) via simple ASCII to DNA conversion. To track the activity of the ligase using flow cytometry, the "Hello" component was attached to polystyrene beads through a biotin-streptavidin interaction while the "World" component was attached to a fluorophore (Figure 3A). The ligase was then

incubated with the target DNA sequences for variable periods of time. The reaction was monitored in a fluorescent activated cell sorting (FACS) machine (Figure 3B). This experimental design enabled us to read-out the ligated DNA sequences ("Hello World") based on fluorescence in a special region in the forward (FSC) vs. side (SSC) scatter outputs of the FACS machine. In summary, when the DNA sequences "Hello" and "World" are absent, the ligase remains inactive (time 0). When the sequences are present, the ligase becomes active and joins the input sequences. This experiment demonstrates that the catalytic nucleic acid ligase can be redesigned without damaging its activity – it is robust to specific changes in the substrate recognition domains.

EXAMPLE 3: PARALLEL ASSEMBLY OF LONG DNA STRANDS FROM SHORT DNA SEQUENCES

[099] In this example, the target DNA sequences are first fragmented into an optimal set of fragments that can be ligated efficiently in parallel by a set of DNAzymes (Figure 5).

EXAMPLE 4: PARALLEL COMBINATORIAL AND ORDERED ASSEMBLY

[0100] In this example, a library of promoters is fused with a library of open reading frames in a single reaction vessel by nucleotide ligases described herein (Figure 6). The resulting ligation products are used in a variety of assays while the nucleotide ligases may be recycled.

EXAMPLE 5: PARALLEL ASSEMBLY ON DNA COATED SOLID SURFACES (REPROGRAMMING)

[0101] Embedding DNA on solid surfaces has numerous applications including generation of microarrays, DNA coated affinity columns and 3-dimensional surfaces such as beads or even other biological molecules (proteins, DNA encoded antibody libraries). The ability to modify DNA sequences already embedded on solid surfaces eliminates the need to know the sequences of DNA that need to be detected by microarrays at the time of printing. That is, instead of printing specific microarray designs for each purpose, one can pre-print a 'universal' microarray that can be reprogrammed to detect new DNA sequences. The reprogramming involves ligation of desired DNA sequences of interest onto the DNA probes pre-embedded on the microarray. This can be extremely useful in the case of a pandemic. For example, in the case of a totally new outbreak, the universal

microarray could be reprogrammed by simply ordering oligos specific to the new pathogen and using a pre-designed set of nucleotide enzymes (Figure 7).

EXAMPLE 6: A SYSTEM FOR ASSESSING DNAZYME PERFORMANCE, PROGRAMMABILITY AND ORTHOGONALITY USING MICROARRAYS.

[0102] To monitor catalytic activity of several DNAzymes (red) in parallel, thousands of variants are synthesized on a microarray chip. Each of these DNAzymes has a self-complementary region to one of their substrate recognition regions. The DNAzyme coated microarrays are then incubated with Cy5 labeled oligonucleotide substrates (green) complementary to the second substrate recognition regions of the DNAzyme variant. Following hybridization, the chip can then be washed leading to removal of any unligated substrates or K-mers. Active DNAzymes are expected to ligate the substrates which are therefore not washed away. The microarray chip can then be scanned and the degree of fluorescence used as a proxy for the activity of a given DNAzyme. (FIG 8B) An alternative approach to (FIG 8A) in which oligonucleotide substrates of different sequences are attached to the microarray surface instead of attachment of DNAzymes. Incubation of the chip with a solution containing a mixture of DNAzymes (red) and Cy5 labeled substrates (green) leads to ligation of the substrates only at spots on the chip containing active DNAzymes. Unligated substrates are removing by a washing step. Scanning of the microarray chips is then used to measure fluorescence from each spot. The incubation steps of both (FIG 8A) and (FIG 8B) can be performed at varying temperatures to determine groups of DNAzymes that are active at specific temperatures, hence, can be used orthogonally. This data can be used to determine the relationship between activity or performance of each DNAzyme and the melting temperature of the duplex formed with their corresponding substrates, thereby enabling the prediction of the activity of DNAzymes for which experimental data may not be available. In some embodiments, the methods in this example are applicable to monitoring activity of nucleic acid operating enzymes including programmable nucleases (e.g. CRISPR/Cas9, meganucleases, ZFNs and TALENs), protein based nucleic acid ligases (e.g. T4 DNA ligase, thermostable ligases), protein based nucleases (restriction endonucleases, ribo-/deoxyribonucleases, nickases, dicers, spliecosomes, etc), integrases, recombinases, transposases, DNA repair enzymes, RNA editing enzymes, ribozymes etc.

EXAMPLE 7: ISOLATION OF LIGATED PRODUCTS

[0103] Reaction mixture can be separated using standard DNA purification and cleaning protocols such as phenol-chloroform extraction. Where highly pure ligation products are desired, purification can be performed using agarose/ polyacrylamide gel electrophoresis to isolate ligated products based on their size. The reaction products can also be isolated using affinity purification. In one instance of affinity purification, the first oligonucleotide can be immobilized onto an affinity column using methods known in the prior art such as the attachment of the oligonucleotide onto an affinity column via a biotin-streptavidin linkage as shown in Figure 9.

[0104] In another instance, the first oligonucleotide or template can be attached to biotin and the second or terminal oligonucleotide attached to a fluorescent dye such as 6-FAM. Ligation reactions can then be purified by fluorescent activated cell sorting (FACS) in which the reactions are mixed with streptavidin coated magnetic beads and injected into a flow cytometry column. Ligated products can then be separated based on their size (due to attachment to beads) and fluorescence.

EXAMPLE 8: PARALLEL MONITORING OF NUCLEOTIDE ENZYME VARIANTS BY SEQUENCING

[0105] The activity of several nucleotide enzyme variants can be monitored in parallel not only using microarrays (Example 6) but also using nucleic acid sequencing approaches. For this to be achieved, nucleotide enzyme variants containing variable substrate recognition sequences or catalytic domains are incubated with an equimolar mixture of templates that are complementary to at least one nucleotide enzyme in the reaction mixture (Figure 10). Following incubation of the reactions, two sets of primers are added to the reaction mixture: i) a set of primers complementary to each possible ligation junctions that are expected given the nucleotide enzyme-template mixture are added to the reaction mixture, and ii) a set of primers complementary to the 3'-terminus of each expected ligation product. The reaction mixture can then be subjected to sequencing using technologies known in prior art. The activity of each nucleotide variant is then estimated as the ratio between the depth of coverage (read count) of its expected ligation products and the observed total coverage of its corresponding templates (ligated and non-ligated) (Figure 10). The methods provided in this example are generalizable to parallel monitoring of the activity of nucleic acid operating enzymes including programmable nucleases (e.g. CRISPR/Cas9, meganucleases,

ZFNs and TALENs), protein based nucleic acid ligases (e.g. T4 DNA ligase, thermostable ligases), protein based nucleases (restriction endonucleases, ribo-/deoxyribonucleases, nickases, dicers, spliecosomes, etc), integrases, recombinases, transposases, DNA repair enzymes, RNA editing enzymes, spliceosomes, etc. For simplicity, the example provided in Figure 10 has only 2 DNAzyme variants (V1 and V2) being monitored. A typical example application of the method would involve 10, 10^2, 10^3, 10^4, 10^5, 10^6 or more variants.

EXAMPLE 9: CONTROLLED ASSEMBLY OF FRAGMENTS

[0106] The capability of programmed DNAzymes to assemble desired target sequences from smaller fragments was examined. To this end, DNAzymes were designed to ligate a given set of fragments, each having a different sequence and length, into determined longer molecules. Specifically, the starting pool consisted of four sequences (Figure 13.A, PCR primer sequences in bold below):

A (60 nt):
GTCGCATCCAAGCGGATCCACATAACCCGTTCAGCCAGGNNNNNNCG
ACTGGCGAACAGC (SEQ ID NO:9)
B (80 nt):
GATCGCACTGGCTNNNNNNAATCGTACTGCAGGGAAAAACGCACGCA
AGAGTNNNNNNGGTGCCTAGTTTCATGGGCAGC (SEQ ID NO:10)
C (100 nt):
GGAACACTATCCGNNNNNNAGGATACGACGTTGTTTGTCGACCGCGCC
ATGGGAAATAAGTGNNNNNNACCGTCATTGATTCTTCTGCCGACATCC
AAGC (SEQ ID NO:11)
D (120 nt):
GGCCCTCGATCAANNNNNNACGTGTGCTGAGTATTCTCAGATTTGAGA
TATCGCTTGCTCGACACAGTCTCTNNNNNNGTTAAGACCCCGCATATC
AGTTGAGCGTGAAGCATCTCAAGC (SEQ ID NO:12)

The different length of the fragments allows for identification of the ligated product by gel electrophoresis, with or without prior PCR amplification. A set of 16 DNAzymes was designed to performed pairwise ligation of these sequences, e.g. DNAzyme ab will ligate fragment A to fragment B.

Controlled ligation into desired sequences was then demonstrated for target molecules ABC, DAC, CBD, ABCD and DBCA (Figure 13.C). In all cases, a

starting mix of the four fragments was provided, as well as the required DNAzymes. Product-specific PCR amplification was performed before gel electrophoresis. These results also provide further evidence for the sequence specificity and programmability of the ligases.

EXAMPLE 10: ASSEMBLY OF LONG DNA MOLECULES

[0107] To investigate whether large DNA molecules could be created using programmable ligation, DNAzymes were designed to ligate a fragment to another copy of itself (Figure 14.A). The resulting products can then be ligated by the same DNAzyme, leading to recursively growing molecules. PCR amplification followed by gel electrophoresis showed that fragments of 10kb and beyond can be achieved through programmed ligation (Figure 14.B). The fact that the size of the resulting molecules increases with the duration of the PCR cycles indicates that the PCR is the limiting step and the original ligation product may be larger still.

EXAMPLE 11: PROGRAMMABLE LIGATION

[0108] The ability of DNAzyme ligases to be transfected into living cells, and ultimately whether ligase activity in vivo could be achieved was examined. This was examined initially on the human endothelial cell line, HEK293.

[0109] Figure 15 shows the preliminary data that suggests in-cell ligation is achievable. A DNAzyme ligase (47nt) and 2 x 50nt substrates was transfected into the cells and rested overnight. They were then assessed ligation activity using a specially designed PCR assay developed for detecting ssDNA in cell lysates. Figure 15(b) shows that lipid based transfection methods are capable of delivering a fluorescent labeled 50nt ssDNA fragment into human cells to levels comparable to a standard GFP expressing plasmid. The ssDNA were also shown to be non-toxic to the cells using standard cell viability assays. Figure 15(c) shows that we can detect a 100nt ligation product in cells transfected with the 47 nt ligase and 2 x 50 nt substrates.

EXAMPLE 12: MODIFIED SSDNA LIGASES

[0110] In addition to standard ssDNA oligonucleotides ODN (oligodeoxynucleotide) ssDNA that contains a modified phosphorothioate backbone were also examined. Modifications such as these, and also locked-nucleic acid variations of DNA structures have been shown to be resistant to endogenous degradation pathways either in cells or in the circulatory system of animals. They are therefore desirable for the gene editing validation and other in vivo applications as can remain active

or stable over longer periods. As shown in FIG 16(a) and (b), a variety of common ssDNA modification strategies do not adversely affect the enzymatic activity of the ligases. See Figure 16 (a) and 16(b).

EXAMPLE 13: RNA-DNA HYBRIDS

[0111] As well as using the DNAzyme ligase to synthesis DNA strands, it has also been shown that RNA-DNA hybrids can be synthesised. These observations provide more flexibility for downstream applications such as tagging of endogenous RNA with DNA substrates as desired. There are also implications for improved delivery aspects if substrates can potentially be expressed off of plasmid vectors as RNA, or engineered into transgenic mouse lines to enable cell type or tissue specific expression of ligase components. Another exciting application is switchable assembly of active mRNA from smaller constituent parts that is capable of protein expression – this provides a wide variety of possible readout mechanisms. See Figure 17.

EXAMPLE 14: RNAZYME LIGASE

It has also been observed that replacing the DNAzyme ligase with RNA does not adversely affect its activity or its ability to ligate ssDNA or DNA-RNA hybrids. This has huge potential in that the ligase can be expressed off of a plasmid vector or delivered into primary cells/tissue using viral vectors such as AAV or lentivirus. This will allow the delivery of programmable ligases into difficult-to-transfect cell types or even tissue. See Figure 17.

CLAIMS

1. A composition comprising at least first and second nucleotide enzymes, wherein each of said nucleotide enzymes comprises:
    a. at least a first and second target binding sequence; and
    b. a catalytic sequence;

    wherein said first and second target binding sequences of said first and second nucleotide enzyme are substantially complementary to polynucleotide target sequences and wherein upon hybridization of polynucleotide target sequences with first and second target binding sequences of any of said nucleotide enzymes, said nucleotide enzyme is capable of catalyzing ligation of said first and second polynucleotide sequences.

2. The composition according to claim 2, further comprising a plurality of polynucleotide target sequences.
3. The composition according to claim 1 or 2, wherein said nucleotide enzyme comprises at least one modified nucleotide.
4. A method of assembling a plurality of polynucleotides comprising combining in a reaction vessel a plurality of polynucleotide target sequences with at least first and second nucleotide enzymes, wherein each of said nucleotide enzymes comprises:
    a. at least a first and second target binding sequences; and
    b. a catalytic sequence;

    wherein said first and second target binding sequences of said nucleotide enzyme hybridize with substantially complementary sequences in first and second polynucleotide target sequences and wherein upon said hybridization said first and second polynucleotide target sequences are ligated with each other to form a first ligated polynucleotide.

5. The method according to claim 4, wherein at least one of said polynucleotide target sequences comprises RNA, DNA or a modified variant thereof.

6. The method according to claim 4 or 5, wherein said first and second target binding sequences of said second nucleotide enzyme hybridize with substantially complementary sequences in first and third polynucleotide target sequences and wherein upon said hybridization said first and third polynucleotide target sequences are ligated with each other to form a second ligated polynucleotide.
7. The method according to claim 4, 5 or 6, further comprising combining in said reaction vessel a third nucleotide enzyme comprising
    a. at least a first and second target binding sequences; and
    b. a catalytic sequence;

    wherein said first and second target binding sequences of said third nucleotide enzyme hybridize with substantially complementary sequences in second and third polynucleotide target sequences and wherein upon said hybridization said second and third polynucleotide target sequences are ligated with each other to form a third ligated polynucleotide.

8. The method according to claim 4, 5, 6 or 7, wherein a nucleotide enzyme hybridizes with said first and second, said second and third, or said first and third ligated polynucleotide, wherein upon said hybridization, said ligated polynucleotides are ligated to form a fourth ligated polynucleotide.

9. A method of assembling a plurality of polynucleotides comprising combining in a reaction vessel a plurality of polynucleotide target sequences with at least one nucleotide enzyme, wherein each of said nucleotide enzymes comprises:
    a. at least a first and second target binding sequences, wherein said first target binding sequences hybridize with the 5' terminus of the first polynucleotide target sequence and said second target binding sequences hybridize with the 3' terminus of the second polynucleotide target sequence; and
    b. a catalytic sequence;

    wherein said first and second target binding sequences of said first nucleotide enzyme hybridize with substantially complementary sequences in first and second polynucleotide target sequences and wherein upon said hybridization said first and

second polynucleotide target sequences are ligated with each other to form a first ligated polynucleotide.

10. A method of assembling a plurality of polynucleotides comprising:
    a. activating the 3' or 5' terminus of a plurality of target polynucleotides;
    b. combining in a reaction vessel a plurality of activated polynucleotide target sequences with at least first and second nucleotide enzymes, wherein each of said nucleotide enzymes comprises:
        i. at least a first and second target binding sequences, wherein said first target binding sequences hybridize with the 5' terminus of the first polynucleotide target sequence and said second target binding sequences hybridize with the 3' terminus of the second polynucleotide target sequence; and
        ii. a catalytic sequence;

    wherein said first and second target binding sequences of said first nucleotide enzyme hybridize with substantially complementary sequences in first and second polynucleotide target sequences and wherein upon said hybridization said first and second polynucleotide target sequences are ligated with each other to form a first ligated polynucleotide.

11. The method according to claim 4, 9 or 10, wherein said target binding sequences comprise at least 4 contiguous polynucleotides.

12. The method according to claim 11, wherein said target binding sequences comprise at least 6 contiguous polynucleotides.

13. The method according to claim 11 or 12, wherein said target binding sequences are shorter than 30 contiguous polynucleotides.

14. The method according to claim 4, 9 or 10, wherein at least 5 target polynucleotide sequences are ligated into a single ligated polynucleotide sequence.

15. The method according to claim 4, 9 or 10, wherein at least 5 different ligated target polynucleotide sequences are generated in a single reaction.

16. The method according to claim 4, 9 or 10, wherein said polynucleotide target sequences are prepared by a method selected from polynucleotide synthesis and fragmentation of precursor polynucleotide sequences.

17. The method according to claim 4, 9 or 10, wherein one of said ligated target polynucleotides comprises a label.

18. The method according to claim 4, 9 or 10, wherein one of said ligated target polynucleotides is immobilized on a solid support.

19. The method according to claim 18, wherein said solid support is selected from a planar substrate or microsphere.

20. The method according to claim 4, 9 or 10, wherein said target polynucleotides are selected from the group consisting of RNA, DNA, LNA, PNA, capped nucleic acids, partially or fully phosphorothioate-backbone nucleic acids, XNA and FANAS.

21. A method for identifying the catalytic and substrate binding domains of a nucleotide enzyme involving aligning a library of variants of the nucleotide enzymes and identifying the substrate recognition domain by identifying the regions of the nucleotide enzymes that are complementary to the target polynucleotide.

22. A method for designing a catalytic nucleic acid for ligating at least two nucleic acids comprising:

    a. modularizing the catalytic nucleic acid into one or more catalytic domains and at least two substrate recognition domains;
    b. defining the catalytic domain as the region of the catalytic nucleic acid between at least two substrate recognition domains;
    c. defining the substrate recognition domains as the sequences that form complementary base pairs to substrate sequences to be joined;
    d. programming the catalytic nucleic acid to ligate two nucleic acids by designing at least one of the substrate recognition domains to be complementary to the terminus of one of the target nucleic acids and the second substrate recognition domain to be complementary to the terminus of the other target nucleic acid based on Watson-Crick base pairing.

23. The method of claim 22 wherein said defining the substrate recognition domains comprises determining the optimal length of at least one of the substrate binding domains by a method comprising:

a. calculating the thermodynamic stability of Watson-Crick pairs between the substrate recognition domains of a given length and the terminal regions of the target nucleic acid sequences;
b. calculating the probability of secondary structures forming on the substrate recognition sequences;
c. calculating the probability of intra-molecular interactions between the substrate recognition sequences and the catalytic domains;
d. calculating the expected catalytic efficiency of the catalytic nucleic acid sequence given the sequence of the substrate recognition domain;
e. calculating the error rate of the nucleic acid synthesizer for making the catalytic nucleic acid; or
f. calculating the cost of synthesizing the nucleic acid of a given length on the selected nucleic acid synthesizer

24. The method according to claim 23, comprising at least two of a, b, c, d, e or f.

25. The method according to claim 23, comprising at least three of a, b, c, d, e or f.

26. The method according to claim 23, comprising at least four of a, b, c, d, e or f.

27. The method according to claim 23, comprising at least five of a, b, c, d, e or f.

28. The method according to claim 23, comprising a, b, c, d, e and f.

29. A composition comprising at lease one nucleotide enzyme, wherein said nucleotide enzyme comprises:

a. at least a first and second target binding sequence; and
b. a catalytic sequence;

wherein said first and second target binding sequences of said first and second nucleotide enzyme are substantially complementary to polynucleotide target sequences and wherein upon hybridization of polynucleotide target sequences with first and second target binding sequences of any of said nucleotide enzymes, said nucleotide enzyme is capable of catalyzing ligation of said first and second polynucleotide sequences, and wherein said nucleotide enzyme comprises at least one modified base, or at least one ribonucleotide.

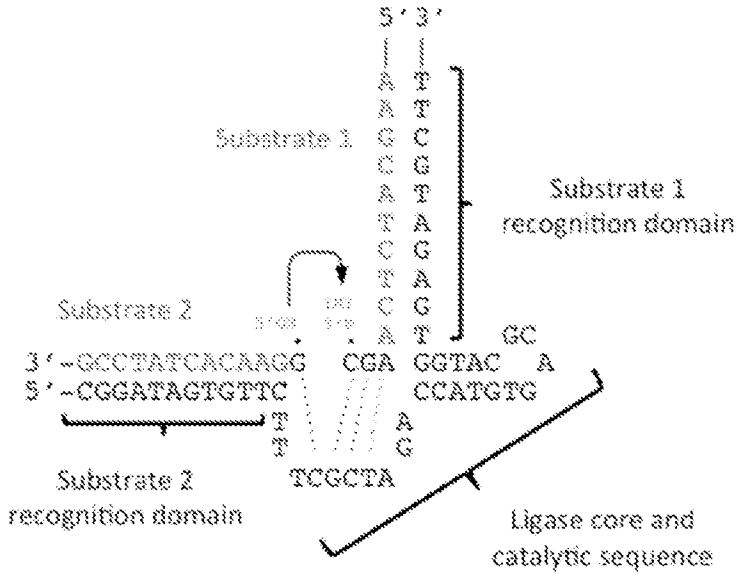

Figure 1

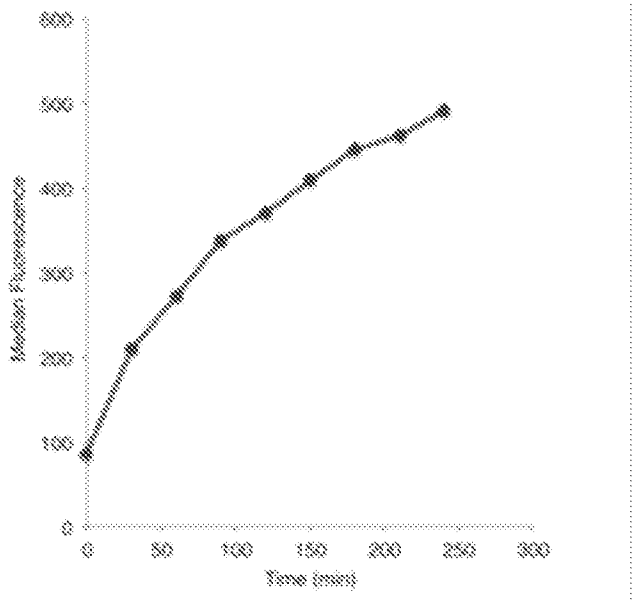

Figure 2

	E47 ligase:
SEQ ID NO:5	CCTGTTTCATGAGACCATGTGACGCATGGCCCG
	S1, Hello:
SEQ ID NO:6	/5'BiotinTEG/GAACGCGGGCGCGCTTGTACGCGGGTG AGCCAGCGGGTATGTCGGTATGTGAGCGGGCTGGCA GGCTCGCGAGACGGCAGGTGCGCTTGCCGGCGAGCG GGCAGGCATGCCAGCTTGTGAGCCAGCGGGTAC/3'Ph os/
	S2, World:
SEQ ID NO:7	5'GACGGTGAGTACGTCGGTGAGCTTGCTGGC AGGCCTGCGGGTATGCTGGCAGGCTAGCTAGTGAGC AGGCTAGCCTACTCGACAGCGGGCTAGCTAGCTTGG GTGCTTGTACGCTAGCGAACAGTAACCCAA/3'6-FAM/

Figure 3

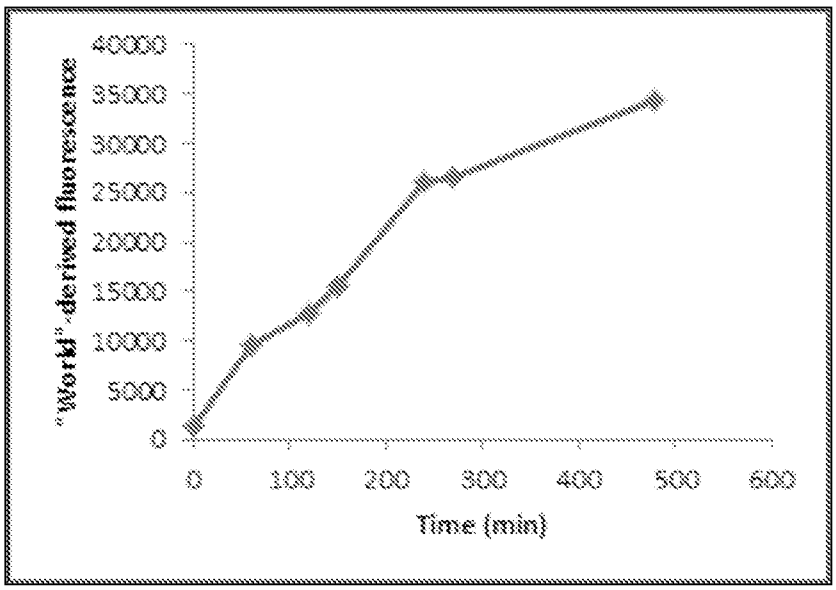

Figure 4

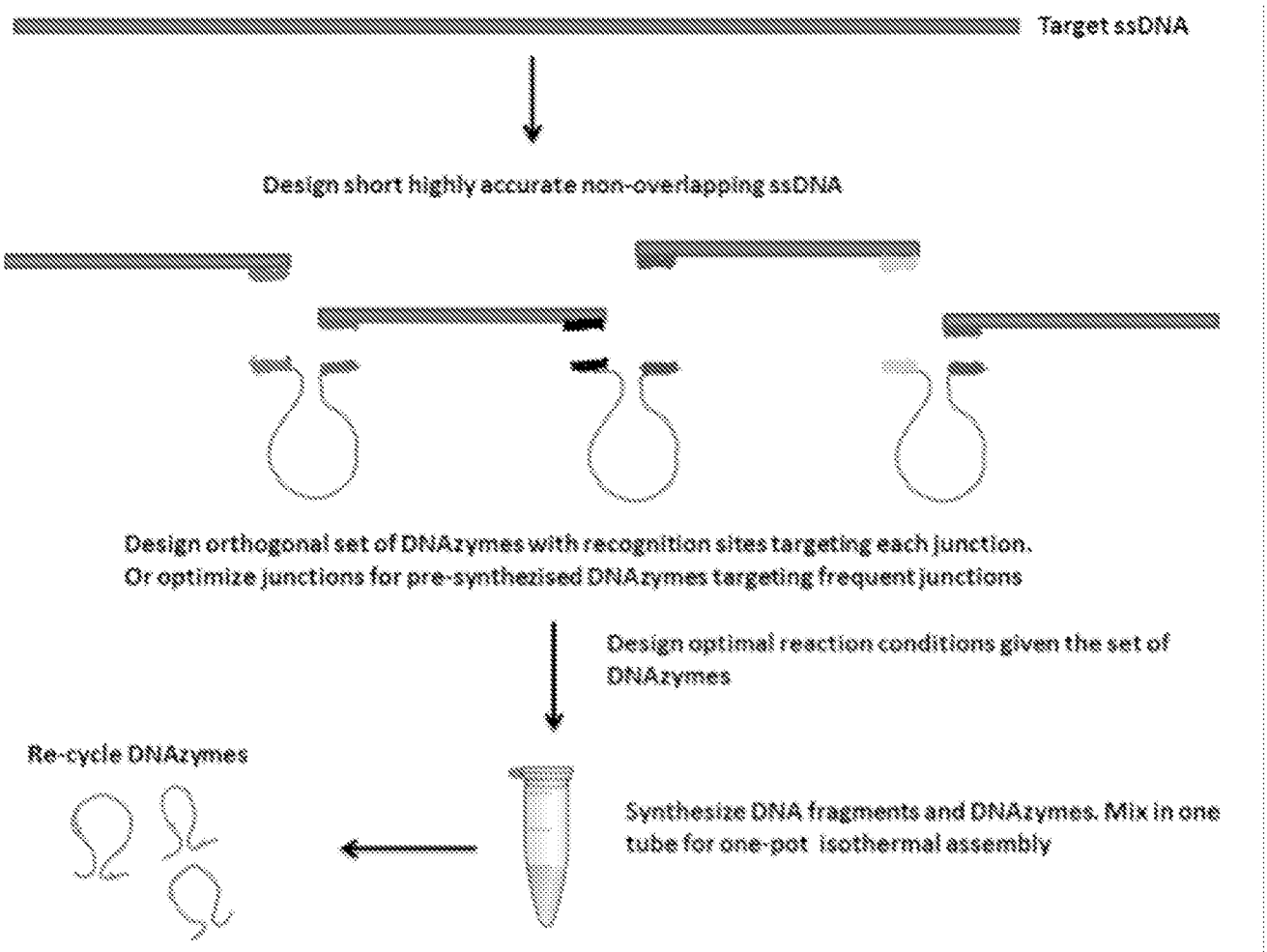
Target ssDNA
Design short highly accurate non-overlapping ssDNA
Design orthogonal set of DNAzymes with recognition sites targeting each junction.
Or optimize junctions for pre-synthesized DNAzymes targeting frequent junctions
Design optimal reaction conditions given the set of DNAzymes
Re-cycle DNAzymes
Synthesize DNA fragments and DNAzymes. Mix in one tube for one-pot isothermal assembly

Figure 5

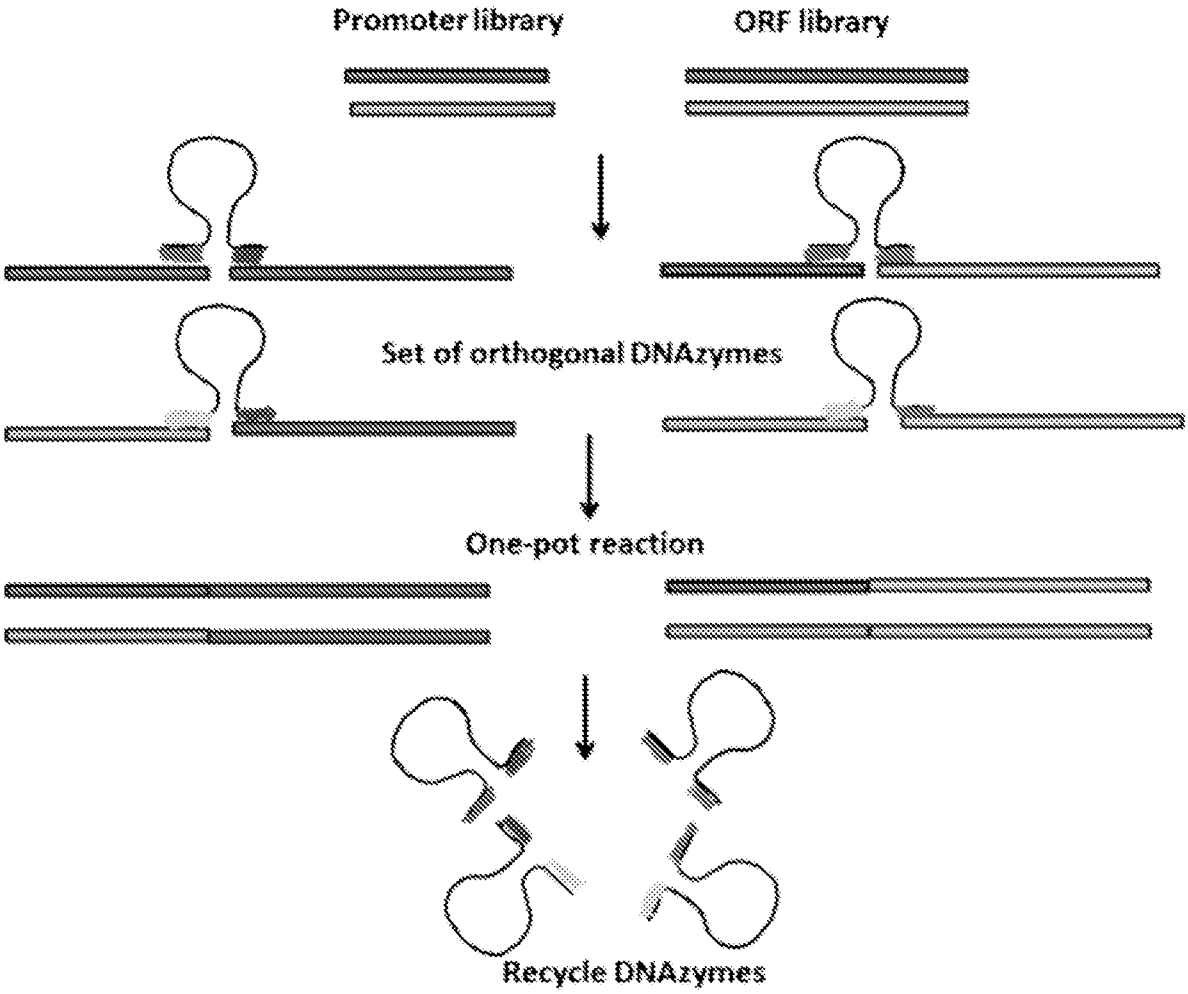
Promoter library
ORF library
Set of orthogonal DNAzymes
One-pot reaction
Recycle DNAzymes

Figure 6

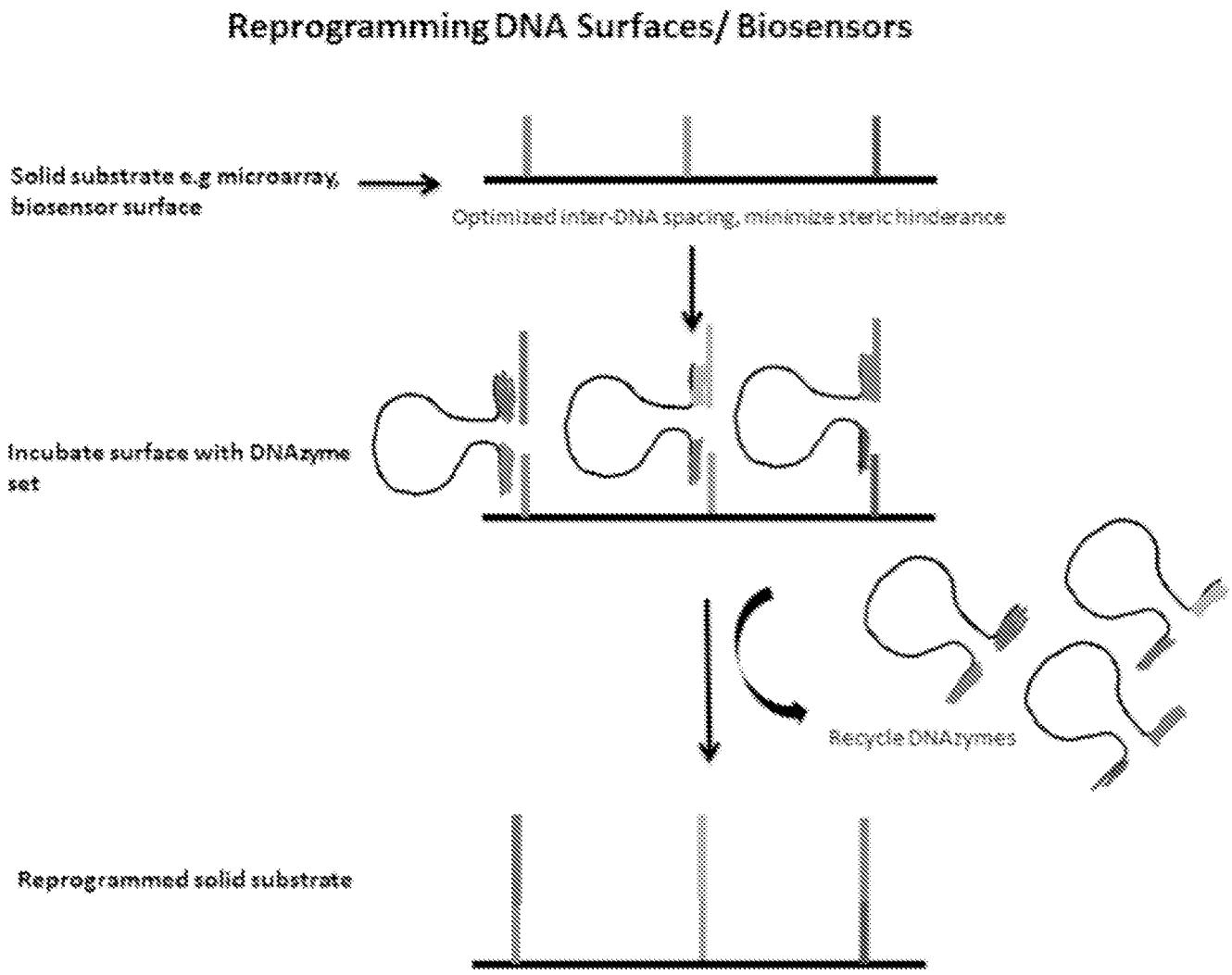
Reprogramming DNA Surfaces/ Biosensors
Solid substrate e.g microarray, biosensor surface
Optimized inter-DNA spacing, minimize steric hinderance
Incubate surface with DNAzyme set
Recycle DNAzymes
Reprogrammed solid substrate

Figure 7

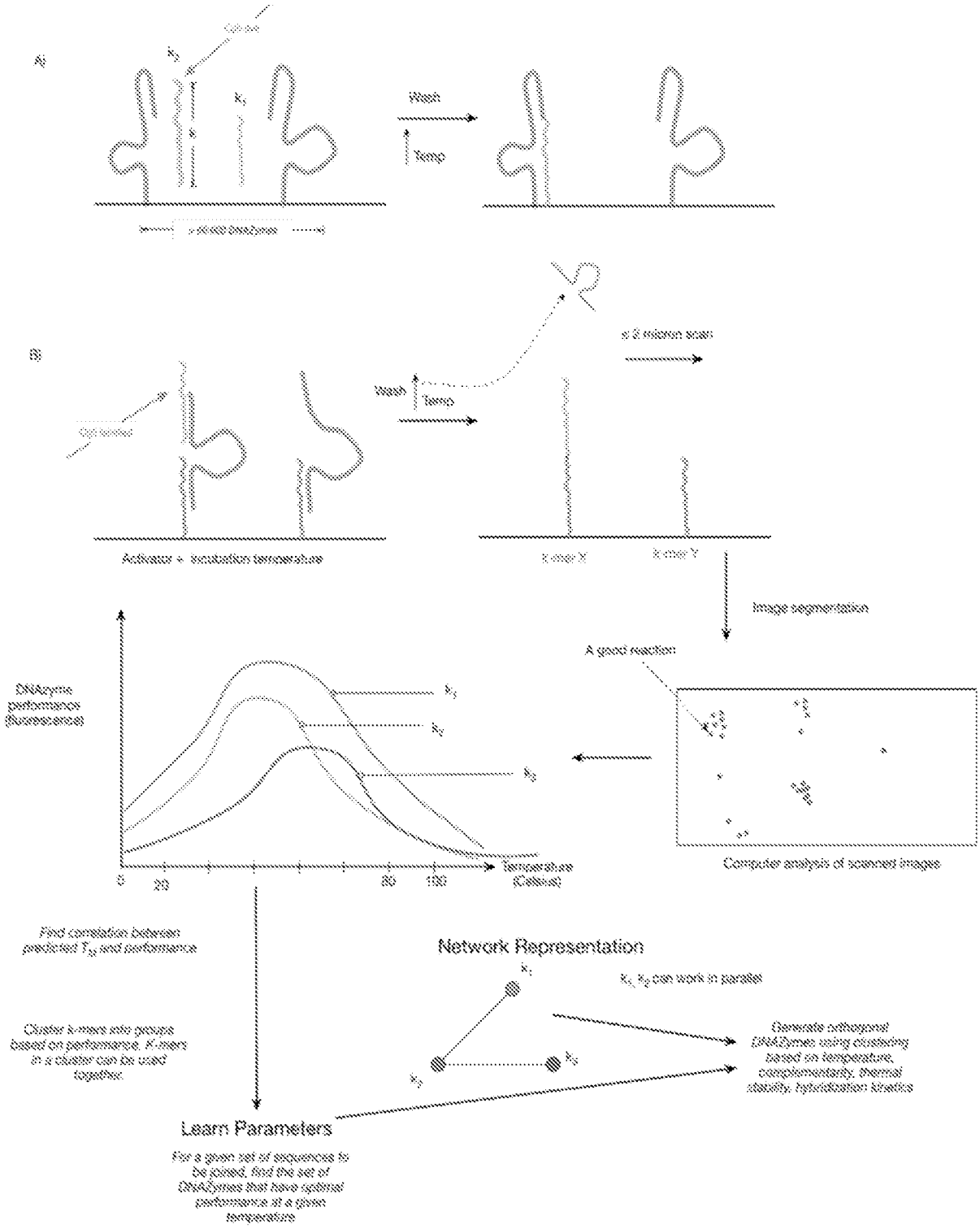
A)
Wash
Temp
B)
Wash
Temp
Image segmentation
A good reaction
DNAzyme performance (fluorescence)
Temperature (Celsius)
Computer analysis of scanned images
Find correlation between predicted T_M and performance
Cluster k-mers into groups based on performance. K-mers in a cluster can be used together.
Network Representation
k_1, k_2 can work in parallel
Generate orthogonal DNAZymes using clustering based on temperature, complementarity, thermal stability, hybridization kinetics
Learn Parameters
For a given set of sequences to be picked, find the set of DNAZymes that have optimal performance at a given temperature

Figure 8

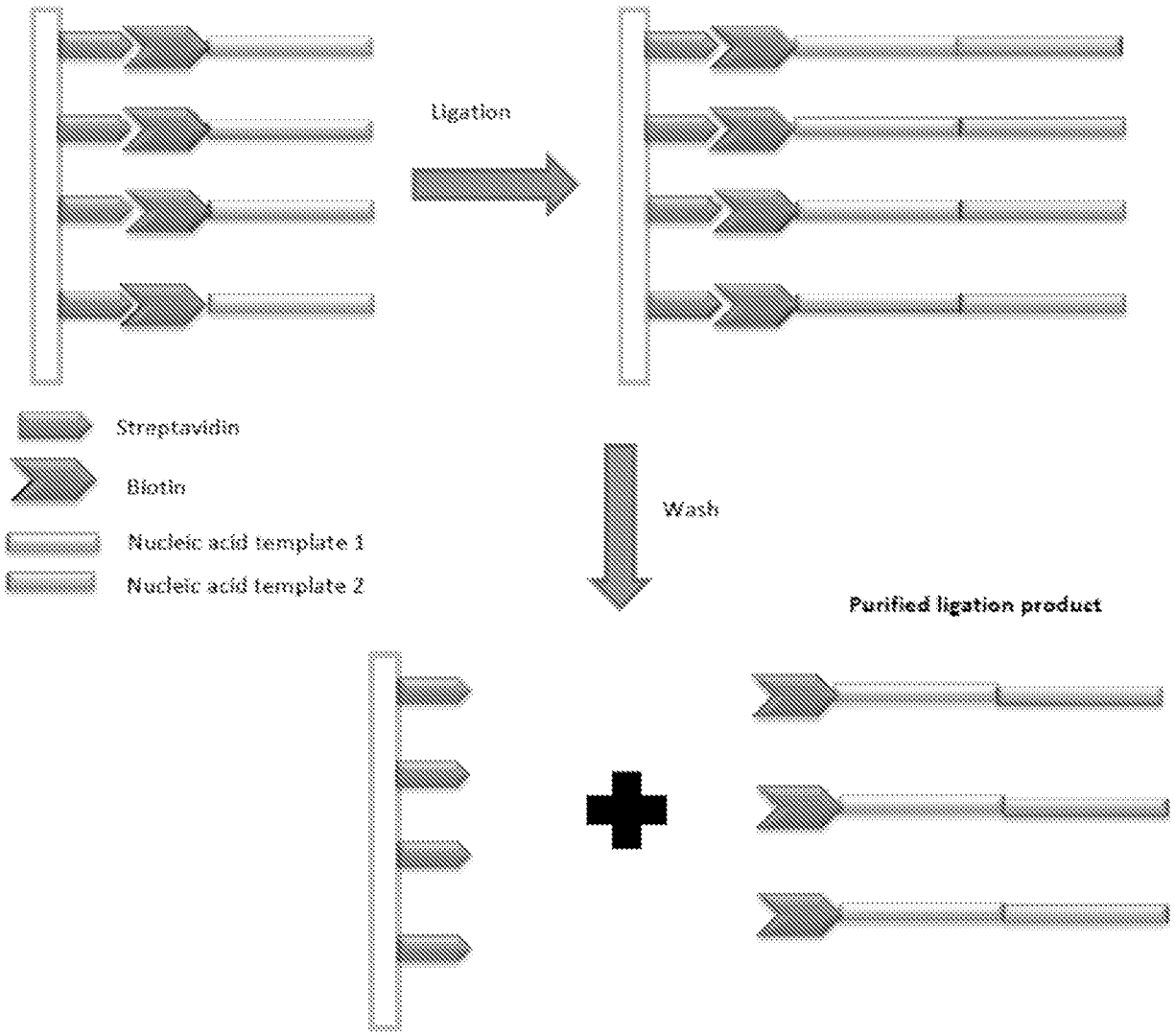
Ligation
Streptavidin
Biotin
Nucleic acid template 1
Nucleic acid template 2
Wash
Purified ligation product

Figure 9

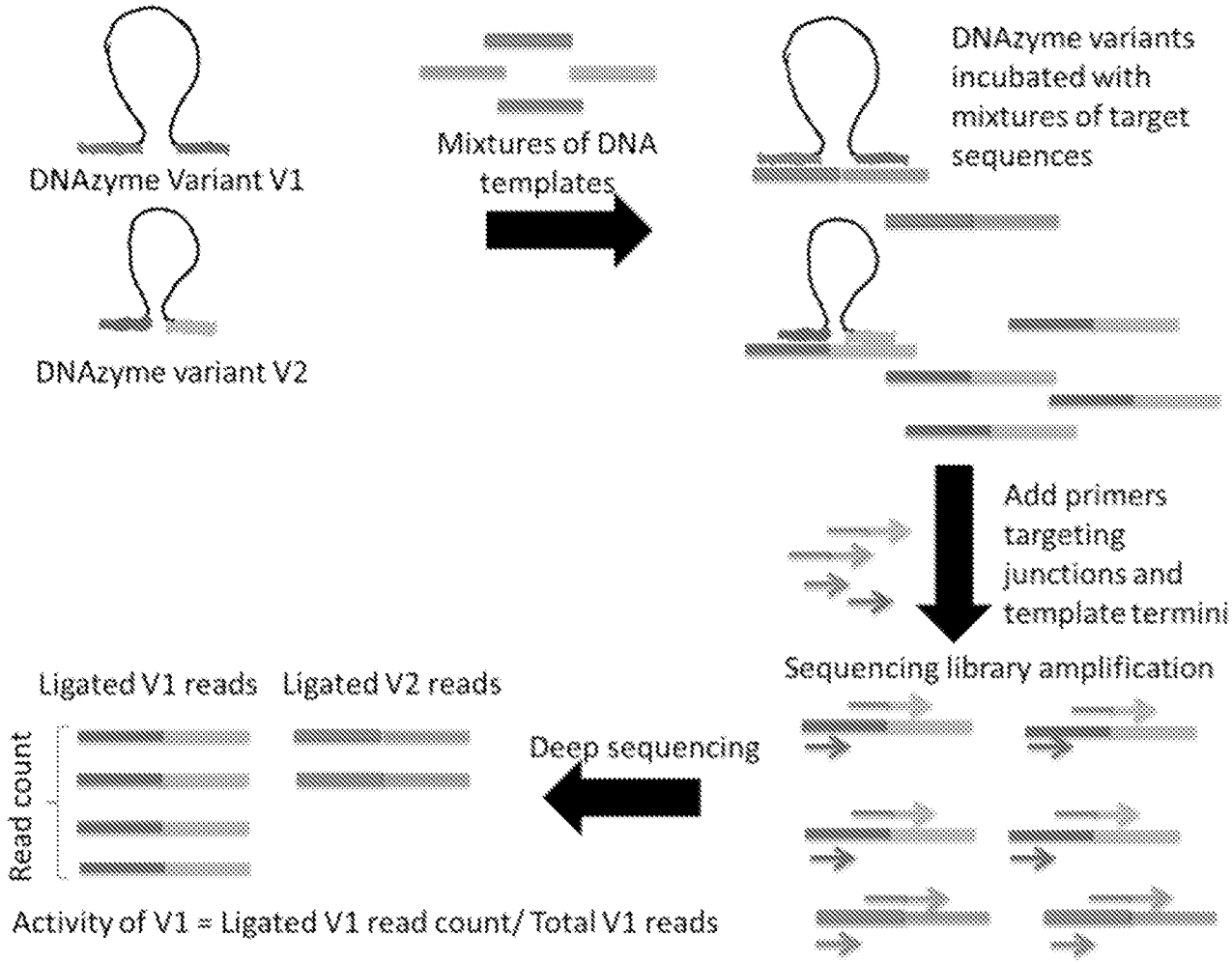
DNAzyme Variant V1
DNAzyme variant V2
Mixtures of DNA templates
DNAzyme variants incubated with mixtures of target sequences
Add primers targeting junctions and template termini
Sequencing library amplification
Deep sequencing
Ligated V1 reads
Ligated V2 reads
Read count
Activity of V1 ≈ Ligated V1 read count/ Total V1 reads

Figure 10

>Target sequence
GTCGACATTGATTATTGATAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAAC
TTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAAC
GCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCAT
ATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACT
TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGG
ATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG
GACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCA
GAGCTGGTTTAGTGAACCGTCAGATCGCCTAGCGCTAGCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGACG
GTACCGCGGGCCCGGGATCCACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACC
GGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCG SEQ ID NO: 8

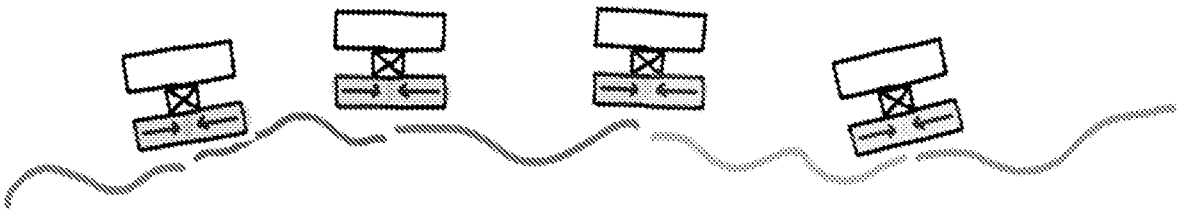

FIG 11

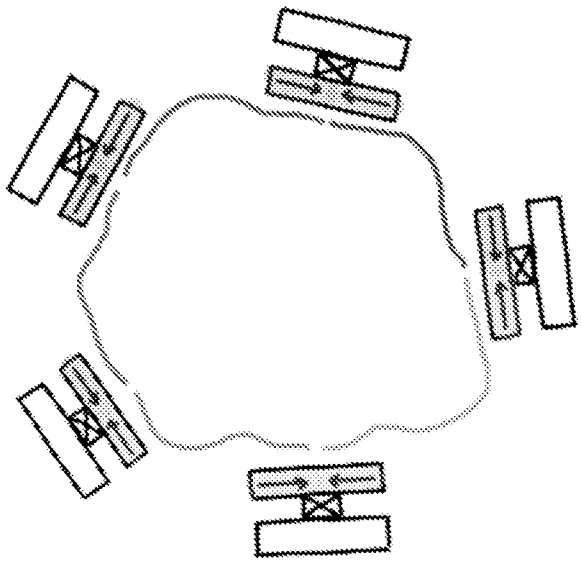

FIG 12

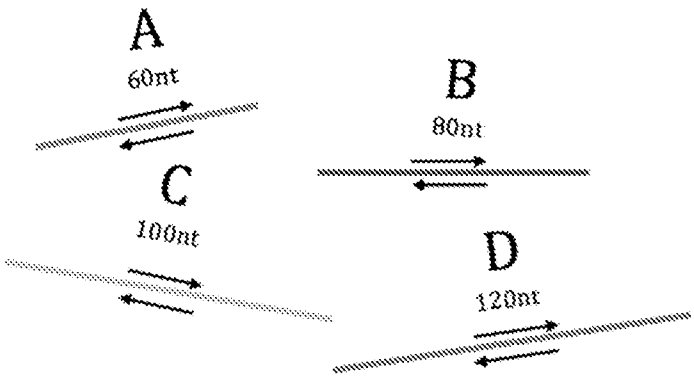

FIG 13A

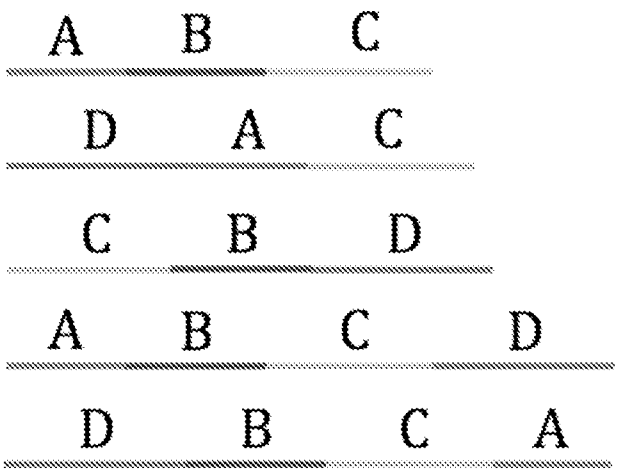

FIG 13B

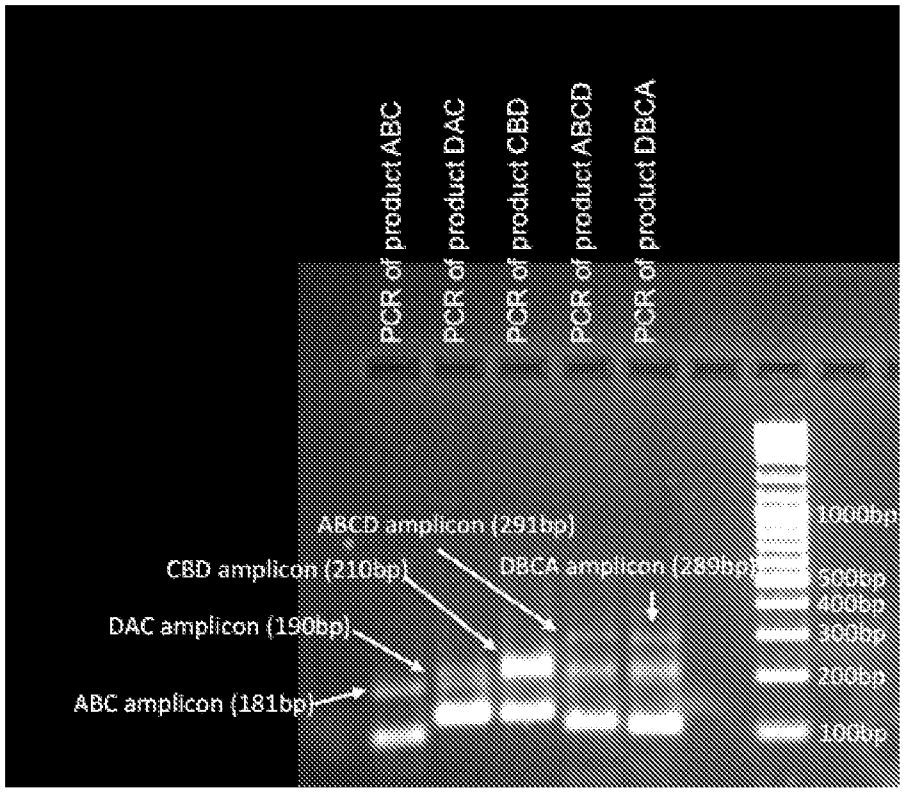
PCR of product ABC
PCR of product DAC
PCR of product CBD
PCR of product ABCD
PCR of product DBCA
ABCD amplicon (291bp)
CBD amplicon (210bp)
DBCA amplicon (289bp)
DAC amplicon (190bp)
ABC amplicon (181bp)
1000bp
500bp
400bp
300bp
200bp
100bp

FIG 13C

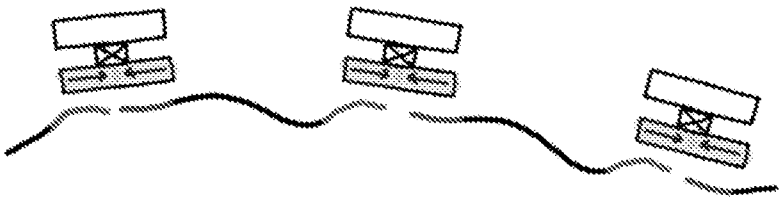

FIG 14A

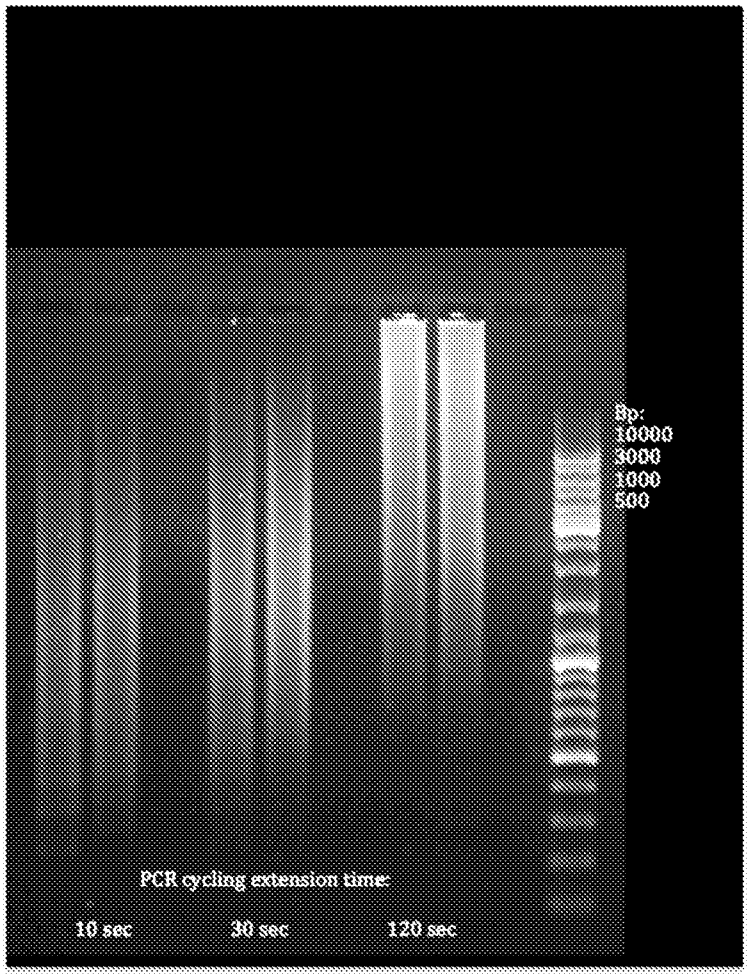

FIG 14B

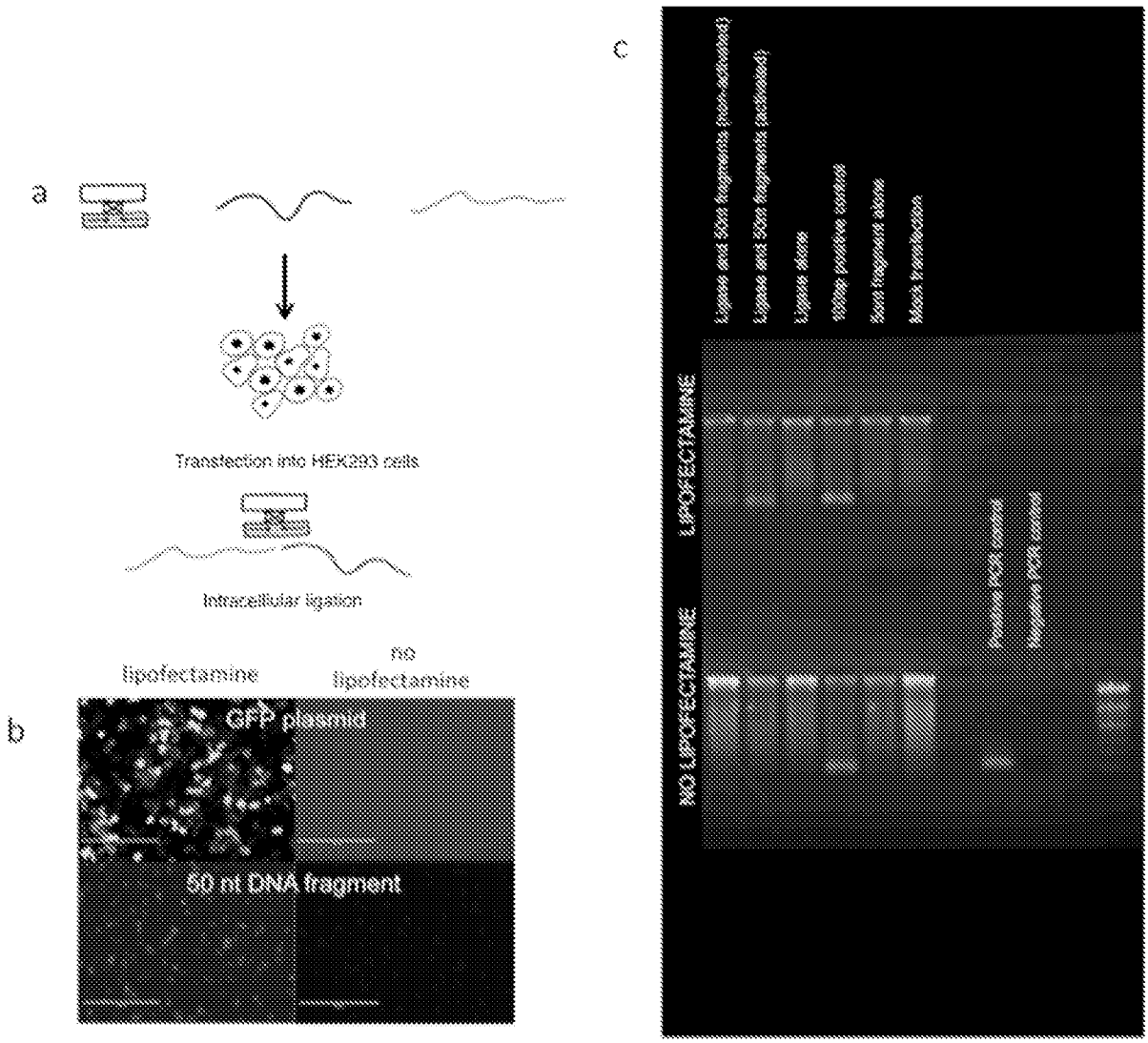
a
Transfection into HEK293 cells
Intracellular ligation
b
lipofectamine
no lipofectamine
GFP plasmid
50 nt DNA fragment
c
Ligase and 50nt fragments (non-annealed)
Ligase and 50nt fragments (annealed)
Ligase alone
Kilo positive control
Kan fragment alone
Mock transfection
LIPOFECTAMINE
NO LIPOFECTAMINE

FIG 15

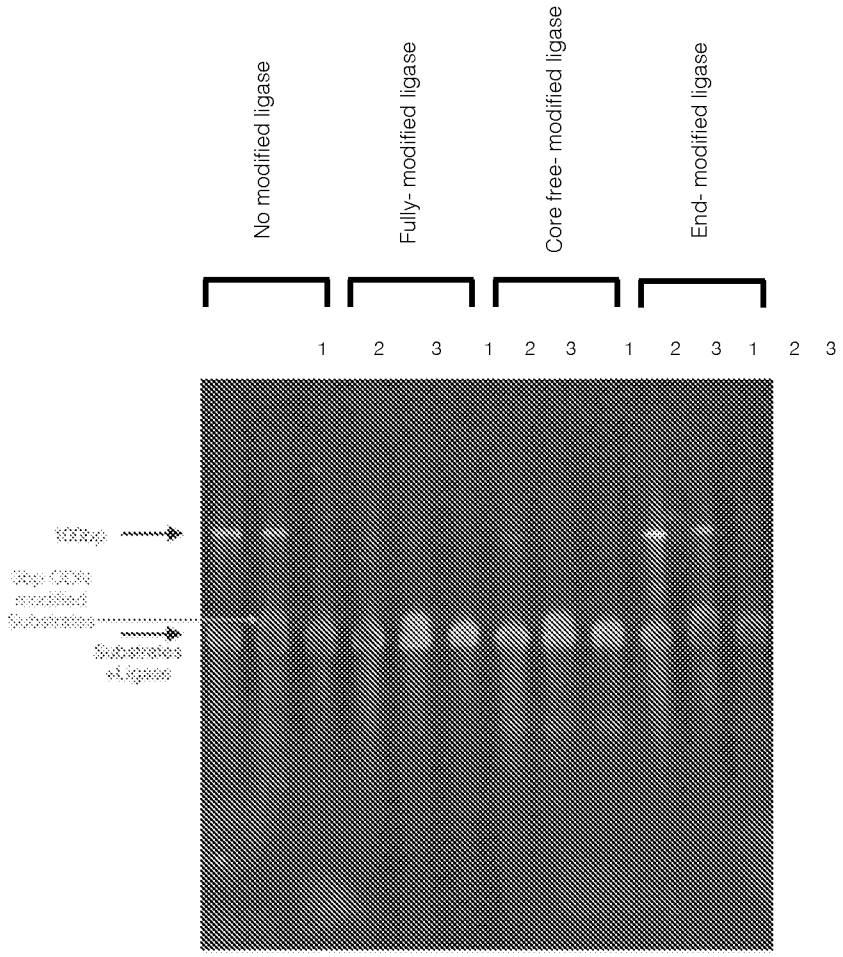
No modified ligase
Fully- modified ligase
Core free- modified ligase
End- modified ligase
1 2 3 1 2 3 1 2 3 1 2 3

FIG 16A

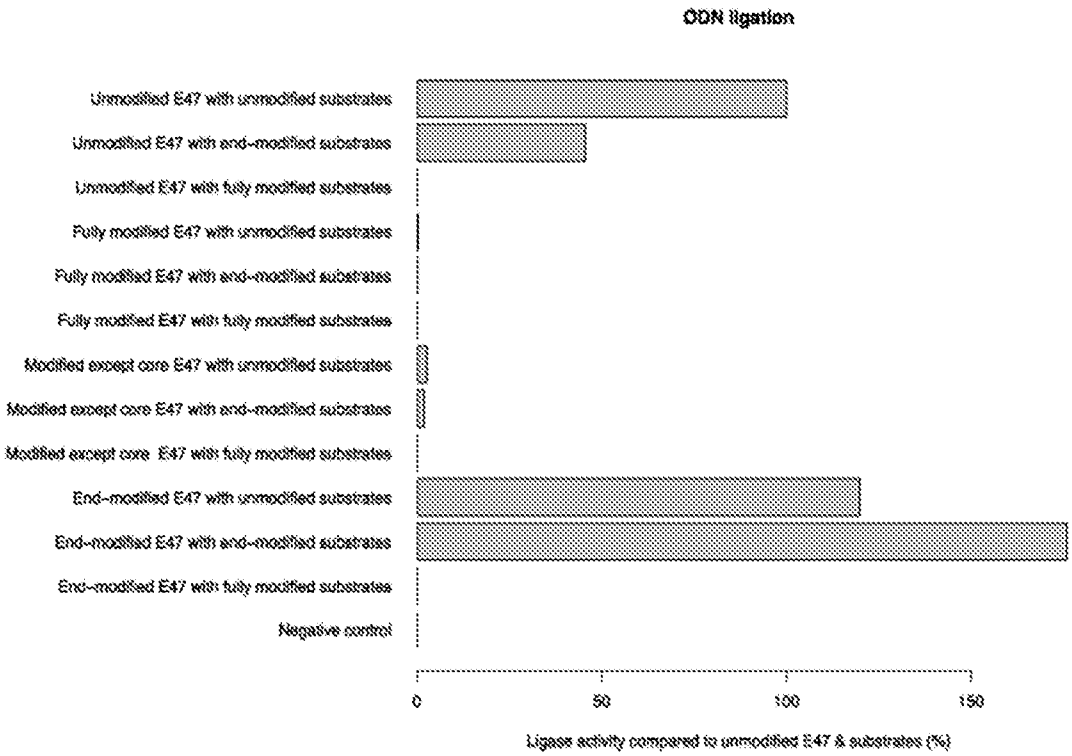

FIG 16B

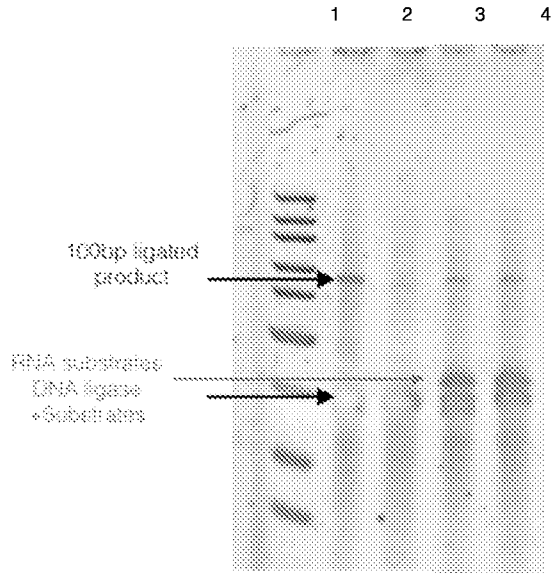

FIG 17