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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 July 2012 (26.07.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/099832 A2

(51) International Patent Classification:
C12N 15/10 (2006.01)

(21) International Application Number:
PCT/US2012/021465

(22) International Filing Date:
17 January 2012 (17.01.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/433,502	17 January 2011 (17.01.2011)	US
61/433,488	17 January 2011 (17.01.2011)	US
61/474,205	11 April 2011 (11.04.2011)	US
61/474,168	11 April 2011 (11.04.2011)	US

(71) Applicant *(for all designated States except US)*: **LIFE TECHNOLOGIES CORPORATION** [US/US]; 5791 Van Allen Way, Carlsbad, California 92008 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **HENDRICKS, Stephen** [US/US]; Life Technologies Corporation, 5791 Van Allen Way, Carlsbad, California 92008 (US).

(74) Agent: **GERMAN, Roberta**; Life Technologies Corporation, 5791 Van Allen Way, Carlsbad, California 92008 (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, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

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

WO 2012/099832 A2

(54) **Title**: ENZYMATIC LIGATION OF NUCLEIC ACIDS

(57) **Abstract**: Methods, assays, compositions and kits for the ligation of short polynucleotides are presented herein. The short polynucleotides are optionally no more than 7 nucleotides in length, and can be as short as 3 or 4 nucleotides in length. The ligation is optionally performed by CV ligase.

ENZYMATIC LIGATION OF NUCLEIC ACIDS

[0001] This application claims priority to the following U.S. provisional applications, each of which is incorporated by reference in its entirety: App. No. 61/474,205 filed April 11, 2011; App. No. 61/474,168 filed April 11, 2011; App. No. 61/433,502 filed January 17, 2011; and App. No. 61/433,488 filed January 17, 2011.

Background

[0002] DNA ligases can join polynucleotides together, for example by catalyzing the formation of a phosphodiester bond at single- or double-stranded breaks on duplex DNA. Ligases can be sensitive to the degree of hybridization between opposing nucleic acid strands in a duplex. For example, successful ligation can occur less frequently (or not at all) where a strand to be ligated to an adjacent strand is in a duplex and is not complementary to its opposing strand in the duplex. In some cases a single nucleotide mismatch between strands in a duplex can significantly impair or prevent ligation. The capacity of ligases for discrimination based on hybridization, including single nucleotide discrimination, has led to the development of ligase-mediated detection techniques (*e.g.*, Landegren, U., Bioessays, 15(11):761-765 (1993), and Barany, PNAS USA, 88(1):189-193 (1991)). Ligase-based linear signal amplification known as LDR (*i.e.*, ligase detection reaction), combined with PCR (*i.e.*, polymerase chain reaction)-based gene specific target amplification, has been proven to be a useful tool in cancer and disease gene mutation detection. PCR/LDR techniques typically rely on two properties of a DNA ligase: (i) specificity, and (ii) thermostability.

Summary

[0003] This application relates to ligation reagents and methods. Among other things, methods of and reagents for ligating nucleic acids to other nucleic acids are provided, including ligation of two polynucleotides. Either or both polynucleotides can be single-stranded or double-stranded. One or both polynucleotides can be a short oligonucleotide. In some embodiments oligonucleotides and/or polynucleotides of different lengths can be ligated to each other. The ligation can be enzymatic, and ligations can be template-

dependent or template-independent. The nucleic acids that are ligated or involved in the ligation reaction can be labeled or unlabeled and immobilized or in solution.

[0004] Some embodiments involve template-independent ligation of double-stranded or single-stranded polynucleotides (e.g., oligonucleotides). The single-stranded polynucleotides are optionally not hybridized to another polynucleotide. In some embodiments, an oligonucleotide can be hybridized to or otherwise associated with all or a portion of an overhang or other single-stranded region of a duplex nucleic acid and ligated to a free end of a strand of the duplex or to another oligonucleotide that is hybridized or otherwise associated with an overhang or other single-stranded portion of the duplex.

[0005] Some embodiments involve ligation of single-stranded polynucleotides (e.g., oligonucleotides). Optionally, the single-stranded polynucleotides are hybridized hybridized adjacent or near each other on another single polynucleotide. In some embodiments, an oligonucleotide can be hybridized to or otherwise associated with all or a portion of an overhang or other single-stranded region of a duplex nucleic acid and ligated to a free end of a strand of the duplex or to another oligonucleotide that is hybridized or otherwise associated with an overhang or other single-stranded portion of the duplex.

[0006] In some aspects, methods and reagents are provided for hybridizing or otherwise associating a first oligonucleotide and a second oligonucleotide to a third oligonucleotide or to a polynucleotide such that the termini of the first and second oligonucleotides are adjacent to or near each other. Such hybridization or association can occur sequentially, simultaneously, or substantially simultaneously. The terminus of the first oligonucleotide can be ligated to the adjacent or nearby terminus of the second oligonucleotide.

[0007] Nucleotide base mismatches between a first and/or second oligonucleotide and a third oligonucleotide or polynucleotide can affect the efficiency of ligation. For example, mismatches at the terminal position of either or both a first and second oligonucleotides can affect ligation efficiency, reducing the probability of successful ligation or precluded ligation entirely. Mismatches at other or at multiple positions can also affect ligation efficiency, reducing the probability of successful ligation or precluding ligation entirely.

[0008] Also provided are methods and reagents for performing multiple ligations sequentially, in parallel, or both sequentially and in parallel.

[0009] Optionally, one or more of the primer, probe or template is labeled. For example the probe can be labeled. Else the primer or template is labeled.

[0010] In some embodiments, methods of ligation are provided that provide information about the sequence of a nucleic acid. For example, in some aspects a ligation can be performed in the presence of multiple oligonucleotides that are at least partially complementary to a target region on a template. Oligonucleotide probes can be used that hybridize or otherwise associate with a template adjacent to or near a terminus of a primer or probe that is hybridized to or otherwise associated with a template. Multiple oligonucleotide probes, each at least partially complementary to a region of a template can be used to determine sequence information in a template-dependent manner as is known in the art. For example, oligonucleotide ligation is used to determine nucleic acid sequence information in the SOLiD System (Life Technologies-Applied Biosystems, Carlsbad, CA). According to some embodiments, sequence information is determined ligating oligonucleotide probes to an oligonucleotide primer in a template sequence-dependent manner, for example by using a SFL. The oligonucleotide probes can be a set of multiple probes having different sequences and distinguishing labels, and the primer and probes can have lengths of not more than 8, 7, 6, 5, 4, 3 or 2 nucleotides. The labels can provide information about the sequence of the probe.

[0011] In some embodiments, ligation can be performed by a "small footprint ligase" (herein "SFL") that can ligate short polynucleotides. SFLs can be used in each of the embodiments of ligations discussed above and in the remainder of this disclosure, as well as in other embodiments of ligations known to person of skill in the art. For example, in some embodiments an SFL can ligate the termini of a first oligonucleotide and a second oligonucleotide. The first oligonucleotide can be a primer and the second oligonucleotide can be a probe, each hybridizing or otherwise associating with a portion of a third oligonucleotide or a polynucleotide.

[0012] In some embodiments, the SFL can ligate oligonucleotides that are 8, 7, 6, 5, 4, 3 or 2 nucleotides in length to a polynucleotide. Ligation of such oligonucleotides can be to oligonucleotides of the same length or of different length or to a polynucleotide. For example, an oligonucleotide of 2 or 3 nucleotides in length can be ligated to an oligonucleotide of 2, 3, 4, 5, 6, 7, 8 or more nucleotides in length or to longer oligonucleotides or to a polynucleotide.

[0013] Also provided are kit comprising a small footprint ligase ("SFL") or functional variant or fragment or derivative thereof. Exemplary SFLs are identified herein, and their sequences provided. Optionally, the kit can also include one or more oligonucleotide probes less than 12 nucleotides in length, (e.g., not more than 8, 6, 5, 4, 3 or 2 nucleotides in length). Optionally, the kit includes CV ligase and one or more oligonucleotides probes less than 6 nucleotides in length.

Brief Description of the Drawings

[0014] **Figure 1**. (A) Extended P1 covalent bead ligation substrate. Shown is a ligation substrate designed to mimic the conditions used in SOLiD™ sequencing. Figure 1(A) represents DNA covalently linked to a bead (such as a magnetic bead, e.g., of 0.1 to 1 μm diameter). DNA on this bead was enriched by PCR using the 40 nucleotide P1 sequence. To the region GCGGATGTACGGTACAGCAG, a 20-mer complementary "primer" is annealed that has a 5' PO_4 which reacts with the 3'OH group of the SOLiD probe, as well as a 3' fluorescent label for detection using capillary electrophoresis. To the overhang sequence CGAATAGA, complementary probes can hybridize to demonstrate the ligation reaction. (B) P1 covalent bead ligation substrate. Shown is a similar ligation substrate to that in Figure 1(A), however the reaction in this case will occur much closer to the bead surface. Probes are hybridized to the overhang region AGTCGGTGAT, where the underlined residues are opposite the inosine triplet of probes having the structure Dye-5' III (s)-xy-NNN 3', as described in further detail herein.

[0015] **Figure 2**: Sequences of exemplary ligases. The GenBank ID and source organism are provided where applicable. Three artificial ligases are also provided. DLX differs from Hin DNA Ligase at 1 amino acid -- designated as the underlined emboldened letter. DLXd differs from Hin DNA Ligase at 2 amino acids -- designated as the underlined emboldened letters. DLXd2 is 22 amino acids shorter than Hin DNA Ligase and differs from Hin DNA Ligase at 2 amino acids, designated as the underlined emboldened letters.

[0016] **Figure 3**: Ligation of short oligonucleotides with various ligases. Ligation of short oligonucleotides with (A) CV ligase, (B) DLXd ligase and (C) MnM ligase. Forward ligation reactions were performed under the following conditions : 2.0 μM ligase, 2-5 μM short oligo, 2.0 nM primer/template (tethered to magnetic beads) and proceeded for 20 minutes at 15°C. The ligation efficiency was calculated as the as the ratio of peak areas determined by CE where a FAM labeled primer was used. Efficiency = ligated/(ligated + unligated).

[0017] **Figure 4**: Ligation of short oligonucleotides with various ligases. Ligation of short oligonucleotides with (A) CV ligase, (B) DLXd ligase and (C) MnM ligase. Reverse ligation reactions were performed under the following conditions : 2.0 μM ligase, 2-5 μM short oligo, 2.0 nM primer/template (tethered to magnetic beads) and proceeded for 20 minutes at 15°C. The ligation efficiency was calculated as the as the ratio of peak areas determined by CE where a FAM labeled primer was used. Efficiency = ligated/(ligated + unligated).

[0018] **Figure 5**: Ligation of 2-mers. Ligation reaction in the forward direction was performed under the following conditions : 2.0 μM DLXd, 123 μM dinucleotide (5'-CG-3'), 2.0 nM primer/template (tethered to magnetic beads) and proceeded for 20 minutes at 15°C. The ligation efficiency was calculated as the as the ratio of peak areas determined by CE where a FAM labeled primer was used. Efficiency = ligated/(ligated + unligated).

Definitions

[0019] "Degenerate", with respect to a position in a polynucleotide that is one of a population of polynucleotides, means that the identity of the base of the nucleoside occupying that position varies among different members of the population. A population of polynucleotides in this context is optionally a mixture of polynucleotides within a single continuous phase (e.g., a fluid). The "position" can be designated by a numerical value assigned to one or more nucleotides in a polynucleotide, generally with respect to the 5' or 3' end. For example, the terminal nucleotide at the 3' end of an extension probe may be assigned position 1. Thus in a pool of extension probes of structure 3'-XXXNXXXX-5', the N is at position 4. A position is said to be k-fold degenerate if it can be occupied by nucleosides having any of k different identities. For example, a position that can be occupied by nucleosides comprising either of 4 different bases is 4-fold degenerate.

[0020] Along similar lines, it should be understood that a statement that a result has occurred (e.g., ligation, binding) is intended to indicate that the result has occurred at a significant or substantial level or an enhanced level compared to when it has not occurred. For example. Ligation is said to have <u>not</u> occurred if it is not significant, insubstantial or greatly reduced (e.g., reduced by at least 80%, 90 %, 95% or 99% compared to when ligation does occur (e.g., under the conditions described in the last paragraph).

[0021] The terms “microparticle,” "beads" "microbeads", etc., refer to particles (optionally but not necessarily spherical in shape) having a smallest cross-sectional length (e.g., diameter) of 50 microns or less, preferably 10 microns or less, 3 microns or less, approximately 1 micron or less, approximately 0.5 microns or less, e.g., approximately 0.1, 0.2, 0.3, or 0.4 microns, or smaller (e.g., under 1 nanometer, about 1-10 nanometer, about 10-100 nanometers, or about 100-500 nanometers). Microparticles (e.g., Dynabeads from from Dynal, Oslo, Norway) may be made of a variety of inorganic or organic materials including, but not limited to, glass (e.g., controlled pore glass), silica, zirconia, cross-linked polystyrene, polyacrylate, polymethymethacrylate, titanium dioxide, latex, polystyrene, etc. Magnetization can facilitate collection and concentration of the microparticle-attached reagents (e.g., polynucleotides or ligases) after amplification, and facilitates additional steps (e.g., washes, reagent removal, etc.). In certain embodiments of the invention a population of microparticles having different shapes sizes and/or colors can be used. The microparticles can optionally be encoded, e.g., with quantum dots such that each microparticle can be individually or uniquely identified.

[0022] The term “sequence” refers to sequence information about a polynucleotide or polypeptide or any portion of the polynucleotide or polypeptide that is two or more units (nucleotides or amino acids) long. The term can also be used as a reference to the polynucleotide or polypeptide molecule itself or a relevant portion thereof. Polynucleotide sequence information relates to the succession of nucleotide bases on the polynucleotide, and in a polypeptide relates to the succession of amino acid side chains in the polypeptide or portion thereof. For example, if the polynucleotide contains bases Adenine, Guanine, Cytosine, Thymine, or Uracil, the polynucleotide sequence can be represented by a corresponding succession of letters A, G, C, T, or U), e.g., a DNA or RNA molecule. Sequences shown herein are presented in a 5’→3’ orientation unless otherwise indicated.

[0023] “Perfectly matched duplex” in reference to probes and template polynucleotides means that one forms a double stranded structure with the other such that each nucleoside in the double stranded structure undergoes Watson-Crick basepairing with a nucleoside on the other. The term also comprehends the pairing of nucleoside analogs, such as deoxyinosine, nucleosides with 2-aminopurine bases, and the like, that may be employed to reduce the degeneracy of the probes, whether or not such pairing involves formation of hydrogen bonds.

[0024] The term “polymorphism” is given its ordinary meaning in the art and refers to a difference in genome sequence among individuals of the same species. A “single nucleotide polymorphism” (SNP) refers to a polymorphism at a single position.

[0025] “Probes”, “oligonucleotides” or "primers" are intended to be interchangeable terms herein, so that any one of these can be taken as a reference to another. These are polynucleotides not necessarily limited to any length. Where so wished, these can be less than 100 nucleotides long, sometimes less than 30 nucleotides long, e.g., less than 20 nucleotides, optionally less than 12 nucleotides, for example less than eight nucleotides in length. In some cases, these are 2, 3, 4, 5, 6, 7, 8, or more nucleotides in length. In some cases, these are 3 or 4 nucleotides in length.

[0026] A "polynucleotide," also called a "nucleic acid," is a linear polymer of two or more nucleotides joined by covalent internucleosidic linkages, or variant or functional fragment thereof. A sequence of letters, such as "ATGCCTG," is intended to represent a polynucleotide sequence in the 5'→3' order from left to right unless otherwise specified. In naturally occuring examples of these, the internucleoside linkage is typically a phosphodiester bond. However, other examples optionally comprise other internucleoside linkages, such as phosphorothiolate linkages and may or may not comprise a phosphate group. In other cases, the polynucleotide can contain non-nucleotidic backbones, for example, polyamide (e.g., peptide nucleic acids (PNAs)) and polymorpholino and other synthetic sequence-specific nucleic acid polymers providing that the polymers contain nucleobases in a configuration which allows for base pairing and base stacking, such as is found in DNA and RNA.

[0027] As used herein, "polynucleotide," "oligonucleotide", "probe", "primer", "template", "nucleic acid" and the like can be taken to refer to a populations or pools of individual molecules that are substantially identical across their entire length or across a relevant portion of interest. For example, the term “template” can indicate a plurality of template molecules that are substantially identical, etc. In the case of polynucleotides that are degenerate at one or more positions, it will be appreciated that the degenerate polynucleotide comprises a plurality of polynucleotide molecules, which have sequences that are substantially identical only at the nondegenerate position(s) and differ in sequence at the degenerate positions. Thus, reference to "a" polynucleotide (e.g., "a" primer, probe, oligonucleotide, template, etc.) can be taken to mean a population of polynucleotide molecules that are substantially identical over at least a portion of interest, such that the plural nature of the population need not be explicitly indicated, but can if so desired. These terms are also intended to provide adequate support for a claim that explicitly specifies a single polynucleotide molecule itself. It will be understood that members of a population need not be 100% identical, e.g., a certain number of “errors” may occur during the course of synthesis. Preferably at least 90%, at least 95%, at least 99%, or more of the members of a population are substantially identical. Preferably the percent identity of at least 95% or more preferably at least 99% of the members of the population to a reference nucleic acid molecule is at least 98%, 99%, 99.9% or greater. Percent identity may be computed by comparing two optimally aligned sequences, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions, and multiplying the result by 100 to yield the percentage of sequence identity. It will be appreciated that in certain instances a nucleic acid molecule such as a template, probe, primer, etc., may be a portion of a larger nucleic acid molecule that also contains a portion that does not serve a template, probe, or primer function. In that case individual members of a population need not be substantially identical with respect to that portion.

[0028] The nucleotides of a polynucleotide can have any combination of bases, including those mentioned herein, for example uracil, adenine, thymine, cytosine, guanine, inosine, xathanine hypoxathanine, isocytosine, isoguanine, etc. Optionally, the polynucleotide is a DNA having the nucleotide bases A, C, T and/or G. Optionally, the polynucleotide is an RNA having the nucleotide bases A, C, T and/or U.

[0029] Polynucleotides include double- and single-stranded DNA, as well as double- and single-stranded RNA, DNA:RNA hybrids, peptide-nucleic acids (PNAs) and hybrids between PNAs and DNA or RNA, and also include known types of modifications, for example, labels which are known in the art, methylation, "caps," substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), with negatively charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), and with positively charged linkages (e.g., aminoalklyphosphoramidates, aminoalkylphosphotriesters), those containing pendant moieties, such as, for example, proteins (including nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide or oligonucleotide. Polynucleotides can optionally be attached to one or more non-nucleotide moieties such as labels and other small molecules, large molecules such proteins, lipids, sugars, and solid or semi-solid supports, for example through either the 5' or 3' end. Labels include any moiety that is detectable using a detection method of choice, and thus renders the attached nucleotide or polynucleotide similarly detectable using a detection method of choice. Optionally, the label emits electromagnetic radiation that is optically detectable or visible. In some cases, the nucleotide or polynucleotide is not attached to a label, and the presence of the nucleotide or polynucleotide is directly detected, and/or the generation of byproducts of ligation such as PPi or NMN is detected. Optionally, the presence of the nucleotide, polynucleotide or byproduct is sensed by a chemical field-effect transistor, e.g., where the charge on the gate electrode is generated by a chemical process. Optionally, the chemical field-effect transistor is an ion-sensitive field-effect transistor.

[0030] Where two or more reagents are labeled, the labels are preferably distinguishable from each other with the detection method of choice. For example, the labels can be spectrally resolvable, i.e., distinguishable on the basis of their spectral characteristics, particularly fluorescence emission wavelength, under conditions of operation. In other instances, the label may comprise a signal-generating compound (SGC). A SGC is optionally a substance that it itself detectable in an assay of choice, or capable of reacting to form a chemical or physical entity (i.e., a reaction product) that is detectable in an assay of choice. Representative examples of reaction products include precipitates, fluorescent signals, compounds having a color, and the like. Representative SGC include e.g., bioluminescent compounds (e.g., luciferase), fluorophores (e.g., below), bioluminescent and chemiluminescent compounds, radioisotopes (e.g., ^{131}I, ^{125}I, ^{14}C, ^{3}H, ^{35}S, ^{32}P and the like), enzymes (e.g., below), binding proteins (e.g., biotin, avidin, streptavidin and the like), magnetic particles, chemically reactive compounds (e.g., colored stains), labeled oligonucleotides; molecular probes (e.g., CY3, Research Organics, Inc.), and the like. Representative fluorophores include fluorescein isothiocyanate, succinyl fluorescein, rhodamine B, lissamine, 9,10-diphenlyanthracene, perylene, rubrene, pyrene and fluorescent derivatives thereof such as isocyanate, isothiocyanate, acid chloride or sulfonyl chloride, umbelliferone, rare earth chelates of lanthanides such as Europium (Eu) and the like. Signal generating compounds also include SGC whose products are detectable by fluorescent and chemiluminescent wavelengths, e.g., sequencing dyes, luciferase, fluorescence emitting metals such as ^{152}Eu, or others of the lanthanide series; compounds such as luminol, isoluminol, acridinium salts, and the like; bioluminescent compounds such as luciferin; fluorescent proteins (e.g., GFP or variants thereof); and the like. The subject SGC are optionally detectable using a visual or optical method; preferably, with a method amenable to automation such as a spectrophotometric method, a fluorescence method, a chemiluminescent method, an electrical nanometric method involving e.g., a change in conductance, impedance, resistance and the like and a magnetic field method. Some SGC's are optionally detectable with the naked eye or with a signal detection apparatus when at an appropriate concentration.

[0031] A "nucleotide" refers to a nucleotide, nucleosideor analog thereof. Optionally, the nucleotide is an N- or C-glycoside of a purine or pyrimidine base. (e.g., deoxyribonucleoside containing 2-deoxy-D-ribose or ribonucleoside containing D-ribose). Examples of other analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides.

[0032] Nucleotide bases or nucleobases usually have a substituted or unsubstituted parent aromatic ring or rings. In certain embodiments, the aromatic ring or rings contain at least one nitrogen atom. In certain embodiments, the nucleotide base is capable of forming Watson-Crick and/or Hoogsteen hydrogen bonds with an appropriately complementary nucleotide base. Exemplary nucleotide bases and analogs thereof include, but are not limited to, purines such as 2-aminopurine, 2,6-diaminopurine, adenine (A), ethenoadenine, N6-Δ2-isopentenyladenine (6iA), N6-Δ2-isopentenyl-2-methylthioadenine (2ms6iA), N6-methyladenine, guanine (G), isoguanine, N2-dimethylguanine (dmG), 7-methylguanine (7 mG), 2-thiopyrimidine, 6-thioguanine (6sG) hypoxanthine and O6-methylguanine; 7-deaza-purines such as 7-deazaadenine (7-deaza-A) and 7-deazaguanine (7-deaza-G); pyrimidines such as cytosine (C), 5-propynylcytosine, isocytosine, thymine (T), 4-thiothymine (4sT), 5,6-dihydrothymine, O4-methylthymine, uracil (U), 4-thiouracil (4sU) and 5,6-dihydrouracil (dihydrouracil; D); indoles such as nitroindole and 4-methylindole; pyrroles such as nitropyrrole; nebularine; base (Y); etc. In certain embodiments, nucleotide bases are universal nucleotide bases. Additional exemplary nucleotide bases can be found, e.g., in Fasman, 1989, Practical Handbook of Biochemistry and Molecular Biology, pp. 385-394, CRC

Press, Boca Raton, Fla., and the references cited therein. A nucleoside is usually a compound having a nucleotide base covalently linked to the C-1′ carbon of a pentose sugar. In certain embodiments, the linkage is via a heteroaromatic ring nitrogen. Typical pentose sugars include, but are not limited to, those pentoses in which one or more of the carbon atoms are each independently substituted with one or more of the same or different —R, —OR, —NRR or halogen groups, where each R is independently hydrogen, (C1-C6) alkyl or (C5-C14) aryl. The pentose sugar may be saturated or unsaturated. Exemplary pentose sugars and analogs thereof include, but are not limited to, ribose, 2′-deoxyribose, 2′-(C1-C6)alkoxyribose, 2′-(C5-C14)aryloxyribose, 2′,3′-dideoxyribose, 2′,3′-didehydroribose, 2′-deoxy-3′-haloribose, 2′-deoxy-3′-fluororibose, 2′-deoxy-3′-chlororibose, 2′-deoxy-3′-aminoribose, 2′-deoxy-3′-(C1-C6)alkylribose, 2′-deoxy -3′-(C1-C6)alkoxyribose and 2′-deoxy-3′-(C5-C14)aryloxyribose. One or more of the pentose carbons of a nucleoside may be substituted with a phosphate ester, as disclosed in US Pat. No 7255994. In certain embodiments, the nucleosides are those in which the nucleotide base is a purine, a 7-deazapurine, a pyrimidine, a universal nucleotide base, a specific nucleotide base, or an analog thereof. Nucleotide analogs include derivatives in which the pentose sugar and/or the nucleotide base and/or one or more of the phosphate esters of a nucleoside may be replaced with its respective analog. Exemplary pentose sugar analogs and nucleotide base analog are described above. Exemplary phosphate ester analogs include, but are not limited to, alkylphosphonates, methylphosphonates, phosphoramidates, phosphotriesters, phosphorothioates, phosphorodithioates, phosphoroselenoates, phosphorodiselenoates, phosphoroanilothioates, phosphoroanilidates, phosphoroamidates, boronophosphates, etc., and may include associated counterions. Other nucleotide analogs are nucleotide analog monomers which can be polymerized into polynucleotide analogs in which the DNA/RNA phosphate ester and/or sugar phosphate ester backbone is replaced with a different type of linkage. Exemplary polynucleotide analogs include, but are not limited to, peptide nucleic acids, in which the sugar phosphate backbone of the polynucleotide is replaced by a peptide backbone.

[0033] The internucleoside linkages can be a phosphodiester linkage, although other linkages (e.g., scissile linkages which can be substantially cleaved under conditions in which phosphodiester linkages are not substantially cleaved) can be used. For example, a linkage that contains an AP endonuclease sensitive site, for example an abasic residue, a residue containing a damaged base that is a substrate for removal by a DNA glycosylase, or another residue or linkage that is a substrate for cleavage by an AP endonuclease, or a disaccharide nucleoside.

[0034] The adjectival term "hybridized" optionally refers to two polynucleotides which are bonded to each other by two or more sequentially adjacent base pairings. The term "hybridization" refers to the process by which polynucleotides become hybridized to each other. Two single-stranded polynucleotides can be regarded as "complementary" if when hybridized together the longer polynucleotide forms a single-stranded overhang and the shorter polynucleotide can be efficiently ligated to a third adjacent polynucleotide that forms a perfectly-matched duplex with the single-stranded overhang. Where the single-stranded overhang is less than eight nucleotides, it can be arbitrarily lengthened to eight nucleotides by adding a random combination of nucleotides to the overhang.

[0035] Similarly, nucleotide residues can be regarded as complementary if when both are base-paired with each other within two hybridized polynucleotides, either nucleotide can be ligated in a template-driven ligation reaction when situated as the terminal nucleotide in its polynucleotide. Nucleotides that are efficiently incorporated by DNA polymerases opposite each other during DNA replication under physiological conditions are also considered complementary. In an embodiment, complementary nucleotides can form base pairs with each other, such as the A-T/U and G-C base pairs formed through specific Watson-Crick type hydrogen bonding between the nucleobases of nucleotides and/or polynucleotides positions antiparallel to each other. The complementarity of other artificial base pairs can be based on other types of hydrogen bonding and/or hydrophobicity of bases and/or shape complementarity between bases.

[0036] In appropriate instances, polynucleotides can be regarded as complementary when they can undergo cumulative base pairing at two or more individual corresponding positions in antiparallel orientation, as in a hybridized duplex. Optionally there can be “complete” or “total” complementarity between a first and second polynucleotide sequence where each nucleotide in the first polynucleotide sequence can undergo a stabilizing base pairing interaction with a nucleotide in the corresponding antiparallel position on the second polynucleotide. "Partial" complementarity describes polynucleotide sequences in which at least 20%, but less than 100%, of the residues of one polynucleotide are complementary to residues in the other polynucleotide. A "mismatch" is present at any position in which the two opposed nucleotides are not complementary. In some ligation assays, a polynucleotide can undergo substantial template-dependent ligation even when it has one or more mismatches to its hybridized template. Optionally, the polynucleotide has no more than 4, 3, or 2 mismatches, e.g., 0 or 1 mismatch, with its template. In some assays, the polynucleotide will not undergo substantial template-dependent ligation unless it is at least 60% complementary, e.g., at least about 70%, 80%, 85%, 90%, 95%, 99% or 100% complementary to its template.

[0037] As used herein, a "biological sample" refers to a sample of tissue or fluid isolated from an individual, including but not limited to, for example, plasma, serum, spinal fluid, semen, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, blood cells, tumors, organs, and also samples of in vitro cell culture constituents (including but not limited to conditioned medium resulting from the growth of cells in cell culture medium, putatively virally infected cells, recombinant cells, and cell components).

[0038] Sequence identity (also called homology) refers to similarity in sequence of two or more sequences (e.g., nucleotide or polypeptide sequences). In the context of two or more homologous sequences, the percent identity or homology of the sequences or subsequences thereof indicates the percentage of all monomeric units (e.g., nucleotides or amino acids) that are the same (i.e., about 70% identity, preferably 75%, 80%, 85%, 90%, 95% or 99% identity). The percent identity can be over a specified region, when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection. Sequences are said to be "substantially identical" when there is at least 90% identity at the amino acid level or at the nucleotide level. Preferably, the identity exists over a region that is at least about 2, 3, 4, 5, 6, 7, 8, 10, 12, 15, 20, 25, 50, or 100 residues in length, or across the entire length of at least one compared sequence. A preferred algorithm for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al, Nuc. Acids Res. 25:3389-3402 (1977). Other methods include the algorithms of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), and Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), etc. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions.

[0039] In the claims, any active verb (or its gerund) is intended to indicate the corresponding actual or attempted action, even if no actual action occurs. For example, the verb "hybridize" and gerund form "hybridizing" and the like refer to actual hybridization or to attempted hybridization by contacting nucleic acid sequences under conditions suitable for hybridization, even if no actual hybridization occurs. Similarly, "detecting" and "detection" when used in the claims refer to actual detection or to attempted detection, even if no target is actually detected.

[0040] "Nonspecific hybridization" is used to refer to any unintended or insignificant hybridization, for example hybridization to an unintended polynucleotide sequence other than the intended target polynucleotide sequence. The uninentended polynucleotide sequence can be on the same or different polynucleotide from the intended target. In some cases, the only intended hybridization can be from Watson-Crick base pairing between two polynucleotides. Other kinds of intended base pairings can include base pairing between corresponding analogs of such nucleotides or between iso-cytidine and iso-guanine. In some cases where hybridization is only intended between complementary bases, any bonding between non-complementary bases is considered to be non-specific hybridization.

[0041] In reference to ligation of two polynucleotides, the "proximal" terminus of either polynucleotide is the terminus that is intended to be ligated to the other polynucleotide. This is generally the terminus that is contacted by the active site of the ligase, or the terminus that is eventually ligated to the other polynucleotide, while the opposite terminus is the "distal" terminus. The terminal nucleotide residue at the proximal terminus can be termed the proximal nucleotide, and the proximal nucleotide position optionally designated as position 1, or -1 depending on which side of the ligation site we are referring to, the penultimate nucleotide position as position 2 or -2, etc. With reference to two adjacently-hybridized polynucleotides, the proximal terminus is generally the terminus of one polynucleotide that is closer to the other polynucleotide. In some non-limiting instances of template-dependent ligation, the proximal termini of both polynucleotides are hybridized adjacently to each other.

[0042] “Support”, as used herein, refers to a structure or matrix on or in which ligation reagents, e.g., nucleic acid molecules, microparticles, and the like may be immobilized so that they are significantly or entirely prevented from diffusing freely or moving with respect to one another. The reagents can for example be placed in contact with the support, and optionally covalently or noncovalently attached or partially/completely embedded

[0043] A “universal base”, as used herein, is a base that is complementary to more than one other base. Fully universal bases can pair with any of the bases typically found in naturally occurring nucleic acids. The base need not be equally capable of pairing with each of the naturally occurring bases. Alternatively, the universal base may pair only or selectively with two or more bases but not all bases. Optionally the universal base pairs only or selectively with purines, or alternatively with pyrimidines. If so desired, two or more universal bases can be included at a particular position in a probe. A number of universal bases are known in the art including, but not limited to, inosine, hypoxanthine, 3-nitropyrrole, 4- nitroindole, 5-nitroindole, 4-nitrobenzimidazole, 5-nitroindazole, 8-aza-7-deazaadenine, 6H,8H-3,4-dihydropyrimido[4,5-c][1,2]oxazin-7-one, 2-amino-6-methoxyaminopurine, etc. Hypoxanthine is one preferred fully univeral base. Nucleosides comprising hypoxanthine include, but are not limited to, inosine, isoinosine, 2'-deoxyinosine, and 7-deaza-2'-deoxyinosine, 2-aza-2'deoxyinosine.

[0044] "Purified" generally refers to isolation of a substance (compound, polynucleotide, protein, polypeptide, polypeptide composition) such that the substance comprises a significant percent, such as a higher proportion than it is naturally found (e.g., greater than 2%, greater than 5%, greater than 10%, greater than 20%, greater than 50%, or more, sometimes more than 90%, 95% or 99%) of the sample in which it resides. In certain embodiments, a substantially purified component comprises at least 50%, 80%-85%, or 90-95% of the sample. Techniques for purifying polynucleotides and polypeptides of interest are well-known in the art and include, for example, ion-exchange chromatography, affinity chromatography and sedimentation according to density. "Isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original or naturally-occurring environment of the polynucleotide.

[0045] Exemplary ligases comprise a polypeptide. The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues, or any variant or functional fragment thereof. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers.

[0046] The term "amino acid" includes naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, .gamma.-carboxyglutamate, and O-phosphoserine. Amino acid analogs refer to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., a carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

[0047] Variants or derivatives of a given nucleotide sequence or polypeptide sequence are optionally conservatively modified variants. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences.

[0048] As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention. (see e.g., Creighton, Proteins (1984)).

Sequence identity (also called homology) refers to similarity in sequence of two or more sequences (e.g., nucleotide or polypeptide sequences). In the context of two or more homologous sequences, the percent identity or homology of the sequences or subsequences thereof indicates the percentage of all monomeric units (e.g., nucleotides or amino acids) that are the same (i.e., about 70% identity, preferably 75%, 80%, 85%, 90%, 95% or 99% identity). The percent identity can be over a specified region, when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection. Sequences are said to be "substantially identical" when there is at least 90% identity at the amino acid level or at the nucleotide level. This definition also refers to the complement of a test sequence. Preferably, the identity exists over a region that is at least about 25, 50, or 100 residues in length, or across the entire length of at least one compared sequence. A preferred algorithm for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al, Nuc. Acids Res. 25:3389-3402 (1977). Other methods include the algorithms of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), and Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), etc. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements specifically hybridize to each other under stringent conditions, such as those described herein. Nucleic acids that do not specifically hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical.

[0049] Two polynucleotides selectively (or specifically) hybridize to each other if they bind significantly or detectably to each other under stringent hybridization conditions when present in a complex polynucleotide mixture such as total cellular or library DNA. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization. Optionally, stringent conditions are selected to be about 5-10° C. lower than the thermal melting point for the specific sequence at a defined ionic strength pH. Stringent conditions are optionally in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C. for short probes (e.g., 10 to 50 nucleotides) and at least about 60° C. for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5.times.SSC, and 1% SDS, incubating at 42° C., or, 5.times.SSC, 1% SDS, incubating at 65° C., with wash in 0.2.times.SSC, and 0.1% SDS at 65° C.

Detailed Description

[0050] Among other novel and surprising information presented herein, the novel and surprising enzymatic ligation of short polynucleotides is presented herein.

1) LIGATIONS

[0051] Ligation herein refers to the enzymatic formation of a covalent bond between the termini of two or more or polynucleotides strands. “Ligation” involves the formation of a covalent bond or linkage between the 5’ and 3’ termini of two or more nucleic acids, e.g. oligonucleotides and/or polynucleotides, optionally in a template-driven reaction. The nature of the bond or linkage may vary widely and the ligation is preferably achieved enzymatically. The nature of the bond or linkage can vary widely. Non-limiting exemplary ligations are carried out enzymatically to form a phosphodiester linkage between a 5′ terminal nucleotide of a polynucleotide strand with a 3′ terminal nucleotide of a polynucleotide.

[0052] The ligation can be one or more of the following types of ligation described herein. A first type of enzymatic ligation involves the formation of a covalent bond between a first terminus of a first polynucleotide strand and a second different terminus of a second polynucleotide strand. The first and second polynucleotide termini can be on different polynucleotide strands, or can both be on the same polynucleotide strand (resulting in circularization). Optionally, the first and second polynucleotide strands are not both hybridized to a third polynucleotide. Optionally, the termini of the first and second polynucleotide strand are joined irrespective of their sequences (e.g., blunt-end ligation, or non-homologous end joining). In another variation, two double-stranded polynucleotides with protruding single-stranded portions that are complementary to each other can be ligated (e.g., cohesive-end ligation). A third type of ligation (template-dependent ligation) is described further below.

[0053] In any one of the methods described herein, the polynucleotide strands can be in single-stranded format or are hybridized to complementary strands in double-stranded format. In blunt-end ligation, both polynucleotide strands to be ligated are hybridized to two different complementary strands such that no overhang exists.

[0054] Methods are provided to ligate two polynucleotides. An exemplary ligation method achieves ligation between a first terminus of a first polynucleotide sequence and a second terminus of a second polynucleotide sequence. For ease of reference, the first polynucleotide sequence is called the "initializing probe" or "primer," the second polynucleotide called the "extension probe" or "probe." A third polynucleotide sequence that is optionally present (e.g., in template-dependent ligation) is termed the template or target sequence. Any one or more of the primer (initializing probe), probe (extension probe) and/or template sequences can be located on the same polynucleotide strand, or on different polynucleotide strands. In a one-strand system, the primer, probe and template are all on the same polynucleotide strand. In an example of a two-strand system, either the probe or primer (but not both) are on the same strand as the template (e.g., hybridization between the template sequence and the primer or probe sequences forms a stem-loop structure or hairpin). In another example of a two-strand system, the probe or primer are both on the same strand. In a three-strand system, the template, primer and probe are all on separate polynucleotide strands. Any ligation method described herein can be performed in a one-strand, two-strand or three-strand system.

[0055] Optionally, ligation converts a linear polynucleotide strand into a circular polynucleotide strand (e.g., in a one-stranded to two-stranded system). Optionally, ligation reduces the number of polynucleotide strands by one (e.g., in a two-stranded or three-stranded system).

[0056] Ligation optionally creates a bond between a terminal nucleotide of the probe with the terminal nucleotide of the primer. Optionally, the proximal terminus of the primer and/or probe is ligated. The terminal nucleotide of the primer can be the 5' terminal nucleotide and the terminal nucleotide of the extension probe can be the 3' terminal nucleotide. Alternatively, the terminal nucleotide of the primer can be the 3' terminal nucleotide and the terminal nucleotide of the extension probe can be the 5' terminal nucleotide. The 5' terminus of a polynucleotide for example has the fifth carbon in the sugar-ring of the deoxyribose or ribose at its terminus, optionally with a phosphate group attached to it, where the phosphate group is capable of forming a phosphodiester bond with a 3' terminal nucleotide. The 3' terminal nucleotide optionally has a 3′-hydroxyl group that is capable of forming a phosphodiester bond with a 5' terminal nucleotide. Ligation optionally results in the formation of a phophodiester bond.

[0057] Any of the polynucleotides no matter how designated (e.g., as "probes" or "primers" or "templates") can be of any sequence, any length, in any form and from any source. The polynucleotides can comprise a naturally-occurring sequence or be highly homologous to a naturally-occurring sequence, and/or derived from a naturally occurring sequence. The naturally occurring sequence can be any portion of a gene, a regulatory sequence, genomic DNA or fragment, cDNA, RNA including mRNA and rRNA, or others. The polynucleotides can optionally comprise any artificial sequence as well. The polynucleotide can be derived or obtained from a sample such as a diagnostic sample. The polynucleotide can be a secondary product of a reaction – for example a ligation product from a ligation reaction or assay such as those described herein, an extended probe from a PCR reaction, or PCR amplification product, ("amplicon"), the product of an invasive cleavage reaction, etc. The polynucleotide can have a 5' phosphate, or can alternatively lack a 5' phosphate.

[0058] The ligation product of any one reaction can optionally be subjected to further ligation and/or non-ligation reactions in turn. For example, the ligation product can be used as the primer (initializing probe) or extension probe and/or template in a subsequent ligation. Also for example, it can be used as a template and/or primer for a polymerase extension reaction, such as in PCR. The probe, primer, template and/or ligation product optionally can be subjected to any one or more modifications before or after ligation. For example, the probe, primer, template and/or ligation product can be cleaved enzymatically or chemically (for example at scissile linkages), treated with exo- or endonucleases, kinases, phosphatases, etc. The ends of a double-stranded product can be blunt-ended or filled in, capped, or adenylated, etc.

[0059] Optionally, the probe is not more than 20 consecutive nucleotides long, for example not more than 15, 12, 10, 8, 7 consecutive nucleotides, preferably not more than 6, 5, 4 3, or 2 consecutive nucleotides. Optionally, the probe is at least 2, 3, 4, 5, 6 or 7 nucleotides long. In some instances the probe is from any of the specified minimum lengths (e.g., 2, 3, 4, 5, 6 or 7 nucleotides long) and is not more than 20, 15, 12, 10, 8, 7 nucleotides, preferably not more than 6, 5, 4 3, or 2 nucleotides. In some examples, the probe is a 2-mer, 3-mer, 4-mer, 5-mer, 6-mer, 7-mer, 8-mer, 9-mer, 10-mer, 11-mer, 12-mer, 13-mer, 14-mer, 15-mer or 20-mer, or any combination of such oligonucleotides.

[0060] Optionally, the primer is not more than 20 consecutive nucleotides long, for example not more than 15, 12, 10, 8, 7 consecutive nucleotides, preferably not more than 6, 5, 4, 3, or 2 consecutive nucleotides. In some examples, the primer is one or more 2-mer, 3-mer, 4-mer, 5-mer, 6-mer, 7-mer, 8-mer, 9-mer, 10-mer, 11-mer, 12-mer, 13-mer, 14-mer, 15-mer or 20-mer, or any combination of such oligonucleotides.

[0061] It should be understood that the probe or primer (or template, if present) can be "mixed" or "composite," i.e., comprising a mixture of one or more polynucleotides of different sequences.

[0062] Optionally, the ligation is performed using CV ligase in combination with one or more probes that is at least 3 nucleotides long and not more than 6 or 5 nucleotides long. Optionally, the one or more probes are at least 3 nucleotides long and not more than 4 nucleotides long. In other examples the one or more probes are at least 4 nucleotides long and not more than 5 nucleotides long. Alternatively all of the one or more probes probe can be 3-mers. Otherwise all of the one or more probes probe can be 4-mers.

[0063] When the ligation is performed using DLX, DLXd, DLXd2 ligase, then optionally ligation is performed with one or more probes that is at least 2 or 3 nucleotides long and not more than 6 or 5 nucleotides long. Optionally, the one or more probes are at least 3 nucleotides long and not more than 4 nucleotides long. In other examples the one or more probes are at least 4 nucleotides long and not more than 5 nucleotides long. Alternatively all of the one or more probes probe can be 3-mers. Otherwise all of the one or more probes probe can be 4-mers.

[0064] The ligation can be a template-dependent ligation. In template-dependent ligation, ligation between a primer sequence and a probe sequence occurs upon hybridization of at least a portion of either or both sequences to a template sequence. In some instances, both probes must hybridize to the template for significant ligation to occur. In a typical example, template-dependent ligation cannot take place unless both polynucleotides are hybridized to the template sequence. The portion of the primer or probe that is hybridized to the target sequence is generally at least two nucleotides long. The hybridized portion is optionally not more than 20 consecutive nucleotides long, for example not more than 15, 12, 10, 8, 7 consecutive nucleotides, preferably not more than 6, 5, 4, 3, or 2 consecutive nucleotides. The hybridized portion is optionally a terminal portion of the first or second polynucleotide (e.g., a portion that includes the 5' or 3' terminal nucleotide). For example, the hybridized portion can consist of the terminal 2, 3, 4, 5, 6, 7, 8, 10, 15 or 20 nucleotides of the 5' or 3' end.

[0065] Optionally, ligation occurs when no mismatch is present within one or more hybridized portions. In other cases, ligation occurs when one, two or three mismatches can be present within one or more hybridized portions. In some cases ligation does not occur when the terminal nucleotide and/or second-most terminal nucleotide and/or third-most terminal nucleotide is mismatched. As mentioned, the terminal nucleotides can be the 5'- or 3'- terminal nucleotides of the polynucleotide.

[0066] Optionally the template, if present, is not more than 11 nucleotides in length, for example not more 10, 9, 8, 7, 6, 5, or 4 nucleotides. Optionally the template is one or more N-mers, where N is 4, 5, 6, 7, 8, 9, 10 or 11.

[0067] Optionally, template-dependent ligation of a nucleic acid comprises: a) providing a first oligonucleotide having less than 6 nucleotides; b) providing a second oligonucleotide; c) bringing the 3' termini of one of the first and second oligonucleotides into proximity with the 5' termini of the other oligonucleotide; and d) ligating the first and the second oligonucleotides. Optionally, the first oligonucleotide has a length of 5 nucleotides. Optionally, the first oligonucleotide has a length of 4 nucleotides. Optionally, the first oligonucleotide has a length of 3 nucleotides. Optionally, the ligation is performed using a small-footprint ligase (SFL). Optionally, the second oligonucleotide includes a sequence complementary to a portion of a template nucleic acid. Optionally, the second oligonucleotide is hybridized to the template nucleic acid at the region of complementarity. Optionally, the first oligonucleotide has a sequence complementary to the template nucleic acid. Optionally, the first oligonucleotide is hybridized to the template nucleic acid at the region of complementarity, and wherein a terminus of the first oligonucleotide is adjacent to a terminus of the second oligonucleotide.

[0068] In some variations, (e.g., "nick ligation" or "template-dependent" ligation), both primer and probe must hybridize adjacently to each other on the template for ligation to occur. Optionally, the probe and primer are adjacently hybridized and can be ligated only when a terminal nucleotide of the primer is hybridized to a first nucleotide of the template and a terminal nucleotide of the extension probe is hybridized to a second nucleotide of the template, where the first and second nucleotides on the template are not separated by an intervening nucleotide of the template. In other embodiments, intervening nucleotides may be present between the first and second nucleotides on the template (optionally a few nucleotides, e.g., not more than 1, 2, 3, 5, 10 or 15 nucleotides). In such embodiments, a "gap-filling" step can be performed to extend the 3' terminus of the probe or probe before it can be ligated to the 5' terminus of the other.

[0069] Optionally, at least one of the probe, the template (if the ligation is template-dependent) and/or the primer is immobilized while another of these three is labeled. For example in ligation sequencing the template and/or primer can be immobilized and the probe can be labeled.

[0070] A probe can for example be N nucleotide residues in length, where N is from 2 to 10, e.g., 2, 3, or 4. N can also be less than 6, for example if the proximal terminus of the probe is its 3' terminus.

[0071] Optionally, ligation is a "forward" ligation (i.e., ligation of the 3' terminus of the probe to the 5' terminus of the primer). Alternatively, "reverse" ligation can be achieved, where the 5' terminus of the probe is ligated to the 3' terminus of the primer.

[0072] A probe can also be of length N, and can comprise a proximal portion that is perfectly hybridized to the template and is L nucleotides long, where the probe's L+1th nucleotide is mismatched with the template. L can be, for example, from 2 to 8, and furthermore can be less than 6 if the proximal terminus of the probe is its 3' terminus. In other embodiments, L can be 2, 3, 4, 5, 6, 7 or 8.

[0073] The ligation can be repeated at least once, for any desired number of cycles. Optionally, any ligation product of a previous ligation reaction is used as the primer of a next ligation. Optionally, all ligations extend the ligation product in the same direction. For example, all ligation reactions can be forward ligations or all can be reverse ligations. In other embodiments, some ligations can be forward and some reverse. Optionally any unligated primer is rendered unligatable before initiating the next ligation. If so desired (e.g., in ligation sequencing), the method further comprises detecting whether the probe has ligated to the primer before repeating the next ligation reaction.

[0074] Where so desired, any ligation product of the previous ligation reaction can be used as the template of the next ligation reaction, e.g., in a ligase chain reaction.

[0075] Optionally, the 5' end of the probe less than 6 nucleotides long is ligated to the 3' end of the primer by a SFL such as CV ligase. For example the probe is 2, 3 or 4 nucleotides long. If so desired, the primer is also a short oligonucleotides. For example, the primer can be less than 6 nucleotides, e.g., 3 or 4 nucleotides long.

[0076] The ligation should produce a significant or detectable amount of ligation product. Optionally, the efficiency of ligation is at least 5%, sometimes at 10%, 20%, 30%, 50%, 60%, 70%, 75%, 80%, 85%, 90% or 95%. The efficiency of ligation (in percentage terms) can in some embodiment be regarded as the percent portion of ligation reagent that is ligated to form ligation product at the end of the ligation reaction. Efficiency is optionally determined after a ligation reaction has reach equilibrium such that increasing the ligation time will not result in a substantial increase of ligation product formed. Generally, a ligation reaction can be said to have reached equilibrium after 20 minutes. For example, a ligation reaction in which 90% of the primer or probe is ligated can be said to have proceeded at 90% efficiency. The ligation reagent used to measure ligation efficiency is optionally whichever reagent that is in lower concentration than the others. Optionally, the other reagents and conditions are non-limiting to the ligation efficiency (e.g., other reagents are present in excess or at a concentration that is at least equal to or higher than the limiting reagent).

[0077] In some embodiments, the SFL can ligate a short probe that is shorter than N nucleotides at least X% as efficiently as the SFL can ligate the corresponding N-mer. N is optionally 4, 5, 6, 7, 8, 10, 12, 15 or 20. In some embodiments N is 6 or 7. X is optionally at least 30%, 50%, 60%, 70%, 80% or 90%. In some embodiments of template-dependent ligation, the length of short probe is Y residues, and the Y proximal residues of the corresponding longer N-mer are identical to the probe, and the distal N-Y residues of the corresponding longer N-mer are perfectly complementary to the template. Y is for example 2, 3, 4, 5 or 6.

[0078] In some embodiments, the SFL can ligate a short probe that is shorter than N nucleotides when the short probe is conjugated to a dye approximately as efficiently as the SFL can ligate the unconjugated probe. N is optionally 4, 5, 6, 7, 8, 10, 12, 15 or 20. In some embodiments N is 6 or 7. In other embodiments, the SFL can ligate the short probe conjugated to a dye at least 30%, 50%, 60%, 70%, 80% or 90% as efficiently as the SFL can ligate the unconjugated probe. Exemplary dyes include Cy5.

[0079] It is understood that ligation efficiency (whether expressed in absolute or relative terms) may increase or decrease depending on the exact reaction conditions used. Optionally, the ligation efficiency is measured when the primer or probe is at a concentration of about 10^{-9} to 10^{-4} M, for example at about 10^{-8}, 10^{-7}, 10^{-6}, or 10^{-5} M. In certain applications, for example sequencing applications, a working concentration often used in the range of 10^{-8} to 10^{-5}, e.g., 10^{-7}or 10^{-6} M of probe. Where a mixture of probes are used, the probe concentration is the concentration of only those probes that are capable of being ligated in that particular ligation reaction (e.g., probe(s) complementary to the template). Thus other probes that are not capable of participating in the ligation reaction are optionally not considered when calculating the concentration of the probe of interest. Optionally, the concentration of probe is between 1 picomolar and 1 millimolar, for example about 0.01-100 µM, e.g., 1-10 µM, e.g., 1-5 µM. In the case of 2-mers, the concentration is optionally increased to 10-1000 µM, for example about 100 µM. Optionally, the concentration of ligase is between 1 picomolar and 1 millimolar, e.g., 1-2 micromolar.. Optionally, the ligation assay is performed at 15-35°C, for example 15 °C, 20 °C, 25 °C or 30 °C. Where two or more reagents are involved and the concentration of one particular reagent is specified, the other reagents are optionally in excess and/or not at a concentration that is limiting for the ligation.

[0080] Ligation can be performed under in vitro conditions that have been experimentally determined to be suitable or optimal for ligase activity. Preferably, the reaction conditions of choice are (i) substantially similar to in vivo or physiological conditions in which a naturally-occurring form of the ligase being used is active, or (ii) have been experimentally determined to result in a ligation efficiency that is comparable to or better than the efficiency obtained using conditions of type (i). If exemplary in vitro ligation conditions are specified herein for a particular SFL, then substantially similar reaction conditions are generally appropriate for that particular ligase. In other embodiments, the conditions are such that the reference ligation assay produces significant or detectable ligation within 30 minutes, within 10 minutes, within 1 minute, or within ten seconds. Another non-limiting example of a significant or detectable efficiency of ligation generates in the range of 100 pM of ligation product, optionally about 1000 pM or 10,000 pM.

[0081] Optionally, relative efficiency can be expressed as relative percent efficiency, which can be calculated as A/B x 100, where A is the percent of test reagent (e.g., probe) ligated in a test assay and B is the percent of the reference reagent (e.g., reference probe) that is ligated in a reference assay. Where the relative efficiency of ligation is specified in comparative or relative terms by comparison to a reference ligation assay, it is implicit that all other reagents and conditions (e.g., temperature, concentration of all reagents, pH, concentration of requisite ions such as Mg++ and Mn++, concentration of requisite cofactors such as NAD and/or ATP, salts, buffers, molar concentrations of all reagents, including enzyme, template, probe, primer, oligonucleotides, etc) are otherwise kept identical. For example, a proviso that a SFL can ligate a short (e.g., less than 6 nucleotides) probe at least X% as efficiently as the SFL can ligate a corresponding octanucleotide, can be taken to mean that the two different ligation assays all reagents except for the probes (e.g., primer, template, enzymes and any other reagents) and all reaction conditions (e.g., temperature, reagent concentrations, concentrations of any other reagents, etc) are kept the same for practical purposes.

[0082] Optionally, the ligase has a better ligation efficiency than T4 DNA ligase, for example in any method described herein. In an embodiment, the ligation efficiency is higher than that of T4 DNA ligase for the same mix of probe(s), primer(s) and template(s), in any ligation method and chosen conditions, including any described herein. The ligation efficiency is for example at least 5% higher than T4 ligase, optionally at least 10%, 15% or 20% higher. The increase in efficiency should be statistically reliable and significant, e.g., with a confidence interval of at least 95%, 99%, 99.9%, 99.99%, or 99.999999%. In an example, the SFL shows higher efficiency than T4

DNA ligase when ligating a probe of 8 nucleotides or less to a primer in the forward or reverse direction. The ligase, template, probe, primer and/or ligation assay conditions are for example any described herein.

[0083] The ligase optionally has good ligation fidelity. The ligation fidelity can be assessed as the percentage of incorrect ligation events for a given combination of ligase, template, primer and probe. An incorrect ligation event is one in which the probe or the primer is not perfectly complementary to the template, or the ligation product is not perfectly complementary to the template. In an embodiment, the ligation fidelity is higher than that of T4 DNA ligase for the same mix of probe(s), primer(s) and template(s), in any ligation method and chosen conditions, including any described herein. The ligation fidelity is for example at least 5% higher than T4 ligase, optionally at least 10%, 15% or 20% higher. The increase in fidelity should be statistically reliable and significant, e.g., with a confidence interval of at least 95%, 99%, 99.9%, 99.99%, or 99.999999%. In an example, the SFL shows higher fidelity than T4 DNA ligase when ligating a probe of 8 nucleotides or less to a primer in the forward or reverse direction. The ligase, template, probe, primer and/or ligation assay conditions are for example any described herein.

2) SMALL FOOTPRINT LIGASES

[0084] Optionally, the enzymatic ligation of polynucleotides is achieved by a small footprint ligase (SFL). For example, any ligation method provided herein can use a ligase shown in Table 1A, 1B or 1C. A SFL is a ligase that can ligate short polynucleotides. As used herein, the term "ligase" is intended to include any fragment or variant or derivative of that ligase. The fragment or variant or derivative optionally possesses one or more functional activities of a ligase. A SFL optionally comprises a polypeptide having any one or more of the following activities: (1) nucleophilic attack on ATP or NAD^+ resulting in release of PPi or NMN and formation of a covalent ligase-adenylate intermediate; (2) transferring the adenylate to the 5'-end of the 5'-phosphate-terminated DNA strand to form DNA-adenylate (e.g., the 5'-phosphate oxygen of the DNA strand attacks the phosphorus of ligase-adenylate); and (3) formation of a covalent bond joining the polynucleotide termini and liberation of AMP (e.g., by the attack by the 3'-OH on DNA-adenylate). Optionally, the SFL can mediate any one or more of the following bond transformations: from phosphoanhydride (ATP) to phosphoramidate (ligase-adenylate); from phosphoramidate (ligase-adenylate) to phosphoanhydride (DNA-adenylate); or from phosphoanhydride (DNA-adenylate) to phosphodiester (sealed DNA). Thus, exemplary SFLs can comprise a polypeptide sequence that is homologous to or a variant of a known SFL sequence or any portion thereof. Exemplary SFLs optionally have amino acid sequence identity of at least 70%, optionally at least 85%, optionally at least 90, 95%, 97% or 99%, with a known ligase or known SFL.

[0085] Representative examples of SFLs include CV ligase, DLX, DLXd, DLXd2 and MnM ligase. A preferred SFL is Chlorella Virus ligase. Some exemplary ligases are identified and their GI or accession numbers are provided in Table 1A below:

Table 1A

PRK08224		
Organism	**Protein name**	**Accession**
B.Acidobacteria		
Candidatus Koribacter versatilis Ellin345Candidatus Koribacter (1 proteins)	ATP-Dependent DNA Ligase	YP_592504
Candidatus Solibacter usitatus Ellin6076Candidatus Solibacter (1 proteins)	ATP-Dependent DNA Ligase	YP_826317
C.Actinobacteria		
Acidothermus cellulolyticus 11BAcidothermus (1 proteins)	ATP-Dependent DNA Ligase	YP_873134
Actinosynnema mirum DSM 43827Actinosynnema (1 proteins)	ATP-Dependent DNA Ligase	YP_003099374
Arthrobacter aurescens TC1Arthrobacter (3 proteins)	ATP-Dependent DNA Ligase	YP_949544
Arthrobacter chlorophenolicus A6Arthrobacter (3 proteins)	ATP-Dependent DNA Ligase	YP_002489901
Arthrobacter sp. FB24Arthrobacter (3 proteins)	ATP-Dependent DNA Ligase	YP_833558
Beutenbergia cavernae DSM 12333Beutenbergia (1 proteins)	ATP-Dependent DNA Ligase	YP_002880505
Catenulispora acidiphila DSM 44928Catenulispora (2 proteins)	ATP-Dependent DNA Ligase	YP_003116519

Catenulispora acidiphila DSM 44928Catenulispora (2 proteins)	ATP-Dependent DNA Ligase	YP_003116565
Frankia alni ACN14aFrankia (2 proteins)	ATP-Dependent DNA Ligase	YP_712338
Frankia sp. EAN1pecFrankia (2 proteins)	ATP-Dependent DNA Ligase	YP_001509433
Kineococcus radiotolerans SRS30216Kineococcus (1 proteins)	ATP-Dependent DNA Ligase	YP_001360406
Kytococcus sedentarius DSM 20547Kytococcus (1 proteins)	ATP-Dependent DNA Ligase	YP_003149340
Mycobacterium abscessus ATCC 19977Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_001701033
Mycobacterium avium 104Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_879648
Mycobacterium avium subsp. paratuberculosis K-10Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	NP_959275
Mycobacterium gilvum PYR-GCKMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_001132524
Mycobacterium gilvum PYR-GCKMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_001132543
Mycobacterium marinum MMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_001853525
Mycobacterium smegmatis str. MC2 155Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_890520
Mycobacterium smegmatis str. MC2 155Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_890521
Mycobacterium sp. JLSMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_001073538
Mycobacterium sp. JLSMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_001073546
Mycobacterium sp. JLSMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_001073574
Mycobacterium ulcerans Agy99Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_907815
Mycobacterium vanbaalenii PYR-	ATP-Dependent DNA	YP_956315

1Mycobacterium (26 proteins)	Ligase	
Mycobacterium vanbaalenii PYR-1Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_956321
Mycobacterium tuberculosis H37RaMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_001285120
Mycobacterium tuberculosis H37RvMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	NP_218248
Mycobacterium bovis AF2122/97Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	NP_857396
Mycobacterium bovis BCG str. Pasteur 1173P2Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_979871
Mycobacterium bovis BCG str. Tokyo 172Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_002646832
Mycobacterium tuberculosis CDC1551Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	NP_338389
Mycobacterium tuberculosis F11Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_001289690
Mycobacterium tuberculosis KZN 1435Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_003033779
Mycobacterium sp. KMSMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_940984
Mycobacterium sp. MCSMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_642076
Mycobacterium sp. KMSMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_941006
Mycobacterium sp. MCSMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_642099
Nakamurella multipartita DSM 44233Nakamurella (1 proteins)	ATP-Dependent DNA Ligase	YP_003200226
Nocardia farcinica IFM 10152Nocardia (1 proteins)	ATP-Dependent DNA Ligase	YP_118771
Nocardioides sp. JS614Nocardioides (1 proteins)	ATP-Dependent DNA Ligase	YP_922117
Rhodococcus erythropolis PR4Rhodococcus (3 proteins)	ATP-Dependent DNA Ligase	YP_002768423
Rhodococcus jostii RHA1Rhodococcus (3	ATP-Dependent DNA	YP_705046

proteins)	Ligase	
Rhodococcus opacus B4Rhodococcus (3 proteins)	ATP-Dependent DNA Ligase	YP_002782360
Saccharopolyspora erythraea NRRL 2338Saccharopolyspora (1 proteins)	ATP-Dependent DNA Ligase	YP_001104098
Salinispora arenicola CNS-205Salinispora (2 proteins)	ATP-Dependent DNA Ligase	YP_001536378
Salinispora tropica CNB-440Salinispora (2 proteins)	ATP-Dependent DNA Ligase	YP_001158390
Streptomyces avermitilis MA-4680Streptomyces (3 proteins)	ATP-Dependent DNA Ligase	NP_822873
Streptomyces coelicolor A3(2)Streptomyces (3 proteins)	ATP-Dependent DNA Ligase	NP_630780
Streptomyces griseus subsp. griseus NBRC 13350Streptomyces (3 proteins)	ATP-Dependent DNA Ligase	YP_001822536
F.Chlamydiae/Verrucomicrobia		
Opitutus terrae PB90-1Opitutus (1 proteins)	ATP-Dependent DNA Ligase	YP_001821013
N.Alphaproteobacteria		
Bradyrhizobium japonicum USDA 110Bradyrhizobium (3 proteins)	ATP-Dependent DNA Ligase	NP_771853
Bradyrhizobium sp. BTAi1Bradyrhizobium (3 proteins)	ATP-Dependent DNA Ligase	YP_001236299
Bradyrhizobium sp. ORS278Bradyrhizobium (3 proteins)	ATP-Dependent DNA Ligase	YP_001202307
Chelativorans sp. BNC1Chelativorans (1 proteins)	ATP-Dependent DNA Ligase	YP_675242
Mesorhizobium loti MAFF303099Mesorhizobium (2 proteins)	ATP-Dependent DNA Ligase	NP_108245
Mesorhizobium loti MAFF303099 (plasmid)Mesorhizobium (2 proteins)	ATP-Dependent DNA Ligase	NP_109531
Methylocella silvestris BL2Methylocella (1 proteins)	ATP-Dependent DNA Ligase	YP_002363964
Nitrobacter hamburgensis X14Nitrobacter (1	ATP-Dependent DNA	YP_579055

proteins)	Ligase	
Phenylobacterium zucineum HLK1Phenylobacterium (1 proteins)	ATP-Dependent DNA Ligase	YP_002131547
Rhizobium leguminosarum bv. trifolii WSM2304 (plasmid)Rhizobium (2 proteins)	ATP-Dependent DNA Ligase	YP_002279148
Rhizobium sp. NGR234 (plasmid)Rhizobium (2 proteins)	ATP-Dependent DNA Ligase	YP_002823307
	ATP-Dependent DNA Ligase	
Sinorhizobium medicae WSM419 (plasmid)Sinorhizobium (2 proteins)	ATP-Dependent DNA Ligase	YP_001312861
Sinorhizobium meliloti 1021 (plasmid)Sinorhizobium (2 proteins)	ATP-Dependent DNA Ligase	NP_436551
P.Deltaproteobacteria		
Anaeromyxobacter dehalogenans 2CP-1Anaeromyxobacter (5 proteins)	ATP-Dependent DNA Ligase	YP_002491286
Anaeromyxobacter dehalogenans 2CP-CAnaeromyxobacter (5 proteins)	ATP-Dependent DNA Ligase	YP_464028
Anaeromyxobacter sp. Fw109-5Anaeromyxobacter (5 proteins)	ATP-Dependent DNA Ligase	YP_001378773
Anaeromyxobacter sp. Fw109-5Anaeromyxobacter (5 proteins)	ATP-Dependent DNA Ligase	YP_001381200
Anaeromyxobacter sp. KAnaeromyxobacter (5 proteins)	ATP-Dependent DNA Ligase	YP_002133229
PRK09125		
Organism	**Protein name**	**Accession**
O.Betaproteobacteria		
Acidovorax sp. JS42Acidovorax (1 proteins)	DNA ligase	YP_986978
Aromatoleum aromaticum EbN1Aromatoleum (1 proteins)	DNA ligase	YP_161050
Azoarcus sp. BH72Azoarcus (1 proteins)	DNA ligase	YP_934633
Candidatus Accumulibacter phosphatis clade IIA str. UW-1Candidatus Accumulibacter (1 proteins)	DNA ligase	YP_003169249
Dechloromonas aromatica RCBDechloromonas (1 proteins)	DNA ligase	YP_284461

Diaphorobacter sp. TPSYDiaphorobacter (1 proteins)	DNA ligase	YP_002553689
Herminiimonas arsenicoxydansHerminiimonas (1 proteins)	DNA ligase	YP_001100009
Leptothrix cholodnii SP-6Leptothrix (1 proteins)	DNA ligase	YP_001791742
Methylibium petroleiphilum PM1Methylibium (1 proteins)	DNA ligase	YP_001020556
Neisseria gonorrhoeae FA 1090Neisseria (7 proteins)	DNA ligase	YP_209054
Neisseria gonorrhoeae NCCP11945Neisseria (7 proteins)	DNA ligase	YP_002002827
Neisseria meningitidis 053442Neisseria (7 proteins)	DNA ligase	YP_001598310
Neisseria meningitidis FAM18Neisseria (7 proteins)	DNA ligase	YP_975951
Neisseria meningitidis MC58Neisseria (7 proteins)	DNA ligase	NP_275038
Neisseria meningitidis Z2491Neisseria (7 proteins)	DNA ligase	YP_002341892
Neisseria meningitidis alpha14Neisseria (7 proteins)	DNA ligase	YP_003082363
Polaromonas naphthalenivorans CJ2Polaromonas (2 proteins)	DNA ligase	YP_982249
Polaromonas sp. JS666Polaromonas (2 proteins)	DNA ligase	YP_549233
Rhodoferax ferrireducens T118Rhodoferax (1 proteins)	DNA ligase	YP_522700
Thauera sp. MZ1TThauera (1 proteins)	DNA ligase	YP_002353773
Thiobacillus denitrificans ATCC 25259Thiobacillus (1 proteins)	DNA ligase	YP_314570
Variovorax paradoxus S110Variovorax (1 proteins)	DNA ligase	YP_002944627
Verminephrobacter eiseniae EF01-2Verminephrobacter (1 proteins)	DNA ligase	YP_998235

P.Deltaproteobacteria		
Desulfobacterium autotrophicum HRM2Desulfobacterium (1 proteins)	LigA2	YP_002604477
Myxococcus xanthus DK 1622Myxococcus (1 proteins)	DNA ligase	YP_628883
Q.Epsilonproteobacteria		
Arcobacter butzleri RM4018Arcobacter (1 proteins)	ATP-dependent DNA ligase	YP_001489632
Campylobacter concisus 13826Campylobacter (10 proteins)	ATP-dependent DNA ligase	YP_001467307
Campylobacter curvus 525.92Campylobacter (10 proteins)	ATP-dependent DNA ligase	YP_001407695
Campylobacter fetus subsp. fetus 82-40Campylobacter (10 proteins)	ATP-dependent DNA ligase	YP_892536
Campylobacter hominis ATCC BAA-381Campylobacter (10 proteins)	ATP-dependent DNA ligase	YP_001406356
Campylobacter jejuni RM1221Campylobacter (10 proteins)	ATP-dependent DNA ligase	YP_179811
Campylobacter jejuni subsp. doylei 269.97Campylobacter (10 proteins)	ATP-dependent DNA ligase	YP_001398949
Campylobacter jejuni subsp. jejuni 81-176Campylobacter (10 proteins)	ATP-dependent DNA ligase	YP_001001312
Campylobacter jejuni subsp. jejuni 81116Campylobacter (10 proteins)	ATP-dependent DNA ligase	YP_001483144
Campylobacter jejuni subsp. jejuni NCTC 11168Campylobacter (10 proteins)	ATP-dependent DNA ligase	YP_002345037
Campylobacter lari RM2100Campylobacter (10 proteins)	ATP-dependent DNA ligase	YP_002574655
Sulfurimonas denitrificans DSM 1251Sulfurimonas (1 proteins)	DNA ligase	YP_393098
R.Gammaproteobacteria		
Actinobacillus succinogenes 130ZActinobacillus (1 proteins)	ATP-dependent DNA ligase	YP_001344487
Aggregatibacter actinomycetemcomitans D11S-1Aggregatibacter (2 proteins)	ATP-dependent DNA ligase	YP_003256304
Aggregatibacter aphrophilus	ATP-dependent DNA ligase	YP_003007537

NJ8700Aggregatibacter (2 proteins)		
Alcanivorax borkumensis SK2Alcanivorax (1 proteins)	ATP-dependent DNA ligase	YP_694422
Aliivibrio salmonicida LFI1238Aliivibrio (3 proteins)	ATP-dependent DNA ligase	YP_002262821
Vibrio fischeri ES114Aliivibrio (3 proteins)	ATP-dependent DNA ligase	YP_204833
Vibrio fischeri MJ11Aliivibrio (3 proteins)	ATP-dependent DNA ligase	YP_002156265
Alteromonas macleodii 'Deep ecotype'Alteromonas (1 proteins)	ATP-dependent DNA ligase	YP_002127707
Mannheimia succiniciproducens MBEL55EBasfia (1 proteins)	ATP-dependent DNA ligase	YP_088131
Colwellia psychrerythraea 34HColwellia (1 proteins)	ATP-dependent DNA ligase	YP_271053
Haemophilus influenzae 86-028NPHaemophilus (3 proteins)	ATP-dependent DNA ligase	YP_248841
Haemophilus influenzae PittEEHaemophilus (3 proteins)	ATP-dependent DNA ligase	YP_001290961
Haemophilus influenzae PittGGHaemophilus (3 proteins)	ATP-dependent DNA ligase	YP_001293088
Haemophilus somnus 129PTHistophilus (2 proteins)	ATP-dependent DNA ligase	YP_719536
Haemophilus somnus 2336Histophilus (2 proteins)	ATP-dependent DNA ligase	YP_001783642
Idiomarina loihiensis L2TRIdiomarina (1 proteins)	ATP-dependent DNA ligase	YP_156435
Marinobacter aquaeolei VT8Marinobacter (1 proteins)	ATP-dependent DNA ligase	YP_960951
Marinomonas sp. MWYL1Marinomonas (1 proteins)	ATP-dependent DNA ligase	YP_001341226
Photobacterium profundum SS9Photobacterium (1 proteins)	ATP-dependent DNA ligase	YP_132765
Pseudoalteromonas atlantica T6cPseudoalteromonas (2 proteins)	ATP-dependent DNA ligase	YP_659659
Pseudoalteromonas haloplanktis TAC125Pseudoalteromonas (2 proteins)	ATP-dependent DNA ligase	YP_340675

Psychromonas ingrahamii 37Psychromonas (1 proteins)	ATP-dependent DNA ligase	YP_942593
Shewanella amazonensis SB2BShewanella (18 proteins)	ATP-dependent DNA ligase	YP_927870
Shewanella baltica OS155Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_001050227
Shewanella baltica OS185Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_001366044
Shewanella baltica OS195Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_001554317
Shewanella baltica OS223Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_002358357
Shewanella frigidimarina NCIMB 400Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_750174
Shewanella halifaxensis HAW-EB4Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_001673966
Shewanella loihica PV-4Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_001093713
Shewanella oneidensis MR-1Shewanella (18 proteins)	ATP-dependent DNA ligase	NP_717802
Shewanella pealeana ATCC 700345Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_001502366
Shewanella piezotolerans WP3Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_002312387
Shewanella sediminis HAW-EB3Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_001474374
Shewanella sp. ANA-3Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_870036
Shewanella sp. MR-4Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_734321
Shewanella sp. MR-7Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_738313
Shewanella woodyi ATCC 51908Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_001760369
Shewanella putrefaciens CN-32Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_001183272
Shewanella sp. W3-18-1Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_963655

Thiomicrospira crunogena XCL-2Thiomicrospira (1 proteins)	ATP-dependent DNA ligase	YP_391938
Vibrio cholerae MJ-1236Vibrio (9 proteins)	ATP-dependent DNA ligase	YP_002878565
Vibrio parahaemolyticus RIMD 2210633Vibrio (9 proteins)	DNA ligase	NP_797856
Vibrio splendidus LGP32Vibrio (9 proteins)	DNA ligase	YP_002417130
Vibrio vulnificus CMCP6Vibrio (9 proteins)	DNA ligase	NP_761477
Vibrio vulnificus YJ016Vibrio (9 proteins)	DNA ligase	NP_934427
Vibrio cholerae M66-2Vibrio (9 proteins)	DNA ligase	YP_002810248
Vibrio cholerae O1 biovar El Tor str. N16961Vibrio (9 proteins)	DNA ligase	NP_231182
Vibrio cholerae O395Vibrio (9 proteins)	DNA ligase	YP_001217094
Vibrio cholerae O395Vibrio (9 proteins)	DNA ligase	YP_002819900
PHA0454		
Organism	**Protein name**	**Accession**
b.Viruses		
Enterobacteria phage 13aT7-like viruses (16 proteins)	ATP-dependent DNA ligase	YP_002003941
Enterobacteria phage BA14T7-like viruses (16 proteins)	ATP-dependent DNA ligase	YP_002003458
Enterobacteria phage EcoDS1T7-like viruses (16 proteins)	ATP-dependent DNA ligase	YP_002003747
Enterobacteria phage K1FT7-like viruses (16 proteins)	ATP-dependent DNA ligase	YP_338096
Enterobacteria phage T3T7-like viruses (16 proteins)	ATP-dependent DNA ligase	NP_523305
Enterobacteria phage T7T7-like viruses (16 proteins)	ATP-dependent DNA ligase	NP_041963
Klebsiella phage K11T7-like viruses (16 proteins)	ATP-dependent DNA ligase	YP_002003797
Kluyvera phage Kvp1T7-like viruses (16 proteins)	DNA ligase	YP_002308390
Morganella phage MmP1T7-like viruses (16	ATP-dependent DNA ligase	YP_002048633

proteins)		
Pseudomonas phage gh-1T7-like viruses (16 proteins)	ATP-dependent DNA ligase	NP_813751
Salmonella phage phiSG-JL2T7-like viruses (16 proteins)	ATP-dependent DNA ligase	YP_001949754
Vibriophage VP4T7-like viruses (16 proteins)	ATP-dependent DNA ligase	YP_249578
Yersinia pestis phage phiA1122T7-like viruses (16 proteins)	ATP-dependent DNA ligase	NP_848267
Yersinia phage BerlinT7-like viruses (16 proteins)	ATP-dependent DNA ligase	YP_918989
Yersinia phage Yepe2T7-like viruses (16 proteins)	ATP-dependent DNA ligase	YP_002003318
Yersinia phage phiYeO3-12T7-like viruses (16 proteins)	ATP-dependent DNA ligase	NP_052075
Enterobacteria phage LKA1phiKMV-like viruses (7 proteins)	ATP-dependent DNA ligase	YP_001522868
Enterobacteria phage phiKMVphiKMV-like viruses (7 proteins)	ATP-dependent DNA ligase	NP_877456
Pseudomonas phage LKD16phiKMV-like viruses (7 proteins)	ATP-dependent DNA ligase	YP_001522807
Pseudomonas phage LUZ19phiKMV-like viruses (7 proteins)	ATP-dependent DNA ligase	YP_001671961
Pseudomonas phage PT2phiKMV-like viruses (7 proteins)	ATP-dependent DNA ligase	YP_002117800
Pseudomonas phage PT5phiKMV-like viruses (7 proteins)	ATP-dependent DNA ligase	YP_002117741
Pseudomonas phage phikF77phiKMV-like viruses (7 proteins)	ATP-dependent DNA ligase	YP_002727838
CLSZ2445448		
Organism	**Protein name**	**Accession**
a.Eukaryota		
Paramecium tetraurelia strain d4-2Paramecium (5 proteins)	DNA ligase	XP_001347270
Paramecium tetraurelia strain d4-2Paramecium (5 proteins)	hypothetical protein	XP_001422985

Paramecium tetraurelia strain d4-2Paramecium (5 proteins)	hypothetical protein	XP_001431968
Paramecium tetraurelia strain d4-2Paramecium (5 proteins)	hypothetical protein	XP_001435874
Paramecium tetraurelia strain d4-2Paramecium (5 proteins)	hypothetical protein	XP_001460273
Tetrahymena thermophilaTetrahymena (1 proteins)	ATP dependent DNA ligase	XP_001011861
CLSP2344013		
Organism	**Protein name**	**Accession**
b.Viruses		
Enterobacteria phage Felix 01unclassified Myoviridae (4 proteins)	Putative DNA ligase	NP_944942
Enterobacteria phage WV8unclassified Myoviridae (4 proteins)	hypothetical protein	YP_002922879
Erwinia phage phiEa21-4unclassified Myoviridae (4 proteins)	putative DNA ligase	YP_002456101
Escherichia phage rv5unclassified Myoviridae (4 proteins)	ATP-dependent DNA ligase	YP_002003586
PRK07636		
Organism	**Protein name**	**Accession**
J.Firmicutes		
Bacillus clausii KSM-K16Bacillus	ATP-dependent DNA ligase	YP_176304
Bacillus subtilis subsp. subtilis str. 168Bacillus	ATP-dependent DNA ligase	NP_389932
Bacillus licheniformis ATCC 14580Bacillus	ATP-dependent DNA ligase	YP_078721
Bacillus licheniformis ATCC 14580Bacillus	ATP-dependent DNA ligase	YP_091132
Geobacillus sp. Y412MC10Geobacillus	ATP dependent DNA ligase	YP_003240778
Paenibacillus sp. JDR-2Paenibacillus	ATP dependent DNA ligase	YP_003009892
CLSK2551528		
Organism	**Protein name**	**Accession**
J.Firmicutes		

Geobacillus sp. Y412MC10Geobacillus (1 proteins)	ATP dependent DNA ligase	YP_003245332
Paenibacillus sp. JDR-2Paenibacillus (1 proteins)	ATP dependent DNA ligase	YP_003013681
CLSK2532515		
Organism	**Protein name**	**Accession**
B.Acidobacteria		
Candidatus Solibacter usitatus Ellin6076Candidatus Solibacter (1 proteins)	ATP dependent DNA ligase	YP_829024
E.Bacteroides/Chlorobi		
Flavobacteriaceae bacterium 3519-10unclassified Flavobacteriaceae (1 proteins)	ATP-dependent DNA ligase	YP_003095681
CLSK2470953		
Organism	**Protein name**	**Accession**
C.Actinobacteria		
Arthrobacter chlorophenolicus A6 (plasmid)Arthrobacter (2 proteins)	ATP dependent DNA ligase	YP_002478051
Arthrobacter chlorophenolicus A6 (plasmid)Arthrobacter (2 proteins)	ATP dependent DNA ligase	YP_002478427
Nocardioides sp. JS614Nocardioides (1 proteins)	ATP dependent DNA ligase	YP_923463
CLSK2469924		
Organism	**Protein name**	**Accession**
J.Firmicutes		
Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446Alicyclobacillus (1 proteins)	ATP dependent DNA ligase	YP_003185050
Brevibacillus brevis NBRC 100599Brevibacillus (1 proteins)	putative ATP-dependent DNA ligase	YP_002773127
CLSK2340991		
Organism	**Protein name**	**Accession**
N.Alphaproteobacteria		
Phenylobacterium zucineum HLK1 (plasmid)Phenylobacterium (2 proteins)	ATP dependent DNA ligase	YP_002128561
Phenylobacterium zucineum HLK1	ATP-dependent DNA ligase	YP_002128631

(plasmid)Phenylobacterium (2 proteins)		
CLSK2333706		
Organism	**Protein name**	**Accession**
J.Firmicutes		
Candidatus Desulforudis audaxviator MP104CCandidatus Desulforudis (1 proteins)	ATP dependent DNA ligase	YP_001716762
Natranaerobius thermophilus JW/NM-WN-LFNatranaerobius (1 proteins)	ATP dependent DNA ligase	YP_001916325
CLSK2303611		
Organism	**Protein name**	**Accession**
C.Actinobacteria		
Streptomyces coelicolor A3(2)Streptomyces (2 proteins)	ATP-dependent DNA ligase	NP_631399
Streptomyces griseus subsp. griseus NBRC 13350Streptomyces (2 proteins)	putative ATP-dependent DNA ligase	YP_001828202
CLSK962101		
Organism	**Protein name**	**Accession**
C.Actinobacteria		
Nocardioides sp. JS614Nocardioides (1 proteins)	ATP dependent DNA ligase	YP_922436
Salinispora arenicola CNS-205Salinispora (2 proteins)	DNA polymerase LigD ligase region	YP_001539124
Salinispora tropica CNB-440Salinispora (2 proteins)	ATP dependent DNA ligase	YP_001160776
CLSK915249		
Organism	**Protein name**	**Accession**
C.Actinobacteria See CLSK2303611 above		
Streptomyces avermitilis MA-4680 (plasmid)Streptomyces (2 proteins)	putative ATP-dependint DNA ligase	NP_828839
Streptomyces sp. HK1 (plasmid)Streptomyces (2 proteins)	putative ATP-dependent DNA ligase	YP_001661618
CLSK899085		
Organism	**Protein name**	**Accession**
O.Betaproteobacteria		

Burkholderia cenocepacia HI2424 (plasmid)Burkholderia (3 proteins)	ATP dependent DNA ligase	YP_840498
Burkholderia cenocepacia J2315 (plasmid)Burkholderia (3 proteins)	putative ligase	YP_002235530
Burkholderia multivorans ATCC 17616 (plasmid)Burkholderia (3 proteins)	DNA polymerase LigD ligase subunit	YP_001573706
R.Gammaproteobacteria		
Pseudomonas putida (plasmid)Pseudomonas (1 proteins)	putative ligase	NP_542805
CLSK862724		
Organism	**Protein name**	**Accession**
A.Archaea		
Archaeoglobus fulgidus DSM 4304Archaeoglobus (1 proteins)	DNA ligase, putative	NP_070553
J.Firmicutes		
Desulfotomaculum reducens MI-1Desulfotomaculum (1 proteins)	ATP dependent DNA ligase	YP_001113345
Moorella thermoacetica ATCC 39073Moorella (1 proteins)	ATP dependent DNA ligase, central	YP_430340
Pelotomaculum thermopropionicum SIPelotomaculum (1 proteins)	ATP-dependent DNA ligase	YP_001211793
Thermoanaerobacter pseudethanolicus ATCC 33223Thermoanaerobacter (2 proteins)	ATP dependent DNA ligase	YP_001664477
Thermoanaerobacter sp. X514Thermoanaerobacter (2 proteins)	ATP dependent DNA ligase	YP_001662589
CLSK820690		
Organism	**Protein name**	**Accession**
A.Archaea		
uncultured methanogenic archaeon RC-Ienvironmental samples (1 proteins)	ATP-dependent DNA ligase	YP_686457
C.Actinobacteria		
Mycobacterium avium 104Mycobacterium (2 proteins)	DNA polymerase LigD ligase subunit	YP_882332
Mycobacterium avium subsp. paratuberculosis	hypothetical protein	NP_960263

K-10Mycobacterium (2 proteins)		
Saccharopolyspora erythraea NRRL 2338Saccharopolyspora (1 proteins)	DNA ligase, ATP-dependent	YP_001107793
N.Alphaproteobacteria		
Bradyrhizobium japonicum USDA 110Bradyrhizobium (2 proteins)	DNA ligase	NP_774671
Bradyrhizobium sp. BTAi1Bradyrhizobium (2 proteins)	putative ATP-dependent DNA ligase	YP_001243518
CLSK808255		
Organism	**Protein name**	**Accession**
N.Alphaproteobacteria		
Sinorhizobium medicae WSM419Sinorhizobium (2 proteins)	DNA polymerase LigD ligase region	YP_001326990
Sinorhizobium meliloti 1021 (plasmid)Sinorhizobium (2 proteins)	putative ATP-dependent DNA ligase protein	NP_437750
CLSK806855		
Organism	**Protein name**	**Accession**
N.Alphaproteobacteria		
Agrobacterium tumefaciens str. C58 (plasmid)Agrobacterium (3 proteins)	ATP-dependent DNA ligase	NP_395985
Agrobacterium tumefaciens str. C58 (plasmid)Agrobacterium (3 proteins)	ATP-dependent DNA ligase	NP_396032
Agrobacterium tumefaciens str. C58 (plasmid)Agrobacterium (3 proteins)	ATP-dependent DNA ligase	NP_396609
Rhizobium etli CFN 42 (plasmid)Rhizobium (10 proteins)	putative DNA ligase (ATP) protein	YP_472413
Rhizobium etli CIAT 652Rhizobium (10 proteins)	probable ATP-dependent DNA ligase protein	YP_001977317
Rhizobium etli CIAT 652 (plasmid)Rhizobium (10 proteins)	putative ATP-dependent DNA ligase protein	YP_001985803
Rhizobium leguminosarum bv. trifolii WSM1325 (plasmid)Rhizobium (10 proteins)	DNA polymerase LigD, ligase domain protein	YP_002973496
Rhizobium leguminosarum bv. trifolii WSM1325 (plasmid)Rhizobium (10 proteins)	DNA polymerase LigD, ligase domain protein	YP_002984974

Rhizobium leguminosarum bv. trifolii WSM1325 (plasmid)Rhizobium (10 proteins)	DNA polymerase LigD, ligase domain protein	YP_002984992
Rhizobium leguminosarum bv. trifolii WSM2304Rhizobium (10 proteins)	DNA polymerase LigD, ligase domain protein	YP_002281897
Rhizobium leguminosarum bv. trifolii WSM2304 (plasmid)Rhizobium (10 proteins)	DNA polymerase LigD, ligase domain protein	YP_002278005
Rhizobium leguminosarum bv. viciae 3841 (plasmid)Rhizobium (10 proteins)	putative DNA ligase	YP_764723
Rhizobium leguminosarum bv. viciae 3841 (plasmid)Rhizobium (10 proteins)	putative DNA ligase	YP_771149
Sinorhizobium meliloti (plasmid)Sinorhizobium (2 proteins)	putative ATP-dependent DNA ligase	YP_001965531
Sinorhizobium meliloti 1021 (plasmid)Sinorhizobium (2 proteins)	ATP-dependent DNA ligase	NP_435470
CLSK762775		
Organism	**Protein name**	**Accession**
N.Alphaproteobacteria		
Rhodococcus jostii RHA1 (plasmid)Rhodococcus (2 proteins)	ATP-dependent DNA ligase	YP_708952
Rhodococcus opacus B4 (plasmid)Rhodococcus (2 proteins)	putative ATP-dependent DNA ligase	YP_002776886
CLSK761995		
Organism	**Protein name**	**Accession**
N.Alphaproteobacteria		
Nitrobacter hamburgensis X14Nitrobacter (1 proteins)	ATP dependent DNA ligase	YP_579015
Rhodopseudomonas palustris BisB5Rhodopseudomonas (2 proteins)	ATP dependent DNA ligase	YP_569297
Rhodopseudomonas palustris TIE-1Rhodopseudomonas (2 proteins)	DNA polymerase LigD, ligase domain protein	YP_001991309
CLSK523944		
Organism	**Protein name**	**Accession**
R.Gammaproteobacteria		
Pseudomonas fluorescens (plasmid)Pseudomonas (3 proteins)	putative ATP-dependent DNA ligase	YP_002887417
Pseudomonas putida (plasmid)Pseudomonas (3 proteins)	putative ligase fragment	NP_863069

Pseudomonas sp. ND6 (plasmid)Pseudomonas (3 proteins)	ATP-dependent DNA ligase	NP_943185
CLSK390680		
Organism	**Protein name**	**Accession**
R.Gammaproteobacteria		
Mesorhizobium loti MAFF303099Mesorhizobium (3 proteins)	ATP-dependent DNA ligase	NP_108227
Mesorhizobium loti MAFF303099Mesorhizobium (3 proteins)	hypothetical protein	NP_108282
Mesorhizobium loti MAFF303099 (plasmid)Mesorhizobium (3 proteins)	DNA ligase-like protein	NP_109396

A subset of ligases of interest is in Table 1B below.

Table 1B

PRK08224		
B.Acidobacteria		
Bacteria; Fibrobacteres/Acidobacteria group; Acidobacteria; unclassifed Acidobacteria; Candidatus Koribacter; Candidatus Koribacter versatilis		
Candidatus Solibacter usitatus Ellin6076Candidatus Solibacter (1 proteins)	ATP-Dependent DNA Ligase	YP_826317
C.Actinobacteria		
Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium marinum		
Mycobacterium gilvum PYR-GCKMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_001132524
Mycobacterium vanbaalenii PYR-1Mycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_956315
Mycobacterium sp. MCSMycobacterium (26 proteins)	ATP-Dependent DNA Ligase	YP_642076
F.Chlamydiae/Verrucomicrobia		
Bacteria; Chlamydiae/Verrucomicrobia group; Verrucomicrobia; Opitutae; Opitutales; Opitutaceae; Opitutus; Opitutus terrae		

Opitutus terrae PB90-1Opitutus (1 proteins)	ATP-Dependent DNA Ligase	YP_001821013
PRK09125		
Organism	**Protein name**	**Accession**
O.Betaproteobacteria		
Neisseria meningitidis Z2491Neisseria (7 proteins)	DNA ligase	YP_002341892
Thiobacillus denitrificans ATCC 25259Thiobacillus (1 proteins)	DNA ligase	YP_314570
Variovorax paradoxus S110Variovorax (1 proteins)	DNA ligase	YP_002944627
Verminephrobacter eiseniae EF01-2Verminephrobacter (1 proteins)	DNA ligase	YP_998235
	_	_
P.Deltaproteobacteria		
Desulfobacterium autotrophicum HRM2Desulfobacterium (1 proteins)	LigA2	YP_002604477
Myxococcus xanthus DK 1622Myxococcus (1 proteins)	DNA ligase	YP_628883
	_	_
Q.Epsilonproteobacteria		
Campylobacter jejuni subsp. jejuni NCTC 11168Campylobacter (10 proteins)	ATP-dependent DNA ligase	YP_002345037
Sulfurimonas denitrificans DSM 1251Sulfurimonas (1 proteins)	DNA ligase	YP_393098
	_	_
R.Gammaproteobacteria		
Aggregatibacter aphrophilus NJ8700Aggregatibacter (2 proteins)	ATP-dependent DNA ligase	YP_003007537
Haemophilus influenzae PittEEHaemophilus (3 proteins)	ATP-dependent DNA ligase	YP_001290961
Shewanella baltica OS195Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_001554317
Shewanella loihica PV-4Shewanella (18 proteins)	ATP-dependent DNA ligase	YP_001093713
Vibrio cholerae M66-2Vibrio (9 proteins)	DNA ligase	YP_002810248
	_	_
PHA0454		
Organism	**Protein name**	**Accession**
b.Viruses		
Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae; Autographivirinae; phiKMV-like viruses		
Pseudomonas phage LKD16phiKMV-like viruses (7 proteins)	ATP-dependent DNA ligase	YP_001522807
CLSZ2445448		

Organism	Protein name	Accession
a.Eukaryota		
Eukaryota; Alveolata; Ciliophora; Intramacronucleata; Oligohymenophorea; Peniculida; Parameciidae; Paramecium; Paramecium tetraurelia		
Paramecium tetraurelia strain d4-2Paramecium (5 proteins)	DNA ligase	XP_001347270
PRK07636		
Organism	**Protein name**	**Accession**
J.Firmicutes		
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus clausii		
Bacillus subtilis subsp. subtilis str. 168Bacillus	ATP-dependent DNA ligase	NP_389932
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Geobacillus		
Geobacillus sp. Y412MC10Geobacillus	ATP dependent DNA ligase	YP_003240778
CLSK2551528		
Organism	**Protein name**	**Accession**
J.Firmicutes		
Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Geobacillus		
Geobacillus sp. Y412MC10Geobacillus (1 proteins)	ATP dependent DNA ligase	YP_003245332
	-	-
CLSK2470953		
Organism	**Protein name**	**Accession**
C.Actinobacteria		
Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae; Actinomycetales; Micrococcineae; Micrococcaceae; Arthrobacter; Arthrobacter chlorophenolicus		
Arthrobacter chlorophenolicus A6 (plasmid)Arthrobacter (2 proteins)	ATP dependent DNA ligase	YP_002478427
CLSK2469924		
Organism	**Protein name**	**Accession**
J.Firmicutes		
Bacteria; Firmicutes; Bacilli; Bacillales; Alicyclobacillaceae; Alicyclobacillus; Alicyclobacillus acidocaldarius; Alicyclobacillus acidocaldarius subsp. acidocaldarius		
Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446Alicyclobacillus	ATP dependent DNA ligase	YP_003185050
CLSK2340991		

Organism	Protein name	Accession
N.Alphaproteobacteria		
Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales; Caulobacteraceae; Phenylobacterium; Phenylobacterium zucineum		
Phenylobacterium zucineum HLK1 (plasmid)Phenylobacterium (2 proteins)	ATP-dependent DNA ligase	YP_002128631
	-	-
CLSK2333706		
Organism	**Protein name**	**Accession**
J.Firmicutes		
Bacteria; Firmicutes; Clostridia; Clostridiales; Peptococcaceae; Candidatus Desulforudis; Candidatus Desulforudis audaxviator		
Candidatus Desulforudis audaxviator MP104CCandidatus Desulforudis (1 proteins)	ATP dependent DNA ligase	YP_001716762
	-	-
CLSK962101		
Organism	**Protein name**	**Accession**
C.Actinobacteria		
Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae; Actinomycetales; Micromonosporineae; Micromonosporaceae; Salinispora; Salinispora arenicola	-	-
Salinispora arenicola CNS-205Salinispora (2 proteins)	DNA polymerase LigD ligase region	YP_001539124
Salinispora tropica CNB-440Salinispora (2 proteins)	ATP dependent DNA ligase	YP_001160776
	-	-
CLSK915249		
Organism	**Protein name**	**Accession**
C.Actinobacteria **See CLSK2303611 above**		
Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces; Streptomyces coelicolor		
Streptomyces avermitilis MA-4680 (plasmid)Streptomyces (2 proteins)	putative ATP-dependint DNA ligase	NP_828839
Streptomyces sp. HK1 (plasmid)Streptomyces (2 proteins)	putative ATP-dependent DNA ligase	YP_001661618
CLSK862724		
Organism	**Protein name**	**Accession**
A.Archaea		
Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales; Archaeoglobaceae; Archaeoglobus; Archaeoglobus fulgidus		

Archaeoglobus fulgidus DSM 4304Archaeoglobus (1 proteins)	DNA ligase, putative	NP_070553
J.Firmicutes		
Pelotomaculum thermopropionicum SIPelotomaculum (1 proteins)	ATP-dependent DNA ligase	YP_001211793
Thermoanaerobacter pseudethanolicus ATCC 33223Thermoanaerobacter (2 proteins)	ATP dependent DNA ligase	YP_001664477
CLSK820690		
Organism	**Protein name**	**Accession**
A.Archaea		
Archaea; Euryarchaeota; environmental samples		
uncultured methanogenic archaeon RC-Ienvironmental samples (1 proteins)	ATP-dependent DNA ligase	YP_686457
N.Alphaproteobacteria		
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium; Bradyrhizobium japonicum		
Bradyrhizobium japonicum USDA 110Bradyrhizobium (2 proteins)	DNA ligase	NP_774671
Bradyrhizobium sp. BTAi1Bradyrhizobium (2 proteins)	putative ATP-dependent DNA ligase	YP_001243518
CLSK808255		
Organism	**Protein name**	**Accession**
N.Alphaproteobacteria		
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium; Sinorhizobium medicae		
Sinorhizobium medicae WSM419Sinorhizobium (2 proteins)	DNA polymerase LigD ligase region	YP_001326990
Sinorhizobium meliloti 1021 (plasmid)Sinorhizobium (2 proteins)	putative ATP-dependent DNA ligase protein	NP_437750
CLSK806855		
Organism	**Protein name**	**Accession**
N.Alphaproteobacteria		
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium; Agrobacterium tumefaciens		
Agrobacterium tumefaciens str. C58 (plasmid)Agrobacterium (3 proteins)	ATP-dependent DNA ligase	NP_396032
Rhizobium leguminosarum bv. trifolii WSM1325	DNA polymerase	YP_002973496

(plasmid)Rhizobium (10 proteins)	LigD, ligase domain protein	
Rhizobium leguminosarum bv. trifolii WSM2304 (plasmid)Rhizobium (10 proteins)	DNA polymerase LigD, ligase domain protein	YP_002278005
CLSK390680		
Organism	**Protein name**	**Accession**
N.Alphaproteobacteria		
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Phyllobacteriaceae; Mesorhizobium; Mesorhizobium loti		
Mesorhizobium loti MAFF303099Mesorhizobium (3 proteins)	hypothetical protein	NP_108282

Some exemplary ligases are identified and their GI or accession numbers are provided in Table 1C below:

TABLE 1C

CV DNA Ligase, GenBank ID AAC96909.1, from Paramecium bursaria Chlorella virus 1: MAITKPLLAATLENIEDVQFPCLATPKIDGIRSVKQTQMLSRTFKPIRNSVMNRLL TELLPEGSDGEISIEGATFQDTTSAVMTGHKMYNAKFSYYWFDYVTDDPLKKYI DRVEDMKNYITVHPHILEHAQVKIIPLIPVEINNITELLQYERDVLSKGFEGVMIRK PDGKYKFGRSTLKEGILLKMKQFKDAEATIISMTALFKNTNTKTKDNFGYSKRST HKSGKVEEDVMGSIEVDYDGVVFSIGTGFDADQRRDFWQNKESYIGKMVKFKY FEMGSKDCPRFPVFIGIRHEEDR
MnM DNA Ligase, GenBank ID YP_333052.1, from Burkholderia pseudomallei 1710b (equivalent sequence to ABA50091) MSGVPYGFKPNLAATLTKPELIKFPVWASPKIDGIRCVFFGGVAYSRSLKPIPNPV VQEFAKAYANLLEGLDGELTVGSPTDANCMQNSMAVMSKAAAPDFTFHVFDW FHPAQAHIEFWQRSDVVEDRIVQFYDRYPEVDIRAAPQVLCTSLAHLDTNEARW LADGYEGMMIRDHCGRYKFGRSTEREGGLVKVKRFTDAEAIVIGFEEEMHNANE AKRDATGRTERSTSKAGLHGKGTLGALVVKNERGIVFNIGTGFTAAQRADYWA NHPSLFGKMVKFKHFDHGTVDAPRHPVFIGFRHPEDM
Hin DNA Ligase, GenBank ID P44121, from Haemophilus influenza MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQ LLTRQGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS<u>T</u>ITKSFKGDGWEKLKLYV FDVPDAEGNLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGE

GVVVRNPNAPYERKRSSQILKLKTAR**G**EECTVIAHHKGKGQFENVMGALTCKN HRGEFKIGSGFNLNERENPPPIGSVITYKYRGITNSGKPRFATYWREKK
DLX DNA Ligase, artificial ligase derived from Hin DNA ligase from Haemophilus influenza: MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQ LLTRQGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS**S**ITKSFKGDGWEKLKLYV FDVPDAEGNLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGE GVVVRNPNAPYERKRSSQILKLKTARGEECTVIAHHKGKGQFENVMGALTCKNH RGEFKIGSGFNLNERENPPPIGSVITYKYRGITNSGKPRFATYWREKK
DLXd DNA Ligase, artificial ligase derived from Hin D ligase from Haemophilus influenza: MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQ LLTRQGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS**S**ITKSFKGDGWEKLKLYV FDVPDAEGNLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGE GVVVRNPNAPYERKRSSQILKLKTAR**D**EECTVIAHHKGKGQFENVMGALTCKNH RGEFKIGSGFNLNERENPPPIGSVITYKYRGITNSGKPRFATYWREKK
DLXd2 DNA Ligase (Gammaproteobacteria, Haemophilus influenza) (modified) MLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTRQGQRLSPPAYFIKDFPPFA IDGELFSERNHFEEIS**S**ITKSFKGDGWEKLKLYVFDVPDAEGNLFERLAKLKAHLL EHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGEGVVVRNPNAPYERKRSSQILKL KTAR**D**EECTVIAHHKGKGQFENVMGALTCKNHRGEFKIGSGFNLNERENPPPIGS VITYKYRGITNSGKPRFATYWREKK

[0086] The SFL is in one aspect an enzyme that can mediate the formation of a covalent bond between two polynucleotide termini, e.g., a 3'-OH terminus and a 5'-PO4 terminus are joined together to form a phosphodiester bond. In some instances, DNA ligation entails any one or more of three sequential nucleotidyl transfer steps, discussed below. All three chemical steps depend on a divalent cation cofactor. In one aspect, the SFL is an ATP-dependent ligase or a NAD^{+}-dependent ligase.

[0087] Optionally, the SFL is Chlorella virus DNA ligase (ChVLig). Ho, et al., J Virol, 71(3):1931-19374 (1997) or functional fragment or variant thereof. For example the SFL can comprise any one or more domains characteristic of a ligase, e.g., an N-terminal nucleotidyltransferase (NTase) domain and/or a C-terminal OB domain. The OB domain optionally comprises a five-stranded antiparallel beta-barrel plus an alpha-helix. Within the NTase domain is an adenylate-binding pocket composed of the six peptide motifs that define the covalent NTase enzyme family of polynucleotide ligases. Optionally, the NTase domain can comprise any one or more of the ligase amino acid motifs I, Ia, III, IIIa, IV, V, and VI preferably all six motifs. Motif I (e.g., KxDGxR or a "KXDG" motif) optionally contains a lysine. Exemplary sequences for each motif in CV ligase are ATPKIDGIR (motif I), SRT (motif Ia), EGSDGEIS (motif III), YWFDY (motif IIIa), EGVMIR (motif IV), LLKMK (motif V). Motif 1 preferably contains a lysine residue. Other examples of motif I include CELKLDGLA, VEHKVDGLS, CEPKLDGLA, CELKLDGVA, AEIKYDGVR, CEYKYDGQR, VDYKYDGER, FEIKYDGAR, FEGKWDGYR, AREKIHGTN, ACEKVHGTN, ILTKEDGSL, and VEEKVDGYN. Examples of motif Ia include TRG, SRT, SRR, SRN, SRS, KRT, KRS, SKG and TRG. Examples of motif III include LEVRGEVF, VEVRGECY, LEVRGEVY, LEARGEAF, FMLDGELM, EGSDGEIS, FILDTEAV, FIIEGEIV, AIVEGELV, VVLDGEAV, YQVFGEFA, LVLNGELF, FTANFEFV and LILVGEMA. Examples of motif IIIa include FCYGV, FLYTV, TFYAL, ICHGL, NAYGI, FVYGL, KLYAI, YWFDY, YAFDI, FLFDL, NLFDV, WAFDL, YVFDI, FAFDI, ILLNA, AND FLFDV. Examples of motif IV include DGVVIK, DGIVIK, DGVVVK, DGTVLK, EGLIVK, EGVMIR, EGLMVK, EGVMVK, EGLMAK, EGVIAK, EGYVLK, EGVVIR, EGYVAV, and EGIIMK. Examples of motif V include AVAFK, AIAYK, ALAYK, AIAYK,WWKMK, LLKMK, WLKLK, WIKLK, WLKIK, WVKDK, AIKCK, IIKLR, HFKIK and IVKYV. The SFL optionally comprises all six motifs. Optionally all six motifs are found together in a naturally-occurring ligase, such as a SFL identified herein. Optionally, the SFL is not an RNA-capping enzyme.

[0088] The ligase optionally comprises any functional portion of a SFL. The ligase can be homologous to a SFL or any functional portion thereof, for example more than 75%, 85%, 90%, 95% or 99% homologous at the amino acid level.

[0089] Optionally, any of the ligation methods described herein is performed by an SFL that is not CV ligase or a functional fragment or derivative thereof. Optionally, the SFL is not T4 DNA ligase or functional fragment or variant thereof. Optionally, the SFL is less than 70%, 80% or 90% identical to T4 DNA ligase. In some examples, the SFL is not a ligase that was known by effective filing date of this application to efficiently ligate oligonucleotides shorter than 6 nucleotides in length. In some examples, the SFL is not a ligase that was known by January 11, 2011, to efficiently ligate oligonucleotides shorter than 6 nucleotides in length.

[0090] In a typical assay, a probe that is 2, 3, 4, 5 or 6 nucleotides in length is ligated in template-dependent manner to a primer using an SFL. Optionally, the 3' end of the probe is ligated to the 5' end of the primer, or vice versa. Optionally, the SFL is CV ligase. The efficiency of ligation is for example more than 5%, 10%, 20%, 30%, or 50%. In other instances a probe that is 6, 7, 8, 9, 10, 11 or 12 nucleotides in length is ligated in template-dependent manner to a primer using an SFL. Optionally, the 5' end of the probe is ligated to the 3' end of the primer, or vice versa. Optionally, the SFL is CV ligase. The efficiency of ligation is for example more than 80%, 90% or 95%.

[0091] In any ligation described herein, ligation product is optionally detected. Ligation can be detected by any known method or method described herein. For example the primer and/or template can be immobilized on a support and the probe labeled. The label on the ligated immobilized probe can be detected after non-ligated probe has been washed away. In other embodiments any one or more of the probe, primer or template is labeled. Any one or more of these reagents can optionally be immobilized.

[0092] In any ligation herein, the ligation can be repeated for a desired number of cycles, for example any reagent that has been subjected to a first ligation cycle is used as a reagent in a next ligation cycle. For example the primer, probe or template of a first ligation cycle can be used as primer, probe or template in a next ligation cycle. In some embodiments (e.g., ligation sequencing) the primer of a first cycle is used as primer in the next cycle. In other instances, the primer of a first cycle can be used a probe or template in the next, the probe of a first cycle can be used as primer or template, or the template of a fist cycle can be used as primer or probe the next.

[0093] The next cycle can be designed such that both ligated and unligated reagent of a first cycle can act as reagent in the next cycle. For example, the ligated and unligated primer of a first ligation cycle can be used as primer in the next cycle. Alternatively, the next cycle can be designed such that only a ligation product of the first cycle can act as a ligation reagent in the next cycle, and reagents that remain unligated in the first cycle will not act as reagents in the next cycle. In some examples, the unligated reagents of the previous cycle are "capped" or otherwise rendered unligatable before the next cycle of ligation is executed.

[0094] Any one or more ligases and/or ligation methods encompassed by the invention can optionally be used in one or more ligation assay formats, and/or are performed in the context of one or more specific ligation applications. Non-limiting examples of assay formats include: oligonucleotide ligation assay (OLA), a ligase chain reaction (LCR), a ligase detection reaction (LDR) and combination assays such as the OLA coupled with the polymerase chain reaction (PCR), e.g., OLA-PCR and PCR-OLA, the Combined Chain Reaction (CCR; a combination of PCR and LCR) and PCR-LDR (see, e.g., Landegren et al., Science 241:1077-80, 1988; Barany, Proc. Natl. Acad. Sci. 88:189-93, 1991; Grossman etal., Nucl. Acids Res. 22(21):4527-34, 1994; Bi and Stambrook, Nucl. Acids Res. 25(14):2949-51, 1997; Zirvi et al., Nucl. Acids Res., 27(24):e40, 1999; U.S. Pat. No. 4,988,617; and PCT Publication Nos. WO 97/31256 and WO 01/92579. Non-limiting examples of specific applications include: amplification of template, detection and/or quantification of the presence of a particular nucleic acid, e.g., in a diagnostic sample, ligation sequencing, single nucleotide polymorphism (SNP) analysis, SNP genotyping, mutation detection, identification of single copy genes, detecting microsatellite repeat sequences, and DNA adduct mapping, among other things. See also Whitely et al, U.S. Pat. No. 4,883,750; Letsinger et al, U.S. Pat. No. 5,476,930; Fung et al, U.S. Pat. No. 5,593,826; Kool, U.S. Pat. No. 5,426,180; Landegren et al, U.S. Pat. No. 5,871,921; Xu and Kool, Nucleic Acids Research, 27: 875-881 (1999); Higgins et al, Methods in Enzymology, 68: 50-71 (1979); Engler et al, The Enzymes, 15-3-29 (1982); and Namsaraev, U.S. patent publication 2004/0110213.

[0095] In an embodiment, the method of ligation comprises or consists of a proximity ligation assay (PLA). PLAs typically involve at least three steps. The first step is typically the binding of first and second probes (e.g., antibody probes) to a ligand (e.g., a protein of interest) such that the probes are in close proximity to another. Each of the probes typically contain an oligonucleotide. The oligonucleotides are brought into proximity to one another with the binding of the probes and, in the second step, are then ligated to one another (e.g., the ligation event). The ligated oligonucleotides may then be amplified and detected to determine the presence of the ligand with a sample (e.g., a biological sample).

[0096] An exemplary PLA assay comprises the steps of determining the presence or absence of a target protein in a sample, comprising (a) contacting the target protein with at least a first and second binding agent, each binding agent having binding specificity for the protein and being adjoined to at least one polynucleotide, (b) ligating the oligonucleotides on the first and second binding agent to one another using a ligase to produce a target nucleic acid and amplifying the target nucleic acid; (c) detecting whether amplified target nucleic acid is present or not, and (d) concluding that the target protein is present in the sample if a significant amount of amplified target nucleic acid is detected, and/or concluding that the target protein is absent from the sample if a significant amount of amplified target nucleic acid is not detected. Optionally, step (d) alternatively or additionally comprises measuring or otherwise assessing the amount of amplified target nucleic acid, and taking the amount of amplified nucleic acid as an indicator of the amount of target protein present in the sample.

[0097] In any methods described herein, polynucleotide ligation probes can be used for detection of a target nucleic acid. The ligation probe is optionally capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (i.e., A, G, C, or T) or modified bases (7-deazaguanosine, inosine, etc.). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labeled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labeled such as with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the select sequence or subsequence.

[0098] The oligonucleotide ligation assay (OLA) is a convenient, highly-stringent method that permits distinction among known DNA sequence variants (Landegren, 1988). Multiplex analysis of highly polymorphic loci is useful for identification of individuals, e.g., for paternity testing and in forensic science, organ transplant donor-receiver matching, genetic disease diagnosis, prognosis, and pre-natal counseling, and other genetic-based testing which depend on the discrimination of single-base differences at a multiplicity of loci (Delahunty, 1996). For example, different assays where two PNA-DNA chimeras, a wild-type (WT) sequence chimera and a mutant sequence chimera, bear different fluorescent dyes. Only when the mutant sequence is present in the target sample, will the mutant sequence chimera ligate to the adjacently annealed second probe (oligo) if the mutant base pair is at the ligation site.

a) Sequence determination

[0099] In some embodiments, successive cycles of ligation can be performed, where the ligation product of a previous cycle is used as primer, probe and/or template in a succeeding cycle.

[00100] An exemplary use of such repetitive ligation is ligation sequencing. A template to be sequenced optionally contains a primer-binding region and polynucleotide region of unknown sequence. A primer with a ligatable terminus (e.g., a free 3' OH group or 5' phosphate group) is annealed to the primer binding region. An extension probe is hybridized to the template adjacently to the primer. The proximal nucleotide of the probe forms a complementary base pair with unknown nucleotide in the template. The extension probe is then ligated to the primer, optionally with an SFL, resulting in an extended duplex. Following ligation, the label attached to extension probe is optionally detected. The process is repeated for a desired number of cycles, using the ligation product of one cycle as the primer of the next cycle. Optionally, the extension probe has a non-ligatable distal terminus, and is then cleaved to provide a ligatable extended duplex.

[00101] In one example, the 3' terminus of the primer is ligated to the 5' terminus of the extension probe. The optional step of cleavage can be done for example at a phosphorothiolate linkage using $AgNO_3$ or another salt that provides Ag^+ ions, leaving a phosphate group at the 3' end of the extended duplex. Phosphatase treatment is used to generate an extendable probe terminus on the extended duplex.

[00102] In one-base encoding, the label corresponds to the identity of nucleotide X. Thus nucleotide Y is identified as the nucleotide complementary to nucleotide X. In other encodings, the label does not have a one-to-one correspondence with the identity of any particular nucleotide in the template.

[00103] Also provided are probe families for use in various ligation methods herein, e.g., sequencing. The probe families are optionally characterized in that each probe family comprises a plurality of labeled oligonucleotide probes of different sequence and, at each position in the sequence, a probe family comprises at least 2 probes having different bases at that position. Probes in each probe family comprise the same label. Preferably the probes comprise a scissile internucleoside linkage. The scissile linkage can be located anywhere in the probe. Preferably the probes have a moiety that is not extendable by ligase at one terminus. Preferably the probes are labeled at a position between the scissile linkage and the moiety that is not extendable by ligase, such that cleavage of the scissile linkage following ligation of a probe to an extendable probe

terminus results in an unlabeled portion that is ligated to the extendable probe terminus and a labeled portion that is no longer attached to the unlabeled portion.

[00104] In multiple-base encodings, the probes in each probe family preferably comprise at least j nucleosides X, wherein j is at least 2, and wherein each X is at least 2-fold degenerate among the probes in the probe family. Probes in each probe family further comprise at least k nucleosides N, wherein k is at least 2, and wherein N represents any nucleoside. In general, j + k is equal to or less than 100, typically less than or equal to 30. Nucleosides X can be located anywhere in the probe. Nucleosides X need not be located at contiguous positions. Similarly nucleosides N need not be located at contiguous positions. In other words, nucleosides X and N can be interspersed. Nevertheless, nucleosides X can be considered to have a 5'→3' sequence, with the understanding that the nucleosides need not be contiguous. For example, nucleosides X in a probe of structure $X_ANX_GNNX_CN$ would be considered to have the sequence AGC. Similarly, nucleosides N can be considered to have a sequence.

[00105] Nucleosides X can be identical or different but are not independently selected, i.e., the identity of each X is constrained by the identity of one or more other nucleosides X in the probe. Thus in general only certain combinations of nucleosides X are present in any particular probe and within the probes in any particular probe family. In other words, in each probe, the sequences of nucleosides X can only represent a subset of all possible sequences of length j. Thus the identity of one or more nucleotides in X limits the possible identities for one or more of the other nucleosides.

[00106] Nucleosides N are preferably independently selected and can be A, G, C, or T (or, optionally, a degeneracy-reducing nucleoside). Preferably the sequence of nucleosides N represents all possible sequences of length k, except that one or more N may be a degeneracy-reducing nucleoside. The probes thus contain two portions, of which the portion consisting of nucleosides N is referred to as the unconstrained portion and the portion consisting of nucleosides X is referred to as the constrained portion. As described above, the portions need not be contiguous nucleosides. Probes that contain a constrained portion and an unconstrained portion are referred to herein as partially constrained probes. Preferably one or more nucleosides in the constrained portion is at the proximal end of the probes, i.e., at the end that contains the nucleoside that will be ligated to the extendable probe terminus, which can be either the 5' or 3' end of the oligonucleotide probe in different embodiments of the invention.

[00107] Since the constrained portion of any oligonucleotide probe can only have certain sequences, knowing the identity of one or more of the nucleosides in the constrained portion of a probe, either by itself, or optionally in combination with the identity of the label on the probe, provides information about one or more of the other nucleotides in the constrained portion. The information may or may not be sufficient to precisely identify one or more of the other nucleosides, but it will be sufficient to eliminate one or more possible nucleotide combinations and/or permutations in the constrained portion. Optionally, the information is not sufficient to eliminate any possible identity of any one individual nucleotide in the constrained portion. In certain preferred embodiments, knowing the identity of one nucleoside in the constrained portion of a probe is sufficient to precisely identify each of the other nucleosides in the constrained portion, i.e., to determine the identity and order of the nucleosides that comprise the constrained portion.

[00108] As in the one-base-encoding sequencing methods described above, the most proximal nucleoside in an extension probe that is complementary to the template is ligated to an extendable terminus of an initializing oligonucleotide (in the first cycle of extension, ligation, and detection) and to an extendable terminus of an extended oligonucleotide probe in subsequent cycles of extension, ligation, and detection. Detection of the associated label determines the name of the probe family to which the newly ligated probe belongs. Since each position in the constrained portion of the probe is at least 2-fold degenerate, the name of the probe family does not in itself identify any nucleotide in the constrained portion. However, since the sequence of the constrained portion is one of a subset of all possible sequences of length j, identifying the probe family does eliminate certain possibilities for the sequence of the constrained portion. The constrained portion of the probe constitutes its sequence determining portion. Therefore, eliminating one or more possibilities for the identity of one or more nucleosides in the constrained portion of the probe by identifying the probe family to which it belongs eliminates one or more possibilities for the identity of a nucleotide in the template to which the extension probe hybridizes. In preferred embodiments of the invention the partially constrained probes comprise a scissile linkage between any two nucleosides.

[00109] In certain embodiments the partially constrained probes have the general structure $(X)_j(N)_k$, in which X represents a nucleoside, $(X)_j$ is at least 2-fold degenerate at each position such that X can be any of at least 2 nucleosides having different basepairing specificities, N represents any nucleoside, j is at least 2, k is between 1 and 100, and at least one N or X other than the X at the probe terminus comprises a detectable moiety. Preferably $(N)_k$ is independently 4-fold degenerate at each position so that, in each probe, $(N)_k$ represents all possible sequences of length k, except that one or more positions in $(N)_k$ may be occupied by a degeneracy-reducing nucleotide. Nucleosides in $(X)_j$ can be identical or different but are not independently selected. In other words, in each probe, $(X)_j$ can only represent a subset of all possible sequences of length j. Thus the identity of one or more nucleotides in $(X)_j$ limits the possible identities for one or more of the other nucleosides. The probes thus contain two portions, of which $(N)_k$ is the unconstrained portion and $(X)_j$ is the constrained portion.

[00110] In certain preferred embodiments of the invention the partially constrained probes have the structure 5'-$(X)_j(N)_kN_B$*-3' or 3'-$(X)_j(N)_kN_B$*-5', wherein N represents any nucleoside, N_B represents a moiety that is not extendable by ligase, * represents a detectable moiety, $(X)_j$ is a constrained portion of the probe that is at least 2-fold degenerate at each position, nucleosides in $(X)_j$ can be identical or different but are not independently selected, at least one internucleoside linkage is a scissile linkage, j is at least 2, and k is between 1 and 100, with the proviso that a detectable moiety may be present on any nucleoside N or X other than the X at the probe terminus instead of, or in addition to, N_B. The scissile linkage can be between two nucleosides in $(X)_j$, between the most distal nucleotide in $(X)_j$ and the most proximal nucleoside in $(N)_k$, between nucleosides within $(N)_k$, or between the terminal nucleoside in $(N)_k$ and N_B. Preferably the scissile linkage is a phosphorothiolate linkage.

[00111] A plurality of probe families is referred to as a "collection" of probe families. Probes in each probe family in a collection of probe families are labeled with a label that is distinguishable from labels used to label other probe families in the collection. Each probe family preferably has its own defined set of sequences, which optionally do not overlap with any other probe family. Preferably the combination of sets of defined sequences for probe families in a collection of probe families includes all possible sequences of the length of the sequence-determining portion. Preferably a collection of probe families comprises or consists of 4 distinguishably labeled probe families.

[00112] Preferably the sequence-determining portions of the probes in each probe family are the same length, and preferably the sequence-determining portions of probe families in a collection of probe families are of the same length. Preferably the sequencing-determining portion is a constrained portion that is at least 2 nucleosides in length.

[00113] In some instances, a series of ligation cycles starting from a particular primer allows partial determination of a sequence, i.e., the identification of individual nucleotides spaced apart from one another in a template. Optionally, in order to gather more complete information, a plurality of reactions is performed in which each reaction utilizes a different initializing oligonucleotide i. The initializing oligonucleotides i bind to different portions of the binding region. Preferably the initializing oligonucleotides bind at positions such the extendable termini of the different initializing oligonucleotides are offset by 1 or more nucleotides from each other when hybridized to the template. For example, as shown in Fig. 3, sequencing reactions 1...N are performed. Initializing oligonucleotides i_1... i_n have the same length and bind such that their terminal nucleotides hybridize to successive adjacent positions in the template. Extension probes e_1...e_n thus bind at successive adjacent regions of the template and are ligated to the extendable termini of the initializing oligonucleotides. Terminal nucleotide of probe e_n ligated to i_n is complementary to nucleotide of polynucleotide region, i.e., the first unknown polynucleotide in the template. In the second cycle of extension, ligation, and detection, terminal nucleotide of probe e_{12} is complementary to nucleotide of polynucleotide region, i.e., the second nucleotide of unknown sequence. Likewise, terminal nucleotides of extension probes ligated to duplexes initialized with initializing oligonucleotides i_2, i_3, i_4, and so on, will be complementary to the third, fourth, and fifth nucleotides of unknown sequence. It will be appreciated that the initializing oligonucleotides may bind to regions progressively further away from the polynucleotide region rather than progressively closer to it.

[00114] The spacer function of the non-terminal nucleotides of the extension probes allows the acquisition of sequence information at positions in the template that are considerably removed from the position at which the initializing oligonucleotide binds without requiring a correspondingly large number of cycles to be performed on any given template.

a. *Capping*

[00115] In any one or more repetitive ligation methods herein, it is possible that fewer than all probes with extendable termini participate in a successful ligation reaction in each cycle of extension, ligation, and cleavage. It will be appreciated that if such probes participated in succeeding cycles, the accuracy of each nucleotide identification step could progressively decline. In certain embodiments of the invention a capping step is included to prevent those extendable termini that do not undergo ligation from participating in future cycles. When sequencing in the 5'→3' direction using extension probes containing a 3'-O-P-S-5' phosphorothiolate linkage, capping may be performed by extending the unligated extendable termini with a DNA polymerase and a non-extendable moiety, e.g., a chain-terminating nucleotide such as a dideoxynucleotide or a nucleotide with a blocking moiety attached, e.g., following the ligation or detection step. When sequencing in the 3'→5' direction using extension probes containing a 3'-S-P-O-5' phosphorothiolate linkage, capping may be performed, e.g., by treating the template with a phosphatase, e.g., following ligation or detection. Other capping methods may also be used.

[00116] It is contemplated that any ligation assay or method herein can be highly multiplexed such that a large number of assays (e.g., greater than 1,000) can be performed in parallel, e.g., simultaneously.

[00117] Any multiplexed ligation assay can optionally be conducted on solid substrates where one or more ligation reagents (e.g., the ligase, template, probe and/or primer) can be immobilized on a solid support or surface. Optionally, the primers, probe and/or template can be attached to different portions of a solid substrate in the form of an array. Optionally, the template, primer or probe can be covalently attached to the solid substrate.

[00118] Optionally, one or more ligation reagents are labeled (e.g., template, probe and/or primer) so that ligation products may be discriminated from each other. Alternatively, ligation products can be distinguished based on: (i) size using electrophoresis or chromatography and/or (ii) detectable labels (Grossman, 1994). For example, multiplexed ligation assays can be conducted on a single sample in a single vessel.

[00119] Any of reagents herein (e.g., primer, probe and/or template) can be optionally immobilized on a solid phase. The solid phase optionally comprises a surface to which one or more reactants may be attached electrostatically, hydrophobically, or covalently. Representative solid phases include e.g.: nylon 6; nylon 66; polystyrene; latex beads; magnetic beads; glass beads; polyethylene; polypropylene; polybutylene; butadiene-styrene copolymers; silastic rubber; polyesters; polyamides; cellulose and derivatives; acrylates; methacrylates; polyvinyl; vinyl chloride; polyvinyl chloride; polyvinyl fluoride; copolymers of polystyrene; silica gel; silica wafers glass; agarose; dextrans; liposomes; insoluble protein metals; and, nitrocellulose. Representative solid phases can be formed into appropriate articles such as beads (e.g., microparticles), tubes, strips, disks, filter papers, plates and the like.

[00120] In an embodiment, any one of the following can be attached to a solid support (e.g., a bead): the SFL, one or more polynucleotide reagents (e.g., an initializing probe, an extension probe, a target DNA or template, a 3' terminal or 5' terminal polynucleotide).

[00121] If so desired, one or more labels can be coupled to any reagent of interest, e.g., the SFL or an antibody that binds specifically to the SFL, one or more polynucleotides (e.g., the template, the 5'-terminal polynucleotide, the 3'-terminal polynucleotide, the template, the initializing probe, the extension probe, etc.

Compositions

[00122] Also provided is a polypeptide comprising or consisting essentially of the amino acid sequence of an SFL provided herein, generally provided as GenBank Accession Nos. in the Tables, or a fragment or variant thereof. The fragment or variant thereof preferably can ligate short oligonucleotides at an efficiency comparable to its parent enzyme. Optionally, the variants or fragments of the SFLs are at least 70%, 80, 90%, 95%, 99% identical to the parent SFL. In an example, the SFL is a ligase derived from Hin DNA ligase (e.g., DLX, DLXd or DLXd2) or any fragment or variant thereof that still retains one or more mutant residues shown in Figure 2, and/or has one or more C-terminal amino acids deleted, e.g., 22 C-terminal amino acids deleted. For example, the mutant Hin DNA ligase is at least 70% identical to Hin D ligase sequence provided in Figure 2 or in GenBank Accession No. P44121, which ligase comprises an amino acid mutation at position 193 of the Hin D ligase sequence provided in Figure 2 or in GenBank Accession No. P44121. Optionally the amino acid mutation consists of changing the glycine at position 193 to aspartic acid or glutamic acid.

[00123] Also provided is a nucleic acid encoding the novel SFLs or fragments or variants described herein. Also provided are genes of the novel SFL which comprise sequences that direct expression of the SFL. Also provided are nucleic acid vectors comprising a nucleic acid encoding the novel SFL and a host cell comprising such a vector. Also provided is an antibody that can bind specifically to the novel SFL but not to its naturally-occurring parent.

[00124] Optionally, the novel SFL or nucleic acid or vector or host cell or antibody is purified, isolated or recombinant.

[00125] Also provided is a method of making any one or more ligases described herein comprising: expressing the SFL in a host cell, e.g., by culturing said host cells under conditions such that the ligase is expressed; and optionally recovering or purifiying said ligase.

Kits

[00126] Also provided are kits comprising components for performing ligation reactions according to the disclosure. Kits can contain an SFL or functional variant or fragment thereof. Kits can alternatively or additionally include one or more oligonucleotide probes less than 12 nucleotides in length, (e.g., less than 8, 7, 6, 5, 4, 3 or 2 nucleotides in length), and/or primer oligonucleotides. In some embodiments, a kit can include CV ligase and one or more oligonucleotides probes less than 6 nucleotides in length. One or more probes for example comprises a 5'-phosphate and/or a label. The label is optionally attached to the 5' terminus, or alternatively to the 3' terminus. Optionally, one or more probes are cleavable.

[00127] The kits optionally provide one or more primers, probes and/or templates described herein. The primers and/or probes can have any one or more features or characteristics described herein. For example the probes can comprise a scissile (e.g., phosphorothiolate) linkage. The kits may contain a cleavage reagent suitable for cleaving phosphororothiolate linkages, e.g., $AgNO3$ and appropriate buffers in which to perform the cleavage. The probes can comprise a trigger residue such as a nucleoside containing a damaged base or an abasic residue. The kits may contain a cleavage reagent suitable for cleaving a linkage between a nucleoside and an adjacent abasic residue and/or a reagent suitable for removing a damaged base from a polynucleotide, e.g., a DNA glycosylase. Certain kits contain oligonucleotide probes that comprise a disaccharide nucleotide and contain periodate as a cleavage reagent. In certain embodiments the kits contain a collection of distinguishably labeled oligonucleotide probe families or collections described herein. Optionally the kit does not comprise a polymerase. Optionally the kit does not comprise an enzyme other than a ligase, phosphatase or kinase. Optionally the kit does not comprise an enzyme other than a SFL that has one or more activities mentioned herein. The kits may include ligation reagents (e.g., ligase, buffers, etc.) and instructions for practicing the particular embodiment of the invention. Appropriate buffers for the other enzymes that may be used, e.g., phosphatase, polymerases, may be included. In some cases, these buffers may be identical. Kits may also include a support, e.g. magnetic beads, which are either pre-attached to primers, primers or templates or are derivatised to be capable of attaching to such molecules. The beads may be functionalized with a primer for performing PCR amplification. Other optional components include washing solutions; vectors for inserting templates for PCR amplification; PCR reagents such as amplification primers, thermostable polymerase, nucleotides; reagents for preparing an emulsion; reagents for preparing a gel, etc. Kits can also comprise a plurality of distinguishably labeled probes, which can be labeled such that the identity of the label provides information about the sequence of the probe. Kits can also comprise other or additional compositions or reagents for use in ligation sequencing determination protocol. Optionally, the identity of the label provides the exact identity of a nucleotide occupying one degenerate position in the probe. For example, the identity of the label optionally eliminates one or more combinations or permutations of nucleotides at two or more degenerate positions. The probes are for example degenerate at two or more constrained positions, where the identity of a nucleotide at one constrained position eliminates at least one possible identity of a nucleotide at another constrained position. Optionally, knowing the identity of a constrained residue in that probe further eliminates at least one possible identity for another constrained residue in the probe.

[00128] Unless otherwise apparent from the context, any feature can be claimed in combination with any other, or be claimed as not present in combination with another feature. A feature can be for example any variation, step, feature, property, composition, method, step, degree, level, component, material, substance, element, mode, variable, aspect, measure, amount or embodiment.

[00129] Many features described herein are intended to be optional. If any feature is not explicitly indicated as being necessary, it is to be regarded as optional. Non-limiting examples of language indicating that a feature is optional include terms such as "variation," "where," "while," "when," "optionally," "include," "preferred," "especial," "recommended," "advisable," "particular," "should," "alternative," "typical," "representative," "various," "such as," "the like," "can," "may," "example," "embodiment," or " in an aspect" or "if" or any combination and/or variation of such terms.

[00130] Any indication that a feature is optional is intended to provide adequate support for claims that include closed or exclusive or negative language (e.g., under 35 U.S.C. 112 or Art. 83 and 84 of EPC). Exclusive language specifically excludes the particular recited feature from including any additional subject matter. For example, if it is indicated that A can be drug X, such language is intended to provide support for a claim that explicitly specifies that A consists of X alone, or that A does not include any other drugs besides X. "Negative" language explicitly excludes the optional feature itself from the scope of the claims. For example, if it is indicated that element A can be X, such language is intended to provide support for a claim that explicitly specifies that A is not X.

[00131] Non-limiting examples exclusive or negative terms include "only," "solely," "consisting of," "consisting essentially of," "alone," "without", "in the absence of (e.g., other items of the same type, structure and/or function)" "excluding," "not", "doesn't", "cannot," or any combination and/or variation of such language.

[00132] Similarly, referents such as "a," "an," "said," or "the," are intended to support explicitly single or plural referents where so desired. Non-limiting examples of plural referents include "at least one," "one or more," "more than one," "two or more," "a multiplicity," "a plurality," "any combination of," "any permutation of," "any one or more of," etc.

[00133] All publications and patents cited in this specification are herein incorporated by reference as if each individual publication or patent were specifically and individually indicated to be incorporated by reference. Genbank records referenced by Genbank ID or accession number, particularly any polypeptide sequence, polynucleotide sequences or annotation thereof, are incorporated by reference herein. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

[00134] The following examples are provided for illustrative purposes and are not intended to limit the invention.

EXAMPLES

Example 1

[00135] This example describes ligation of short probes.

[00136] The oligonucleotide reagents were as follows. The short oligos in forward ligation (i.e., ligation of the 3' terminus of the probe to the 5' terminus of the primer) were: 8mer : 5'-CTCATTCG-3'; 6mer : 5'-TCATTCG-3' 5mer : 5'-ATTCG-3'; 4mer : 5'-TTCG-3'; 3mer : 5'-TCG-3'. Short oligos used for reverse ligation (i.e., ligation of the 5' terminus of the probe to the 3' terminus of the primer) were the same as forward ligation oligos except they have a 5' phosphate. All oligonucleotides were obtained from Integrated DNA Technologies. The forward primer was: PO4-5'-CTGCTGTACCGTACATCCGC-3'-6FAM**.** The reverse primer was: 5'-FAM-CTGCCCCGGGTTCCTCATTCTCT-3'

[00137] Ligase Substrate Preparation: MyOne magnetic beads labeled with carboxylic acid, were obtained from Invitrogen. 5' amino labeled, 41-base oligonucleotide (CCA CTA CGC CTC CGC TTT CCT CTC TAT GGG CAG TCG GTG AT) was coupled to MyOne beads using standard amine coupling chemistry as taught in Nakajima et al., Bioconjugate Chem. 1995, 6(1),123-130 (Figure 1b). PCR was used to extend this oligonucleotide using (CTG CCC CGG GTT CCT CAT TCT CTA TTC GCT GCT GTA CCG TAC ATC CGC CTT GGC CGT ACA GCA GAT CAC CGA CTG CCC ATA GAG AGG). The following 105 base ligation substrate was now tethered to MyOne beads (CCA CTA CGC CTC CGC TTT CCT CTC TAT GGG CAG TCG GTG ATC TGC TGT ACG GCC AAG GCG GAT GTA CGG TAC AGC AGC GAA TAG AGA ATG AGG AAC CCG GGG CAG, Figure 1a). The amount of 105-base oligo tethered to the bead was measured by hybridization of a fluorescently labeled "primer" oligo, with 3' fluorescein label (6FAM) for detection by capillary electrophoresis. The amount of fluorescence denatured from the template was compared to a standard curve.

[00138] Ligation reactions were performed in buffer containing 50 mM Tris-HCl, 10 mM MgCL2, 1 mM ATP, 1 mM DTT and 5% w/v PEG 8000 at pH 7.5 (ligase buffer). Primer oligonucleotide was annealed to template in ligase buffer by heating to 85 °C for 10 minutes followed by slow cooling to room temperature. In the case of forward ligation, the primer was 5' phosphorylated and a 3' 6FAM labeled. In the case of reverse ligation, the primer was 5' 6FAM labeled. Ligation reactions were performed in a 96 well plate using 2 nM template/primer, 2.0 μM ligase, 2-5 μM short oligonucleotide in a total volume of 50 μL. The ligation reactions proceeded for 20 minutes at 15 °C. The reactions were stopped using 1% v/v SDS followed by magnetic separation of the templated beads from the solution. Beads were washed three times in 100 μL of 1% SDS prior to washing with 100 uL of 100% formamide. 5 μL of the supernatant containing denatured, FAM labeled primer and ligation product were then analyzed by capillary electrophoresis. The ligation efficiency was calculated as the ratio of peak areas determined by CE, where FAM labeled, unligated and ligated primer were separated, i.e., Efficiency = ligated /(ligated + unligated).

[00139] Shown in Figures 3-5 are a series of ligations of probes of different lengths with a variety of SFLs. In each panel, the ligation reaction is plotted as a function of the probe length. These data clearly demonstrate that small footprint ligases can ligate probe substrates as short as two nucleotides. By utilizing SFLs, the size of a sequencing probe set (e.g., for ligation sequencing in SOLiD™) can be reduced from the current set of 1024 8-mers to as few as 16 (in the case of di-mers), 64 (in the case of trimers), or 256 in the case of tetramers.

[00140] In another experiment, the ligation efficiency of SF Ligase was compared using the same experimental system as above to T4 DNA ligase, using 8-mer ligation sequencing probes that were conjugated or not conjugated to dyes. SFL the highest overall ligation efficiency for both the dark (chase) probe and the labeled sequencing probe. This is true at the 200 nM concentration of probe, which is currently used in SOLiD. Specifically, a mixture of 1024 probes is used in SOLiD, where each individual probe of the 1024 probes is present at a final working concentration of 200 nM. The other concentrations were 2000 nM of enzyme and 1 nM bead template, where ligation was performed at 15°C for 30 minutes. Results are shown in Table 2.

Table 2

Maximum ligation efficiency of 8mer probes by various ligase enzymes

Ligase	**Dark**	**Cy 5**
SFL	0.91 ± 0.01	0.82 ± 0.01
T4	0.67 ± 0.02	0.62 ± 0.02

Example 2

[00141] The fidelity of a small-footprint ligase (CV ligase) was determined to be surprisingly better than T4 DNA ligase in DNA sequencing by ligation on the SOLiD sequencing system. A 35-nucleotide DNA template was covalently attached to beads, such each magnetic bead had about 100,000 covalently bound DNA molecules on average. Approximately 40 million beads were subjected to a ligation sequencing reaction at 20 °C. The sequencing reaction ligated 8-mer oligonucleotide "probes" labeled with fluorescent dye onto a primer, and a total of seven consecutive ligation or probing events were performed on each template molecule. The approximate number of ligation events were therefore 100,000 (i.e., number of template molecules per bead) x 40,000,000 (i.e., number of beads) x 7 (number of ligation events per template) = 2.8 x 10^{13}.

[00142] Results are given in Table 3 below. 0MM= Percentage of all ligation products (which are the result of seven consecutive ligation events) with no mismatches; 3MM = Percentage of all ligation products (which are the result of seven consecutive ligation events) with less than 3 mismatches.

Table 3

Ligase	**0MM**	**3MM**
T4	20%	57%
SF	27%	65%

The sequencing by ligation data demonstrates that 20% of the time the ligation product was perfectly complementary to the template and formed by seven consecutive ligation events that were all correct. An estimate of the ligation fidelity (expressed as the percentage probability that a ligation event is correct, or "p") can be calculated from the equation: 0MM rate = $(p/100)^n$ where p is the percentage ligation fidelity and n is the number of ligation events per final ligation product (here, seven). Thus for T4 DNA ligase, the ligation fidelity is 79.5% and for SFL the ligation fidelity is 83%.

We claim:

1. A method of ligation comprising enzymatically ligating an oligonucleotide ("probe") and a polynucleotide ("primer"), wherein: the probe is N nucleotide residues in length, where N is from 2 to 7.

2. The method of claim 1, comprising enzymatically ligating an oligonucleotide ("probe") and a polynucleotide ("primer"), wherein: the probe is N nucleotide residues in length, where N is 2 or 3.

3. The method of claim 1, comprising enzymatically ligating a polynucleotide ("primer") to an oligonucleotide ("probe"), wherein (i) the probe is 4 or 5 nucleotide residues in length, and (ii) the ligase is not Paramecium Bursaria Chlorella Virus ligase (CV ligase) or a derivative thereof, or the 3' end of the probe is ligated to the 5' end of the primer, and (iii) more than 30% of the probe or primer is ligated.

4. The method of claim 1, comprising enzymatically ligating a polynucleotide ("primer") to an oligonucleotide ("probe"), wherein (i) the probe is 6 or 7 nucleotide residues in length, and (ii) the ligase is not CV ligase or a derivative thereof, and (iii) more than 50% of the probe or primer is ligated.

5. The method of any one of the preceding claims, wherein ligation is performed by a ligase listed in Table 1A, 1B or 1C.

6. The method of any one of the preceding claims, wherein the probe and the primer are not hybridized to the same polynucleotide ("template") if they are different polynucleotides,

7. The method of claim 5, wherein the probe and primer are single-stranded.

8. The method of claim 7, wherein the ligation is performed prior to amplification in a proximity ligation assay.

9. The method of any one of claims 1-4, comprising ligating a proximal terminus of the probe to a proximal terminus of the primer, wherein the probe and the primer are both hybridized to a polynucleotide ("template") such that their proximal termini are adjacently hybridized to each other.

10. A method of claim 1, wherein:

a) the probe and the primer are both hybridized to a polynucleotide ("template") such that their proximal termini are adjacently hybridized to each other, where the proximal terminus of the probe is ligated to the adjacently-hybridized proximal terminus of the primer;
b) the probe is N nucleotide residues in length, where N is 2 or 3.

11. A method of template-dependent ligation comprising enzymatically ligating an oligonucleotide ("probe") and a polynucleotide ("primer"), wherein:

a) the probe and primer are both hybridized to the template such that their proximal termini are adjacent to each other;
b) the probe is of length N, and comprises a proximal portion, wherein the proximal portionis (i) perfectly hybridized to the template and is (ii) L nucleotides long, where the probe's L+1th nucleotide is mismatched with the template; and
c) L is from 2 to 8, and furthermore is less than 6 if the proximal terminus of the probe is its 3' terminus.

12. The method of any one of the preceding claims, wherein ligation is achieved with a small footprint ligase ("SFL") at substantial efficiency, and optionally

(a) at least 10% of the probe or primer is ligated, or

(b) the SFL ligates the proximal termini of the primer and the probe at least 50% as efficiently as the SFL can ligate the proximal termini of the primer and a corresponding octanucleotide whose N proximal residues are identical to the probe, and whose distal 12-N residues are perfectly complementary to the template.

13. The method of any one of the preceding claims, wherein the SFL is CV ligase, or alternatively the SFL is not CV ligase or a derivative thereof.

14. The method of any one of the preceding claims, wherein the primer is not more than 6 nucleotides in length, for example not more 5, 4, 3 or 2 nucleotides.

15. The method of any one of the preceding claims, wherein the template, if present, is not more than 11 nucleotides in length, for example not more 10, 9, 8, 7, 6, 5, 4, 3 or 2 nucleotides.

16. The method of any one of the preceding claims, wherein the proximal terminus of the probe is its 3' terminus and the proximal terminus of the primer is its 5' terminus.

17. The method of any one of claims 1 to 14, wherein the proximal terminus of the probe is its 5' terminus and the proximal terminus of the primer is its 3' terminus.

18. The method of claim 16 or 17, wherein N is 4.

19. The method of claim 16 or 17, wherein N is 3.

20. The method of claim 16 or 17, wherein N is 2.

21. The method of any one of the preceding claims, wherein the probe is labeled.

22. The method of any one of the preceding claims, wherein the primer or template is labeled.

23. The method of any one of the preceding claims, wherein at least two of the probe, primer and template are labeled.The method of any one of the preceding claims, wherein at least one of the probe, the template and/or the primer is immobilized.

24. The method of any one of the preceding claims, wherein the ligation is repeated at least once.

25. The method of claim 22, wherein any ligation product of a previous ligation reaction is used as the primer of a next ligation, where optionally any unligated primer is rendered unligatable before initiating the next ligation.

26. The method of claim 22, further comprising detecting whether the probe has ligated to the primer before repeating the ligation.

27. The method of claim 24, wherein the method provides information about the template sequence.

28. The method of claim 24, wherein ligation is performed in the presence of multiple probes that are at least partially complementary to the same target region on the template adjacent to the proximal terminus of the primer.

29. The method of claim 26, wherein at least two of the multiple probes are labeled distinguishably from each other.

30. The method of claim 27, wherein identifying the label on a ligated probe provides information about the template sequence.

31. The method of claim 22, wherein any ligation product of the previous ligation reaction is used as the template of the next ligation reaction.

32. The method of claim 24, wherein the method comprises a ligase chain reaction.

33. The method of any one of claims 1-16 and 18-32, comprising enzymatically ligating the 5' end of the probe and the 3' end of the primer with CV ligase, wherein N is less than 6.

34. The method of claim 33, wherein N is 2, 3 or 4.

35. A method of template-dependent ligation of an oligonucleotide ("probe") with a polynucleotide ("primer"), where the length of the probe is less than 8 nucleotides.

36. The method of claim 35, wherein the primer is also less than 8 nucleotides, for example not more 6, 5, 4, 3 or 2 nucleotides.

37. The method of claim 35 or 36, wherein the primer is 4 nucleotides long.

38. The method of claim 35 or 36, wherein the primer is 3 nucleotides long.

39. The method of any one of claims 35 to 38, wherein the SFL is CV ligase.

40. A kit comprising a small footprint ligase ("SFL") and an oligonucleotide probe less than N nucleotides in length, where N is from 2 to 7.

41. The kit of claim 40, wherein the SFL is not CV ligase or a derivative thereof and N is less than 6.

42. The kit of any one of claims 40-41, wherein the probe comprises a 5'-phosphate.

43. The kit of any one of claims 40-42, wherein the probe comprises a label.

44. The kit of claim 43, wherein the label is attached to the 5' terminus.

45. The kit of claim 43, wherein the label is attached to the 3' terminus.

46. The kit of any one of claims 40-45, wherein the probe is cleavable.

47. The kit of any one of claims 40-46, wherein the kit comprises a plurality of distinguishably labeled probes.

48. The kit of claim 47, wherein the probes in said plurality are degenerate at one or more nucleotide positions.

49. The kit of claim 48, wherein the probes are labeled such that the identity of the label provides information about a nucleotide occupying a degenerate position in the probe.

50. The kit of claim 49, wherein the identity of the label provides the exact identity of a nucleotide occupying one degenerate position in the probe.

51. The kit of claim 48, wherein the identity of the label eliminates one or more combinations or permutations of nucleotides at two or more degenerate positions.

52. The kit of any one of claims 47 to 51, wherein the probes are degenerate at two or more constrained positions, where the identity of a nucleotide at one constrained position eliminates at least one possible identity of a nucleotide at another constrained position.

53. The kit of claim 52, wherein knowing the identity of a constrained residue in that probe further eliminates at least one possible identity for another constrained residue in the probe.

54. A mutant Hin (Haemophilus Influenzae) DNA ligase comprising a sequence of DLX, DLXd or DLXd2 shown in Table 1C.

55. A mutant Hin DNA ligase consisting essentially of a sequence of DLX, DLXd or DLXd2 shown in Table 1C.

56. A mutant Hin DNA ligase that has at least 70% identity to Hin D ligase sequence provided in Table 1C or in GenBank Accession No. P44121, which ligase comprises an amino acid mutation at position 193 of Hin D ligase sequence provided in Table 1C or in

GenBank Accession No. P44121, optionally wherein said amino acid mutation consists of changing the glycine at position 193 to aspartic acid or glutamic acid.

57. The ligase of claim 56, wherein the ligase has at least 80% identity to Hin D ligase.

58. The ligase of claim 56, wherein the ligase is at least 90% identical to Hin D ligase.

59. The ligase of claim 56, wherein the ligase is at least 95% identical to Hin D ligase.

60. The ligase of claim 56, wherein the ligase is at least 99% identical to Hin D ligase.

61. The ligase of any one of claims 54-60, wherein the ligase is isolated, purified or recombinant.

62. A nucleic acid molecule encoding the ligase of any one of claims 54-61.

63. The nucleic acid of claim 62, wherein the nucleic acid is isolated, purified or recombinant.

64. A vector comprising the nucleic acid molecule of claim 62.

65. The vector of claim 64, wherein the vector is isolated, purified or recombinant.

66. A host cell comprising the vector of claim 64.

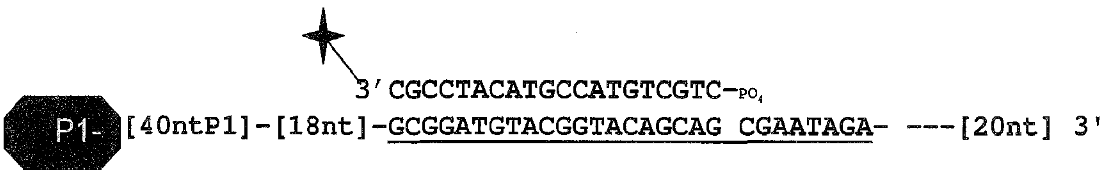
67. The host cell of claim 66, wherein the host cell is isolated, purified or recombinant.

68. A method of making a ligase comprising:

   a. culturing the isolated recombinant host cell of claim 66 or 67 under conditions such that the ligase is expressed; and
   b. recovering said ligase.

FIGURE 1

(A)


3' CGCCTACATGCCATGTCGTC-PO4
P1-
[40ntP1]-[18nt]-GCGGATGTACGGTACAGCAG CGAATAGA- ---[20nt] 3'

(B)


3' AGGAGAGATACCCGPO4
5'
- CCACTACGCCTCCGCTT-TCCTCTCTATGGGC-AGTCGGTGAT-3'

FIGURE 2

CV DNA Ligase, GenBank ID AAC96909.1, from Paramecium bursaria Chlorella virus 1:

MAITKPLLAATLENIEDVQFPCLATPKIDGIRSVKQTQMLSRTFKPIRNSVMNRLLTELL
PEGSDGEISIEGATFQDTTSAVMTGHKMYNAKFSYYWFDYVTDDPLKKYIDRVEDMK
NYITVHPHILEHAQVKIIPLIPVEINNITELLQYERDVLSKGFEGVMIRKPDGKYKFGRST
LKEGILLKMKQFKDAEATIISMTALFKNTNTKTKDNFGYSKRSTHKSGKVEEDVMGSIE
VDYDGVVFSIGTGFDADQRRDFWQNKESYIGKMVKFKYFEMGSKDCPRFPVFIGIRHE
EDR

MnM DNA Ligase, GenBank ID YP_333052.1, from Burkholderia pseudomallei 1710b (equivalent sequence to ABA50091)

MSGVPYGFKPNLAATLTKPELIKFPVWASPKIDGIRCVFFGGVAYSRSLKPIPNPVVQEF
AKAYANLLEGLDGELTVGSPTDANCMQNSMAVMSKAAAPDFTFHVFDWFHPAQAHI
EFWQRSDVVEDRIVQFYDRYPEVDIRAAPQVLCTSLAHLDTNEARWLADGYEGMMIR
DHCGRYKFGRSTEREGGLVKVKRFTDAEAIVIGFEEEMHNANEAKRDATGRTERSTSK
AGLHGKGTLGALVVKNERGIVFNIGTGFTAAQRADYWANHPSLFGKMVKFKHFDHGT
VDAPRHPVFIGFRHPEDM

Hin DNA Ligase, GenBank ID P44121, from Haemophilus influenza

MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTR
QGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS**T**ITKSFKGDGWEKLKLYVFDVPDAE
GNLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGEGVVVRNPNAP
YERKRSSQILKLKTAR**G**EECTVIAHHKGKGQFENVMGALTCKNHRGEFKIGSGFNLNE
RENPPPIGSVITYKYRGITNSGKPRFATYWREKK

DLX DNA Ligase, artificial ligase derived from Hin DNA ligase from Haemophilus influenza:

MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTR
QGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS**S**ITKSFKGDGWEKLKLYVFDVPDAEG
NLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGEGVVVRNPNAPY
ERKRSSQILKLKTARGEECTVIAHHKGKGQFENVMGALTCKNHRGEFKIGSGFNLNER
ENPPPIGSVITYKYRGITNSGKPRFATYWREKK

DLXd DNA Ligase, artificial ligase derived from Hin D ligase from Haemophilus influenza:

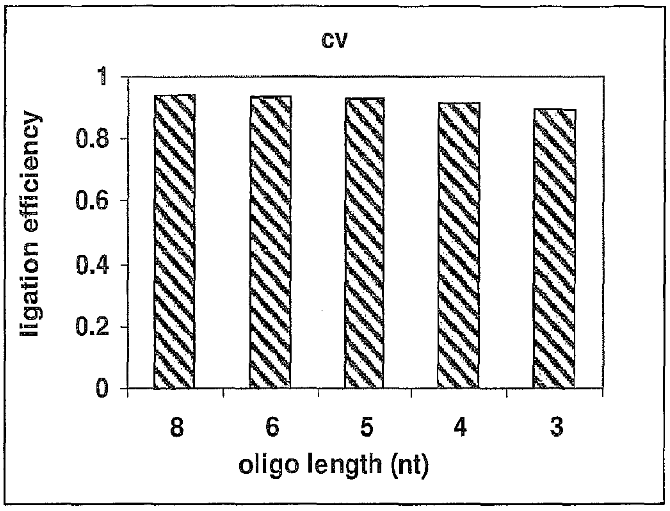
MKFYRTLLLFFASSFAFANSDLMLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTR
QGQRLSPPAYFIKDFPPFAIDGELFSERNHFEEIS**S**ITKSFKGDGWEKLKLYVFDVPDAEG
NLFERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVENLQGEGVVVRNPNAPY
ERKRSSQILKLKTAR**D**EECTVIAHHKGKGQFENVMGALTCKNHRGEFKIGSGFNLNER
ENPPPIGSVITYKYRGITNSGKPRFATYWREKK

DLXd2 DNA Ligase (Gammaproteobacteria, Haemophilus influenza) (modified)

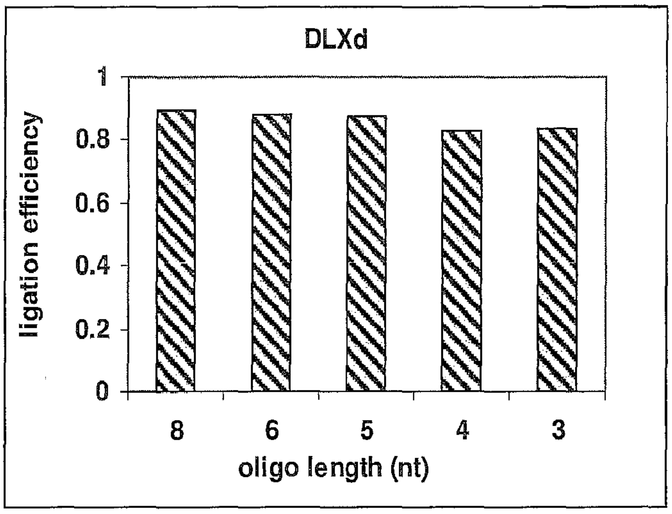
MLLHTYNNQPIEGWVMSEKLDGVRGYWNGKQLLTRQGQRLSPPAYFIKDFPPFAIDGE
LFSERNHFEEIS**S**ITKSFKGDGWEKLKLYVFDVPDAEGNLFERLAKLKAHLLEHPTTYIE
IIEQIPVKDKTHLYQFLAQVENLQGEGVVVRNPNAPYERKRSSQILKLKTAR**D**EECTVIA
HHKGKGQFENVMGALTCKNHRGEFKIGSGFNLNERENPPPIGSVITYKYRGITNSGKPR
FATYWREKK

FIGURE 3

(A)


cv
ligation efficiency
1
0.8
0.6
0.4
0.2
0
8
6
5
4
3
oligo length (nt)

(B)


DLXd
ligation efficiency
1
0.8
0.6
0.4
0.2
0
8
6
5
4
3
oligo length (nt)

(C)

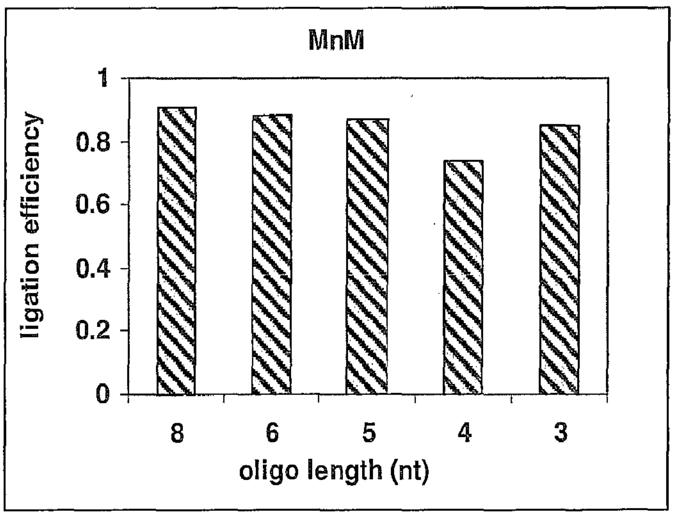

MnM
ligation efficiency
1
0.8
0.6
0.4
0.2
0
8
6
5
4
3
oligo length (nt)

FIGURE 4

(A)

(B)

(C)

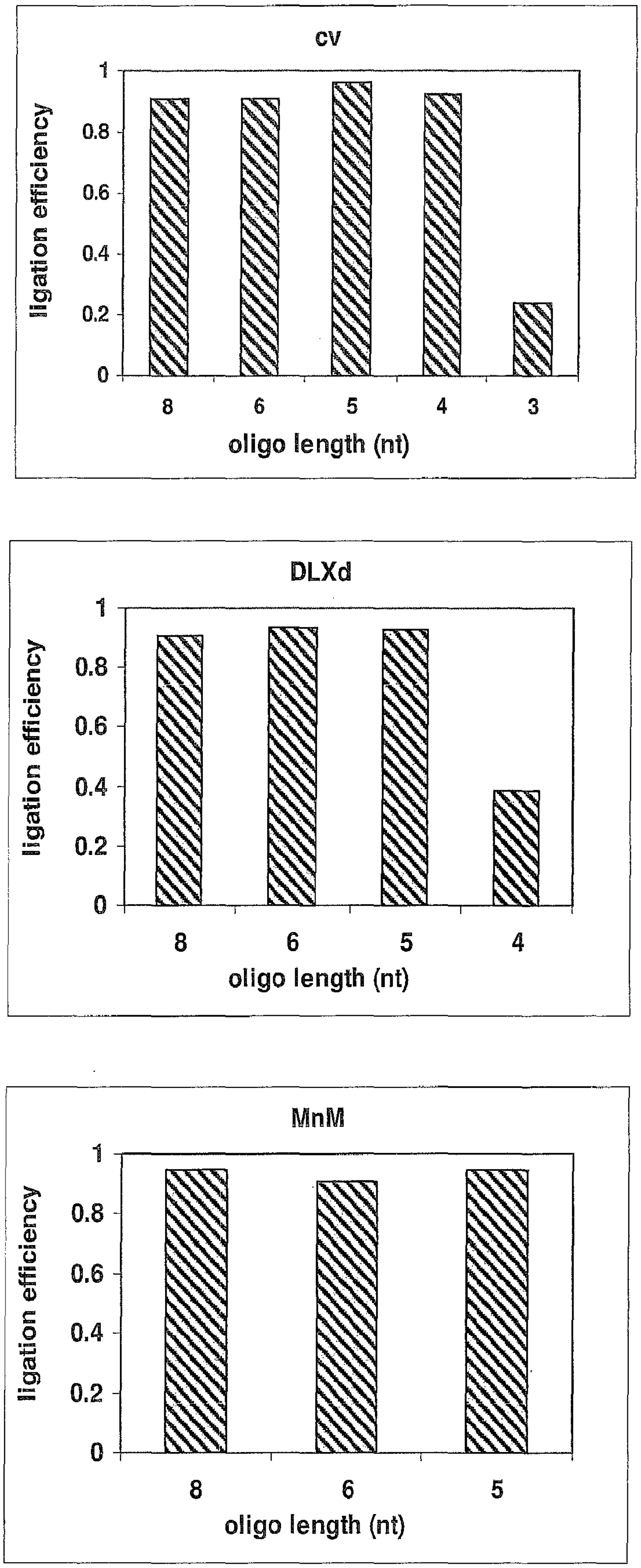

cv
ligation efficiency
1
0.8
0.6
0.4
0.2
0
8
6
5
4
3
oligo length (nt)
DLXd
ligation efficiency
1
0.8
0.6
0.4
0.2
0
8
6
5
4
oligo length (nt)
MnM
ligation efficiency
1
0.8
0.6
0.4
0.2
0
8
6
5
oligo length (nt)

FIGURE 5