

METHODS FOR TAGGING DNA-ENCODED LIBRARIES

Cross-reference to Related Applications

5 This application claims benefit of U.S. Provisional Application Nos. 61/531,820, filed September 7, 2011, and 61/536,929, filed September 20, 2011, each of which is hereby incorporated by reference.

Background of the Invention

10 In general, this invention relates to DNA-encoded libraries of compounds and methods of using and creating such libraries. The invention also relates to compositions for use in such libraries.

 DNA-encoded combinatorial libraries afford many benefits for drug discovery. These libraries can provide a large number of diverse compounds that can be rapidly screened and interrogated. To further increase complexity, various steps of the discovery process can be programmed and automated. These steps include the use of multi-step, split-and-pool synthesis to add building blocks to atomic or polyatomic scaffolds and the use of enzymatic and/or chemical ligation to add DNA tags that encode both
15 the synthetic steps and the building blocks.

 Despite these benefits, numerous issues can arise when very large or complex libraries must be synthesized and deconvoluted. As the size of the library increases, improved methods may be needed to provide high yields of tag ligation. To create libraries under diverse reaction conditions, stable ligated
20 nucleotide constructs would be beneficial, such as constructs that are stable under conditions of high pH and elevated temperature. To simplify deconvolution of tags, the sequence of the tags could be recognized by DNA- or RNA-dependent polymerases, such that tag population demographics can be determined by template-dependent polymerization and sequence determination. Difficulties may arise when creating a library having all of these beneficial attributes. Accordingly, there exists a need for
25 improved, more robust methods of screening and identifying small compounds in DNA-encoded libraries.

Summary of the Invention

 The present invention features methods of creating libraries, where the method includes one or more conditions that improve single-stranded ligation of tags, and compositions for use in creating
30 libraries. Exemplary conditions include the use of one or more 2'-substituted bases within the tags, such as 2'-O-methyl or 2'-fluoro; the use of tags of particular length; the use of one or more enzymes; optionally, the inclusion of error-recognition capabilities in the tag design; and/or the use of one or more agents during ligation.

 Accordingly, the invention features a method of tagging a first library including an
35 oligonucleotide-encoded chemical entity, the method including: (i) providing a headpiece having a first functional group and a second functional group, where the headpiece includes at least one 2'-substituted nucleotide; (ii) binding the first functional group of the headpiece to a first component of the chemical

entity, where the headpiece is directly connected to the first component or the headpiece is indirectly connected to the first component by a bifunctional linker (e.g., a poly ethylene glycol linker or – $(\text{CH}_2\text{CH}_2\text{O})_n\text{CH}_2\text{CH}_2-$, where n is an integer from 1 to 50); and (iii) binding the second functional group of the headpiece to a first building block tag to form a complex, where steps (ii) and (iii) can be performed in any order and where the first building block tag encodes for the binding reaction of step (ii), thereby providing a tagged library.

In some embodiments, the headpiece includes a 2'-substituted nucleotide at one or more of the 5'-terminus, the 3'-terminus, or the internal position of the headpiece. In particular embodiments, the headpiece includes the 2'-substituted nucleotide and the second functional group at the 5'-terminus or at the 3'-terminus.

In other embodiments, the first building block tag includes at least one (e.g., at least two, three, four, five, or more) 2'-substituted nucleotides. In particular embodiments, the first building block tag includes a 2'-substituted nucleotide at one or more of the 5'-terminus, the 3'-terminus, or the internal position of the first building block tag (e.g., a 2'-O-methyl nucleotide or a 2'-fluoro nucleotide at both of the 5'- and 3'-termini). In some embodiments, the first building block tag includes a protecting group at the 3'-terminus or at the 5'-terminus.

In any of the embodiments described herein, the 2'-substituted nucleotide is a 2'-O-methyl nucleotide (e.g., 2'-O-methyl guanine or 2'-O-methyl uracil) or a 2'-fluoro nucleotide (e.g., 2'-fluoro guanine, or 2'-fluoro uracil).

In any of the above embodiments, step (ii) may include joining, binding, or operatively associating the headpiece directly to the first component (e.g., a scaffold or a first building block). In yet other embodiments, step (ii) includes binding the headpiece indirectly to the first component (e.g., a scaffold or a first building block) via a bifunctional linker (e.g., the method includes binding the headpiece with the first functional group of the linker and binding the first component with the second functional group of the linker).

In any of the above embodiments, the method may further include (iv) binding a second building block tag to the 5'-terminus or 3'-terminus of the complex; and (v) binding a second component (e.g., a first building block or a second building block) of the chemical library to the first component, where steps (iv) and (v) can be performed in any order. In some embodiments, the second building block tag encodes for the binding reaction of step (v). In other embodiments, step (iv) may include binding the second building block tag to the 5'-terminus of the complex; the complex includes a phosphate group at the 5'-terminus; and the second building block tag includes a hydroxyl group at both of the 3'- and 5'-termini. In other embodiments, step (iv) may further include purifying the complex and reacting the complex with a polynucleotide kinase to form a phosphate group on the 5'-terminus prior to binding the second building block tag. In other embodiments, step (iv) may include binding the second building block tag to the 3'-terminus of the complex; the complex includes a protecting group at the 3'-terminus; and the second building block tag includes a phosphate group at the 5'-terminus and a protecting group at the 3'-

terminus. In yet other embodiments, step (iv) may further include reacting the complex with a hydrolyzing agent to release the protecting group from the complex prior to binding the second building block tag to the complex.

In further embodiments, the second building block tag includes a 2'-substituted nucleotide (e.g., a 2'-O-methyl nucleotide or a 2'-fluoro nucleotide) at one or more of the 5'-terminus, the 3'-terminus, or the internal position of the second building block tag (e.g., a 2'-O-methyl nucleotide and/or a 2'-fluoro nucleotide at both of the 5'- and 3'-termini).

In some embodiments, step (iv) may include the use of an RNA ligase (e.g., T4 RNA ligase) and/or a DNA ligase (e.g., a ssDNA ligase) to bind the second building block tag to the complex (e.g., may include the use of both RNA ligase and the DNA ligase).

In other embodiments, step (iii) may include the use of an RNA ligase (e.g., T4 RNA ligase) and/or a DNA ligase (e.g., ssDNA ligase) to bind the headpiece to the first building block tag (e.g., may include the use of both RNA ligase and the DNA ligase).

In further embodiments, step (iii) and/or step (iv), if present, may include the use of poly ethylene glycol and/or one or more soluble multivalent cations (e.g., magnesium chloride, manganese (II) chloride, or hexamine cobalt (III) chloride). In some embodiments, the poly ethylene glycol is in an amount from about 25% (w/v) to about 35% (w/v) (e.g., from about 25% (w/v) to about 30 % (w/v), from about 30 % (w/v) to about 35% (w/v), or about 30% (w/v)). In other embodiments, the poly ethylene glycol has an average molecular weight from about 3,000 to about 5,500 Daltons (e.g., about 4,600 Daltons). In other embodiments, the one or more soluble multivalent cations are in an amount of from about 0.05 mM to about 10.5 mM (e.g., from 0.05 mM to 0.5 mM, from 0.05 mM to 0.75 mM, from 0.05 mM to 1.0 mM, from 0.05 mM to 1.5 mM, from 0.05 mM to 2.0 mM, from 0.05 mM to 3.0 mM, from 0.05 mM to 4.0 mM, from 0.05 mM to 5.0 mM, from 0.05 mM to 6.0 mM, from 0.05 mM to 7.0 mM, from 0.05 mM to 8.0 mM, from 0.05 mM to 9.0 mM, from 0.05 mM to 10.0 mM, from 0.1 mM to 0.5 mM, from 0.1 mM to 0.75 mM, from 0.1 mM to 1.0 mM, from 0.1 mM to 1.5 mM, from 0.1 mM to 2.0 mM, from 0.1 mM to 3.0 mM, from 0.1 mM to 4.0 mM, from 0.1 mM to 5.0 mM, from 0.1 mM to 6.0 mM, from 0.1 mM to 7.0 mM, from 0.1 mM to 8.0 mM, from 0.1 mM to 9.0 mM, from 0.1 mM to 10.0 mM, from 0.1 mM to 10.5 mM, from 0.5 mM to 0.75 mM, from 0.5 mM to 1.0 mM, from 0.5 mM to 1.5 mM, from 0.5 mM to 2.0 mM, from 0.5 mM to 3.0 mM, from 0.5 mM to 4.0 mM, from 0.5 mM to 5.0 mM, from 0.5 mM to 6.0 mM, from 0.5 mM to 7.0 mM, from 0.5 mM to 8.0 mM, from 0.5 mM to 9.0 mM, from 0.5 mM to 10.0 mM, from 0.5 mM to 10.5 mM, from 0.75 mM to 1.0 mM, from 0.75 mM to 1.5 mM, from 0.75 mM to 2.0 mM, from 0.75 mM to 3.0 mM, from 0.75 mM to 4.0 mM, from 0.75 mM to 5.0 mM, from 0.75 mM to 6.0 mM, from 0.75 mM to 7.0 mM, from 0.75 mM to 8.0 mM, from 0.75 mM to 9.0 mM, from 0.75 mM to 10.0 mM, from 0.75 mM to 10.5 mM, from 1.0 mM to 1.5 mM, from 1.0 mM to 2.0 mM, from 1.0 mM to 3.0 mM, from 1.0 mM to 4.0 mM, from 1.0 mM to 5.0 mM, from 1.0 mM to 6.0 mM, from 1.0 mM to 7.0 mM, from 1.0 mM to 8.0 mM, from 1.0 mM to 9.0 mM, from 1.0 mM to 10.0 mM, from 1.0 mM to 10.5 mM, from 1.5 mM to 2.0 mM, from 1.5 mM to 3.0 mM, from 1.5 mM to 4.0 mM,

from 1.5 mM to 5.0 mM, from 1.5 mM to 6.0 mM, from 1.5 mM to 7.0 mM, from 1.5 mM to 8.0 mM, from 1.5 mM to 9.0 mM, from 1.5 mM to 10.0 mM, from 1.5 mM to 10.5 mM, from 2.0 mM to 3.0 mM, from 2.0 mM to 4.0 mM, from 2.0 mM to 5.0 mM, from 2.0 mM to 6.0 mM, from 2.0 mM to 7.0 mM, from 2.0 mM to 8.0 mM, from 2.0 mM to 9.0 mM, from 2.0 mM to 10.0 mM, and from 2.0 mM to 10.5 mM). In some embodiments, one or more multivalent cations are in an amount of about 1 mM (e.g., from 0.5 mM to 1.5 mM). In a particular embodiment, the multivalent cation is in the form of hexamine cobalt (III) chloride.

In other embodiments, the method further includes separating the complex from any unreacted tag or unreacted headpiece before any one of binding steps (ii)-(v). In other embodiments, the method further includes purifying the complex before any one of binding steps (ii)-(v). In other embodiments, the method further includes binding one or more additional components (e.g., a scaffold or a first building block) and one or more additional building block tags, in any order and after any one of binding step (ii)-(v).

The invention also features a method of tagging a first library including an oligonucleotide-encoded chemical entity, the method including: (i) providing a headpiece having a first functional group and a second functional group, where the headpiece includes a 2'-substituted nucleotide at the 5'-terminus, optionally one or more nucleotides at the internal position of the headpiece, and a protecting group at the 2'-position and/or the 3'-position at the 3'-terminus; (ii) binding the first functional group of the headpiece to a first component of the chemical entity, where the headpiece is directly connected to the first component or the headpiece is indirectly connected to the first component by a bifunctional linker; and (iii) binding the second functional group of the headpiece to a first building block tag, where the first building block tag includes a 2'-substituted nucleotide and a hydroxyl group at the 5'-terminus, optionally one or more nucleotides at the internal position of the tag, and a 2'-substituted nucleotide and a hydroxyl group at the 3'-terminus; where steps (ii) and (iii) can be performed in any order and where the first building block tag encodes for the binding reaction of step (ii), thereby providing a tagged library.

In some embodiments, the 2'-substituted nucleotide is a 2'-O-methyl nucleotide (e.g., 2'-O-methyl guanine) or a 2'-fluoro nucleotide (e.g., 2'-fluoro guanine). In other embodiments, one or more nucleotides at the internal position of the headpiece are 2'-deoxynucleotides. In yet other embodiments, the bifunctional linker is a poly ethylene glycol linker (e.g., $-(CH_2CH_2O)_nCH_2CH_2-$, where n is an integer from 1 to 50).

In other embodiments, one or more nucleotides (e.g., one or more 2'-deoxynucleotides) are present at the internal position of the headpiece or the tag.

In some embodiments, step (iii) may include the use of one or more soluble multivalent cations (e.g., magnesium chloride, manganese (II) chloride, or hexamine cobalt (III) chloride), poly ethylene glycol (e.g., having an average molecular weight of about 4,600 Daltons), and RNA ligase (e.g., T4 RNA ligase).

In another aspect, the invention features methods to identify and/or discover a chemical entity, the method including tagging a first library including an oligonucleotide-encoded chemical entity (e.g., including steps (i) to (iii) and optionally including steps (iv) to (v)) and selecting for a particular characteristic or function (e.g., selecting for binding to a protein target including exposing the
5 oligonucleotide-encoded chemical entity or chemical entity to the protein target and selecting the one or more oligonucleotide-encoded chemical entities or chemical entities that bind to the protein target (e.g., by using size exclusion chromatography)). The invention also features a complex including a headpiece and a building block tag, where the tag includes from 5 to 20 nucleotides, a 2'-substituted nucleotide at the 5'-terminus, and a 2'-substituted nucleotide at the 3'-terminus. In some embodiments, the 2'-
10 substituted nucleotide at the 5'-terminus and/or 3'-terminus is a 2'-O-methyl nucleotide (e.g., 2'-O-methyl guanine or 2'-O-methyl uracil) or a 2'-fluoro nucleotide (e.g., 2'-fluoro guanine or 2'-fluoro uracil). In particular embodiments, the headpiece includes a hairpin structure. In some embodiments, the headpiece includes a 2'-substituted nucleotide at one or more of the 5'-terminus, the 3'-terminus, or the internal position of the headpiece. In other embodiments, the headpiece further includes a preadenylated
15 5'-terminus. In yet other embodiments, the headpiece includes from 5 to 20 nucleotides.

In any of the above embodiments, the headpiece, the first building block tag, the second building block tag, or the one or more additional building block tags, if present, includes a preadenylated 5'-terminus.

In any of the above embodiments, the method further includes binding one or more (e.g., one,
20 two, three, four, five, six, seven, eight, nine, or ten) additional building block tags to the complex and binding one or more (e.g., one, two, three, four, five, six, seven, eight, nine, or ten) additional components (e.g., scaffolds or building blocks) to the complex, where the one or more additional building block tag encodes for the one or more additional components or encodes for the binding reaction of one or more additional components, thereby providing a tagged library.

In any of the above embodiments, the 2'-substituted nucleotide is a 2'-O-methyl nucleotide, such
25 as 2'-O-methyl guanine, 2'-O-methyl uracil, 2'-O-methyl adenosine, 2'-O-methyl thymidine, 2'-O-methyl inosine, 2'-O-methyl cytidine, or 2'-O-methyl diamino purine. Alternatively, in any of the above embodiments, the 2'-substituted nucleotide is a 2'-fluoro nucleotide, such as 2'-fluoro guanine, 2'-fluoro uracil, 2'-fluoro adenosine, 2'-fluoro thymidine, 2'-fluoro inosine, 2'-fluoro cytidine, or 2'-fluoro
30 diamino purine.

In any of the above embodiments, the RNA ligase is T4 RNA ligase and/or the DNA ligase is a ssDNA ligase.

In any of the above embodiments, the method includes a plurality of headpieces. In some
embodiments of this method, each headpiece of the plurality of headpieces includes an identical sequence
35 region and a different encoding region. In particular embodiments, the identical sequence region is a primer binding region. In other embodiments, the different encoding region is an initial building block tag that encodes for the headpiece or for an addition of an initial component.

In any of the above embodiments, binding in at least one of steps (ii)-(iv), if present, includes enzyme ligation and/or chemical ligation. In some embodiments, enzymatic ligation includes use of an RNA ligase (e.g., T4 RNA ligase) or a DNA ligase (e.g., ssDNA ligase). In other embodiments, enzymatic ligation includes use of an RNA ligase (e.g., T4 RNA ligase) and a DNA ligase (e.g., ssDNA ligase). In some embodiments, chemical ligation includes use of one or more chemically co-reactive pairs (e.g., a pair including an optionally substituted alkynyl group with an optionally substituted azido group; a pair including an optionally substituted diene having a 4π electron system (e.g., an optionally substituted 1,3-unsaturated compound, such as optionally substituted 1,3-butadiene, 1-methoxy-3-trimethylsilyloxy-1,3-butadiene, cyclopentadiene, cyclohexadiene, or furan) with an optionally substituted dienophile or an optionally substituted heterodienophile having a 2π electron system (e.g., an optionally substituted alkenyl group or an optionally substituted alkynyl group); a pair including a nucleophile (e.g., an optionally substituted amine or an optionally substituted thiol) with a strained heterocyclyl electrophile (e.g., optionally substituted epoxide, aziridine, aziridinium ion, or episulfonium ion); a pair including a phosphorothioate group with an iodo group (e.g., a phosphorothioate group at the 3'-terminus and an iodo group at the 5'-terminus); or a pair including an aldehyde group with an amino group (e.g., a primary amino or a secondary amino group, including a hydrazido group)). In particular embodiments, the chemically co-reactive pair produces a resultant spacer having a length from about 4 to about 24 atoms (e.g., from about 4 to about 10 atoms). In other embodiments, chemical ligation includes use of a phosphorothioate group (e.g., at the 3'-terminus) and an iodo group (e.g., at the 5'-terminus). In further embodiments, chemical ligation includes a splint oligonucleotide in the binding reaction. In some embodiments, the chemical ligation includes use of a phosphorothioate group (e.g., at the 3'-terminus of the headpiece, the first building block tag, the second building block tag, the one or more additional building block tags, the library-identifying tag, the use tag, and/or the origin tag, if present), an iodo group (e.g., at the 5'-terminus of the headpiece, the first building block tag, the second building block tag, the one or more additional building block tags, the library-identifying tag, the use tag, and/or the origin tag, if present), and a splint oligonucleotide in the binding reaction, where the use avoids use of one or more protecting groups. In other embodiments, chemical ligation of multiple tags comprises alternating use of orthogonal chemically co-reactive pairs (e.g., any two or more chemically co-reactive pairs described herein) for ligating successive tags.

In any of the above embodiments, the headpiece may include a single-stranded (e.g., hairpin) structure.

In any of the above embodiments, the headpiece, the first building block tag, the second building block tag, the one or more additional building block tags, the library-identifying tag, the use tag, and/or the origin tag, if present, includes a sequence that is substantially identical (e.g., at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical) to any sequence described herein (e.g., the sequence in any one of SEQ ID NOs: 6-21, 26, 27, or 29-31), or a sequence that is complementary to a sequence that is substantially identical (e.g., at least 50%, 60%, 70%, 75%, 80%,

85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical) to any sequence described herein (e.g., the sequence in any one of SEQ ID NOs: 6-21, 26, 27, or 29-31). In particular embodiments, the first building block tag, the second building block tag, the one or more additional building block tags, the library-identifying tag, the use tag, and/or the origin tag, if present, further includes a sequence that is substantially identical (e.g., at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical) to the sequence of SEQ ID NO: 1 or SEQ ID NO: 2.

In any of the above embodiments, the methods or complexes include only single-stranded molecules, where the headpiece, the first building block tag, the second building block tag, and/or the one or more additional building block tags are single-stranded. In some embodiments, one or more of the single-stranded molecules have a hairpin structure. In particular embodiments, the headpiece includes a hairpin structure and the one or more building block tags do not include a hairpin structure.

In any of the above embodiments, the method further comprises one or more optional steps to diversify the library or to interrogate the members of the library, as described herein. In some embodiments, the method further comprises identifying a small drug-like library member that binds or inactivates a protein of therapeutic interest. In other embodiments, the method further comprises contacting a member of the library with a biological target under conditions suitable for at least one member of the library to bind to the target, removing one or more library members that do not bind to the target, and analyzing the one or more oligonucleotide tags associated with them.

As described herein, the use of single-stranded molecules (e.g., including hairpin molecules) could have numerous benefits. Accordingly, in any of the embodiments described herein, the methods and complexes include a headpiece, one or more building block tags, a complex, a chemical entity, a molecule, or any member of a tagged library having decreased mass, increased solubility (e.g., in an organic solvent), decreased cost, increased reactivity, increased target accessibility, decreased hydrodynamic radius, and/or increased accuracy of analytical assessments, as compared to a method including one or more double-stranded molecules (e.g., a double-stranded headpiece or a double-stranded building block tag). In some embodiments, each of the building block tags (e.g., the first building block tag, the second building block tag, and/or one or more additional building block tags, if present) has about the same mass (e.g., each building block tag has a mass that is about +/- 10% from the average mass between two or more building block tags). In particular embodiments, the building block tag has a decreased mass (e.g., less than about 15,000 Daltons, about 14,000 Daltons, about 13,000 Daltons, about 12,000 Daltons, about 11,000 Daltons, about 10,000 Daltons, about 9,000 Daltons, about 8,000 Daltons, about 7,500 Daltons, about 7,000 Daltons, about 6,000 Daltons, about 6,500 Daltons, about 5,000 Daltons, about 5,500 Daltons, about 4,000 Daltons, about 4,500 Daltons, or about 3,000 Daltons) compared to a double-stranded tag (e.g., a double-stranded tag having a mass of about 15,000 Daltons, about 14,000 Daltons, about 13,000 Daltons, or about 12,000 Daltons). In other embodiments, the building block tag has a reduced length compared to a double-stranded tag (e.g., a double-stranded tag having a length of less than about 20 nucleotides, less than about 19 nucleotides, less than about 18

nucleotides, less than about 17 nucleotides, less than about 16 nucleotides, less than about 15 nucleotides, less than about 14 nucleotides, less than about 13 nucleotides, less than about 12 nucleotides, less than about 11 nucleotides, less than about 10 nucleotides, less than about 9 nucleotides, less than about 8 nucleotides, or less than about 7 nucleotides). In some embodiments, one or more building block tags or members of the library lack a primer binding region and/or a constant region (e.g., during a selection step, such as selection using size exclusion chromatography). In some embodiments, one or more building block tags or members of the library have a reduced constant region (e.g., a length less than about 30 nucleotides, less than about 25 nucleotides, less than about 20 nucleotides, less than about 19 nucleotides, less than about 18 nucleotides, less than about 17 nucleotides, less than about 16 nucleotides, less than about 15 nucleotides, less than about 14 nucleotides, less than about 13 nucleotides, less than about 12 nucleotides, less than about 11 nucleotides, less than about 10 nucleotides, less than about 9 nucleotides, less than about 8 nucleotides, or less than about 7 nucleotides). In other embodiments, the methods include a headpiece that encodes for a molecule, a portion of a chemical entity, a binding reaction (e.g., chemical or enzymatic ligation) of a step, or the identity of a library, where the encoding headpiece eliminates the need of an additional building block tag to encode such information.

In any of the above embodiments, an oligonucleotide (e.g., the headpiece, the first building block tag, the second building block tag, and/or one or more additional building block tags, if present) encodes for the identity of the library. In some embodiments, the oligonucleotide (e.g., the headpiece, the first building block tag, the second building block tag, and/or one or more additional building block tags, if present) includes a first library-identifying sequence, where the sequence encodes for the identity of the first library. In particular embodiments, the oligonucleotide is a first library-identifying tag. In some embodiments, the method includes providing a first library-identifying tag, where the tag includes a sequence that encodes for a first library, and/or binding the first library-identifying tag to the complex. In some embodiments, the method includes providing a second library and combining the first library with a second library. In further embodiments, the method includes providing a second library-identifying tag, where the tag includes a sequence that encodes for a second library.

In any of the above embodiments, an oligonucleotide (e.g., a headpiece and/or one or more building blocks) encodes for the use of the member of the library (e.g., use in a selection step or a binding step, as described herein). In some embodiments, the oligonucleotide (e.g., the headpiece, the first building block tag, the second building block tag, and/or one or more additional building block tags, if present) includes a use sequence, where the sequence encodes for use of a subset of members in the library in one or more steps (e.g., a selection step and/or a binding step). In particular embodiments, the oligonucleotide is a use tag including a use sequence. In some embodiments, an oligonucleotide (e.g., a headpiece and/or one or more building blocks) encodes for the origin of the member of the library (e.g., in a particular part of the library). In some embodiments, the oligonucleotide (e.g., the headpiece, the first building block tag, the second building block tag, and/or one or more additional building block tags, if present) includes an origin sequence (e.g., a random degenerate sequence having a length of about 10,

9, 8, 7, or 6 nucleotides), where the sequence encodes for the origin of the member in the library. In particular embodiments, the oligonucleotide is an origin tag including an origin sequence. In some embodiments, the method further includes joining, binding, or operatively associating a use tag and/or an origin tag to the complex.

5 In any of the above embodiments, the methods, compositions, and complexes optionally include a tailpiece, where the tailpiece includes one or more of a library-identifying sequence, a use sequence, or an origin sequence, as described herein. In particular embodiments, the methods further include joining, binding, or operatively associating the tailpiece (e.g., including one or more of a library-identifying sequence, a use sequence, or an origin sequence) to the complex.

10 In any of the above embodiments, the methods, compositions, and complexes, or portions thereof (e.g., the headpiece, the first building block tag, the second building block tag, and/or the one or more additional building block tags, if present), includes a modified phosphate group (e.g., a phosphorothioate or a 5'-N-phosphoramidite linkage) between the terminal nucleotide at the 3'-terminus and the nucleotide adjacent to the terminal nucleotide. In particular embodiments, the modified phosphate group minimizes
15 shuffling during enzymatic ligation between two oligonucleotides (e.g., minimizes inclusion of an additional nucleotide or excision of a nucleotide in the final product or complex, as compared to the sequences of two oligonucleotides to be ligated, such as between a headpiece to a building block tag or between a first building block tag and a second building block tag), as compared to ligation between two oligonucleotides (e.g., a headpiece and a building block tag or a first building block tag and a second
20 building block tag) lacking the modified phosphate group. In some embodiments, the complex may include a phosphorothioate or a triazole group.

In any of the above embodiments, the methods, compositions, and complexes, or portions thereof (e.g., the headpiece, the first building block tag, the second building block tag, and/or the one or more additional building block tags, if present), includes a modification that supports solubility in semi-,
25 reduced-, or non-aqueous (e.g., organic) conditions. In some embodiments, the bifunctional linker, headpiece, or one or more building block tags is modified to increase solubility of a member of said DNA-encoded chemical library in organic conditions. In some embodiments, the modification is one or more of an alkyl chain, a polyethylene glycol unit, a branched species with positive charges, or a hydrophobic ring structure. In some embodiments, the modification includes one or more modified
30 nucleotides having a hydrophobic moiety (e.g., modified at the C5 positions of T or C bases with aliphatic chains, such as in 5'-dimethoxytrityl-N4-diisobutylaminomethylidene-5-(1-propynyl)-2'-deoxycytidine, 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite; 5'-dimethoxytrityl-5-(1-propynyl)-2'-deoxyuridine, 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite; 5'-dimethoxytrityl-5-fluoro-2'-deoxyuridine, 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite; and 5'-dimethoxytrityl-
35 5-(pyren-1-yl-ethynyl)-2'-deoxyuridine, or 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite) or an insertion having a hydrophobic moiety (e.g., an azobenzene). In some embodiments, the member of the library has an octanol:water coefficient from about 1.0 to about 2.5 (e.g., about 1.0 to about 1.5, about 1.0

to about 2.0, about 1.3 to about 1.5, about 1.3 to about 2.0, about 1.3 to about 2.5, about 1.5 to about 2.0, about 1.5 to about 2.5, or about 2.0 to about 2.5).

In any of the above embodiments, the headpiece, the tailpiece, the first building block tag, the second building block tag, the one or more additional building block tags, the library-identifying tag, the use tag, and/or the origin tag, if present, may include from 5 to 20 nucleotides (e.g., from 5 to 7 nucleotides, from 5 to 8 nucleotides, from 5 to 9 nucleotides, from 5 to 10 nucleotides, from 5 to 11 nucleotides, from 5 to 12 nucleotides, from 5 to 13 nucleotides, from 5 to 14 nucleotides, from 5 to 15 nucleotides, from 5 to 16 nucleotides, from 5 to 17 nucleotides, from 5 to 18 nucleotides, from 5 to 19 nucleotides, from 6 to 7 nucleotides, from 6 to 8 nucleotides, from 6 to 9 nucleotides, from 6 to 10 nucleotides, from 6 to 11 nucleotides, from 6 to 12 nucleotides, from 6 to 13 nucleotides, from 6 to 14 nucleotides, from 6 to 15 nucleotides, from 6 to 16 nucleotides, from 6 to 17 nucleotides, from 6 to 18 nucleotides, from 6 to 19 nucleotides, from 6 to 20 nucleotides, from 7 to 8 nucleotides, from 7 to 9 nucleotides, from 7 to 10 nucleotides, from 7 to 11 nucleotides, from 7 to 12 nucleotides, from 7 to 13 nucleotides, from 7 to 14 nucleotides, from 7 to 15 nucleotides, from 7 to 16 nucleotides, from 7 to 17 nucleotides, from 7 to 18 nucleotides, from 7 to 19 nucleotides, from 7 to 20 nucleotides, from 8 to 9 nucleotides, from 8 to 10 nucleotides, from 8 to 11 nucleotides, from 8 to 12 nucleotides, from 8 to 13 nucleotides, from 8 to 14 nucleotides, from 8 to 15 nucleotides, from 8 to 16 nucleotides, from 8 to 17 nucleotides, from 8 to 18 nucleotides, from 8 to 19 nucleotides, from 8 to 20 nucleotides, from 9 to 10 nucleotides, from 9 to 11 nucleotides, from 9 to 12 nucleotides, from 9 to 13 nucleotides, from 9 to 14 nucleotides, from 9 to 15 nucleotides, from 9 to 16 nucleotides, from 9 to 17 nucleotides, from 9 to 18 nucleotides, from 9 to 19 nucleotides, from 9 to 20 nucleotides, from 10 to 11 nucleotides, from 10 to 12 nucleotides, from 10 to 13 nucleotides, from 10 to 14 nucleotides, from 10 to 15 nucleotides, from 10 to 16 nucleotides, from 10 to 17 nucleotides, from 10 to 18 nucleotides, from 10 to 19 nucleotides, from 10 to 20 nucleotides, from 11 to 12 nucleotides, from 11 to 13 nucleotides, from 11 to 14 nucleotides, from 11 to 15 nucleotides, from 11 to 16 nucleotides, from 11 to 17 nucleotides, from 11 to 18 nucleotides, from 11 to 19 nucleotides, from 11 to 20 nucleotides, from 12 to 13 nucleotides, from 12 to 14 nucleotides, from 12 to 15 nucleotides, from 12 to 16 nucleotides, from 12 to 17 nucleotides, from 12 to 18 nucleotides, from 12 to 19 nucleotides, from 12 to 20 nucleotides, from 13 to 14 nucleotides, from 13 to 15 nucleotides, from 13 to 16 nucleotides, from 13 to 17 nucleotides, from 13 to 18 nucleotides, from 13 to 19 nucleotides, from 13 to 20 nucleotides, from 14 to 15 nucleotides, from 14 to 16 nucleotides, from 14 to 17 nucleotides, from 14 to 18 nucleotides, from 14 to 19 nucleotides, from 14 to 20 nucleotides, from 15 to 16 nucleotides, from 15 to 17 nucleotides, from 15 to 18 nucleotides, from 15 to 19 nucleotides, from 15 to 20 nucleotides, from 16 to 17 nucleotides, from 16 to 18 nucleotides, from 16 to 19 nucleotides, from 16 to 20 nucleotides, from 17 to 18 nucleotides, from 17 to 19 nucleotides, from 17 to 20 nucleotides, from 18 to 19 nucleotides, from 18 to 20 nucleotides, and from 19 to 20 nucleotides). In particular embodiments, the headpiece, the first building block tag, the second building block tag, the one or more additional building block tags, the library-identifying tag, the use tag, and/or

the origin tag, if present, have a length of less than 20 nucleotides (e.g., less than 19 nucleotides, less than 18 nucleotides, less than 17 nucleotides, less than 16 nucleotides, less than 15 nucleotides, less than 14 nucleotides, less than 13 nucleotides, less than 12 nucleotides, less than 11 nucleotides, less than 10 nucleotides, less than 9 nucleotides, less than 8 nucleotides, or less than 7 nucleotides).

5 In particular embodiments, the first building block tag and the second building block tag include the same number of nucleotides. In other embodiments, either the first building block tag or the second building block tag includes more than 8 nucleotides (e.g., more than 9 nucleotides, more than 10 nucleotides, more than 11 nucleotides, more than 12 nucleotides, more than 13 nucleotides, more than 14 nucleotides, and more than 15 nucleotides). In some embodiments, the first building block tag is a donor
 10 tag (e.g., as defined herein) having from 8 to 20 nucleotides (e.g., from 8 to 9 nucleotides, from 8 to 10 nucleotides, from 8 to 11 nucleotides, from 8 to 12 nucleotides, from 8 to 13 nucleotides, from 8 to 14 nucleotides, from 8 to 15 nucleotides, from 8 to 16 nucleotides, from 8 to 17 nucleotides, from 8 to 18 nucleotides, from 8 to 19 nucleotides, from 8 to 20 nucleotides, from 9 to 10 nucleotides, from 9 to 11 nucleotides, from 9 to 12 nucleotides, from 9 to 13 nucleotides, from 9 to 14 nucleotides, from 9 to 15
 15 nucleotides, from 9 to 16 nucleotides, from 9 to 17 nucleotides, from 9 to 18 nucleotides, from 9 to 19 nucleotides, from 9 to 20 nucleotides, from 10 to 11 nucleotides, from 10 to 12 nucleotides, from 10 to 13 nucleotides, from 10 to 14 nucleotides, from 10 to 15 nucleotides, from 10 to 16 nucleotides, from 10 to 17 nucleotides, from 10 to 18 nucleotides, from 10 to 19 nucleotides, from 10 to 20 nucleotides, from 11 to 12 nucleotides, from 11 to 13 nucleotides, from 11 to 14 nucleotides, from 11 to 15 nucleotides, from
 20 11 to 16 nucleotides, from 11 to 17 nucleotides, from 11 to 18 nucleotides, from 11 to 19 nucleotides, from 11 to 20 nucleotides, from 12 to 13 nucleotides, from 12 to 14 nucleotides, from 12 to 15 nucleotides, from 12 to 16 nucleotides, from 12 to 17 nucleotides, from 12 to 18 nucleotides, from 12 to 19 nucleotides, from 12 to 20 nucleotides, from 13 to 14 nucleotides, from 13 to 15 nucleotides, from 13 to 16 nucleotides, from 13 to 17 nucleotides, from 13 to 18 nucleotides, from 13 to 19 nucleotides, from
 25 13 to 20 nucleotides, from 14 to 15 nucleotides, from 14 to 16 nucleotides, from 14 to 17 nucleotides, from 14 to 18 nucleotides, from 14 to 19 nucleotides, from 14 to 20 nucleotides, from 15 to 16 nucleotides, from 15 to 17 nucleotides, from 15 to 18 nucleotides, from 15 to 19 nucleotides, from 15 to 20 nucleotides, from 16 to 17 nucleotides, from 16 to 18 nucleotides, from 16 to 19 nucleotides, from 16 to 20 nucleotides, from 17 to 18 nucleotides, from 17 to 19 nucleotides, from 17 to 20 nucleotides, from
 30 18 to 19 nucleotides, from 18 to 20 nucleotides, and from 19 to 20 nucleotides).

Definitions

By “2’-substituted nucleotide” is meant a nucleotide base having a substitution at the 2’-position of ribose in the base.

35 By “about” is meant +/- 10% of the recited value.

By “bifunctional” is meant having two reactive groups that allow for binding of two chemical moieties. For example, a bifunctional linker is a linker, as described herein, having two reactive groups that allow for binding of a headpiece and a chemical entity

By “binding” is meant attaching by a covalent bond or a non-covalent bond. Non-covalent bonds include those formed by van der Waals forces, hydrogen bonds, ionic bonds, entrapment or physical encapsulation, absorption, adsorption, and/or other intermolecular forces. Binding can be effectuated by any useful means, such as by enzymatic binding (e.g., enzymatic ligation) or by chemical binding (e.g., chemical ligation).

By “building block” is meant a structural unit of a chemical entity, where the unit is directly linked to other chemical structural units or indirectly linked through the scaffold. When the chemical entity is polymeric or oligomeric, the building blocks are the monomeric units of the polymer or oligomer. Building blocks can have one or more diversity nodes that allow for the addition of one or more other building blocks or scaffolds. In most cases, each diversity node is a functional group capable of reacting with one or more building blocks or scaffolds to form a chemical entity. Generally, the building blocks have at least two diversity nodes (or reactive functional groups), but some building blocks may have one diversity node (or reactive functional group). Alternatively, the encoded chemical or binding steps may include several chemical components (e.g., multi-component condensation reactions or multi-step processes). Reactive groups on two different building blocks should be complementary, i.e., capable of reacting together to form a covalent or a non-covalent bond.

By “building block tag” is meant an oligonucleotide portion of the library that encodes the addition (e.g., by a binding reaction) of a component (i.e., a scaffold or a building block), the headpiece in the library, the identity of the library, the use of the library, and/or the origin of a library member. By “acceptor tag” is meant a building block tag having a reactive entity (e.g., a hydroxyl group at the 3'-terminus in the case of enzymatic ligation). By “donor tag” is meant a building block tag having an entity capable of reacting with the reactive entity on the acceptor tag (e.g., a phosphoryl group at the 5'-terminus in the case of enzymatic ligation).

By “chemical entity” is meant a compound comprising one or more building blocks and optionally a scaffold. The chemical entity can be any small molecule or peptide drug or drug candidate designed or built to have one or more desired characteristics, e.g., capacity to bind a biological target, solubility, availability of hydrogen bond donors and acceptors, rotational degrees of freedom of the bonds, positive charge, negative charge, and the like. In certain embodiments, the chemical entity can be reacted further as a bifunctional or trifunctional (or greater) entity.

By “chemically co-reactive pair” is meant a pair of reactive groups that participates in a modular reaction with high yield and a high thermodynamic gain, thus producing a spacer. Exemplary reactions and chemically co-reactive pairs include a Huisgen 1,3-dipolar cycloaddition reaction with a pair of an optionally substituted alkynyl group and an optionally substituted azido group; a Diels-Alder reaction with a pair of an optionally substituted diene having a 4π electron system and an optionally substituted

dienophile or an optionally substituted heterodienophile having a 2π electron system; a ring opening reaction with a nucleophile and a strained heterocyclyl electrophile; a splint ligation reaction with a phosphorothioate group and an iodo group; and a reductive amination reaction with an aldehyde group and an amino group, as described herein.

5 By “complex” or “ligated complex” is meant a headpiece that is operatively associated with a chemical entity and/or one or more oligonucleotide tags by a covalent bond or a non-covalent bond. The complex can optionally include a bifunctional linker between the chemical entity and the headpiece.

By “component” of a chemical entity is meant either a scaffold or a building block.

10 By “diversity node” is meant a functional group at a position in the scaffold or the building block that allows for adding another building block.

By “headpiece” is meant a starting oligonucleotide for library synthesis that is operatively linked to a component of a chemical entity and to a building block tag. Optionally, a bifunctional linker connects the headpiece to the component.

15 By “library” is meant a collection of molecules or chemical entities. Optionally, the molecules or chemical entities are bound to one or more oligonucleotides that encodes for the molecules or portions of the chemical entity.

By “linker” is meant a chemical connecting entity that links the headpiece to a chemical entity.

20 By “multivalent cation” is meant a cation capable of forming more than one bond with more than one ligand or anion. The multivalent cation can form either an ionic complex or a coordination complex. Exemplary multivalent cations include those from the alkali earth metals (e.g., magnesium) and transition metals (e.g., manganese (II) or cobalt (III)), and those that are optionally bound to one or more anions and/or one or more univalent or polydentate ligands, such as chloride, amine, and/or ethylenediamine.

25 By “oligonucleotide” is meant a polymer of nucleotides having a 5'-terminus, a 3'-terminus, and one or more nucleotides at the internal position between the 5'- and 3'-termini. The oligonucleotide may include DNA, RNA, or any derivative thereof known in the art that can be synthesized and used for base-pair recognition. The oligonucleotide does not have to have contiguous bases but can be interspersed with linker moieties. The oligonucleotide polymer may include natural bases (e.g., adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, deoxycytidine, inosine, or diamino purine), base analogs (e.g., 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 30 3-methyl adenosine, C5-propynylcytidine, C5-propynyluridine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-methylcytidine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine), modified nucleotides (e.g., 2'-substituted nucleotides, such as 2'-O-methylated bases and 2'-fluoro bases), intercalated bases, modified sugars (e.g., 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose), and/or modified phosphate groups (e.g., 35 phosphorothioates and 5'-N-phosphoramidite linkages). Other modified bases are described herein. By “acceptor oligonucleotide” is meant an oligonucleotide having a reactive entity (e.g., a hydroxyl group at the 3'-terminus in the case of enzymatic ligation or an optionally substituted azido group in the case of

chemical ligation). By “donor oligonucleotide” is meant an oligonucleotide having an entity capable of reacting with the reactive entity on the acceptor oligonucleotide (e.g., a phosphoryl group at the 5'-terminus in the case of enzymatic ligation or an optionally substituted alkynyl group in the case of chemical ligation).

5 By “operatively linked” or “operatively associated” is meant that two or more chemical structures are directly or indirectly linked together in such a way as to remain linked through the various manipulations they are expected to undergo. Typically, the chemical entity and the headpiece are operatively linked in an indirect manner (e.g., covalently via an appropriate linker). For example, the linker may be a bifunctional moiety with a site of attachment for chemical entity and a site of attachment
10 for the headpiece. In addition, the chemical entity and the oligonucleotide tag can be operatively linked directly or indirectly (e.g., covalently via an appropriate linker).

By “protecting group” is meant a group intended to protect the 3'-terminus or 5'-terminus of an oligonucleotide against undesirable reactions during one or more binding steps of tagging a DNA-encoded library. Commonly used protecting groups are disclosed in Greene, “Protective Groups in
15 Organic Synthesis,” 4th Edition (John Wiley & Sons, New York, 2007), which is incorporated herein by reference. Exemplary protecting groups include irreversible protecting groups, such as dideoxynucleotides and dideoxynucleosides (ddNTP or ddN), and, more preferably, reversible protecting groups for hydroxyl groups, such as ester groups (e.g., O-(α -methoxyethyl)ester, O-isovaleryl ester, and O-levulinyl ester), trityl groups (e.g., dimethoxytrityl and monomethoxytrityl), xanthenyl groups (e.g., 9-
20 phenylxanthen-9-yl and 9-(p-methoxyphenyl)xanthen-9-yl), acyl groups (e.g., phenoxyacetyl and acetyl), and silyl groups (e.g., t-butyltrimethylsilyl).

By “purifying” is meant removing any unreacted product or any agent present in a reaction mixture that may reduce the activity of a chemical or biological agent to be used in a successive step. Purifying can include one or more of chromatographic separation, electrophoretic separation, and
25 precipitation of the unreacted product or reagent to be removed.

By “scaffold” is meant a chemical moiety that displays one or more diversity nodes in a particular special geometry. Diversity nodes are typically attached to the scaffold during library synthesis, but in some cases one diversity node can be attached to the scaffold prior to library synthesis (e.g., addition of one or more building blocks and/or one or more tags). In some embodiments, the
30 scaffold is derivatized such that it can be orthogonally deprotected during library synthesis and subsequently reacted with different diversity nodes.

By “small molecule” drug or “small molecule” drug candidate is meant a molecule that has a molecular weight below about 1,000 Daltons. Small molecules may be organic or inorganic, isolated (e.g., from compound libraries or natural sources), or obtained by derivatization of known compounds.

35 By “substantial identity” or “substantially identical” is meant a polypeptide or polynucleotide sequence that has the same polypeptide or polynucleotide sequence, respectively, as a reference sequence, or has a specified percentage of amino acid residues or nucleotides, respectively, that are the same at the

corresponding location within a reference sequence when the two sequences are optimally aligned. For example, an amino acid sequence that is “substantially identical” to a reference sequence has at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to the reference amino acid sequence. For polypeptides, the length of comparison sequences will generally be at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 contiguous amino acids, more preferably at least 25, 50, 75, 90, 100, 150, 200, 250, 300, or 350 contiguous amino acids, and most preferably the full-length amino acid sequence. For nucleic acids, the length of comparison sequences will generally be at least 5 contiguous nucleotides, preferably at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 contiguous nucleotides, and most preferably the full length nucleotide sequence. Sequence identity may be measured using sequence analysis software on the default setting (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Such software may match similar sequences by assigning degrees of homology to various substitutions, deletions, and other modifications.

By “tailpiece” is meant an oligonucleotide portion of the library that is attached to the complex after the addition of all of the building block tags and encodes for the identity of the library, the use of the library, and/or the origin of a library member.

Other features and advantages of the invention will be apparent from the following Detailed Description and the claims.

Brief Description of the Drawings

Figure 1 shows an exemplary method for the general synthesis of chemical libraries using single-stranded DNA tags that are joined sequentially by means of enzymatic and/or chemical ligation. “BB” refers to building block.

Figures 2A-2B show exemplary methods for single-stranded DNA tagging of libraries using enzymatic ligation. Figure 2A shows an exemplary method for tagging libraries using single-stranded enzymatic ligation with a protected (re-installed) 5'-monophosphate (5'-P) oligonucleotide, where gray boxes refer to 2'-OMe nucleotides, “X” refers to a protecting group or a component of a chemical entity, and “PNK” refers to polynucleotide kinase. Figure 2B shows an exemplary method for tagging libraries using single-stranded ligation with a protected 3'-OH oligonucleotide, where black boxes attached to -O- refer to a protecting group of the 3'-OH terminus and “LC” refers to liquid chromatographic separation of the protecting group.

Figure 3 shows an exemplary method for tagging libraries using single-stranded ligation with a 5'-preadenylated (labeled “5'-App”) oligonucleotide (headpiece) with a 3'-terminus that is blocked, e.g., by a chemical entity (labeled “X-3”). This method can be used to ligate a 5'-phosphorylated oligonucleotide tag (labeled “Tag A”) to the headpiece and additional tags having a 3'-OH terminus (labeled “Tag B” and “Tag C”) to the complex in the presence of ATP.

Figures 4A-4E show exemplary complexes, each having a headpiece, a linker, and a small molecule including a scaffold ("S") and diversity nodes A, B, and C. The dark gray boxes refer to 2'-OMe nucleotides, and the dotted lines refer to the presence of one or more complementary bases. Figures 4A-4B are schematics for complexes having a single-stranded linear oligonucleotide headpiece, where the linker and small molecule are connected to the 3'-terminus (Figure 4A) or the 5'-terminus (Figure 4B) of the headpiece. Figures 4C-4D are schematics for complexes having a single-stranded hairpin oligonucleotide headpiece, where the linker and small molecule are connected to the internal position (Figure 4C) or the 3'-terminus (Figure 4D) of the headpiece. Figure 4E shows an exemplary method for tagging libraries having a hairpin oligonucleotide headpiece, where the star refers to a chemical moiety and "Y" at the 3'-terminus refers to a protecting group. Oligonucleotide tags are labeled 1-4, and the adapter sequence is the black line at the 5'-terminus.

Figures 5A-5C show oligonucleotide ligation by T4 RNA ligase or CircLigase™ ssDNA ligase. Figure 5A is a schematic of the enzymatic ligation reaction. The donor oligonucleotide is 5'-phosphorylated and carries a 3'-fluorescein label, imitating a headpiece with a chemical library at 3' end. The acceptor oligonucleotide is not phosphorylated. Figure 5B shows gel electrophoresis analysis of a ligation reaction on an 8M urea/ 15% polyacrylamide gel (PAAG). "SM" refers to fluorescently labeled donor, "Product" refers to ligation product, and "Adenylated donor" refers to 5' App-Donor, as described above. Figure 5C shows high yield ligation achieved for T4 RNA ligase at high enzyme and oligonucleotide concentrations.

Figures 6A-6B represent optimization of PEG molecular weight (Figure 6A) and concentration (Figure 6B) to achieve maximal ligation yield by T4 RNA ligase. Reaction conditions are as described above for Figures 5A-5C. Figure 6A is graph quantifying the electrophoretic analysis of a ligation reaction with MNA/DNA 15mer donor and acceptor tags after incubation for 5 hours or 20 hours with 25% (w/v) PEG having a molecular weight from 300 to 20,000 (20K). Figure 6B shows the effect of concentration on ligation after incubation for 18-20 hours in the presence of 5% to 45% (w/v) of PEG4600.

Figures 7A-7B show a correlation between ligation efficiency by CircLigase™ (Figure 7A) and T4 RNA ligase (Figure 7B) and length of the donor or acceptor oligonucleotides. Figure 7A depicts a graph quantifying the effect of the acceptor length on ligation yield in the CircLigase™ ligation reaction. Figure 7B depicts a graph and a table quantifying the effect of nucleotide length of the acceptor and donor MNA/DNA tags on single-stranded ligation with T4 RNA ligase. These data represent an average of two independent experiments obtained by densitometry of fluorescent gels at 450 nm excitation.

Figures 8A-8B are LC-MS spectra for a MNA/DNA tag before and after phosphorylation. Data are shown for 15mer tag 5'-HO-mUAC GTA TAC GAC TGmG-OH-3' (SEQ ID NO: 13) (at 250 μM) before (Figure 8A) and after (Figure 8B) reaction with T4 polynucleotide kinase (50 units per 5 nmole of tag).

Figure 9 shows an electrophoretic gel for sequential single-stranded ligation of tags A-C. The 3'-terminus included fluorescein to represent a library compound (or chemical entity), and the asterisk (*) indicates purification of the ligated product (or complex) prior to phosphorylation.

Figures 10A-10B show schematics of a "chemically co-reactive pair" reaction between donor and acceptor oligonucleotides resulting in a 5-atom "short" spacer (Figure 10A) and a 24-atom "long" spacer (Figure 10B).

Figures 11A-11E show results of reverse transcription (RT) and PCR analysis of 75mer DNA templates containing a short or a long single spacer, as depicted in Figures 10A-10B. Figure 11A is a schematic of the RT reaction. LC-MS spectra of the RT were recorded at both 260 nm and 650 nm for the control 75mer DNA template (Figure 11B), the 75mer DNA template containing a single 5-atom ("short") spacer (Figure 11C), and the 75mer DNA template containing a single 24-atom ("long") spacer (Figure 11D). Figure 11E shows RT-PCR analysis of the control 75mer DNA template ("templ75"), a 75mer DNA template with a 5-atom spacer ("short click"), and a 75mer DNA template with a 24-atom spacer ("long click").

Figures 12A-12G show the results of a chemical ligation reaction between a 5'-iodo-modified DNA oligonucleotide and a 3'-phosphorothioate DNA oligonucleotide in the presence or absence of a complementary splint oligonucleotide. Figure 12A shows an exemplary schematic of the reaction. The 5'-iodo oligonucleotide is labeled with 6-FAM at 3'-terminus, while the 3'-phosphorothioate oligonucleotide is labeled with Cy5 at the 5'-terminus. Figure 12B shows a gel electrophoresis analysis of the ligation reactions in the presence (+spl) or absence (-spl) of a complementary splint. CCy5 and CFL indicate visible bands of Cy5 and fluorescein-labeled starting material, respectively. Figure 12C shows a time course of the splinted ligation reaction under the above conditions, which was quantified using Cy5 (635 nm) and fluorescein (450 nm) detection. Figure 12D shows LC-MS analysis of the ligation of CFL and CCy5 in the absence (top, at 260 nm, 495 nm, and 650 nm) and presence (bottom, at 260 nm, 495 nm, and 650 nm) of a splint, where ligation reactions were incubated for seven days. Figure 12E shows LC-MS analysis of the ligation of CFL and CCy5 in the absence a splint (at 260 nm, 495 nm, and 650 nm), where ligation reactions were incubated for eight days. Figure 12F shows MS analysis of reaction of CFL oligonucleotide with piperidine, where this reaction was intended to displace iodine. Reaction conditions included oligonucleotides at 100 μ M, piperidine at 40 mM (400 equivalents) in 100 mM borate buffer, pH 9.5, for 20 hrs at room temperature (left); and oligonucleotides at 400 μ M, piperidine at 2 M (4,000 equivalents) in 200 mM borate buffer, pH 9.5, for 2 hrs at 65°C (right). Figure 12G shows MS analysis of a splinted ligation reaction of CFL and CCy5 oligonucleotides at 50 μ M performed in the presence of 400 equivalents of piperidine in 100 mM borate buffer, pH 9.5, for 20 hrs at room temperature.

Figures 13A-13C shows the use of modified oligonucleotides to minimize shuffling. Figure 13A shows an LC-MS analysis of a single-stranded ligation reaction of a 5'-phosphorylated headpiece ssHP (3,636 Da) and a tag (tag 15; 2,469 Da) having 2'-O methyl nucleotides. The LC-MS analysis showed

three peaks: peak 1 for the tag (2,469 Da); peak 2 for the adenylated headpiece (3,965 Da); and peak 3 having two (in some instances three) sub-peaks containing products with molecular weights of 6,089 Da (expected ligation product); 5,769 Da (expected 6,089 Da - 320 Da); and 6,409 Da (expected 6,089 Da + 320 Da). This mass difference of 320 Da corresponds exactly to either removal or addition of an extra 2'-O-Me C nucleotide. Figures 13B-1 to 13B-3 show a non-limiting, proposed mechanism of the nucleotide shuffling, where about 90% of the reaction provides the expected (normal) ligation product and about 10% of the reaction provides aberrant ligation products ("Product -1 nt" and "Product + 1 nt"). Figure 13C shows an LC-MS analysis of ligation of headpiece HP-PS with tag 15. The headpiece HP-PS has the sequence the headpiece ssHP but includes a phosphorothioate linkage at the 5'-terminus. LC analysis showed three peaks: peak 1 for the tag (2,469), peak 2 for the adenylated headpiece (3,984), and peak 3 for a single ligation product (6,107) with almost no nucleotide shuffling observed. Traces of +/- 320 peaks likely correspond to the oxidative conversion of the phosphorothioate linkage into a native phosphodiester linkage or are due to incomplete sulfurization.

Figure 14 is a graph showing separation of library members using size exclusion chromatography, where target-bound library members (left on graph) elute at a shorter time than unbound library members (right on graph).

Figure 15A is an exemplary schematic showing the chemical ligation of encoding DNA tags using a single chemistry that is not splint-dependent, e.g. 5'-azido/3'-alkynyl. The reactive groups are present on the 3' and 5' ends of each tag (Tag A, B, and C), and one of the reactive groups on either end (for example, the 3' end) is protected to prevent the cyclization, polymerization, or wrong-cycle ligation of the tags. The cycle of tag ligation includes chemical ligation, followed by deprotection of the remaining functional group to render the growing ligated entity competent for the next cycle of ligation. Each cycle also includes addition of one or more building blocks (BBA, BBB, and BBC, which are encoded by Tag A, B, and C, respectively). The chemical ligation process can optionally include addition of a tailpiece.

Figure 15B is an exemplary schematic showing the chemical ligation of encoding DNA tags using a single chemistry that is splint-dependent. The template-dependent nature of this approach reduces the frequency of occurrence of tag polymerization, tag cyclization, as well as of mistagging events. Similar to Figure 15A, this schematic includes tags (Tag A, B, and C) and one or more building blocks encoded by tags (BBA, BBB, and BBC).

Figure 15C is an exemplary schematic showing the use of a succession of chemically ligated tags as a template for template-dependent polymerization, generating cDNA that is competent for PCR amplification and sequencing, as well as using a template-dependent polymerase capable of reading through the chemically ligated junctions.

Figure 16A is an exemplary schematic showing the chemical ligation of encoding DNA tags using TIPS-protected alkynyl tags and "click" chemistry. Each cycle of library synthesis includes Cu(I)-catalyzed chemical ligation of the TIPS-protected tag to the deprotected alkyne from the previous cycle.

After the ligation, the TIPS group is removed (deprotected), thereby activating the alkyne for the next chemical ligation step.

Figure 16B shows the structure of DMT-succinyl-3'-O-TIPS-propargyl uridine CPG that is used to initiate solid-phase synthesis of oligonucleotides bearing 3'-O-TIPS-propargyl uridine at the 3'-terminus.

Figure 16C is an exemplary schematic showing the use of a succession of "click" chemically ligated tags as a template for template-dependent polymerization, generating cDNA that is competent for PCR amplification and sequencing, as well as using a template-dependent polymerase capable of reading through the "click" chemically ligated junctions.

Figures 17A-17C show the synthesis of 5'-biotinylated, "single-click" templates Y55 and Y185. Figure 17A provides an exemplary schematic. Figure 17B and Figure 17C show LC-MS analysis of Y55 and Y185, respectively.

Figures 18A-18C provide an exemplary assay for the "read-through" of a "single-click" template. Figure 18A shows a schematic, where FAM-labeled primer is annealed to the biotinylated template and is incubated with the template-dependent polymerase, according to the manufacturer's recommended conditions. The complexes are subsequently incubated with streptavidin beads, washed, eluted with NaOH, and then neutralized. After neutralization the samples are analyzed by LC-MS. Figure 18B and Figure 18C show LC-MS data of the Klenow fragment copying of templates Y55 and Y185, respectively.

Figures 19A-19D provides the synthesis of 5'-biotinylated "double-click" template YDC and "triple-click template" YTC using a TIPS-protected alkynyl tag. Figures 19A and 19B show exemplary schematics for this synthesis. Figures 19C and 19D show LC-MS analysis of the YDC and YTC templates respectively.

Figures 20A-20C provide an exemplary click "read-through" assay using "double-click" and "triple-click" templates. Figure 20A is a schematic, where FAM-labeled primer is annealed to the biotinylated template and is incubated with Klenow fragment of *E.coli* DNA polymerase I according to the manufacturer's recommended reaction conditions. The complexes are incubated with streptavidin beads, washed, eluted with NaOH, and neutralized. After the neutralization, the samples are assayed by LC-MS. Figures 20B and 20C show LC-MS data of the Klenow fragment copying of the templates YDC and YTC, respectively.

Figure 21 is a graph showing the efficiency of the click "read-through" using "single-click", "double-click" and "triple-click" templates in comparison to a control "no-click" DNA template. These data were obtained using the "read-through" assay described herein, and the yields were measured by LC MS analysis by comparison to an internal standard.

Figures 22A-22C provide exemplary schematics of chemical ligation with orthogonal chemistry. Figure 22A is a schematic of the chemical ligation strategy for DNA encoding tags that (i) utilizes two successive orthogonal chemistries for (ii) available read-through strategies. Each tag contains two

orthogonal reactive groups, indicated by differing symbols for the 5'-terminus and the 3'-terminus of each tag. In each successive cycle of chemical ligation, an orthogonal chemistry is used. This strategy reduces the frequency of occurrence of mistagging events and may also be used without the protection of the reactive terminal groups. Figure 22B is a schematic of the template-dependent polymerization "read-though" of a template generated by the orthogonal chemical ligation of orthogonal DNA tags to generate cDNA from which the sequence of the tags can be deduced. Figure 22C is the same as Figure 22B but includes a self-priming tailpiece, which may be rendered double-stranded by restriction digestion to facilitate strand-separation during PCR amplification.

Figure 23 is an exemplary schematic showing the chemical ligation strategy for DNA encoding tags that utilizes two specific successive orthogonal chemistries. Each tag contains click-reactive and phosphorothioate/iodo-reactive groups. Tags bearing orthogonal reactive groups at their 3' and 5' ends cannot polymerize and have a reduced frequency of occurrence of mistagging events. Without wishing to be limited, this approach may eliminate the need for the TIPS-protection of the 3'-alkyne. In cycle A, the 5'-iodo/3'-alkynyl tag is ligated using splint-dependent ligation to the 3'-phosphorothioate headpiece, leaving a reactive 3' alkyne for the next cycle of chemical ligation to a 5'-azido/3'-phosphorothioate tag. The orthogonal ligation cycles may be repeated as many times as is desired.

Figures 24A-24B show the protection and use of 3'-phosphorothioate/5'-iodo groups on DNA tags. Figure 24A shows an exemplary schematic for using protecting groups (PG) for these tags. Figure 24B shows an exemplary scheme for use of 3'-phosphorothioate/5'-iodo tags to chemically ligate succession of encoding DNA tags that encode a chemical library covalently installed upon the 5'-terminus.

Figures 25A-25B show the protection and use of 3'-phosphorothioate groups on DNA tags. Figure 25A shows the scheme for protection of these groups. Figure 25B shows the scheme for use of 3'-phosphorothioate/5'-azido and 3'-propargyl/5'-iodo tags to chemically ligate a succession of orthogonal encoding DNA tags that encode a chemical library covalently installed upon the 5'-terminus.

Detailed Description

The invention features methods of using single-stranded ligation to install oligonucleotide tags onto chemical entity-oligonucleotide complexes. This method can be used to create diverse libraries of selectable chemical entities by establishing an encoded relationship between particular tags and particular chemical reactions or building blocks. To identify one or more chemical entities, the oligonucleotide tags can be amplified, cloned, sequenced, and correlated by using the established relationship. In particular, reaction conditions that promote single-stranded ligation of tags were identified. These conditions include the use of one or more 2'-substituted nucleotides (e.g., 2'-O-methyl nucleotides or 2'-fluoro nucleotides) within the tags, the use of tags of particular length (e.g., between 5 and 15 nucleotides), the use of one or more enzymes (e.g., RNA ligase and/or DNA ligase), and/or the use of one or more agents during ligation (e.g., poly ethylene glycol and/or a soluble multivalent cation, such as $\text{Co}(\text{NH}_3)_6\text{Cl}_3$).

These methods additionally include methods of chemically joining oligonucleotides, such that the sequence of the joined oligonucleotide product may be utilized as a template for a template-dependent polymerase reaction. Methods of creating and tagging libraries of these complexes are described in detail below.

5

Methods for tagging encoded libraries

This invention features a method for operatively linking oligonucleotide tags with chemical entities, such that encoding relationships may be established between the sequence of the tag and the structural units (or building blocks) of the chemical entity. In particular, the identity and/or history of a chemical entity can be inferred from the sequence of bases in the oligonucleotide. Using this method, a library including diverse chemical entities or members (e.g., small molecules or peptides) can be addressed with a particular tag sequence.

Generally, these methods include the use of a headpiece, which has at least one functional group that may be elaborated chemically and at least one functional group to which a single-stranded oligonucleotide may be bound (or ligated). Binding can be effectuated by any useful means, such as by enzymatic binding (e.g., ligation with one or more of an RNA ligase and/or a DNA ligase) or by chemical binding (e.g., by a substitution reaction between two functional groups, such as a nucleophile and a leaving group).

To create numerous chemical entities within the library, a solution containing the headpiece can be divided into multiple aliquots and then placed into a multiplicity of physically separate compartments, such as the wells of a multiwell plate. Generally, this is the "split" step. Within each compartment or well, successive chemical reaction and ligation steps are performed with a single-stranded tag within each aliquot. The relationship between the chemical reaction conditions and the sequence of the single-stranded tag are recorded. The reaction and ligation steps may be performed in any order. Then, the reacted and ligated aliquots are combined or "pooled," and optionally purification may be performed at this point. These split and pool steps can be optionally repeated.

Next, the library can be tested and/or selected for a particular characteristic or function, as described herein. For example, the mixture of tagged chemical entities can be separated into at least two populations, where the first population binds to a particular biological target and the second population does not. The first population can then be selectively captured (e.g., by eluting on a column providing the target of interest or by incubating the aliquot with the target of interest) and, optionally, further analyzed or tested, such as with optional washing, purification, negative selection, positive selection, or separation steps.

Finally, the chemical histories of one or more members (or chemical entities) within the selected population can be determined by the sequence of the operatively linked oligonucleotide. Upon correlating the sequence with the particular building block, this method can identify the individual members of the library with the selected characteristic (e.g., an increased tendency to bind to the target

protein and thereby elicit a therapeutic effect). For further testing and optimization, candidate therapeutic compounds may then be prepared by synthesizing the identified library members with or without their associated oligonucleotide tags.

Figures 1-3 provide various exemplary methods for tagging libraries using single-stranded ligation with a headpiece, where tags can be ligated on the 5'-terminus or the 3'-terminus of the headpiece. To control the order in which the tags are ligated and to reduce side reactions, these methods ensure that only one reactive 5'-terminus and one reactive 3'-terminus are present during ligation. Furthermore, these exemplary methods use 2'-substituted nucleotides (e.g., mixed 2'-deoxy/2'-O-methyl nucleotides) in the tags, and these tags act as templates for a DNA- or RNA-dependent polymerase capable of polymerizing nucleotides in a template-dependent fashion. Without wishing to be limited by theory, the use of one or more 2'-substituted nucleotides (e.g., 2'-O-methyl nucleotides and/or 2'-fluoro nucleotides) within a tag could promote ligation by RNA ligase by more closely resembling RNA, while preserving both the physical and chemical robustness of the recording medium as well as the ability to extract sequence information using template-dependent polymerization.

Figure 1 provides an exemplary method for reducing side reactions, where the ligated complex and tags are designed to avoid unwanted reactions between reactive 3'-OH and 5'-monophosphate ("5'-P") groups. In particular, this scheme depicts the phosphorylation-ligation cycle approach. During ligation, only one 3'-OH group (in the tag) and one 5'-P group (in the headpiece) are available, and, thus, only one ligation event is possible. Following the ligation and purification steps, a 5'-OH group is formed in the complex, and this group can be converted into a 5'-P for adding subsequent oligonucleotide tags. The 3'-terminus of the complex is blocked by X, which can be a protecting group or a component of a chemical entity (e.g., optionally including a linker that acts as a spacer between the chemical entity and the headpiece).

As shown in Figure 1, the exemplary method includes ligation of building block tag 1 ("tag 1") to the 5'-terminus of the headpiece, thereby creating a complex, and performing successive ligations to the 5'-terminus of the complex. The reactive 5'-terminus is a phosphate group on the complex, and the reactive 3'-terminus is a hydroxyl group on the tags. After the addition of each tag, the ligated complex is separated from the unreacted, unligated headpiece and tags and from other reagents (e.g., phosphate, cobalt, or other reagents present during the ligation step). Separation can be accomplished by any useful method (e.g., by chromatographic or electrophoretic separation of ligated and non-ligated products or by precipitation of a reagent). Then, the ligated complex is exposed to an agent (e.g., a polynucleotide kinase or a chemical phosphorylating agent) to form a phosphate group on the 5'-terminus of the complex. The separation and phosphorylation steps may be performed in either order. In particular, if a kinase is used in the phosphorylation step, the kinase should be inactivated or removed prior to the addition of the subsequent tags that may also contain a 5'-OH group, or any reagents that can inhibit the kinase should be removed from the reaction mixture prior to the phosphorylation step.

In another embodiment, the method includes binding successive tags from the 3'-terminus of the preceding ligated complex. In this method, the ligated complex lacks a reactive 3'-OH group immediately after the ligation step but contains a group that can be converted into a 3'-OH group (e.g., by release of a protecting group). Figure 2A provides a schematic showing an exemplary method for tagging the 3'-terminus of a complex, and Figure 2B provides an exemplary reaction scheme for a protected 3'-terminus that contains convertible 3'-OH group upon release of the 3'-linked protecting group. As shown in Figure 2A, building block tag 1 ("tag 1") has a 3'-protected group. In the first step, the exemplary method includes ligation of the tag to the 3'-terminus of the headpiece, thereby creating a complex. Successive ligations are performed to the 3'-terminus of the complex. The reactive 5'-terminus is a phosphate group on the tag, and the reactive 3'-terminus is a hydroxyl group on the complex. After the addition of each tag, the ligated complex is deprotected (e.g., by the addition of a hydrolyzing agent) to release the 3'-protecting group.

In yet another embodiment, the method includes binding successive tags by using a 5'-preadenylated (5'-App) oligonucleotide and a ligase (e.g., T4 RNA ligase). In the presence of ATP, T4 RNA ligase will use the ATP cofactor to form an adenylated intermediate prior to ligation. In the absence of ATP, T4 RNA ligase will only ligate preadenylated oligonucleotides, and possible side reactions with 5'-P oligonucleotides will not occur. Thus, single-stranded ligation with reduced side reactions can be performed with a chemically synthesized 5'-App oligonucleotide in the presence of 5'-monophosphorylated tag, where the 5'-App oligonucleotide can be ligated to a headpiece prior to tagging or to a complex formed after multiple rounds of tagging.

Figure 3 provides a schematic showing an exemplary method for tagging the 5'-terminus of a preadenylated headpiece. Adenylation of the donor nucleotide at the 5'-phosphate group is the first step in the ligation reaction, and this reaction generally requires one molecule of ATP. In the second step, the 3'-OH group of the acceptor oligonucleotide reacts with the adenylated donor and forms a diester bond between two oligonucleotides, thus releasing one AMP molecule. The chemically adenylated 5'-phosphate group of the donor oligonucleotide imitates a product of the first step of the ligation reaction and can be ligated to the second oligonucleotide in the absence of ATP. In the following scheme, a 5'-App headpiece is ligated to the 3'-OH group of a 5'-phosphorylated oligonucleotide tag (labeled "Tag A"). Due to the presence of the adenylated 5'-terminus of the oligonucleotide, ligation can occur in the absence of ATP. Under these conditions, the 5'-phosphate group of Tag A does not serve as a ligation donor. Building block Tag B can be ligated by providing a nucleotide having a 3'-OH terminus (labeled "Tag B") in the presence of ATP, and additional tags (labeled "Tag C") can be included.

In Figure 3, the 3'-terminus of the headpiece can be blocked with any protecting group (e.g., an irreversible protecting group, such as ddN, or a reversible protecting group). In the first step, the method includes ligation of the tag to the 5'-terminus of the headpiece in the absence of ATP, thereby creating a complex. Successive ligations are performed to the 5'-terminus of the complex in the presence of ATP. This method can be modified in order to perform successive ligation to the 3'-terminus of a complex. For

example, the method can include the use of a 5'-preadenylated tag and a headpiece having a reactive 3'-OH terminus. This method may further require blocking the 3'-terminus of the tag to avoid cross-reactions between tags, such as the method described above and in Figure 2.

The general method provided in Figure 3 can be modified by replacing the primer with a headpiece. In this case, the headpiece has to be adenylated chemically at the 5'-terminus, and Tag A is phosphorylated at 5'-terminus. Ligation of this phosphorylated Tag A to the adenylated headpiece occurs in the same standard conditions, described herein, but omitting ATP. By using this ligation condition, the ligation of phosphorylated 5' terminus can be prevented. In the next step, ligation of Tag B requires that this tag have a free hydroxyl group at 5'-terminus (i.e., non-phosphorylated). Successive ligation reactions can be performed in the presence of ATP, followed by phosphorylation of the 5'-terminus of the resulting oligonucleotide if further extension of the tags (e.g., Tag C in Figure 3) is desired.

The methods described herein can include any number of optional steps to diversify the library or to interrogate the members of the library. For any tagging method described herein (e.g., as in Figures 1-3), successive "n" number of tags can be added with additional "n" number of ligation, separation, and/or phosphorylation steps. Exemplary optional steps include restriction of library members using one or more restriction endonucleases; ligation of one or more adapter sequences to one or both of the library termini, e.g., such as one or more adapter sequences to provide a priming sequence for amplification and sequencing or to provide a label, such as biotin, for immobilization of the sequence; reverse-transcription or transcription, optionally followed by reverse-transcription, of the assembled tags in the complex using a reverse transcriptase, transcriptase, or another template-dependent polymerase; amplification of the assembled tags in the complex using, e.g., PCR; generation of clonal isolates of one or more populations of assembled tags in the complex, e.g., by use of bacterial transformation, emulsion formation, dilution, surface capture techniques, etc.; amplification of clonal isolates of one or more populations of assembled tag in the complex, e.g., by using clonal isolates as templates for template-dependent polymerization of nucleotides; and sequence determination of clonal isolates of one or more populations of assembled tags in the complex, e.g., by using clonal isolates as templates for template-dependent polymerization with fluorescently labeled nucleotides. Additional methods for amplifying and sequencing the oligonucleotide tags are described herein.

These methods can be used to identify and discover any number of chemical entities with a particular characteristic or function, e.g., in a selection step. The desired characteristic or function may be used as the basis for partitioning the library into at least two parts with the concomitant enrichment of at least one of the members or related members in the library with the desired function. In particular embodiments, the method comprises identifying a small drug-like library member that binds or inactivates a protein of therapeutic interest. In another embodiment, a sequence of chemical reactions is designed, and a set of building blocks is chosen so that the reaction of the chosen building blocks under the defined chemical conditions will generate a combinatorial plurality of molecules (or a library of molecules), where one or more molecules may have utility as a therapeutic agent for a particular protein.

For example, the chemical reactions and building blocks are chosen to create a library having structural groups commonly present in kinase inhibitors. In any of these instances, the tags encode the chemical history of the library member and, in each case, a collection of chemical possibilities may be represented by any particular tag combination.

5 In one embodiment, the library of chemical entities, or a portion thereof, is contacted with a biological target under conditions suitable for at least one member of the library to bind to the target, followed by removal of library members that do not bind to the target, and analyzing the one or more oligonucleotide tags associated with them. This method can optionally include amplifying the tags by methods known in the art. Exemplary biological targets include enzymes (e.g., kinases, phosphatases, 10 methylases, demethylases, proteases, and DNA repair enzymes), proteins involved in protein:protein interactions (e.g., ligands for receptors), receptor targets (e.g., GPCRs and RTKs), ion channels, bacteria, viruses, parasites, DNA, RNA, prions, and carbohydrates.

In another embodiment, the chemical entities that bind to a target are not subjected to amplification but are analyzed directly. Exemplary methods of analysis include microarray analysis, 15 including evanescent resonance photonic crystal analysis; bead-based methods for deconvoluting tags (e.g., by using his-tags); label-free photonic crystal biosensor analysis (e.g., a BIND® Reader from SRU Biosystems, Inc., Woburn, MA); or hybridization-based approaches (e.g. by using arrays of immobilized oligonucleotides complementary to sequences present in the library of tags).

In addition, chemically co-reactive pairs (or functional groups) can be readily included in solid- 20 phase oligonucleotide synthesis schemes and will support the efficient chemical ligation of oligonucleotides. In addition, the resultant ligated oligonucleotides can act as templates for template-dependent polymerization with one or more polymerases. Accordingly, any of the binding steps described herein for tagging encoded libraries can be modified to include one or more of enzymatic ligation and/or chemical ligation techniques. Exemplary ligation techniques include enzyme ligation, 25 such as use of one or more RNA ligases and/or DNA ligases; and chemical ligation, such as use of chemically co-reactive pairs (e.g., a pair including optionally substituted alkynyl and azido functional groups).

Furthermore, one or more libraries can be combined in a split-and-mix step. In order to permit mixing of two or more libraries, the library member may contain one or more library-identifying 30 sequences, such as in a library-identifying tag, in a ligated building block tag, or as part of the headpiece sequence, as described herein.

Methods having reduced mass

Much of the motivation for single-stranded encoding strategies arises from the reduced mass of a 35 single-stranded tag when compared to a double-stranded tag. Reduced mass potentially confers several benefits including increased solubility, decreased cost, increased reactivity, increased target accessibility, decreased hydrodynamic radius, increased accuracy of analytical assessments, etc. In addition to using a

single-stranded tagging methodology, further reductions in mass can be achieved by including the use of one or more of the following: one or more tags having a reduced length, constant mass tag sets, an encoding headpiece, one or more members of a library lacking a primer binding region and/or a constant region, one or more members of a library having a reduced constant region, or any other methodologies described herein.

To minimize the mass of the members in the library, the length of one or more building block tags can be reduced, such as to a length that is as short as possible to encode each split size. In particular, the tags can be less than 20 nucleotides (e.g., less than 19 nucleotides, less than 18 nucleotides, less than 17 nucleotides, less than 16 nucleotides, less than 15 nucleotides, less than 14 nucleotides, less than 13 nucleotides, less than 12 nucleotides, less than 11 nucleotides, less than 10 nucleotides, less than 9 nucleotides, less than 8 nucleotides, or less than 7 nucleotides). As described below in the Examples, shorter tags (e.g., about 10 nucleotides or shorter) can be used for tag ligation.

Constant mass strategies can also be used, which could aid in analysis during library synthesis. In addition, constant mass tag sets could permit the recognition of all single error occurrences (e.g., errors arising from misreading a sequence or from chemical or enzymatic ligation of a tag) and most multiple error occurrences. The relationship between the length of a constant mass single-stranded tag set and encoding ability (e.g., minimum lengths to support specific building block split sizes or library identities, etc.) is outlined below in Table 1. Accordingly, use of constant mass tag sets could be used to provide beneficial encoding ability, while maintaining error recognition during library formation.

Table 1

Length	Base #1	Base #2	Base #3	Base #4	Combinations
1	1	0	0	0	1
2	1	1	0	0	2
3	1	1	1	0	6
4	1	1	1	1	24
5	2	1	1	1	60
6	2	2	1	1	180
7	2	2	2	1	630
8	2	2	2	2	2,520
9	3	2	2	2	7,560
10	3	3	2	2	25,200
11	3	3	3	2	92,400
12	3	3	3	3	369,600
13	4	3	3	3	1,201,200
14	4	4	3	3	4,204,200
15	4	4	4	3	15,765,750
16	4	4	4	4	63,063,000
17	5	4	4	4	214,414,200
18	5	5	4	4	771,891,120
19	5	5	5	4	2,933,186,256
20	5	5	5	5	11,732,745,024

To minimize mass in the library, the headpiece can be used not only to link the chemical moiety and a tag but to also encode for the identity of a particular library or for a particular step. For example, the headpiece can encode information, e.g., a plurality of headpieces that encode the first split(s) or the identity of the library, such as by using a particular sequence related to a specific library.

In addition, primer binding (e.g., constant) regions from the library of DNA-encoded chemical entities can be excluded during the selection step(s). Then, these regions can be added after selection by, e.g., single-stranded ligation. One exemplary strategy would include providing a chemical entity at the 5'-terminus of an encoding oligonucleotide, selecting a particular chemical entity based on any useful particular characteristic or function, and ligating a tailpiece oligonucleotide to the 3'-terminus of the encoding oligonucleotide that includes a primer binding sequence and may optionally contain one or more tags, e.g. a "use" tag, an "origin" tag, etc., as described herein. This primer binding sequence could then be used to initiate template-dependent polymerization to generate cDNA (or cRNA) that is complementary to the selected library member. The cDNA or cRNA would then be ligated at its 3'-terminus to an oligonucleotide that contains a primer binding sequence and, now that the encoding information is flanked on both sides by primer binding sequences, the oligonucleotide may be sequenced and/or amplified using established approaches, such as any described herein.

Mass may further be minimized by omitting or reducing the size of one or more constant sequences that separate encoding tags. Single-stranded ligation requires no complementary relationship between the ends to be ligated or between these ends and a splint. Therefore, no fixed sequence is required to support enzymatic ligation. Short fixed regions between tags may be useful for informatic parsing of tags or other *in silico* deconvolution processes.

Oligonucleotide tags

The oligonucleotide tags described herein (e.g., a building block tag or a portion of a headpiece) can be used to encode any useful information, such as a molecule, a portion of a chemical entity, the addition of a component (e.g., a scaffold or a building block), a headpiece in the library, the identity of the library, the use of one or more library members (e.g., use of the members in an aliquot of a library), and/or the origin of a library member (e.g., by use of an origin sequence).

Any sequence in an oligonucleotide can be used to encode any information. Thus, one oligonucleotide sequence can serve more than one purpose, such as to encode two or more types of information or to provide a starting oligonucleotide that also encodes for one or more types of information. For example, the first building block tag can encode for the addition of a first building block, as well as for the identification of the library. In another example, a headpiece can be used to provide a starting oligonucleotide that operatively links a chemical entity to a building block tag, where the headpiece additionally includes a sequence that encodes for the identity of the library (i.e., the library-identifying sequence). Accordingly, any of the information described herein can be encoded in separate oligonucleotide tags or can be combined and encoded in the same oligonucleotide sequence (e.g., an oligonucleotide tag, such as a building block tag, or a headpiece).

A building block sequence encodes for the identity of a building block and/or the type of binding reaction conducted with a building block. This building block sequence is included in a building block tag, where the tag can optionally include one or more types of sequence described below (e.g., a library-identifying sequence, a use sequence, and/or an origin sequence).

A library-identifying sequence encodes for the identity of a particular library. In order to permit mixing of two or more libraries, a library member may contain one or more library-identifying sequences, such as in a library-identifying tag (i.e., an oligonucleotide including a library-identifying sequence), in a ligated building block tag, in a part of the headpiece sequence, or in a tailpiece sequence. These library-identifying sequences can be used to deduce encoding relationships, where the sequence of the tag is translated and correlated with chemical (synthesis) history information. Accordingly, these library-identifying sequences permit the mixing of two or more libraries together for selection, amplification, purification, sequencing, etc.

A use sequence encodes the history (i.e., use) of one or more library members in an individual aliquot of a library. For example, separate aliquots may be treated with different reaction conditions, building blocks, and/or selection steps. In particular, this sequence may be used to identify such aliquots

and deduce their history (use) and thereby permit the mixing together of aliquots of the same library with different histories (uses) (e.g., distinct selection experiments) for the purposes of the mixing together of samples together for selection, amplification, purification, sequencing, etc. These use sequences can be included in a headpiece, a tailpiece, a building block tag, a use tag (i.e., an oligonucleotide including a use sequence), or any other tag described herein (e.g., a library-identifying tag or an origin tag).

An origin sequence is a degenerate (random) oligonucleotide sequence of any useful length (e.g., about six oligonucleotides) that encodes for the origin of the library member. This sequence serves to stochastically subdivide library members that are otherwise identical in all respects into entities distinguishable by sequence information, such that observations of amplification products derived from unique progenitor templates (e.g., selected library members) can be distinguished from observations of multiple amplification products derived from the same progenitor template (e.g., a selected library member). For example, after library formation and prior to the selection step, each library member can include a different origin sequence, such as in an origin tag. After selection, selected library members can be amplified to produce amplification products, and the portion of the library member expected to include the origin sequence (e.g., in the origin tag) can be observed and compared with the origin sequence in each of the other library members. As the origin sequences are degenerate, each amplification product of each library member should have a different origin sequence. However, an observation of the same origin sequence in the amplification product could indicate a source of error, such as an amplification error or a cyclization error in the sequence that produces repeated sequences, and the starting point or source of these errors can be traced by observing the origin sequence at each step (e.g., at each selection step or amplification step) of using the library. These origin sequences can be included in a headpiece, a tailpiece, a building block tag, an origin tag (i.e., an oligonucleotide including an origin sequence), or any other tag described herein (e.g., a library-identifying tag or a use tag).

Any of the types of sequences described herein can be included in the headpiece. For example, the headpiece can include one or more of a building block sequence, a library-identifying sequence, a use sequence, or an origin sequence.

Any of these sequences described herein can be included in a tailpiece. For example, the tailpiece can include one or more of a library-identifying sequence, a use sequence, or an origin sequence.

These sequences can include any modification described herein for oligonucleotides, such as one or more modifications that promote solubility in organic solvents (e.g., any described herein, such as for the headpiece), that provide an analog of the natural phosphodiester linkage (e.g., a phosphorothioate analog), or that provide one or more non-natural oligonucleotides (e.g., 2'-substituted nucleotides, such as 2'-O-methylated nucleotides and 2'-fluoro nucleotides, or any described herein).

These sequences can include any characteristics described herein for oligonucleotides. For example, these sequences can be included in tag that is less than 20 nucleotides (e.g., as described herein). In other examples, the tags including one or more of these sequences have about the same mass (e.g., each tag has a mass that is about +/- 10% from the average mass between two or more tags); lack a

primer binding (e.g., constant) region; lack a constant region; or have a constant region of reduced length (e.g., a length less than 30 nucleotides, less than 25 nucleotides, less than 20 nucleotides, less than 19 nucleotides, less than 18 nucleotides, less than 17 nucleotides, less than 16 nucleotides, less than 15 nucleotides, less than 14 nucleotides, less than 13 nucleotides, less than 12 nucleotides, less than 11 nucleotides, less than 10 nucleotides, less than 9 nucleotides, less than 8 nucleotides, or less than 7 nucleotides).

Sequencing strategies for libraries and oligonucleotides of this length may optionally include concatenation or catenation strategies to increase read fidelity or sequencing depth, respectively. In particular, the selection of encoded libraries that lack primer binding regions has been described in the literature for SELEX, such as described in Jarosch et al., *Nucleic Acids Res.* 34: e86 (2006), which is incorporated herein by reference. For example, a library member can be modified (e.g., after a selection step) to include a first adapter sequence on the 5'-terminus of the complex and a second adapter sequence on the 3'-terminus of the complex, where the first sequence is substantially complementary to the second sequence and result in forming a duplex. To further improve yield, two fixed dangling nucleotides (e.g., CC) are added to the 5'-terminus. In particular embodiments, the first adapter sequence is 5'-GTGCTGC-3' (SEQ ID NO: 1), and the second adapter sequence is 5'-GCAGCACCC-3' (SEQ ID NO: 2).

Headpiece

In the library, the headpiece operatively links each chemical entity to its encoding oligonucleotide tag. Generally, the headpiece is a starting oligonucleotide having two functional groups that can be further derivatized, where the first functional group operatively links the chemical entity (or a component thereof) to the headpiece and the second functional group operatively links one or more tags to the headpiece. A linker can optionally be used as a spacer between the headpiece and the chemical entity.

The functional groups of the headpiece can be used to form a covalent bond with a component of the chemical entity and another covalent bond with a tag. The component can be any part of the small molecule, such as a scaffold having diversity nodes or a building block. Alternatively, the headpiece can be derivatized to provide a linker (i.e., a spacer separating the headpiece from the small molecule to be formed in the library) terminating in a functional group (e.g., a hydroxyl, amine, carboxyl, sulfhydryl, alkynyl, azido, or phosphate group), which is used to form the covalent linkage with a component of the chemical entity. The linker can be attached to the 5'-terminus, at one of the internal positions, or to the 3'-terminus of the headpiece. When the linker is attached to one of the internal positions, the linker can be operatively linked to a derivatized base (e.g., the C5 position of uridine) or placed internally within the oligonucleotide using standard techniques known in the art. Exemplary linkers are described herein.

The headpiece can have any useful structure. The headpiece can be, e.g., 1 to 100 nucleotides in length, preferably 5 to 20 nucleotides in length, and most preferably 5 to 15 nucleotides in length. The

headpiece can be single-stranded or double-stranded and can consist of natural or modified nucleotides, as described herein. Particular exemplary embodiments of the headpiece are described in Figures 4A-4D. For example, the chemical moiety can be operatively linked to the 3'-terminus (Figure 4A) or 5'-terminus (Figure 4B) of the headpiece. In particular embodiments, the headpiece includes a hairpin structure
5 formed by complementary bases within the sequence. For example, the chemical moiety can be operatively linked to the internal position (Figure 4C), the 3'-terminus (Figure 4D), or the 5'-terminus of the headpiece.

Generally, the headpiece includes a non-complementary sequence on the 5'- or 3'- terminus that allows for binding an oligonucleotide tag by polymerization, enzymatic ligation, or chemical reaction. In
10 Figure 4E, the exemplary headpiece allows for ligation of oligonucleotide tags (labeled 1-4), and the method includes purification and phosphorylation steps. After the addition of tag 4, an additional adapter sequence can be added to the 5'-terminus of tag 4. Exemplary adapter sequences include a primer binding sequence or a sequence having a label (e.g., biotin). In cases where many building blocks and corresponding tags are used (e.g., 100 tags), a mix-and-split strategy may be employed during the
15 oligonucleotide synthesis step to create the necessary number of tags. Such mix-and-split strategies for DNA synthesis are known in the art. The resultant library members can be amplified by PCR following selection for binding entities versus a target(s) of interest.

The headpiece or the complex can optionally include one or more primer binding sequences. For example, the headpiece has a sequence in the loop region of the hairpin that serves as a primer binding
20 region for amplification, where the primer binding region has a higher melting temperature for its complementary primer (e.g., which can include flanking identifier regions) than for a sequence in the headpiece. In other embodiments, the complex includes two primer binding sequences (e.g., to enable a PCR reaction) on either side of one or more tags that encode one or more building blocks. Alternatively, the headpiece may contain one primer binding sequence on the 5'- or 3'-terminus. In other embodiments,
25 the headpiece is a hairpin, and the loop region forms a primer binding site or the primer binding site is introduced through hybridization of an oligonucleotide to the headpiece on the 3' side of the loop. A primer oligonucleotide, containing a region homologous to the 3'-terminus of the headpiece and carrying a primer binding region on its 5'-terminus (e.g., to enable a PCR reaction) may be hybridized to the headpiece and may contain a tag that encodes a building block or the addition of a building block. The
30 primer oligonucleotide may contain additional information, such as a region of randomized nucleotides, e.g., 2 to 16 nucleotides in length, which is included for bioinformatics analysis.

The headpiece can optionally include a hairpin structure, where this structure can be achieved by any useful method. For example, the headpiece can include complementary bases that form
intermolecular base pairing partners, such as by Watson-Crick DNA base pairing (e.g., adenine-thymine
35 and guanine-cytosine) and/or by wobble base pairing (e.g., guanine-uracil, inosine-uracil, inosine-adenine, and inosine-cytosine). In another example, the headpiece can include modified or substituted nucleotides that can form higher affinity duplex formations compared to unmodified nucleotides, such

modified or substituted nucleotides being known in the art. In yet another example, the headpiece includes one or more crosslinked bases to form the hairpin structure. For example, bases within a single strand or bases in different double strands can be crosslinked, e.g., by using psoralen.

The headpiece or complex can optionally include one or more labels that allow for detection. For example, the headpiece, one or more oligonucleotide tags, and/or one or more primer sequences can include an isotope, a radioimaging agent, a marker, a tracer, a fluorescent label (e.g., rhodamine or fluorescein), a chemiluminescent label, a quantum dot, and a reporter molecule (e.g., biotin or a his-tag).

In other embodiments, the headpiece or tag may be modified to support solubility in semi-, reduced-, or non-aqueous (e.g., organic) conditions. Nucleotide bases of the headpiece or tag can be rendered more hydrophobic by modifying, for example, the C5 positions of T or C bases with aliphatic chains without significantly disrupting their ability to hydrogen bond to their complementary bases. Exemplary modified or substituted nucleotides are 5'-dimethoxytrityl-N4-diisobutylaminomethylidene-5-(1-propynyl)-2'-deoxycytidine, 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite; 5'-dimethoxytrityl-5-(1-propynyl)-2'-deoxyuridine, 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite; 5'-dimethoxytrityl-5-fluoro-2'-deoxyuridine, 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite; and 5'-dimethoxytrityl-5-(pyren-1-yl-ethynyl)-2'-deoxyuridine, or 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite.

In addition, the headpiece oligonucleotide can be interspersed with modifications that promote solubility in organic solvents. For example, azobenzene phosphoramidite can introduce a hydrophobic moiety into the headpiece design. Such insertions of hydrophobic amidites into the headpiece can occur anywhere in the molecule. However, the insertion cannot interfere with subsequent tagging using additional DNA tags during the library synthesis or ensuing PCR once a selection is complete or microarray analysis, if used for tag deconvolution. Such additions to the headpiece design described herein would render the headpiece soluble in, for example, 15%, 25%, 30%, 50%, 75%, 90 %, 95%, 98%, 99%, or 100% organic solvent. Thus, addition of hydrophobic residues into the headpiece design allows for improved solubility in semi- or non-aqueous (e.g., organic) conditions, while rendering the headpiece competent for oligonucleotide tagging. Furthermore, DNA tags that are subsequently introduced into the library can also be modified at the C5 position of T or C bases such that they also render the library more hydrophobic and soluble in organic solvents for subsequent steps of library synthesis.

In particular embodiments, the headpiece and the first building block tag can be the same entity, i.e., a plurality of headpiece-tag entities can be constructed that all share common parts (e.g., a primer binding region) and all differ in another part (e.g., encoding region). These may be utilized in the "split" step and pooled after the event they are encoding has occurred.

In particular embodiments, the headpiece can encode information, e.g., by including a sequence that encodes the first split(s) step or a sequence that encodes the identity of the library, such as by using a particular sequence related to a specific library.

Enzymatic ligation and chemical ligation techniques

Various ligation techniques can be used to add scaffolds, building blocks, linkers, building block tags, and/or the headpiece to produce a complex. Accordingly, any of the binding steps described herein can include any useful ligation techniques, such as enzyme ligation and/or chemical ligation. These binding steps can include the addition of one or more building block tags to the headpiece or complex; the addition of a linker to the headpiece; and the addition of one or more scaffolds or building blocks to the headpiece or complex. In particular embodiments, the ligation techniques used for any oligonucleotide provide a resultant product that can be transcribed and/or reverse transcribed to allow for decoding of the library or for template-dependent polymerization with one or more DNA or RNA polymerases.

Generally, enzyme ligation produces an oligonucleotide having a native phosphodiester bond that can be transcribed and/or reverse transcribed. Exemplary methods of enzyme ligation are provided herein and include the use of one or more RNA or DNA ligases, such as T4 RNA ligase, T4 DNA ligase, CircLigaseTM ssDNA ligase, CircLigaseTM II ssDNA ligase, and ThermoPhageTM ssDNA ligase (Prokaryote Ltd., Reykjavik, Iceland).

Chemical ligation can also be used to produce oligonucleotides capable of being transcribed or reverse transcribed. One benefit of chemical ligation is that solid phase synthesis of such oligonucleotides can be optimized to support efficient ligation yield. However, the efficacy of a chemical ligation technique to provide oligonucleotides capable of being transcribed or reverse transcribed may need to be tested. This efficacy can be tested by any useful method, such as liquid chromatography-mass spectrometry, RT-PCR analysis, and/or PCR analysis. Examples of these methods are provided in Example 5.

In particular embodiments, chemical ligation includes the use of one or more chemically co-reactive pairs to provide a spacer that can be transcribed or reverse transcribed. In particular, reactions suitable for chemically co-reactive pairs are preferred candidates for the cyclization process (Kolb et al., *Angew. Chem. Int. Ed.*, 40:2004-2021 (2001); Van der Eycken et al., *QSAR Comb. Sci.*, 26:1115-1326 (2007)). Exemplary chemically co-reactive pairs are a pair including an optionally substituted alkynyl group and an optionally substituted azido group to form a triazole spacer via a Huisgen 1,3-dipolar cycloaddition reaction; an optionally substituted diene having a 4π electron system (e.g., an optionally substituted 1,3-unsaturated compound, such as optionally substituted 1,3-butadiene, 1-methoxy-3-trimethylsilyloxy-1,3-butadiene, cyclopentadiene, cyclohexadiene, or furan) and an optionally substituted dienophile or an optionally substituted heterodienophile having a 2π electron system (e.g., an optionally substituted alkenyl group or an optionally substituted alkynyl group) to form a cycloalkenyl spacer via a Diels-Alder reaction; a nucleophile (e.g., an optionally substituted amine or an optionally substituted thiol) with a strained heterocyclic electrophile (e.g., optionally substituted epoxide, aziridine, aziridinium ion, or episulfonium ion) to form a heteroalkyl spacer via a ring opening reaction; a phosphorothioate group with an iodo group, such as in a splinted ligation of an oligonucleotide containing 5'-iodo dT with

a 3'-phosphorothioate oligonucleotide; and an aldehyde group and an amino group, such as a reaction of a 3'-aldehyde-modified oligonucleotide, which can optionally be obtained by oxidizing a commercially available 3'-glyceryl-modified oligonucleotide, with 5'-amino oligonucleotide (i.e., in a reductive amination reaction) or a 5'-hydrazido oligonucleotide.

5 In other embodiments, chemical ligation includes introducing an analog of the phosphodiester bond, e.g., for post-selection PCR analysis and sequencing. Exemplary analogs of a phosphodiester include a phosphorothioate linkage (e.g., as introduced by use of a phosphorothioate group and a leaving group, such as an iodo group), a phosphoramidate linkage, or a phosphorodithioate linkage (e.g., as introduced by use of a phosphorodithioate group and a leaving group, such as an iodo group).

10 **Reaction conditions to promote enzymatic ligation or chemical ligation**

The invention also features one or more reaction conditions that promote enzymatic or chemical ligation between the headpiece and a tag or between two tags. These reaction conditions include using modified nucleotides within the tag, as described herein; using donor tags and acceptor tags having
15 different lengths and varying the concentration of the tags; using different types of ligases, as well as combinations thereof (e.g., CircLigaseTM DNA ligase and/or T4 RNA ligase), and varying their concentration; using poly ethylene glycols (PEGs) having different molecular weights and varying their concentration; use of non-PEG crowding agents (e.g., betaine or bovine serum albumin); varying the temperature and duration for ligation; varying the concentration of various agents, including ATP,
20 Co(NH₃)₆Cl₃, and yeast inorganic pyrophosphate; using enzymatically or chemically phosphorylated oligonucleotide tags; using 3'-protected tags; and using preadenylated tags. These reaction conditions also include chemical ligations.

The headpiece and/or tags can include one or more modified or substituted nucleotides. In preferred embodiments, the headpiece and/or tags include one or more modified or substituted
25 nucleotides that promote enzymatic ligation, such as 2'-O-methyl nucleotides (e.g., 2'-O-methyl guanine or 2'-O-methyl uracil), 2'-fluoro nucleotides, or any other modified nucleotides that are utilized as a substrate for ligation. Alternatively, the headpiece and/or tags are modified to include one or more chemically reactive groups to support chemical ligation (e.g. an optionally substituted alkynyl group and an optionally substituted azido group). Optionally, the tag oligonucleotides are functionalized at both
30 termini with chemically reactive groups, and, optionally, one of these termini is protected, such that the groups may be addressed independently and side-reactions may be reduced (e.g., reduced polymerization side-reactions).

Enzymatic ligation can include one or more ligases. Exemplary ligases include CircLigaseTM ssDNA ligase (EPICENTRE Biotechnologies, Madison, WI), CircLigaseTM II ssDNA ligase (also from
35 EPICENTRE Biotechnologies), ThermoPhageTM ssDNA ligase (Prokaryote Ltd., Reykjavik, Iceland), T4 RNA ligase, and T4 DNA ligase. In preferred embodiments, ligation includes the use of an RNA ligase

or a combination of an RNA ligase and a DNA ligase. Ligation can further include one or more soluble multivalent cations, such as $\text{Co}(\text{NH}_3)_6\text{Cl}_3$, in combination with one or more ligases.

Before or after the ligation step, the complex can be purified for three reasons. First, the complex can be purified to remove unreacted headpiece or tags that may result in cross-reactions and introduce “noise” into the encoding process. Second, the complex can be purified to remove any reagents or unreacted starting material that can inhibit or lower the ligation activity of a ligase. For example, phosphate may result in lowered ligation activity. Third, entities that are introduced into a chemical or ligation step may need to be removed to enable the subsequent chemical or ligation step. Methods of purifying the complex are described herein.

Enzymatic and chemical ligation can include poly ethylene glycol having an average molecular weight of more than 300 Daltons (e.g., more than 600 Daltons, 3,000 Daltons, 4,000 Daltons, or 4,500 Daltons). In particular embodiments, the poly ethylene glycol has an average molecular weight from about 3,000 Daltons to 9,000 Daltons (e.g., from 3,000 Daltons to 8,000 Daltons, from 3,000 Daltons to 7,000 Daltons, from 3,000 Daltons to 6,000 Daltons, and from 3,000 Daltons to 5,000 Daltons). In preferred embodiments, the poly ethylene glycol has an average molecular weight from about 3,000 Daltons to about 6,000 Daltons (e.g., from 3,300 Daltons to 4,500 Daltons, from 3,300 Daltons to 5,000 Daltons, from 3,300 Daltons to 5,500 Daltons, from 3,300 Daltons to 6,000 Daltons, from 3,500 Daltons to 4,500 Daltons, from 3,500 Daltons to 5,000 Daltons, from 3,500 Daltons to 5,500 Daltons, and from 3,500 Daltons to 6,000 Daltons, such as 4,600 Daltons). Poly ethylene glycol can be present in any useful amount, such as from about 25% (w/v) to about 35% (w/v), such as 30% (w/v).

In a preferred embodiment of this invention, the building block tags are installed by ligation of a single-stranded oligonucleotide to a single-stranded oligonucleotide using the ligation protocol outlined below:

Headpiece:	25 μM (5' terminus: 5'-monophospho/2'-OMe G, intervening nucleotides: 2'-deoxy, and 3' terminus: 2'-blocked/3'-blocked)
Building Block Tag:	25 μM (5'-terminus: 2'-OMe/5'-OH G, intervening nucleotides: 2'-deoxy, and 3'-terminus: 3'-OH/2'-OMe)
$\text{Co}(\text{NH}_3)_6\text{Cl}_3$:	1 mM
PEG 4600:	30% (w/v)
T4 RNA Ligase (Promega):	1.5 units/ μl
Yeast Inorganic Pyrophosphatase:	0.0025 units/ μl
Tris:	50 mM
MgCl_2 :	10 mM
ATP:	1 mM
pH:	7.5
Water:	Balance

In further embodiments, the protocol includes incubation at 37°C for 20 hours. For the purposes of actual library construction, higher concentration of headpiece, tags, and/or ligase may be used, and such modifications to these concentrations would be apparent to those skilled in the art.

5 Methods for encoding chemical entities within a library

The methods of the invention can be used to synthesize a library having a diverse number of chemical entities that are encoded by oligonucleotide tags. Examples of building blocks and encoding DNA tags are found in U.S. Patent Application Publication No. 2007/0224607, hereby incorporated by reference.

10 Each chemical entity is formed from one or more building blocks and optionally a scaffold. The scaffold serves to provide one or more diversity nodes in a particular geometry (e.g., a triazine to provide three nodes spatially arranged around a heteroaryl ring or a linear geometry).

The building blocks and their encoding tags can be added directly or indirectly (e.g., via a linker) to the headpiece to form a complex. When the headpiece includes a linker, the building block or scaffold is added to the end of the linker. When the linker is absent, the building block can be added directly to the headpiece or the building block itself can include a linker that reacts with a functional group of the headpiece. Exemplary linkers and headpieces are described herein.

The scaffold can be added in any useful way. For example, the scaffold can be added to the end of the linker or the headpiece, and successive building blocks can be added to the available diversity nodes of the scaffold. In another example, building block A_n is first added to the linker or the headpiece, and then the diversity node of scaffold S is reacted with a functional group in building block A_n . Oligonucleotide tags encoding a particular scaffold can optionally be added to the headpiece or the complex. For example, S_n is added to the complex in n reaction vessels, where n is an integer more than one, and tag S_n (i.e., tag $S_1, S_2, \dots S_{n-1}, S_n$) is bound to the functional group of the complex.

25 Building blocks can be added in multiple, synthetic steps. For example, an aliquot of the headpiece, optionally having an attached linker, is separated into n reaction vessels, where n is an integer of two or greater. In the first step, building block A_n is added to each n reaction vessel (i.e., building block $A_1, A_2, \dots A_{n-1}, A_n$ is added to reaction vessel 1, 2, $\dots$ $n-1, n$), where n is an integer and each building block A_n is unique. In the second step, scaffold S is added to each reaction vessel to form an A_n - S complex. Optionally, scaffold S_n can be added to each reaction vessel to form an A_n - S_n complex, where n is an integer of more than two, and each scaffold S_n can be unique. In the third step, building block B_n is added to each n reaction vessel containing the A_n - S complex (i.e., building block $B_1, B_2, \dots B_{n-1}, B_n$ is added to reaction vessel 1, 2, $\dots$ $n-1, n$ containing the A_1 - S, A_2 - $S, \dots A_{n-1}$ - S, A_n - S complex), where each building block B_n is unique. In further steps, building block C_n can be added to each n reaction vessel containing the B_n - A_n - S complex (i.e., building block $C_1, C_2, \dots C_{n-1}, C_n$ is added to reaction vessel 1, 2, $\dots$ $n-1, n$ containing the B_1 - A_1 - $S \dots B_n$ - A_n - S complex), where each building block C_n is unique. The

resulting library will have n^3 number of complexes having n^3 tags. In this manner, additional synthetic steps can be used to bind additional building blocks to further diversify the library.

After forming the library, the resultant complexes can optionally be purified and subjected to a polymerization or ligation reaction using one or more primers. This general strategy can be expanded to include additional diversity nodes and building blocks (e.g., D, E, F, etc.). For example, the first diversity node is reacted with building blocks and/or S and encoded by an oligonucleotide tag. Then, additional building blocks are reacted with the resultant complex, and the subsequent diversity node is derivatized by additional building blocks, which is encoded by the primer used for the polymerization or ligation reaction

To form an encoded library, oligonucleotide tags are added to the complex after or before each synthetic step. For example, before or after the addition of building block A_n to each reaction vessel, tag A_n is bound to the functional group of the headpiece (i.e., tag $A_1, A_2, \dots A_{n-1}, A_n$ is added to reaction vessel 1, 2, $\dots$ $n-1, n$ containing the headpiece). Each tag A_n has a distinct sequence that correlates with each unique building block A_n , and determining the sequence of tag A_n provides the chemical structure of building block A_n . In this manner, additional tags are used to encode for additional building blocks or additional scaffolds.

Furthermore, the last tag added to the complex can either include a primer sequence or provide a functional group to allow for binding (e.g., by ligation) of a primer sequence. The primer sequence can be used for amplifying and/or sequencing the oligonucleotide tags of the complex. Exemplary methods for amplifying and for sequencing include polymerase chain reaction (PCR), linear chain amplification (LCR), rolling circle amplification (RCA), or any other method known in the art to amplify or determine nucleic acid sequences.

Using these methods, large libraries can be formed having a large number of encoded chemical entities. For example, a headpiece is reacted with a linker and building block A_n , which includes 1,000 different variants (i.e., $n = 1,000$). For each building block A_n , a DNA tag A_n is ligated or primer extended to the headpiece. These reactions may be performed in a 1,000-well plate or 10 x 100 well plates. All reactions may be pooled, optionally purified, and split into a second set of plates. Next, the same procedure may be performed with building block B_n , which also include 1,000 different variants. A DNA tag B_n may be ligated to the A_n -headpiece complex, and all reactions may be pooled. The resultant library includes 1,000 x 1,000 combinations of A_n x B_n (i.e., 1,000,000 compounds) tagged by 1,000,000 different combinations of tags. The same approach may be extended to add building blocks C_n, D_n, E_n , etc. The generated library may then be used to identify compounds that bind to the target. The structure of the chemical entities that bind to the library can optionally be assessed by PCR and sequencing of the DNA tags to identify the compounds that were enriched.

This method can be modified to avoid tagging after the addition of each building block or to avoid pooling (or mixing). For example, the method can be modified by adding building block A_n to n reaction vessels, where n is an integer of more than one, and adding the identical building block B_1 to

each reaction well. Here, B_1 is identical for each chemical entity, and, therefore, an oligonucleotide tag encoding this building block is not needed. After adding a building block, the complexes may be pooled or not pooled. For example, the library is not pooled following the final step of building block addition, and the pools are screened individually to identify compound(s) that bind to a target. To avoid pooling all of the reactions after synthesis, a BIND[®] Reader (from SRU Biosystems, Inc.), for example, may be used to monitor binding on a sensor surface in high throughput format (e.g., 384 well plates and 1,536 well plates). For example, building block A_n may be encoded with DNA tag A_n , and building block B_n may be encoded by its position within the well plate. Candidate compounds can then be identified by using a binding assay (e.g., using a BIND[®] Biosensor, also available by SRU Biosystems, Inc., or using an ELISA assay) and by analyzing the A_n tags by sequencing, microarray analysis and/or restriction digest analysis. This analysis allows for the identification of combinations of building blocks A_n and B_n that produce the desired molecules.

The method of amplifying can optionally include forming a water-in-oil emulsion to create a plurality of aqueous microreactors. The reaction conditions (e.g., concentration of complex and size of microreactors) can be adjusted to provide, on average, a microreactor having at least one member of a library of compounds. Each microreactor can also contain the target, a single bead capable of binding to a complex or a portion of the complex (e.g., one or more tags) and/or binding the target, and an amplification reaction solution having one or more necessary reagents to perform nucleic acid amplification. After amplifying the tag in the microreactors, the amplified copies of the tag will bind to the beads in the microreactors, and the coated beads can be identified by any useful method.

Once the building blocks from the first library that bind to the target of interest have been identified, a second library may be prepared in an iterative fashion. For example, one or two additional nodes of diversity can be added, and the second library is created and sampled, as described herein. This process can be repeated as many times as necessary to create molecules with desired molecular and pharmaceutical properties.

Various ligation techniques can be used to add the scaffold, building blocks, linkers, and building block tags. Accordingly, any of the binding steps described herein can include any useful ligation technique or techniques. Exemplary ligation techniques include enzymatic ligation, such as use of one of more RNA ligases and/or DNA ligases, as described herein; and chemical ligation, such as use of chemically co-reactive pairs, as described herein.

Scaffold and building blocks

The scaffold S can be a single atom or a molecular scaffold. Exemplary single atom scaffolds include a carbon atom, a boron atom, a nitrogen atom, or a phosphorus atom, etc. Exemplary polyatomic scaffolds include a cycloalkyl group, a cycloalkenyl group, a heterocycloalkyl group, a heterocycloalkenyl group, an aryl group, or a heteroaryl group. Particular embodiments of a heteroaryl scaffold include a triazine, such as 1,3,5-triazine, 1,2,3-triazine, or 1,2,4-triazine; a pyrimidine; a

pyrazine; a pyridazine; a furan; a pyrrole; a pyrroline; a pyrrolidine; an oxazole; a pyrazole; an isoxazole; a pyran; a pyridine; an indole; an indazole; or a purine.

The scaffold S can be operatively linked to the tag by any useful method. In one example, S is a triazine that is linked directly to the headpiece. To obtain this exemplary scaffold, trichlorotriazine (i.e., a chlorinated precursor of triazine having three chlorines) is reacted with a nucleophilic group of the headpiece. Using this method, S has three positions having chlorine that are available for substitution, where two positions are available diversity nodes and one position is attached to the headpiece. Next, building block A_n is added to a diversity node of the scaffold, and tag A_n encoding for building block A_n ("tag A_n") is ligated to the headpiece, where these two steps can be performed in any order. Then, building block B_n is added to the remaining diversity node, and tag B_n encoding for building block B_n is ligated to the end of tag A_n. In another example, S is a triazine that is operatively linked to the linker of a tag, where trichlorotriazine is reacted with a nucleophilic group (e.g., an amino group) of a PEG, aliphatic, or aromatic linker of a tag. Building blocks and associated tags can be added, as described above.

In yet another example, S is a triazine that is operatively linked to building block A_n. To obtain this scaffold, building block A_n having two diversity nodes (e.g., an electrophilic group and a nucleophilic group, such as an Fmoc-amino acid) is reacted with the nucleophilic group of a linker (e.g., the terminal group of a PEG, aliphatic, or aromatic linker, which is attached to a headpiece). Then, trichlorotriazine is reacted with a nucleophilic group of building block A_n. Using this method, all three chlorine positions of S are used as diversity nodes for building blocks. As described herein, additional building blocks and tags can be added, and additional scaffolds S_n can be added.

Exemplary building block A_n's include, e.g., amino acids (e.g., alpha-, beta-, gamma-, delta-, and epsilon- amino acids, as well as derivatives of natural and unnatural amino acids), chemically co-reactive reactants (e.g., azide or alkyne chains) with an amine, or a thiol reactant, or combinations thereof. The choice of building block A_n depends on, for example, the nature of the reactive group used in the linker, the nature of a scaffold moiety, and the solvent used for the chemical synthesis.

Exemplary building block B_n's and C_n's include any useful structural unit of a chemical entity, such as optionally substituted aromatic groups (e.g., optionally substituted phenyl or benzyl), optionally substituted heterocyclyl groups (e.g., optionally substituted quinoliny, isoquinoliny, indolyl, isoindolyl, azaindolyl, benzimidazolyl, azabenzimidazolyl, benzisoxazolyl, pyridinyl, piperidyl, or pyrrolidinyl), optionally substituted alkyl groups (e.g., optionally substituted linear or branched C₁₋₆ alkyl groups or optionally substituted C₁₋₆ aminoalkyl groups), or optionally substituted carbocyclyl groups (e.g., optionally substituted cyclopropyl, cyclohexyl, or cyclohexenyl). Particularly useful building block B_n's and C_n's include those with one or more reactive groups, such as an optionally substituted group (e.g., any described herein) having one or optional substituents that are reactive groups or can be chemically modified to form reactive groups. Exemplary reactive groups include one or more of amine (-NR₂, where each R is, independently, H or an optionally substituted C₁₋₆ alkyl), hydroxy, alkoxy (-OR, where R is an

optionally substituted C₁₋₆ alkyl, such as methoxy), carboxy (-COOH), amide, or chemically co-reactive substituents. A restriction site may be introduced, for example, in tag B_n or C_n, where a complex can be identified by performing PCR and restriction digest with one of the corresponding restriction enzymes.

Linkers

The bifunctional linker between the headpiece and the chemical entity can be varied to provide an appropriate spacer and/or to increase the solubility of the headpiece in organic solvent. A wide variety of
5 linkers are commercially available that can couple the headpiece with the small molecule library. The linker typically consists of linear or branched chains and may include a C₁₋₁₀ alkyl, a heteroalkyl of 1 to 10 atoms, a C₂₋₁₀ alkenyl, a C₂₋₁₀ alkynyl, C₅₋₁₀ aryl, a cyclic or polycyclic system of 3 to 20 atoms, a phosphodiester, a peptide, an oligosaccharide, an oligonucleotide, an oligomer, a polymer, or a poly alkyl glycol (e.g., a poly ethylene glycol, such as -(CH₂CH₂O)_nCH₂CH₂-, where n is an integer from 1 to 50),
10 or combinations thereof.

The bifunctional linker may provide an appropriate spacer between the headpiece and a chemical entity of the library. In certain embodiments, the bifunctional linker includes three parts. Part 1 may be a reactive group, which forms a covalent bond with DNA, such as, e.g., a carboxylic acid, preferably activated by a N-hydroxy succinimide (NHS) ester to react with an amino group on the DNA (e.g.,
15 amino-modified dT), an amidite to modify the 5' or 3'-terminus of a single-stranded headpiece (achieved by means of standard oligonucleotide chemistry), chemically co-reactive pairs (e.g., azido-alkyne cycloaddition in the presence of Cu(I) catalyst, or any described herein), or thiol reactive groups. Part 2 may also be a reactive group, which forms a covalent bond with the chemical entity, either building block A_n or a scaffold. Such a reactive group could be, e.g., an amine, a thiol, an azide, or an alkyne. Part 3
20 may be a chemically inert spacer of variable length, introduced between Part 1 and 2. Such a spacer can be a chain of ethylene glycol units (e.g., PEGs of different lengths), an alkane, an alkene, a polyene chain, or a peptide chain. The linker can contain branches or inserts with hydrophobic moieties (such as, e.g., benzene rings) to improve solubility of the headpiece in organic solvents, as well as fluorescent moieties (e.g. fluorescein or Cy-3) used for library detection purposes. Hydrophobic residues in the headpiece
25 design may be varied with the linker design to facilitate library synthesis in organic solvents. For example, the headpiece and linker combination is designed to have appropriate residues wherein the octanol:water coefficient (P_{oct}) is from, e.g., 1.0 to 2.5.

Linkers can be empirically selected for a given small molecule library design, such that the library can be synthesized in organic solvent, for example, in 15%, 25%, 30%, 50%, 75%, 90%, 95%,
30 98%, 99%, or 100% organic solvent. The linker can be varied using model reactions prior to library synthesis to select the appropriate chain length that solubilizes the headpiece in an organic solvent. Exemplary linkers include those having increased alkyl chain length, increased poly ethylene glycol units, branched species with positive charges (to neutralize the negative phosphate charges on the headpiece), or increased amounts of hydrophobicity (for example, addition of benzene ring structures).

Examples of commercially available linkers include amino-carboxylic linkers, such as those being peptides (e.g., Z-Gly-Gly-Gly-Osu (N-alpha-benzylloxycarbonyl-(Glycine)₃-N-succinimidyl ester) or Z-Gly-Gly-Gly-Gly-Gly-Gly-Osu (N-alpha-benzylloxycarbonyl-(Glycine)₆-N-succinimidyl ester, SEQ ID NO: 3)), PEG (e.g., Fmoc-aminoPEG2000-NHS or amino-PEG (12-24)-NHS), or alkane acid chains (e.g., Boc-ε-aminocaproic acid-Osu); chemically co-reactive pair linkers, such as those chemically co-reactive pairs described herein in combination with a peptide moiety (e.g., azidohomoalanine-Gly-Gly-Gly-OSu (SEQ ID NO: 4) or propargylglycine-Gly-Gly-Gly-OSu (SEQ ID NO: 5)), PEG (e.g., azido-PEG-NHS), or an alkane acid chain moiety (e.g., 5-azidopentanoic acid, (S)-2-(azidomethyl)-1-Boc-pyrrolidine, 4-azidoaniline, or 4-azido-butan-1-oic acid N-hydroxysuccinimide ester); thiol-reactive linkers, such as those being PEG (e.g., SM(PEG)_n NHS-PEG-maleimide), alkane chains (e.g., 3-(pyridin-2-yl-disulfanyl)-propionic acid-Osu or sulfosuccinimidyl 6-(3'-[2-pyridyldithio]-propionamido)hexanoate)); and amidites for oligonucleotide synthesis, such as amino modifiers (e.g., 6-(trifluoroacetyl-amino)-hexyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite), thiol modifiers (e.g., S-trityl-6-mercaptohexyl-1-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite, or chemically co-reactive pair modifiers (e.g., 6-hexyn-1-yl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite, 3-dimethoxytrityloxy-2-(3-(3-propargyloxypropanamido)propanamido)propyl-1-O-succinoyl, long chain alkylamino CPG, or 4-azido-butan-1-oic acid N-hydroxysuccinimide ester)). Additional linkers are known in the art, and those that can be used during library synthesis include, but are not limited to, 5'-O-dimethoxytrityl-1',2'-dideoxyribose-3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite; 9-O-dimethoxytrityl-triethylene glycol,1-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite; 3-(4,4'-dimethoxytrityloxy)propyl-1-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite; and 18-O-dimethoxytrityl hexaethyleneglycol,1-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite. Any of the linkers herein can be added in tandem to one another in different combinations to generate linkers of different desired lengths.

Linkers may also be branched, where branched linkers are well known in the art and examples can consist of symmetric or asymmetric doublers or a symmetric trebler. See, for example, Newcome et al., *Dendritic Molecules: Concepts, Synthesis, Perspectives*, VCH Publishers (1996); Boussif et al., *Proc. Natl. Acad. Sci. USA* 92:7297-7301 (1995); and Jansen et al., *Science* 266:1226 (1994).

Example 1

General strategy to improve single-stranded ligation of DNA tags

Various reaction conditions were explored to improve single-stranded ligation of tags to form an encoded library. These reaction conditions included using modified nucleotides within the tag (e.g., use of one or more nucleotides having a 2'-OMe group to form a MNA/DNA tag, where "MNA" refers to an oligonucleotide having at least one 2'-O-methyl nucleotide); using donor tags and acceptor tags having different lengths and varying the concentration of the tags; using different types of ligases, as well as combinations thereof (e.g., CircLigaseTM ssDNA ligase and/or T4 RNA ligase), and varying their

concentration; purifying the complex by removing unreacted starting materials; using poly ethylene glycols (PEGs) having different molecular weights and varying their concentration; varying the temperature and duration for reaction, such as ligation; varying the concentration of various agents, including ATP, $\text{Co}(\text{NH}_3)_6\text{Cl}_3$, and yeast inorganic pyrophosphate; using enzymatically or chemically phosphorylated oligonucleotide tags; using 3'-protected tags; and using 5'-chemically adenylated tags.

After a thorough analysis of different conditions, optimal combinations of parameters that provided up to 90% ligation efficiency (e.g., Figure 5C), as determined by the fraction of ligated final product to un-ligated starting reactant ("fraction ligated"), were found. A scheme of the ligation reaction using ligase is shown in Figure 5A, and a typical denaturing polyacrylamide gel electrophoresis is shown in Figure 5B. The donor oligonucleotide was labeled at the 3'-terminus and could be detected on a gel by scanning at 450 nm excitation on a StormTM 800 PhosphorImager. The gel depicts an unligated donor (or starting material) and the ligated product. In particular, the adenylated donor can be resolved and distinguished from the starting material on this gel.

Table 2 provides ligation efficiencies measured as a function of the composition of the oligonucleotide (i.e., oligonucleotides with all DNA nucleotides versus oligonucleotides with at least one 2'-O-methyl nucleotide, labeled "MNA") and the type of ligase (i.e., RNA ligase versus ssDNA ligase). These ligation experiments included the following tags: an all-DNA donor having the sequence of 5'-P-GCT GTG CAG GTA GAG TGC-6-FAM-3' (SEQ ID NO: 6); a 5'-MNA-DNA donor having the sequence of 5'-P-mGCT GTG CAG GTA GAG TGC-6-FAM-3' (SEQ ID NO: 7); an all-MNA donor having the sequence of 5'-P-mGmUmG mCmAmG mGmUmA mGmAmG mUmGmC-6-FAM-3' (SEQ ID NO: 8); a DNA-3'MNA acceptor having the sequence of 5'-HO-TAC GTA TAC GAC TGmG-OH-3' (SEQ ID NO: 9); an all-DNA acceptor having the sequence of 5'-HO-GCA GAC TAC GTA TAC GAC TGG-OH-3' (SEQ ID NO: 10); and an all-MNA acceptor having the sequence of 5'-HO-mUmAmC mGmUmA mUmAmC mGmAmC mUmGmG-OH-3' (SEQ ID NO: 11), where "m" indicates a 2'-OMe base, "P" indicates a phosphorylated nucleotide, and "FAM" indicates fluorescein.

Ligation efficiencies were calculated from gel densitometry data as the ratio between the intensity from the ligation product and the sum of the intensity from the ligation product and the unligated starting material. The reaction conditions for T4 RNA ligase included the following: 5 μM each of donor and acceptor oligonucleotides (15-18 nucleotides (nts) long) in a buffer solution containing 50 mM Tris HCl, 10 mM MgCl_2 , 1 mM hexamine cobalt chloride, 1 mM ATP, 25% PEG4600, and 5 units of T4 RNA ligase (NEB- new units) at pH 7.5. The reactions were incubated at 37°C for 16 hours. The reaction conditions for CircLigaseTM included the following: 5 μM each of donor and acceptor oligonucleotides (length 15 or 18 nts) incubated in a buffer containing 50 mM MOPS (pH 7.5), 10 mM KCl, 5 mM MgCl_2 , 1 mM DTT, 0.05 mM ATP, 2.5 mM MnCl_2 , and 25% (w/v) PEG 8000 with 20 units of CircLigaseTM (Epicentre) at 50°C for 16 hours. The reactions were resolved on 8M urea/15% PAAG, followed by densitometry using excitation at 450 nm.

Table 2

Donor	Acceptor	T4 RNA ligase	CircLigase™
All-DNA	All-DNA	9 %	89%
All-DNA	All-MNA	14%	68%
All-DNA	DNA-3'MNA	46%	85%
All-MNA	All-DNA	11%	84%
All-MNA	All-MNA	20%	29%
All-MNA	DNA-3'MNA	32%	73%
5'-MNA-DNA	All-DNA	29%	90%
5'-MNA-DNA	All-MNA	16 %	46%
5'-MNA-DNA	DNA-3'MNA	69%	81%

Generally, CircLigase™ produced higher ligation yields than T4 RNA ligase (Table 2). When both donor and acceptor were DNA/MNA hybrid oligonucleotides, efficient ligation was achieved with T4 RNA ligase.

Figure 5C shows high yield ligation achieved for T4 RNA ligase at high enzyme and oligonucleotide concentrations. The reaction conditions included following: 250 μM each of donor and acceptor oligonucleotides in a buffer containing 50 mM Tris HCl, 10 mM MgCl₂, 1 mM hexamine cobalt chloride, 2.5 mM ATP, 30% (w/v) PEG4600, pH 7.5, different amounts of T4 RNA ligase at 40 units/ μL (NEB- new units), and 0.1 unit of yeast inorganic pyrophosphatase. The reactions were incubated at 37°C for 5 and 20 hours and resolved on 8M urea/15% PAAG, followed by densitometry using excitation at 450 nm.

Overall, these data suggest that enzymatic ligation can be optimized by including one or more modified 2'-nucleotides and/or by using an RNA or DNA ligase. Further details for several other tested conditions, such as PEG or tag length, that can contribute to ligation efficiency are discussed below.

Example 2

Effect of PEG on single-stranded ligation

To determine the effect of PEG molecular weight (MW) on ligation, single-stranded tags were ligated with 25% (w/v) of PEG having a MW from 300 to 20,000 Daltons. As shown in Figure 6A, 80% or greater ligation was observed for PEG having a MW of 3,350, 4,000, 6,000, 8,000, and 20,000. These ligation experiments included the following tags: a 15mer donor having the sequence of 5'-P-mGTG CAG GTA GAG TGC-6-FAM-3' (SEQ ID NO: 12) and a 15mer acceptor having the sequence of 5'-HO-mUAC GTA TAC GAC TGmG-OH-3' (SEQ ID NO: 13). These oligonucleotide tags were DNA sequences with one or two terminal 2'-O-methyl (2'-OMe) RNA bases (e.g., 2'-OMe-U (mU) or 2'-OMe-G (mG)).

Experiments were also conducted to determine the effect of PEG concentration. Single-stranded tags were ligated with various concentration of PEG having a MW of 4,600 Daltons (PEG4600). As shown in Figure 6B, 70% or greater ligation, on average, was observed for 25% (w/v) to 35% (w/v) PEG4600.

Example 3

Effect of tag length on single-stranded ligation

To determine the effect of tag length on ligation, acceptor and donor tags of various lengths were constructed. For CircLigase™ experiments, a 15mer donor having the sequence 5'-P-mGTG CAG GTA GAG TGC-6-FAM-3' (SEQ ID NO: 12) was used and paired with 10, 12, 14, 16, and 18mer DNA acceptor oligonucleotides. For T4 RNA ligase experiments, the tags included one or more 2'-OMe-bases (designated as being MNA/DNA tags). Table 3 provides the sequence for the three donor tags (15mer, 8mer, and 5mer) and the three acceptor tags (15mer, 8mer, and 5mer).

Table 3

Oligonucleotide tag	Sequence*
15mer donor	5'-P-mGTG CAG GTA GAG TGC-6-FAM-3' (SEQ ID NO: 12)
15mer acceptor	5'-HO-mUAC GTA TAC GAC TGmG-OH-3' (SEQ ID NO: 13)
8mer donor	5'-P-mGT GAG TGC-6-FAM-3' (SEQ ID NO: 14)
8mer acceptor	5'-HO-C A GAC TGmG-OH-3' (SEQ ID NO: 15)
5mer donor	5'-P-mGT GAC-6-FAM-3' (SEQ ID NO: 16)
5mer acceptor	5'-HO-mAC TGmG-OH-3' (SEQ ID NO: 17)

* "m" indicates a 2'-OMe base, "P" indicates a phosphorylated nucleotide, and "FAM" indicates fluorescein.

The extent of ligation was analyzed by densitometry of electrophoretic gels (Figures 7A-7B). The results of the CircLigase™ reactions indicate a strong dependence of ligation yield on the length of the acceptor oligonucleotide (Figure 7A). The highest ligation yield was observed with an 18mer acceptor (62%), while ligation yield with a 10mer acceptor was lower than 10%. The results of the T4 RNA ligase reactions indicate that the combination of an 8mer acceptor with an 8mer donor provided the highest yield and that combinations having a 15mer donor with any of the tested acceptors provided yields greater than 75% (Figure 7B). If a library includes shorter tags (i.e., about 10mer or shorter), then T4 RNA ligase may be preferred for tag ligation. In other cases, ligation can be further optimized by using CircLigase™ or a combination of T4 RNA ligase and CircLigase™.

Example 4

Effect of purification on single-stranded ligation

To determine the effect of purification on ligation, single-stranded tags were ligated to imitate the library synthetic process. For these experiments, the tags included 15mer donor and 15mer acceptor tags, as provided above in Table 3. The chemical entity was bound to the 3'-terminus of the library, where the chemical entity was fluorescein in this example to aid in visualization. As shown in Figure 9 (right), successive tags were ligated to the 5'-OH group of the complex after phosphorylation by T4 PNK.

Experiments were also conducted by purifying the ligated product (i.e., the complex) prior to the PNK reaction, where particular agents useful in the ligation reaction (e.g., phosphate, cobalt, and/or unreacted tags) can inhibit the phosphorylation reaction with PNK or reduce ligation yield. As shown in Figure 9 (left), purifying the complex (i.e., minimal precipitation) prior to the PNK reaction increased ligation (see data marked with *, indicating purification). Figures 8A-8B show LC-MS spectra for a 15mer MNA/DNA tag before and after phosphorylation. The presence or absence of DTT had no effect on phosphorylation.

Example 5

Chemically co-reactive pair ligation and reverse transcription of junctions

The methods described herein can further include chemically co-reactive pair ligation techniques, as well as enzyme ligation techniques. Accordingly, as an example of chemical ligation, an exemplary chemically co-reactive pair (i.e., an alkyne and an azido pair in a cycloaddition reaction) in two variants: a short chemically co-reactive pair and a long chemically co-reactive pair, was used.

Materials

In a first variant, a short chemically co-reactive pair (Figure 10A) was used. The pair included (i) an oligonucleotide having the sequence 5'-GCG TGA ACA TGC ATC TCC CGT ATG CGT ACA GTC CAT T/propargylG/-3' ("5end3propargyl," SEQ ID NO: 18) and (ii) an oligonucleotide having the sequence 5'-/azidoT/ATA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG-3' ("3end5azido," SEQ ID NO: 19). This pair of oligonucleotides was prepared by TriLink BioTechnologies, Inc. (San Diego, CA). These oligonucleotides were designed to produce a short spacer between two oligonucleotides upon ligation, where the linker would be 5 atoms long (counting from the C3'-position of the 5end3propargyl oligonucleotide to the C5'-position of the 3end5azido oligonucleotide). In addition, the 5'-azido oligonucleotide (3end5azido) was prepared by converting the iodo group in the corresponding 5'-iodo oligonucleotide into an azido group.

In a second variant, a long chemically co-reactive pair (Figure 10B) was used. The pair included (i) an oligonucleotide having the sequence 5'-GCG TGA ACA TGC ATC TCC CGT ATG CGT ACA GTC CAT TG/spacer7-azide/-3' ("5end3azide," SEQ ID NO: 20) and (ii) an oligonucleotide having the sequence 5'-/hexynyl/TA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG-3' ("3end5hexynyl," SEQ ID NO: 21). This pair of oligonucleotides was prepared by Integrated DNA Technologies, Inc. (IDT DNA, San Diego, CA, and Coralville, IA). The 5end3azide oligonucleotide was prepared by reacting an azidobutyrate N-hydroxysuccinimide ester with a 3'-amino-modifier C7 (2-dimethoxytrityl oxymethyl-6-fluorenylmethoxycarbonylamino-hexane-1-succinoyl-long chain alkylamino), which was introduced during oligonucleotide column synthesis. This pair was designed to produce a 24 atom long spacer between the oligonucleotides (counting from the C3'-position of the 5end3azide oligonucleotide to the C5'-position of the 3end5hexynyl oligonucleotide).

For reverse transcription (as shown by the schematic in Figure 11A), the primers and templates included the following: a reverse transcription primer having the sequence of 5'-Cy5/ CAG TAC GCA AGC TCG-3' ("Cy5s_primer15," SEQ ID NO: 22); a control template having the sequence of 5'-GCG TGA ACA TGC ATC TCC CGT ATG CGT ACA GTC CAT TGT ATA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG-3' ("templ75," SEQ ID NO: 23); a 5'-PCR primer having the sequence of 5'-GCG TGA ACA TGC ATC TCC-3' (SEQ ID NO: 24); and a 3'-PCR primer having the sequence of 5'-CAG TAC GCA AGC TCG CC-3' (SEQ ID NO: 25), where these sequences were obtained from IDT DNA. A Cy5-labeled DNA primer was used for the experiments to enable separate detection of the reverse transcription products by LC.

Experimental Conditions

For the chemically co-reactive pair ligations, 1 mM solutions of chemically co-reactive pairs, such as 5end3propargyl+3end5azido (short) or 5end3azide+3end5hexynyl (long), were incubated for 12 hours in the presence of 100 equivalents of TBTA ligand (tris-[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine) and 50 equivalents of CuBr in a water/dimethyl acetate mixture. Following the reaction, an excess of EDTA was added, and the reaction mixtures were desalted using Zeba Spin Desalting Columns (Invitrogen Corp., Carlsbad, CA) and then ethanol precipitated. For the reverse transcription reactions, the templates were purified on a 15% polyacrylamide gel containing 8M urea.

Liquid chromatography-mass spectrometry (LC-MS) was performed on a Thermo Scientific LCQ Fleet using an ACE 3 C18-300 (50 x 2.1 mm) column and a 5 minute gradient of 5-35% of buffer B using buffer A (1% hexafluoroisopropanol (HFIP), 0.1% di-isopropylethyl amine (DIEA), 10μM EDTA in water) and buffer B (0.075% HFIP, 0.0375% DIEA, 10 μM EDTA, 65% acetonitrile/35% water). LC was monitored at 260 nm and 650 nm. MS was detected in the negative mode, and mass peak deconvolution was performed using ProMass software.

Reverse transcription reactions were performed using ThermoScript™ RT (Invitrogen Corp.), according to the manufacturer's protocol, at 50°C for 1-2 hours. The results were analyzed by LC-MS and by PCR. PCR was performed using Platinum® SuperMix and resolved on 4% agarose E-Gels (both from Invitrogen Corp.). Eleven and eighteen cycles of PCR were performed with or without a preceding RT reaction. The 75mer template was not reverse transcribed and used directly for the PCR amplification.

Results and discussion

In both the ligations forming a short spacer and a long spacer, reaction yields were high, close to quantitative, as analyzed by LC-MS. Accordingly, chemical ligation provides a high yield technique to bind or operatively associate a headpiece to one or more building block tags.

For a viable chemical ligation strategy to produce DNA-encoded libraries, the resultant complex should be capable of undergoing PCR or RT-PCR for further sequencing applications. While PCR and

RT-PCR may not be an issue with enzymatically ligated tags, such as described above, unnatural chemical linkers may be difficult to process by RNA or DNA polymerases. The data provided in Figures. 11B-11E suggest that oligonucleotides having a spacer of particular lengths can be transcribed and/or reverse transcribed.

5 In the case of a chemically co-reactive pair linker resulting in a triazole-linked oligonucleotide, a dependence on the length of the linker was observed. For the short chemically co-reactive pair, the resultant template was reverse transcribed and analyzed by LC-MS. LC analysis revealed three major absorption peaks at 2.79 min., 3.47 min., and 3.62 min. for 260 nm, where the peaks at 3.47 min. and 3.62 min. also provided absorption peaks at 650 nm. MS analysis of the peak at 3.47 min. showed only the
10 presence of the template 23097.3 (calc'd 23098.8), and the peak at 3.62 min. contained a template (23098.0) and a fully extended primer (23670.8, calc'd: 23671.6) at an approximately 1.7:1 ratio, suggesting a 50-60% yield for this RT reaction (Figure 11C). For comparison, reverse transcription (RT) of the control having an all-DNA template produced the extended primer (peak 23068.9) in an amount roughly equivalent to the template (23078.7), suggesting close to a 100% yield (Figure 11B).

15 For the long chemically co-reactive pair, LC of the RT reaction showed two absorption peaks at 2.77 min and 3.43 min for 260 nm, where the peak at 3.43 min also provided absorption peaks at 650 nm, i.e., contained a Cy5 labeled material, which is the expected RT product. MS analysis of the peak at 3.43 min. revealed the template (observed 23526.6, calc'd: 23534.1), as well as the Cy5 primer extended to the linker (11569.1). No full length product was observed by LC-MS, indicating that the RT reaction did not
20 occur in a measureable amount (Figure 11D).

RT-PCR was performed with the templates described above and revealed that only the short linker yielded reverse transcription product, albeit at 5-10 lower efficiency (Figure 11E). Efficiency of the RT was estimated to be about 2-fold lower than the template (templ75). For example, the PCR product of the short ligated template around 2-fold lower after RT and around 5-10 times lower without
25 RT, as compared to the PCR product of the all-DNA template 75 (templ75). Accordingly, these data provide support for the use of chemical ligation to produce a complex that can be reverse transcribed and/or transcribed, and chemically ligated headpieces and/or tags can be used in any of the binding steps described herein to produce encoded libraries.

30 Example 6

Ligation of 3'-phosphorothioate oligonucleotides with 5'-iodo oligonucleotides

To determine the flexibility of the methods described herein, the ligation efficiency of oligonucleotides having other modifications were determined. In particular, analogs of the natural phosphodiester linkage (e.g., a phosphorothioate analog) could provide an alternative moiety for post-
35 selection PCR analysis and sequencing.

The following oligonucleotides were synthesized by TriLink BioTechnologies, Inc. (San Diego, CA): (i) 5'-Cy5/ CGA TAT ACA CAC TGG CGA GCT/thiophosphate/-3' ("CCy5," SEQ ID NO: 26),

(ii) 5'-IododT/ GC GTA CTG AGC/6-FAM/-3' ("CFL," SEQ ID NO: 27), as shown in Figure 12A, and (iii) a splint oligonucleotide having the sequence of CAG TAC GCA AGC TCG CC ("spl," SEQ ID NO: 28). Ligation reactions were performed with 100 μ M of each reactant oligonucleotide in a buffer containing 50mM Tris HCl (pH 7.0), 100 mM NaCl, and 10 mM MgCl₂ ("ligation buffer") at room temperature. The ligation reactions were supplemented by either of the following: 100 μ M of the splint oligonucleotide, 10 mM Co(NH₃)₆Cl₃, 40% (w/v) of PEG4000, or 80% (w/v) of PEG300. The reaction was allowed to progress for up to 48 hours. Ligation products were analyzed by LC-MS using detection at 260 nm, 495 nm, and 650 nm, as well as by an 8M urea/ 15% polyacrylamide gel (PAAG) that was further scanned at 450 and 635 nm excitation on a StormTM 800 PhosphorImager.

In the absence of the splint oligonucleotide, no ligation was observed (Figure 12B, lanes labeled "-spl"). In the presence of the splint oligonucleotide, ligation occurred and reached around 60% of fraction ligated after 48 hours (Figures 12B-12C). LC-MS revealed several peaks in the chromatogram, with a peak at 3.00 min absorbing at 260 nm, 495 nm, and 650 nm. MS of this peak showed mostly the product of ligation at 11539.6 Da (calc'd 11540) with less than 10% of CCy5 oligonucleotide at 7329.8 Da (calc'd 7329.1). Low levels of ligation were detected in the presence of PEGs and hexamine cobalt, where hexamine cobalt caused precipitation of the Cy5-labeled oligonucleotide. These data suggest that headpieces and/or tags having modified phosphate groups (e.g., modified phosphodiester linkages, such as phosphorothioate linkages) can be used in any of the binding steps described herein to produce encoded libraries.

In order to further study the iodo-phosphorothioate ligation reaction, the ligation of 5'-I dT-oligo-3'-FAM (CFL) and 5'-Cy5-oligo-3'-PS (CCy5) was performed in the absence and presence of a splint under different reaction conditions.

In a first set of conditions, ligation experiments were conducted with incubation for seven to eight days. These experiments were performed in the same ligation buffer as above with 50 μ M of each oligonucleotide and incubated for a week at room temperature. Figure 12D shows LC-MS analysis of the ligation of CFL and CCy5 in the absence (top) and presence (bottom) of a splint (positive control), where ligation reactions were incubated for seven days. Three LC traces were recorded for each reaction at 260 nm (to detect all nucleic acids), at 495 nm (to detect the CFL oligonucleotide and the ligation product), and at 650 nm (to detect the CCy5 oligonucleotide and the ligation product).

In the absence of the splint, no ligation occurred, and only starting materials CFL (4339 Da) and CCy5 (7329 Da) were detected (Figure 12D, top). When the splint oligonucleotide was present for seven days, a characteristic peak was observed in 495 nm channel with a retention time of 2.98 min, which corresponds to the ligated product (11542 Da) (Figure 12D, bottom). This peak overlapped with that for the CCy5 oligonucleotide observed at the 650 nm channel and, thus, was indistinguishable from CCy5 at 650 nm.

Figure 12E shows the LC-MS analysis of CFL and CCy5 in the absence of a splint, where ligation reactions were incubated for eight days at 400 μ M of each oligonucleotide. No ligation product

was detected. Peak 1 (at 495 nm) contained CFL starting material (4339 Da), as well as traces of the loss of iodine product (4211 Da) and an unknown degradation product (4271 Da, possibly ethyl mercaptane displacement). Peak 2 (at 650 nm) contained CCy5 starting material (7329 Da) and oxidized CCy5 oligonucleotide (7317 Da). Peak 3 (at 650 nm) contained dimerized CCy5 (14663 Da).

5 In a second set of conditions, iodine displacement reactions were conducted in the presence of piperidine and at a pH higher than 7.0. Figure 12F shows MS analysis for a reaction of CFL oligonucleotide with piperidine, where this reaction was intended to displace the terminal iodine present in CFL. One reaction condition included oligonucleotides at 100 μ M, piperidine at 40 mM (400 equivalents) in 100 mM borate buffer, pH 9.5, for 20 hrs at room temperature (data shown in left panel of
10 Figure 12F); and another reaction condition included oligonucleotides at 400 μ M, piperidine at 2 M (4,000 equivalents) in 200 mM borate buffer, pH 9.5, for 2 hrs at 65°C (data shown in right panel of Figure 12F).

In the reaction condition including 40 mM of piperidine (Figure 12F, left), no piperidine displacement was observed, and a small amount of hydrolysis product was detected (4229 Da). In
15 addition, traces of the loss of iodine (4211 Da) and unknown degradation product (4271 Da) were observed. In the reaction condition including 2 M of piperidine (Figure 12F, right), piperidine displacement of iodine was observed (4296 Da), and the amount of starting material was substantially diminished (4339 Da). In addition, peaks corresponding to hydrolysis of iodine (by displacement of OH) or impurity (4229 Da) and loss of iodine (4214 Da) were also observed. These data show that the
20 presence of an amine (e.g., as part of chemical library synthesis) will not detrimentally effect the oligonucleotide portion of the library members and/or interfere with this ligation strategy.

In a third set of conditions, splint ligation reactions were conducted in the presence of piperidine and at a pH higher than 7.0. Figure 12G shows a splint ligation reaction of CFL and CCy5 oligonucleotides at 50 μ M performed in the presence of 400 equivalents of piperidine in 100 mM borate
25 buffer, pH 9.5, for 20 hrs at room temperature. The characteristic peak detected in the LC trace (at 495 nm) contained predominantly the product of ligation at 11541.3 Da (calc'd 11540 Da). Based on these results, it can be concluded that that piperidine does not impair enzymatic ligation and that the presence of other amines (e.g., as part of chemical library synthesis) will likely not interfere with this ligation strategy.

30 Taking together, these data indicate that this ligation strategy can be performed under various reaction conditions that are suitable for a broad range of chemical transformations, including extended incubation times, elevated pH conditions, and/or presence of one or more amines. Thus, the present methods can be useful for developing library members with diverse reaction conditions and precluding the necessity of buffer exchange, such as precipitation or other resource-intensive methods.

Example 7

Minimization of shuffling with modified nucleotides

During single-stranded enzymatic ligation with T4 RNA ligase, low to moderate extent of terminal nucleotide shuffling can occur. Shuffling can result in the inclusion or excision of a nucleotide, where the final product or complex includes or excludes a nucleotide compared to the expected ligated sequence (i.e., a sequence having the complete sequence for both the acceptor and donor oligonucleotides).

Though low levels of shuffling can be tolerated, shuffling can be minimized by including a modified phosphate group. In particular, the modified phosphate group is a phosphorothioate linkage between the terminal nucleotide at the 3'-terminus of an acceptor oligonucleotide and the nucleotide adjacent to the terminal nucleotide. By using such a phosphorothioate linkage, shuffling was greatly reduced. Only residual shuffling was detected by mass spectrometry, where shuffling likely arose due to incomplete conversion of the native phosphodiester linkage into the phosphorothioate linkage or to low levels of oxidation of the phosphorothioate linkage followed by conversion into the native phosphodiester linkage. Taking together this data and the ligation data in Example 6, one or more modified phosphate groups (e.g., a phosphorothioate or a 5'-N-phosphoramidite linkage) could be included in any oligonucleotide sequence described herein (e.g., between the terminal nucleotide at the 3'-terminus of a headpiece, a complex, a building block tag, or any tag described herein, and the nucleotide adjacent to the terminal nucleotide) to minimize shuffling during single-stranded ligation.

A single stranded headpiece (ssHP, 3636 Da) was phosphorylated at the 5'-terminus and modified with a hexylamine linker at the 3'-terminus to provide the sequence of 5'-P-mCGAGTCACGTC/Aminohex/-3' (SEQ ID NO: 29). The headpiece was ligated to a tag (tag 15, XTAGSS000015, 2469 Da) having the sequence of 5'-mCAGTGTCmA-3' (SEQ ID NO: 30), where mC and mA indicate 2'-O methyl nucleotides. LC-MS analysis (Figure 13A) revealed that the ligation product peak contained up to three species, which was partially separated by LC and had the following molecular weights: 6089 Da (expected), 5769 Da (-320 Da from expected) and 6409 Da (+320 Da from expected). This mass difference of 320 Da corresponds exactly to either removal or addition of an extra O-Me C nucleotide ("terminal nucleotide shuffling").

Experiments with other terminal O-Me nucleotides, as well as terminal 2'-fluoro nucleotides, confirmed that shuffling likely occurs by cleavage of the 5'-terminal nucleotide of the donor oligonucleotide, probably after adenylation of the latter. The mechanism of this event is unknown. Without being limited by mechanism, Figure 13B illustrates a possible scheme for nucleotide reshuffling during T4 RNA ligase reaction between a headpiece and a tag, where one of skill in the art would understand that this reaction could occur between any donor and acceptor oligonucleotides (e.g., between two tags, where one tag is the donor oligonucleotide and the other tag is the acceptor oligonucleotide).

Generally, the majority of the ligation reaction with T4 RNA ligase (T4Rnl1) provides the expected (normal) ligation product having the combined sequence of both the donor and acceptor

oligonucleotides (Figure 13B-1, reaction on left). A small minority of the reaction provides aberrant ligation products (Figure 13B-1, reaction on right), where these aberrant products include those having the removal or addition of a terminal nucleotide ("Product -1 nt" and "Product + 1 nt," respectively, in Figure 13B-2).

Without being limited by mechanism, cleavage of the donor oligonucleotide ("headpiece" or "HP" in Figure 13B-1) may occur by reacting with the 3'-OH group of the acceptor ("tag"), thereby providing a 5'-phosphorylated donor lacking one nucleotide ("HP-1 nt") and an adenylated nucleotide with an accessible 3'-OH group ("1 nt"). Figure 13B-2 shows two exemplary schemes for the reaction between the headpiece (HP), tag, HP-1 nt, and 1 nt. To provide a product with an excised terminal nucleotide (Figure 13B-2, left), the 5'-phosphorylated donor lacking one nucleotide (HP-1 nt) acts a substrate for the ligation event. This HP-1 nt headpiece is re-adenylated by T4 RNA ligase (to provide "Adenylated HP-1 nt" in Figure 13B-2) and ligated to the tag, resulting in a ligation product minus one nucleotide ("Product-1 nt"). To provide a product with an additional terminal nucleotide (Figure 13B-2, left), the adenylated nucleotide (1 nt) likely serves as a substrate for ligation to the tag, thereby producing an oligonucleotide having one nucleotide longer than the acceptor ("Tag+1 nt"). This Tag+1 nt oligonucleotide likely serves as an acceptor for the unaltered headpiece, where this reaction provides a ligation product having an additional nucleotide ("Product+1 nt"). LC-MS analyses of "Product", "Product-1 nt", and "Product+1 nt" were performed (Figure 13B-3). When an aberrant tag and an aberrant headpiece (i.e., Tag+1 nt and HP-1 nt, respectively) recombine, then the resultant ligation product is indistinguishable from the expected product.

To further study the mechanism of terminal nucleotide reshuffling, a headpiece (HP-PS) having the sequence of 5'-P-mC*GAGTCACGTC/Aminohex/-3' (SEQ ID NO: 31) was prepared. Headpiece HP-PS has the same sequence as ssHP but contains one modification, namely the first phosphodiester linkage between 5'-terminal nucleotide mC and the following G was synthesized as a phosphorothioate linkage (one non-bridging phosphate oxygen was substituted by a sulfur). LC-MS analysis of the HP-PS ligation to tag 15 revealed that shuffling was almost completely inhibited (Figure 13C). Traces of +/- 320 peaks likely correspond to the oxidative conversion of the phosphorothioate linkage into native phosphodiester linkages or incomplete sulfurization.

Example 8

Size exclusion chromatography of library members

Libraries of chemical entities that are generated using short, single-stranded oligonucleotides as encoding elements are well suited for the enrichment of binders via size exclusion chromatography (SEC). SEC is chromatographic technique that separates molecules on the basis of size, where larger molecules having higher molecular weight flow through the column faster than smaller molecules having lower molecular weight.

Complexes of proteins and ssDNA library members can be readily separated from unbound library members using SEC. Figure 14 is an ultraviolet trace from an SEC experiment in which a small molecule covalently attached to short ssDNA (a range of oligonucleotides with defined lengths in the 20-50 mer range) was mixed with a protein target known to bind the small molecule. The peaks that elute first from the column, in the 11-13 minute time range, represent target-associated library members. The later peaks, eluting from 14-17 minutes, represent unbound library members. The ratio of protein target to library molecule was 2:1, so approximately 50% of the library molecules should associate with the protein in the early eluting fraction, as observed in Figure 14. Libraries with larger, double-stranded oligonucleotide coding regions cannot be selected using this method since the unbound library members co-migrate with the bound library members on SEC. Thus, small molecule libraries attached to encoding single-stranded oligonucleotides in the 20-50mer length range enable the use of a powerful separation technique that has the potential to significantly increase the signal-to-noise ratio required for the effective selection of small molecule binders to one or more targets, e.g., novel protein targets that are optionally untagged and/or wild-type protein. In particular, these approaches allow for identifying target-binding chemical entities in encoded combinatorially-generated libraries without the need for tagging or immobilizing the target (e.g., a protein target).

Example 9

Encoding with chemically ligated DNA tags using the same chemistry for each ligation step

Encoding DNA tags can be ligated enzymatically or chemically. A general approach to chemical DNA tag ligation is illustrated in Figure 15A. Each tag bears co-complementary reactive groups on its 5' and 3' ends. In order to prevent polymerization or cyclization of the tags, either (i) protection of one or both reactive groups (Figure 15A), e.g., in case of TIPS-protected 3' alkynes, or (ii) splint-dependent ligation chemistry (Figure 15B), e.g., in the case of 5'-iodo/3'-phosphorothioate ligation, is used. For (i), unligated tags can be removed or capped after each library cycle to prevent mistagging or polymerization of the deprotected tag. This step may be optional for (ii), but may still be included. Primer extension reactions, using polymerase enzymes that are capable of reading through chemically ligated junctions, can also be performed to demonstrate that ligated tags are readable and therefore the encoded information is recoverable by post-selection amplification and sequencing (Figure 15C).

A library tagging strategy that implements ligation of the tags using "click-chemistry" (Cu(I) catalyzed azide/alkyne cycloaddition) is shown in Figure 16A. The implementation of this strategy relies on the ability of precise successive ligation of the tags, avoiding mistagging, and tag polymerizations, as well as the ability to copy the chemically ligated DNA into amplifiable natural DNA (cDNA) for post-selection amplification and sequencing (Figure 16C).

To achieve accurate tag ligation triisopropylsilyl (TIPS)-protected 3' propargyl nucleotides, (synthesized from propargyl U in the form of a CPG matrix used for oligonucleotide synthesis) was used (Figure 16B). The TIPS protecting group can be specifically removed by treatment with

tetrabutylammonium fluoride (TBAF) in DMF at 60°C for 1-4 hours. As a result, the ligation during library synthesis includes a 5'-azido/3'-TIPS-propargyl nucleotide (Tag A) reacting with the 3'-propargyl of the headpiece through a click reaction. After purification, the previous cycle is treated with TBAF to remove TIPS and generate the reactive alkyne which in turn reacts with the next cycle tag. The procedure is repeated for as many cycles as it is necessary to produce 2, 3 or 4 or more successively installed encoding tags (Figure 16A).

Materials and methods

Oligos: The following oligos were synthesized by Trilink Biotechnologies, San Diego CA: ss-HP-alkyne: 5'-NH₂-TCG AAT GAC TCC GAT AT (3'-Propargyl G)-3' (SEQ ID NO: 32); ss-azido-TP: 5'-azido dT ATA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG -3' (SEQ ID NO: 33); and B-azido: 5' azido dT ACA CAC TGG CGA GCT TGC GTA CTG -3' (SEQ ID NO: 34).

ClickTag-TIPS: 5'-azido dT AT GCG TAC AGT CC (propargyl U-TIPS)-3' (SEQ ID NO: 35) and 5'-Dimethoxytrityl 2'-succinyl 3'-O-(triisopropyl silyl) Propargyl uridine cpg were synthesized by Prime Organics, Woburn MA.

The following oligos were synthesized by IDT DNA technologies, Coralville, IA: FAM-click-primer: (5'-6-FAM) CAG TAC GCA AGC TCG CC -3' (SEQ ID NO: 36) and Cy5-click-primer: (5'-Cy5) CAG TAC GCA AGC TCG CC -3' (SEQ ID NO: 37).

DNA55-control: /5'Biotin-TEG//ispC3//ispC3/-TCGAATGACTCCGATATGT ATA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG -3' (SEQ ID NO: 38).

rDNA55-control: /5Bio-TEG//ispC3//ispC3/-TCGAATGACTCCGATAT(riboG)T ATA GCG CGA TAT ACA CAC TGG CGA GCT TGC GTA CTG -3' (SEQ ID NO: 39)

Synthesis of the templates: In the following examples, the phrase "chemically ligated tags", or control sequences related to them, are referred to as "templates" because the subsequent step ("reading") utilizes them as templates for template-dependent polymerization.

Tag ligation: To a solution of 1 equivalent (1 mM) of ssHP-alkyne and 1 equivalent (1 mM) of ss-azidoTP in 500 mM pH 7.0 phosphate buffer, was added a solution of pre-mixed 2 eq of Cu(II)Acetate (to a final concentration of 2 mM), 4 eq of sodium ascorbate (to a final concentration of 4 mM), 1 eq TBTA (to a final concentration of 1 mM) in DMF/water. The mixture was incubated at room temperature overnight. After LC-MS confirmation of the completion of the reaction, the reaction was precipitated using salt/ethanol.

"Single click" templates Y55 and Y185 were synthesized by the reaction of ss-HP-alkyne with ss-azido-TP and B-azido, respectively. Double and triple click templates (YDC and YTC) were synthesized by click ligation of ss-HP-alkyne with ClickTag-TIPS, followed by deprotection of TIPS using TBAF (tetrabutylammonium fluoride) in DMF at 60°C for an hour, followed by click ligation with ss-azido TP. For triple click template (YTC), the ligation and deprotection of ClickTag-TIPS was repeated twice.

The templates were reacted with biotin-(EG)₄-NHS and desalted (Figure 17A). The final products were purified by RP HPLC and/or on a 15-20% polyacryl amide gel/ 8M urea and analyzed by LC-MS.

Enzymes: The following DNA polymerases with their reaction buffers were purchased from New England Biolabs: Klenow fragment of E. coli DNA polymerase I, Klenow fragment (exo-), E. coli DNA polymerase I, Terminator™, 9°N™, Superscript III™.

Streptavidin magnetic Dynabeads® M280 were purchased from Invitrogen.

Template-dependent polymerization assessment: Each template (5 μM) was incubated with 1 equivalent of either Cy5 or FAM Click-primer in 40 to 50 μL of the corresponding 1x reaction buffer and each enzyme, using reaction conditions according to the manufacturer's guidelines for 1 hour. Certain reactions (such as SSII or SSIII transcriptions) were additionally supplemented with 1 mM MnCl₂. The product of the reaction was loaded on 125 μL of pre-washed SA beads for 30 minutes with shaking. The beads were then collected, and the flowthrough was discarded. Beads were washed with 1 mL of Tris-buffered saline (pH 7.0) and eluted with 35 μL of 100 mM NaOH. The eluate was immediately neutralized by adding 10 μL of 1 M Tris HCl, pH 7.0. The products were analyzed using LC-MS.

Results and discussion

Template Preparation: Each template, Y55, Y185 (Figures 17B and 17C), YDC and YTC (Figure 19) was synthesized and purified to greater than 85% purity (the major impurity being unbiotinylated template). LC-MS revealed the following MWs for the templates: Y55 17,624 (calculated 17,619) Da; YDC 22,228 (calculated 22,228) Da; and YTC 26,832 (calculated 26,837) Da.

The single click templates Y55 and Y185 (Figures 17B and 17C) were synthesized from oligonucleotides that bear only one click chemistry functionality (alkyne or azide). The efficiency of the click reaction (chemical ligation) was over 90% in an overnight reaction using Cu(I) catalyst generated *in situ*.

Templates YDC and YTC (Figures 19A-19D) serve to demonstrate successive chemical ligations. Both YDC and YTC use individual tags which simultaneously contain both azido and TIPS-protected alkyne functionalities. Template YTC demonstrates three successive cycles of tagging as may be used to encode three steps of chemical library generation.

All of the above templates were tested for primer extension through and beyond the click-ligation linkages to demonstrate that ligated tags are readable, and therefore that encoded information is recoverable.

Template-dependent polymerization using "single-click" template Y55: A large set of polymerases was tested to read through a triazole click linkage (Figure 18A). Initial experiments were performed using Cy5-click-primer. In later experiments FAM-click-primer was used. The fluorophore had no effect on the copying of the template, i.e., the results were equivalent using either primer. As a control template DNA55-control and rDNA55-control were used (to test the effect of a single ribonucleotide in the template, since propargyl-G used for a click ligation is a ribonucleotide derivative).

Expected full length products in all three templates have the same molecular weight, which is 17446 (FAM primer) (Figure 18B) or 17443 (Cy5 primer). A small amount of the product which corresponds to primer extension up to, but stopping at, the click ligation linkage (11880 Da) was also observed for some polymerases.

- 5 A set of polymerases that can produce substantial degree of read-through of the click linkage (production of full-length cDNA) were discovered and are tabulated below.

Full-length cDNA yields of over 50%
Klenow fragment of <i>E. coli</i> DNA polymerase I
Klenow fragment (exo-)
<i>E. coli</i> DNA polymerase I
Therminator™
9°N™
Superscript III™ supplemented with 1 mM MnCl ₂

- The highest yields (over 80% read-through at a single click junction) were achieved when using
 10 Klenow fragment with incubation at 37°C (Figure 18B). Somewhat lower yield was observed using *E. coli* DNA polymerase I. 50% yields with Therminator™ and 9°N™ polymerases, as well as Klenow fragment exo- were achieved.

- Superscript III™ reverse transcriptase produced about 50% yield of cDNA when the buffer was supplemented with 1 mM MnCl₂. However, manganese caused the mis-incorporation of nucleotides
 15 which was observed by MS, i.e., polymerization fidelity was reduced.

- Template-dependent polymerization using "single-click" template Y185:* Template Y185 features the same primer binding site as all templates used in this example, except, due to a different tailpiece B-azido, the distance between the last nucleotide of the primer binding site to the click linkage is 8 nucleotides, as compared to 20 nucleotides in Y55 and all other templates. The template was used to test
 20 whether transcription of a click linkage was still possible when the enzyme was in initiation-early elongation conformation. Klenow was capable of copying the Y185 template with similar efficiency to Y55, opening the possibility of reducing the length of the click-ligated encoding tags (Figure 18C).

- Template-dependent polymerization using double and triple click-ligated templates YDC and YTC:* After establishing that the Klenow fragment was the most efficient enzyme to read through the click
 25 ligation linkages under the assay condition employed, cDNA using YDC and YTC templates (Figures 20A-20C) were also generated. Primer extension reactions with both YDC and YTC templates produced full length products. Other observed products, which composed around 10-15% of total reaction output, corresponded to partially extended primer, stalled at each click junction, such as e.g., 11880 Da and 16236 Da. The yields were measured by LC-MS analysis in the presence of the internal standard and

were about 80-90% per junction (i.e., around 85% for 1 click, 55% for 2-click and 50% for 3-click templates, see Figure 21).

The product of YDC transcription lacked 1 dA nucleotide (calculated 22110, observed 27197 Da; -313 dA Figure 20B) and the product of YTC transcription lacked 2 dA nucleotides (calculated 26773, observed 26147; -626 2xdA) (Figure 20C). This correlates with the number of propargyl U nucleotides in the template. Without wishing to be limited by mechanism, it can be hypothesized that Klenow skipped over those U's in the context of T-triazole-U junction. In contrast, the propargyl G nucleotide in the 1st click junction was correctly copied.

Example 10

Use of 3'-phosphorothioate/5'-iodo tags to chemically ligate a succession of encoding DNA tags that encode a chemical library covalently installed upon the 5'-terminus

Protection of 3'-phosphorothioate on tag: As shown in Figure 24A, a 5'-iodo-3'-phosphorothioate tag (1 eq.) was dissolved in water to give a final concentration of 5 mM. Subsequently, vinyl methyl sulfone (20 eq.) was added and the reaction was incubated at room temperature overnight. Upon completion of the reaction, the product was precipitated by ethanol.

Library synthesis (Figure 24B)

Cycle A: To each well in the split was added single-stranded DNA headpiece (1 eq., 1 mM solution in 500 mM pH 9.5 borate buffer), one cycle A protected tag (1.5 eq.), and splint (1.2 eq.). The chemical ligation was incubated at room temperature overnight. To each well (in the split) was then added one Fmoc amino acid (100 eq.), followed by 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (100 eq.). The chemical reaction was incubated at room temperature overnight. Upon completion, all wells were pooled and the products precipitated using ethanol. The cycle A pool was purified using LC and lyophilized to dryness, and then dissolved in water to give a 1 mM final concentration and piperidine (10% v/v) was added to perform the deprotection of cycle A tag (60°C, 2h). The deprotected product was precipitated again using ethanol.

Cycle B: The deprotected cycle A pool was dissolved in 500 mM, pH 9.5, borate buffer to give a 1mM concentration and then split into separate reaction wells (1 eq. of cycle A product in each well). To each well was added one cycle B protected tag (1.5 eq.), and splint (1.2 eq.). The chemical ligation was incubated at room temperature overnight. To each well (in the split) was added a mixture of one formyl acid (100 eq.), diisopropyl carbodiimide (100 eq.) and 1-hydroxy-7-aza-benzotriazole (100 eq.). The chemical reaction was incubated at room temperature overnight. Upon completion, all wells were pooled and the products precipitated using ethanol. The cycle B pool was purified using LC and lyophilized to dryness, and then dissolved in water to give a 1 mM final concentration and piperidine (10% v/v) was added to perform the deprotection of cycle B tag (60°C, 2h). The deprotected product was precipitated again using ethanol.

Cycle C: The deprotected cycle B pool was dissolved in 500 mM pH 5.5 phosphate buffer to give a 1 mM concentration and then split into separate reaction wells (1 eq. of cycle B product in each well). To each well was added one cycle C tag (1.5 eq.) and splint (1.2 eq.). The chemical ligation was incubated at room temperature overnight. To each well (in the split) was added an amine (80 eq.) and sodium cyanoborohydride (80 eq.). The chemical reaction was incubated at 60°C for 16h. Upon completion, all wells were pooled and the products precipitated using ethanol. The cycle C pool was purified using LC and lyophilized to dryness.

Example 11

Encoding with chemically ligated DNA tags using a pair of orthogonal chemistries for each successive tag ligation step

Another approach for generation of chemically ligated encoding DNA tags is the use of a pair of orthogonal chemistries for successive ligations (Figure 22A). Tags that bear orthogonal reactive groups at their ends will not tag polymerize or cyclize, and the orthogonal nature of successive ligation steps will reduce the frequency of mistagging events. Such approaches require (i) having at least two orthogonal chemistries available for oligonucleotide conjugation, and (ii) available read-through strategy for each of the junctions thus created (Figures 22B and 22C). This approach may also obviate the need for the use of protection groups or capping steps, thereby simplifying the tag ligation process.

Orthogonal chemical ligation strategy utilizing 5'-Azido/3'-Alkynyl and 5'-Iodo/3'-Phosphorothioate ligation for successive steps: An example of the use of two orthogonal chemistries tag ligation is the combination of 5'-azido/3'-alkynyl and 5'-iodo/3'-phosphorothioate ligations. Figure 23 shows an exemplary schematic of the synthesis of a 3-cycle orthogonal chemical ligation tagging strategy using these successive ligation chemistries. Figures 25A-25B show an example of the use of 3'-phosphorothioate/5'-azido and 3'-propargyl/5'-iodo tags to chemically ligate a succession of orthogonal encoding DNA tags that encode a chemical library covalently installed upon the 5'-terminus.

Protection of 3'-phosphorothioate on tags: As shown in Figure 25A, a 5'-azido-3'-phosphorothioate tag (1 eq.) was dissolved in water to give a final concentration of 5 mM. Subsequently, vinyl methyl sulfone (20 eq.) was added and the reaction was incubated at room temperature overnight. Upon completion of the reaction, the product was precipitated by ethanol.

Library synthesis (Figure 25B)

Cycle A: To each well in the split was added single stranded DNA headpiece (1 eq., 1 mM solution in 500 mM pH 9.5 borate buffer), one cycle A tag (1.5 eq.), and splint (1.2 eq.). The chemical ligation was incubated at room temperature overnight. To each well (in the split) was then added one Fmoc amino acid (100 eq.), followed by 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium

chloride (100 eq.). The chemical reaction was incubated at room temperature overnight. Upon completion, all wells were pooled and the products precipitated using ethanol. The cycle A pool was purified using LC and lyophilized to dryness. Fmoc deprotection was performed on cycle A pool by treating the pool (1mM in water) with piperidine (10% v/v) for 2h at room temperature. The deprotected product was precipitated again using ethanol.

Cycle B: The purified cycle A pool was dissolved in 500 mM, pH 7.0 phosphate buffer to give a 1 mM concentration and then split into separate reaction wells (1 eq. of cycle A product in each well). To each well was added one cycle B protected tag (1.2 eq.), copper (II) acetate (2 eq.), sodium ascorbate (4 eq.), and tris-(benzyltriazolylmethyl)amine (1 eq.). The chemical ligation was incubated at room temperature overnight. Upon completion, the products were precipitated (in the split) using ethanol and then diluted to a 1 mM concentration using 500 mM, pH 9.5 borate buffer. To each well (in the split) was then added a mixture of one formyl acid (100 eq.), diisopropyl carbodiimide (100 eq.), and 1-hydroxy-7-aza-benzotriazole (100 eq.). The chemical reaction was incubated at room temperature overnight. Upon completion, all wells were pooled and the products precipitated using ethanol. The cycle B pool was then dissolved in water to give a 1 mM final concentration, and piperidine (10% v/v) was added to perform the deprotection of cycle B tag (room temperature, 18h). The deprotected product was precipitated again using ethanol. The deprotected Cycle B pool was purified using LC and lyophilized to dryness.

Cycle C: The purified cycle B pool was dissolved in 500 mM, pH 5.5 phosphate buffer to give a 1 mM concentration and then split into separate reaction wells (1 eq. of cycle B product in each well). To each well was added one cycle C tag (1.5 eq.) and splint (1.2 eq.). The chemical ligation was incubated at room temperature overnight. To each well (in the split) was added an amine (80 eq.) and sodium cyanoborohydride (80 eq.). The chemical reaction was incubated at 60°C for 16h. Upon completion, all wells were pooled and the products precipitated using ethanol. The cycle C pool was purified using LC and lyophilized to dryness.

Other embodiments

All publications, patent applications, and patents mentioned in this specification are herein incorporated by reference.

Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific desired embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the fields of medicine, pharmacology, or related fields are intended to be within the scope of the invention.

Claims

1. A method of tagging a first library comprising an oligonucleotide-encoded chemical entity, said method comprising:

- (i) providing a headpiece having a first functional group and a second functional group, wherein said headpiece comprises at least one 2'-substituted nucleotide;
- (ii) binding said first functional group of said headpiece to a first component of said chemical entity, wherein said headpiece is directly connected to said first component or said headpiece is indirectly connected to said first component by a bifunctional linker; and
- (iii) binding said second functional group of said headpiece to a first building block tag to form a complex,

wherein said steps (ii) and (iii) can be performed in any order and wherein said first building block tag encodes for the binding reaction of said step (ii), thereby providing a tagged library.

2. The method of claim 1, wherein said headpiece comprises a 2'-substituted nucleotide at one or more of the 5'-terminus, the 3'-terminus, or the internal position of said headpiece.

3. The method of claim 2, wherein said headpiece comprises said 2'-substituted nucleotide and said second functional group at the 5'-terminus or at the 3'-terminus.

4. The method of any one of claims 1-3, wherein said first building block tag comprises at least one 2'-substituted nucleotide.

5. The method of claim 4, wherein said first building block tag comprises a 2'-substituted nucleotide at one or more of the 5'-terminus, the 3'-terminus, or the internal position of said first building block tag.

6. The method of claim 5, wherein said first building block tag comprises said 2'-substituted nucleotide at both of the 5'- and 3'-termini.

7. The method of any one of claims 1-6, wherein said 2'-O-substituted nucleotide is a 2'-O-methyl nucleotide or a 2'-fluoro nucleotide.

8. The method of claim 7, wherein said 2'-O-substituted nucleotide is said 2'-O-methyl nucleotide selected from 2'-O-methyl guanine or 2'-O-methyl uracil.
9. The method of claim 7, wherein 2'-O-substituted nucleotide is said 2'-fluoro nucleotide selected from 2'-fluoro guanine or 2'-fluoro uracil.
10. The method of any one of claims 1-9, wherein said first building block tag comprises a protecting group at the 3'-terminus or at the 5'-terminus.
11. The method of any one of claims 1-10, wherein step (ii) comprises binding said headpiece directly to said first component.
12. The method of claim 11, wherein said component is a scaffold or a first building block.
13. The method of any one of claims 1-10, wherein step (ii) comprises binding said headpiece indirectly to said first component via a bifunctional linker.
14. The method of claim 13, wherein said method comprises binding said headpiece with the first functional group of said linker and binding said first component with the second functional group of said linker.
15. The method of claim 14, wherein said first component is a scaffold or a first building block.
16. The method of any one of claims 1-15, further comprising
 - (iv) binding a second building block tag to the 5'-terminus or 3'-terminus of said complex;and
 - (v) binding a second component of said chemical library to said first component, wherein said steps (iv) and (v) can be performed in any order.
17. The method of claim 16, wherein step (iv) comprises binding said second building block tag to the 5'-terminus of said complex; said complex comprises a phosphate group at the 5'-terminus; and said second building block tag comprises a hydroxyl group at both of the 3'- and 5'-termini.

18. The method of claim 17, wherein step (iv) further comprises purifying said complex and reacting said complex with a polynucleotide kinase to form a phosphate group on the 5'-terminus prior to binding said second building block tag.

19. The method of claim 16, wherein step (iv) comprises binding said second building block tag to the 3'-terminus of said complex; said complex comprises a protecting group at the 3'-terminus; and said second building block tag comprises a phosphate group at the 5'-terminus and a protecting group at the 3'-terminus.

20. The method of claim 19, wherein step (iv) further comprises reacting said complex with a hydrolyzing agent to release said protecting group from said complex prior to binding said second building block tag to said complex.

21. The method of any one of claims 16-20, wherein said second building block tag comprises a 2'-substituted nucleotide at one or more of the 5'-terminus, the 3'-terminus, or the internal position of said second building block tag.

22. The method of any one of claims 16-21, wherein said second component is a first building block or a second building block.

23. The method of any one of claims 16-22, wherein said step (iv) comprises an RNA ligase and/or a DNA ligase to bind said second building block tag to said complex.

24. The method of claim 23, wherein said step (iv) comprises said RNA ligase and said RNA ligase is T4 RNA ligase.

25. The method of claim 23, wherein said step (iv) comprises said DNA ligase and said DNA ligase is ssDNA ligase

26. The method of claim 23, wherein said step (iv) comprises said RNA ligase and said DNA ligase.

27. The method of any one of claims 1-26, wherein said step (iii) comprises an RNA ligase and/or a DNA ligase to bind said headpiece to said first building block tag.

28. The method of claim 27, wherein said step (iii) comprises said RNA ligase and said RNA ligase is T4 RNA ligase.

29. The method of claim 27, wherein said step (iii) comprises said DNA ligase and said DNA ligase is ssDNA ligase.

30. The method of claim 27, wherein said step (iii) comprises said RNA ligase and said DNA ligase.

31. The method of any one of claims 1-30, wherein said step (iii) and/or said step (iv), if present, comprises poly ethylene glycol and/or one or more soluble multivalent cations.

32. The method of claim 31, wherein said step (iii) and/or said step (iv), if present, comprises said poly ethylene in an amount from about 25% (w/v) to about 35% (w/v).

33. The method of claim 32, wherein said poly ethylene glycol has an average molecular weight from about 3,000 to about 5,500 Daltons.

34. The method of claim 33, wherein said poly ethylene glycol has an average molecular weight of about 4,600 Daltons.

35. The method of any one of claims 31-34, wherein said step (iii) and/or said step (iv), if present, comprises said one or more soluble multivalent cations in an amount of from about 0.05 mM to about 10.5 mM.

36. The method of claim 35, wherein said one or more multivalent cations are selected from the group consisting of magnesium chloride, manganese (II) chloride, and hexamine cobalt (III) chloride.

37. The method of claim 35 or 36, wherein said one or more multivalent cations are in an amount of from about 0.5 mM to about 1.5 mM.

38. The method of any one of claims 1-37, wherein said method further comprises separating said complex from any unreacted tag or unreacted headpiece before any one of binding steps (ii)-(v).

39. The method of any one of claims 1-38, wherein said method further comprises purifying said complex before any one of binding steps (ii)-(v).

40. The method of any one of claims 1-39, wherein said method further comprises binding one or more additional building block tags to said complex and binding one or more additional components to said complex.

41. The method of any one of claims 1-40, wherein said headpiece, said first building block tag, said second building block tag, or said one or more additional building block tags, if present, comprises a preadenylated 5'-terminus.

42. The method of any one of claims 1-41, wherein said headpiece comprises a hairpin structure.

43. The method of any one of claims 1-42, wherein said headpiece, said first building block tag, said second building block tag, and/or said one or more additional building block tags, if present, comprises from 5 to 20 nucleotides.

44. The method of claim 43, wherein said first building block tag, said second building block tag, and/or said one or more additional building block tags, if present, has about the same mass.

45. The method of any one of claims 1-44, wherein said headpiece, said first building block tag, said second building block tag, and/or said one or more additional building block tags, if present, further comprises a first library-identifying sequence.

46. The method of any one of claims 1-44, wherein said method further comprises binding a first library-identifying tag to said complex.

47. The method of claim 45 or 46, further comprising providing a second library and combining said first library with said second library.

48. The method of any one of claims 1-47, wherein said headpiece, said first building block tag, said second building block tag, and/or said one or more additional building block tags, if present, further comprises a use sequence and/or an origin sequence.

49. The method of any one of claims 1-47, wherein said method further comprises binding a use tag and/or an origin tag to said complex.

50. The method of any one of claims 1-49, wherein said method further comprises binding a tailpiece to said complex.

51. A method of tagging a first library comprising an oligonucleotide-encoded chemical entity, said method comprising:

- (i) providing a headpiece having a first functional group and a second functional group, wherein said headpiece comprises a 2'-substituted nucleotide at the 5'-terminus, optionally one or more nucleotides at the internal position of said headpiece, and a protecting group at the 2'-position and/or the 3'-position at the 3'-terminus;
 - (ii) binding said first functional group of said headpiece to a first component of said chemical entity, wherein said headpiece is directly connected to said first component or said headpiece is indirectly connected to said first component by a bifunctional linker; and
 - (iii) binding said second functional group of said headpiece to a first building block tag, wherein said first building block tag comprises a 2'-substituted nucleotide and a hydroxyl group at the 5'-terminus, optionally one or more nucleotides at the internal position of said tag, and a 2'-substituted nucleotide and a hydroxyl group at the 3'-terminus;
- wherein said steps (ii) and (iii) can be performed in any order and wherein said first building block tag encodes for the binding reaction of said step (ii), thereby providing a tagged library.

52. The method of claim 51, wherein said 2'-substituted nucleotide is a 2'-O-methyl nucleotide of 2'-O-methyl guanine.

53. The method of claim 51 or claim 52, wherein said one or more nucleotides are present at the internal position of said headpiece.

54. The method of any one of claims 51-53, wherein said one or more nucleotides are present at the internal position of said tag.

55. The method of claim 53 or claim 54, wherein said one or more nucleotides of said headpiece and/or said tag are one or more 2'-deoxynucleotides.

56. The method of any one of claims 51-55, wherein said step (iii) comprises one or more soluble multivalent cations, poly ethylene glycol, and RNA ligase.

57. The method of claim 56, wherein said poly ethylene glycol has an average molecular weight of about 4,600 Daltons and said RNA ligase is T4 RNA ligase.

58. The method of claim 56 or claim 57, wherein said one or more multivalent cations are selected from the group consisting of magnesium chloride, manganese (II) chloride, and hexamine cobalt (III) chloride.

59. The method of any one of claims 51-57, wherein said headpiece and/or said first building block tag comprises from 5 to 20 nucleotides.

60. The method of claim 59, wherein said first building block tag, said second building block tag, and/or said one or more additional building block tags, if present, has about the same mass.

61. The method of any one of claims 51-60, wherein said headpiece, said first building block tag, said second building block tag, and/or said one or more additional building block tags, if present, further comprises a first library-identifying sequence.

62. The method of any one of claims 51-60, wherein said method further comprises binding a first library-identifying tag to said complex.

63. The method of claim 61 or 62, further comprising providing a second library and combining said first library with said second library.

64. The method of any one of claims 51-63, wherein said headpiece, said first building block tag, said second building block tag, and/or said one or more additional building block tags, if present, further comprises a use sequence or an origin sequence.

65. The method of any one of claims 51-63, wherein said method further comprises binding a use tag and/or an origin tag to said complex.

66. The method of any one of claims 51-65, wherein said method further comprises binding a tailpiece to said complex.

67. The method of claim 1 or 51, wherein said 2'-substituted nucleotide is selected from the group consisting of 2'-O-methyl guanine, 2'-O-methyl uracil, 2'-O-methyl adenosine, 2'-O-methyl thymidine, 2'-O-methyl inosine, 2'-O-methyl cytidine, 2'-O-methyl diamino purine, 2'-fluoro guanine, 2'-fluoro uracil, 2'-fluoro adenosine, 2'-fluoro thymidine, 2'-fluoro inosine, 2'-fluoro cytidine, and 2'-fluoro diamino purine.

68. The method of any one of claims 1-67, wherein said method comprises a plurality of headpieces.

69. The method of claim 68, wherein each headpiece of said plurality of headpieces comprises an identical sequence region and a different encoding region.

70. The method of claim 69, wherein said identical sequence region is a primer binding region.

71. The method of claim 69 or claim 70, wherein said different encoding region is an initial building block tag that encodes for said headpiece or for an addition of an initial component.

72. The method of any one of claims 1-71, wherein said binding in at least one of steps (ii)-(iv), if present, comprises enzymatic ligation.

73. The method of claim 72, wherein said enzymatic ligation comprises use of an RNA ligase and/or a DNA ligase.

74. The method of any one of claims 1-73, wherein said binding in at least one of steps (ii)-(iv), if present, comprises chemical ligation.

75. The method of claim 74, wherein said chemical ligation comprises use of one or more chemically co-reactive pairs.

76. The method of claim 75, wherein said chemically co-reactive pair is an optionally substituted alkynyl group and an optionally substituted azido group.

77. The method of claim 75, wherein said chemically co-reactive pair is a phosphorothioate group and an iodo group.

78. The method of claim 77, wherein said phosphorothioate group is at the 5'-terminus of an oligonucleotide and said iodo group is at the 3'-terminus of an oligonucleotide.

79. The method of claim 77 or 78, wherein said chemical ligation further comprises a splint oligonucleotide in the binding reaction between said chemically co-reactive pair.

80. The method of any one of claims 75-79, wherein said chemically co-reactive pair produces a spacer having a length from about 4 to about 24 atoms.

81. The method of claim 80, wherein said spacer has a length from about 4 to about 10 atoms.

82. The method of any one of claims 74-81, wherein said chemical ligation of one or more building block tags comprises orthogonal chemically co-reactive pairs for ligating successive building block tags.

83. The method of claim 82, wherein said orthogonal chemically co-reactive pairs comprises an (i) optionally substituted alkynyl group and an optionally substituted azido group and (ii) a phosphorothioate group and an iodo group.

84. The method of any one of claims 1-83, wherein said complex, said headpiece, said first building block tag, said second building block tag, and/or said one or more additional building block tags, if present, comprises a modified phosphate group between the terminal nucleotide at the 3'-terminus and the nucleotide adjacent to said terminal nucleotide.

85. A complex comprising a headpiece and a building block tag, wherein said tag comprises from 5 to 20 nucleotides, a 2'-substituted nucleotide at the 5'-terminus, and a 2'-substituted nucleotide at the 3'-terminus.

86. The complex of claim 85, wherein said 2'-substituted nucleotide at the 5'-terminus and/or 3'-terminus is a 2'-O-methyl nucleotide or a 2'-fluoro nucleotide.

87. The complex of claim 85 or 86, wherein said headpiece comprises a hairpin structure.

88. The complex of any one of claims 85-87, wherein said headpiece comprises a 2'-substituted nucleotide at one or more of the 5'-terminus, the 3'-terminus, or the internal position of said headpiece.

89. The complex of any one of claims 85-88, wherein said headpiece further comprises a preadenylated 5'-terminus.

90. The complex of any one of claims 85-89, wherein said 2'-substituted nucleotide is a 2'-O-methyl nucleotide or a 2'-fluoro nucleotide.

91. The complex of claim 90, wherein said 2'-substituted nucleotide is selected from the group consisting of 2'-O-methyl guanine, 2'-O-methyl uracil, 2'-O-methyl adenosine, 2'-O-methyl thymidine, 2'-O-methyl inosine, 2'-O-methyl cytidine, 2'-O-methyl diamino purine, 2'-fluoro guanine, 2'-fluoro uracil, 2'-fluoro adenosine, 2'-fluoro thymidine, 2'-fluoro inosine, 2'-fluoro cytidine, and 2'-fluoro diamino purine.

92. The complex of any one of claims 85-91, further comprising a tailpiece.

93. The complex of any one of claims 85-92, wherein said complex, said headpiece, or said building block tag comprises a modified phosphate group between the terminal nucleotide at the 3'-terminus and the nucleotide adjacent to said terminal nucleotide.

94. The complex of any one of claims 85-93, wherein said complex comprises a phosphorothioate or a triazole group.