METHODS FOR TAGGING DNA-ENCODED LIBRARIES

Cross-reference to Related Applications

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

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 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 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 improved, more robust methods of screening and identifying small compounds in DNA-encoded libraries.

Summary

The present disclosure 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 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.

In one aspect, there is provided a method of tagging a library comprising an oligonucleotide-encoded small molecule or peptide, said method comprising:

(i) providing a single-stranded oligonucleotide headpiece having a first functional group and a second functional group;

(ii) reacting said first functional group of said headpiece to a first building block of said small molecule or peptide, wherein said headpiece is directly connected to said first building block or said headpiece is indirectly connected to said first building block by a bifunctional linker; and

(iii) ligating said second functional group of said headpiece to a single-stranded oligonucleotide first building block tag, wherein said ligating comprises chemical ligation of one or more chemically co-reactive pairs selected from

(a) a phosphorothioate group and an iodo group; or

(b) an alkyne group and an azido group,

thereby forming a complex,

wherein steps (ii) and (iii) can be performed in any order; and

wherein said first building block tag encodes for the reaction of step (ii).

Accordingly, the disclosure 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 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 –$(CH_2CH_2O)_nCH_2CH_2$-, 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 examples, 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 examples, the headpiece includes the 2'-substituted nucleotide and the second functional group at the 5'-terminus or at the 3'-terminus.

In other examples, the first building block tag includes at least one (e.g., at least two, three, four, five, or more) 2'-substituted nucleotides. In particular examples, 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 examples, the first building block tag includes a protecting group at the 3'-terminus or at the 5'-terminus.

In any of the examples 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 examples, 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 examples, 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 examples, 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 examples, the second building block tag encodes for the binding reaction of step (v). In other examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 example, the multivalent cation is in the form of hexamine cobalt (III) chloride.

In other examples, the method further includes separating the complex from any unreacted tag or unreacted headpiece before any one of binding steps (ii)-(v). In other examples, the method further includes purifying the complex before any one of binding steps (ii)-(v). In other examples, 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 disclosure 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 examples, 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 examples, one or more nucleotides at the internal position of the headpiece are 2'-deoxynucleotides. In yet other examples, the bifunctional linker is a poly ethylene glycol linker (e.g., $-(CH_2CH_2O)_n CH_2CH_2-$, where n is an integer from 1 to 50).

In other examples, 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 examples, 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).

Disclosed are 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 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 disclosure 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 examples, the 2'-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 examples, the headpiece includes a hairpin structure. In some examples, 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 examples, the headpiece further includes a preadenylated 5'-terminus. In yet other examples, the headpiece includes from 5 to 20 nucleotides.

In any of the above examples, 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 examples, the method further includes binding one or more (e.g., one, 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 examples, the 2'-substituted nucleotide is a 2'-O-methyl nucleotide, such 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 examples, 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 diamino purine.

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

In any of the above examples, the method includes a plurality of headpieces. In some examples of this method, each headpiece of the plurality of headpieces includes an identical sequence region and a different encoding region. In particular examples, the identical sequence region is a primer binding region. In other examples, 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 examples, binding in at least one of steps (ii)-(iv), if present, includes enzyme ligation and/or chemical ligation. In some examples, enzymatic ligation includes use of an RNA ligase (e.g., T4 RNA ligase) or a DNA ligase (e.g., ssDNA ligase). In other examples, enzymatic ligation includes use of an RNA ligase (e.g., T4 RNA ligase) and a DNA ligase (e.g., ssDNA ligase). In some examples, 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 examples, 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 examples, 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 examples, chemical ligation includes a splint oligonucleotide in the binding reaction. In some examples, 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 examples, 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 examples, the headpiece may include a single-stranded (e.g., hairpin) structure.

In any of the above examples, 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 examples, 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 examples, 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 examples, one or more of the single-stranded molecules have a hairpin structure. In particular examples, 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 examples, 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 examples, the method further comprises identifying a small drug-like library member that binds or inactivates a protein of therapeutic interest. In other examples, 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 examples 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, the oligonucleotide is a first library-identifying tag. In some examples, 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 examples, the method includes providing a second library and combining the first library with a second library. In further examples, 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 examples, 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 examples, 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 examples, the oligonucleotide is a use tag including a use sequence. In some examples, 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 examples, 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 examples, the oligonucleotide is an origin tag including an origin sequence. In some examples, the method further includes joining, binding, or operatively associating a use tag and/or an origin tag to the complex.

In any of the above examples, 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 examples, 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.

In any of the above examples, 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 examples, the modified phosphate group minimizes 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 building block tag) lacking the modified phosphate group. In some examples, the complex may include a phosphorothioate or a triazole group.

In any of the above examples, 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-, reduced-, or non-aqueous (e.g., organic) conditions. In some examples, 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 examples, 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 examples, the modification includes one or more modified 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-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 examples, 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 examples, 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 examples, 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).

In particular examples, the first building block tag and the second building block tag include the same number of nucleotides. In other examples, 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 examples, the first building block tag is a donor 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 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).

Definitions

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

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 examples, 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

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).

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 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 a 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 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-phenylxanthen-9-yl and 9-(p-methoxyphenyl)xanthen-9-yl), acyl groups (e.g., phenoxyacetyl and acetyl), and silyl groups (e.g., t-butyldimethylsilyl).

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 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 examples, the 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.

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

CLAIMS

1. A method of tagging a library comprising an oligonucleotide-encoded small molecule or peptide, said method comprising:

(i) providing a single-stranded oligonucleotide headpiece having a first functional group and a second functional group;

(ii) reacting said first functional group of said headpiece to a first building block of said small molecule or peptide, wherein said headpiece is directly connected to said first building block or said headpiece is indirectly connected to said first building block by a bifunctional linker; and

(iii) ligating said second functional group of said headpiece to a single-stranded oligonucleotide first building block tag, wherein said ligating comprises chemical ligation of one or more chemically co-reactive pairs selected from:

(a) a phosphorothioate group and an iodo group; or

(b) an alkyne group and an azido group,

thereby forming a complex,

wherein steps (ii) and (iii) can be performed in any order; and

wherein said first building block tag encodes for the reaction of step (ii).

2. The method of claim 1, wherein the headpiece is indirectly connected to the first building block by a bifunctional linker.

3. The method of claim 1 or 2, wherein the method further comprises:

(iv) ligating a second building block tag to the 5'-terminus or 3'-terminus of the complex; and

(v) reacting a second building block of the small molecule or peptide to the first building block of the small molecule or peptide,

wherein steps (iv) and (v) can be performed in any order; and

wherein said ligating of step (iv) comprises chemical ligation of one or more chemically co-reactive pairs selected from:

(a) a phosphorothioate group and an iodo group; or

(b) an alkyne group and an azido group.

4. The method of claim 3, wherein the chemical ligation further comprises the use of a splint oligonucleotide.

5. The method of claim 1 or 2, wherein the method further comprises:

(iv) ligating one or more additional building block tags to the 5'-terminus or 3'-terminus of the complex; and

(v) reacting one or more additional building blocks of the small molecule or peptide to the first building block of the small molecule or peptide,

wherein steps (iv) and (v) can be performed in any order; and

wherein said ligating of step (iv) comprises chemical ligation of one or more chemically co-reactive pairs selected from:

(a) a phosphorothioate group and an iodo group; or

(b) an alkyne group and an azido group.

6. The method of claim 5, wherein the chemical ligation further comprises the use of a splint oligonucleotide.

7. The method of any one of claims 3 to 6, wherein the method further comprises separating the complex from any unreacted tag or unreacted headpiece prior to steps (iv) or (v).

8. The method of any one of claims 3 to 7, wherein the method further comprises purifying the complex prior to steps (iv) or (v).

9. The method of any one of claims 1 to 8, wherein the chemical ligation comprises a phosphorothioate group and an iodo group.

10. The method of claim 9, wherein the phosphorothioate group is at the 5'-terminus of an oligonucleotide and the iodo group is at the 3'-terminus of an oligonucleotide.

11. The method of claim 9, wherein the phosphorothioate group is at the 3'-terminus of an oligonucleotide and the iodo group is at the 5'-terminus of an oligonucleotide.

12. The method of claim 9, wherein the second functional group of the oligonucleotide headpiece is a phosphorothioate group.

13. The method of any one of claims 1 to 12, wherein the chemical ligation comprises an alkyne group and an azido group.

14. The method of claim 13, wherein the alkyne group is at the 5'-terminus of an oligonucleotide and the azido group is at the 3'-terminus of an oligonucleotide.

15. The method of claim 13, wherein the alkyne group is at the 3'-terminus of an oligonucleotide and the azido group is at the 5'-terminus of an oligonucleotide.

16. The method of claim 13, wherein the second functional group of the oligonucleotide headpiece is an alkyne group.

17. The method of any one of claims 1 to 16, wherein the headpiece comprises a hairpin oligonucleotide.

18. The method of any one of claims 1 to 17, wherein the headpiece, the first building block tag, the second building block tag, and/or the one or more additional building block tags, if present, comprises from 5 to 20 nucleotides.

19. The method of claim 18, wherein the headpiece, the first building block tag, the second building block tag, and/or the one or more additional building block tags, if present, has about the same mass.

20. The method of any one of claims 1 to 19, wherein the method further comprises binding a tailpiece to the complex.