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 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 small molecule or peptide, 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) ligating said second functional group of said headpiece to a first building block tag to form a complex, wherein said ligating comprises chemical ligation,

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

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

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 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 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., $-(CH_2CH_2O)_n CH_2CH_2-$, 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).

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 embodiments, 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 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 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, 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 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 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 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.

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 disclosure 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.

CLAIMS

1. A method of tagging a library comprising an oligonucleotide-encoded small molecule or peptide, 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 small molecule or peptide, 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) ligating said second functional group of said headpiece to a first building block tag to form a complex, wherein said ligating comprises chemical ligation,

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

2. The method of claim 1, wherein said headpiece is indirectly connected to said first component by a bifunctional linker.

3. The method of claim 1 or 2, wherein said method comprises separating the complex from any unreacted tag or unreacted headpiece prior to steps (ii) or (iii).

4. The method of any one of claims 1 to 3, wherein said method comprises purifying the complex prior to steps (ii) or (iii).

5. The method of any one of claims 1 to 4, 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) binding one or more additional components of said small molecule or peptide,

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

6. The method of claim 5, wherein said ligating comprises chemical ligation.

7. The method of claim 5 or 6, wherein said method comprises separating the complex from any unreacted tag or unreacted headpiece prior to steps (ii) or (iii).

8. The method of any one of claims 5 to 7, wherein said method comprises purifying the complex prior to steps (ii) or (iii).

9. The method of any one of claims 1 to 8, wherein said chemical ligation comprises the use of one or more chemically co-reactive pairs.

10. The method of claim 9, wherein said one or more chemically co-reactive pairs is an optionally substituted alkynyl group and an optionally substituted azido group; an optionally substituted diene having a 4 π-electron system and an optionally substituted dienophile or an optionally substituted heterodienophile having a 2 π-electron system; a nucleophile and a strained heterocyclyl electrophile; a phosphorothioate group and an iodo group; or an amino group and an aldehyde.

11. The method of any one of claims 1 to 10, wherein the headpiece comprises a hairpin oligonucleotide.

12. The method of any one of claims 1 to 11, wherein the headpiece or a building block tag encodes for the identity of the library.

13. The method of any one of claims 1 to 12, wherein the headpiece or a building block tag encodes for the use of the member of the library.

14. The method of any one of claims 1 to 13, wherein the method further comprises:
(vi) binding a tailpiece to the complex.

15. The method of claim 14, wherein the tailpiece encodes the identity of the library, the use of the member of the library, or the origin of the member of the library.

16. The method of any one of claims 1 to 15, wherein the complex comprises a modification that supports solubility in organic conditions.