



- (51) International Patent Classification:
C40B 40/06 (2006.01)
- (21) International Application Number:
PCT/US2013/050303
- (22) International Filing Date:
12 July 2013 (12.07.2013)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
61/671,406 13 July 2012 (13.07.2012) US
- (71) Applicant: X-CHEM, INC. [US/US]; 100 Beaver Street,
Suite 101, Waltham, MA 02453 (US).
- (72) Inventors: KEEFE, Anthony, D.; 9 Bellis Circle, Apt. 6,
Cambridge, 02140 (US). LITOVCHICK, Alexander; 468
Dutton Road, Sudbury, MA 01776 (US). CLARK, Mat-
thew, A.; 36 Clelland Road, Lexington, MA 02421 (US).
- (74) Agent: ELBING, Karen, L.; Clark & Elbing LLP, 101
Federal Street, 15th Floor, Boston, MA 02110 (US).
- (81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC,
SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))



WO 2014/012010 A1

(54) Title: DNA-ENCODED LIBRARIES HAVING ENCODING OLIGONUCLEOTIDE LINKAGES NOT READABLE BY POLYMERASES

(57) Abstract: The present invention relates to complexes of oligonucleotide-encoded libraries and methods of tagging and using such libraries. In particular, the oligonucleotides and methods can include complexes having at least one linkage for which a polymerase has reduced ability to read or translocate through.

**DNA-ENCODED LIBRARIES HAVING ENCODING OLIGONUCLEOTIDE LINKAGES NOT
READABLE BY POLYMERASES**

5

Cross-Reference to Related Applications

This application claims benefit of U.S. Provisional Application No. 61/671,406, filed July 13, 2012, which is hereby incorporated by reference in its entirety.

10

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 oligonucleotide 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 encoding oligonucleotide tag ligation using robust and rapid methodologies. To create libraries under diverse reaction conditions, stable oligonucleotide 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 oligonucleotide-encoded libraries.

20

25

Summary of the Invention

The present invention features complexes for use in DNA-encoded libraries, where these complexes have one or more linkages for which a polymerase has reduced ability to read or translocate through. In order to discover the identity of encoded regions (e.g., headpiece, one or more tags, and/or tailpiece), the linkage is either reversed before it is exposed to the polymerase or alternatively bypassed by the polymerase. This linkage can be bypassed in any useful manner that may not capture all of the sequence information of the tags (e.g., such as in the 5'- or 3'- connectors, as described herein) but captures the encoding sequence information of the tags. Such linkages could result in lower incidences of mis-tagging, expand the number and kind of linkages to be used in making complexes and libraries, and provide non-enzymatic methods for creating and screening DNA-encoded complexes. Additional advantages of these complexes and methods are described herein.

30

35

Accordingly, in one aspect, the invention features a complex including: a chemical entity including one or more scaffolds or one or more building blocks; a first oligonucleotide tag encoding the identity of at least one of the one or more scaffolds or building blocks; and a headpiece having a first functional group and a second functional group, where the first functional group operatively associates with the chemical entity and the second functional

40

group operatively associates with the first tag via a first linkage for which a polymerase has reduced ability to read or translocate through.

In another aspect, the invention features a complex including: a chemical entity including one or more scaffolds or one or more building blocks; n number of oligonucleotide tags having $n-1$ linkages, where n is an integer between 1 and about 10, and where each of the linkages is between two adjacent tags and each tag encodes the identity of at least one of the one or more scaffolds or building blocks; and a headpiece having a first functional group operatively associated with the chemical entity and a second functional group operatively associated with at least one of the n number of tags via a first linkage, where a polymerase has reduced ability to read or translocate through at least one of the first linkage and $n-1$ linkages.

In some embodiments, n is between 1 and 2, 1 and 3, 1 and 4, 1 and 5, 1 and 6, 1 and 7, 1 and 8, 1 and 9, 1 and 10, 1 and 12, 1 and 15, 1 and 18, 1 and 20, 2 and 3, 2 and 4, 2 and 5, 2 and 6, 2 and 7, 2 and 8, 2 and 9, 2 and 10, 2 and 12, 2 and 15, 2 and 18, 2 and 20, 3 and 4, 3 and 5, 3 and 6, 3 and 7, 3 and 8, 3 and 9, 3 and 10, 3 and 12, 3 and 15, 3 and 18, 3 and 20, 4 and 5, 4 and 6, 4 and 7, 4 and 8, 4 and 9, 4 and 10, 4 and 12, 4 and 15, 4 and 18, or 4 and 20.

In some embodiments, the polymerase has reduced ability to read or translocate through at least about 10% (e.g., about 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or even 100%, as compared to control) of the first linkage and $n-1$ linkages. In particular embodiments, the polymerase has reduced ability to read or translocate through about 10% to about 100% of the first linkage and $n-1$ linkages (e.g., 20% to 100%, 25% to 100%, 50% to 100%, 75% to 100%, 90% to 100%, 95% to 100%, 10% to 95%, 20% to 95%, 25% to 95%, 50% to 95%, 75% to 95%, 90% to 95%, 10% to 90%, 20% to 90%, 25% to 90%, 50% to 90%, or 75% to 90%, as compared to control (e.g., as compared to a control oligonucleotide lacking the linkage)).

In some embodiments, one or more tags include a 5'-connector at the 5'-terminus of the one or more tags and a 3'-connector at the 3'-terminus of the one or more tags. In particular embodiments, each 5'-connector and/or each 3'-connector includes the same sequence. In other embodiments, each 5'-connector and/or each 3'-connector include a different sequence.

In some embodiments, the first linkage and/or $n-1$ linkages are at least about 3 angstroms in length (e.g., at least about 5, 8, 10, 15, 20, 30, 35, 40, 50, 60, 65 or 70 angstroms). In some embodiments, the first linkage and/or $n-1$ linkages are less than about 30 angstroms in length (e.g., less than about 25, 20, 15, or 10 angstroms).

In some embodiments, a DNA polymerase and/or an RNA polymerase (e.g., as described herein) has reduced ability to read or translocate through the first linkage and/or $n-1$ linkages.

In some embodiments, less than about 10% (e.g., about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%) of the first linkage and $n-1$ linkages include an enzymatic linkage. In some embodiments, the first linkage and $n-1$ linkages include between 0% to 90% enzymatic linkages (e.g., about 0% to 40%, 0% to 45%, 0% to 50%, 0% to 55%, 0% to 60%, 0% to 65%, 0% to 70%, 0% to 75%, 0% to 80%, 0% to 85%, 0% to 90%, 0% to 95%, 0% to 96%, 0% to 97%, 0% to 98%, 0% to 99%, 5% to 40%, 5% to 45%, 5% to 50%, 5% to 55%, 5% to 60%, 5% to 65%, 5% to 70%, 5% to 75%, 5% to 80%, 5% to 85%, 5% to 90%, 5% to 95%, 5% to 96%, 5% to 97%, 5% to 98%, 5% to 99%, 10% to 40%, 10% to 45%, 10% to 50%, 10% to 55%, 10% to 60%, 10% to 65%, 10% to 70%, 10% to 75%, 10% to 80%, 10% to 85%, 10% to 90%, 10% to 95%, 10% to 96%, 10% to 97%, 10% to 98%, 10% to 99%, 15% to 40%, 15% to 45%, 15% to 50%, 15% to 55%, 15% to 60%, 15% to 65%, 15% to 70%, 15% to 75%, 15% to 80%, 15% to 85%, 15% to 90%, 15% to 95%, 15% to 96%, 15% to 97%, 15% to 98%, 15% to 99%, 20% to 40%, 20% to 45%, 20% to 50%, 20% to 55%, 20% to 60%, 20% to

65%, 20% to 70%, 20% to 75%, 20% to 80%, 20% to 85%, 20% to 90%, 20% to 95%, 20% to 96%, 20% to 97%, 20% to 98%, or 20% to 99%).

In some embodiments, the first linkage and/or *n-1* linkages include a chemical linkage (e.g., a chemical-reactive group, a photo-reactive group, an intercalating moiety, or a cross-linking oligonucleotide). In particular
5
embodiments, at least one (e.g., two, three, four, five, or more) chemical-reactive group, photo-reactive group, or intercalating moiety is present in a 5'-connector at or in proximity to the 5'-terminus of the tag and/or in a 3'-connector at or in proximity to the 3'-terminus of the tag. In other embodiments, the sequence of at least one of the 5'-connector is complementary to the sequence of the adjacent 3'-connector or identical or sufficiently similar to allow for hybridization to a complementary oligonucleotide. In some embodiments, at least 10% (e.g., about 20%,
10
25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or even 100%) of the first linkage and/or *n-1* linkages are chemical linkages. In other embodiments, about 10% to about 100% of the first linkage and *n-1* linkages (e.g., 20% to 100%, 25% to 100%, 50% to 100%, 75% to 100%, 90% to 100%, 95% to 100%, 10% to 95%, 20% to 95%, 25% to 95%, 50% to 95%, 75% to 95%, 90% to 95%, 10% to 90%, 20% to 90%, 25% to 90%, 50% to 90%, or 75% to 90%) are chemical linkages.

In some embodiments, the chemical-reactive group is selected from a pair of an optionally substituted
15
alkynyl group and an optionally substituted azido group; a pair of 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 pair of a nucleophile and a strained heterocyclyl electrophile; a pair of an optionally substituted amino group and an aldehyde or a ketone group; a pair of an optionally substituted amino group and a carboxylic
20
acid group; a pair of an optionally substituted hydrazine and an aldehyde or a ketone group; a pair of an optionally substituted hydroxylamine and an aldehyde or a ketone group; a pair of a nucleophile and an optionally substituted alkyl halide; a platinum complex; an alkylating agent; or a furan-modified nucleotide (e.g., any described herein).

In some embodiments, the photo-reactive group includes an intercalating moiety, a psoralen derivative, an optionally substituted cyanovinylcarbazole group (e.g., a 3-cyanovinylcarbazole group, such as 3-
25
cyanovinylcarbazole-1'- β -deoxyribose-5'-triphosphate), an optionally substituted vinylcarbazole group (e.g., a amidovinylcarbazole group, a carboxyvinylcarbazole group, or a C₂₋₇ alkoxy carbonyl vinylcarbazole group, as described herein), an optionally substituted cyanovinyl group, an optionally substituted acrylamide group, an optionally substituted diazirine group, an optionally substituted benzophenone, or an optionally substituted azide group (e.g., any described herein).

In some embodiments, the intercalating moiety is a psoralen derivative (e.g., psoralen, 8-methoxypsoralen, or 4-hydroxymethyl-4,5,8-trimethyl-psoralen (HMT-psoralen)), an alkaloid derivative (e.g., berberine, palmatine, coralyne, sanguinarine (e.g., iminium or alkanolamine forms thereof, or aristololactam- β -D-glucoside), an ethidium
30
cation (e.g., ethidium bromide), an acridine derivative (e.g., proflavine, acriflavine, or amsacrine), an anthracycline derivative (e.g., doxorubicin, epirubicin, daunorubicin (daunomycin), idarubicin, and aclarubicin), or thalidomide.

In some embodiments, the chemical linkage includes the cross-linking oligonucleotide, where the sequence
35
of at least five nucleotides at the 5'-terminus of the cross-linking oligonucleotide is complementary to the sequence of at least five nucleotides at the 3'-terminus of one or more tags or identical or sufficiently similar to allow for hybridization to a complementary oligonucleotide, and where the sequence of at least five nucleotides at the 3'-terminus of the cross-linking oligonucleotide is complementary to the sequence of at least five nucleotides at the 5'-
40
terminus of one or more tags or identical or sufficiently similar to allow for hybridization to a complementary oligonucleotide. In particular embodiments, the 3'-terminus of one or more tags includes a 3'-connector. In particular embodiments, the 5'-terminus of one or more tags includes a 5'-connector.

In some embodiments, the 5'-terminus and/or 3'-terminus of the cross-linking oligonucleotide include a reversible co-reactive group (e.g., a cyanovinylcarbazole group, a cyanovinyl group, an acrylamide group, a thiol group, or a vinyl sulfone group, as described herein).

5 In some embodiments, the 3'-connector and/or 5'-connector include a reversible co-reactive group (e.g., a cyanovinylcarbazole group, a cyanovinyl group, an acrylamide group, a thiol group, or a vinyl sulfone group, as described herein).

In some embodiments, the chemical entity is operatively associated to the headpiece via a bifunctional spacer (e.g., any described herein). In other embodiments, the chemical entity is covalently attached to the headpiece. In particular embodiments, the headpiece includes an oligonucleotide selected from the group consisting of a double-stranded oligonucleotide, a single-stranded oligonucleotide, or a hairpin oligonucleotide. In some
10 embodiments, the headpiece includes a primer-binding region.

In any of the embodiments described herein, the complex or method further includes one or more first library-identifying tag(s), use tag(s), and/or origin tag(s). In some embodiments, the complex or method includes between 2 to 20 tags (e.g., between 2 to 10 building block or scaffold tags, one first library-identifying tag, one
15 optional use tag, and one origin tag). In some embodiments, the complex or method includes from 5 to 75 nucleotides (e.g., as described herein, such as about 40 or 50 nucleotides).

In any of the embodiments described herein, each of the tags within an individual tag set includes about the same mass.

In any of the embodiments described herein, the complex includes RNA, DNA, modified DNA, and/or
20 modified RNA (e.g., PNA, LNA, GNA, TNA, or a mixture thereof within the same oligonucleotide, as described herein).

In any of the embodiments described herein, the complex or method further includes a tailpiece.

In another aspect, the invention features a library including one or more complexes described herein. In some embodiments, the library includes a plurality of headpieces. In other embodiments, each headpiece of the
25 plurality of headpieces includes an identical sequence region (e.g., a primer-binding region) and a different encoding region (e.g., a first tag that encodes for use of the library, origin of the library, identity of the library, history of the library, a linkage, a spacer, or addition of a first component or an oligonucleotide sequence that facilitates hybridization, amplification, or sequencing technologies). In particular embodiments, the library includes between
30 about 10^2 to 10^{20} complexes (e.g., about 10^2 to 10^3 , 10^2 to 10^4 , 10^2 to 10^5 , 10^2 to 10^6 , 10^2 to 10^7 , 10^2 to 10^8 , 10^2 to 10^9 , 10^2 to 10^{10} , 10^2 to 10^{11} , 10^2 to 10^{12} , 10^2 to 10^{13} , 10^2 to 10^{14} , 10^2 to 10^{15} , 10^2 to 10^{16} , 10^2 to 10^{17} , 10^2 to 10^{18} , 10^2 to 10^{19} , 10^4 to 10^5 , 10^4 to 10^6 , 10^4 to 10^7 , 10^4 to 10^8 , 10^4 to 10^9 , 10^4 to 10^{10} , 10^4 to 10^{11} , 10^4 to 10^{12} , 10^4 to 10^{13} , 10^4 to 10^{14} , 10^4 to 10^{15} , 10^4 to 10^{16} , 10^4 to 10^{17} , 10^4 to 10^{18} , 10^4 to 10^{19} , 10^4 to 10^{20} , 10^5 to 10^6 , 10^5 to 10^7 , 10^5 to 10^8 , 10^5 to 10^9 , 10^5 to 10^{10} , 10^5 to 10^{11} , 10^5 to 10^{12} , 10^5 to 10^{13} , 10^5 to 10^{14} , 10^5 to 10^{15} , 10^5 to 10^{16} , 10^5 to 10^{17} , 10^5 to 10^{18} , 10^5 to 10^{19} , or 10^5 to 10^{20} complexes). In some embodiments, each complex is different.

35 In another aspect, the invention features a method of tagging a first library including an encoded chemical entity, the method including: (a) providing a headpiece having a first functional group and a second functional group (e.g., where the headpiece optionally encodes information); (b) 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 spacer; and (c) binding the second
40 functional group of the headpiece to a first oligonucleotide tag via a first linkage to form a complex, where a polymerase has reduced ability to read or translocate through the first linkage; where the steps (b) and (c) can be

performed in any order and where the first tag encodes for the binding reaction of the step (b), thereby providing a tagged library.

In yet another aspect, the invention features a method of tagging a first library including an encoded chemical entity, the method including: (a) providing a headpiece having a first functional group and a second functional group (e.g., where the headpiece optionally encodes information); (b) 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 spacer; (c) binding the second functional group of the headpiece to a first oligonucleotide tag via a first linkage; (d) binding n_c number of additional components of the chemical entity, where n_c is an integer between 1 and 10; and (e) binding n_t number of additional oligonucleotide tags having n_t linkages to form a complex, where n_t is an integer between 1 and 10, where each of the linkages is between two adjacent tags and where each tag encodes the identity of at least one of the components; where a polymerase has reduced ability to read or translocate through at least one of the first linkage and n_t linkages; and where the steps (b) and (c) can be performed in any order and where the first tag encodes for the binding reaction of the step (b); where the steps (d) and (e) can be performed in any order and where each additional tag encodes for the binding reaction of each additional component of the step (d), thereby providing a tagged library.

In some embodiments, n_c and n_t are each independently an integer between 1 and 2, 1 and 3, 1 and 4, 1 and 5, 1 and 6, 1 and 7, 1 and 8, 1 and 9, 1 and 10, 1 and 12, 1 and 15, 1 and 18, 1 and 20, 2 and 3, 2 and 4, 2 and 5, 2 and 6, 2 and 7, 2 and 8, 2 and 9, 2 and 10, 2 and 12, 2 and 15, 2 and 18, 2 and 20, 3 and 4, 3 and 5, 3 and 6, 3 and 7, 3 and 8, 3 and 9, 3 and 10, 3 and 12, 3 and 15, 3 and 18, 3 and 20, 4 and 5, 4 and 6, 4 and 7, 4 and 8, 4 and 9, 4 and 10, 4 and 12, 4 and 15, 4 and 18, or 4 and 20.

In some embodiments, the polymerase has reduced ability to read or translocate through at least about 10% of the first linkage and n_t linkages (e.g., about 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or even 100%, as compared to control, of the first linkage and n_t linkages). In particular embodiments, the reduced ability to read or translocate through the first linkage and n_t linkages is between about 10 to about 100% (e.g., 20% to 100%, 25% to 100%, 50% to 100%, 75% to 100%, 90% to 100%, 95% to 100%, 10% to 95%, 20% to 95%, 25% to 95%, 50% to 95%, 75% to 95%, 90% to 95%, 10% to 90%, 20% to 90%, 25% to 90%, 50% to 90%, or 75% to 90%, as compared to control (e.g., as compared to a control oligonucleotide lacking the linkage)).

In some embodiments, the first component and/or additional components include a scaffold or a building block.

In some embodiments, step (b) includes binding the headpiece indirectly to the first component via a bifunctional spacer.

In some embodiments, the methods further include binding a use tag, an origin tag, and/or a first library-identifying tag to the complex. In particular embodiments, the method further includes providing a second library and combining the first library with the second library. In some embodiments, the method further includes binding a tailpiece to the complex.

In some embodiments, the binding step includes a chemical-reactive group, a photo-reactive group, an intercalating moiety, or a cross-linking oligonucleotide (e.g., as described herein for any of the complexes herein). In particular embodiments, the method further includes performing split-and-pool synthesis in one or more steps, separating one or more complexes, and/or purifying one or more complexes.

In any of the methods described herein, the method results in any of the complexes described herein or any of the libraries described herein.

In yet another aspect, the invention features a method of screening a plurality of chemical entities, the method including: (a) contacting a target with any complex described herein and/or a library described herein; and (b) selecting one or more complexes having a predetermined characteristic for the target, as compared to a control, thereby screening the chemical entity.

5 In some embodiments, the predetermined characteristic includes increased binding for the target (e.g., a biological target, as described herein), as compared to a control.

In some embodiments, prior to step (a) one or more operations are performed selected from the list consisting of: annealing one or more relay primers with said complex, wherein said one or more relay primers span said first linkage and/or *n-1* linkages; extending said relay primers using a polymerase to produce oligonucleotide
10 fragments; and ligating said oligonucleotide fragments to produce a template.

In another aspect, the invention features a method of determining the nucleotide sequence of any complex described herein, the method including: (a) annealing one or more relay primers with the complex, where the one or more relay primers span the first linkage and/or *n-1* linkages; (b) extending the relay primers using a polymerase to produce oligonucleotide fragments; (c) ligating the oligonucleotide fragments to produce a template; (d) optionally
15 amplifying the template by polymerase chain reaction to produce an amplified mix; and (e) sequencing the optionally amplified mix to determine the sequence of the complex.

In some embodiments, the complex includes a 5'-connector at the 5'-terminus of one or more tags and a 3'-connector at the 3'-terminus of one or more tags, and where each of the one or more relay primers hybridizes to adjacent 5'- and 3'-connectors. In particular embodiments, each 5'-connector includes the same sequence, and/or
20 each 3'-connector includes the same sequence. In some embodiments, the sequence of at least one of the 5'-connector is complementary to the sequence of the adjacent 3'-connector (e.g., to form a duplex between the 5'- and 3'-connectors under hybridization conditions) or identical or sufficiently similar to allow for hybridization to a complementary oligonucleotide. In further embodiments, the method includes hybridizing the 5'-connector to the adjacent 3'-connector. In some embodiments, each relay primer for a specific junction (e.g., thereby forming a
25 three-helix junction) between the 5'- and 3'-connectors includes the same sequence.

In another aspect, the invention features a method of determining the nucleotide sequence of any complex described herein, the method including: (a) providing the complex including the chemical linkage, where the chemical linkage is a cross-linking oligonucleotide that spans the junction between two adjacent tags and that
30 operatively associates with the two adjacent tags via one or more reversible co-reactive groups; (b) releasing the cross-linking oligonucleotide (e.g., any described herein) to produce a template; (c) optionally amplifying the template by polymerase chain reaction to produce an amplified mix; and (d) sequencing the optionally amplified mix to determine the sequence of the complex. In some embodiments, a portion of the cross-linking oligonucleotide hybridizes to the termini of two adjacent tags, thereby producing a nick or a gap between the two adjacent tags.

In further embodiments, the method includes ligating the tags using a chemical process or an enzymatic
35 process (e.g., 5'-phosphorylation) prior to step (b).

In some embodiments, the method further includes repairing the template prior to the optional amplifying step or the sequencing step. In particular embodiments, the repairing step includes modifying at least one of the first linkage and/or *n-1* linkages to be a repaired linkage capable of being read or translocated through by a polymerase. In some embodiments, the polymerase can read or translocate through at least 50% (e.g., at least 60%, 65%, 70%,
40 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or even 100%) of the repaired linkages (e.g., as compared to a control, such as a control oligonucleotide having the first linkage and *n-1* linkages that have not been repaired). In particular embodiments, the polymerase can read or translocate through about 10% to about 100% of the repaired

linkages (e.g., 20% to 100%, 25% to 100%, 50% to 100%, 75% to 100%, 90% to 100%, 95% to 100%, 10% to 95%, 20% to 95%, 25% to 95%, 50% to 95%, 75% to 95%, 90% to 95%, 10% to 90%, 20% to 90%, 25% to 90%, 50% to 90%, and 75% to 90%, as compared to control (e.g., as compared to a control oligonucleotide having the first linkage and *n-1* linkages that have not been repaired)). In particular embodiments, the repairing step includes providing the template with a repair enzyme (e.g., a photolyase, a glycosylase, an endonuclease, a Flap endonuclease, an apurinic/apyrimidinic (AP) endonuclease, a poly ADP ribose polymerase, or a methyltransferase, as described herein).

In any of the above embodiments, the methods or complexes may include only single-stranded molecules, where the headpiece, the first tag, and/or the one or more additional tags are single-stranded.

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 the target.

As described herein, the use of single-stranded molecules could have numerous benefits. Accordingly, in any of the embodiments described herein, the methods and complexes include a headpiece, one or more 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 tag). In some embodiments, each of the tags within a set of tags (e.g., the first tag or a subsequent tag, if present) has about the same mass (e.g., each tag has a mass that is about +/- 10% from the average mass between two or more tags). In particular embodiments, the 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 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 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 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 tag to encode such information.

5 In any of the above embodiments, an oligonucleotide (e.g., the headpiece, the first tag, and/or one or more additional tags, if present) encodes for the identity of the library. In some embodiments, the oligonucleotide (e.g., the headpiece, the first tag, and/or one or more additional 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 some embodiments, more than two libraries are combined (e.g., three, four, five, six, seven, eight, nine, ten, or more libraries).

15 In any of the above embodiments, the encoded information is provided in one or more tags or in a combination of more than one tag. In some embodiments, the encoded information is represented by more than one tag (e.g., two, three, four, five, six, seven, eight, nine, ten, or more tags). In some embodiments, the encoded information is represented by more than one tag, where all encoding tags are contained within the encoding sequence (e.g., by using of a specific tag combination to encode information). In some embodiments, the encoded information is represented by more than one tag, where less than all encoding tags are contained within the encoding sequence (e.g., by using one tag from a set of more than one individual tag to encode within an individual encoding sequence). In some embodiments, the encoded information is represented orthogonally, where encoded information is represented by a combination of more than one tag with less than all of the encoding information being contained within an individual library member, such that more than one corresponding library member needs to be sequenced in order to deconvolute the encoded information. In some embodiments, more than one chemical building block is represented by a single tag (e.g., for a racemic building block, such as two, three, four, five, six, seven, eight, nine, ten, or more building blocks represented by a single tag).

25 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 tag, and/or one or more additional tags, if present) include 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 oligonucleotide tags) 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 tag, and/or one or more additional tags, if present) includes an origin sequence (e.g., a random or 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, as compared to other members of 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.

40 In any of the embodiments herein, 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.

In any of the above embodiments, the methods, compositions, and complexes, or portions thereof (e.g., the headpiece, the first tag, and/or the one or more additional tags, if present), includes a modification that support
5 solubility in semi-, reduced-, or non-aqueous (e.g., organic) conditions. In some embodiments, the bifunctional spacer, headpiece, or one or more 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
10 embodiments, 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-
15 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 tag, the one or more additional tags,
20 the library-identifying tag, the use tag, and/or the origin tag, if present, may include from about 5 to about 75 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 5 to 20 nucleotides, from 5 to 30 nucleotides, from 5 to 40 nucleotides, from 5 to 50 nucleotides, from 5 to 60
25 nucleotides, from 5 to 70 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,
30 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
35 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 10 to 30 nucleotides, from 10 to 40 nucleotides, from 10 to 50 nucleotides, from 10 to 60 nucleotides, from 10 to 70
40 nucleotides, from 10 to 75 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, from 19 to 20 nucleotides, from 20 to 30 nucleotides, from 20 to 40 nucleotides, from 20 to 50 nucleotides, from 20 to 60 nucleotides, from 20 to 70 nucleotides, from 20 to 75 nucleotides, from 30 to 40 nucleotides, from 30 to 50 nucleotides, from 30 to 60 nucleotides, from 30 to 70 nucleotides, from 30 to 75 nucleotides, from 40 to 50 nucleotides, from 40 to 60 nucleotides, from 40 to 70 nucleotides, from 40 to 75 nucleotides, from 50 to 60 nucleotides, from 50 to 70 nucleotides, from 50 to 75 nucleotides, from 60 to 70 nucleotides, from 60 to 75 nucleotides, and from 70 to 75 nucleotides). In particular
embodiments, the headpiece, the first tag, the second tag, the one or more additional 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 any of the above embodiments, the encoding sequence (e.g., the headpiece, the tailpiece, the first tag, the one or more additional tags, the library-identifying tag, the use tag, and/or the origin tag, if present) may include more than 20 nucleotides (e.g., more than 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, or 75 nucleotides).

Definitions

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

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.

By “bifunctional spacer” is meant a spacing moiety having two reactive groups that allow for binding of a chemical entity and the encoding information of the complex. In one non-limiting example, the bifunctional spacer is provided between the chemical entity and a tag. In another non-limiting example, the bifunctional spacer is provided between the chemical entity and a headpiece. Exemplary bifunctional spacers are provided herein.

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 to provide an enzymatic linkage) or by chemical binding (e.g., chemical ligation to provide a chemical linkage).

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.

5 By “chemical entity” is meant a compound comprising one or more building blocks and optionally one or more scaffolds. 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
10 greater) entity.

By “chemical-reactive group” is meant a reactive group that participates in a modular reaction, thus producing a linkage. Exemplary reactions and reactive groups include those selected from 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 4 π -electron system and an
15 optionally substituted dienophile or an optionally substituted heterodienophile having a 2 π -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.

By “complementary” is meant a sequence capable of hybridizing, as defined herein, to form secondary
20 structure (a duplex or a double-stranded portion of a nucleic acid molecule). The complementarity need not be perfect but may include one or more mismatches at one, two, three, or more nucleotides. For example, complementary sequence may contain nucleobases that can form hydrogen bonds according to Watson-Crick base-pairing rules (e.g., G with C, A with T or A with U) or other hydrogen bonding motifs (e.g., diaminopurine with T, 5-methyl C with G, 2-thiothymidine with A, inosine with C, pseudoisocytosine with G). The sequence and its
25 complementary sequence can be present in the same oligonucleotide or in different oligonucleotides.

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 spacer between the chemical entity and the headpiece.

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

30 By “connector” of an oligonucleotide tag is meant a portion of the tag at or in proximity to the 5'- or 3'-terminus having a fixed sequence. A 5'-connector is located at or in proximity to the 5'-terminus of an oligonucleotide, and a 3'-connector is located at or in proximity to the 3'-terminus of an oligonucleotide. When present in a complex, each 5'-connector may be the same or different, and each 3'-connector may be the same or different. In an exemplary, non-limiting complex having more than one tags, each tag can include a 5'-connector and a 3'-connector, where each 5'-connector has the same sequence and each 3'-connector has the same sequence
35 (e.g., where the sequence of the 5'-connector can be the same or different from the sequence of the 3'-connector). In another exemplary, non-limiting complex, the sequence of the 5'-connector is designed to be complementary, as defined herein, to the sequence of the 3'-connector (e.g., to allow for hybridization between 5'- and 3'-connectors). The connector can optionally include one or more groups allowing for a linkage (e.g., a linkage for which a
40 polymerase has reduced ability to read or translocate through, such as a chemical linkage).

By “constant” or “fixed constant” sequence is meant a sequence of an oligonucleotide that does not encode information. Non-limiting, exemplary portions of a complex having a constant sequence include a primer-binding

region, a 5'-connector, or a 3'-connector. The headpiece of the invention can encode information (thus, a tag) or alternatively not encode information (thus, a constant sequence). Similarly, the tailpiece of the invention can encode or not encode information.

By "cross-linking oligonucleotide" is meant an oligonucleotide that operatively associates, as defined
5 herein, at a particular junction between two adjacent tags in a complex. In a non-limiting example, one terminus of the cross-linking oligonucleotide hybridizes to the 3'-connector of a first tag, and the other terminus of the cross-linking oligonucleotide hybridizes to the 5'-connector of a second tag that is adjacent to the first tag. Exemplary, non-limiting embodiments of cross-linking oligonucleotides include those having one or more reactive groups (e.g., a chemical-reactive group, a photo-reactive group, an intercalating moiety, or a reversible co-reactive group, or any
10 described herein) that operatively associates with adjacent tags or connectors of adjacent tags.

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 chemical structure for library synthesis that is operatively linked to a component of a chemical entity and to a tag, e.g., a starting oligonucleotide. Optionally a headpiece may contain few or no
15 nucleotides, but may provide a point at which they may be operatively associated. Optionally, a bifunctional spacer connects the headpiece to the component.

By "hybridize" is meant to pair to form a double-stranded molecule between complementary oligonucleotides, or portions thereof, under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507.) For example, high
20 stringency hybridization can be obtained with a salt concentration ordinarily less than about 750 mM NaCl and 75 mM trisodium citrate, less than about 500 mM NaCl and 50 mM trisodium citrate, or less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide or at least about 50% formamide. High stringency hybridization temperature conditions will ordinarily include
25 temperatures of at least about 30°C, 37°C, or 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In one embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In an alternative embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM
30 trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a further alternative embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

For most applications, washing steps that follow hybridization will also vary in stringency. Wash
35 stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, high stringency salt concentrations for the wash steps may be, e.g., less than about 30 mM NaCl and 3 mM trisodium citrate, or less than about 15 mM NaCl and 1.5 mM trisodium citrate. High stringency temperature conditions for the wash steps will ordinarily include a temperature of, e.g., at least about 25°C, 42°C, or 68°C. In one embodiment, wash steps will
40 occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In an alternative embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a further alternative embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on

these conditions will be readily apparent to those skilled in the art. Hybridization techniques are well known to those skilled in the art and are described, for example, in Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel et al. (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York); and Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

By “intercalating moiety” is meant a reactive group that results in inclusion of a moiety between two or more nucleotides. In a non-limiting example, the intercalating moiety reacts with one or more nucleotides to form inter- or intra-strand cross-links between duplex or triplex oligonucleotides. Exemplary, non-limiting intercalating moieties are described herein.

By “junction” is meant a nick (lack of an internucleotide bond) or a gap (lack of one or more nucleotides) between two adjacent tags in a complex. The junction can also be between two adjacent connectors present in two adjacent tags (e.g., between the 3'-connector of a first tag and the 5'-connector of a second tag that is adjacent to the first tag).

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 “linkage” is meant a chemical connecting entity that allows for operatively associating two or more chemical structures, where the linkage is present between the headpiece and one or more tags, between two tags, or between a tag and a tailpiece. The chemical connecting entity can be a non-covalent bond (e.g., as described herein), a covalent bond, or a reaction product between two functional groups. By “chemical linkage” is meant a linkage formed by a non-enzymatic, chemical reaction between two functional groups. Exemplary, non-limiting functional groups include a chemical-reactive group, a photo-reactive group, an intercalating moiety, or a cross-linking oligonucleotide (e.g., as described herein). By “enzymatic linkage” is meant an internucleotide or internucleoside linkage formed by an enzyme. Exemplary, non-limiting enzymes include a kinase, a polymerase, a ligase, or combinations thereof. By a linkage “for which a polymerase has reduced ability to read or translocate through” is meant a linkage, when present in an oligonucleotide template, that provides a reduced amount of elongated and/or amplified products by a polymerase, as compared to a control oligonucleotide lacking the linkage. Exemplary, non-limiting methods for determining such a linkage include primer extension as assessed by PCR analysis (e.g., quantitative PCR), RT-PCR analysis, liquid chromatography-mass spectrometry, sequence demographics, or other methods. Exemplary, non-limiting polymerases include DNA polymerases and RNA polymerases, such as DNA polymerase I, DNA polymerase II, DNA polymerase III, DNA polymerase VI, Taq DNA polymerase, Deep VentR™ DNA Polymerase (high-fidelity thermophilic DNA polymerase, available from New England Biolabs), T7 DNA polymerase, T4 DNA polymerase, RNA polymerase I, RNA polymerase II, RNA polymerase III, or T7 RNA polymerase.

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.

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 and nucleotide (e.g., modified DNA or RNA) 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, 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 bases (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; hexose; anhydrohexitol; altritol; mannitol; cyclohexanyl; cyclohexenyl; morpholino that also has a phosphoramidate backbone; locked nucleic acids (LNA, e.g., where the 2'-hydroxyl of the ribose is connected by a C₁₋₆ alkylene or C₁₋₆ heteroalkylene bridge to the 4'-carbon of the same ribose sugar, where exemplary bridges included methylene, propylene, ether, or amino bridges); glycol nucleic acid (GNA, e.g., R-GNA or S-GNA, where ribose is replaced by glycol units attached to phosphodiester bonds); threose nucleic acid (TNA, where ribose is replaced with α -L-threofuranosyl-(3' → 2')); and/or replacement of the oxygen in ribose (e.g., with S, Se, or alkylene, such as methylene or ethylene)), modified backbones (e.g., peptide nucleic acid (PNA), where 2-amino-ethyl-glycine linkages replace the ribose and phosphodiester backbone), and/or modified phosphate groups (e.g., phosphorothioates, 5'-N-phosphoramidites, phosphoroselenates, boranophosphates, boranophosphate esters, hydrogen phosphonates, phosphoramidates, phosphorodiamidates, alkyl or aryl phosphonates, phosphotriesters, bridged phosphoramidates, bridged phosphorothioates, and bridged methylene-phosphonates). The oligonucleotide can be single-stranded (e.g., hairpin), double-stranded, or possess other secondary or tertiary structures (e.g., stem-loop structures, double helixes, triplexes, quadruplexes, etc.).

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 associated in an indirect manner (e.g., covalently via an appropriate spacer). For example, the spacer may be a bifunctional moiety with a site of attachment for chemical entity and a site of attachment for the headpiece.

By "photo-reactive group" is meant a reactive group that participates in a reaction caused by absorption of ultraviolet, visible, or infrared radiation, thus producing a linkage. Exemplary, non-limiting photo-reactive groups are described herein.

By "protecting group" is meant a group intended to protect the 3'-terminus or 5'-terminus of an oligonucleotide or to protect one or more functional groups of the chemical entity, scaffold, or building block against undesirable reactions during one or more binding steps of making, tagging, or using an oligonucleotide-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 for oligonucleotides 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-butyltrimethylsilyl). Exemplary, non-limiting protecting groups for chemical entities, scaffolds, and building blocks include N-protecting groups to protect an amino group against undesirable reactions during synthetic procedure (e.g., acyl; aryloyl; carbamyl groups, such as formyl, acetyl, propionyl, pivaloyl, t-butylacetyl, 2-chloroacetyl, 2-

bromoacetyl, trifluoroacetyl, trichloroacetyl, phthalyl, o-nitrophenoxyacetyl, α -chlorobutyryl, benzoyl, 4-chlorobenzoyl, 4-bromobenzoyl, 4-nitrobenzoyl, and chiral auxiliaries, such as protected or unprotected D, L or D, L-amino acids, such as alanine, leucine, phenylalanine, and the like; sulfonyl-containing groups, such as benzenesulfonyl, p-toluenesulfonyl, and the like; carbamate forming groups, such as benzyloxycarbonyl, p-chlorobenzoyloxycarbonyl, p-methoxybenzyloxycarbonyl, p-nitrobenzyloxycarbonyl, 2-nitrobenzyloxycarbonyl, p-bromobenzoyloxycarbonyl, 3,4-dimethoxybenzyloxycarbonyl, 3,5 dimethoxybenzyl oxycarbonyl, 2,4-dimethoxybenzyloxycarbonyl, 4 methoxybenzyloxycarbonyl, 2-nitro-4,5-dimethoxybenzyloxycarbonyl, 3,4,5-trimethoxybenzyloxycarbonyl, 1-(p-biphenyl)-1-methylethoxycarbonyl, α,α -dimethyl-3,5-dimethoxybenzyloxycarbonyl, benzhydryloxy carbonyl, t-butyloxycarbonyl, diisopropylmethoxycarbonyl, isopropylloxycarbonyl, ethoxycarbonyl, methoxycarbonyl, allyloxycarbonyl, 2,2,2-trichloroethoxycarbonyl, phenoxycarbonyl, 4-nitrophenoxy carbonyl, fluorenyl-9-methoxycarbonyl, cyclopentylloxycarbonyl, adamantylloxycarbonyl, cyclohexylloxycarbonyl, phenylthiocarbonyl, and the like; alkaryl groups, such as benzyl, triphenylmethyl, benzyloxymethyl, and the like; and silyl groups such as trimethylsilyl, and the like; where preferred N-protecting groups are formyl, acetyl, benzoyl, pivaloyl, t-butylacetyl, alanyl, phenylsulfonyl, benzyl, t-butylloxycarbonyl (Boc), and benzyloxycarbonyl (Cbz)); O-protecting groups to protect a hydroxyl group against undesirable reactions during synthetic procedure (e.g., alkylcarbonyl groups, such as acyl, acetyl, pivaloyl, and the like; optionally substituted arylcarbonyl groups, such as benzoyl; silyl groups, such as trimethylsilyl (TMS), tert-butyl dimethylsilyl (TBDMS), tri-iso-propylsilyloxymethyl (TOM), triisopropylsilyl (TIPS), and the like; ether-forming groups with the hydroxyl, such methyl, methoxymethyl, tetrahydropyranyl, benzyl, p-methoxybenzyl, trityl, and the like; alkoxy-carbonyls, such as methoxycarbonyl, ethoxycarbonyl, isopropoxycarbonyl, n-isopropoxycarbonyl, n-butyloxycarbonyl, isobutyloxycarbonyl, sec-butyloxycarbonyl, t-butyloxycarbonyl, 2-ethylhexylloxycarbonyl, cyclohexylloxycarbonyl, methylloxycarbonyl, and the like; alkoxyalkoxy-carbonyl groups, such as methoxymethoxycarbonyl, ethoxymethoxycarbonyl, 2-methoxyethoxycarbonyl, 2-ethoxyethoxycarbonyl, 2-butoxyethoxycarbonyl, 2-methoxyethoxymethoxycarbonyl, allyloxycarbonyl, propargyloxycarbonyl, 2-butenoxycarbonyl, 3-methyl-2-butenoxycarbonyl, and the like; haloalkoxy-carbonyls, such as 2-chloroethoxycarbonyl, 2-chloroethoxycarbonyl, 2,2,2-trichloroethoxycarbonyl, and the like; optionally substituted arylalkoxy-carbonyl groups, such as benzyloxycarbonyl, p-methylbenzyloxycarbonyl, p-methoxybenzyloxycarbonyl, p-nitrobenzyloxycarbonyl, 2,4-dinitrobenzyloxycarbonyl, 3,5-dimethylbenzyloxycarbonyl, p-chlorobenzoyloxycarbonyl, p-bromobenzoyloxy-carbonyl, and the like; and optionally substituted aryloxycarbonyl groups, such as phenoxycarbonyl, p-nitrophenoxycarbonyl, o-nitrophenoxycarbonyl, 2,4-dinitrophenoxycarbonyl, p-methyl-phenoxycarbonyl, m-methylphenoxycarbonyl, o-bromophenoxycarbonyl, 3,5-dimethylphenoxycarbonyl, p-chlorophenoxycarbonyl, 2-chloro-4-nitrophenoxy-carbonyl, and the like); carbonyl-protecting groups (e.g., acetal and ketal groups, such as dimethyl acetal, 1,3-dioxolane, and the like; acylal groups; and dithiane groups, such as 1,3-dithianes, 1,3-dithiolane, and the like); carboxylic acid-protecting groups (e.g., ester groups, such as methyl ester, benzyl ester, t-butyl ester, orthoesters, and the like; silyl groups, such as trimethylsilyl, as well as any described herein; and oxazoline groups); and phosphate-protecting groups (e.g., optionally substituted ester groups, such as methyl ester, isopropyl ester, 2-cyanoethyl ester, allyl ester, t-butyl ester, benzyl ester, fluorenylmethyl ester, 2-(trimethylsilyl)ethyl ester, 2-(methylsulfonyl)ethyl ester, 2,2,2-trichloroethyl ester, 3',5'-dimethoxybenzoic ester, p-hydroxyphenacyl ester, and the like).

By "proximity" or "in proximity" to a terminus of an oligonucleotide is meant near or closer to the stated terminus than the other remaining terminus. For example, a moiety or group in proximity to the 3'-terminus of an oligonucleotide is near or closer to the 3'-terminus than the 5'-terminus. In particular embodiments, a moiety or

group in proximity to the 3'-terminus of an oligonucleotide is one, two, three, four, five, six, seven, eight, nine, ten, fifteen, or more nucleotides from the 3'-terminus. In other embodiments, a moiety or group in proximity to the 5'-terminus of an oligonucleotide is one, two, three, four, five, six, seven, eight, nine, ten, fifteen, or more nucleotides from the 5'-terminus.

5 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 "reversible co-reactive group" is meant a reactive group that participates in a reversible reaction. Exemplary, non-limiting reactive groups include photo-reactive groups, where exposure to a particular absorption radiation results in a linkage between the photo-reactive groups and exposure to a different, particular absorption radiation results in cleavage of the formed linkage (e.g., a cyanovinylcarbazole group, a cyanovinyl group, and an acrylamide group). Another exemplary, non-limiting reactive group includes redox-reactive groups, where such groups can be reversibly reduced or oxidized (e.g., a thiol group).

15 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 scaffold is derivatized such that it can be orthogonally deprotected during library synthesis and subsequently reacted with different diversity nodes.

20 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 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 "tag" or "oligonucleotide tag" is meant an oligonucleotide portion of the library at least part of which encodes information. Non-limiting examples of such information include the addition (e.g., by a binding reaction) of a component (i.e., a scaffold or a building block, as in a scaffold tag or a building block tag, respectively), the headpiece in the library, the identity of the library (i.e., as in an identity tag), the use of the library (i.e., as in a use tag), and/or the origin of a library member (i.e., as in an origin tag).

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

5 By “primer” is meant an oligonucleotide that is capable of annealing to an oligonucleotide template and then being extended by a polymerase in a template-dependent manner.

By “relay primer” is meant an oligonucleotide that is capable of annealing to an oligonucleotide template that contains, in the region of the template to which the primer is hybridized, at least one internucleotide linkage that reduces the ability of a polymerase to read or translocate through. Upon hybridization, one or more relay primers allow for extension by a polymerase in a template dependent manner.

10 By “recombination,” as used herein, is meant the generation of a polymerase product as a result of at least two distinct hybridization events.

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

15 Brief Description of the Drawings

Figures 1A-1B shows exemplary complexes. Figure 1A shows a complex (from the 5' to 3' direction) having a chemical entity (star), a fixed constant sequence (e.g., a headpiece), followed by three variable encoding sequences (e.g., three tags), and another fixed constant sequence (e.g., a tailpiece), where the linkage between the fixed constant sequences and variable encoding sequences are readable. The complex can be further extended using
20 a polymerase with a primer (hybridized to the 3'-terminus of the complex), optionally amplified with PCR, and sequenced. Figure 1B shows a complex having unreadable linkages. The complex (from the 5' to 3' direction) has a chemical entity (star), a fixed constant sequence (e.g., a headpiece), followed by three variable encoding sequences (e.g., three tags), and another fixed constant sequence (e.g., a tailpiece), where the linkage between the fixed constant sequences and variable encoding sequences are unreadable. Each of the variable encoding sequences
25 includes a 5'-connector and a 3'-connector that are fixed constant sequences. Because the linkages are unreadable, relay primers are used to span the linkages and to allow for extension with a polymerase, thus forming oligonucleotide fragments. The fragments are then ligated using ligase, optionally amplified with PCR, and sequenced. The reading process (e.g., optional extension, ligation, optional amplification, and sequencing) may optionally be entirely performed after selection. Alternatively, the optional extension and ligation part of the
30 reading process may be performed prior to selection.

Figure 2 show an exemplary complex having an unreadable linkage formed by psoralen. The schematic provides (from the 5' to 3' direction) a variable encoding sequence, a 3'-connector, a 5'-connector, and another variable encoding sequence. The 3'-connector and 5'-connector hybridizes to form a duplex, and psoralen (an intercalating, photo-reactive moiety) forms an unreadable linkage between the 5'- and 3'-connectors. The relay
35 primer hybridizes to the 5'- and 3'-connectors and spans the unreadable linkage. The readable portions of the complex (e.g., the variable encoding sequence) are extended with a polymerase to form oligonucleotide fragments. The fragments and relay primers are ligated using ligase, optionally amplified with PCR, and sequenced. The reading process (e.g., extension, ligation, optional amplification, and sequencing) may optionally be entirely performed after selection. Alternatively, the extension and ligation part of the reading process may be performed
40 prior to selection.

Figure 3 show an exemplary complex having unreadable linkages formed by a cross-linking oligonucleotide. In the top schematic, the complex includes (from the 5' to 3' direction) a chemical entity (star), a

fixed constant sequence (e.g., a headpiece), a cross-linking oligonucleotide that is cross-linked via a reversible co-reactive group (e.g., cyanovinylcarbazole groups cross-linked at about 366 nm, cross-links marked with X), and followed by a variable encoding sequence having fixed constant sequences at the 5'- and 3'-termini (e.g., a tag having 5'- and 3'-connectors). The combination of the cross-linking oligonucleotide and variable encoding sequence are repeated three more times. Then, the final cross-linking oligonucleotide and another fixed constant sequence (e.g., a tailpiece) follows. The cross-links are marked with X. Next, the complex is reacted with a chemical process and/or an enzymatic process (e.g., a ligase and optionally a kinase) to ligate the tags. Next, the cross-linking oligonucleotides are released to form a template (e.g., using absorption at about 312 nm). Finally, the template is optionally amplified with PCR and sequenced. The reading process (e.g., optional kinase step, ligation, optional amplification, and sequencing) may optionally be entirely performed after selection. Alternatively, the optional kinase step and ligation step of the reading process may be performed prior to selection.

Figure 4 shows an exemplary reversible reaction with reversible co-reactive groups (a pair of a cyanovinylcarbazole group and a thymidine) and use of these groups permits the formation of cross-links between the cross-linking oligonucleotide and the fixed constant sequence of the oligonucleotide tag. Reversible co-reactive groups (marked by X and X') are present in the 3'-connector, 5'-connector, and cross-linking oligonucleotide.

Figure 5 shows Oligonucleotide CNVKJ, containing two cyanovinyl carbazole modifications, CNVK2_P_TagB, CNVK_2_NP_TagB, and CNVK2_TagA.

Figure 6 shows the results of electrophoresis of the products of photochemical cross-linking of Oligonucleotides CNVKJ, CNVK2_P_TagB and CNVK2_TagA.

Figure 7 shows the LCMS of purified CNVK2_P_TagB and CNVK2_TagA.

Figure 8 shows the LCMS of the recovery of the starting Double_CNVK_template.

Figure 9 shows the LCMS of treatment of a complex of CNVK2_P_TagB, CNVK2_TagA and CNVKJ with T4 DNA ligase.

Figure 10 shows the LCMS results of the phosphorylation-ligation reaction of photochemically cross-linked non-phosphorylated model tag pair CNVK2_NP_TagB with CNVK2_TagA and CNVKJ.

Figure 11 shows Oligonucleotide 5PSO2_A9_TA, modified with C2-Psoralen at the 5'-terminus and Oligonucleotide PSO_HP_A9_TCT.

Figure 12 shows the LCMS of the photochemical oligonucleotide cross-linking using psoralen within a stem to model one tagging event of Oligonucleotide 5PSO2_A9_TA, modified with C2-Psoralen at the 5'-terminus and Oligonucleotide PSO_HP_A9_TCT.

Figure 13 shows Oligonucleotides Tag1_PsoCVU and SplintC_PsoC2 and Oligonucleotide 5PSOC2_A9_GA (modified at 5'-terminus with C2 psoralen).

Figure 14 shows a schematic of psoralen photoligation.

Figure 15 shows a gel of the time-course of psoralen photoligation.

Figure 16 shows the LCMS of the photoligated conjugate of 5PsoC2_A9_GA and Tag1_PsoCVU.

Figure 17 shows a 5'-biotinylated oligonucleotide 5Bio_Tag_PsoCVU and a 5'-Psoralen oligonucleotide 5PsoC2_A9_GA along with a splint oligonucleotide SplintC_PsoC2.

Figure 18 shows the LCMS of a Psoralen-Thymidine photochemically ligated oligonucleotide conjugate (Bio_Pso).

Figure 19 shows the LCMS of cDNA generated from Bio_Pso using terminal and non-terminal primer extension and ligation.

Figure 20 shows the LCMS of no-ligation control of cDNA generated from Bio_Pso using terminal and non-terminal primer extension.

Figure 21 shows Oligonucleotides CVU_G and CVU_A and Oligonucleotides Tag1_PsoCVU, SplintC_CVU and SplintA_CVU.

5 **Figure 22** shows the structure of the photochemical ligation junction formed between an irradiated 5'-5-(Carboxy)vinyl-2'-deoxyuridine and a 3'-Thymidine.

Figure 23 shows a gel of the time-course of CVU photoligation at 365 nm.

Figure 24 shows the LCMS of photochemical ligation product of CVU_G with Tag1_PsoCVU.

10 **Figure 25** shows the LCMS of a Carboxyvinyl Uridine-Thymidine photochemically-ligated oligonucleotide conjugate (Bio_CVU).

Figure 26 shows the LCMS of Bio_CVU template minus a single dA nucleotide.

Figure 27 shows Oligonucleotides S_IDTN₃ and S_IDTalkyne.

Figure 28 shows LCMS analysis of the gel purified conjugate of S_IDTN₃ and S_IDTalkyne.

15 **Figure 29** shows TKR_Central and S_IDTN₃ oligos and 5'-dibenzo-cyclooctyne (DBCO)-modified oligo TKR_DBCO_S.

Figure 30 shows the LCMS time course of the copper-free click reaction between TKR_DBCO_S and TKR_Central.

Figure 31 shows the LCMS of the 2cl_s conjugate.

Figure 32 shows Conjugate_Click_S and Conjugate_Click_L.

20 **Figure 33** shows splint-dependent-recombination-mediated cDNA generation/PCR amplification of Conjugate_Click_L and Conjugate_Click_S by Deep Vent (exo-) polymerase in the presence and absence of ePsplint. 24 cycles, 200 pM starting conjugate concentration. 1-Marker; 2- no template; 3- no template + ePsplint; 4- Conjugate_Click_S; 5- Conjugate_Click_S + ePsplint; 6- Conjugate_Click_L; 5- Conjugate_Click_L + ePsplint.

25 **Figure 34** shows the proposed mechanism of splint-dependent-recombination-mediated cDNA generation/PCR amplification.

Figure 35 shows splint-dependent-recombination-mediated cDNA generation/PCR amplification of Conjugate_Click_S and Conjugate_Click_L by Deep Vent exo- polymerase both separately and mixed together in free solution. All reactions supplemented with ePsplint. 1-Marker; 2- no template; 3- Conjugate_Click_S; 4- Conjugate_Click_L; 5- Conjugate_Click_S and Conjugate_Click_L.

30 **Figure 36** shows the possible mechanism of the recombination of the splint-dependent-recombination-mediated cDNA generation/PCR amplification products in free solution with consequent shuffling (loss) of tag associations (encoded information).

35 **Figure 37** shows recombination-mediated cDNA generation/PCR amplification of Conjugate_Click_S and Conjugate_Click_L mixture by Deep Vent (exo-) polymerase in free solution and in emulsified aqueous compartments with conditions as follows: 30 seconds at 94°C, 30 seconds at 52°C and 60 seconds at 72°C for 35 cycles, and initial conjugate concentration of 5pM;. 1-Marker; 2- free solution, 3- emulsified aqueous compartments.

Figure 38 shows TKR_S and TKR_L conjugates and their LCMS traces.

40 **Figure 39** shows Deep Vent (exo-) Polymerase repeat-dependent-recombination-mediated cDNA generation/PCR amplification of TKR_S and TKR_L. 1- Marker, 2- no template, 3- TKR_S, 4- TKR_L, 5- TKR L and TKR_S. Conditions: 5 uL of 10x Thermopol buffer; 2.5 uL of a 10 uM solution of each forward and reverse primer; 2.5 uL 10 mM dNTP mix (NEB); 40 pM final concentration of either conjugate or their 1:1 mixture, 2.5 uL

Deep Vent (exo-) Polymerase and water to 50 uL. The reaction was cycled 21 times as follows: 30 seconds at 94°C, 30 seconds at 52°C and 60 seconds at 72°C.

Figure 40 shows sequence data derived from the cloned amplification product of the repeat-dependent-recombination-mediated cDNA generation/PCR amplification of mixed TKR_S and TKR_L in free solution.

5 **Figure 41** shows TKR_2_click_S and TKR_2_click_L conjugates and their LCMS traces.

Figure 42 shows Deep Vent (exo-) Polymerase repeat-dependent-recombination-mediated cDNA generation/PCR amplification of 2_click_S and 2_click_L in free solution. 1- Marker; 2- No template control; 3- 2_click_S; 4- 2_click_L; 5-1:1 mixture of 2_click_S and 2_click_L. Conditions: 5 uL of 10x Thermopol buffer; 2.5 uL of a 10 uM solution of each forward and reverse primer; 2.5 uL 10 mM dNTP mix (NEB); 40 pM final concentration of either conjugate or their 1:1 mixture, 2.5 uL Deep Vent (exo-) Polymerase and water to 50 uL. The reaction was cycled 22 times as follows: 30 seconds at 94°C, 30 seconds at 52°C and 60 seconds at 72°C.

Figure 43 shows sequence data derived from the cloned amplification product of the repeat-dependent-recombination-mediated cDNA generation/PCR amplification product of TKR_2_click_S and TKR_2_click_L mixture in free solution.

15

Detailed Description

We have developed complexes having at least one linkage (e.g., a chemical linkage) for which a polymerase has reduced ability to read or translocate through. The ability to tag libraries without being constrained by readability of a polymerase greatly expands the range of tagging methods that can be used. Possible, non-limiting advantages include decreased mass of tags and/or library members; increased solubility of library members in various reaction conditions (e.g., in aqueous and/or organic conditions suitable for both chemical entity and tag synthesis); increased stability (e.g., in aqueous and/or organic reaction conditions, where such oligonucleotide tags and linkages may be more resistant to particular reaction conditions, as compared to tag without such linkages); decreased cost (e.g., reduce use of expensive enzyme reagents); increased ease of use (e.g., use of cyanovinylcarbazole cross-linking occurs over a broad pH range of 5.5-9.5, which permits cross-linking under reaction conditions suitable for forming the chemical entity-portion of the complex); reduced number of reaction steps and reduced use of reagent(s) (e.g., reduced use of a buffer exchange step, such as during combinatorial split reactions, where thousands of individual, small aliquots need to be independently processed; a precipitation step, or a pH-modification step); increased fidelity (e.g., when hybridization-mediated methods are used to reduce the frequency of occurrence of mis-tagging events); and increased compatibility with reaction conditions for forming the chemical entity (e.g., allows for use of orthogonal functionalities, where unique, tagging chemistries (e.g., use of UV irradiation) can be used that will not occur during post-split/mix step(s), which will also reduce the frequency of mis-tagging events).

Because such linkages may have reduced readability by a polymerase, we have also developed methods to allow for sequencing such complexes. These methods include use of relay primers to span the linkages and/or use of releasable cross-linking oligonucleotides. These complexes and methods 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. Methods of creating and tagging libraries of these complexes are described in detail below.

40

Complexes

This invention features a complex including a chemical entity, one or more tags, and a headpiece operatively associated with the chemical entity and one or more tags. At least one of the linkages between the headpiece and the tag or between two tags is a linkage for which a polymerase has reduced ability to read or translocate through. The chemical entities, headpieces, tags, linkages, and bifunctional spacers are further described below.

Chemical Entities

The chemical entities or members (e.g., small molecules or peptides) of the invention can include one or more building blocks and optionally include one or more scaffolds.

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 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), chemical-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 quinolinyl, isoquinolinyl, 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 chemical-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.

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 bifunctional spacer can optionally be used as a spacing moiety 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 spacer (i.e., a spacing moiety 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 spacer can be attached to the 5'-terminus, at one of the internal positions, or to the 3'-terminus of the headpiece. When the spacer is attached to one of the internal positions, the spacer 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 spacers 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. For example, the chemical moiety can be operatively linked to the 3'-terminus or 5'-terminus of the headpiece. In particular embodiments, the headpiece includes a hairpin structure formed by complementary bases within the sequence. For example, the chemical moiety can be operatively linked to the internal position, the 3'-terminus, or the 5'-terminus of the headpiece.

Generally, the headpiece includes a non-self-complementary sequence on the 5'- or 3'- terminus that allows for binding an oligonucleotide tag by polymerization, enzymatic ligation, or chemical reaction. The headpiece can allow for ligation of oligonucleotide tags and optional purification and phosphorylation steps. After the addition of the last tag, an additional adapter sequence can be added to the 5'-terminus of the last tag. 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), a mix-and-split strategy may be employed during the 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 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, 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 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 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 cross-linked bases to form the hairpin structure. For example, bases within a single strand or bases in different double strands can be cross-linked, 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 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.

Oligonucleotide tags

The oligonucleotide tags described herein (e.g., a tag or a portion of a headpiece or a portion of a tailpiece) 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 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 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 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 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 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 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, stochastically-generated) 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 multiple amplicons derived from the same template molecule. When it is desired to determine the statistics and demographics of the population of encoding tags prior to amplification, as opposed to post-amplification, the origin tag may be used. These origin sequences can be included in a headpiece, a tailpiece, a 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.

Any of tags described herein can include a connector at or in proximity to the 5'- or 3'-terminus having a fixed sequence. Connectors facilitate the formation of linkages (e.g., chemical linkages) by providing a reactive group (e.g., a chemical-reactive group or a photo-reactive group) or by providing a site for an agent that allows for a linkage (e.g., an agent of an intercalating moiety or a reversible reactive group in the connector(s) or cross-linking oligonucleotide). Each 5'-connector may be the same or different, and each 3'-connector may be the same or different. In an exemplary, non-limiting complex having more than one tags, each tag can include a 5'-connector and a 3'-connector, where each 5'-connector has the same sequence and each 3'-connector has the same sequence (e.g., where the sequence of the 5'-connector can be the same or different from the sequence of the 3'-connector). The connector provides a sequence that can be used for one or more linkages. To allow for binding of a relay primer or for hybridizing a cross-linking oligonucleotide, the connector can include one or more functional groups allowing for a linkage (e.g., a linkage for which a polymerase has reduced ability to read or translocate through, such as a chemical linkage).

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 within a specific set of tags that encode a specific variable); 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).

Linkages

The linkages of the invention are present between oligonucleotides that encode information (e.g., such as between the headpiece and a tag, between two tags, or between a tag and a tailpiece), where such linkages include any linkage for which a polymerase has reduced ability to read or translocate through. Exemplary linkages include chemical linkages including one or more of a chemical-reactive group, a photo-reactive group, an intercalating moiety, a cross-linking oligonucleotide, or a reversible co-reactive group.

A linkage may be tested to determine whether a polymerase has reduced ability to read or translocate through that linkage. This ability can be tested by any useful method, such as liquid chromatography-mass spectrometry, RT-PCR analysis, sequence demographics, and/or PCR analysis.

In particular embodiments, chemical ligation includes the use of one or more chemical-reactive pairs to provide a linkage. Exemplary chemical-reactive pairs are a pair including an optionally substituted alkynyl group and an optionally substituted azido group to form a triazole via a Huisgen 1,3-dipolar cycloaddition reaction; an optionally substituted diene having a 4 π -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 2 π -electron system (e.g., an optionally substituted alkenyl group or an optionally substituted alkynyl group) to form a cycloalkenyl via a Diels-Alder reaction; 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) to form a heteroalkyl 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; an optionally substituted amino group with an aldehyde group or a ketone 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; a pair of an optionally substituted amino group and a carboxylic acid group or a thiol group (e.g., with or without the use of succinimidyl trans-4-

5 (maleimidylmethyl)cyclohexane-1-carboxylate (SMCC) or 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDAC); a pair of an optionally substituted hydrazine and an aldehyde or a ketone group; a pair of an optionally substituted hydroxylamine and an aldehyde or a ketone group; or a pair of a nucleophile and an optionally substituted alkyl halide.

10 Platinum complexes, alkylating agents, or furan-modified nucleotides can also be used as a chemical-reactive group to form inter- or intra-strand linkages. Such agents can be used between two oligonucleotides and can optionally be present in the cross-linking oligonucleotide.

Exemplary, non-limiting platinum complexes include cisplatin (cis-diamminedichloroplatinum(II), e.g., to form GG intra-strand linkages), transplatin (trans-diamminedichloroplatinum(II), e.g., to form GXG inter-strand

15 linkages, where X can be any nucleotide), carboplatin, picolatin (ZD0473), ormaplatin, or oxaliplatin to form, e.g., GC, CG, AG, or GG linkages. Any of these linkages can be inter- or intra-strand linkages.

Exemplary, non-limiting alkylating agents include nitrogen mustard (mechlorethamine, e.g., to form GG linkages), chlorambucil, melphalan, cyclophosphamide, prodrug forms of cyclophosphamide (e.g., 4-

20 hydroperoxycyclophosphamide and ifosfamide), 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU, carmustine), an aziridine (e.g., mitomycin C, triethylenemelamine, or triethylenethiophosphoramidate (thio-tepa) to form GG or AG linkages), hexamethylmelamine, an alkyl sulfonate (e.g., busulphan to form GG linkages), or a nitrosourea (e.g., 2-chloroethylnitrosourea to form GG or CG linkages, such as carmustine (BCNU), chlorozotocin, lomustine (CCNU), and semustine (methyl-CCNU)). Any of these linkages can be inter- or intra-strand linkages.

Furan-modified nucleotides can also be used to form linkages. Upon in situ oxidation (e.g., with N-

25 bromosuccinimide (NBS)), the furan moiety forms a reactive oxo-enal derivative that reacts with a complementary base to form an inter-strand linkage. In some embodiments, the furan-modified nucleotides forms linkages with a complementary A or C nucleotide. Exemplary, non-limiting furan-modified nucleotides include any 2'-(furan-2-yl)propanoylamino-modified nucleotide; or an acyclic, modified nucleotides of 2-(furan-2-yl)ethyl glycol nucleic acid.

Photo-reactive groups can also be used as a reactive group. Exemplary, non-limiting photo-reactive groups include an intercalating moiety, a psoralen derivative (e.g., psoralen, HMT-psoralen, or 8-methoxypsoralen), an optionally substituted cyanovinylcarbazole group, an optionally substituted vinylcarbazole group, an optionally substituted cyanovinyl group, an optionally substituted acrylamide group, an optionally substituted diazine group, an optionally substituted benzophenone (e.g., succinimidyl ester of 4-benzoylbenzoic acid or benzophenone

35 isothiocyanate), an optionally substituted 5-(carboxy)vinyl-uridine group (e.g., 5-(carboxy)vinyl-2'-deoxyuridine), or an optionally substituted azide group (e.g., an aryl azide or a halogenated aryl azide, such as succinimidyl ester of 4-azido-2,3,5,6-tetrafluorobenzoic acid (ATFB)).

Intercalating moieties can also be used as a reactive group. Exemplary, non-limiting intercalating moieties include a psoralen derivative, an alkaloid derivative (e.g., berberine, palmatine, coralyne, sanguinarine (e.g.,

40 iminium or alkanolamine forms thereof), or aristololactam- β -D-glucoside), an ethidium cation (e.g., ethidium bromide), an acridine derivative (e.g., proflavine, acriflavine, or amsacrine), an anthracycline derivative (e.g., doxorubicin, epirubicin, daunorubicin (daunomycin), idarubicin, and aclarubicin), or thalidomide.

For a cross-linking oligonucleotide, any useful reactive group (e.g., described herein) can be used to form inter- or intra-strand linkages. Exemplary reactive groups include chemical-reactive group, a photo-reactive group, an intercalating moiety, and a reversible co-reactive group. Cross-linking agents for use with cross-linking oligonucleotides include, without limitation, alkylating agents (e.g., as described herein), cisplatin (cis-
5 diamminedichloroplatinum(II)), trans-diamminedichloroplatinum(II), psoralen, HMT-psoralen, 8-methoxypsoralen, furan-modified nucleotides, 2-fluoro-deoxyinosine (2-F-dI), 5-bromo-deoxycytosine (5-Br-dC), 5-bromo deoxyuridine (5-Br-dU), 5-iodo-deoxycytosine (5-I-dC), 5-iodo-deoxyuridine (5-I-dU), succinimidyl trans-4- (maleimidylmethyl)cyclohexane-1-carboxylate, SMCC, EDAC, or succinimidyl acetylthioacetate (SATA).

Oligonucleotides can also be modified to contain thiol moieties that can be reacted with a variety of thiol
10 reactive groups such as maleimides, halogens, iodoacetamides and thus can be used for cross-linking two oligonucleotides. The thiol groups can be linked to the 5'- or the 3'- terminus of an oligonucleotide.

For inter-strand cross-linking between duplex oligonucleotides at a pyrimidine (e.g., thymidine) position, the intercalating, photo-reactive moiety psoralen can be chosen. Psoralen intercalates into the duplex and forms covalent inter-strand cross-links with pyrimidines, preferentially at 5'-TpA sites, upon irradiation with ultraviolet
15 light (about 254 nm). The psoralen moiety can be covalently attached to a modified oligonucleotide (e.g., by an alkane chain, such as a C₁₋₁₀ alkyl, or a polyethylene glycol group, such as as -(CH₂CH₂O)_nCH₂CH₂-, where n is an integer from 1 to 50). Exemplary psoralen derivatives can also be used, where non-limiting derivatives include 4'- (hydroxyethoxymethyl)-4,5',8-trimethylpsoralen (HMT-psoralen) and 8-methoxypsoralen.

Various portions of the cross-linking oligonucleotide can be modified to introduce a linkage. For example,
20 terminal phosphorothioates in oligonucleotides can also be used for linking two adjacent oligonucleotides. Halogenated uracils/cytosines can also be used as cross-linker modifications in the oligonucleotide. For example, 2-fluoro-deoxyinosine (2-F-dI) modified oligonucleotides can be reacted with disulfide-containing diamines or thiopropylamines to form disulfide linkages.

As described below, reversible co-reactive groups include those selected from a cyanovinylcarbazole
25 group, a cyanovinyl group, an acrylamide group, a thiol group, or a sulfonylethyl thioethers. An optionally substituted cyanovinylcarbazole (CNV) group can also be used in oligonucleotides to cross-link to a pyrimidine base (e.g., cytosine, thymine, and uracil, as well as modified bases thereof) in complementary strands. CNV groups promote [2+2] cycloaddition with the adjacent pyrimidine base upon irradiation at 366 nm, which results in an inter-strand cross-link. Irradiation at 312 nm reverses the cross-link and thus provides a method for reversible cross-
30 linking of oligonucleotide strands. A non-limiting CNV group is 3-cyanovinylcarbozaole, which can be included as a carboxyvinylcarbazole nucleotide (e.g., as 3-carboxyvinylcarbazole-1'-β-deoxyribose-5'-triphosphate).

The CNV group can be modified to replace the reactive cyano group with another reactive group to provide an optionally substituted vinylcarbazole group. Exemplary non-limiting reactive groups for a vinylcarbazole group include an amide group of -CONR_{N1}R_{N2}, where each R_{N1} and R_{N2} can be the same or different and is independently
35 H and C₁₋₆ alkyl, e.g., -CONH₂; a carboxyl group of -CO₂H; or a C₂₋₇ alkoxy carbonyl group (e.g., methoxycarbonyl). Furthermore, the reactive group can be located on the alpha or beta carbon of the vinyl group. Exemplary vinylcarbazole groups include a cyanovinylcarbazole group, as described herein; an amidovinylcarbazole group (e.g., an amidovinylcarbazole nucleotide, such as 3-amidovinylcarbazole-1'-β-deoxyribose-5'-triphosphate); a carboxyvinylcarbazole group (e.g., a carboxyvinylcarbazole nucleotide, such as 3-carboxyvinylcarbazole-1'-β-
40 deoxyribose-5'-triphosphate); and a C₂₋₇ alkoxy carbonyl vinylcarbazole group (e.g., an alkoxy carbonyl vinylcarbazole nucleotide, such as 3-methoxycarbonyl vinylcarbazole-1'-β-deoxyribose-5'-triphosphate). Additional optionally substituted vinylcarbazole groups and nucleotides having such groups are

provided in the chemical formulas of U.S. Patent No. 7,972,792 and Yoshimura and Fujimoto, *Org. Lett.* 10:3227-3230 (2008), which are both hereby incorporated by reference in their entirety.

Other reversible reactive groups include a thiol group and another thiol group to form a disulfide, as well as a thiol group and a vinyl sulfone group to form a sulfonylethyl thioethers. Thiol-thiol groups can optionally include a linkage formed by a reaction with bis-((N-iodoacetyl)piperazinyl)sulfonerhodamine. Other reversible reactive groups (e.g., such as some photo-reactive groups) include optionally substituted benzophenone groups. A non-limiting example is benzophenone uracil (BPU), which can be used for site- and sequence-selective formation of an interstrand cross-link of BPU-containing oligonucleotide duplexes. This cross-link can be reversed upon heating, providing a method for the reversible cross-linking of two oligonucleotide strands.

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).

For any of the groups described herein (e.g., a chemical-reactive group, a photo-reactive group, an intercalating moiety, a cross-linking oligonucleotide, or a reversible co-reactive group), the group can be incorporated at or in proximity to the terminus of an oligonucleotide or between the 5'- and 3'-termini. Furthermore, one or more groups can be present in each oligonucleotide. When pairs of reactive groups are required, then oligonucleotides can be designed to facilitate a reaction between the pair of groups. In the non-limiting example of a cyanovinylcarbazole group that co-reacts with a pyrimidine base, the first oligonucleotide can be designed to include the cyanovinylcarbazole group at or in proximity to the 5'-terminus. In this example, a second oligonucleotide can be designed to be complementary to the first oligonucleotide and to include the co-reactive pyrimidine base at a position that aligns with the cyanovinylcarbazole group when the first and second oligonucleotide hybridizes. Any of the groups herein and any of the oligonucleotides having one or more groups can be designed to facilitate reaction between the groups to form one or more linkages.

Bifunctional spacers

The bifunctional spacer between the headpiece and the chemical entity can be varied to provide an appropriate spacing moiety and/or to increase the solubility of the headpiece in organic solvent. A wide variety of spacers are commercially available that can couple the headpiece with the small molecule library. The spacer 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), or combinations thereof.

The bifunctional spacer may provide an appropriate spacing moiety between the headpiece and a chemical entity of the library. In certain embodiments, the bifunctional spacer 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., amino-modified dT), an amidite to modify the 5' or 3'-terminus of a single-stranded headpiece (achieved by means of standard oligonucleotide chemistry), chemical-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 may be a chemically inert spacing moiety of variable length, introduced between Part 1 and 2. Such a spacing moiety 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 spacer 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 design may be varied with the spacer design to facilitate library synthesis in organic solvents. For example, the headpiece and spacer combination is designed to have appropriate residues wherein the octanol:water coefficient (P_{oc}) is from, e.g., 1.0 to 2.5.

Spacers 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%, 98%, 99%, or 100% organic solvent. The spacer 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 spacers 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 spacers include amino-carboxylic spacers, such as those being peptides (e.g., Z-Gly-Gly-Gly-Osu (N-alpha-benzyloxycarbonyl-(Glycine)₃-N-succinimidyl ester) or Z-Gly-Gly-Gly-Gly-Gly-Gly-Osu (N-alpha-benzyloxycarbonyl-(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); chemical-reactive pair spacers, such as those chemical-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 spacers, 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-pyridyl]dithio)-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 chemical-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 spacers 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 spacers herein can be added in tandem to one another in different combinations to generate spacers of different desired lengths.

Spacers may also be branched, where branched spacers 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).

Enzymatic ligation and chemical ligation techniques

Various ligation techniques can be used to add scaffolds, building blocks, spacers, linkages, 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 enzymatic ligation and/or chemical ligation. These binding steps can include the addition of one or more tags to the headpiece or complex; the addition of a spacer 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, enzymatic 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, CircLigase™ ssDNA ligase, CircLigase™ II ssDNA ligase, and ThermoPhage™ ssDNA ligase (Prokazyme Ltd., Reykjavik, Iceland).

Chemical ligation can also be used to produce oligonucleotides capable of being transcribed or reverse transcribed. 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. In particular embodiments, chemical ligation includes the use of one or more chemical-reactive pairs to provide a spacing moiety that can be transcribed or reverse transcribed. In particular, reactions suitable for chemical-reactive pairs are preferred candidates for the ligation process (Kolb et al., *Angew. Chem. Int. Ed.*, 40:2004-2021 (2001); Van der Eycken et al., *QSAR Comb. Sci.*, 26:1115-1326 (2007)). In one embodiment, the ligated oligonucleotides contain a linkage that polymerases have a reduced ability to read or translocate through, e.g. an “unreadable” linkage.

Reaction conditions to promote enzymatic ligation or chemical ligation

The methods described herein can include 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 different lengths and varying the concentration of the tags; using different types of ligases, as well as combinations thereof (e.g., CircLigase™ 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, $\text{Co}(\text{NH}_3)_6\text{Cl}_3$, and yeast inorganic pyrophosphate; using enzymatically or chemically phosphorylated oligonucleotide tags; using 3'-protected tags; and using 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 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 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 CircLigase™ ssDNA ligase (EPICENTRE Biotechnologies, Madison, WI), CircLigase™ II ssDNA ligase (also from EPICENTRE Biotechnologies), ThermoPhage™ 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 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); 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; and 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. Chemical ligation methodologies can include any chemical method that enables the operative association of two encoding tags, whether or not a polymerase is able to read or translocate through the operatively-associated linkage.

Methods for determining the nucleotide sequence of a complex

This invention features a method for determining the nucleotide sequence of a complex, 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.

Figures 1-4 provide various exemplary methods for determining the nucleotide sequence of a complex. These figures provide exemplary methods to span unreadable linkage and/or release cross-linking oligonucleotides, which result in the formation of an oligonucleotide template.

Figure 1b provides an exemplary method using relay primers. In particular, the relay primers hybridize to the fixed constant sequence portions (e.g., the 5'- and 3'-connectors), where each sequence for the relay primers can be the same or different. The relay primers span the unreadable linkage, and extension proceeds along the readable portions of the complex (i.e., the encoding sequence, e.g., one or more tags) to produce oligonucleotide fragments. Then, the relay primers and fragments can be ligated to form a template, which can be optionally amplified and sequenced.

Figures 2-4 provide exemplary unreadable linkages. Figure 2 provides a linkage formed by psoralen, which reacts within a duplex formed by adjacent 5'- and 3'-connectors. Figure 3 provides a linkage formed by a cross-linked oligonucleotide, which hybridized to adjacent 5'- and 3'-connectors.

Any of the linkages described herein can be reversible or irreversible. Reversible linkages include photo-reactive linkages (e.g., a cyanovinylcarbazole group and thymidine, as in Figure 4) and redox linkages. Additional linkages are described herein.

In an alternative embodiment, an "unreadable" linkage can be enzymatically repaired in order to generate a readable or at least translocatable linkage. Enzymatic repair processes are well known to those skilled in the art and include, but are not limited to, pyrimidine (e.g., thymidine) dimer repair mechanisms (e.g., using a photolyase or a glycosylase (e.g., T4 pyrimidine dimer glycosylase (PDG))), base excision repair mechanisms (e.g., using a glycosylase, an apurinic/apyrimidinic (AP) endonuclease, a Flap endonuclease, or a poly ADP ribose polymerase (e.g., human apurinic/apyrimidinic (AP) endonuclease, APE 1; endonuclease III (Nth) protein; endonuclease IV; endonuclease V; formamidopyrimidine [fapy]-DNA glycosylase (Fpg); human 8-oxoguanine glycosylase 1 (α isoform) (hOGG1); human endonuclease VIII-like 1 (hNEIL1); uracil-DNA glycosylase (UDG); human single-strand selective monofunctional uracil DNA glycosylase (SMUG1); and human alkyladenine DNA glycosylase (hAAG)), which can be optionally combined with one or more endonucleases, DNA or RNA polymerases, and/or a ligases for the repair), methylation repair mechanisms (e.g., using a methyl guanine methyltransferase), AP repair mechanisms (e.g., using an apurinic/apyrimidinic (AP) endonuclease (e.g., APE 1; endonuclease III; endonuclease IV; endonuclease V; Fpg; hOGG1; and hNEIL1), which can be optionally combined with one or more endonucleases, DNA or RNA polymerases, and/or a ligases for the repair), nucleotide excision repair mechanisms (e.g., using excision repair cross-complementing proteins or excision nucleases, which can be optionally combined with one or more endonucleases, DNA or RNA polymerases, and/or a ligases for the repair), and mismatch repair mechanisms (e.g., using an endonuclease (e.g., T7 endonuclease I; MutS, MutH, and/or MutL), which can be optionally combined with one or more exonucleases, endonucleases, helicases, DNA or RNA polymerases, and/or ligases for the repair). Commercial enzyme mixtures are available to readily provide these kinds of repair mechanisms, e.g., PreCR® Repair Mix (New England Biolabs Inc., Ipswich MA), which includes Taq DNA Ligase, Endonuclease IV, Bst DNA Polymerase, Fpg, Uracil-DNA Glycosylase (UDG), T4 PDG (T4 Endonuclease V), and Endonuclease VIII.

Methods for tagging encoded libraries

This invention features a method for operatively associating 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 encoded 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 (e.g., by negative selection or positive selection). 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.

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, 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 member-associated encoding oligonucleotides using one or more restriction endonucleases; repair of the associated encoding oligonucleotides, e.g., with any repair enzyme, such as those described herein; ligation of one or more adapter sequences to one or both of the termini for library member-associated encoding oligonucleotides, 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 oligonucleotide 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.

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 the target. This method can optionally include amplifying the tags by methods known in the art. Exemplary biological targets include enzymes (e.g., kinases, phosphatases, 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, 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, chemical-reactive pairs (or functional groups) can be readily included in solid-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, such as use of one or more RNA ligases and/or DNA ligases; and chemical ligation, such as use of chemical-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 sequences, such as in a library-identifying tag, in a ligated 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 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 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.

5 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
10 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
15 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
20 processes.

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
25 U.S. Patent Application Publication No. 2007/0224607, hereby incorporated by reference.

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 spacer) to the
30 headpiece to form a complex. When the headpiece includes a spacer, the building block or scaffold is added to the end of the spacer. When the spacer is absent, the building block can be added directly to the headpiece or the building block itself can include a spacer that reacts with a functional group of the headpiece. Exemplary spacers 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
35 spacer 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 spacer 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
40 functional group of the complex.

Building blocks can be added in multiple, synthetic steps. For example, an aliquot of the headpiece, optionally having an attached spacer, 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, e.g., to a tailpiece. 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-binding sequence or provide a functional group to allow for binding (e.g., by ligation) of a primer-binding sequence. The primer-binding 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 spacer 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, spacers, linkages, and 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 or more RNA ligases and/or DNA ligases, as described herein; and chemical ligation, such as use of chemical-reactive pairs, as described herein.

Example 1

General strategy for complexes

The complexes of the invention can be formed to include one or more linkages which a polymerase has reduced ability to read or translocate through. For example, complexes can be formed to include (from the 5' to 3' direction) a chemical entity (star), a fixed constant sequence (e.g., a headpiece), followed by three variable encoding sequences (e.g., three tags), and another fixed constant sequence (e.g., a tailpiece), where the linkage between the fixed constant sequences and variable encoding sequences can be unreadable (Figure 1B).

Unreadable linkages can be formed by any useful moiety or functional group. A non-limiting example is psoralen (Figure 2), which can form an unreadable linkage between the 5' - and 3'-connectors.

Unreadable linkages can be formed by a cross-linking oligonucleotide with a reversible co-reactive group of a cyanovinylcarbazole group. The complex can include (from the 5' to 3' direction) a chemical entity (star), a fixed constant sequence (e.g., a headpiece), a cross-linking oligonucleotide that is cross-linked via a reversible co-reactive group of a cyanovinylcarbazole group (cross-linked at about 366 nm, marked with X), and followed by a variable encoding sequence having fixed constant sequences at the 5' - and 3'-termini (e.g., a tag having 5' - and 3'-connectors) (Figure 3). An exemplary reversible reaction with reversible co-reactive groups (a pair of a

cyanovinylcarbazole group and a thymidine), which can be present in the 3'-connector, 5'-connector, and cross-linking oligonucleotide (Figure 4). Other unreadable linkages may be rendered deconvolutable by the use of repair enzyme(s).

As part of the reading or deconvolution process, the complex can be reacted with a chemical process and/or an enzymatic process (e.g., a ligase and optionally a kinase) to ligate the tags. When the junction between adjacent tags is a nick, then ligase can be used to ligate the tags. Optionally, a kinase can be used to convert terminal 5'-hydroxyl groups to phosphate groups. When the junction between adjacent tags is a gap requiring extension with one or more nucleotides, then a polymerase can be used to extend across the gap prior to the ligation of the tags (optionally including kinase). Optionally the polymerase has reduced strand-displacement activity. Next, the cross-linking oligonucleotides are released to form a template (e.g., using irradiation at about 312 nm). Finally, the template is optionally amplified with PCR and sequenced.

For any of these complexes, each of the variable encoding sequences can optionally include a 5'-connector and a 3'-connector that are fixed constant sequences. Because the linkages are unreadable, relay primers can be used to span the linkages and to allow for extension with a polymerase, thus forming oligonucleotide fragments. The fragments can then be ligated using ligase, optionally amplified with PCR, and sequenced.

Alternatively encoding sequences on either side of an unreadable linkage may be copied into a single cDNA sequence using a recombination-mediated process in which recombination events are favored to occur for sequences derived from the same ancestral template sequence. This mechanism preserves tag association information and renders these associations deducible from derived sequence data. One possible method of generating unreadable linkages that may become templates for recombination-prone cDNA is to establish repeat homologous regions either side of the linkage.

Example 2

Photochemical oligonucleotide cross-linking using cyanovinyl carbazole to model one tagging event

Oligonucleotide CNVKJ, containing two cyanovinyl carbazole modifications, was obtained from Nihon Techno Service (Japan) (Figure 5). Oligonucleotides CNVK2_P_TagB, CNVK2_NP_TagB and CNVK2_TagA were obtained from Integrated DNA Technologies (IDT) (IA). Each oligonucleotide was dissolved in water to a concentration of 1 mM.

Photochemical cross-linking conditions were as follows: Oligonucleotides CNVKJ, CNVK2_P_TagB and CNVK2_TagA and were mixed at equimolar ratio at 100 uM each in 500 mM phosphate buffer at pH 7.0 in 20 uL or 50 uL aliquots in 1.5 mL natural color polypropylene microcentrifuge tubes (Fisher Scientific, 02-682-550). Each reaction was then irradiated using a UVL-21 (4 Watt) lamp (95-0018-02) (Ultra Violet Products, CA) at 365 nm for 2 hours at 4°C. The products of the reactions were then analyzed by electrophoresis on a 10% denaturing PAAG as well as by LCMS (Thermo Fisher Scientific) (Figure 6). Over 70% of the starting material was converted into the photochemically cross-linked oligonucleotide conjugate with an observed MW of 16,894 Da (calculated 16,890 Da). When the same cross-linking reaction was performed in 500 mM phosphate buffer at pH 5.5 or 500 mM Borate buffer at pH 9.5 similar results and yields were observed.

The photochemically cross-linked oligonucleotide conjugate of CNVK2_P_TagB, CNVK2_TagA and CNVKJ was purified on a 10% denaturing PAAG and was un-crosslinked by heating at 80°C in 50 uL aliquots in 1.5 mL natural color polypropylene microcentrifuge tubes (Fisher Scientific, 02-682-550) with irradiation (at 80°C) using a Spectroline E-Series EB-160C (6 Watt) lamp (Spectronics Corp., NY) at 312 nm for 1 hour. The products were analyzed by LCMS. The reaction quantitatively yields unaltered starting oligonucleotide model tags

CNVK2_P_TagB observed MW 6,578 Da (calculated 6,579 Da) and CNVK2_TagA observed MW 5,558 Da (calculated 5,558 Da) (Figure 7).

In order to study the dissociation of multiple CNVK-containing cross-linked oligonucleotides, we adapted the above protocol to prepare a complex with two CNVK oligos, CNVKJ and Splint_14_CNVK, (Splint_14_CNVK is CGAXCGTGTCAXCG, where X is CNVK, synthesized by Biosearch Technologies, CA), photochemically crosslinked to a single oligonucleotide, Double_CNVK_Template (AAAAAAGTCGTGACACGTCGGAAAAAAAAAAAAACGGTGACACGGTCGAAAAAA, IDT (IA)). This complex was analyzed by LCMS: observed MW 25,079 Da (calculated 25,055 Da). Then it was dissolved in water to 100 uM concentration and was irradiated by UV 312 nm for 1 hour at 80°C, again as described above. The products were analyzed by LCMS to find the recovery of the starting Double_CNVK_template: observed MW 16,492 Da (calculated 16,492 Da) (Figure 8).

Example 3

Ligation and photochemical un-cross-linking of cyanovinylcarbazole photochemically cross-linked oligonucleotides to model the recovery of sequence information from photochemically cross-linked tags

A complex of CNVK2_P_TagB, CNVK2_TagA and CNVKJ was prepared, photochemically cross-linked, and gel-purified as described in Example 2.

The photochemically cross-linked model tag pair was dissolved in 1x T4 DNA Ligase buffer (NEB, MA) and was incubated with T4 DNA ligase (NEB, MA) for 1 hour at 37°C. The product of this reaction was then uncross-linked as described Example 2 and was analyzed by denaturing gel electrophoresis and LCMS (Figure 9). The product of the ligation reaction was observed in high yield, observed MW 12,119 Da (calculated 12,119 Da).

The corresponding photochemically cross-linked non-phosphorylated model tag pair CNVK2_NP_TagB with CNVK2_TagA and CNVKJ was also prepared, as described above. The LCMS analysis of this complex revealed an observed MW of 16,503 Da (calculated 16,499 Da, Figure 10). The photochemically cross-linked complex was dissolved in 1x T4 Ligase buffer (NEB) and was incubated with a mixture of T4 Polynucleotide Kinase and T4 DNA Ligase (NEB) for 1 hour at 37°C. The product of this coupled phosphorylation-ligation reaction was then photochemically uncross-linked by irradiation with UV at 312 nm for 1 hour at 80°C as described in Example 2. The product of this phosphorylation-ligation reaction was observed by denaturing gel electrophoresis and LCMS, revealing an observed MW of 11,894 Da (calculated MW 11,896 Da (additional phosphorylation at the 5' end of the ligated product)) (Figure 10).

Example 4

Photochemical oligonucleotide cross-linking using psoralen within a stem to model one tagging event

Oligonucleotide 5PSO2_A9_TA, modified with C2-Psoralen at the 5'-terminus was obtained from Biosearch, CA (Figure 11). Oligonucleotide PSO_HP_A9_TCT was obtained from IDT (IA) (Figure 11). Both oligonucleotides were dissolved to 0.5 mM concentration each in 500 mM phosphate buffer at pH 7.0 and were annealed by heating to 95°C and subsequently cooling to room temperature. The reaction was then irradiated with UV light at 365 nm at 4°C using a UVL-21 compact UV lamp (UVP) for 30 minutes. The products were analyzed by LCMS. We observed rapid formation of the photochemically cross-linked product with an observed MW of 17,643 Da (calculated MW 17,643.4 Da, see Figure 12).

Example 5

Photochemical oligonucleotide ligation using psoralen to model one tagging event

Oligonucleotides Tag1_PsoCVU and SplintC_PsoC2 were obtained from IDT (IA) (Figure 13).

Oligonucleotide 5PSOC2_A9_GA (modified at 5'-terminus with C2 psoralen) was obtained from Biosearch (CA) (Figure 13).

The photochemically ligated oligonucleotide conjugate of Tag1_PsoCVU and 5PSOC2_A9_GA was prepared as follows. Photoligation is achieved by the photochemical reaction of the 5'-psoralen pyrone moiety of one oligonucleotide with the 3' thymidine of a second oligonucleotide in the presence of a third oligonucleotide designed to hybridize to both of the first and second oligonucleotides and co-locate their reactive termini (Figure 14).

Photochemical ligation conditions were as follows: Oligonucleotides 5PsoC2_A9_GA (final 1 mM) and Tag1_PsoCVU (final 1.1 mM) and SplintC_PsoC2 (final 1.1 mM) were combined in a 10 uL photoligation reaction in 500 mM Phosphate buffer at pH 7.0 in 1.5 mL natural color polypropylene microcentrifuge tubes (Fisher Scientific, 02-682-550). The reaction mixture was first heated to 95°C followed by slow cooling to room temperature. The reaction was then irradiated with UV light at 365 nm at 4°C using a UVL-21 compact UV lamp (UVP). Aliquots of 1 uL were taken over a time-course and were analyzed on a 15% TBE/8M Urea denaturing polyacrylamide gel (Figure 15). The gel was visualized and photographed using the UV shadowing of a fluorescent TLC plate. The photoligated conjugate of 5PsoC2_A9_GA and Tag1_PsoCVU formed at high yield. After 24 hours the yield is approximately 90% (Figure 15).

The 24-hour sample was analyzed by LCMS (Figure 16). The major LC peak corresponded to the photoligated conjugate of 5PsoC2_A9_GA and Tag1_PsoCVU with an observed MW of 15,252.4 Da (calculated MW 15,243.1 Da).

Example 6

cDNA generation from non-polymerase-readable photochemically-ligated psoralen-conjugated model tags using terminal primer extension, non-terminal primer extension, and ligation to model the reading of a single tagging event

A Psoralen-Thymidine photochemically ligated oligonucleotide conjugate (Bio_Pso) was generated using a method similar to that described in Example 5 with the following two model tag oligonucleotides: a 5'-biotinylated oligonucleotide 5Bio_Tag_PsoCVU and a 5'-Psoralen oligonucleotide 5PsoC2_A9_GA along with a splint oligonucleotide SplintC_PsoC2, all three oligonucleotides were obtained from IDT (IA) and are shown in Figure 17. After 24 hrs of the reaction Bio_Pso was purified on a 10% TBE-Urea denaturing PAAG and analyzed by LCMS, revealing a product with an observed MW of 15,834.0 Da (calculated MW 15,812.7 Da) (Figure 18).

The Psoralen-Thymidine photochemically ligated oligonucleotide conjugate Bio_Pso was used as a template to generate cDNA by the enzymatic extension of both a terminal oligonucleotide primer (FAMprimer) and a non-terminal oligonucleotide primer (Phos-SplintC_PsoC2) with T4 DNA polymerase (NEB, MA). Phos-SplintC_PsoC2 was generated by the 5'-phosphorylation of SplintC_PsoC2 using T4 PNK (NEB, MA). The cDNA generation was conducted in a 100 uL reaction containing 10 uM each of FAMprimer, Phos-SplintC_PsoC2 and Bio_Pso in 1x T4 DNA ligase buffer (NEB, MA), supplemented with 1 mM of each dNTP and 10 uL of T4 DNA polymerase (NEB, MA). The reaction mixture was incubated at 37°C for 1 hour and was then supplemented with 0.5 mM ATP and 5 uL of T4 DNA ligase (2,000 u/uL, NEB, MA) and then incubated for another 1 hour at 37°C.

The reaction mixture was then incubated with 200 μ L of Streptavidin-coated DynaBeads M280 (Invitrogen, pre-washed with PBS for 1 hour at room temperature), the beads were washed with 1 mL of PBS and the product was eluted with 35 μ L of 100 mM NaOH. The eluent was then neutralized by the addition of 5 μ L of 1 M Tris HCl pH 7.0 and was analyzed by LCMS.

5 LCMS analysis of the resultant product, with detection at 495nm, showed that approximately 50% of the FAMprimer had been extended to the full-length complementary sequence to the Bio_Pso template minus a single dA nucleotide (observed MW 15,170 Da, expected full-length MW 15,458 Da). Most of the rest of the FAMprimer was extended up until the photochemical ligation junction (observed MW 11,154 Da, calculated MW 11,154 Da) (Figure 19).

10 When the cDNA-generation, described above, was performed without addition of the ligase enzyme, the full-length complementary sequence (minus a single dA) was not observed (Figure 20).

Example 7

Photochemical oligonucleotide ligation using carboxyvinyl uridine to model one tagging event

15 Oligonucleotides CVU_G and CVU_A were synthesized by Trilink (CA) (Figure 21). Both oligonucleotides are 5'-modified with a 5-(Carboxy)vinyl-2'-deoxyuridine (Figure 21). Oligonucleotides Tag1_PsoCVU, SplintC_CVU and SplintA_CVU were obtained from IDT DNA (IA) (Figure 21).

The structure of the photochemical ligation junction formed between an irradiated 5'-5-(Carboxy)vinyl-2'-deoxyuridine and a 3'-Thymidine is shown in Figure 22.

20 A 10 μ L photochemical ligation reaction mixture was generated with 1 mM CVU_G, 1.1 mM Tag1_PsoCVU and 1.1 mM SplintC_CVU dissolved in 500 mM Sodium Phosphate buffer pH 7.0 in a 1.5 ml natural color polypropylene microcentrifuge tubes (Fisher Scientific, 02-682-550). The reaction mixture was heated to 95°C followed by slow cooling to the room temperature. The reaction was then irradiated by UV light at 365 nm for 24 hours at 4°C using a UVL-21 compact UV lamp (UVP). Aliquots of 1 μ L were taken over a time-course and
25 were electrophoresed on a 15% TBE/ 8M Urea denaturing polyacrylamide gel (Figure 23). The gel was visualized and photographed by UV shadowing of a fluorescent TLC plate. After 24 hours of irradiation over 90% of the CVU_G oligonucleotide had been photochemically ligated.

The 24-hour sample was analyzed by LCMS (Figure 24). The major peak corresponds to the photochemical ligation product of CVU_G with Tag1_PsoCVU. The observed MW was 15,260.6 Da (calculated
30 MW 15,238.7 Da).

Example 8

cDNA generation from non-polymerase-readable photochemically-ligated carboxyvinyl uridine-conjugated model tags using terminal primer extension, non-terminal primer extension, and ligation to model the reading 35 of a single tagging event

A Carboxyvinyl Uridine-Thymidine photochemically-ligated oligonucleotide conjugate (Bio_CVU) was generated using a method similar to that described in Example 7 with the following two model tag oligonucleotides: A 5'-5-Carboxyvinyl Uridine oligonucleotide 5'- CVU_A and a 5'-Biotinylated oligonucleotide 5Bio_Tag_PsoCVU along with a splint oligonucleotide SplintA_CVU, all three oligonucleotides were obtained from IDT (IA) and are
40 shown in Figures 21 and 25. The reaction product was gel-purified as described in Example 6 and LCMS analysis revealed a product with an observed MW of 15,811.8 Da (calculated MW 15,792.3 Da) (Figure 25).

The Carboxyvinyl Uridine -Thymidine photochemically-ligated oligonucleotide conjugate Bio_CVU was used as a template to generate cDNA by the enzymatic extension of both a terminal oligonucleotide primer (FAMprimer) and a non-terminal oligonucleotide primer (Phos-SplintA_CVU) with T4 DNA polymerase (NEB, MA). Phos-SplintA_CVU was generated by the 5'-phosphorylation of SplintA_CVU using T4 PNK (NEB, MA).

5 The cDNA generation was conducted in a 100 uL reaction containing 10 uM each of FAMprimer, Phos-SplintA_CVU and Bio_CVU in 1x T4 DNA ligase buffer (NEB, MA), supplemented with 1 mM of each dNTP and 10 uL of T4 DNA polymerase (NEB, MA). The reaction mixture was incubated at 37°C for 1 hour and was then supplemented with 0.5 mM ATP and 5 uL of T4 DNA ligase (2,000 u/uL, NEB, MA) and then incubated for another 1 hour at 37°C.

10 The reaction mixture was then incubated with 200 uL of Streptavidin-coated DynaBeads M280 (Invitrogen, pre-washed with PBS for 1 hour at room temperature), the beads were washed with 1mL of PBS and the product was eluted with 35 ul of 100 mM NaOH. The eluent was then neutralized by the addition of 5 uL of 1 M Tris HCl pH 7.0 and was analyzed by LCMS.

15 LCMS analysis of the resultant product, with detection at 495nm, showed that approximately 80% of the FAMprimer had been extended to the full-length complementary sequence to the Bio_CVU template minus a single dA nucleotide (observed MW 15,176 Da, calculated full-length MW 15,458 Da). Most of the rest of the FAMprimer was extended up until three nucleotides before the photochemical ligation junction (observed MW 10,560 Da, calculated MW 10,559 Da) (Figure 26).

20 When the cDNA-generation, described above, was performed without addition of the ligase enzyme, the full-length complementary sequence (minus a single dA) was not observed.

Example 9

Oligonucleotide ligation using a non-polymerase-readable click-conjugation to model one tagging event

Oligonucleotides S_IDTN₃ and S_IDTalkyne were obtained from IDT (IA) (Fig 27).

25 The 3'-azide modification of oligonucleotide S_IDTN₃ is an IDT proprietary modification (3' IDT azide modification /3AzideN/ (MW 350)).

Oligonucleotides containing the /3AzideN/ modification were HPLC-purified.

Oligonucleotides containing the Hexynyl modification were desalted and used without further purification.

30 Click-conjugation of these model tags was achieved by a Cu (I) catalyzed 3+1 azido alkyne cycloaddition as follows:

Stock solutions were prepared as follows: Cu(OAc)₂·H₂O (FW 181.6), 10 mM in DMF; Sodium Ascorbate (FW 198.1), 20 mM in H₂O and Tris-(BenzylTriazolylmethyl)Amine (TBTA) (FW 530.6), 10 mM in DMF. To an Eppendorf tube was added 10 uL pH 7.0 phosphate buffer (500 mM aqueous solution), 10 uL of S_IDTalkyne (1 mM aqueous solution, 10 nmol, 1 equivalent) and 10 uL S-IDTN₃ (1 mM aqueous solution, 10 nmol, 1 molar equivalent). To this solution was added 5 uL of Cu_Pre-mix (2 molar equivalents of Cu(OAc)₂, 4 molar equivalents of sodium ascorbate, 1 molar equivalent of TBTA). The reaction was incubated at room temperature overnight. The product was analyzed by LCMS (Thermo) (Figure 28) and the MW was found to be 26,304.8 Da (calcd: 26311.2 Da). The conjugate was further purified by electrophoresis using 10% PAAG/8M urea. This conjugation product was named Conjugate_Click_S.

40

Example 10**Oligonucleotide ligation using a pair of orthogonal non-polymerase-readable click-conjugations to model a pair of orthogonal tagging events**

TKR_Central and S_IDTN₃ oligos were obtained from IDT (Figure 29). The 5'-dibenzo-cyclooctyne (DBCO)-modified oligo TKR_DBCO_S was obtained from Biosearch Technologies (CA) (Figure 29).

Firstly, a click conjugation was performed with TKR_Central and TKR_DBCO_S using a cyclooctyne-based copper-free click reaction. Equal molar amounts of each oligonucleotide (10 uL of 1 mM aqueous solutions) were mixed together in 0.2M sodium phosphate buffer at pH 7.0. The reaction progress was monitored by LCMS. The reaction was complete in 30 to 60 min, yielding over 90% of the pure conjugate. MS indicated a MW of 25,336 Da (calculated 25,341.8 Da) (Figure 30).

Secondly, a click conjugation was performed between the conjugation product of TKR_Central and TKR_DBCO_S with S_IDTN₃ using the Cu (I) catalyzed click procedure described in Example 9. The resulting final conjugate, 2c1_S was purified on a 10% TBE/8M urea polyacrylamide gel and was analyzed by LCMS (Thermo) revealing the expected molecular weight for 2c1_S 38578.1 Da, (calculated 38560.4 Da) (Figure 31).

Example 11**PCR-amplification of cDNA derived from non-polymerase-readable click-conjugated model tags using non-terminal-primer-mediated recombination in free solution and in emulsified aqueous droplets to model the reading of sequence information from chemically conjugated tags**

Conjugate_Click_S (Figure 32), was prepared by the conjugation of S_IDTN₃ and S_IDTalkyne, as described in Example 9.

Conjugate_Click_L (Figure 32) was prepared by the conjugation of L_IDTN₃:
 TGCGGTCTAACTGTCTAGGCACTTGTTGCGTTTGCCAGTGTGAGGAATGAACAGG/3AzideN/ and
 L_IDTalkyne:
 /5hexynyl/TACCGAATTGCCCTTCCTCGTACAGTTCTAAGGCGCTTGGACACCACTTCAATCGGGTGCTATG
 CTT (IDT, IA) similarly to the conjugation protocol described in Example 10. LCMS after gel purification confirmed the MW as 37,582.6 Da (calculated 37,584.4 Da).

Oligodeoxynucleotide primers, including forward primer: TGCGGTCTAACTGTCTA, reverse primer: AAG CATAGCACCCGATT and ePsplint: GGCAATTCGGTACCTGTTTCATTCC were obtained from (IDT, IA).

The conjugates Conjugate_Click_L and Conjugate_Click_S were each diluted to 1 nM and were used as templates for a splint-dependent-recombination-mediated cDNA generation/PCR amplification using Deep Vent (exo-) DNA polymerase (NEB, MA). A 50 uL Deep Vent (exo-) reaction was prepared as follows: 5 uL of 10x Thermopol buffer; 2.5 uL of a 10 uM solution of each forward and reverse primer, 1 uL of 100 nM ePsplint primer; 2.5 uL 10 mM dNTP mix (NEB); 5 uL of a 2 nM dilution of either conjugate or their 1:1 mixture, 2.5 uL Deep Vent (exo-) and water to 50 uL. The reaction was cycled 24 times as follows: 30 seconds at 94°C, 30 seconds at 52°C and 60 seconds at 72°C.

When Conjugate_Click_L or Conjugate_Click_S conjugates are amplified separately in free solution in the presence of both terminal primers and the splint primer "ePsplint" and then evaluated by electrophoresis, each gives rise to a single product of with a mobility that correlates with the length of the template (lanes 5 and 7 in Figure 33). Similar amplifications in the absence of the splint primer "ePsplint" gave no product (lanes 4 and 6 in Figure 33). This suggests that splint-dependent-recombination-mediated cDNA generation/PCR amplification of both

conjugates by Deep Vent (exo-) polymerase is dependent on the presence of ePsplint primer (Figure 33). A possible mechanism for this recombination and amplification is presented in Figure 34.

When both Conjugate_Click_L and Conjugate_Click_S conjugates are mixed and amplified together in free solution in the presence of both terminal primers and the splint primer "ePsplint" and then evaluated by electrophoresis, a ladder of different-mobility products are generated with mobilities that correlate to the lengths of each of the templates as well as with mobilities that correlate to lengths intermediate between each of the templates (lane 5 in Figure 35). A possible mechanism for this recombination and amplification is presented in Figure 36.

When both Conjugate_Click_L and Conjugate_Click_S conjugates are mixed and amplified together in a water-in-oil emulsion in the presence of both terminal primers and the splint primer "ePsplint" at low concentration of conjugates relative to emulsion droplets, and then evaluated by electrophoresis, a pair of products are observed with mobilities that correlate to the lengths of each of the templates (lane 3 in Figure 37). Droplets of the emulsion act as individual microreactors, within which isolated individual conjugates are amplified. The oil/surfactant phase was prepared by mixing 4.5% (vol/vol) of Span 80 (Fluka, cat#85548), 0.4% (vol/vol) of Tween 80 (Sigma, cat#S-8074) and 0.05% (vol/vol) of Triton X-100 (Sigma, cat# T-9284) in light mineral oil (Sigma, cat# M-3516). The aqueous phase (50 uL) was made by mixing 5 uL of 10x Thermopol buffer; 2.5 uL of a solution of each 10 uM primer, 1 uL of 100 nM ePsplint primer; 2.5 uL 10 mM dNTP mix (NEB); 2.5 uL DVexo- (NEB), 2.5 uL of BSA (NEB) and water to 50 uL. A 1:1 mixture of the conjugates Conjugate_Click_L and Conjugate_Click_S were added to the reaction to a final concentration of 5 pM. To 50 uL of the aqueous phase, 150 uL of the non-aqueous phase was then added and emulsification was performed by vortexing for 5 minutes. The recombination/amplification process was conducted as follows: 30 seconds at 94°C, 30 seconds at 52°C and 60 seconds at 72°C for 32 cycles.

Following the splint-dependent-recombination-mediated cDNA generation/PCR amplification, the emulsion was broken by microcentrifugation at 21,000g for 20 minutes, then the non-aqueous phase was removed and the aqueous phase and interphase were extracted into chloroform. The aqueous phase was then evaluated by native electrophoresis (4% agarose minigels, Invitrogen CA).

The results of the emulsion-compartmentalized splint-dependent-recombination-mediated cDNA generation/PCR amplification show that the extent of formation of shuffled PCR products is significantly diminished or nonexistent (Lane 3, Figure 37).

Example 12

PCR amplification of cDNA derived from non-polymerase-readable click-conjugated model tags using repeat-homology-mediated-recombination in free solution to model the reading of a single tagging event

Two conjugates, TKR_S and TKR_L (Figure 38), each containing a single non-polymerase-readable click-conjugation linkage formed between a 3'-azido-modified-oligonucleotide and a 5'-hexynyl-modified-oligonucleotide using Cu(I) catalyzed cycloaddition and the protocol described in Example 9. TKR_S and TKR-L each contain a duplicated 12-mer GGAATGAACAGG sequence that flanks the non-polymerase-readable linkage.

TKR_S was prepared by conjugating the following two IDT DNA oligonucleotides, TKR_S_N₃:

TGCGGTCTAACTGTCTAGTTCACCTTCTCCGGAATGAACAGG/3AzidoN/ (also referred to as S_IDTN₃ in Example 9) and TKR_S_Hex: /5'-Hexynyl/GGAATGAACAGGGCCTGATTGAAGCAATCGGGTGTATGCTT, while TKR_L was prepared by conjugating two IDT DNA oligonucleotides TKR_L_N₃:

TGCGGTCTAACTGTCTAGGCACTTGTTCGTTTGCCAGTGTGATGGACACCACTTGGGAATGAACAGG/3AzideN/ and TKR_L_Hex:

/5'Hexynyl/GGAATGAACAGGGGAATGAACAGGTTCCCTCGTACAGTTCTAAGGCGCTCAATCGGGTGTCTA

TGCTT. Both conjugates were PAGE-purified under denaturing conditions and analyzed by LCMS to reveal molecular weights for TKR_S: 26,447.5 Da (26,449.2 Da calculated), and for TKR_L 37,582.6 Da (37,584.4 Da calculated).

5 The two conjugates were each diluted to 1 nM and were used for a repeat-dependent-recombination-mediated cDNA generation/PCR amplification with Deep Vent (exo-) DNA polymerase (NEB) in free solution with the forward primer TGCGGTCTAACTGTCTA and the reverse primer AAGCATAGCACCCGATT (IDT).

10 The repeat-dependent-recombination-mediated cDNA generation/PCR amplification reaction was performed using Deep Vent (exo-) DNA polymerase (NEB). 5 uL of 10x Thermopol buffer; 2.5 uL of a 10 uM solution of each forward and reverse primer; 2.5 uL 10 mM dNTP mix (NEB); 40 pM final concentration of either conjugate or their 1:1 mixture, 2.5 uL Deep Vent (exo-) Polymerase and water to a final volume of 50 uL. The reaction was thermally cycled 22 times as follows: 30 seconds at 94°C, 30 seconds at 52°C and 60 seconds at 72°C.

The products were visualized on a 4% agarose minigel (Invitrogen) (Figure 39).

15 The Deep Vent (exo-) Polymerase repeat-dependent-recombination-mediated cDNA generation/PCR amplification reaction produced large amounts of amplicon derived from each separately amplified conjugate (Lanes 3 and 4, Figure 39). Amplification of the mixture of the conjugates under the same conditions did not result in the production of any intermediate-length amplification products (Lane 5, Figure 39). To further confirm the lack of shuffling (loss) of tag associations (encoded information) in these amplification products, we cloned and sequenced the amplification products derived from the co-amplified mixture of TKR_S and TKR_L. These sequence data confirmed the complete lack of shuffling of the model tag pairs in the amplification products derived from the co-amplified conjugates (Figure 40).

Example 13

PCR amplification of cDNA derived from unreadable twice-click-conjugated tags using repeat-homology-mediated-recombination in free solution to model the reading of a pair of tagging events

25 Two conjugates, 2_cl_S and 2_cl_L (Figure 41), each containing two unreadable click conjugation linkages, each flanked by different 12-mer repeat sequences (GAAGGCGTTATG-click-GAAGGCGTTATG) and (GGAATGAACAGG-click-GGAATGAACAGG), respectively, were prepared in a two-step procedure similar to that described in Example 10.

30 Firstly, a conjugation was performed using a cyclooctyne-based copper-free click reaction between TKR_central: /5'-Hexynyl/GGAATGAACAGGGTAAGCTGGAGTGAAGGCGTTATG/3azideN/ (IDT) and either TKR_DBCO_S: (DBCO)GAAGGCGTTATGTCCGTACTCTTGCAATCGGGTGCTATGCTT or TKR_DBCO_L (DBCO)GAAGGCGTTATGTGATATCCGTGGTGTCGTGAGTTCCAATCGGGTGCTATGCTT. Both DBCO-containing oligonucleotides were obtained from Biosearch (CA), DBCO is the 5'-dibenzo-cyclooctyne modification for which the phosphoramidite reagent is available from Glen Research (VA)). The products of these conjugations were named TKR_S_N₃ and TKR_L_N₃ respectively.

35 Secondly, a conjugation was performed between the products of the first conjugations, TKR_S_N₃ and TKR_L_N₃, using a Cu (I) catalyzed click reaction. The resulting conjugates, TKR_2_click_S and TKR_2_click_L (Figure 41) were purified using electrophoresis on a 10% TBE/8M urea polyacrylamide gel and were analyzed by LCMS (Thermo) revealing the expected molecular weights for TKR_2_click_S 38,578.1 Da (38,560.4 Da calculated), and for TKR_2_click_L 49,913.3 Da (49,916.7 Da calculated) (Figure 41).

The two conjugates were each diluted to 1 nM and were used for a repeat-dependent-recombination-mediated cDNA generation/PCR amplification with Deep Vent (exo-) Polymerase (NEB). For amplification the

forward primer TGCGGTCTAACTGTCTA and the reverse primer AAGCATAGCACCCGATT (IDT DNA) were used with 5 uL of 10x Thermopol buffer; 2.5 uL of a 10 uM solution of each forward and reverse primer; 2.5 uL 10 mM dNTP mix (NEB); 40 pM final concentration of either conjugate or their 1:1 mixture, 2.5 uL Deep Vent (exo-) Polymerase and water to 50 uL. The reaction was thermally cycled 22 times as follows: 30 seconds at 94°C, 30
5 seconds at 52°C and 60 seconds at 72°C. The amplification products were visualized on 4% agarose minigels (Invitrogen).

The Deep Vent (exo-) polymerase repeat-dependent-recombination-mediated cDNA generation/PCR amplification reaction produced large amounts of amplicon derived from each separately amplified conjugate (Lanes 3 and 4, Figure 42). Amplification of the mixture of the conjugates under the same conditions did not result in the
10 production of any intermediate-length amplification products (Lane 5, Figure 42). To further confirm the lack of shuffling (loss) of tag associations (encoded information) in these amplification products, we cloned and sequenced the amplification products derived from the co-amplified mixture of TKR_2_click_S and TKR_2_click_L. These sequence data confirmed the complete lack of shuffling of the model tag pairs in the amplification products derived from the co-amplified conjugates (Figure 43).

15

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
20 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.

25 What is claimed is:

Claims

1. A complex comprising:
 - (i) a chemical entity comprising one or more scaffolds or one or more building blocks;
 - (ii) a first oligonucleotide tag encoding the identity of at least one of said one or more scaffolds or building blocks; and
 - (iii) a headpiece having a first functional group and a second functional group, wherein said first functional group operatively associates with said chemical entity and said second functional group operatively associates with said first tag via a first linkage for which a polymerase has reduced ability to read or translocate through.
2. A complex comprising:
 - (i) a chemical entity comprising one or more scaffolds or one or more building blocks;
 - (ii) n number of oligonucleotide tags having $n-1$ linkages, wherein n is an integer between 1 and 10, and wherein each of said linkages is between two adjacent tags and each tag encodes the identity of at least one of said one or more scaffolds or building blocks; and
 - (iii) a headpiece having a first functional group operatively associated with said chemical entity and a second functional group operatively associated with at least one of said n number of tags via a first linkage,
wherein a polymerase has reduced ability to read or translocate through at least one of said first linkage and $n-1$ linkages.
3. The complex of claim 2, wherein said polymerase has reduced ability to read or translocate through at least about 10% of said first linkage and $n-1$ linkages.
4. The complex of any one of claims 1 to 3, wherein one or more tags comprise a 5'-connector at the 5'-terminus of said one or more tags and a 3'-connector at the 3'-terminus of said one or more tags.
5. The complex of claim 4, wherein each 5'-connector comprises the same sequence.
6. The complex of claim 4 or 5, wherein each 3'-connector comprises the same sequence.
7. The complex of claim 4, wherein each 5'-connector and/or each 3'-connector comprise a different sequence.
8. The complex of any one of claims 1 to 7, wherein said first linkage and/or $n-1$ linkages are at least about 3 angstroms in length.
9. The complex of any one of claims 1 to 8, wherein said first linkage and/or $n-1$ linkages are less than about 30 angstroms in length.

10. The complex of any one of claims 1 to 9, wherein a DNA polymerase and/or an RNA polymerase has reduced ability to read or translocate through said first linkage and/or *n-I* linkages.

11. The complex of any one of claims 1 to 10, wherein less than about 50% of said first linkage and *n-I* linkages comprise an enzymatic linkage.

12. The complex of any one of claims 1 to 11, wherein said first linkage and/or *n-I* linkages comprise a chemical linkage.

13. The complex of claim 12, wherein said chemical linkage comprises a chemical-reactive group, a photo-reactive group, an intercalating moiety, or a cross-linking oligonucleotide.

14. The complex of claim 13, wherein at least one chemical-reactive group, photo-reactive group, or intercalating moiety is present in a 5'-connector at or in proximity to the 5'-terminus of said tag and/or in a 3'-connector at or in proximity to the 3'-terminus of said tag.

15. The complex of claim 14, wherein the sequence of at least one of said 5'-connector is complementary to the sequence of the adjacent 3'-connector or is identical or is sufficiently similar to allow for hybridization to the same complementary oligonucleotide on either side of the junction in separate hybridization events.

16. The complex of any one of claims 13 to 15, wherein said chemical-reactive group is selected from a pair of an optionally substituted alkynyl group and an optionally substituted azido group; a pair of 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 pair of a nucleophile and a strained heterocyclyl electrophile; a pair of an optionally substituted amino group and an aldehyde or a ketone group; a pair of an optionally substituted amino group and a carboxylic acid group; a pair of an optionally substituted hydrazine and an aldehyde or a ketone group; a pair of an optionally substituted hydroxylamine and an aldehyde or a ketone group; a pair of a nucleophile and an optionally substituted alkyl halide; a platinum complex; an alkylating agent; or a furan-modified nucleotide.

17. The complex of any one of claims 13 to 15, wherein said photo-reactive group comprises an intercalating moiety, a psoralen derivative, an optionally substituted cyanovinylcarbazole group, an optionally substituted vinylcarbazole group, an optionally substituted cyanovinyl group, an optionally substituted acrylamide group, an optionally substituted diazirine group, an optionally substituted benzophenone, an optionally substituted 5-(carboxy)vinyl-uridine group, or an optionally substituted azide group.

18. The complex of any one of claims 13 to 15, wherein said intercalating moiety is a psoralen derivative, an alkaloid derivative, an ethidium cation, an acridine derivative, an anthracycline derivative, or thalidomide.

19. The complex of claim 18, wherein said psoralen derivative is psoralen, 8-methoxypsoralen, or 4-hydroxymethyl-4,5,8-trimethyl-psoralen.

20. The complex of claim 13, wherein said chemical linkage comprises said cross-linking oligonucleotide, wherein the sequence of at least five nucleotides at the 5'-terminus of said cross-linking oligonucleotide is complementary to the sequence of at least five nucleotides at the 3'-terminus of one or more tags, and wherein the sequence of at least five nucleotides at the 3'-terminus of said cross-linking oligonucleotide is complementary to the sequence of at least five nucleotides at the 5'-terminus of one or more tags.
21. The complex of claim 20, wherein said 3'-terminus of one or more tags comprises a 3'-connector.
22. The complex of claim 20 or 21, wherein said 5'-terminus of one or more tags comprises a 5'-connector.
23. The complex of any one of claims 20 to 22, wherein said 5'-terminus and/or 3'-terminus of said cross-linking oligonucleotide comprise a reversible co-reactive group.
24. The complex of claim 21 or 22, wherein said 3'-connector and/or 5'-connector comprise a reversible co-reactive group.
25. The complex of claim 23 or 24, wherein said reversible co-reactive group is a cyanovinylcarbazole group, a cyanovinyl group, an acrylamide group, a thiol group, or a vinyl sulfone group.
26. The complex of any one of claims 1 to 25, wherein said chemical entity is operatively associated to said headpiece via a bifunctional spacer.
27. The complex of any one of claims 1 to 26, wherein said chemical entity is covalently attached to said headpiece.
28. The complex of any one of claims 1 to 27, wherein said headpiece comprises an oligonucleotide selected from the group consisting of a double-stranded oligonucleotide, a single-stranded oligonucleotide, or a hairpin oligonucleotide.
29. The complex of claim 28, wherein said headpiece comprises a primer-binding region.
30. The complex of any one of claims 1 to 29, further comprising one or more first library-identifying tag(s), use tag(s), and/or origin tag(s).
31. The complex of any one of claims 1 to 30, wherein said complex comprises between 2 to 20 tags.
32. The complex of any one of claims 1 to 31, wherein each of said tags comprises from 5 to 75 nucleotides.

33. The complex of any one of claims 1 to 32, wherein each of said tags within an individual tag set comprises about the same mass.
34. The complex of any one of claims 1 to 33, wherein said complex comprises RNA, DNA, modified DNA, and/or modified RNA.
35. The complex of claim 34, wherein said modified DNA or modified RNA is PNA, LNA, GNA, TNA, or a mixture thereof within the same oligonucleotide.
36. The complex of any one of claims 1 to 35, further comprising a tailpiece.
37. A library comprising one or more complexes of any one of claims 1 to 36.
38. The library of claim 37, wherein said library comprises a plurality of headpieces.
39. The library of claim 38, wherein each headpiece of said plurality of headpieces comprises an identical sequence region and a different encoding region.
40. The library of claim 39, wherein said identical sequence region is a primer-binding region.
41. The library of claim 39 or 40, wherein said different encoding region is a first tag that encodes for use of the library, origin of the library, identity of the library, history of the library, a linkage, a spacer, or addition of a first component; or an oligonucleotide sequence that facilitates hybridization, amplification, cloning, or sequencing technologies.
42. The library of any one of claims 37 to 41, comprising between about 10^2 to 10^{20} complexes.
43. The library of any one of claims 37 to 42, wherein each complex is different.
44. A method of tagging a first library comprising an encoded chemical entity, said method comprising:
- (a) providing a headpiece having a first functional group and a second functional group;
 - (b) 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 spacer; and
 - (c) binding said second functional group of said headpiece to a first oligonucleotide tag via a first linkage to form a complex, wherein a polymerase has reduced ability to read or translocate through said first linkage;
- wherein said steps (b) and (c) can be performed in any order and wherein said first tag encodes for the binding reaction of said step (b), thereby providing a tagged library.

45. A method of tagging a first library comprising an encoded chemical entity, said method comprising:
- (a) providing a headpiece having a first functional group and a second functional group;
 - (b) 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 spacer;
 - (c) binding said second functional group of said headpiece to a first oligonucleotide tag via a first linkage;
 - (d) binding n_c number of additional components of said chemical entity, wherein n_c is an integer between 1 and 10; and
 - (e) binding n_t number of additional oligonucleotide tags having n_t linkages to form a complex, wherein n_t is an integer between 1 and 10 and wherein each of said linkages is between two adjacent tags and each tag encodes the identity of at least one of said components;
- wherein a polymerase has reduced ability to read or translocate through at least one of said first linkage and n_t linkages; and
- wherein said steps (b) and (c) can be performed in any order and wherein said first tag encodes for the binding reaction of said step (b); wherein said steps (d) and (e) can be performed in any order and wherein each additional tag encodes for the binding reaction of each additional component of said step (d), thereby providing a tagged library.
46. The method of claim 45, wherein said polymerase has reduced ability to read or translocate through at least about 10% of said first linkage and n_t linkages.
47. The method of any one of claims 44 to 46, wherein said first component and/or additional components comprise a scaffold or a building block.
48. The method of any one of claims 44 to 47, wherein step (b) comprises binding said headpiece indirectly to said first component via a bifunctional spacer.
49. The method of any one of claims 44 to 48, further comprising binding a use tag and/or an origin tag to said complex.
50. The method of any one of claims 44 to 49, further comprising binding a first library-identifying tag to said complex.
51. The method of claim 50, further comprising providing a second library and combining said first library with said second library.
52. The method of any one of claims 44 to 51, further comprising binding a tailpiece to said complex.
53. The method of any one of claims 44 to 52, wherein said binding step comprises a chemical-reactive group, a photo-reactive group, an intercalating moiety, or a cross-linking oligonucleotide.

54. The method of any one of claims 44 to 53, further comprising performing split-and-pool synthesis in one or more steps, separating one or more complexes, and/or purifying one or more complexes.

55. The method of any one of claims 44 to 53, further comprising a complex of any one of claims 1 to 36 and/or a library of any one of claims 37 to 43.

56. A method of screening a plurality of chemical entities, said method comprising:

(a) contacting a target with a complex of any one of claims 1 to 36 and/or a library of any one of claims 37 to 43; and

(b) selecting one or more complexes having a predetermined characteristic for said target, as compared to a control, thereby screening said chemical entity.

57. The method of claim 56, where said predetermined characteristic comprises increased binding for said target, as compared to a control.

58. The method of claims 56 or 57, wherein, prior to step (a), one or more operations are performed, said operations being selected from the list consisting of: annealing one or more relay primers with said complex, wherein said one or more relay primers span said first linkage and/or $n-1$ linkages; extending said relay primers using a polymerase to produce oligonucleotide fragments; and ligating said oligonucleotide fragments to produce a template.

59. A method of determining the nucleotide sequence of a complex of any one of claims 1 to 36, said method comprising:

(a) annealing one or more relay primers with said complex, wherein said one or more relay primers span said first linkage and/or $n-1$ linkages;

(b) extending said relay primers using a polymerase to produce oligonucleotide fragments;

(c) ligating said oligonucleotide fragments to produce a template;

(d) optionally amplifying said template by polymerase chain reaction to produce an amplified mix; and

(e) sequencing said optionally amplified mix to determine the sequence of said complex.

60. The method of claim 59, wherein said complex comprises a 5'-connector at the 5'-terminus of one or more tags and a 3'-connector at the 3'-terminus of one or more tags, and wherein each of said one or more relay primers hybridizes to adjacent 5'- and 3'-connectors.

61. The method of claim 60, wherein each 5'-connector comprises the same sequence.

62. The method of claim 61, wherein each 3'-connector comprises the same sequence.

63. The method of any one of claims 60 to 62, wherein the sequence of at least one of said 5'-connector is complementary to the sequence of the adjacent 3'-connector or is identical or is sufficiently similar to

allow hybridization to the same complementary oligonucleotide on either side of the junction in separate hybridization events.

64. The method of claim 63, further comprising hybridizing said 5'-connector to said adjacent 3'-connector.

65. The method of any one of claims 60 to 64, wherein each relay primer for a specific junction between the 5'- and 3'-connectors comprises the same sequence.

66. A method of determining the nucleotide sequence of a complex of any one of claims 1 to 36, said method comprising:

- (a) providing said complex comprising the chemical linkage, wherein said chemical linkage is a cross-linking oligonucleotide that spans the junction between two adjacent tags and that operatively associates with said two adjacent tags via one or more reversible co-reactive groups;
- (b) releasing said cross-linking oligonucleotide to produce a template;
- (c) optionally amplifying said template by polymerase chain reaction to produce an amplified mix; and
- (d) sequencing said optionally amplified mix to determine the sequence of said complex.

67. The method of claim 66, wherein a portion of said cross-linking oligonucleotide hybridizes to the termini of two adjacent tags, thereby producing a nick or a gap between said two adjacent tags.

68. The method of claim 67, further comprising ligating said tags using a chemical process or an enzymatic process prior to step (b).

69. The method of claim 68, wherein said chemical process or enzymatic process comprises 5'-phosphorylation.

70. The method of any one of claims 59 to 69, further comprising repairing said template prior to said optional amplifying step or said sequencing step.

71. The method of claim 70, wherein said repairing step comprises modifying at least one of said first linkage and/or *n-1* linkages to be a repaired linkage capable of being read or translocated through by a polymerase.

72. The method of claim 71, wherein said polymerase can read or translocate through at least 90% of said repaired linkages.

73. The method of any one of claims 70 to 72, wherein said repairing step comprises providing said template with a repair enzyme.

74. The method of claim 73, wherein said repair enzyme is a photolyase, a glycosylase, an endonuclease, a Flap endonuclease, an apurinic/apyrimidinic (AP) endonuclease, a poly ADP ribose polymerase, or a methyltransferase.

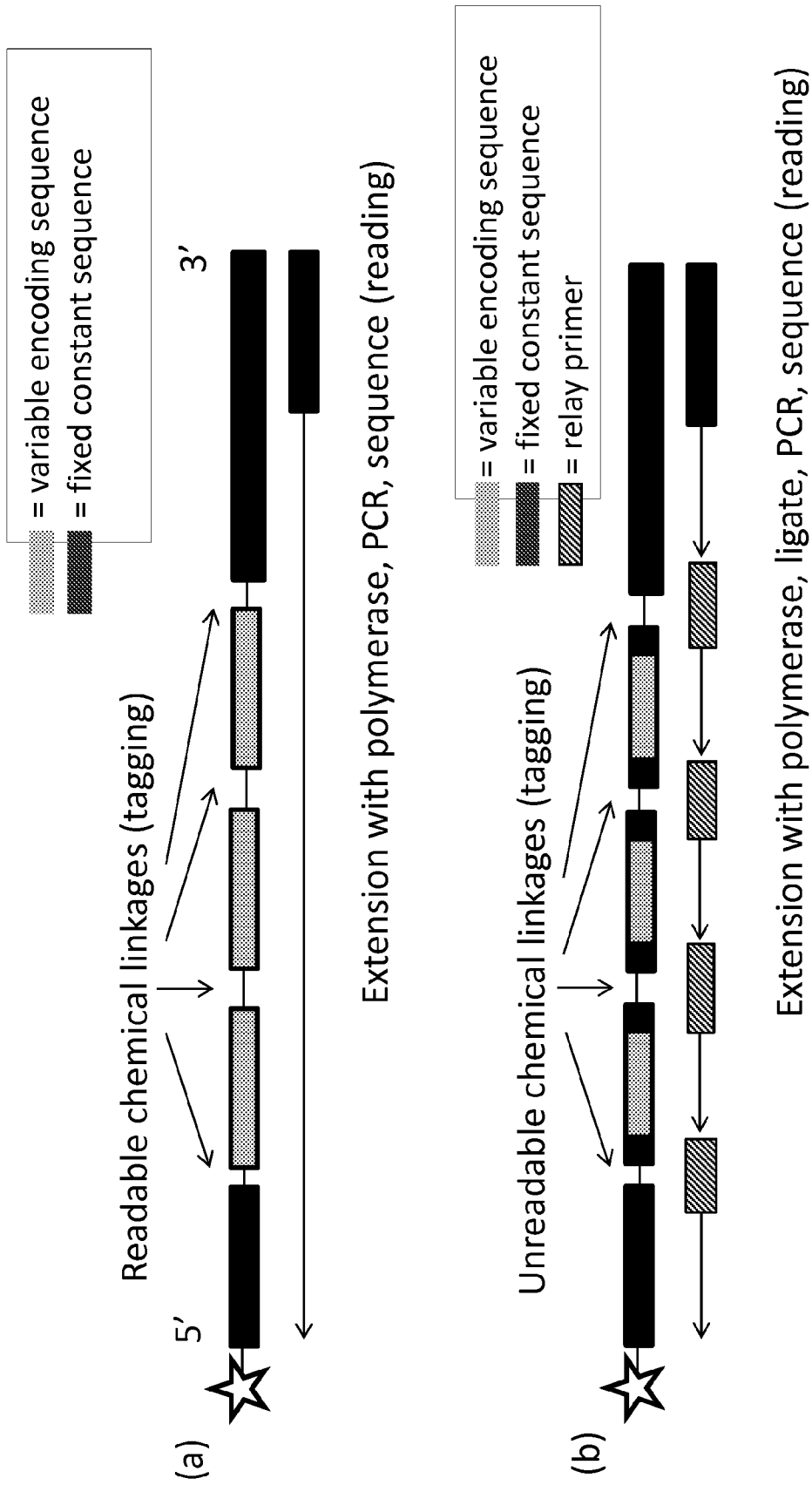


Figure 1

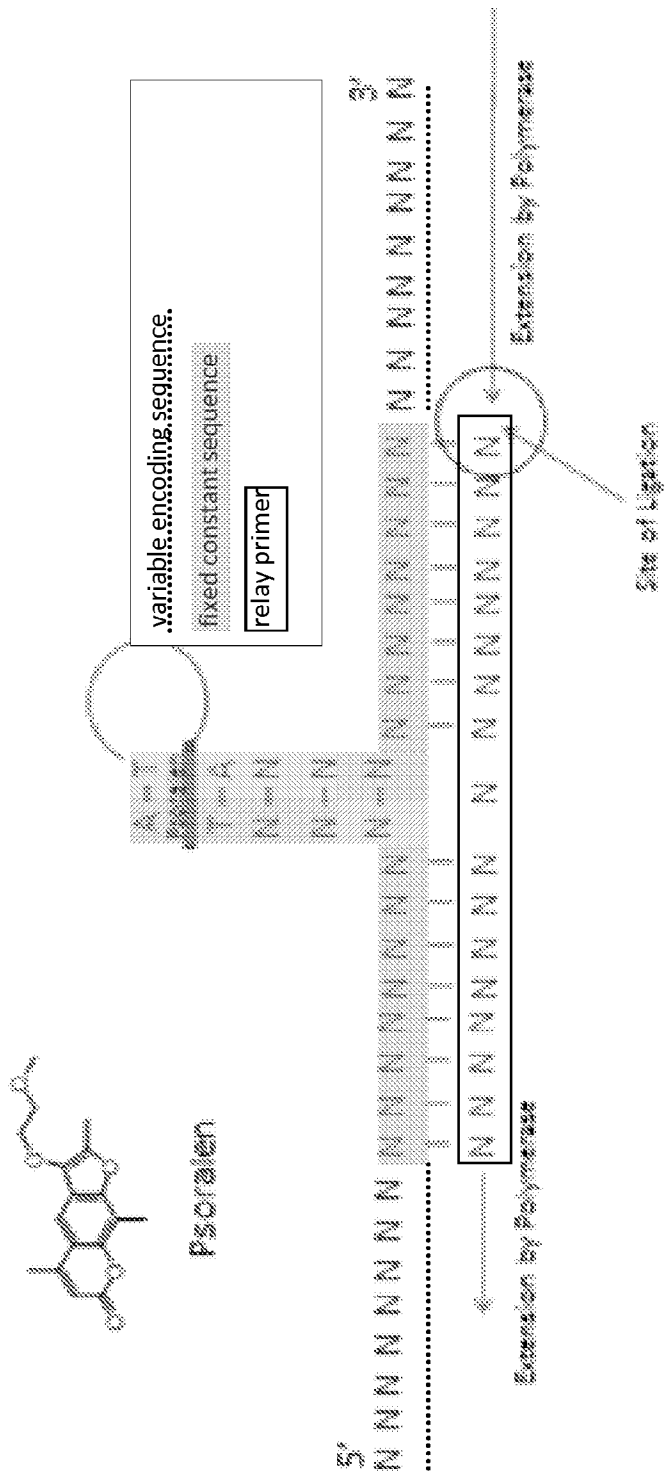


Figure 2

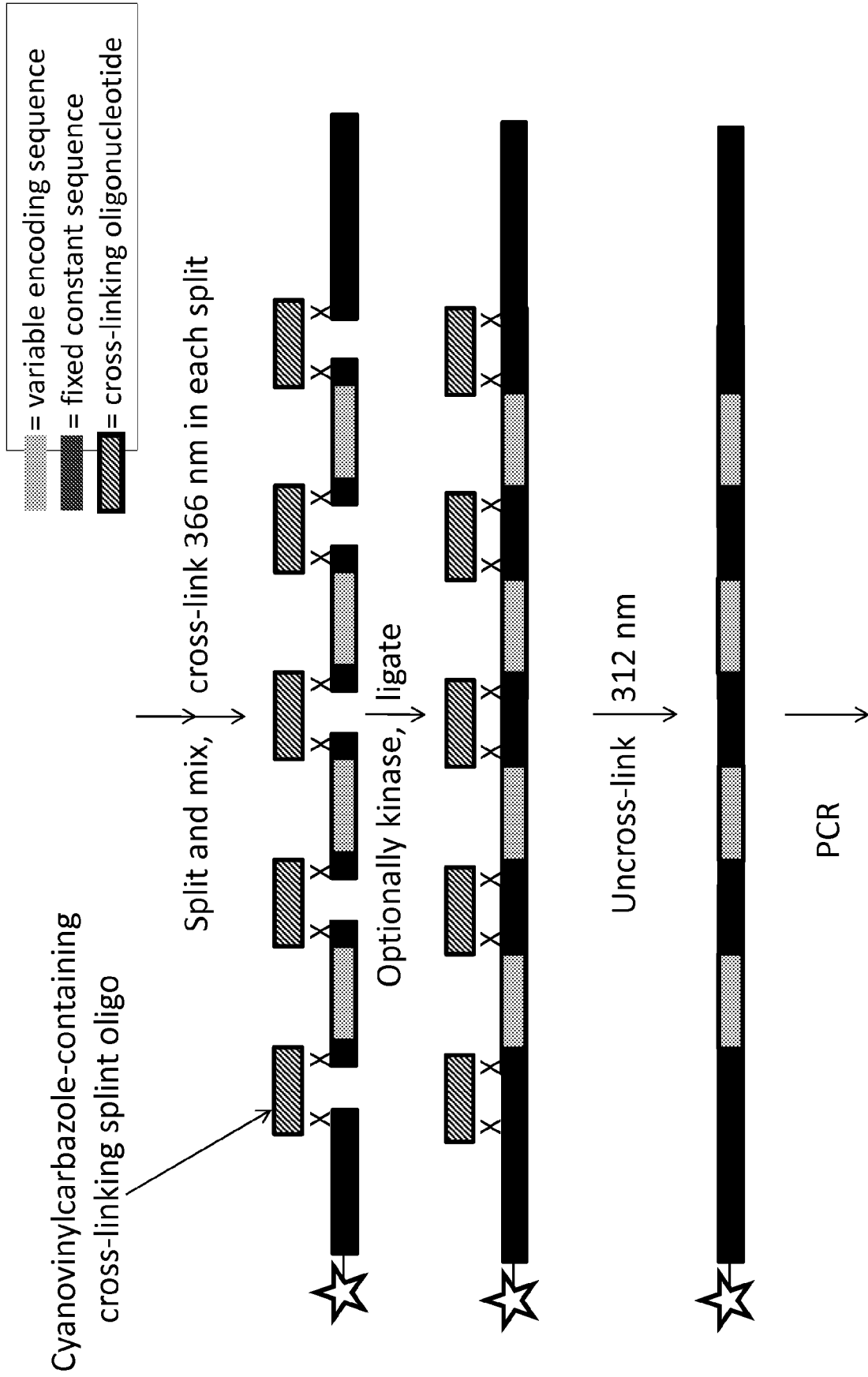


Figure 3

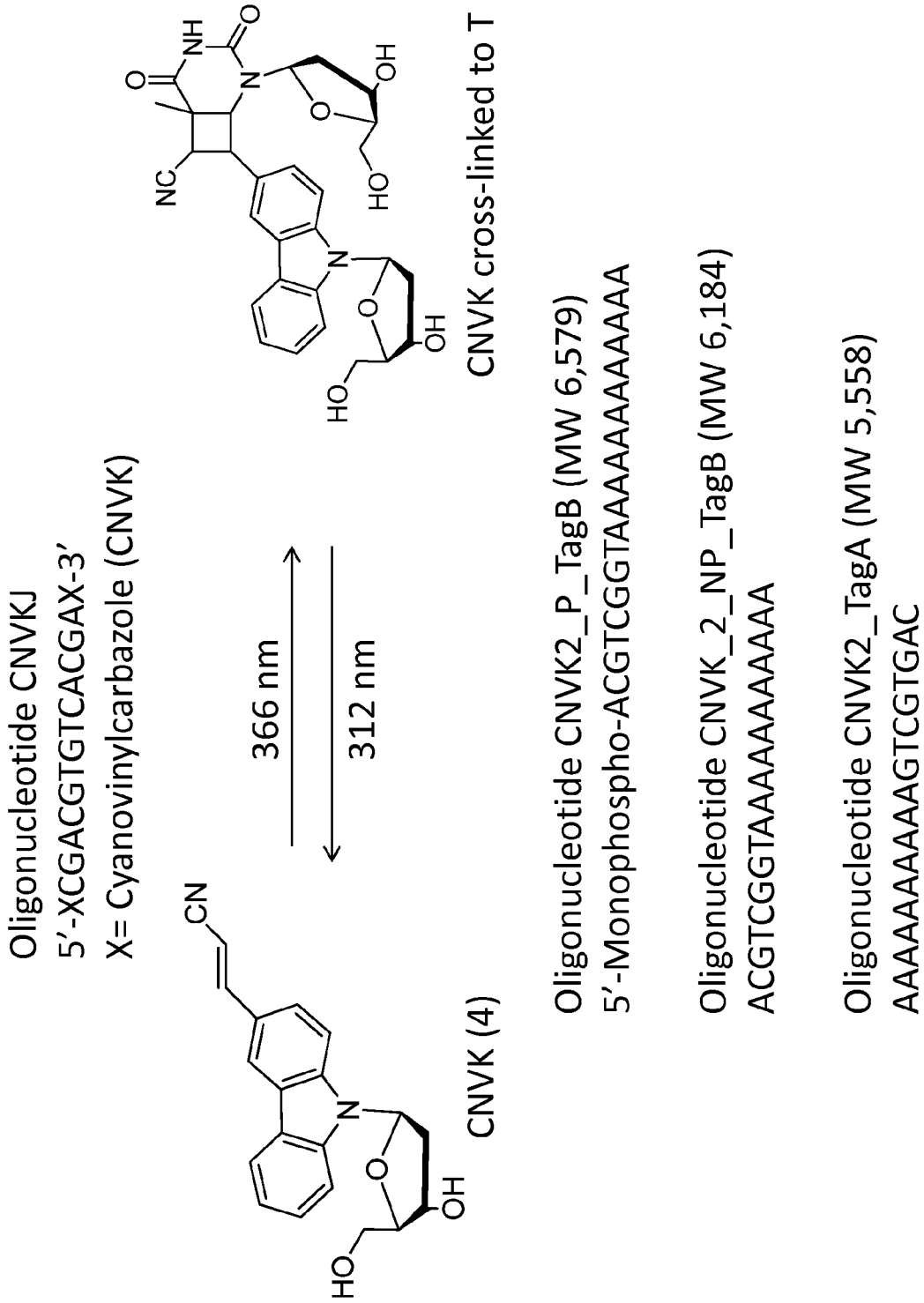


Figure 5

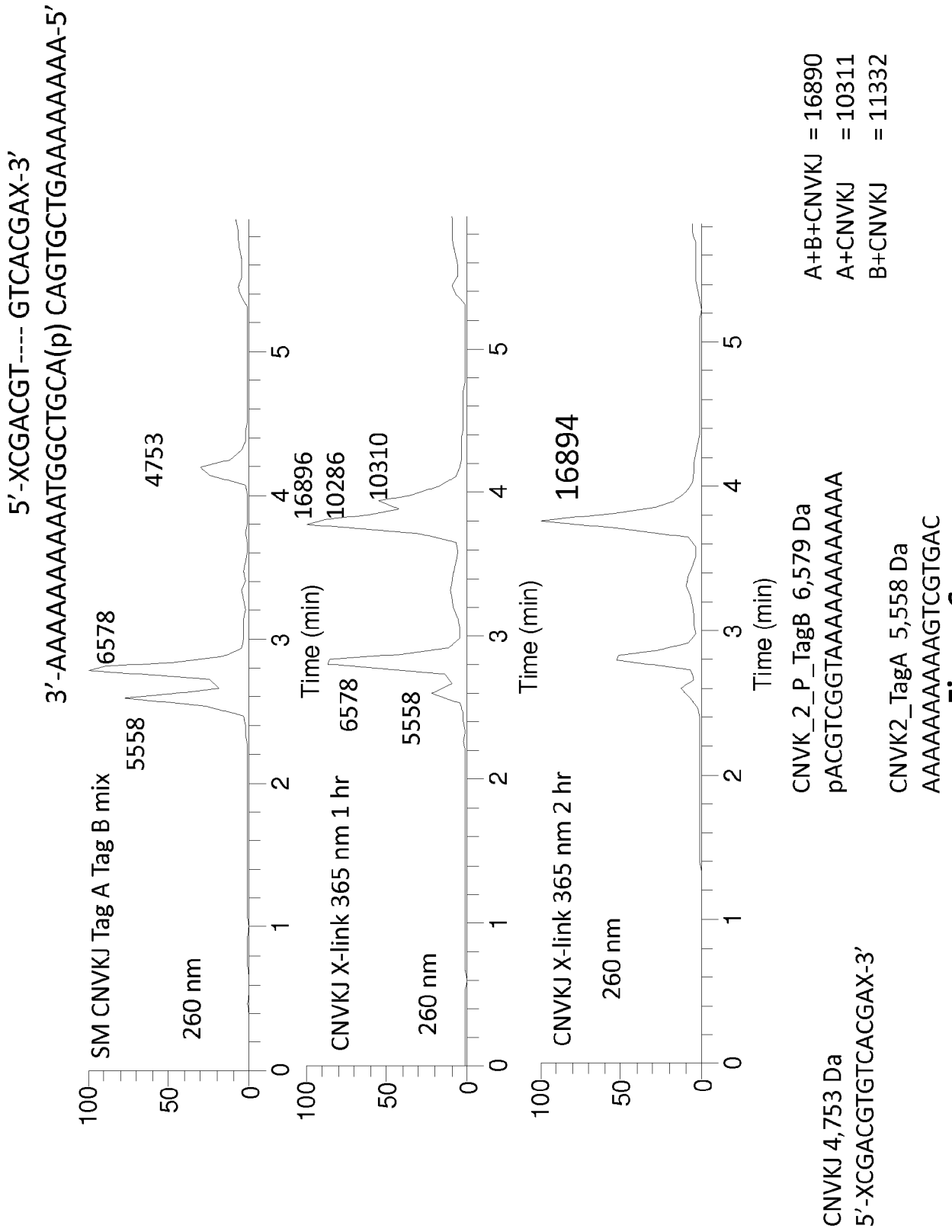
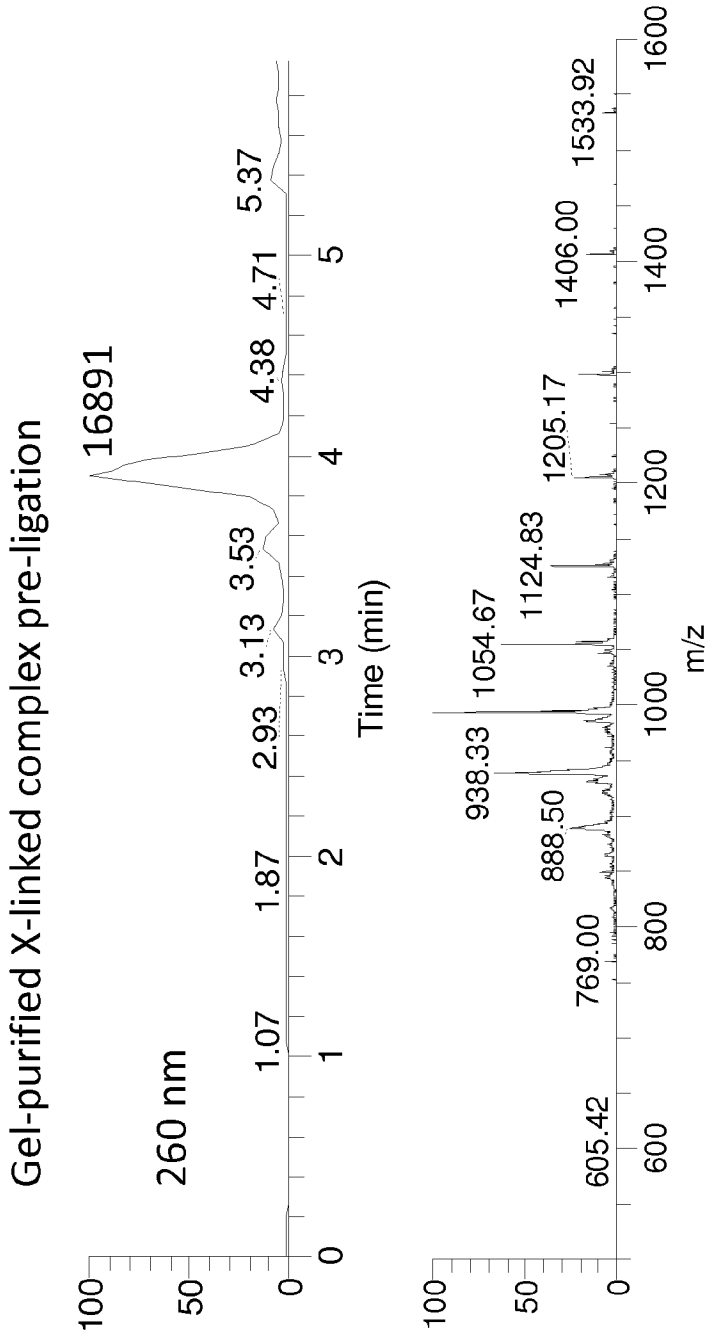


Figure 6



Gel purified X-linked complex pre ligation un-xlinked at 312 nm for 1.5 hr at Room Temp.

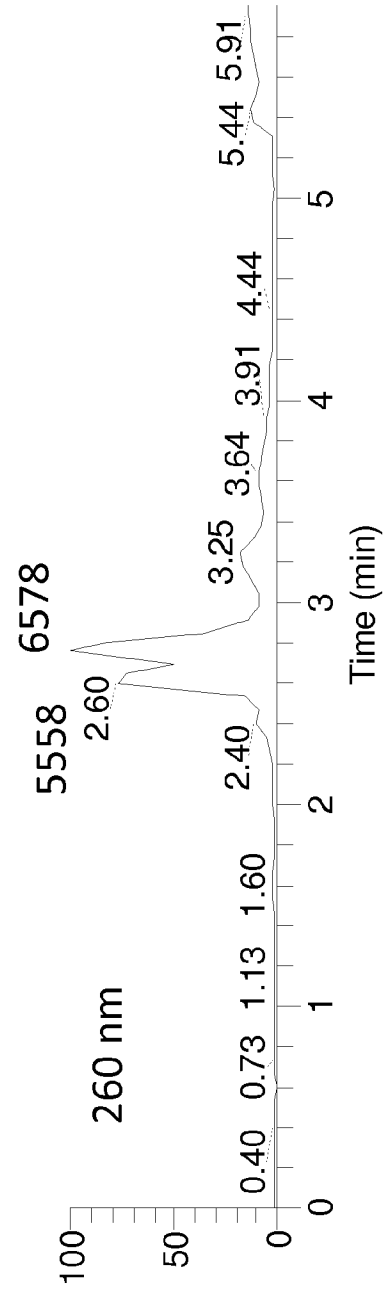


Figure 7

SM:
53mer DNA, crosslinked with 2xCNVK splints (gel-purified) $\xrightarrow{\text{UV 312 nm, 80}^\circ\text{C, 1 hr}}$ 53mer contiguous ssDNA oligonucleotide

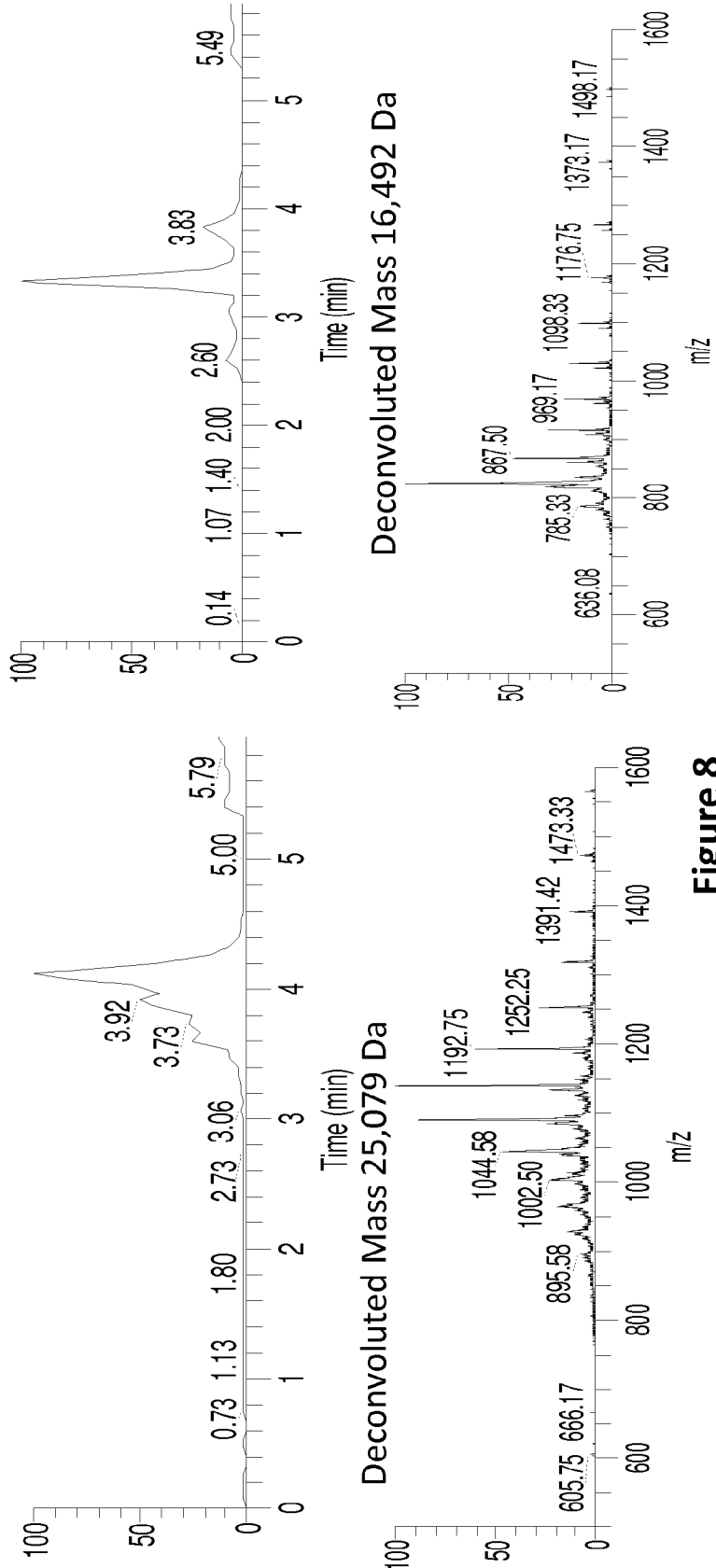


Figure 8

T4 DNA ligation of CNVK X-linked complex followed by un X-link for 1 hour 312 nm at 80°C

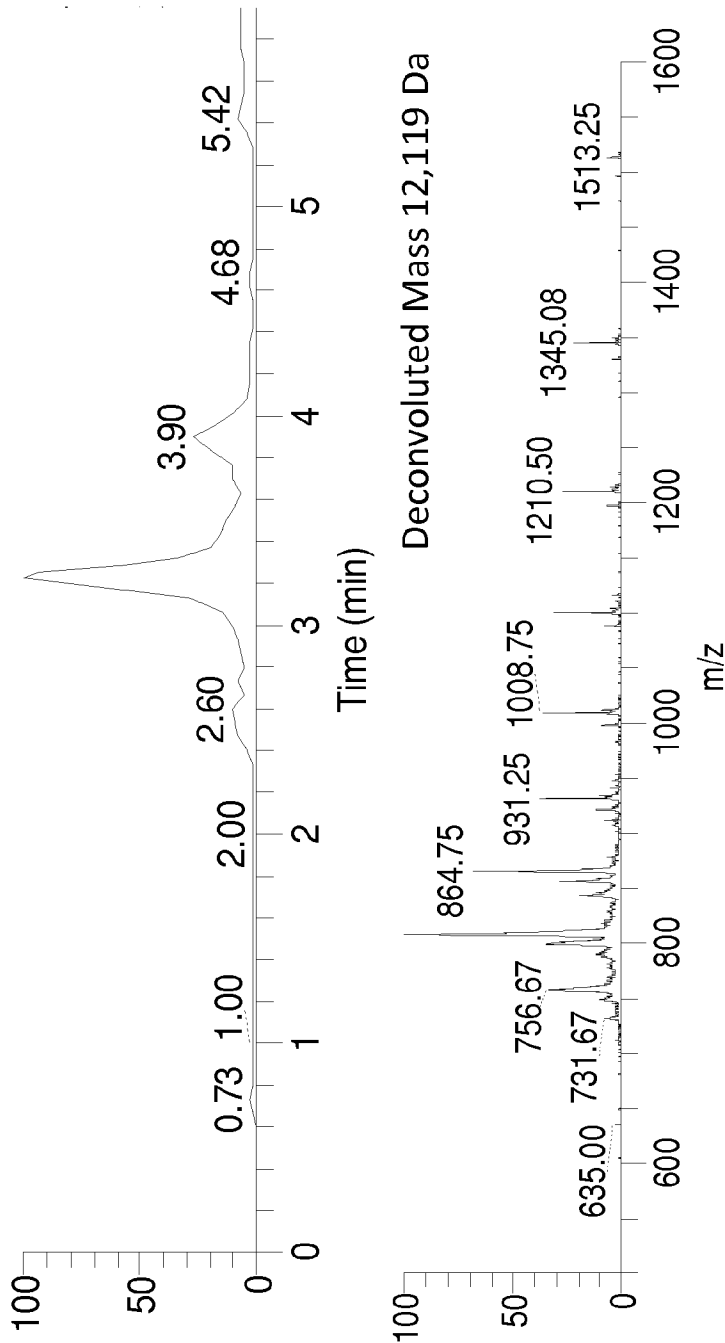


Figure 9

SM:
Tag A + Tag B + splint complex

Product:
Tag A ligated to Tag B

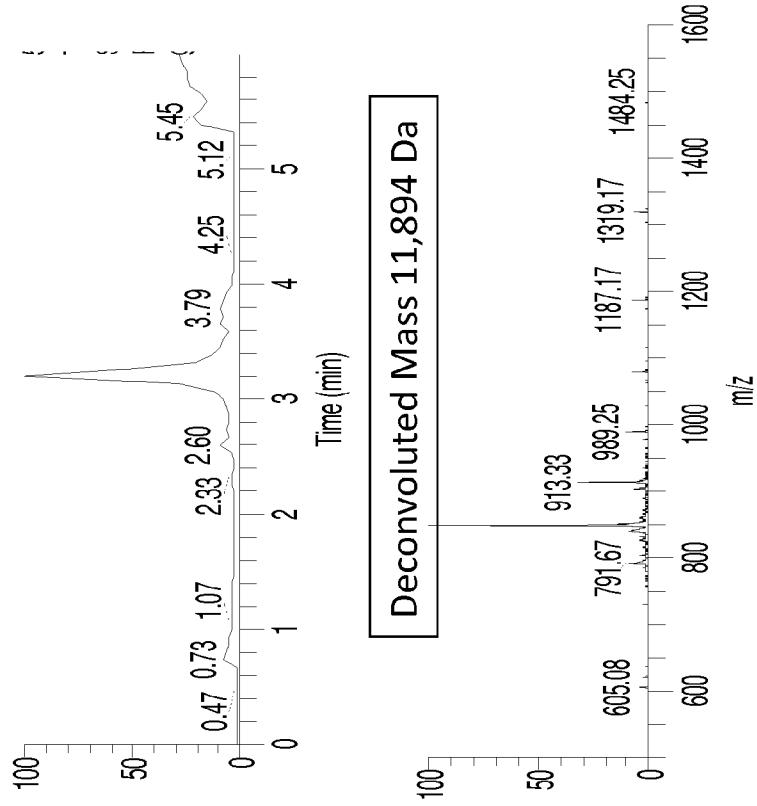
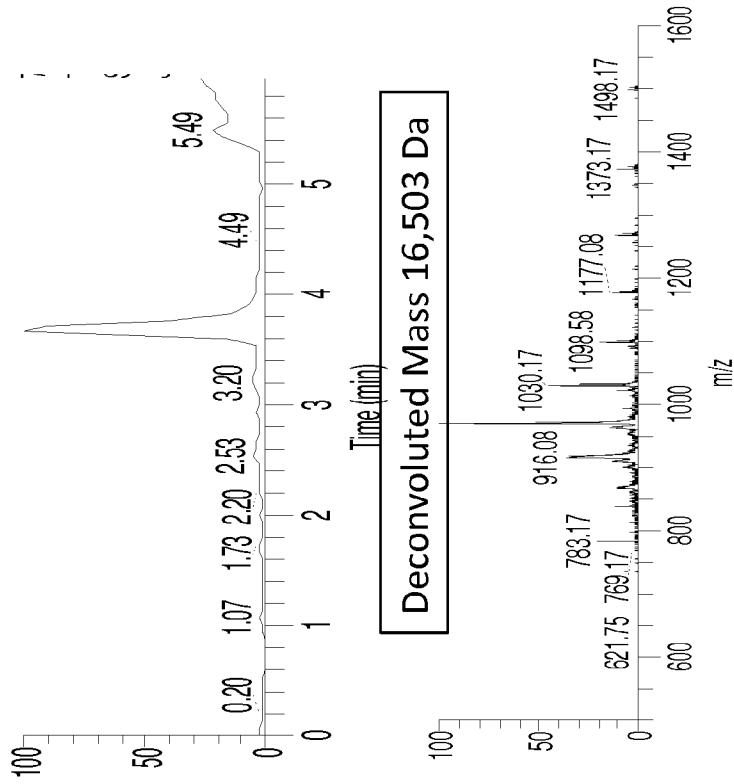


Figure 10

5PSO2_A9_TA (MW 11,556.3 Da)
5'-Psoralen (C2-linked)-TAGCGGATGCAAAAAAAAAAGGCGAGCT TGGTACTG-3'

PSO_HP_A9_TCT (MW 6,087.1 Da)
5'-AAAAAA AAAGCATCCGCTCT -3'

Figure 11

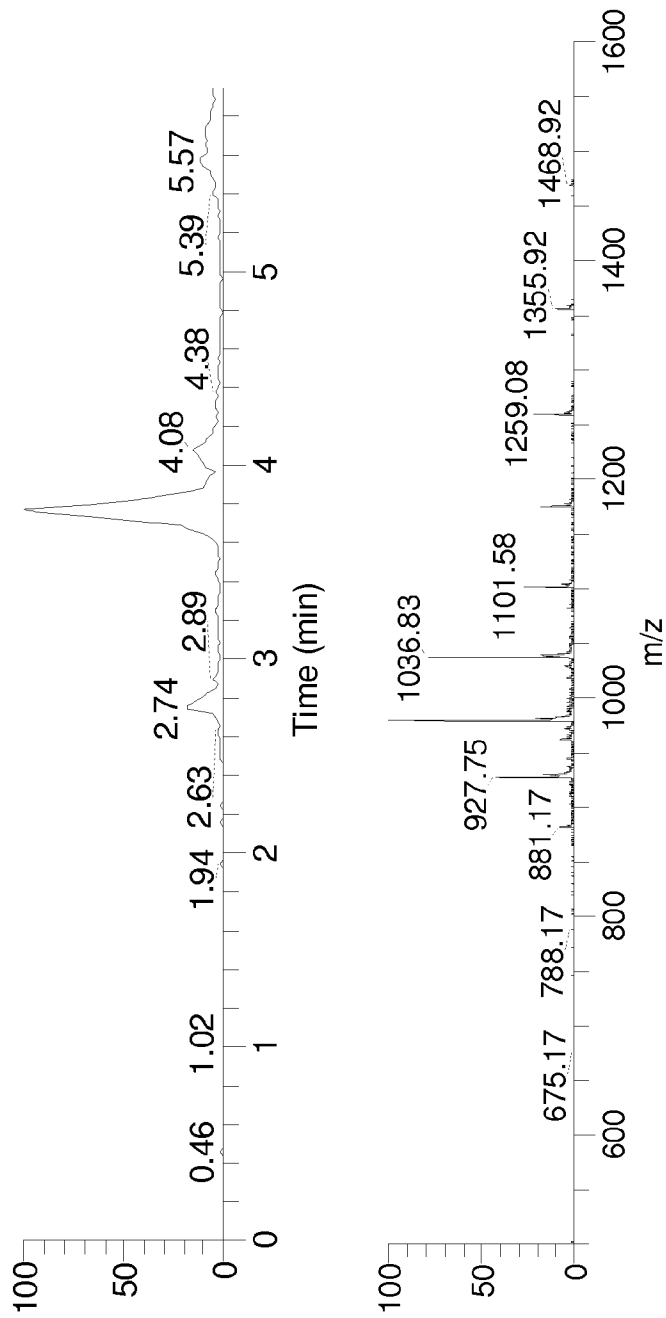
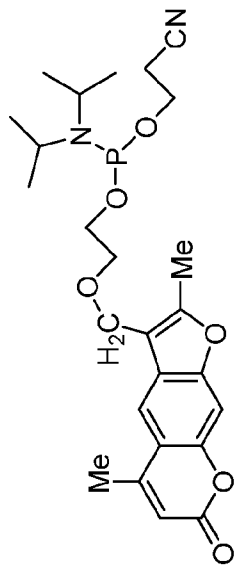


Figure 12



Psoralen (C2) Phosphoramidite

Tag1_PsoCVU (MW 3,662.5 Da)

5'-AAAAACTCAGGT-3'

SplintC_PsoC2 (MW 4,529.0 Da)

5'-TCCGCTCGACCTGAG-3'

5PsoC2_A9_GA (MW 11,580.6 Da)

5'-Psoralen (C2)-GAGCGGATGCCAAAAAAGGCGAGCTTGCGTACTG-3'

Figure 13

5'-NNNNNNNNNT PNNNNNNNNNNNNNNNNNNNNNNNN-3'
3'-NNNAGNNNNNNNN-5'

P = 5'-Psoralen

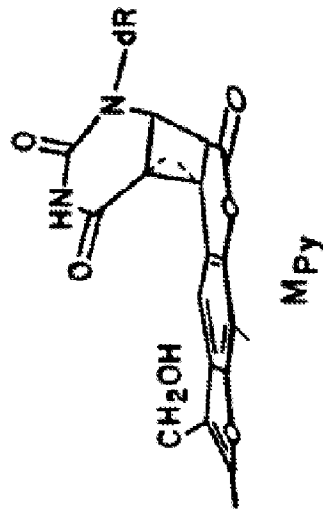


Figure 14

Time-course of Psoralen photoligation

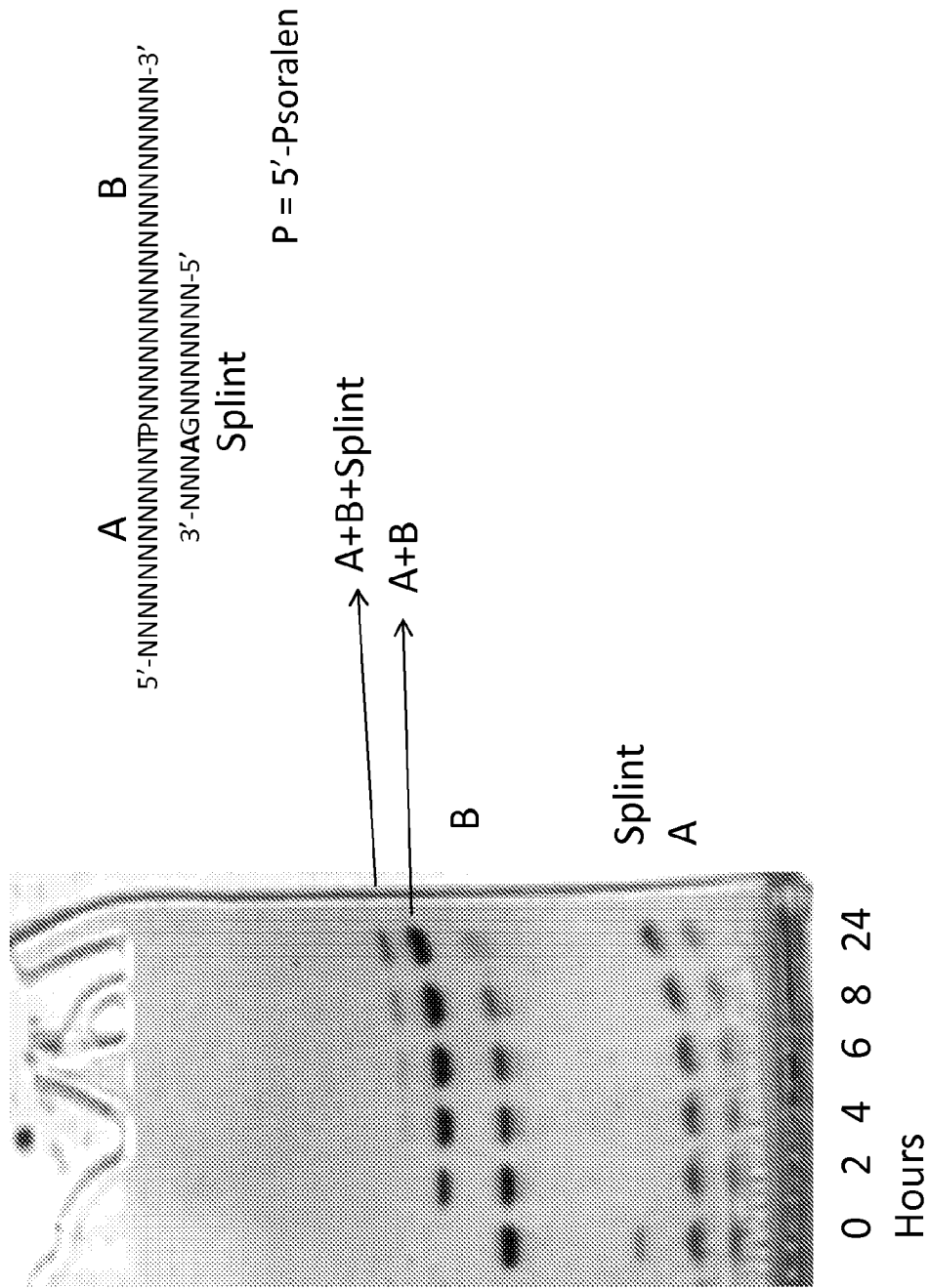


Figure 15

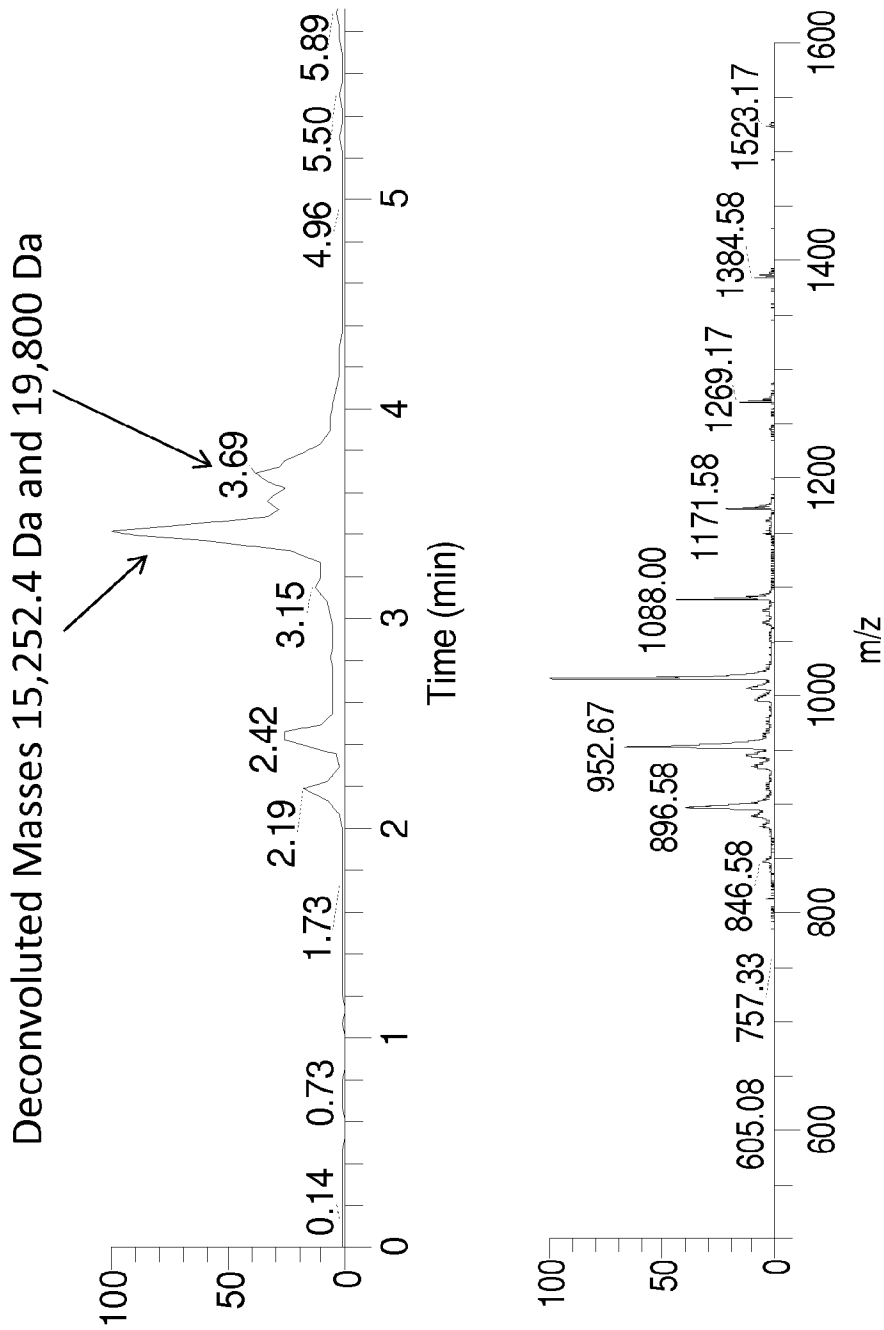


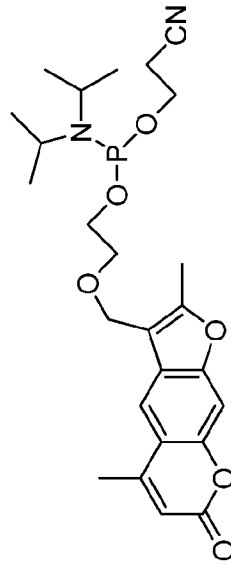
Figure 16

5Bio_Tag_PsoCVU (MW 4,232.1 Da) /5BioTEG/ is a proprietary
 5'-/5BioTEG/AAAAACTCAGGT-3' biotin modification

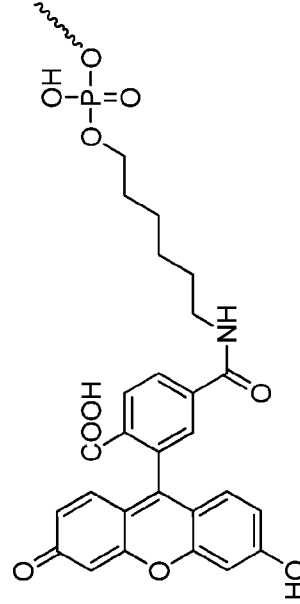
5PsoC2_A9_GA (MW 11,580.6 Da)
 5'-Psoralen (C2)-GAGCGGATGC AAAAAGCGAGCTTGCGTACTG-3'

SplintC_PsoC2 (MW 4,529.0 Da) FAMprimer (MW 5,099.5 Da)
 5'-TCCGCTCGACCTGAG-3' 5'-/56-FAM/AGTACGCAAGCTCGC-3'

Phos-SplintC_PsoC2
 5'-Monophosphate-TCCGCTCGACCTGAG-3'



Psoralen (C2) Phosphoramidite



/56-FAM/ Phosphoramidite

Figure 17

LCMS of gel-purified photochemical conjugate Bio_Pso

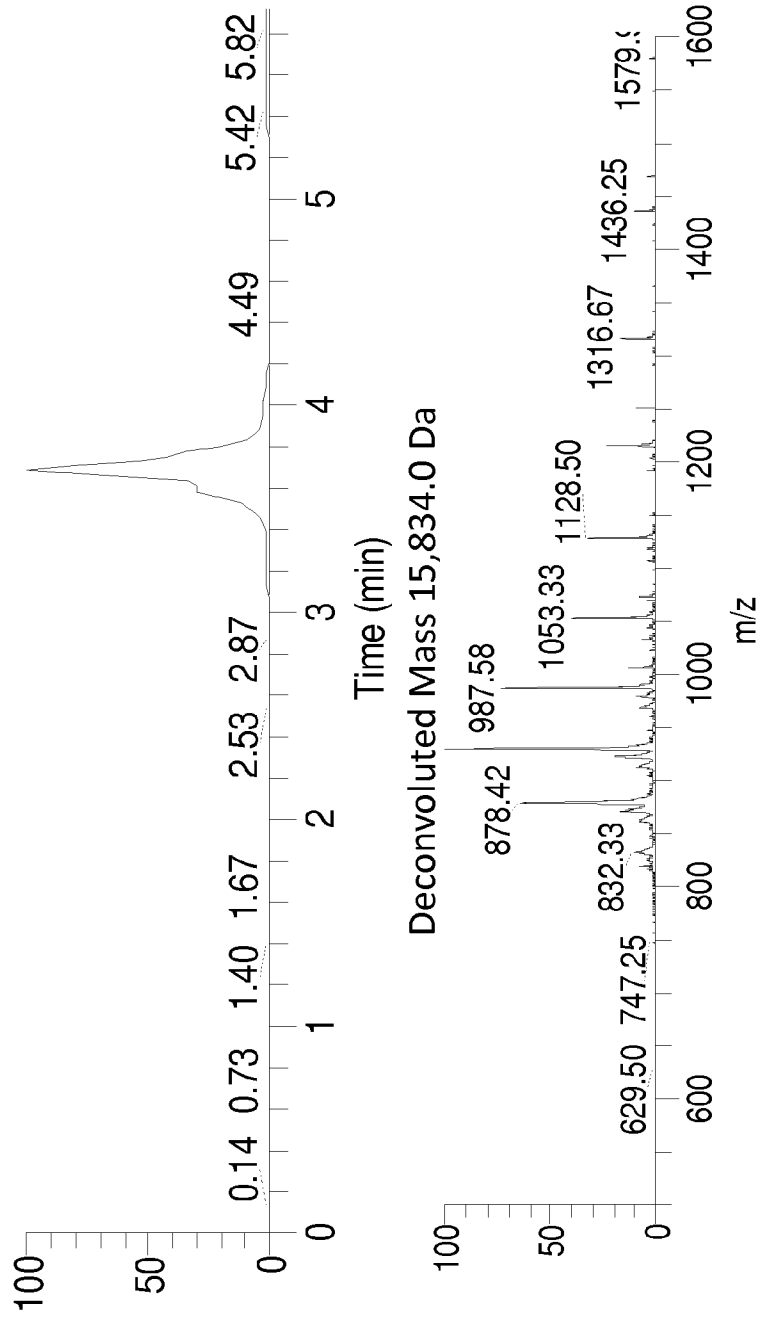


Figure 18

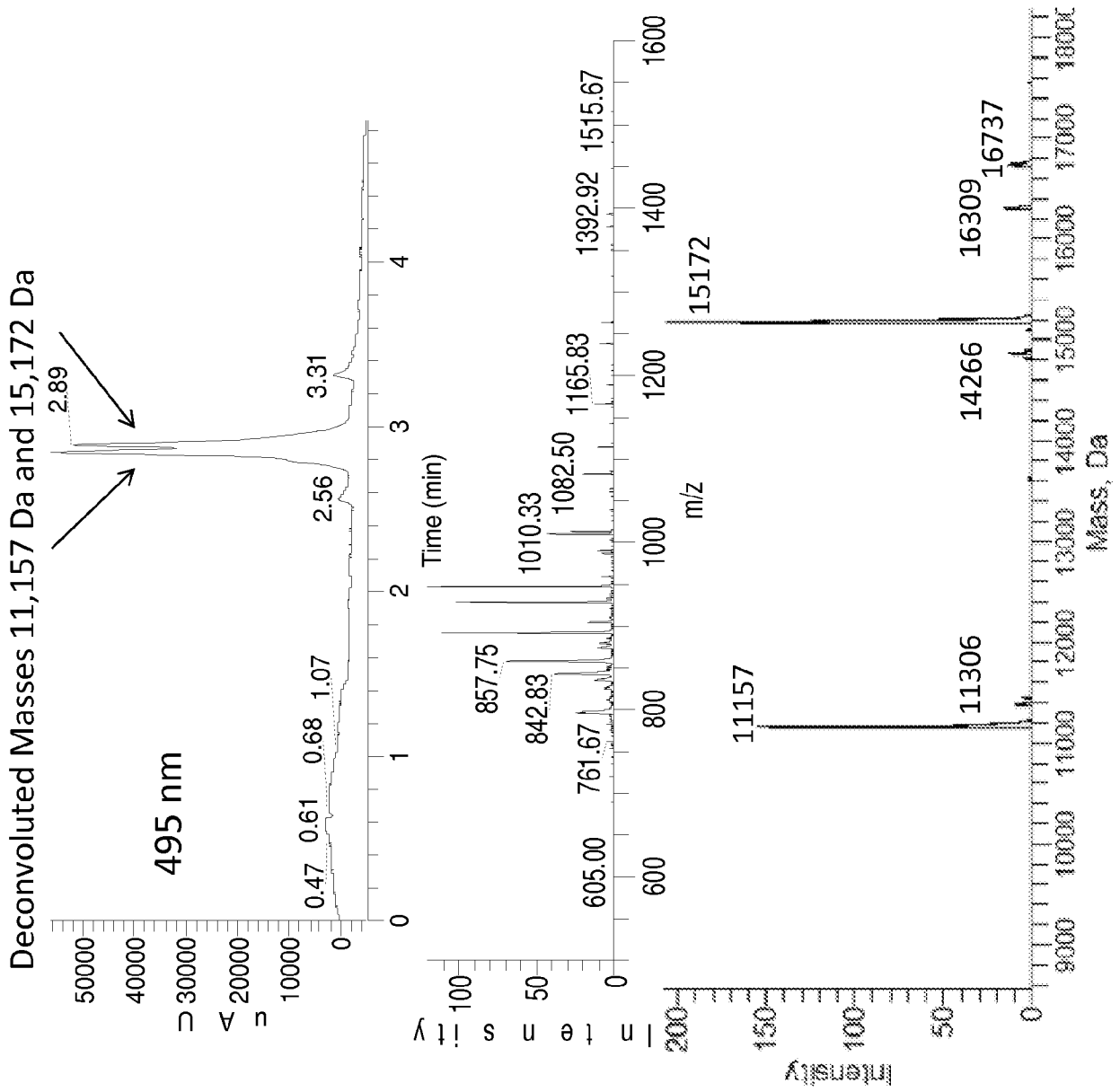


Figure 19

LCMS of no-ligation control of cDNA generated from Bio_Pso using terminal and non-terminal primer extension

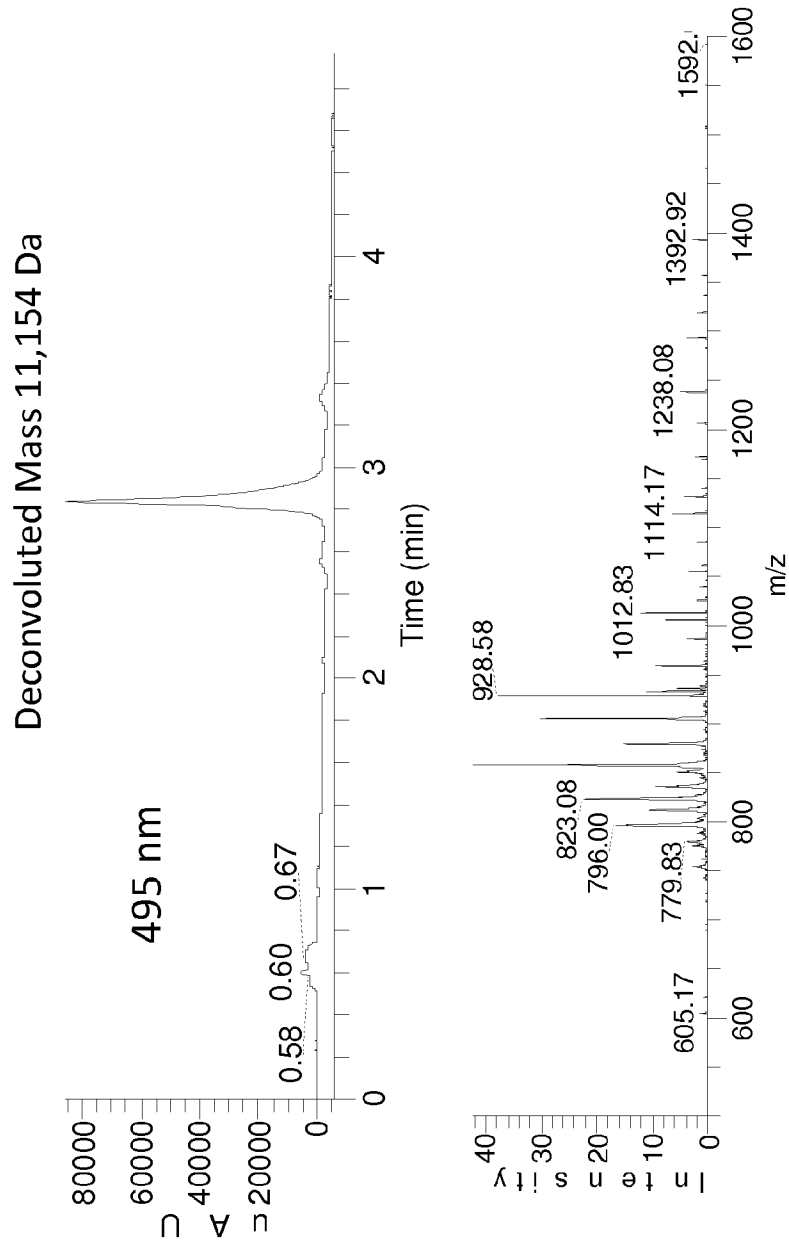
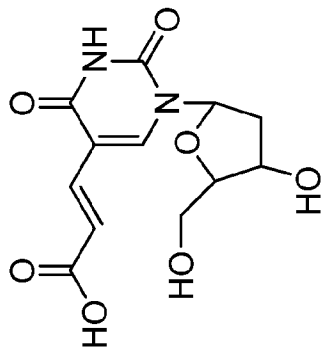


Figure 20



5-(Carboxy)vinyl-2'-deoxyuridine

SplintC_CVU
5'-TCCGCTCAACCTGAG-3'

SplintA_CVU
5'-TCCGCTTAACCTGAG-3'

Tag1_PsoCVU (MW 3,662.5 Da)
5'-AAAAACTCAGGT-3'

CVU_G (MW 11,576.2 Da)
5'- (5-(Carboxy)vinyl-2'-deoxyuridine)GAGCGGATGCAAAAAAGGCGAGCTTGCCTACTG -3'

CVU_A (MW 11560.2 Da)
5'- (5-(Carboxy)vinyl-2'-deoxyuridine)AAGCGGATGCAAAAAAGGCGAGCTTGCCTACTG -3'

Figure 21

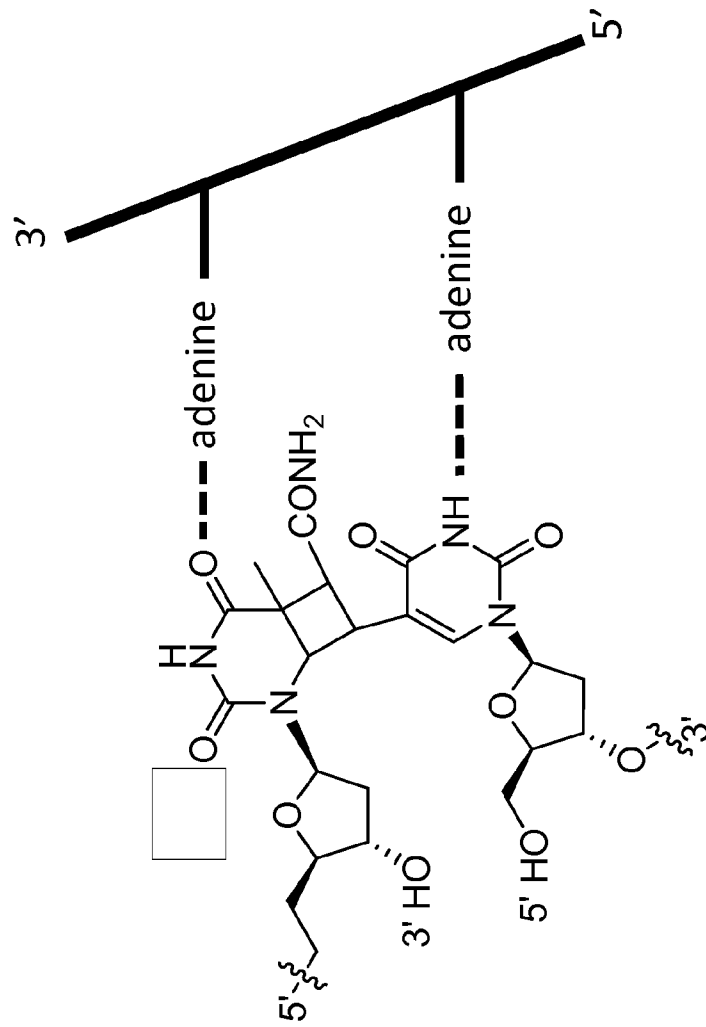


Figure 22

Time-course of CVU photoligation at 365 nm

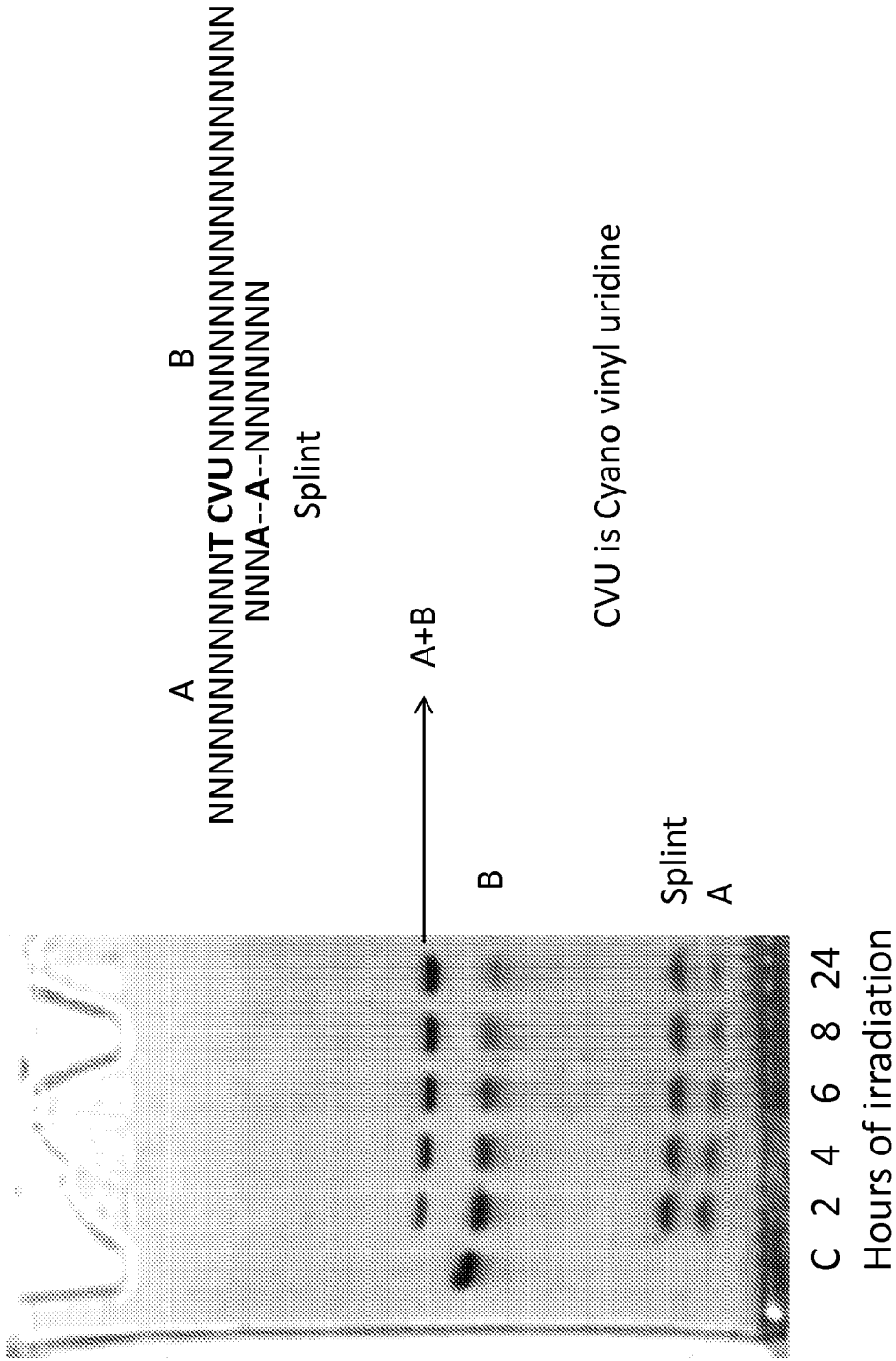


Figure 23

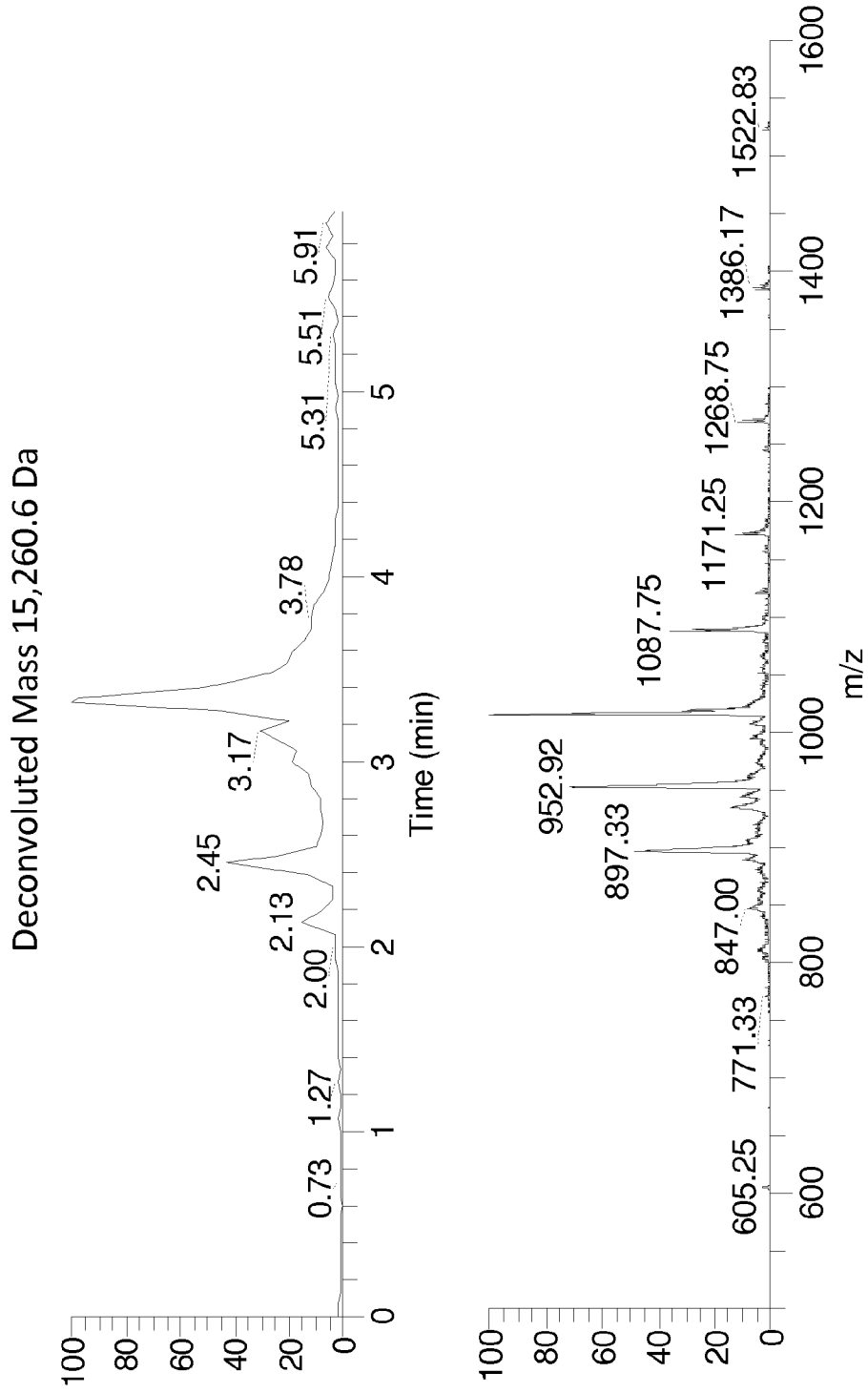
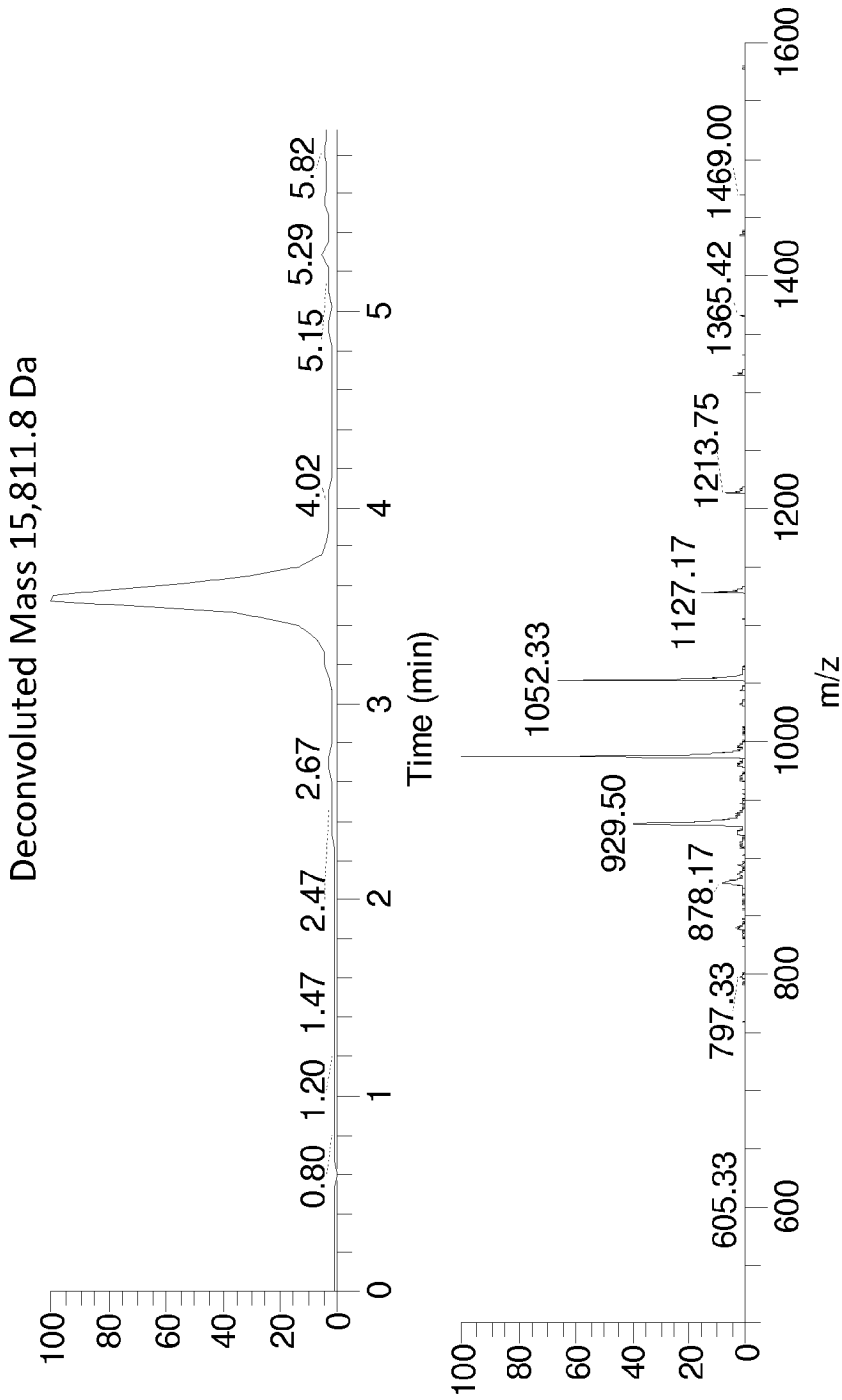


Figure 24

Irradiation for 20 hrs followed by gel-purification



5Bio_Tag_PsoCVU (MW 4,232.1 Da)
/5BioTEG/AAAAA CTCAGGT

Figure 25

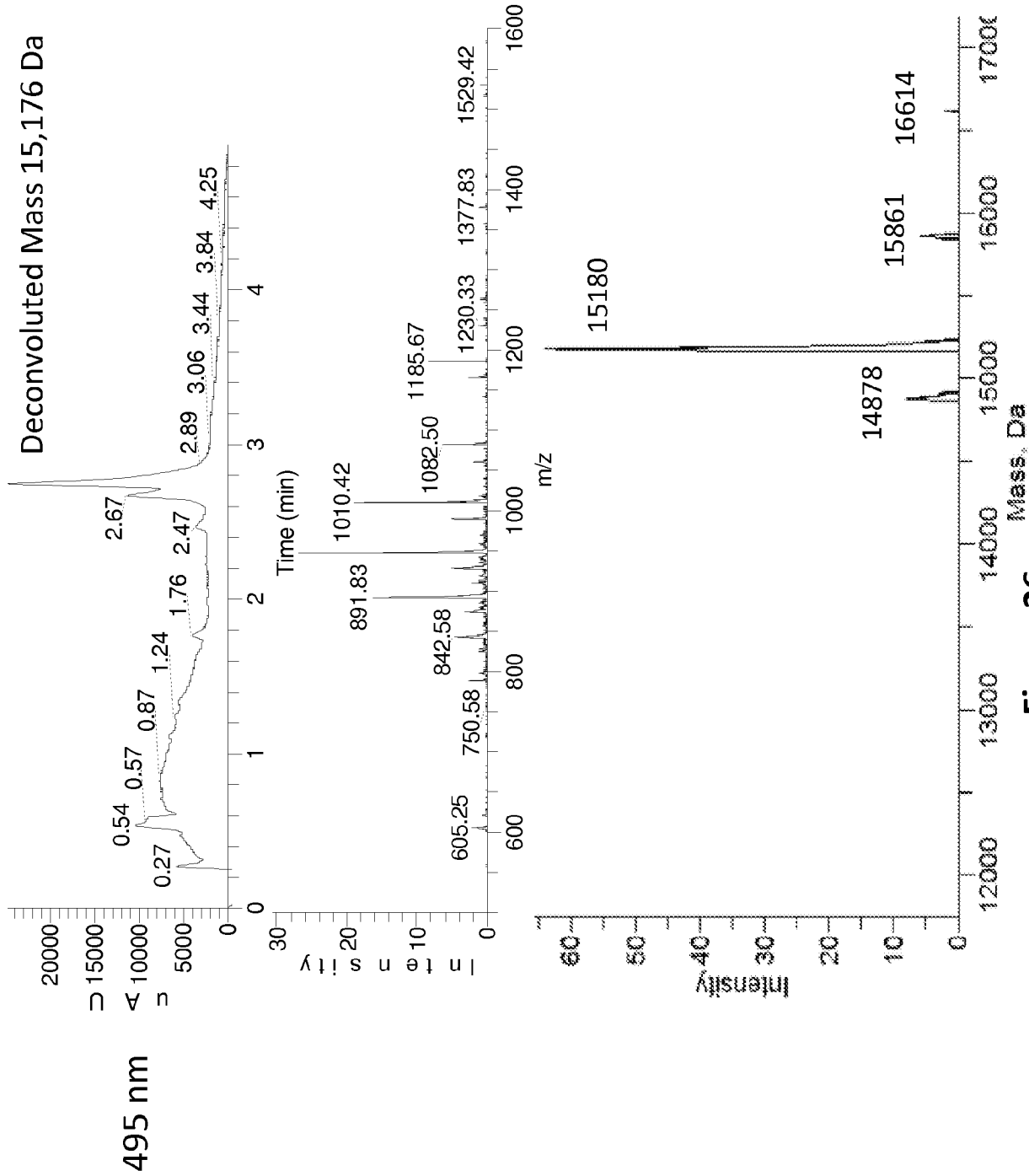
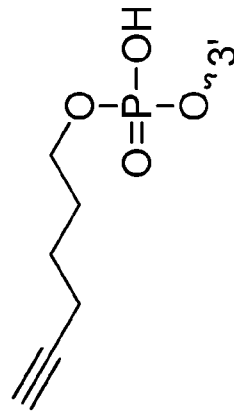


Figure 26

S_IDTN₃
TGCGGTCTAACTGTCTAG TTCACCTTCTCCGGAATGAACAGG/3AzideN/

S_IDTalkyne
/5Hexynyl/TACCGAATTGCCGCCTGATTGAAGCAATCGGGTGCTATGCTT



5Hexynyl Modification

Figure 27

LC/MS analysis of the gel-purified conjugate of S_IDTN₃ and S_IDTalkyne.

Deconvoluted observed mass 26,304.8 Da (calculated 26,311.2 Da)

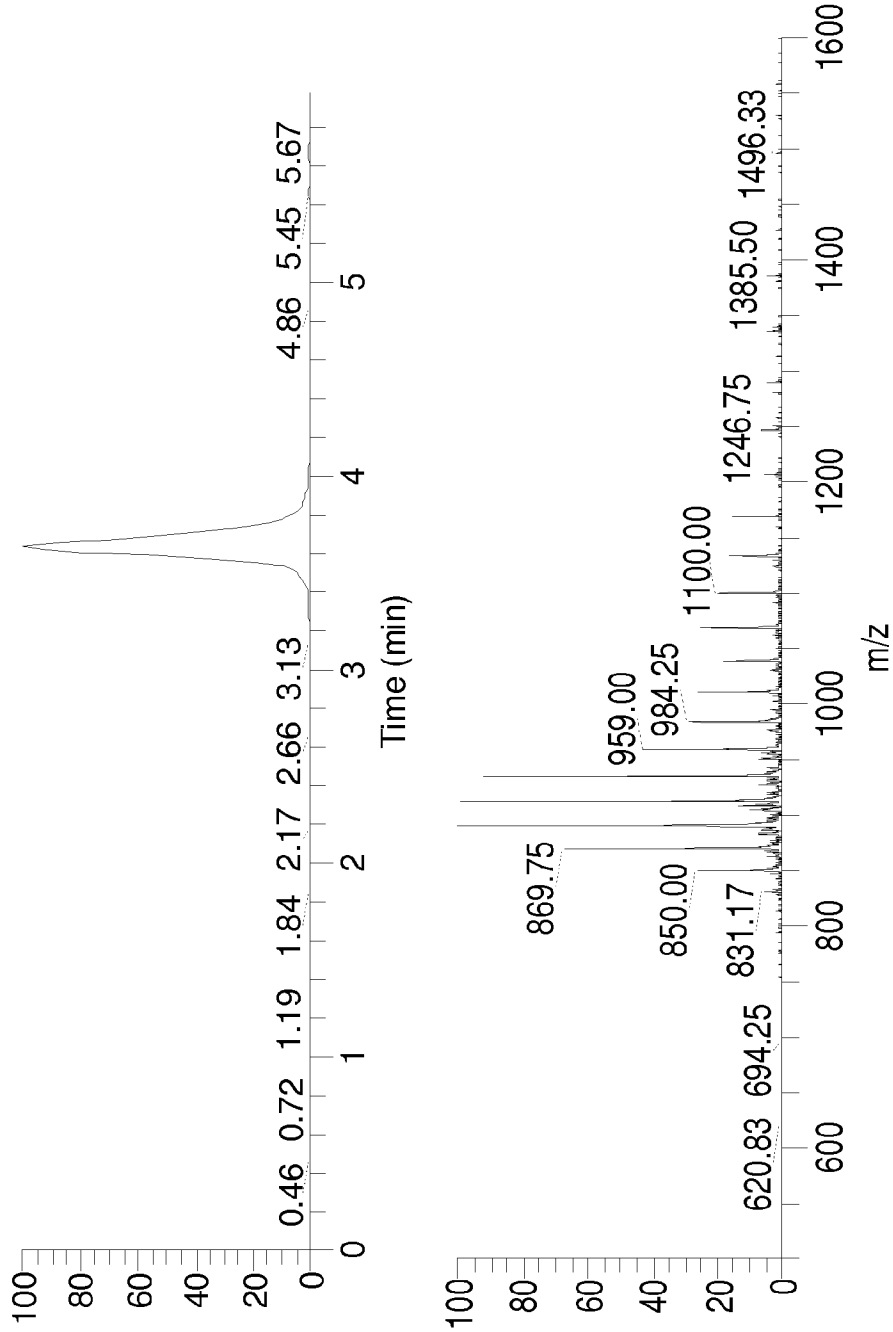
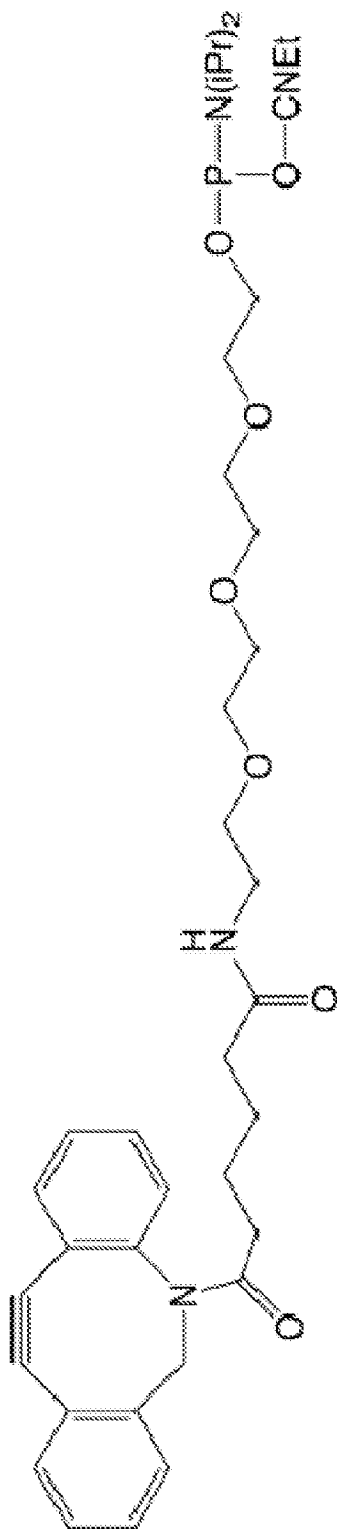


Figure 28

TKR_Central: /5Hexynyl/GGAATGAACAGG **GTAAGCTGGAGTGAAGCGTTATG/3AzideN/**

S_IDTN₃: TGCGGTACTAGTCTAGTTCACCTTCTCCGGAATGAACAGG /3AzideN/

TKR_DBCO_S: /DBCO/GAAGGCGTTATGTCGGTACTCTTGCAATCGGGTGCTATGCTT



DBCO Phosphoramidite

Figure 29

Observed Deconvoluted Mass 25,336 Da (calculated 25,341.8 Da)

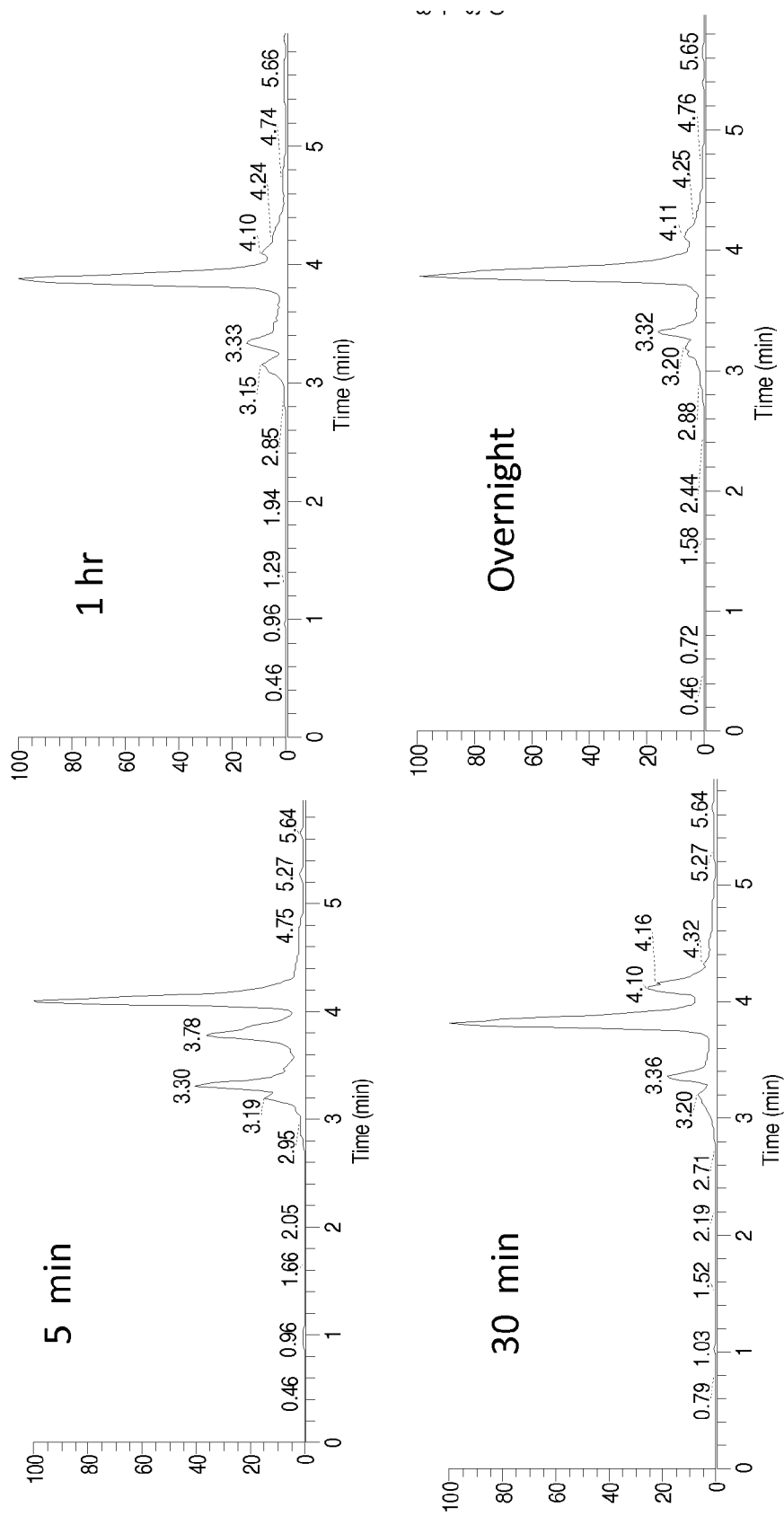


Figure 30

Observed Deconvoluted Mass 38,578.1 Da (calculated 38,560.4 Da)

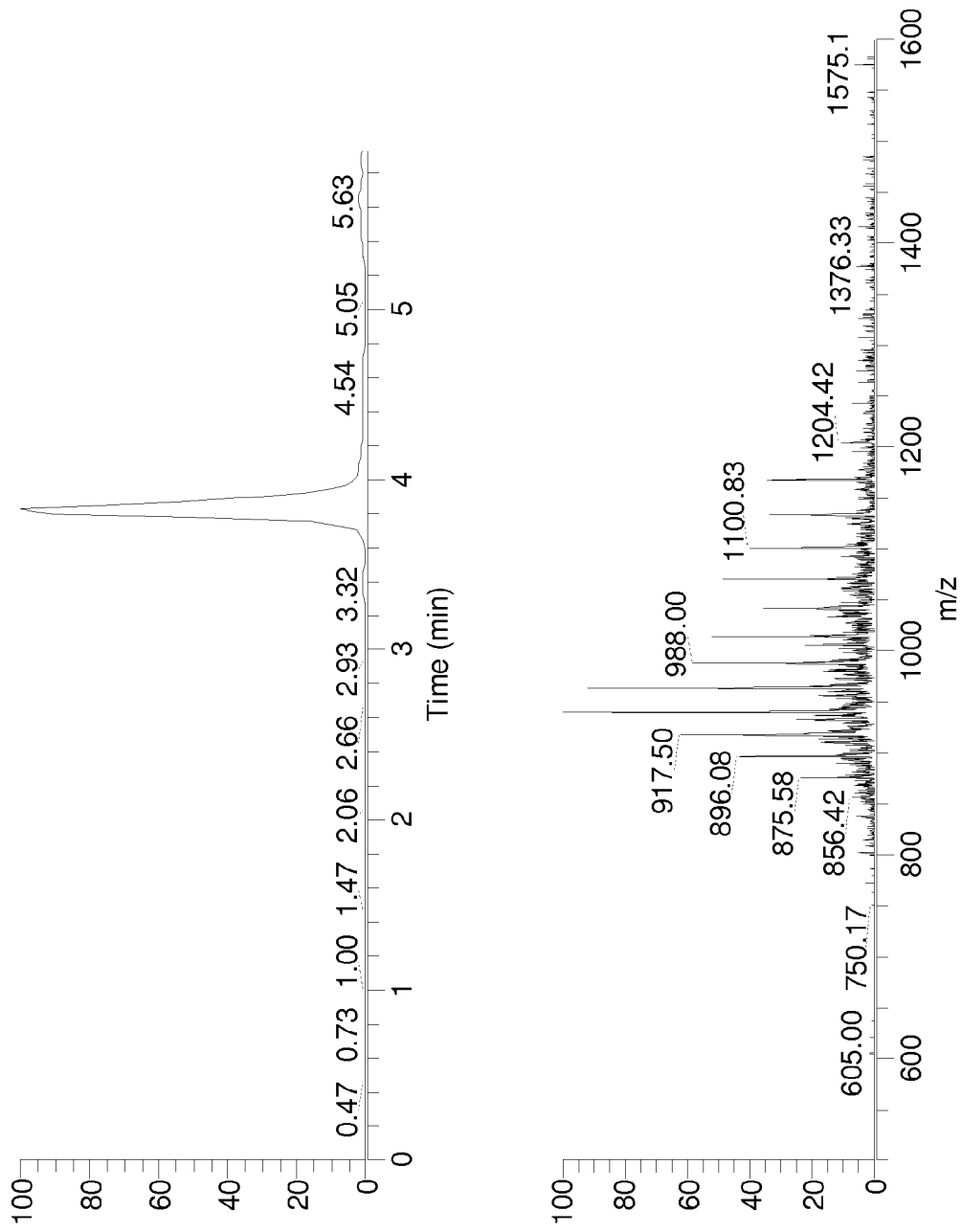
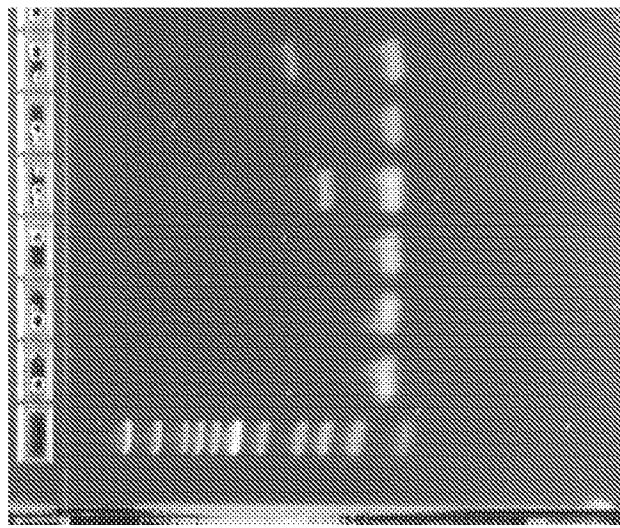


Figure 31

Conjugate_Click_S
TGCGGTCTAACTGTCTAGTTACCTTCTCCGGAATGAACAGG(click)TACCGAATTGCCGCCTGATTGAAG
-CAATCGGGTGCTATGCTT

Conjugate_Click_L
TGCGGTCTAACTGTCTAGGCACCTTGTTCGTTGCCAGTGTGAGGAATGAACAGG(click)TACCGAATTGC
CTTCCTCG-TACAGTTCTAAGGCGCTTGGACACCACTTCAATCGGGTGCTATGCTT

Figure 32



Lane 1 2 3 4 5 6 7

Figure 33

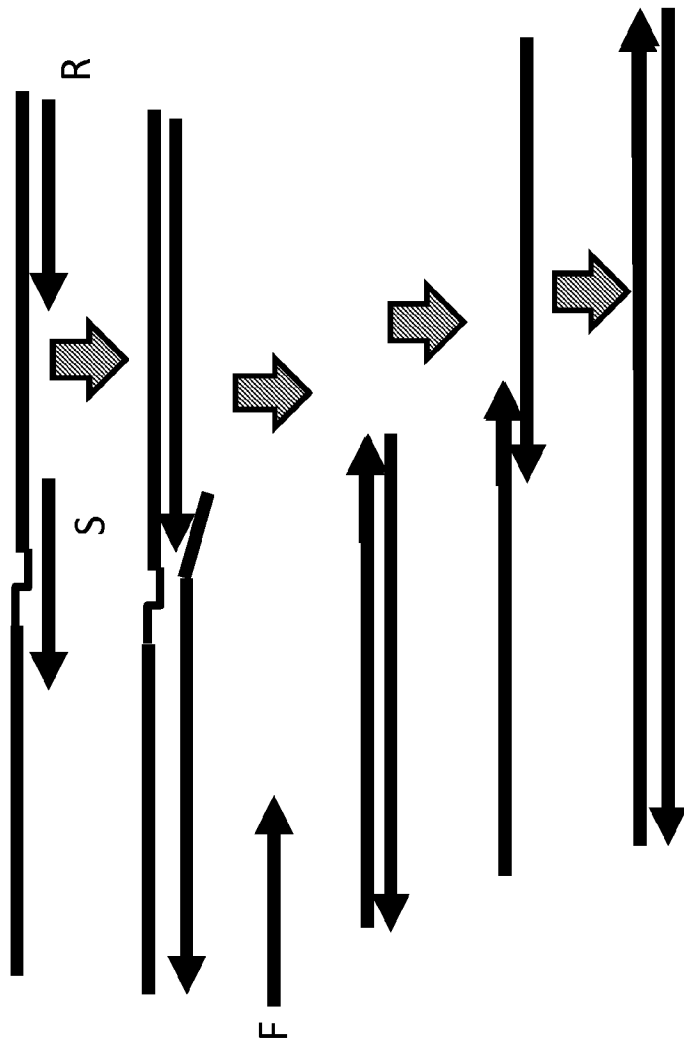
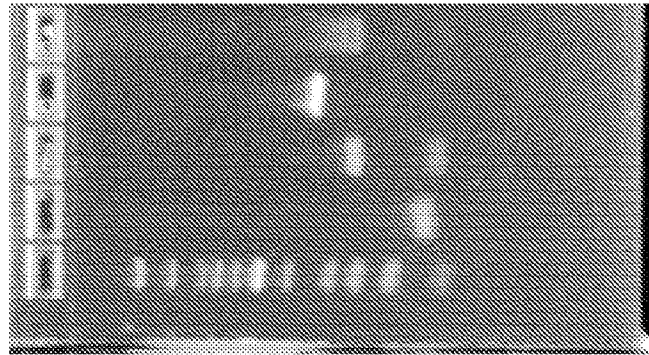


Figure 34



Lane 1 2 3 4 5

Figure 35

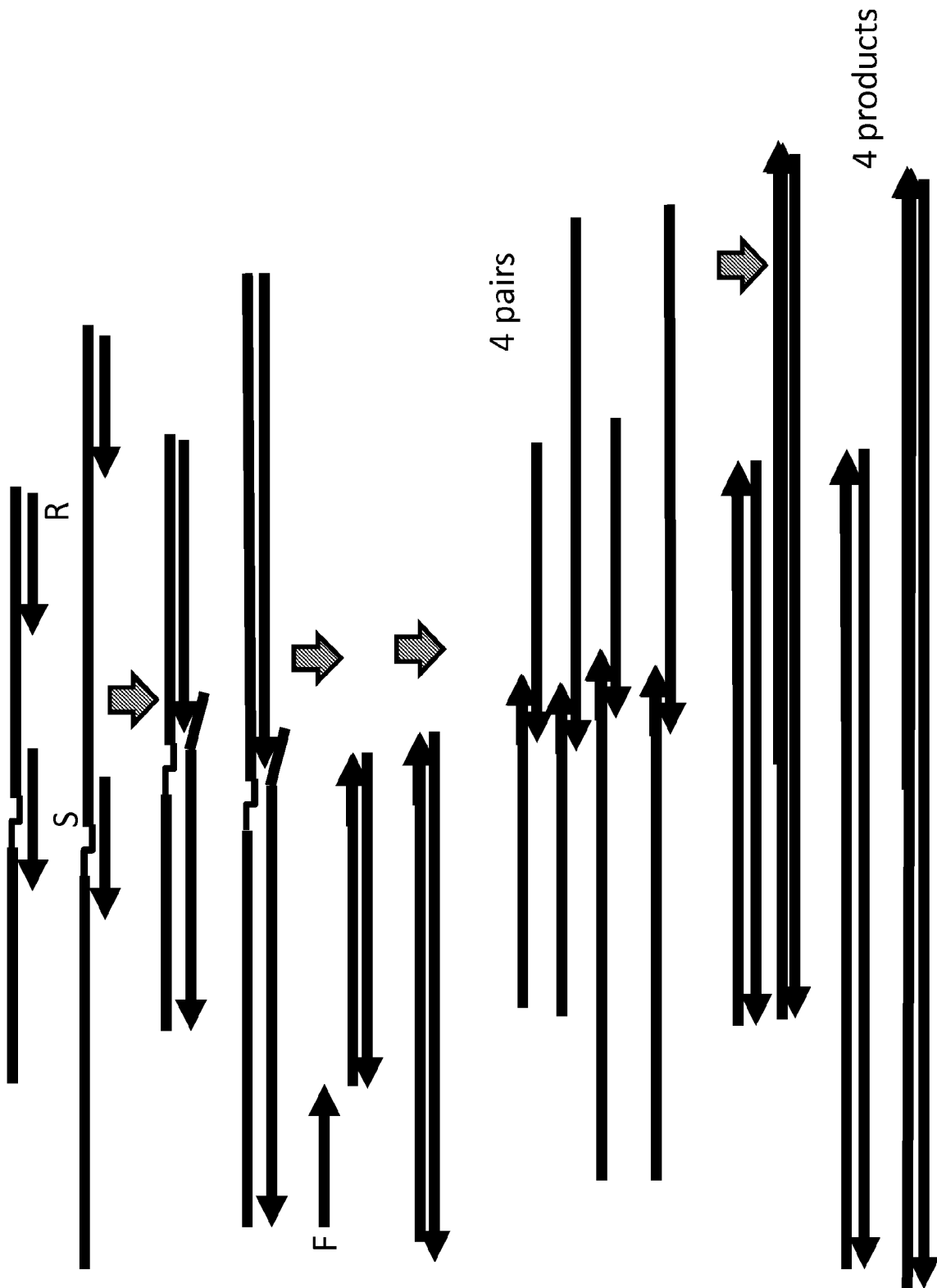


Figure 36

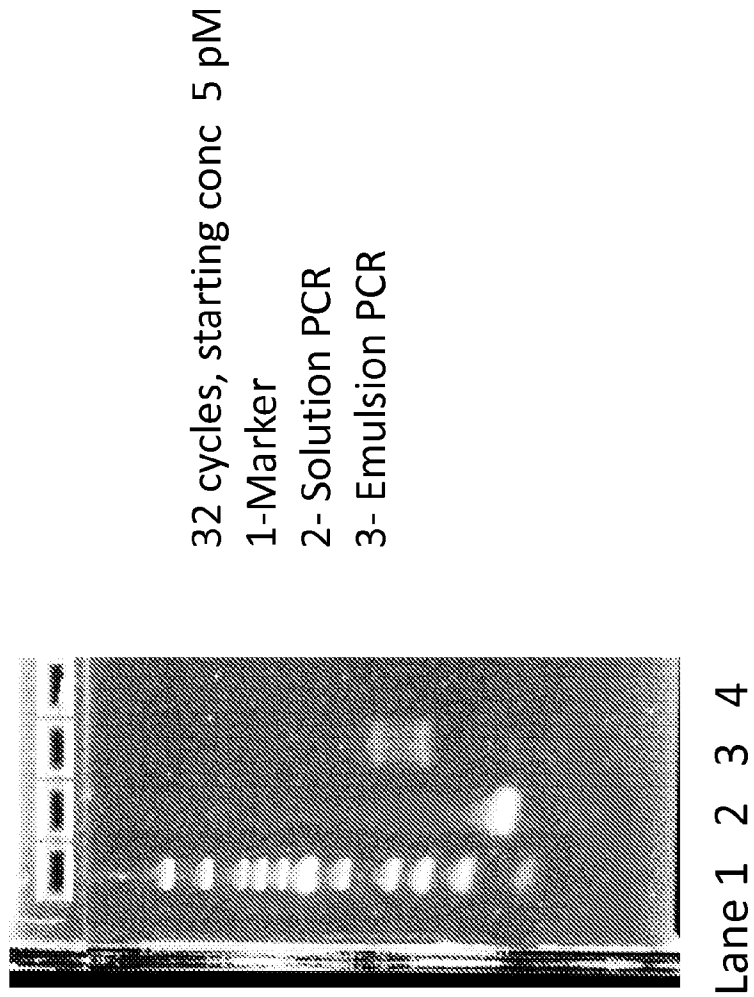


Figure 37

TKR_S:
TGCGGTCTAACTGTCTAGTTCACCTTCTCCGGAAATGAACAGG(click)GGAATGAACAGGGCCCTGATTGAAGCAATCGGGTGCTATGCTT

TKR_L:
TGCGGTCTAACTGTCTAGGACACTTGTTCGTTTGCCAGTGTGATGGACACCCTTGGAAATGAACAGG(click)GGAATGAACAGGGGAATGAACAG
GTTCCCTGACAGTTCTAAGGCGCTCAATCGGGTGCTATGCTT

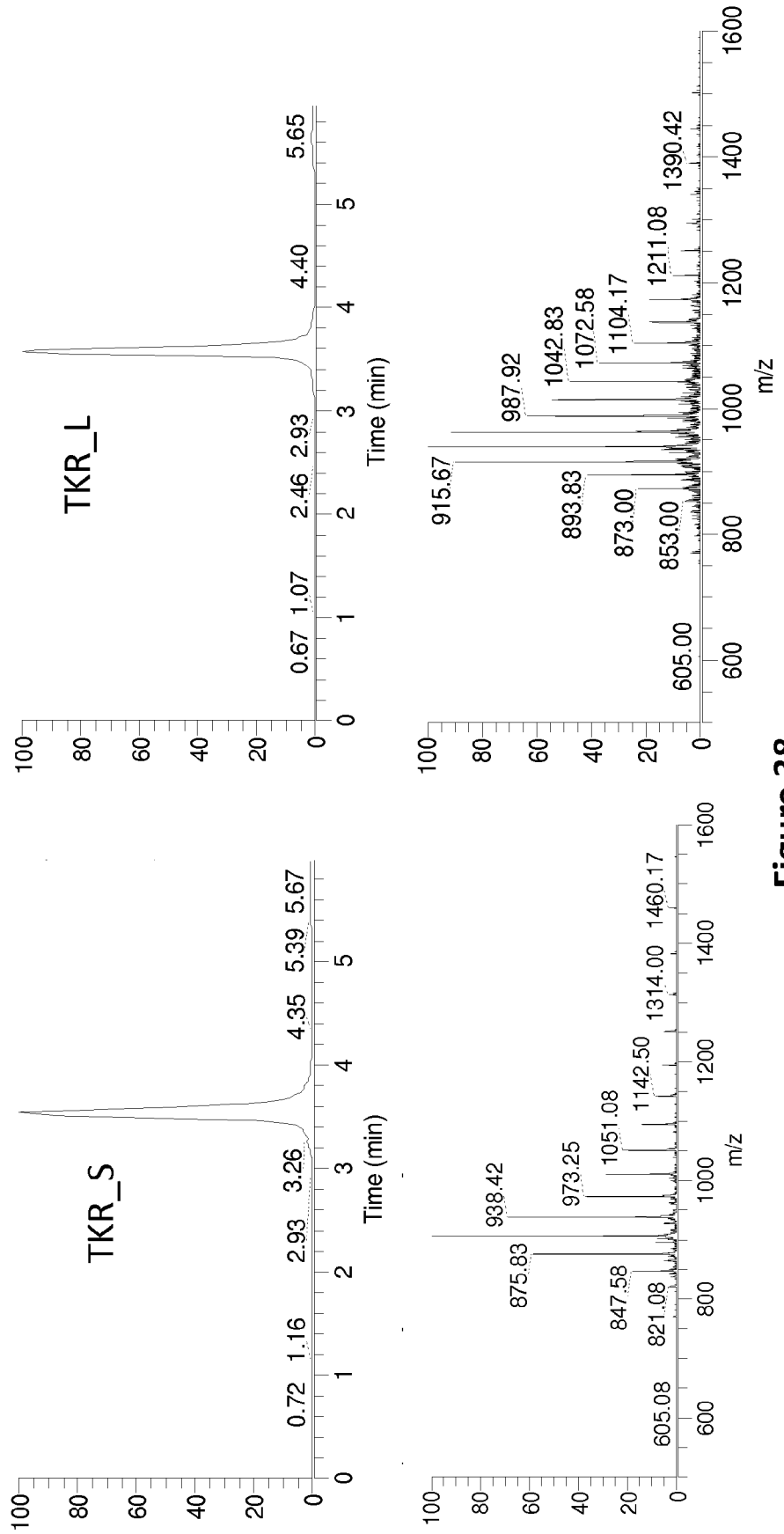
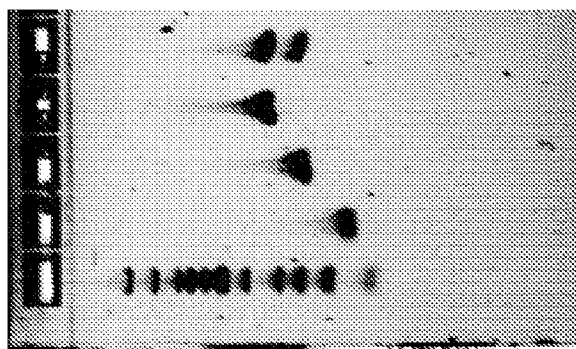


Figure 38



Lane 1 2 3 4 5

Figure 39

TKR_2_click_S
TGCGGTCTAACTGTCTAGTTCACCTTCTCCGGAATGAACAGG(CLICK)GGAATGAACAGGGTAAGCTGGAGTGAAGGGCGTTATG(DBCO)
GAAGGCGTTATGTCGGTACTCTTGCAATCGGGTGCTATGCTT

TKR_2_click_L
TGCGGTCTAACTGTCTAGGCACCTGTTGTTGCCAGTGTGATGGACACCACCTGGAATGAACAGG(CLICK)GGAATGAACAGGGTAAGC
TGGAGTGAAGGCGTTATG(DBCO)GAAGGCGTTATGTGATATCCGTGGTGCATGAGTTCCAATCGGGTGCTATGCTT

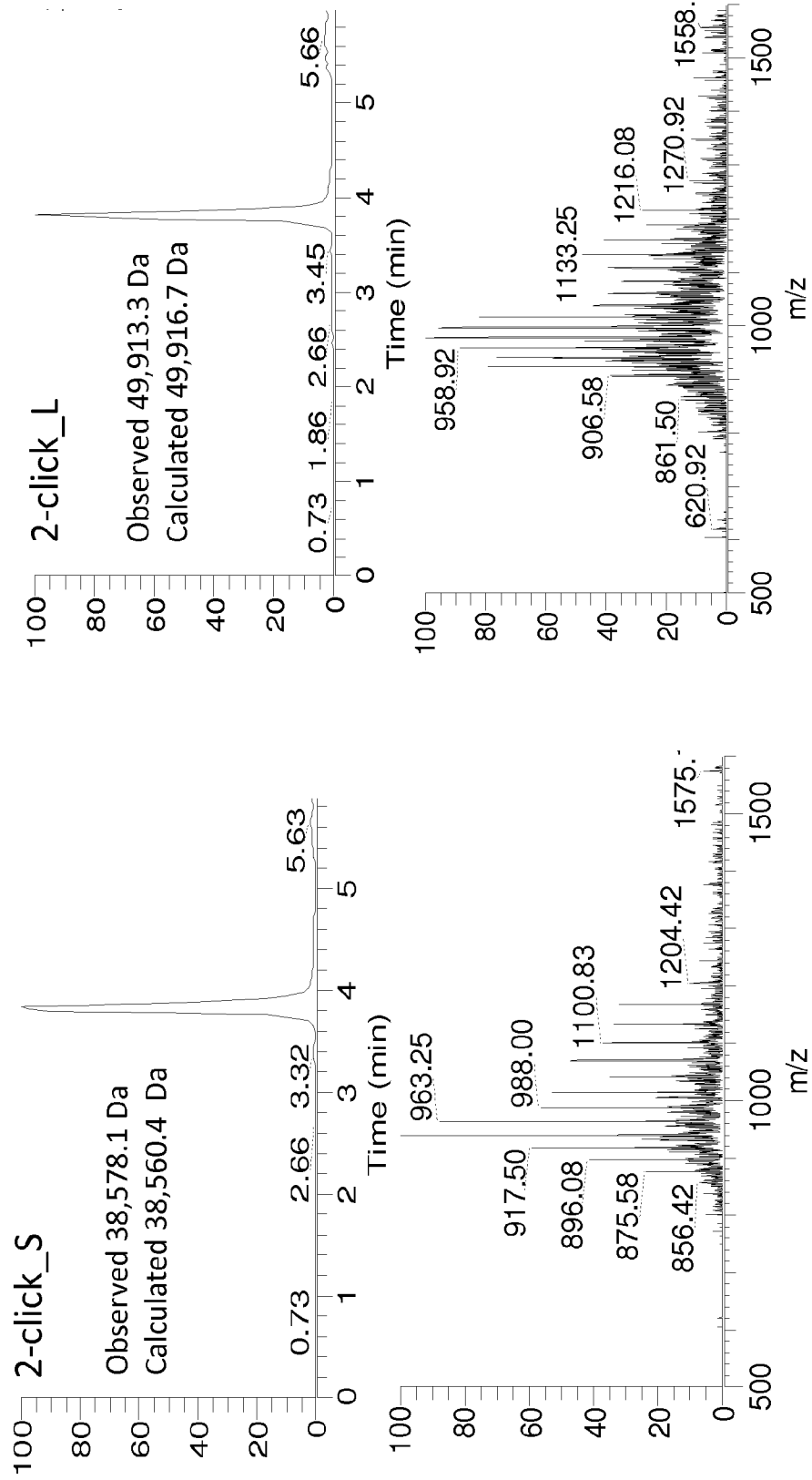
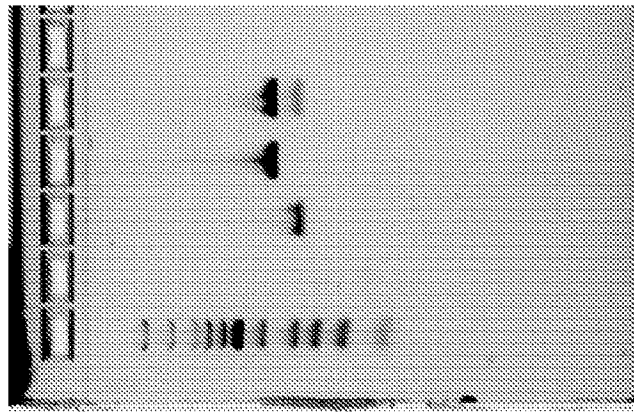


Figure 41



Lane 1 2 3 4 5

Figure 42

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2013/050303

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8)- C40B 40/06 (2013.01)

USPC - 506/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) - C07H 21/00, 21/02; C12Q 1/68; C40B 40/06, 40/08 (2013.01)

USPC - 435/6, 89, 91.1; 506/26, 30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

CPC - C12Q 1/6844, 1/6846; C40B 50/06, 50/08, 50/10 (2013.01)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatBase, Google Patents, PubMed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2012/0053091 A1 (WAGNER) 01 March 2012 (01.03.2012) entire document	1-5, 7, 44-47
Y	US 2011/0262898 A1 (DONG et al) 27 October 2011 (27.10.2011) entire document	1-5, 7, 44-47
A	WO 2011/127933 A1 (GOULIAEV et al) 20 October 2011 (20.10.2011) entire document	1-5, 7, 44-47
A	US 2011/0136697 A1 (MORGAN et al) 09 June 2011 (09.06.2011) entire document	1-5, 7, 44-47

Further documents are listed in the continuation of Box C.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

11 November 2013

Date of mailing of the international search report

29 NOV 2013

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.: 6, 8-43, 48-73
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.