

(19)



(11) Publication number:

SG 173627 A1

(43) Publication date:

29.09.2011

(51) Int. Cl:

;

(12)

Patent Application

(21) Application number: **2011057619**

(71) Applicant:

**X-CHEM, INC. 24 COOLIDGE AVENUE
CAMBRIDGE, MA 02138 MC US**

(22) Date of filing: **16.02.2010**

(30) Priority: **US 61/152,508 13.02.2009**

(72) Inventor:

**WAGNER, RICHARD, W. 24 COOLIDGE
AVENUE CAMBRIDGE, MA 02138 US**

(54) Title:

**METHODS OF CREATING AND SCREENING DNA-
ENCODED LIBRARIES**

(57) Abstract:

The present invention features a number of methods for identifying one or more compounds that bind to a biological target. The methods include synthesizing a library of compounds, wherein the compounds contain a functional moiety having one or more diversity positions. The functional moiety of the compounds is operatively linked to an initiator oligonucleotide that identifies the structure of the functional moiety.

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
19 August 2010 (19.08.2010)

PCT

(10) International Publication Number
WO 2010/094036 A1

(51) International Patent Classification:
C40B 40/08 (2006.01) *C40B 70/00* (2006.01)

(21) International Application Number:
PCT/US2010/024314

(22) International Filing Date:
16 February 2010 (16.02.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/152,508 13 February 2009 (13.02.2009) US

(71) Applicant (for all designated States except US): **X-CHEM, INC.** [US/US]; 24 Coolidge Avenue, Cambridge, MA 02138 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **WAGNER, Richard, W.** [US/US]; 24 Coolidge Avenue, Cambridge, MA 02138 (US).

(74) Agent: **ELBING, Karen, L.**; Clark & Elbing LLP, 101 Federal Street, 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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: METHODS OF CREATING AND SCREENING DNA-ENCODED LIBRARIES

(57) Abstract: The present invention features a number of methods for identifying one or more compounds that bind to a biological target. The methods include synthesizing a library of compounds, wherein the compounds contain a functional moiety having one or more diversity positions. The functional moiety of the compounds is operatively linked to an initiator oligonucleotide that identifies the structure of the functional moiety.



WO 2010/094036 A1

METHODS OF CREATING AND SCREENING DNA-ENCODED LIBRARIES

BACKGROUND OF THE INVENTION

The burgeoning cost of drug discovery has led to the ongoing search for new
5 methods of screening greater chemical space as inexpensively as possible to find
molecules with greater potency and little to no toxicity. Combinatorial chemistry
approaches in the 1980s were originally heralded as being methods to transcend the
drug discovery paradigm, but largely failed due to insufficient library sizes and
inadequate methods of deconvolution. Recently, the use of DNA-displayed
10 combinatorial libraries of small molecules has created a new paradigm shift for the
screening of therapeutic lead compounds.

Morgan et al. (U.S. Patent Application Publication No. 2007/0224607, hereby
incorporated by reference) identifies the major challenges in the use of DNA-
displayed combinatorial approaches in drug discovery: (1) the synthesis of libraries of
15 sufficient complexity and (2) the identification of molecules that are active in the
screens used. In addition, Morgan et al. states that the greater the degree of
complexity of a library, i.e., the number of distinct structures present in the library,
the greater the probability that the library contains molecules with the activity of
interest. Thus, the chemistry employed in library synthesis must be capable of
20 producing vast numbers of compounds within a reasonable time frame. This
approach has been generally successful at identifying molecules with diverse
chemotypes and high affinity. However, a number of issues have surfaced with
respect to generating libraries of enormous complexity and evaluating the sequencing
output on the scale that has been described. For example, purification of a library
25 following multiple chemical transformations (e.g., usually 3 or 4 steps) and biological
transformations (e.g., enzymatic ligation of DNA tags) is cumbersome and results in a
significant amount of "noise" in the library due either to incomplete synthesis of
molecules or to mis-tagging during the ligation step. Furthermore, the amount of
sequencing that is required to interrogate selected populations is striking, usually
30 requiring "nextgeneration" sequencing methods. The latter is due to the fact that
sophisticated genetic tagging schemes embedded in the DNA portion of the library,
together with bioinformatics algorithms for analyzing the "nextgeneration"
sequencing output, are required to sift through the noise and identify hits in the

library. As a result, even with these methodologies, the sequencing is still not advanced enough to fully capture the diversity of sequences (representing both real hits and “noise”) from a given screen.

DNA display of combinatorial small molecule libraries relies on multistep, split-and-pool synthesis of the library, coupled to enzymatic addition of DNA tags that encode both the synthetic step and building block used. Several (e.g., 3 or 4) synthetic steps are typically carried out and encoded, and these include diversity positions (described herein as A, B, and C (Fig. 1)), such as those formed by coupling building blocks with, e.g., amine or carboxylate functional groups onto a chemical scaffold that displays the attached building blocks in defined orientations. One example of a scaffold (S) that is often used in combinatorial libraries is a triazine moiety, which can be orthogonally derivatized in three positions about its ring structure.

The process of library formation can be time consuming, products are often inefficiently purified, and the result is that unknown reactions may occur that create unwanted and/or unknown molecules attached to the DNA. Furthermore, incomplete purification of the library can result in tags cross-contaminating during the ligation steps, resulting in mis-tagging. The end result for screening and sequencing hits from the library is that massively parallel sequencing has to be employed due the inherent “noise” of both DNAs that are attached to molecules that are unintended (e.g., unreacted or side products) or that are mis-tagged. Thus, the efficiency of sequencing is lost.

In some instances, an initiator oligonucleotide, from which the small molecule library is built, contains a primer-binding region for polymerase amplification (e.g., PCR) in the form of a covalently-closed, double-stranded oligonucleotide. This construct is very problematic for performing polymerase reactions, owing to the difficulty of melting the duplex and allowing a primer oligonucleotide to bind and initiate polymerization, which results in an inefficient reaction, reducing yield by 10- to 1000-fold or more.

There exists a need for a more step-wise approach to screening and identifying small molecules that have greater potency and little to no toxicity.

SUMMARY OF THE INVENTION

The present invention features a method for creating and screening simplified DNA-encoded libraries, owing to fewer synthetic steps (e.g., no enzymatic ligation or no covalently closed initiator double-stranded oligonucleotides) and, therefore, substantially less “noise” during the analysis of the encoded oligomers (herein termed “identifier regions”). Thus, sequencing becomes much more efficient, or alternatively, microarray analysis becomes possible, taking into account the inherent biases that can confound interpretation of the data that can be introduced by amplification of the encoding region. We also have identified methods for creating a greater diversity of chemical reactions rather than those simply limited to aqueous conditions to render the DNA-encoded library more hydrophobic and soluble in organic solvents for subsequent steps of library synthesis. In this manner, chemical reactions can be carried out with potentially higher yield, a greater diversity of building blocks, and improved fidelity of the chemical reactions.

Accordingly, the present invention features a method of tagging DNA-encoded chemical libraries by binding a first functional group of a bifunctional linker to an initiator oligonucleotide at the 5' end of the initiator oligonucleotide, wherein the initiator oligonucleotide forms a hairpin structure, and binding a second functional group of the bifunctional linker to a component of the chemical library. The initiator oligonucleotide may include a first identifier region and a second identifier region, such that the second identifier region hybridizes to the first identifier region of the initiator oligonucleotide. The second identifier region may include a fluorescent tag (e.g., a fluorophore or GFP) or biotin label. In addition, the second identifier region is not amplified prior to analysis following a selection step.

In another embodiment, the invention features a method of creating DNA-encoded libraries by (a) creating a first diversity node, (b) encoding the first diversity node in separate vessels, (c) pooling the first diversity node, and (d) splitting the pooled first diversity node into a second set of separate vessels, wherein the first diversity node reacts to form a second diversity node. In certain embodiments, the second diversity node is not encoded and pooled.

In another embodiment, the present invention features a method for creating libraries using semi- or non-aqueous (e.g., organic) chemical reactions with higher yield, a greater diversity of building blocks, and a greater number of chemical

reactions that can be used to create more DNA-tagged combinatorial libraries than previously achieved.

In general, the methods of the present invention provide a set of libraries containing, e.g., one or two diversity positions on a chemical scaffold that can be efficiently generated at high yield, screened to identify preferred individual building blocks or combinations of building blocks that reside at the, e.g., one or two diversity positions, and iteratively diversified at, e.g., a second, third, and/or fourth diversity position to create molecules with improved properties. In addition, the methods described herein allow for an expansive and extensive analysis of the selected compounds having a desired biological property, which, in turn, allows for related compounds with familial structural relationships to be identified (e.g., structure-activity relationships).

By “scaffold” is meant a chemical moiety which displays diversity node(s) in a particular special geometry. Diversity node(s) 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 identifier regions). In some embodiments, the scaffold is derivatized such that it can be orthogonally deprotected during library synthesis and subsequently reacted with different diversity nodes (e.g., using identifier tagging at each step).

By “identifier region” is meant the DNA tag portion of the library that encodes the building block addition to the library.

By “initiator oligonucleotide” is meant the starting oligonucleotide for library synthesis which also contains a covalently attached linker and functional moiety for addition of a diversity node or scaffold. The oligonucleotide can be single- or double-stranded. The oligonucleotide can consist of natural or modified bases.

By “functional moiety” is meant a chemical moiety comprising one or more building blocks that can be selected from any small molecule or designed and built based on desired characteristics of, for example, solubility, availability of hydrogen bond donors and acceptors, rotational degrees of freedom of the bonds, positive charge, negative charge, and the like. The functional moiety must be compatible with chemical modification such that it reacts with the headpiece. In certain embodiments, the functional moiety can be reacted further as a bifunctional or trifunctional (or greater) entity. Functional moieties can also include building blocks that are used at

any of the diversity nodes or positions. Examples of building blocks and encoding DNA tags are found in Tables 1 and 2. See, e.g., U.S. Patent Application Publication No. 2007/0224607, hereby incorporated by reference.

By "building block" is meant a chemical structural unit which is linked to
5 other chemical structural units or can be linked to other such units. When the functional moiety is polymeric or oligomeric, the building blocks are the monomeric units of the polymer or oligomer. Building blocks can also include a scaffold structure (e.g., a scaffold building block) to which is, or can be, attached one or more additional structures (e.g., peripheral building blocks). The building blocks can be
10 any chemical compounds which are complementary (i.e., the building blocks must be able to react together to form a structure comprising two or more building blocks). Typically, all of the building blocks used will have at least two reactive groups, although some of the building blocks used will have only one reactive group each. Reactive groups on two different building blocks should be complementary, i.e.,
15 capable of reacting together to form a covalent bond.

By "linker" is meant a molecule that links the nucleic acid portion of the library to the functional displayed species. Such linkers are known in the art, and those that can be used during library synthesis include, but are not limited to, 5'-O-Dimethoxytrityl-1',2'-Dideoxyribose-3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-
20 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-Dimethoxytritylhexaethyleneglycol,1-
[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite. Such linkers can be added in tandem to one another in different combinations to generate linkers of different
25 desired lengths. By "branched linker" is meant a molecule that links the nucleic acid position of the library to 2 or more identical, functional species of the library. Branched linkers are well known in the art and examples can consist of symmetric or asymmetric doublers (1) and (2) or a symmetric trebler (3). 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);
30 and Jansen et al., *Science* 266: 1226 (1994).

As used herein, the term "oligonucleotide" refers to a polymer of nucleotides. The oligonucleotide may include DNA or any derivative thereof known in the art that

can be synthesized and used for base-pair recognition. The oligonucleotide does not have to have contiguous bases, but can be interspersed with linker moieties. The oligonucleotide polymer may include natural nucleosides (e.g., adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine), nucleoside 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), chemically modified bases, biologically modified bases (e.g., methylated bases), intercalated bases, modified sugars (e.g., 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose), and/or modified phosphate groups (e.g., phosphorothioates and 5'-N-phosphoramidite linkages).

By "operatively linked" is meant that two chemical structures are linked together in such a way as to remain linked through the various manipulations they are expected to undergo. Typically, the functional moiety and the encoding oligonucleotide are linked covalently via an appropriate linking group. For example, the linking group may be a bifunctional moiety with a site of attachment for the encoding oligonucleotide and a site of attachment for the functional moiety.

By "small molecule" is meant a molecule that has a molecular weight below about 1000 Daltons. Small molecules may be organic or inorganic, and may be isolated from, e.g., compound libraries or natural sources, or may be obtained by derivatization of known compounds.

Other features and advantages of the invention will be apparent from the following detailed description, the drawings, the examples, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic illustrating the diversity positions A, B, and C.

Fig. 2 is a schematic of a DNA-encoded chemical library member of Mode 1, showing, in part, the initiator oligonucleotide, which includes a hairpin structure complementary at the identifier region, which has been reacted with A and B diversity nodes. The identifier region for B is being added. In this figure, the "C" diversity node is the potential position for an additional diversity position to be added following the addition of B identifier region.

Fig. 3 is a schematic of a DNA-encoded chemical library member of Mode 1, showing, in part, the initiator oligonucleotide, which includes a sequence in the loop region of the hairpin structure that can serve as a primer binding region for amplification.

5 Fig. 4 is a schematic of a DNA-encoded chemical library member of Mode 1, showing, in part, the initiator oligonucleotide, which includes a non-complementary sequence on the 3' end of the molecule that can serve to bind a second identifier region for either polymerization or for enzymatic ligation.

10 Fig. 5 is a schematic of a DNA-encoded chemical library member of Mode 1, showing, in part, the initiator oligonucleotide, wherein the loop region of the initiator oligonucleotide and at least the identifier region on the 3' side of the loop region can serve to hybridize to a complementary oligonucleotide that also contains a second identifier region.

15 Fig. 6 is a schematic of PCR amplification of the hairpin model, as presented in Figure 5.

Fig. 7 is a schematic of a DNA-encoded chemical library member of Mode 2, showing a hairpin oligonucleotide that is covalently closed (e.g., via a hairpin or chemically) on the distal end to the linker.

20 Fig. 8 is a schematic of a DNA-encoded chemical library member of Mode 2, showing the inclusion of additional diversity nodes.

Fig. 9 is a schematic of a DNA-encoded chemical library member of Mode 2, showing the steps for screening of libraries and methods for deconvoluting the identifier regions.

25 Fig. 10 is a schematic showing oligonucleotides used in library synthesis. Headpiece (HP) was synthesized by IDT DNA and HPLC purified. Arrows indicate the site for BbvCI restriction (underlined) or Nb.BbvCI or Nt.BbvCI nicking digest. Sequences of the DNA tags A1, B1, and C1 (top and bottom strands), the 5' and 3' PCR primers, and the 3' end of the HP are also shown.

30 Fig. 11 is an electrophoretic gel (TBE-urea (15%) gel electrophoresis; UV shadowing on a TLC plate) of the headpiece at different steps of its synthesis. Headpiece HP (IDT DNA) was acylated by Fmoc-amino-PEG2000-NHS (JenKem Technology USA). Lane 1 is the HP (IDT DNA) oligonucleotide (42 nts). Lane 2 is HP acylated with Fmoc-amino-PEG2000-NHS. Following Tris-HCl addition, some

deprotection of Fmoc is observed. Lane 3 is the crude reaction with piperidine, showing complete deprotection of Fmoc. Lane 4 is the same as Lane 3 after desalting on a NAP-5 column and lyophilization. (XC: xylene cyanol (migrates as 60 nt DNA); BPB: bromophenol blue (migrates as 15 nt DNA))

5 Fig. 12 is a schematic showing the steps in model library synthesis. DTAF was conjugated to amino-PEG modified headpiece (HP-1) in the first step. Following this step, a portion of HP-1-DTAF was further acylated with pentylamino-biotin.

 Fig. 13A is a scheme of the ligation of the DNA tags. Fig 13B illustrates a 4% agarose gel of HP-1-DTAF-biotin library at different steps of the DNA tag ligation.
10 M: marker; Lane 1: HP-1-DTAF-biotin; Lane 2: 1 + Tag A only; Lane 3: 1 + Tags A, B, and C, as well as 3'-end oligo ligated. Arrow indicates bright green fluorescence (DTAF). No substantial separation is observed on the gel. Fig. 13C illustrates PCR amplification (24 cycles) of the ligation reactions. M: marker (lowest band is 100);
15 Lane 1: PCR amplification of the green fluorescent band from Lane 1 of Fig. 14B (HP-1-DTAF-biotin + Tag A); Lane 2: PCR amplification of the green fluorescent band from Lane 2 of Fig. 13B (HP-1-DTAF-biotin + all 3 tags and 3'-end oligo);
 Lane 3: PCR amplification of the crude ligation reaction HP-1-DTAF-biotin + all 3 tags; Lane 4: no template control.

 Fig. 14 is a set of electrophoretic gels showing the purification of the XChem
20 model compound and model selection (via a binding interaction between the biotin moiety of the XChem model compound and streptavidin). The gels are 4-12% SDS NuPage gels with MES running buffer. Gels were scanned for green fluorescence using a 450-nm laser. Fig. 14A is a gel showing synthesis and purification steps. Samples were mixed with loading buffer and boiled. M: marker; Lane 1: HP-1 +
25 DTAF; Lanes 2 and 2a: HP-1-DTAF + biotin (two independent reactions); Lanes 3-6 (steps of purification/model selection using streptavidin Dynal beads): Lane 3: flow-through; Lane 4: last wash (washed with water at 80°C for 10 minutes); Lanes 5 and 5': elution with 25 mM EDTA at 90°C (1st and 2nd); Lanes 6 and 6': elution with 25 mM EDTA and 5 mM NaOH at 90°C (1st and 2nd). Fig. 14B is a gel showing binding
30 of HP-1-DTAF-biotin ("library of 1") to streptavidin. Samples were mixed with gel loading buffer and directly loaded onto the gel without boiling. Samples, as in the gel of Fig. 14A, were incubated with an excess of streptavidin in 50 mM NaCl/10 mM Tris HCl, pH 7.0, for 10 minutes. "S" indicates the addition of streptavidin. Samples

5 and 6 were pooled together. Lane 1: HP-1-DTAF; Lane 1S: HP-1-DTAF + streptavidin; Lane 2: HP-1-DTAF-biotin (desalted); Lane 2S: HP-1-DTAF-biotin + streptavidin; Lane 4: last wash (washed with water at 80°C for 10 minutes); Lane 4S: last wash sample + streptavidin; Lane 5+6: pooled samples 5, 5', 6 and 6' (elution fractions from streptavidin beads, purified and selected HP-1-DTAF-biotin; Lane 5+6S': purified and selected HP-1-DTAF-biotin + streptavidin. Note that there is no noticeable difference in migration between different the steps of "library of 1" synthesis. Fig. 14C is a 4% agarose gel of headpiece (Trilink) HP-T, reacted with DTAF. Lane 1: Marker; Lane 2: DTAF; Lane 3 HP-T-DTAF. Left panel: UV visualization of the gel (ethidium bromide staining); Right panel: same gel scanned for fluorescence at excitation wavelength 450 nm (green, fluorescein). Fig 14D is a 4-12% SDS NuPage gel with MES running buffer, showing binding of HP-T-DTAF-biotin to streptavidin. Samples were mixed with gel loading buffer and directly loaded onto the gel without boiling. Samples, as in the gel of Fig. 14A, were incubated with an excess of streptavidin in 50 mM NaCl/10 mM Tris HCl, pH 7.0, for 10 minutes. Lane 1: DTAF; Lane 2: HP-T-DTAF; Lane 3: HP-T-DTAF + streptavidin; Lane 4: HP-T-DTAF-biotin (desalted); Lane 5: HP-T-DTAF-biotin + streptavidin; Lane 6: pooled samples 5, 5', 6 and 6' (elution fractions from streptavidin beads, purified and selected HP-1-DTAF-biotin; Lane 7: purified and selected HP-1-DTAF-biotin + streptavidin.

Fig. 15A is a scheme of the synthesis of the construct for the T7 RNAP intracellular delivery experiment. The V_H dsDNA clone was PCR amplified to append a BsmI site at the 5' end upstream of the T7 promoter. Following restriction digestion and purification, the construct was ligated to HP-1-DTAF-R7 (headpiece modified with DTAF and $(-Arg-\epsilon Ahx)_6-Arg$ peptide). Fig. 15B is an electrophoretic gel of the ligation reaction. Lanes 1 and 2 show different HP-1 samples ligated to V_H ; Lane 3 shows unligated V_H PCR product; and M is the marker. Fig. 15C is an electrophoretic gel showing validation for T7 promoter activity. The gel shows a T7 Megascript (Ambion, Inc.) reaction using samples from Lanes 1-3 of Fig. 15B.

Fig. 16 is an agarose gel electrophoresis of the steps in library 10x10 synthesis. Fig. 16A is a 4% agarose gel of headpiece (Trilink) HP-T ligated with tag A. Lane 1: Marker; Lane 2: HP-T; Lane 3: Tag A annealed; Lane 4: HP-T ligated with tag A; Lane 5: HP-T ligated with tag A and desalted on Zeba column. Fig. 16B is a 2%

agarose gel of HP-T-A ligation with 12 different tags B. Lane M: Marker, Lanes 1 and 9: HP-T-A; Lanes 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 15 and 16: HP-T-A ligation with tags B1- B12. Fig 16C is a 4% agarose gel of the pooled library (library B), with tags A and B1-B12 ligated, after reaction with cyanouric chloride and amines B1-B12.

- 5 Lane 1: Marker; Lane 2: HP-T-A; Lane 3: Library-B pooled and desalted on Zeba columns.

DETAILED DESCRIPTION OF THE INVENTION

The present invention features a number of methods for identifying one or
10 more compounds that bind to a biological target. The methods include synthesizing a library of compounds, wherein the compounds contain a functional moiety having one or more diversity positions. The functional moiety of the compounds is operatively linked to an initiator oligonucleotide that identifies the structure of the functional moiety. In summary, Mode 1 provides a number of methods to preserve the double-
15 stranded character of the dsDNA during library synthesis, which is important during the chemical reaction step, and can be used (as shown in Figs. 2-6) for generating up to two diversity nodes. Mode 2 (Figs. 7-9) anticipates one node of diversity and uses a hairpin oligonucleotide that is covalently closed (e.g., via a hairpin or chemically) on the distal end to the linker. Mode 3 provides methods to create libraries with one,
20 two, three, or more nodes of diversity. Modes 1, 2, and 3 are described in detail below.

Mode 1

The present invention features a method for identifying one or more
25 compounds that bind to a biological target. The method includes synthesizing a library of compounds, wherein the compounds contain a functional moiety having no greater than two diversity positions. The functional moiety of the compounds is operatively linked to an initiator oligonucleotide that identifies the structure of the functional moiety by providing a solution containing A initiator compounds.

30 The initiator oligonucleotide includes a linker L (e.g., polyethylene glycol) with an integer of one or greater, wherein the initiator oligonucleotides contain a functional moiety that includes A building blocks attached to L and separated into A

reaction vessels, wherein A is an integer of two or greater, which is operatively linked to an initiator oligonucleotide that identifies the A building blocks.

In some embodiments, the A building blocks can be further derivatized through a common node S. In other embodiments, A is subsequently transformed
5 with S, S being a scaffold molecule that allows further nodes of diversity introduction. In some embodiments, A-S can be screened directly, representing a single node of diversity. In other embodiments, the A-S reaction vessels (e.g., which may first include a purification of A-S from starting materials) are mixed together and aliquoted into B reaction vessels, wherein B is an integer of one or greater, and reacted with one
10 of B building blocks. A-S-B, still in B reaction vessels, is in some cases reacted with a C building block, where C is an integer of one, is purified, and subjected to a polymerization or ligation reaction using B primers, in which the B primers differ in sequence and identify the B building blocks.

In certain embodiments, A-S can be an integer of one. In one embodiment, A-
15 S can be linked directly to B initiator oligonucleotides, and following reaction of B building blocks, the B reactions are mixed. In certain embodiments, the A-S-B mixture, where B represents the only diversity node, is screened directly, representing a single node of diversity. In other embodiments, the A-S-B mixture, where B represents the only diversity node, is subsequently aliquoted into C reaction vessels,
20 reacted with C building blocks, and subjected to second strand polymerization or ligation reaction using C primers, in which the C primers differ in sequence and identify the C building blocks.

In certain embodiments, B can be an integer of one and A-S is greater than one, in which case A-S, now derivatized with B, is aliquoted into C reaction vessels,
25 reacted with C building blocks, and subjected to second strand polymerization reaction using C primers, in which the C primers differ in sequence and identify the C building blocks. This general strategy can be expanded to include additional diversity nodes (e.g., D, E, F, etc.) so that the first diversity node is reacted with building blocks and/or S and encoded by an initial oligonucleotide, mixed, re-aliquoted into
30 vessels and then the subsequent diversity node is derivatized by building blocks, which is encoded by the primer used for the polymerization or ligation reaction.

In certain embodiments, A can be an integer of one, B can be an integer of one, and C initiator oligonucleotides are used. A-S-B, attached to C initiator

oligonucleotides, is formed in C reaction vessels, reacted with C building blocks, and screened directly.

In certain embodiments, S is reacted first with the initiator oligonucleotide, and A, B and/or C (e.g., or D, E, F, and so on) are subsequently reacted.

5 In certain embodiments, A, B, or C (e.g., or D, E, F, and so on) can contain sites for additional diversity nodes. If this is the case, then S may or may not be used or needed to introduce additional diversity nodes.

In one embodiment, the initiator oligonucleotide includes a hairpin structure complementary at the identifier region (Fig. 2). The identifier region can be, e.g., 2 to
10 100 base pairs in length, preferably 5 to 20 base pairs in length, and most preferably 6 to 12 base pairs in length. The initiator oligonucleotide further includes a sequence in the loop region of the hairpin structure that can serve as a primer binding region for amplification (Fig. 3), such that the primer binding region has a higher melting temperature for its complementary primer (e.g., which can include flanking identifier
15 regions) than the identifier region alone.

In one embodiment, the loop region may include modified bases that can form higher affinity duplex formations than unmodified bases, such modified bases being known in the art (Fig. 3). The initiator oligonucleotide can further include a non-complementary sequence on the 3' end of the molecule that can serve to bind a
20 second identifier region for either polymerization or for enzymatic ligation (Fig. 4). In one embodiment, the strands can be subsequently crosslinked, e.g., using psoralen.

In another embodiment, the loop region and at least the identifier region on the 3' side of the loop region can serve to hybridize to a complementary oligonucleotide that also contains a second identifier region (Fig. 5). In cases where many building
25 blocks and corresponding tags are used (e.g., 100 tags), 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. In one embodiment, the strands can be subsequently crosslinked, e.g., using psoralen. The resultant library members can be amplified by PCR following selection for binding
30 entities versus a target(s) of interest (Fig. 6).

For example, a headpiece, which includes an initiator oligonucleotide, may be reacted with a linker and A, which includes, for example, 1000 different variants. For each A building block, a DNA tag A may be ligated or primer extended to the

headpiece. These reactions may be performed in, e.g., a 1000-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 B building blocks, which also include, for example, 1000 different variants. A DNA tag B may be
5 ligated to the headpiece, and all reactions may be pooled. A library of 1000 x 1000 combinations of A to B (i.e., 1,000,000 compounds), tagged by 1,000,000 different combinations of tags. The same approach may be extended to add variants C, D, E, etc. The generated library may then be used to identify compounds that bind to the target. The composition of the compounds that bind to the library can be assessed by
10 PCR and sequencing of the DNA tags to identify the compounds that were enriched.

Mode 2

In another embodiment (Fig. 7), the method includes synthesizing a library of compounds, wherein the compounds contain a functional moiety having no greater
15 than two diversity positions. The functional moiety of the compounds is operatively linked to an initiator oligonucleotide, which contains a unique genetic sequence that identifies the structure of the functional moiety by providing a solution comprising A initiator compounds, wherein L is an integer of one or greater, where the initiator compounds include a functional moiety having A building blocks separated into A
20 reaction vessels, where, e.g., A is an integer of two or greater, which is operatively linked to an initial oligonucleotide which identifies the A building blocks. In some embodiments, the A building blocks are pre-derivatized with a common S. In other embodiments, A is subsequently transformed with S, S being a scaffold molecule that allows further nodes of diversity introduction. Next, the A-S reaction vessels (which
25 may first include a purification of A-S from starting materials) are mixed together and aliquoted into B reaction vessels, wherein B is an integer of one or greater, and reacted with one of B building blocks. A-S-B, still in B reaction vessels is, in some embodiments, reacted with a C building block, where C is an integer of one, are purified, and kept separate in B vessels for screening. In some embodiments, A-S is
30 an integer of one. In one embodiment, A-S can be linked directly to B initiator oligonucleotides and, following the reaction of B building blocks, the B reactions are mixed and aliquoted into C reaction vessels, reacted with C building blocks, and kept separate in C vessels for screening. In other embodiments, B can be an integer of one

and A-S is greater than one, in which case A-S, now derivatized with B, is aliquoted into C reaction vessels reacted with C building blocks, and kept separate in C vessels for screening. This general strategy can be expanded to include additional diversity nodes (e.g., D, E, F, etc.) so that the first diversity node is reacted with building
5 blocks and/or S and encoded by an initiator oligonucleotide, mixed, re-aliquoted into vessels, and then the subsequent diversity node is derivatized by building blocks and kept in their respective vessels for screening (Fig. 8).

For example, as described in Mode 1, a headpiece, which includes an initiator oligonucleotide, may be reacted with a linker and A building blocks, which include,
10 for example, 1000 different variants. For each A building block, a DNA tag A may be ligated or primer extended to the headpiece. The reactions may be pooled. Next, the same procedure may be performed with B building blocks, but a DNA tag is not added for B. Because B is not coded for, all "B" reactions may be pooled (e.g., 1000 reactions) and a selection step may be performed to identify all A building blocks that
15 produce the desired binding effect with unknown B building blocks. A library of A building blocks identified in the selection step (e.g., 10 A building blocks) may then be reacted with the same 1000 B building blocks, resulting in a screen of 10,000 compounds or less. In this round, DNA tags for B may be added and B building blocks that produce the desired binding effect in combination with the, e.g., 10 A
20 building blocks can be identified, resulting in a step-wise convolution of an initial library of, for example, 1,000,000 compounds. A set of these final compounds may be individually tested to identify the best, e.g., binders, activators, or inhibitors.

To avoid pooling all of the reactions after B synthesis, a BIND Reader (SRU Biosystems), for example, may be used to monitor binding on a sensor surface in high
25 throughput format (e.g., 384 well plates and 1536 well plates). For example, the A building blocks may be encoded with DNA tags and the B building blocks may be position encoded. Binders can then be identified using a BIND sensor, sequencing, and microarray analysis or restriction digest analysis of the A tags. This analysis allows for the identification of combinations of A and B building blocks that produce
30 the desired molecules. Other methods for monitoring binding known to those of skill in the art may be used including, e.g., ELISA.

Modes 1 and 2

The initiator oligonucleotide of Modes 1 and 2 may contain a hairpin structure, complementary at the identifier region. The initiator oligonucleotide further contains a sequence in the loop region of the hairpin structure that can serve as a
5 primer-binding region for amplification, such that the primer binding region has a higher melting temperature for its complementary primer (which can include flanking identifier regions) than the identifier region alone.

In one embodiment, the initiator oligonucleotide includes a linker molecule capable of being functionally reacted with building blocks. The linker molecule can
10 be attached directly to the 5' end of the oligonucleotide through methods known in the art or can be embedded within the molecule, e.g., off of a derivatized base (e.g., the C5 position of uridine), or the linker can be placed in the middle of the oligonucleotide using standard techniques known in the art.

The initiator oligonucleotide may be single-stranded or double-stranded. The
15 formation of a double-stranded oligonucleotide may be achieved through hairpin formation of the oligonucleotide or through cross-linking using, e.g., a psoralen moiety, as known in the art.

The initiator oligonucleotide may contain two primer-binding regions (e.g., to enable a PCR reaction) on either side of the identifier region that encodes the building
20 block. Alternatively, the initiator oligonucleotide may contain one primer-binding site on the 5' end. In other embodiments, the initiator oligonucleotide 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 identifier region on the 3' side of the loop. A primer oligonucleotide, containing a region homologous to the
25 3' end of the initiator oligonucleotide and carrying a primer binding region on its 5' end (e.g., to enable a PCR reaction) may be hybridized to the initiator oligonucleotide, and may contain an identifier region that encodes the building blocks used at one of the diversity positions. The primer oligonucleotide may contain additional information, such as a region of randomized nucleotides, e.g., 2 to 16 nucleotides in
30 length, which is included for bioinformatic analysis.

In one embodiment, the initiator oligonucleotide does not contain a PCR primer-binding site.

In another embodiment, the library of compounds, or a portion thereof, is contacted with a biological target under conditions suitable for at least one member of the library of compounds to bind to the target, followed by removal of library members that do not bind to the target, and analyzing the identifier region or regions.

5 Exemplary biological targets include, e.g., 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, or carbohydrates).

10 In one embodiment, the library of compounds, or a portion thereof, is contacted with a biological target under conditions suitable for at least one member of the library of compounds to bind to the target, followed by removal of library members that do not bind to the target, followed by amplification of the identifier region by methods known in the art, and subsequently analyzing the identifier region
15 or regions by methods known in the art.

In one embodiment the method of amplification of the identifier region can include, e.g., polymerase chain reaction (PCR), linear chain amplification (LCR), rolling circle amplification (RCA), or any other method known in the art to amplify nucleic acid sequences.

20 In a further embodiment, the library of compounds 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.

In another embodiment, the molecules that bind to a target are not subjected to amplification, but are analyzed directly. Methods of analysis include, e.g., microarray
25 analysis or bead-based methods for deconvoluting the identifier regions (Fig. 9). Molecules that bind during the screening step may also be detected by a label-free photonic crystal biosensor.

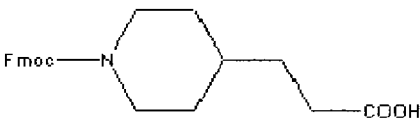
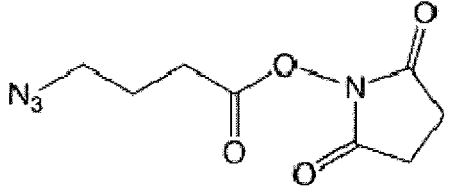
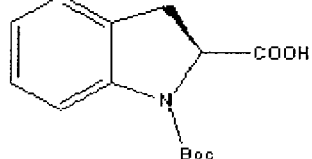
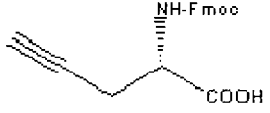
In one embodiment, the initiator oligonucleotide and/or the primer oligonucleotide contain a functional moiety that allows for its detection by, e.g.,
30 fluorescent tags, Q dots, or biotin.

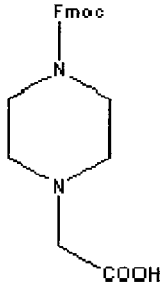
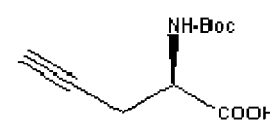
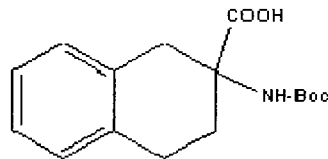
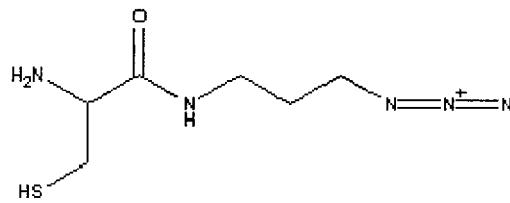
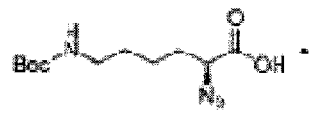
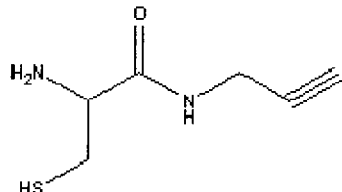
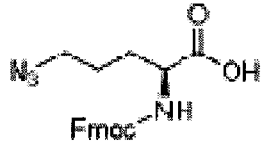
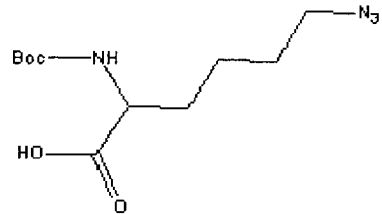
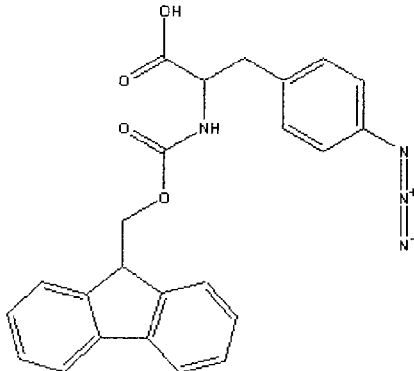
In one embodiment, the microarray analysis uses advanced detection capability, such as, e.g., evanescent resonance photonic crystals.

- In one embodiment, the method of amplifying includes forming a water-in-oil emulsion to create a plurality of aqueous microreactors, wherein at least one of the microreactors has at least one member of a library of compounds that binds to the target, a single bead capable of binding to the encoding oligonucleotide of the at least one member of the library of compounds that binds to the target, and amplification reaction solution containing reagents necessary to perform nucleic acid amplification, amplifying the encoding oligonucleotide in the microreactors to form amplified copies of the encoding oligonucleotide, and binding the amplified copies of the encoding oligonucleotide to the beads in the microreactors.
- 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, in which one or two additional nodes of diversity are added, and the library is created and diversity sampled as described herein. This process can be repeated as many times as necessary to create molecules with desired molecular and pharmaceutical properties.

Exemplary A building blocks include, e.g., amino acids (not limited to alpha-amino acids), click-chemistry reactants (e.g., azide or alkine chains) with an amine, or a thiol reactant. The choice of A building block 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. See, e.g., Table 1.

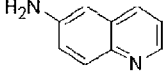
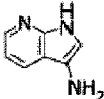
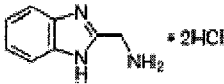
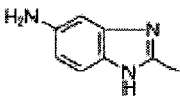
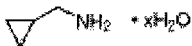
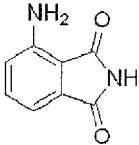
Table 1. Exemplary Position A Building Blocks

 3-(1-Fmoc-piperidine-4-yl)-propionic acid	 4-Azido-butan-1-oic acid N-Hydroxysuccinimide Ester
 Boc-L-indoline-2-carboxylic acid	 Fmoc-L-propargylglycine

 Fmoc-(4-carboxymethyl)piperazine	 Boc-D-propargylglycine•DCHA
 Boc-2-amino-1,2,3,4-tetrahydronaphthalene-2-carboxylic acid	 2-Amino-N-(3-azidopropyl)-3-mercaptopropionamide
 (S)-(-)-2-Azido-6-(Boc-amino)hexanoic acid	 2-Amino-3-mercapto-N-(prop-2-ynyl)propionamide
 (S)-5-Azido-2-(Fmoc-amino)pentanoic acid	 Boc-Lys(N₃)-OH
 Fmoc-4-azidophenylalanine	

Exemplary B and C building blocks are described in Tables 2 and 3, respectively. A restriction site may be introduced, for example, in the B or C position for analysis of the final product and selection by performing PCR and restriction digest with one of the corresponding restriction enzymes.

Table 2. Examples of Position B Building Blocks and Encoding DNA Tags

Chemical Name and Structure	Restriction Site (Restriction Enzyme)	Top Strand Bottom Strand
 6-Aminoquinoline (B1)	T/CCGGA (BspEI)	5'-Phos-CCTCCGGAGA (SEQ ID NO: 1) 5'-Phos-TCCGGAGGAC (SEQ ID NO: 2)
 3-Amino-7-azaindole, 1H-pyrrolo[2,3-b]pyridin-3-ylamine (B2)	GGC/GCC (SfoI)	5'-Phos-CCGGCGCCGA (SEQ ID NO: 3) 5'-Phos-GGCGCCGGAC (SEQ ID NO: 4)
 2-(Aminomethyl) benzimidazole dihydrochloride (B3)	GGTAC/C (KpnI)	5'-Phos-CCGGTACCGA (SEQ ID NO: 5) 5'-Phos-GGTACCGGAC (SEQ ID NO: 6)
 2-Methyl-1H-benzimidazol-5-amine (B4)	CAC/GTG (PmlI)	5'-Phos-CCCACGTGGA (SEQ ID NO: 7) 5'-Phos-CACGTGGGAC (SEQ ID NO: 8)
 (Aminomethyl)cyclopropane (B5)	GAGCT/C (SacI)	5'-Phos-CCGAGCTCGA (SEQ ID NO: 9) 5'-Phos-GAGCTCGGAC (SEQ ID NO: 10)
 3-Aminophthalimide (B6)	G/GATCC (BamHI)	5'-Phos-CCGGATCCGA (SEQ ID NO: 11) 5'-Phos-GGATCCGGAC (SEQ ID NO: 12)

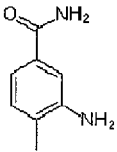
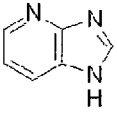
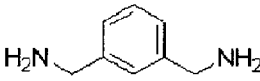
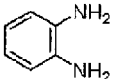
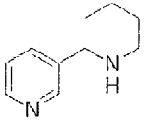
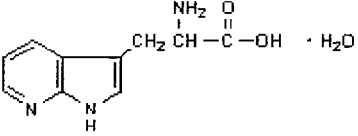
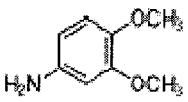
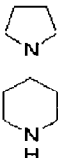
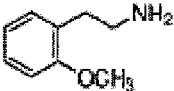
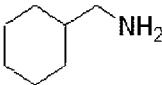
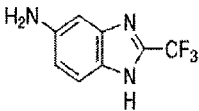
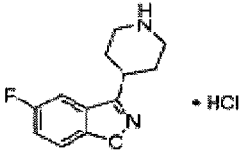
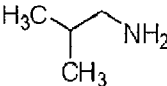
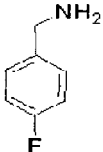
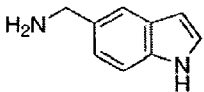
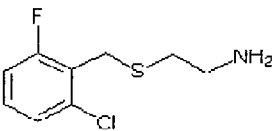
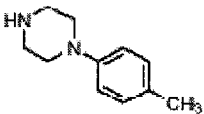
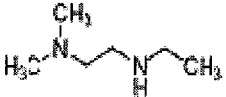
 3-Amino-4-methylbenzamide (B7)	AT/CGAT (BspDI)	5'-Phos-CCATCGATGA (SEQ ID NO: 13) 5'-Phos-ATCGATGGAC (SEQ ID NO: 14)
 4-Azabenzimidazole (B8)	A/AGCTT (HindIII)	5'-Phos-CCAAGCTTGA (SEQ ID NO: 15) 5'-Phos-AAGCTTGGAC (SEQ ID NO: 16)
 m-Xylylenediamine (B9)	A/GATCT (BglII)	5'-Phos-CCAGATCTGA (SEQ ID NO: 17) 5'-Phos-AGATCTGGAC (SEQ ID NO: 18)
 1,2-Phenylenediamine (B10)	G/AATTC (EcoRI)	5'-Phos-CCGAATTCGA (SEQ ID NO: 19) 5'-Phos-GAATTCGGAC (SEQ ID NO: 20)
 Anabasine (B11)	T/GATCA (BclI)	5'-Phos-CCTGATCAGA (SEQ ID NO: 21) 5'-Phos-TGATCAGGAC (SEQ ID NO: 22)
 DL-7-Azatryptophan hydrate (B12)	CA/TATG (NdeI)	5'-Phos-CCCATATGGA (SEQ ID NO: 23) 5'-Phos-CATATGGGAC (SEQ ID NO: 24)

Table 3. Examples of Position C Building Blocks and Encoding DNA Tags

Chemical Name and Structure	Top Strand Bottom Strand
 3,4-Dimethoxyaniline (C1)	5'-Phos-GAACCTGCTT (SEQ ID NO: 25) 5'-Phos-GCAGGTTCTC (SEQ ID NO: 26)

 4-(1-Pyrrolidinyl)piperidine (C2)	5'-Phos-GAAGACGCTT (SEQ ID NO: 27) 5'-Phos-GCGTCTTCTC (SEQ ID NO: 28)
 2-Methoxyphenethylamine (C3)	5'-Phos-GACCAGACTT (SEQ ID NO: 29) 5'-Phos-GTCTGGTCTC (SEQ ID NO: 30)
 Cyclohexanemethylamine (C4)	5'-Phos-GACGACTCTT (SEQ ID NO: 31) 5'-Phos-GAGTCGTCTC (SEQ ID NO: 32)
 2-(1-Cyclohexenyl)ethylamine (C5)	5'-Phos-GACGCTTCTT (SEQ ID NO: 33) 5'-Phos-GAAGCGTCTC (SEQ ID NO: 34)
 5-Amino-2-(trifluoromethyl)benzimidazole (C6)	5'-Phos-GAGCAACCTT (SEQ ID NO: 35) 5'-Phos-GGTTGCTCTC (SEQ ID NO: 36)
 5-Fluoro-3-(4-piperidiny)-1,2- benzisoxazole hydrochloride (C7)	5'-Phos-GAGCCATCTT (SEQ ID NO: 37) 5'-Phos-GATGGCTCTC (SEQ ID NO: 38)
 Isobutylamine (C8)	5'-Phos-GCAACCACTT (SEQ ID NO: 39) 5'-Phos-GTGGTTGCTC (SEQ ID NO: 40)
 4-Fluorobenzylamine (C9)	5'-Phos-GCACAGACTT (SEQ ID NO: 41) 5'-Phos-GTCTGTGCTC (SEQ ID NO: 42)

 5-(Aminomethyl)indole (C10)	5'-Phos-GCGATCACTT (SEQ ID NO: 43) 5'-Phos-GTGATCGCTC (SEQ ID NO: 44)
 2-[(2-chloro-6-fluorobenzyl)thio]ethylamine (C11)	5'-Phos-GCGGTTACTT (SEQ ID NO: 45) 5'-Phos-GTAACCGCTC (SEQ ID NO: 46)
 1-(4-Methylphenyl)piperazine (C12)	5'-Phos-GCATGACCTT (SEQ ID NO: 47) 5'-Phos-GGTCATGCTC (SEQ ID NO: 48)
 N,N-Dimethyl-N'-ethylethylenediamine (C13)	5'-Phos-GCGTACTCTT (SEQ ID NO: 49) 5'-Phos-GAGTACGCTC (SEQ ID NO: 50)

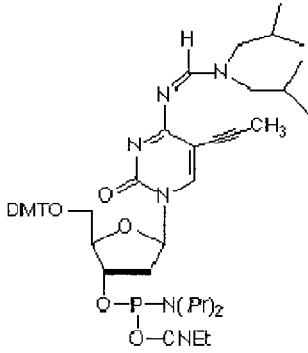
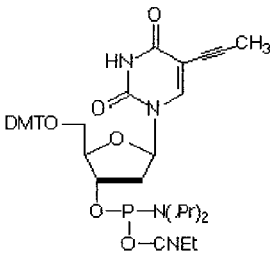
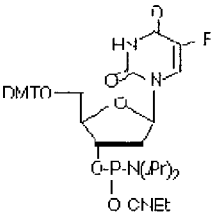
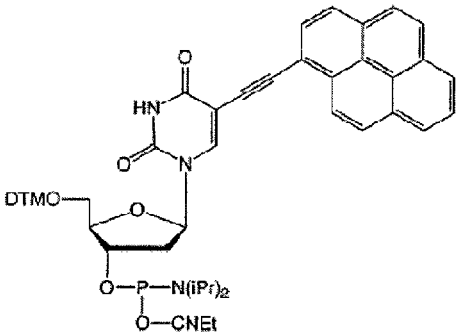
Mode 3

In either of the modes described herein (e.g., Modes 1 and 2), the headpiece oligonucleotide may be modified to support solubility in semi- or non-aqueous (e.g., organic) conditions. The headpiece, in certain embodiments, includes the identifier region. In other embodiments, the headpiece with linker can first be derivatized with a building block (e.g., a functional moiety) or scaffold, and the identifier sequence is then added.

Nucleotide bases of the headpiece 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. See, e.g., Table 4 for examples of modified bases. 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 reactions 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 nucleic acid 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.

Table 4. Exemplary modified nucleotide bases

 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	 5'-Dimethoxytrityl-5-(pyren-1-ylethynyl)-2'-deoxyUridine,3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite

The linker molecule between the headpiece and small molecule library can be varied to increase the solubility of the headpiece in organic solvent. A wide variety of linkers are commercially available that can couple the headpiece with the small molecule library. Linkers are empirically selected for a given small molecule library design (scaffolds and building blocks) such that the library can be synthesized in organic solvent, for example, 15%, 25%, 30%, 50%, 75%, 90 %, 95%, 98%, 99%, or 100% organic solvent. The linker can be varied using model reactions prior to library synthesis to select the appropriate chain length that solubilizes the headpiece in organic solvent. Such linkers may include linkers with, e.g., increased alkyl chain length, increased polyethylene 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).

The linker molecule may provide an appropriate spacer between the headpiece DNA and member of a chemical library. For example, bifunctional linkers may be used. In certain embodiments, bifunctional linkers may include, for example, 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' end of a single-stranded DNA headpiece (achieved by means of standard oligonucleotide chemistry), click chemistry pairs (azide alkyne cycloaddition in the presence of Cu(I) catalyst), or thiol reactive groups. Part 2 may also be a reactive group, which forms a covalent bond with the chemical library, either a building block in the position A or scaffold moiety. Such a reactive group could be, e.g., an amine, a thiol, an azide, or an alkyne for water based reactions or multiple other reactive groups for the organic-based reactions. Part 3 may be a chemically inert spacer of variable length, introduced between Part 1 and 2. Such a spacer can be a chain of ethylene glycol units (e.g., PEGs of different lengths), an alkane, an alkene, polyene chain, or peptide chain. The linker can contain branches or inserts with hydrophobic moieties (such as, e.g., benzene rings) to improve solubility of the headpiece in organic solvents, as well as fluorescent moieties (e.g. fluorescein or Cy-3) used for library detection purposes.

Examples of commercially available linkers include, e.g., amino-carboxylic linkers (e.g., peptides (e.g., Z-Gly-Gly-Gly-Osu or Z-Gly-Gly-Gly-Gly-Gly-

Osu), PEG (e.g., Fmoc-aminoPEG2000-NHS or amino-PEG (12-24)-NHS), or alkane acid chains (e.g., Boc- ϵ -aminocaproic acid-Osu)), click chemistry linkers (e.g., peptides (e.g., azidohomalanine-Gly-Gly-Gly-OSu or propargylglycine-Gly-Gly-Gly-OSu), PEG (e.g., azido-PEG-NHS), or alkane acid chains (e.g., 5-azidopentanoic acid, (S)-2-(azidomethyl)-1-Boc-pyrrolidine, or 4-azido-butan-1-oic acid N-hydroxysuccinimide ester)), thiol-reactive linkers (e.g., PEG (e.g., SM(PEG)_n NHS-PEG-maleimide), alkane chains (e.g., 3-(pyridin-2-yl-disulfanyl)-propionic acid-Osu or sulfosuccinimidyl 6-(3'-[2-pyridyldithio]-propionamido)hexanoate))), amidites for oligonucleotide synthesis (e.g., 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 click chemistry 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)).

Hydrophobic residues in the headpiece design may be varied with the linker design to facilitate library synthesis in organic solvents. For example, the headpiece and linker combination is designed to have appropriate residues wherein the octanol:water coefficient (P_{oct}) is from, e.g., 1.0 to 2.5.

20

EXAMPLES

The following examples are intended to illustrate the invention. They are not meant to limit the invention in any way.

25 Example 1. Preparation of the headpiece (Variant 1)

A phosphorylated oligonucleotide headpiece (oligo HP) having the following sequence was synthesized and HPLC purified by IDT DNA.

5'-(phosphate)TCCTGGCTGAGGCGAGAGTT(dT-C6-NH)
TTCTCTCGCCTCAGCCAGGACC-3' (SEQ ID NO: 51)

30 The oligonucleotide folds into a hairpin (Fig. 10) with an overhang and contains a cleavage site (CCTCAGC) for restriction enzyme BbvCI or nicking versions of this enzyme Nb.BbvCI or Nt.BbvCI, which can cleave either the top or bottom strand (New England BioLabs). In the middle of the hairpin loop, the side

chain C5-amino-modified dT is inserted (dT-C6-NH; "C6" refers to a carbon 6 linker), which was used for the coupling of the amino-PEG linker (PEG2000, approximately 45 ethylene glycol units). The top and bottom strands of the DNA tags A, B, and C were synthesized and purified by IDT DNA and purified by standard
5 desalting. Longer oligonucleotides, such as the 3' end and PCR primers, were synthesized by IDT DNA and HPLC purified.

Ten nanomoles of the oligo HP were dissolved in 50 µl water. A 20-fold molar excess of Fmoc-amino-PEG2000-carboxyl-NHS ester (JenKem Technology USA) was dissolved in 50 µl dimethylformamide (DMF) and was added to the
10 oligonucleotide solution in 2 portions over the course of 2 hours at room temperature (final solvent composition of 50% DMF/50% water). Subsequently, 60 µl of 1 M Tris HCl, pH 7.0 (final concentration of 200 mM), was added to quench the excess of NHS esters, and the solution was incubated for an additional 30 minutes at room temperature. The resulting reaction mixture was diluted to 500 µl with water and was
15 desalted by passing through a NAP-5 column (Sephadex-25, GE Healthcare).

The resulting material was lyophilized and dissolved in 100 µl water. 20 µl of piperidine (to a final concentration of 20%) was added and incubated for 2 hours at room temperature. A cloudy precipitate was formed due to deprotection of the amine and release of the water insoluble Fmoc group. The reaction was then filtered through
20 0.2-µm spin-filters (Millipore) and precipitated using 300 mM sodium acetate by the addition of 3 volumes of ethanol. The Fmoc-protected form of the modified oligonucleotide was found to be soluble in ethanol and isopropanol. Due to high coupling efficiency, the resulting headpiece (HP-1) was used without further purification (Fig. 11).

25

Example 2. Preparation of the headpiece (Variant 2)

A complete headpiece (HP-1) having the following sequence was prepared by Trilink, Inc., following a similar procedure as described above, and RP-HPLC purified.

30 5'-(phosphatc)TCCTGGCTGAGGCGAGAGTT(dT-C6-NH)(X)
TTCTCTCGCCTCAGCCAGGACC-3' (SEQ ID NO: 52)

where X stands for amino-PEG2000.

Example 3. Synthesis of a model library member*Step 1: Coupling of DTAF*

In order to prepare a “library of 1,” a model compound, 5-(4,6-dichlorotriazinyl-aminofluorescein) (DTAF; Anaspec) (Fig. 12), was coupled to the amino group of HP-1. DTAF structurally represents a trichlorotriazine scaffold with one amino compound coupled. To form a library, trichlorotriazine scaffolds can be derivatized with a diversity of building blocks at each of the three chlorine positions. DTAF also provides a fluorescent label to the model library. The reaction (10 µl) was set up as follows. To 5 µl of 400 µM HP-1 dissolved in water, 2 µl of 750 mM borate buffer, pH 9.5, and 1 µl of DMF were added. DTAF was dissolved in DMF to 50 mM and 2 µl was added to the reaction. Final concentrations of the HP-1 and DTAF were 200 µM and 10 mM, respectively, thus generating a 50-fold excess of DTAF. The final DMF concentration was 30%. It was noticed that HP-1 stayed soluble in up to 90% DMF, demonstrating that it was soluble in an organic solvent, e.g., DMF. The reaction was allowed to proceed at 4°C for 16-20 hours. The reaction mixture was then diluted with water to 30-50 µl and desalted on a Zeba spin column (Pierce). No further purification was completed at this point.

Step 2: Coupling of amino-biotin

After DTAF was coupled to HP-1, one more reactive group on the scaffold molecule is still available for modification. We chose an amino-biotin analog, EZ-Link Pentylamine-Biotin (Pierce), to couple at this position in order to generate a model binding compound (Fig. 12). The reaction was set up as follows. 20 µl of the reaction mixture contained around 200 pmol of HP-1-DTAF (Step 1) dissolved in 150 mM borate buffer, pH 9.5, and 10 nmol of pentylamine-biotin. The reaction was allowed to proceed for 4-12 hours at 75°C. The reaction was then purified by desalting on a Zeba spin column, as described above.

Step 3: Ligation of the DNA tags to HP-1-DTAF-biotin

Phosphorylated DNA tags (3' end primer region and 5' and 3' PCR primers) were synthesized by IDT DNA. Oligonucleotide sequences (Fig. 13) are as follows.

DNA Tag A1 (top): 5'-phos-GGAGGACTGT (SEQ ID NO: 53)

DNA Tag A1 (bottom): 5'-phos-AGTCCTCCGG (SEQ ID NO: 54)

DNA Tag B1 (top): 5'-phos-CAGACGACGA (SEQ ID NO: 55)

DNA Tag B1 (bottom): 5'-phos-GTCGTCTGAC (SEQ ID NO: 56)

DNA Tag C1 (top): 5'-phos-CGATGCTCTT (SEQ ID NO: 57)

5 DNA Tag C1 (bottom): 5'-phos-GAGCATCGTC (SEQ ID NO: 58)

3' end (top): 5'-phos-GCTGTGCAGGTAGAGTGC-3' (SEQ ID NO: 59)

3' end (bottom): 5'-AACGACACGTCCATCTCACG (SEQ ID NO: 60)

5' PCR primer: 5'-CTCTCGCCTCAGCCAGGA (SEQ ID NO: 61)

3' PCR primer: 5'-GCACTCTACCTGCACAGC (SEQ ID NO: 62)

10 Equivalent amounts of top and bottom pairs of tags and 3' end oligonucleotides were dissolved in water and annealed by heating to 85°C and ramping down to 4°C in 200 mM NaCl, 50 mM Tris HCl, pH 7.0, buffer.

First, the double-stranded A1 tag was ligated to the headpiece. The ligation reaction (20 µl) contained 2.5 µM of HP-1-DTAF-biotin and 2.5 µM of double-
15 stranded A1 tag in 1x T4 DNA ligase buffer and 60 Weiss units of T4 DNA ligase (New England BioLabs). The reaction was incubated at 16°C for 16 hours. The resulting product did not resolve on any of the tested gels, including different percentages of TBE-urea, NativePage, SDS-PAGE, or 2% and 4% agarose E-gel (Invitrogen, Inc.). Mobility of the oligonucleotide, modified with PEG linker and
20 DTAF-biotin, was mostly determined by the presence of these groups rather than by the DNA itself (data not shown). To test the efficiency of the ligation, we ligated all tags and 3' end oligonucleotides and performed PCR assays of the resulting construct to confirm the ligation efficiency. The ligation reaction (70 µl) contained: 2.5 µM of HP-1-DTAF-biotin; 2.5 µM of each of the annealed double-stranded DNA tags (A1, B1, and C1), as well as the 3' end tag; 1x T4 DNA ligase buffer; and 210 Weiss units
25 of T4 DNA ligase. The reaction was incubated at 16°C for 20 hours.

The reaction mixture was loaded on a 4% agarose gel and the fluorescent band was extracted from the gel. This material was used for the test 24 cycle PCR amplification using primers 5' and 3' as described above. The results are summarized
30 in Fig. 13.

Step 4: Purification of HP-1-DTAF-biotin on streptavidin beads and reaction with streptavidin

Purification of HP-1-DTAF-biotin on streptavidin (SA) Dynal magnetic beads M-280 (Invitrogen) serves as a model for affinity selection for the chemical DNA-
5 tagged library. SA beads were pre-equilibrated in 2x PBS buffer containing 0.05% Triton X-100. 50 pmol of HP-1-DTAF-biotin were loaded on 25 μ l of the pre-washed SA beads for 15 minutes at room temperature with tumbling. The flow-through was collected and the beads were washed 3 times for 30 minutes with 1 ml of the same buffer. A final wash was performed at 80°C for 10 minutes with 30 μ l water
10 (collected). The beads were eluted with 30 μ l of 25 mM EDTA and 5 mM NaOH for 10 minutes at 90°C, and the eluent was immediately neutralized by adding 3 μ l of 1 M Tris HCl, pH 7.0.

For the streptavidin binding experiment, 5 μ l of the elution samples were incubated with an excess of streptavidin in 50 mM NaCl/10 mM Tris HCl, pH 7.0, for
15 10 minutes. The samples were mixed with gel-loading buffer without boiling and resolved on a 4-12% SDS NuPage gel (Invitrogen) using MES running buffer. The results are summarized in Fig. 14.

Example 4. Coupling of H(-Arg- ϵ Ahx)₆-Arg-OH peptide to HP-1-DTAF

20 We have chosen an arginine-rich peptide R7, H(-Arg- ϵ Ahx)₆-Arg-OH (Bachem), to use as another modification for the last reactive group on the triazine scaffold. This is an arginine-aminohexanoic acid cell membrane permeable peptide used for intracellular compound delivery. The reaction was set up similar to the reaction conditions described above: 20 μ l reaction contained around 200 pmol of HP-
25 1-DTAF (Step 1) dissolved in 150 mM borate buffer, pH 9.5, and 10 nmol of R7 peptide. Under these conditions, the side chains of the arginines do not react, and the only reactive amine in the peptide is the N-terminus. The reaction was allowed to proceed for 12 hours at 75°C and was then purified by desalting on a Zeba spin column.

30

Example 5. DNA construct for intracellular T7 RNAP delivery detection experiment

The DNA construct used for the chemical “library of 1” intracellular delivery experiment was prepared from a PCR product of a V_H DNA single clone of ~400 bp featuring a T7 promoter region at the 5’ end and a short antibody constant C_{mu} region close to the 3’ end of the molecule. In order to link the DNA construct to the modified headpiece of the model chemical library, a BsmI restriction site was appended upstream of the T7 promoter region by PCR amplification of the clone. BsmI restriction digest produced a 3’ GG overhang, which allowed ligation to the headpiece (3’ CC overhang). The 5’ primer with BsmI site (underlined) was synthesized by IDT DNA, Inc.

5’-GGATGCCCGAATGCCTAATACGACTCACTATAGGG-
ACAATTACTATTTACAATTACA (SEQ ID NO: 63)

Following PCR amplification, the DNA construct was purified using a PCR purification kit (Invitrogen), and the resulting DNA was digested with 250 U BsmI (New England BioLabs) at 65°C in NEB buffer 4 for 2 hours. The DNA was purified on a 2% agarose gel. The ligation reaction (30 µl) contained 2 pmol of each V_H DNA construct, digested with BsmI, as well as HP-1-DTAF-R7 (arginine-aminohexanoic acid peptide) in 1x T4 DNA ligase buffer and 60 Weiss units of T4 DNA ligase (New England BioLabs). The reaction was incubated at 16°C for 20 hours. Due to high efficiency of the ligation, the material was further used for the intracellular delivery/T7 RNAP experiment without further purification. The results are summarized in Fig. 15.

Example 6. Synthesis of 10x10 Library

Step 1. Ligation of the tag A to the headpiece HP-T

In this exemplary library, only positions B and C are used. One tag A is ligated to HP-T. The tag has the following sequence:

DNA Tag A1 (top): 5’-phos-GGAGGACTGT (SEQ ID NO: 64)

DNA Tag A1 (bottom): 5’-phos-AGTCCTCCGG (SEQ ID NO: 65)

30 nmol of HP-T were mixed with 45 nmol of each Tag A1 top and Tag A1 bottom oligos in 1x T4 DNA ligase buffer and were annealed by heating to 95°C for 1 minute, followed by cooling to 4°C at 0.2°C/second. The sample was then brought to 16°C. 300 Weiss Units of T4 DNA ligase was added and the samples were allowed to

incubate for 16-20 hours at 16°C. Following the ligation, HP-T-A was desalted using a Zeba column (Pierce). See, e.g., Fig. 16A.

Step 2. Ligation of tags B1-B12 and C tags

5 Twelve ligation reactions were set up similar to the ligation reactions described above. In each of 12 tubes, 5 nmol pairs of B1-B12 top and bottom oligos were added to 1x T4 DNA ligase buffer and annealed as described above. HP-T-A was dissolved in 1x T4 DNA ligase buffer. 2.5 nmol of HP-T-A were aliquoted in these 12 tubes. 30 Weiss units of T4 DNA ligase were added to each tube and
10 reactions were allowed to proceed for 20 hours at 16°C. Following the incubation, each reaction mixture was individually desalted on a 0.5 ml Zeba spin column, equilibrated with 150 mM borate buffer, pH 9.0. To each tube, a 20x excess of cyanouric chloride (50 nmol), dissolved in acetonitrile, was added and incubated for 1.5 hours at 4°C. Following this incubation, a 100x excess (250 nmol, i.e., 5x excess
15 relative to cyanouric chloride) of amines B1-B12, dissolved in acetonitrile or DMF, was added in correspondence with the ligated B1-B12 tags. The reaction with amines was allowed to proceed for 20 hours at 4°C. Following this reaction the library was pooled, desalted twice on 2-ml Zeba columns and lyophilized. See, e.g., Figs. 16B and 16C.

20 Like the reactions above, the C tags and amines are added using similar reaction conditions to those described above.

Other Embodiments

25 All publications, patents, and patent applications mentioned in the above specification are hereby incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it should be understood that the invention as claimed should not be unduly limited to
30 such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention.

Other embodiments are in the claims.

What is claimed is:

Claims

1. A method of tagging DNA-encoded chemical libraries, said method comprising binding a first functional group of a bifunctional linker to an initiator
5 oligonucleotide at the 5' end of said initiator oligonucleotide, wherein said initiator oligonucleotide bound to said bifunctional linker forms a hairpin structure, and binding a second functional group of said bifunctional linker to a component of said chemical library.
- 10 2. The method of claim 1, wherein said initiator oligonucleotide comprises a first identifier region.
3. The method of claim 2, wherein said initiator oligonucleotide comprises a second identifier region hybridizes to said first identifier region of the initiator
15 oligonucleotide.
4. The method of claim 3, wherein the second identifier region comprises a fluorescent tag or biotin label.
- 20 5. The method of claim 4, wherein said second identifier region is not amplified prior to analysis following a selection step.
6. The method of any one of claims 1-5, wherein said bifunctional linker, initiator oligonucleotide, first identifier region, or second identifier region is modified
25 to increase solubility of a member of said DNA-encoded chemical library in organic conditions.
7. A method of creating DNA-encoded libraries, said method comprising:
30 (a) creating a first diversity node;
(b) encoding said first diversity node in separate vessels;
(c) pooling said first diversity node; and

(d) splitting said pooled first diversity node into a second set of separate vessels, wherein said first diversity node reacts to form a second diversity node.

5 8. The method of claim 7, wherein the second diversity node is not encoded and pooled.