SUMMARY

The present disclosure 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 disclosure 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 one aspect, the present disclosure refers to a method of tagging DNA-encoded chemical libraries, said method comprising:

(a) binding a first functional group of a bifunctional linker to a headpiece comprising an initiator oligonucleotide at the 5'-end of said initiator oligonucleotide, wherein said initiator oligonucleotide comprises a hairpin structure;

(b) binding a second functional group of the bifunctional linker to a first building block;

(c) reacting a second building block with said first building block; and

(d) binding an identifier region at the 3' end of said headpiece, wherein the identifier region encodes said second building block,

wherein the chemical reactions of the method, including step (c), are performed in organic solvent.

In another example, the disclosure 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 examples, the second diversity node is not encoded and pooled.

In another example, the present disclosure 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 disclosure 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 examples, 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 examples, 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

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 disclosure 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.