

METHODS FOR IDENTIFYING COMPOUNDS BINDING TO A TARGET MOLECULE USING MIRROR-IMAGING OF THE TARGET MOLECULE

Sequence Listing

5 The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on August 23, 2023, is named "50719-070001_Sequence_Listing_8_23_23.xml" and is 3,401 bytes in size.

Background

10 DNA-encoded combinatorial libraries offer many benefits for drug discovery. These libraries can provide a large number of diverse compounds that can be rapidly screened and interrogated. To further increase complexity, various steps of the discovery process can be programmed and automated. These steps include the use of multi-step, split-and-pool synthesis to add building blocks to atomic or polyatomic scaffolds and the use of enzymatic and/or chemical ligation to add DNA tags that encode both the
15 synthetic steps and the building blocks. Despite these and other advantages, improved methods to screen DNA-encoded libraries are desirable.

Summary of the Invention

 The present disclosure provides methods for identifying compounds useful as therapeutic agents
20 and/or useful as starting points for optimization in the development of therapeutic agents. These methods are particularly useful for identifying binding compounds for nucleic acid and nucleic acid-binding protein targets.

 In one aspect, the disclosure provides a method for identifying a compound that binds to a target molecule, the method comprising the steps of:

- 25 (i) providing a set of candidate compounds, wherein each candidate compound comprises a DNA tag encoding the identity of the candidate compound;
- (ii) contacting the set of candidate compounds with a molecule that is the mirror-image of the target molecule; and
- (iii) selecting a binding compound from the set of candidate compounds that binds to the
30 mirror-image of the target molecule, thereby identifying a compound that binds to the target molecule.

 In some embodiments, the method further comprises generating the mirror-image of the binding compound, and thereby identifying a further compound that binds to the target molecule.

 In some embodiments, the method further comprises validating the compound that binds to the target molecule.

35 In some embodiments, validating the compound that binds to the target molecule comprises binding a molecule that is a mirror-image of the binding compound to a non-mirrored target molecule.

 In some embodiments, the target molecule binds to the DNA tag in its naturally-occurring stereochemical form.

 In some embodiments, the mirror-image of the target molecule does not bind to the DNA tag.

40

In some embodiments, the target molecule is a complex comprising a protein and/or a nucleic acid. In some embodiments, the complex comprises more than one protein and/or more than one nucleic acid. In some embodiments, (i) the protein in the complex comprises L-amino acid residues, and (ii) the nucleic acid in the complex comprises D-nucleotides. In some embodiments, the mirror-image of the target molecule is produced by chemical synthesis.

In some embodiments, the target molecule is a protein. In some embodiments, the protein comprises L-amino acid residues. In some embodiments, the protein is a nucleic acid-binding protein. In some embodiments, the nucleic acid-binding protein is a DNA-binding protein. In some embodiments, the nucleic acid-binding protein is an RNA-binding protein. In some embodiments, the mirror-image of the protein is produced by chemical synthesis.

In some embodiments, the target molecule is a nucleic acid.

In some embodiments, the nucleic acid is a DNA. In some embodiments, the DNA is a protein-coding DNA, RNA-coding DNA, exonic DNA, intronic DNA, regulatory DNA (e.g., promoter DNA, enhancer DNA, activator DNA, repressor DNA), pseudogenes, transposons, chromosomal DNA, mitochondrial DNA, centromeric DNA, telomeric DNA, satellite DNA, scaffold DNA, repetitive DNA, DNA encoding expanded repeats, modified DNA, methylated DNA, A-DNA, B-DNA, Z-DNA, single-stranded DNA, or a double-stranded DNA.

In some embodiments, the nucleic acid is a D-DNA. In some embodiments, the mirror-image of the target molecule is an L-DNA.

In some embodiments, the nucleic acid is an RNA. In some embodiments, the RNA is an antisense RNA, circular RNA, long non-coding RNA, microRNA, messenger RNA, PIWI-interacting RNA, ribosomal RNA, small conditional RNA, small nucleolar RNA, small nuclear RNA, transfer RNA, Y RNA, MALAT1-associated small cytoplasmic RNA, mitochondrial RNA, small interfering RNA, guide RNA, CRISPR RNA, trans-activating CRISPR RNA, single guide RNA, short hairpin RNA, enhancer RNA, vault RNA, or an RNA encoding expanded repeats.

In some embodiments, the nucleic acid is a D-RNA. In some embodiments, the mirror-image of the target molecule is an L-RNA.

In some embodiments, the mirror-image of the nucleic acid is produced by an enzymatic reaction.

In some embodiments, the mirror-image of the nucleic acid is produced by chemical synthesis.

In some embodiments, the candidate compounds are small molecules. In some embodiments, the small molecules are optically active compounds.

In some embodiments, the candidate compounds are in stereochemically pure form. In some embodiments, the candidate compounds are in the form of a mixture of stereoisomers. In some embodiments, the candidate compounds are in the form of a racemic mixture.

In some embodiments, the selecting step involves affinity-mediated selection. In some embodiments, the selecting step comprises enriching the set of candidate compounds for the binding compound.

In some embodiments, the selecting step comprises sequencing the DNA tag to identify the binding compound.

In some embodiments, the set of candidate compounds comprises at least 250,000 different compounds. In some embodiments, the set of candidate compounds comprises at least two million

different compounds. In some embodiments, the set of candidate compounds comprises at least five million different compounds. In some embodiments, the set of candidate compounds comprises at least ten million different compounds. In some embodiments, the set of candidate compounds comprises at least twenty-five million different compounds.

- 5 In some embodiments, the compound that binds to the target molecule is a chiral molecule. In some embodiments, the compound that binds to the target molecule is not a chiral molecule.

Definitions

- 10 The term “enrichment,” as used herein refers to the process of increasing the relative concentration of a particular compound within a mixture or library. Preferably, as used herein, a compound is enriched in a set of candidate compounds, by the claimed selection method, by at least 2-fold, at least 5-fold, at least 10-fold, at least 20-fold, at least 50-fold, at least 100-fold, at least 200-fold, at least 500-fold, at least 1,000-fold, at least 2,000-fold, or at least 5,000-fold. In some embodiments, a
15 compound is enriched in a set of candidate compounds, by the claimed selection method, by between 2-fold and 10,000-fold, between 2-fold and 5,000-fold, between 10-fold and 5,000-fold, between 20-fold and 5,000-fold, between 50-fold and 5,000-fold, between 100-fold and 5,000-fold, between 200-fold and 5,000-fold, between 500-fold and 5,000-fold, between 1,000-fold and 5,000-fold, or between 2,000-fold and 5,000-fold.

- 20 The term “mirror-image” as used herein, refers to an isomer that is in a mirror-image relationship with the natural material in chirality.

- The term “aptamer,” as used herein refers to an oligonucleotide that is capable of forming a complex with an intended target molecule. The complexation is target-specific in the sense that other materials which may accompany the target do not complex to the aptamer. It is recognized that
25 complexation and affinity are a matter of degree. However, in this context, “target-specific” means that the aptamer binds to the target molecule with a much higher degree of affinity than it binds to contaminating materials.

- The affinity of a molecule X for its partner Y can generally be represented by the dissociation constant (K_D). Affinity can be measured by common methods known in the art, including those described
30 herein. The term “ K_D ,” as used herein, is intended to refer to the dissociation equilibrium constant of a particular compound-protein or complex-protein interaction. Typically, the compounds of the invention bind to targets with a dissociation equilibrium constant (K_D) of less than about 10^{-6} M, such as less than approximately 10^{-7} M, 10^{-8} M, 10^{-9} M, or 10^{-10} M or even lower, e.g., when determined by surface plasmon resonance (SPR).

- 35 As used herein, the term “set” or “library” refers to a group of 2, 5, 10, 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , 10^{11} , 10^{12} or more different molecules. In some embodiments, at least 10% (e.g., at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99% or 100%) of the compounds in the library are compounds including a DNA tag encoding their identity.

- 40 As used herein the term “DNA tag” refers to a molecule comprising deoxyribonucleic acid bases. In general, a “DNA tag” comprises adenine, guanine, thymine, and/or cytosine bases. In some

embodiments, a “DNA tag” comprises unnatural deoxyribonucleic acid bases or unnatural deoxyribonucleotides, for example, d5SICS or dNaM. In some embodiments, a “DNA tag” comprises modified deoxyribonucleic acid bases. In general, a “DNA tag” can be amplified. In some embodiments, a “DNA tag” is single-stranded. In some embodiments, a “DNA tag” is double-stranded.

5 The term “selective” when used with reference to a compound having an activity is understood by those skilled in the art to mean that the compound discriminates between potential targets. For example, in some embodiments, a compound is said to bind “selectively” to its target if it binds preferentially with that target in the presence of one or more competing alternative targets. In many embodiments, selective interaction is dependent upon the presence of a particular structural feature of the target entity (e.g., an epitope, a cleft, a binding site). It is to be understood that selectivity need not be absolute. In some
10 embodiments, selectivity may be evaluated relative to that of the binding agent for one or more other potential target entities (e.g., competitors). In some embodiments, selectivity is evaluated relative to that of a reference selective binding agent. In some embodiments, selectivity is evaluated relative to that of a reference non-selective binding agent. In some embodiments, the agent or entity does not detectably
15 bind to the competing alternative target under conditions of binding to its target entity. In some embodiments, binding agent binds with higher on-rate, lower off-rate, increased affinity, decreased dissociation, and/or increased stability to its target entity as compared with the competing alternative target(s).

 The term “small molecule” means a low molecular weight organic and/or inorganic compound. In
20 general, a “small molecule” is a molecule that is less than about 5 kilodaltons (kD) in size. In some embodiments, a small molecule is less than about 4 kD, 3 kD, about 2 kD, or about 1 kD. In some embodiments, the small molecule is less than about 800 Daltons (D), about 600 D, about 500 D, about 400 D, about 300 D, about 200 D, or about 100 D. In some embodiments, a small molecule is less than about 2000 g/mol, less than about 1500 g/mol, less than about 1000 g/mol, less than about 800 g/mol, or
25 less than about 500 g/mol. In some embodiments, a small molecule is not a polymer. In some embodiments, a small molecule does not include a polymeric moiety. In some embodiments, a small molecule is not a protein or polypeptide (e.g., is not an oligopeptide or peptide). In some embodiments, a small molecule is not a polynucleotide (e.g., is not an oligonucleotide). In some embodiments, a small molecule is not a polysaccharide. In some embodiments, a small molecule does not comprise a
30 polysaccharide (e.g., is not a glycoprotein, proteoglycan, glycolipid, *etc.*). In some embodiments, a small molecule is not a lipid. In some embodiments, a small molecule is a modulating compound. In some embodiments, a small molecule is biologically active. In some embodiments, a small molecule is detectable (e.g., comprises at least one detectable moiety). In some embodiments, a small molecule is a therapeutic molecule.

35 Those of ordinary skill in the art, reading the present disclosure, will appreciate that certain small molecule compounds described herein may be provided and/or utilized in any of a variety of forms such as, for example, salt forms, protected forms, pro-drug forms, ester forms, isomeric forms (e.g., optical and/or structural isomers), isotopic forms, etc. In some embodiments, reference to a particular compound may relate to a specific form of that compound. In some embodiments, reference to a particular
40 compound may relate to that compound in any form. In some embodiments, where a compound is one that exists or is found in nature, that compound may be provided and/or utilized in accordance in the

present invention in a form different from that in which it exists or is found in nature. Those of ordinary skill in the art will appreciate that a compound preparation including a different level, amount, or ratio of one or more individual forms than a reference preparation or source (e.g., a natural source) of the compound may be considered to be a different form of the compound as described herein. Thus, in some
5 embodiments, for example, a preparation of a single stereoisomer of a compound may be considered to be a different form of the compound than a racemic mixture of the compound; a particular salt of a compound may be considered to be a different form from another salt form of the compound; a preparation containing one conformational isomer ((Z) or (E)) of a double bond may be considered to be a different form from one containing the other conformational isomer ((E) or (Z)) of the double bond; a
10 preparation in which one or more atoms is a different isotope than is present in a reference preparation may be considered to be a different form; etc.

As used herein, the terms “specific binding” or “specific for” or “specific to” refer to an interaction between a binding agent and a target entity. As will be understood by those of ordinary skill, an interaction is considered to be “specific” if it is favored in the presence of alternative interactions, for
15 example, binding with a K_D of less than 10 μM (e.g., less than 5 μM , less than 1 μM , less than 500 nM, less than 200 nM, less than 100 nM, less than 75 nM, less than 50 nM, less than 25 nM, less than 10 nM or 10 nM to 100 nM, 50 nM to 250 nM, 100 nM to 500 nM, 250 nM to 1 μM , 500 nM to 2 μM , 1 μM to 5 μM). In many embodiments, specific interaction is dependent upon the presence of a particular structural feature of the target entity (e.g., an epitope, a cleft, a binding site). It is to be understood that specificity
20 need not be absolute. In some embodiments, specificity may be evaluated relative to that of the binding agent for one or more other potential target entities (e.g., competitors). In some embodiments, specificity is evaluated relative to that of a reference specific binding agent. In some embodiments, specificity is evaluated relative to that of a reference non-specific binding agent.

The term “substantially” refers to the qualitative condition of exhibiting total or near-total extent or
25 degree of a characteristic or property of interest. One of ordinary skill in the biological arts will understand that biological and chemical phenomena rarely, if ever, go to completion and/or proceed to completeness or achieve or avoid an absolute result. The term “substantially” is therefore used herein to capture the potential lack of completeness inherent in many biological and chemical phenomena.

The term “target molecule” refers to a molecule (e.g., DNA, RNA, or protein) that binds with a
30 small molecule. In some embodiments, the target protein participates in a biological pathway associated with a disease, disorder or condition. In some embodiments, a target protein is a naturally-occurring protein. In some such embodiments, a target protein is naturally found in certain mammalian cells (e.g., a mammalian target protein), fungal cells (e.g., a fungal target protein), bacterial cells (e.g., a bacterial target protein), or plant cells (e.g., a plant target protein). Target proteins can be naturally occurring, e.g.,
35 wild type. Alternatively, the target protein can vary from the wild-type protein but still retain biological function, e.g., as an allelic variant, a splice mutant or a biologically active fragment. Exemplary mammalian target proteins are GTPases, GTPase activating protein, Guanine nucleotide-exchange factor, heat shock proteins, ion channels, coiled-coil proteins, kinases, phosphatases, ubiquitin ligases, transcription factors, chromatin modifier/remodelers, proteins with classical protein-protein interaction
40 domains and motifs, or any other proteins that participate in a biological pathway associated with a disease, disorder or condition. In some embodiments, the target molecule is a nucleic acid-binding

protein. In some embodiments, the nucleic acid-binding protein is a DNA-binding protein encoded by a gene selected from the group consisting of ALX3, ALX4, AR, ARF4, ARNT, ARNTL, ARX, ASCL1, ASH1L, ATF2, ATF3, ATF4, ATF7, ATRX, BACH1, BACH2, BARHL2, BARX, BATF, BATF3, BCL6, BCL6B, BHLHA15, BHLHB2, BHLHB3, BHLHE22, BHLHE23, BHLHE40, BHLHE41, BRG1, BRM, BSX, CART1, CBP, CDX1, CDX2, CEBPA, CEBPB, CEBPD, CEBPG, CENPB, CLOCK, CPEB1, CREB1, CREB3, CREB3L1, CREBBP, CREM, CTCF, CTCFL, CUX1, CUX2, DBP, DLX1, DLX2, DLX3, DLX4, DLX5, DLX6, DMBX1, DPRX, DRGX, DUX4, DUXA, E2F1, E2F2, E2F3, E2F4, E2F6, E2F7, E2F8, EBF1, EBF3, EGR1, EGR2, EGR3, EGR4, EHF, ELF1, ELF3, ELF4, ELF5, ELK1, ELK3, ELK4, EMX1, EMX2, EN1, EN2, EOMES, EP300, EPAS1, ERF, ERG, ESR1, ESR2, ESRRA, ESRRB, ESRRG, ESX1, ETS1, ETV1, ETV2, ETV3, ETV4, ETV5, ETV6, EVX1, EVX2, FEV, FIGLA, FLI1, FOS, FOSL1, FOSL2, FOXA1, FOXA2, FOXA3, FOXB1, FOXC1, FOXC2, FOXD2, FOXD3, FOXG1, FOXI1, FOXJ2, FOXJ3, FOXK1, FOXK2, FOXL1, FOXO1, FOXO3, FOXO4, FOXO6, FOXP1, FOXP3, GABPA, GATA1, GATA2, GATA3, GATA4, GATA5, GATA6, GAVPA, GBX1, GBX2, GCM1, GCM2, GFI1, GFI1B, GLI2, GLIS1, GLIS2, GLIS3, GMEB2, GRHL1, GRHL2, GSC, GSC2, GSX1, GSX2, HAND2, HAS5, HES7, HESX1, HEY1, HEY2, HIC1, HIC2, HIF1A, HINFP, HLF, HMBOX1, HMX1, HMX2, HMX3, HNF1A, HNF1B, HNF4A, HNF4G, HOMEZ, HOXA1, HOXA10, HOXA13, HOXA2, HOXA9, HOXAB13, HOXB13, HOXB2, HOXB3, HOXB5, HOXC10, HOXC11, HOXC12, HOXC13, HOXD11, HOXD12, HOXD13, HOXD8, HSF1, HSF2, HSF4, HSFY2, ID4, IRF1, IRF2, IRF3, IRF4, IRF5, IRF7, IRF8, IRF9, IRX2, IRX5, ISL1, ISL2, ISX, JDP2, JUN, JUNB, JUND, KAT2A, KAT6A, KAT6B, KLF1, KLF11, KLF12, KLF13, KLF14, KLF15, KLF16, KLF4, KLF5, KLF9, LBX2, LHX2, LHX6, LHX9, LEF1, LHX2, LHX9, LMX1A, LMX1B, MAF, MAFB, MAFF, MAFG, MAFK, MAX, MECOM, MEF2A, MEF2B, MEF2C, MEF2D, MEIS1, MEIS2, MEIS3, MEOX1, MEOX2, MESP1, MGA, MITF, MIXL1, MLL, MLX, MLXIPL, MNT, MNX1, MSC, MSX1, MSX2, MTF1, MXI1, MYB, MYBL1, MYBL2, MYC, MYCL, MYCN, MYF5, MYF6, MYOD1, MYOG, NEUROD1, NEUROD2, NEUROG2, NFAT5, NFATC1, NFE2, NFE2L2, NFIA, NFIB, NFIC, NFIL3, NFIX, NFKB1, NFKB2, NFYA, NFYB, NFYC, NHLH1, NKX2-3, NKX2-5, NKX2-8, NKX3-1, NKX3-2, NKX6-1, NKX6-2, NOTO, NR1H2, NR1H3, NR1H4, NR2C1, NR2C2, NR2E1, NR2F1, NR2F2, NR2F6, NR3C1, NR3C2, NR4A1, NR4A2, NR5A1, NR5A2, NRF1, NRL, NSD1, OLIG1, OLIG2, OLIG3, ONECUT1, ONECUT2, ONECUT3, OSR2, OTX1, OTX2, PAX1, PAX2, PAX3, PAX4, PAX5, PAX6, PAX7, PAX9, PBX1, PBX2, PBX3, PDX1, PHOX2A, PHOX2B, PITX1, PITX3, PKNX1, PLAG1, POU1F1, POU2F1, POU2F2, POU2F3, POU3F1, POU3F2, POU3F3, POU3F4, POU4F1, POU4F2, POU4F3, POU5F1, POU5F1P1, POU6F2, PPARG, PRDM1, PRDM3, PRDM4, PROP1, PROX1, PRRX1, PRRX2, RARA, RARB, RARG, RAX, RAXL1, RBPJ, REL, RELA, RELB, REST, RFX1, RFX2, RFX3, RFX4, RFX5, RHOXF1, RORA, RUNX1, RUNX2, RUNX3, RXRA, RXRB, RXRG, SCRT1, SCRT2, SETD2, SHOX, SHOX2, SIX1, SIX2, SMAD2, SMAD3, SMAD4, SNAI2, SOX10, SOX13, SOX14, SOX15, SOX17, SOX18, SOX2, SOX21, SOX4, SOX6, SOX7, SOX8, SOX9, SP1, SP2, SP3, SP4, SP8, SPDEF, SPI1, SPIB, SPIC, SREBF1, SREBF2, SRF, SRY, STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, STAT6, T, TAL1, TBR1, TBX1, TBX15, TBX19, TBX2, TBX20, TBX21, TBX4, TBX5, TCF12, TCF3, TCF4, TCF7, TCF7L1, TCF7L2, TEAD1, TEAD3, TEAD4, TEF, TFAP2A, TFAP2B, TFAP2C, TFAP4, TFAP2, TFDP1, TFE3, TFEB, TFEC, TGIF1, TGIF2, TGIF2LX, THAP11, THRA, THRB, TP53, TP63, TP73, TWIST1, UNCX, USF1, USF2, VAX1, VAX2, VENTX, VDR, VSX1, VSX2, WT1, XBP1, YY1, YY2, ZBED1, ZBTB7A, ZBTB7B, ZBTB7C, ZFX, ZIC1, ZIC3, ZIC4, ZNF143, ZNF232, ZNF238, ZNF263, ZNF282, ZNF306, ZNF410,

ZNF435, ZBTB49, ZNF524, ZNF713, ZNF740, ZNF75A, ZNF784, and ZSCAN4. In some embodiments, the nucleic acid-binding protein is an RNA-binding protein encoded by a gene selected from the group consisting of A1CF, A2BP1, ACO1, AKAP1, ANKRD17, BRUNOL4, BRUNOL5, BRUNOL6, C14orf156, CIRBP, CPEB1, CPEB2, CPEB3, CPEB4, CSDA, CSDE1, CSTF2, CSTF2T, CUGBP1, CUGBP2, DAZAP1, DDX43, DPPA5, EIF3G, EIF4B, ELAVL1, ELAVL2, ELAVL3, ELAVL4, ESRP1, ESRP2, EWSR1, FMR1, FUBP1, FUBP3, FUS, FXR1, FXR2, G3BP1, GAPDH, GRSF1, HNRNPA1, HNRNPA2B1, HNRNPA3, HNRNPAB, HNRNPC, HNRNPCL1, HNRNPD, HNRNPF, HNRNPH1, HNRNPH2, HNRNPH3, HNRNPK, HNRNPL, HNRNPM, HNRNPR, HNRPDL, HNRPLL, IGF2BP1, IGF2BP2, IGF2BP3, IREB2, KHDRBS1, KHDRBS2, KHDRBS3, KHSRP, KIN, LARP1, LARP1B, LARP6, LARP7, LIN28, LIN28B, LSM11, MBNL1, MBNL2, MBNL3, MEX3A, MEX3B, MEX3C, MEX3D, MIR1236, MKRN1, MKRN2, MKRN3, MSI1, MYEF2, NCL, NONO, NOVA1, NOVA2, NUFIP1, PABPC1, PABPC1L, PABPC3, PABPC4, PABPN1, PARN, PCBP1, PCBP2, PCBP3, PCBP4, PNPT1, PPIE, PPP1R10, PSPC1, PTBP1, PTBP2, PUF60, PUM1, PUM2, QKI, RALY, RALYL, RBM10, RBM12B, RBM15, RBM15B, RBM16, RBM26, RBM27, RBM3, RBM38, RBM39, RBM4, RBM5, RBM9, RBMS1, RBMS2, RBMS3, RBMX, RBMXL2, RBMY1A1, RC3H1, RC3H2, RDM1, RNF113A, RNF113B, ROD1, SAFB, SAFB2, SART3, SF1, SFPQ, SFRS1, SFRS11, SFRS12, SFRS13A, SFRS15, SFRS2, SFRS3, SFRS4, SFRS5, SFRS6, SFRS7, SFRS9, SLBP, SLTM, SNRNP70, SNRPA, SNRPB, SNRPB2, SNRPN, SRRM1, SSB, SYNCRIP, TAF15, TARDBP, THOC4, TIA1, TIAL1, TNRC4, TNRC6A, TRA2A, TRA2B, TUT1, U2AF1, U2AF2, YBX1, YBX2, YTHDC1, YTHDF1, YTHDF2, YTHDF3, ZC3H12A, ZC3H12B, ZC3H12C, ZC3H14, ZC3H4, ZC3H6, ZC3H7B, ZFP36, ZFP36L1, ZFP36L2, ZFR, ZNF239, ZNF74, ZRANB2, ZRSR1, and ZRSR2. In some embodiments, the target molecule is a nucleic acid. In some embodiments, the nucleic acid is a DNA, for example, a DNA selected from the group consisting of protein-coding DNA, RNA-coding DNA, exonic DNA, intronic DNA, regulatory DNA (e.g., promoter DNA, enhancer DNA, activator DNA, repressor DNA), pseudogenes, transposons, chromosomal DNA, mitochondrial DNA, centromeric DNA, telomeric DNA, satellite DNA, scaffold DNA, repetitive DNA, DNA encoding expanded repeats, modified DNA, methylated DNA, A-DNA, B-DNA, Z-DNA, single-stranded DNA, and double-stranded DNA. In some embodiments, the nucleic acid is an RNA, for example, an RNA selected from the group consisting of antisense RNA, circular RNA, long non-coding RNA, microRNA, messenger RNA, PIWI-interacting RNA, ribosomal RNA, small conditional RNA, small nucleolar RNA, small nuclear RNA, transfer RNA, Y RNA, MALAT1-associated small cytoplasmic RNA, mitochondrial RNA, small interfering RNA, guide RNA, CRISPR RNA, trans-activating CRISPR RNA, single guide RNA, short hairpin RNA, enhancer RNA, vault RNA, RNA encoding expanded repeats, exonic RNA, intronic RNA, non-coding RNA, retrotransposon RNA, pre-mRNA, post-transcriptionally edited RNA, single-stranded RNA, and double-stranded RNA.

35

Brief Description of the Drawings

FIG. 1 is a schematic illustrating selection of a DNA-encoded library against the mirror-image of a target molecule. DEL, DNA-encoded library.

FIGS. 2A and 2B are charts showing fluorescence polarization in millipolarization units (mP) of sulfo-Cy5 conjugates of two compounds, identified by affinity selection, when incubated with increasing concentrations of (GGGGCC)₈ RNA. **FIG. 2A** shows the fluorescence polarization of compound

40

XCMPD026217 when incubated with increasing concentrations of (GGGGCC)₈ RNA. FIG. 2B shows the fluorescence polarization of compound XCMPD026218 when incubated with increasing concentrations of (GGGGCC)₈ RNA.

5

Detailed Description of the Invention

The present disclosure provides methods for identifying compounds useful as therapeutic agents and/or useful as starting points for optimization in the development of therapeutic agents. These methods utilize a mirror-image of the target molecule to select for compounds that bind to the target molecule, thereby enhancing the efficiency of the selection. This method can be combined with any other DNA-encoded library approach, including, without limitation, International Patent Application Publication Nos. WO 2010/094036 A1, WO 2013/036810 A1, WO 2014/012010 A1, WO 2016/109423 A1, WO 2018/195134 A1, and WO 2021/016525 A1, each of which is incorporated by reference in its entirety.

The present disclosure features, for example, a method for identifying a compound that binds to a target molecule, the method comprising the steps of:

- (i) providing a set of candidate compounds, wherein each candidate compound comprises a DNA tag encoding the identity of the candidate compound;
- (ii) contacting the set of candidate compounds with a molecule that is the mirror-image of the target molecule; and
- (iii) selecting a binding compound from the set of candidate compounds that binds to the mirror-image of the target molecule, thereby identifying a compound that binds to the target molecule.

When screening DNA-encoded chemical libraries for candidate compounds that bind to a target molecule, it is important to suppress DNA-mediated binding between the target molecule and the DNA tag attached to the candidate compounds, especially when the target molecule is capable of binding to DNA (e.g., a DNA-binding protein). The present disclosure provides an improved technique to screen DNA-encoded chemical libraries with targets that bind to DNA. The technique minimizes DNA-mediated binding between the target molecule and the DNA tag attached to the candidate compounds, thereby improving the efficiency of the screening.

Encoded Compounds

This invention features methods utilizing encoded chemical entities including a chemical entity, one or more tags, and a headpiece operatively associated with the first chemical entity and one or more tags. The chemical entities, headpieces, tags, linkages, and bifunctional spacers are further described below.

35

Chemical entities

The encoded compounds (e.g., small molecules) utilized in the methods 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; an oxadiazole; a pyrazole; an isoxazole; a pyran; a pyridine; an indole; an indazole; or a purine.

5 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, 10 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 15 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 20 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. Scaffolds can also be constructed during the library synthesis process from atoms present in the building blocks thereby reducing the number and/or proportion of invariant atoms present within a scheme.

25 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.

30 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 C_{1-6} alkyl groups or 35 optionally substituted C_{1-6} 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 ($-NR_2$, where 40 each R is, independently, H or an optionally substituted C_{1-6} alkyl), hydroxy, alkoxy ($-OR$, where R is an optionally substituted C_{1-6} 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 polymerase chain reaction (PCR) and restriction digest with one of the corresponding restriction enzymes.

5 *Headpiece*

In an encoded chemical entity, the headpiece operatively links each chemical entity to its encoding oligonucleotide tag. Generally, the headpiece is a starting oligonucleotide having at least two functional groups that can be further derivatized, where the first functional group operatively links the first chemical entity (or a component thereof) to the headpiece and the second functional group operatively
10 links one or more tags to the headpiece. A bifunctional spacer can optionally be used as a spacing moiety between the headpiece and a chemical entity.

The functional groups of the headpiece can be used to form a covalent bond with a component of a chemical entity and another covalent bond with a tag. The component can be any part of a small molecule, such as a scaffold having diversity nodes or a building block. Alternatively, the headpiece can
15 be derivatized to provide a spacer (e.g., 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
20 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
25 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
35 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.

40 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. Alternatively this region is included in the tailpiece.

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 as 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.

15 *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.

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). Exemplary linkages include phosphodiesters, phosphonates, and phosphorothioates. In some embodiments, a polymerase has reduced ability to read or translocate through one or more linkages. In certain embodiments, chemical linkages include one or more of a chemical-reactive group such as a

monophosphate and/or a hydroxyl 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 some embodiments, chemical ligation includes the use of one or more chemical-reactive pairs to provide a linkage such as a monophosphate and a hydroxyl. As described herein, readable linkages may be synthesized by chemical ligation, for example, by reaction of a monophosphate, a monophosphotioate, or monophosphanate on a 5'- or 3'-terminus with a hydroxyl group on a 5'- or 3'-terminus in the presence of cyanoimidazole and a divalent metal source (e.g., ZnCl₂).

Other 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-(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.

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 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-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-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 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., 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 a 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-diamminedichloroplatinum(II)), trans-diaminedichloroplatinum(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 reactive groups such as maleimides, halogens, and 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 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 -(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, 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 group, a cyanovinyl group, an acrylamide group, a thiol group, or a sulfonylethyl thioether. 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-linking of oligonucleotide strands. A non-limiting CNV group is 3-cyanovinylcarbazole, which can be included as a carboxyvinylcarbazole nucleotide (e.g., as 3-carboxyvinylcarbazole-1'-β-deoxyriboside-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 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-deoxyriboside-5'-triphosphate); a carboxyvinylcarbazole group (e.g., a carboxyvinylcarbazole nucleotide, such as 3-carboxyvinylcarbazole-1'-β-deoxyriboside-5'-triphosphate); and a C₂₋₇ alkoxy carbonyl vinylcarbazole group (e.g., an alkoxy carbonyl vinylcarbazole nucleotide, such as 3-methoxycarbonyl vinylcarbazole-1'-β-deoxyriboside-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 thioether. 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
5 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
10 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'-
15 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 a 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
25 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 $-(CH_2CH_2O)_nCH_2CH_2-$, 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
35 (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
40 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_{oct}) 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 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).

Examples of commercially available spacers include amino-carboxylic spacers, such as those being peptides (e.g., Z-Gly-Gly-Gly-OSu (N-alpha-benzylloxycarbonyl-(Glycine)₃-N-succinimidyl ester) or Z-Gly-Gly-Gly-Gly-Gly-Gly-OSu (N-alpha-benzylloxycarbonyl-(Glycine)₆-N-succinimidyl ester, SEQ ID NO: 1)), 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 or propargylglycine-Gly-Gly-Gly-OSu), 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-ylsulfanyl)-propionic acid-OSu or sulfosuccinimidyl 6-(3'-[2-pyridylthio]-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).

Methods for Determining the Nucleotide Sequence of a Complex

This invention features methods which include determining the nucleotide sequence of a complex, such that encoding relationships may be established between the sequence of the assembled tag sequence 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.

Any of the linkages described herein can be reversible or irreversible. Reversible linkages include photo-reactive linkages (e.g., a cyanovinylcarbazole group and thymidine) 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 Encoding Chemical Entities within a Library

The methods of the invention may utilize 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 U.S. Patent Application Publication No. 2007/0224607, the building blocks and tags of which are 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 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 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 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 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, ..., $n-1, n$), where n is an integer and each building block A_n is unique. In the second step, scaffold 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, ..., $n-1, n$ containing the A_1 -S, A_2 -S, ..., 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, ..., $n-1, n$ containing the B_1 - A_1 -S, ..., 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, ..., $n-1, n$ 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 oligonucleotides tags of the complex. Exemplary methods for amplifying and for sequencing include 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 binding assay e.g. ELISA, SPR, ITC, T_m shift, SEC or similar, 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., ELISA, SPR, ITC, T_m shift, SEC or similar) 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 of more RNA ligases and/or DNA ligases, as described herein; and chemical ligation, such as use of chemical-reactive pairs, as described herein.

Selection Methods

There are multiple established technical methods to determine binding of compounds to target molecules (e.g., proteins or nucleic acids). Any of these methods may be used in the present invention. In some embodiments, the selecting step comprises affinity-mediated selection. In some embodiments, the selecting step comprises enriching the set of candidate compounds for the binding compound. In some embodiments, the selecting step comprises spatially separating the set of candidate compounds into fractions, wherein one or more fractions is enriched with compounds that bind to the mirror-image of the target molecule (binding compounds). In some embodiments, the selection step further comprises isolating the one or more fractions enriched with binding compounds. In some embodiments, the selection step further comprises identifying the binding compounds from the isolated fractions.

Target Molecules

A target molecule (e.g., a D-DNA, a D-RNA, or a protein) is preferably a molecule which mediates a disease condition or a symptom of a disease condition. As such, a desirable therapeutic effect can be achieved by modulating (inhibiting or increasing) its activity.

Target molecules can be naturally occurring, e.g., wild type. Alternatively, a target molecule can vary from the wild-type molecule but still retain biological function, e.g., as an allelic variant, a splice mutant, or a biologically active fragment.

In some embodiments, the target molecule is a nucleic acid. In some embodiments, the target molecule is a DNA, preferably a D-DNA. In some embodiments, the target molecule is an RNA, preferably a D-RNA.

In some embodiments, the target molecule is a protein. In some embodiments, the target molecule is a nucleic acid-binding protein. In some embodiments, the target molecule is a DNA-binding protein. In some embodiments, the target molecule is an RNA-binding protein.

In some embodiments, the target molecule is an enzyme (e.g., a kinase). In some embodiments, a target molecule is a transmembrane protein. In some embodiments, a target molecule has a coiled coil structure. In certain embodiments, a target molecule is one protein of a dimeric complex.

In some embodiments, the target molecule is a protein selected from the group consisting of a GTPase, GTPase activating protein, Guanine nucleotide-exchange factor, heat shock protein, ion channel, coiled-coil protein, kinase, phosphatase, ubiquitin ligase, transcription factor, chromatin modifier/remodeler, protein with classical protein-protein interaction domains and motifs, or any other protein that participate in a biological pathway associated with a disease, disorder or condition. In some

5
embodiments, the target molecule is a nucleic acid-binding protein. In some embodiments, the nucleic acid-binding protein is a DNA-binding protein encoded by a gene selected from the group consisting of

10
ALX3, ALX4, AR, ARF4, ARNT, ARNTL, ARX, ASCL1, ASH1L, ATF2, ATF3, ATF4, ATF7, ATRX, BACH1, BACH2, BARHL2, BARX, BATF, BATF3, BCL6, BCL6B, BHLHA15, BHLHB2, BHLHB3, BHLHE22, BHLHE23, BHLHE40, BHLHE41, BRG1, BRM, BSX, CART1, CBP, CDX1, CDX2, CEBPA, CEBPB, CEBPD, CEBPG, CENPB, CLOCK, CPEB1, CREB1, CREB3, CREB3L1, CREBBP, CREM, CTCF, CTCFL, CUX1, CUX2, DBP, DLX1, DLX2, DLX3, DLX4, DLX5, DLX6, DMBX1, DPRX, DRGX, DUX4, DUXA, E2F1, E2F2, E2F3, E2F4, E2F6, E2F7, E2F8, EBF1, EBF3, EGR1, EGR2, EGR3, EGR4, EHF, ELF1, ELF3, ELF4, ELF5, ELK1, ELK3, ELK4, EMX1, EMX2, EN1, EN2, EOMES, EP300, EPAS1,

15
ERF, ERG, ESR1, ESR2, ESRRA, ESRRB, ESRRG, ESX1, ETS1, ETV1, ETV2, ETV3, ETV4, ETV5, ETV6, EVX1, EVX2, FEV, FIGLA, FLI1, FOS, FOSL1, FOSL2, FOXA1, FOXA2, FOXA3, FOXB1, FOXC1, FOXC2, FOXD2, FOXD3, FOXG1, FOXI1, FOXJ2, FOXJ3, FOXK1, FOXK2, FOXL1, FOXO1, FOXO3, FOXO4, FOXO6, FOXP1, FOXP3, GABPA, GATA1, GATA2, GATA3, GATA4, GATA5, GATA6, GAVPA, GBX1, GBX2, GCM1, GCM2, GF11, GF11B, GLI2, GLIS1, GLIS2, GLIS3, GMEB2, GRHL1, GRHL2, GSC, GSC2, GSX1, GSX2, HAND2, HAS5, HES7, HESX1, HEY1, HEY2, HIC1, HIC2, HIF1A, HINFP, HLF, HMBOX1, HMX1, HMX2, HMX3, HNF1A, HNF1B, HNF4A, HNF4G, HOMEZ, HOXA1, HOXA10, HOXA13, HOXA2, HOXA9, HOXAB13, HOXB13, HOXB2, HOXB3, HOXB5, HOXC10, HOXC11, HOXC12, HOXC13, HOXD11, HOXD12, HOXD13, HOXD8, HSF1, HSF2, HSF4, HSFY2, ID4, IRF1, IRF2, IRF3, IRF4, IRF5, IRF7, IRF8, IRF9, IRX2, IRX5, ISL1, ISL2, ISX, JDP2, JUN, JUNB, JUND,

20
KAT2A, KAT6A, KAT6B, KLF1, KLF11, KLF12, KLF13, KLF14, KLF15, KLF16, KLF4, KLF5, KLF9, LBX2, LHX2, LHX6, LHX9, LEF1, LHX2, LHX9, LMX1A, LMX1B, MAF, MAFB, MAFF, MAFG, MAFK, MAX, MECOM, MEF2A, MEF2B, MEF2C, MEF2D, MEIS1, MEIS2, MEIS3, MEOX1, MEOX2, MESP1, MGA, MITF, MIXL1, MLL, MLX, MLXIPL, MNT, MNX1, MSC, MSX1, MSX2, MTF1, MXI1, MYB, MYBL1, MYBL2, MYC, MYCL, MYCN, MYF5, MYF6, MYOD1, MYOG, NEUROD1, NEUROD2, NEUROG2,

25
NFAT5, NFATC1, NFE2, NFE2L2, NFIA, NFIB, NFIC, NFIL3, NFIX, NFKB1, NFKB2, NFYA, NFYB, NFYC, NHLH1, NKX2-3, NKX2-5, NKX2-8, NKX3-1, NKX3-2, NKX6-1, NKX6-2, NOTO, NR1H2, NR1H3, NR1H4, NR2C1, NR2C2, NR2E1, NR2F1, NR2F2, NR2F6, NR3C1, NR3C2, NR4A1, NR4A2, NR5A1, NR5A2, NRF1, NRL, NSD1, OLIG1, OLIG2, OLIG3, ONECUT1, ONECUT2, ONECUT3, OSR2, OTX1, OTX2, PAX1, PAX2, PAX3, PAX4, PAX5, PAX6, PAX7, PAX9, PBX1, PBX2, PBX3, PDX1, PHOX2A, PHOX2B, PITX1, PITX3, PKNOX1, PLAG1, POU1F1, POU2F1, POU2F2, POU2F3, POU3F1, POU3F2, POU3F3, POU3F4, POU4F1, POU4F2, POU4F3, POU5F1, POU5F1P1, POU6F2, PPARG, PRDM1, PRDM3, PRDM4, PROP1, PROX1, PRRX1, PRRX2, RARA, RARB, RARG, RAX, RAXL1, RBPJ, REL, RELB, REST, RFX1, RFX2, RFX3, RFX4, RFX5, RHOF1, RORA, RUNX1, RUNX2, RUNX3, RXRA, RXRB, RXRG, SCRT1, SCRT2, SETD2, SHOX, SHOX2, SIX1, SIX2, SMAD2, SMAD3, SMAD4,

30
SNAI2, SOX10, SOX13, SOX14, SOX15, SOX17, SOX18, SOX2, SOX21, SOX4, SOX6, SOX7, SOX8, SOX9, SP1, SP2, SP3, SP4, SP8, SPDEF, SPI1, SPIB, SPIC, SREBF1, SREBF2, SRF, SRY, STAT1,

35
40

STAT2, STAT3, STAT4, STAT5A, STAT5B, STAT6, T, TAL1, TBR1, TBX1, TBX15, TBX19, TBX2, TBX20, TBX21, TBX4, TBX5, TCF12, TCF3, TCF4, TCF7, TCF7L1, TCF7L2, TEAD1, TEAD3, TEAD4, TEF, TFAP2A, TFAP2B, TFAP2C, TFAP4, TFCP2, TFDP1, TFE3, TFEB, TFEC, TGIF1, TGIF2, TGIF2LX, THAP11, THRA, THRB, TP53, TP63, TP73, TWIST1, UNCX, USF1, USF2, VAX1, VAX2, VENTX, VDR, VSX1, VSX2, WT1, XBP1, YY1, YY2, ZBED1, ZBTB7A, ZBTB7B, ZBTB7C, ZFX, ZIC1, ZIC3, ZIC4, ZNF143, ZNF232, ZNF238, ZNF263, ZNF282, ZNF306, ZNF410, ZNF435, ZBTB49, ZNF524, ZNF713, ZNF740, ZNF75A, ZNF784, and ZSCAN4. In some embodiments, the nucleic acid-binding protein is an RNA-binding protein encoded by a gene selected from the group consisting of A1CF, A2BP1, ACO1, AKAP1, ANKRD17, BRUNOL4, BRUNOL5, BRUNOL6, C14orf156, CIRBP, CPEB1, CPEB2, CPEB3, CPEB4, CSDA, CSDE1, CSTF2, CSTF2T, CUGBP1, CUGBP2, DAZAP1, DDX43, DPPA5, EIF3G, EIF4B, ELAVL1, ELAVL2, ELAVL3, ELAVL4, ESRP1, ESRP2, EWSR1, FMR1, FUBP1, FUBP3, FUS, FXR1, FXR2, G3BP1, GAPDH, GRSF1, HNRNPA1, HNRNPA2B1, HNRNPA3, HNRNPAB, HNRNPC, HNRNPCL1, HNRNPD, HNRNPF, HNRNPH1, HNRNPH2, HNRNPH3, HNRNPK, HNRNPPL, HNRNPM, HNRNPR, HNRPDL, HNRPLL, IGF2BP1, IGF2BP2, IGF2BP3, IREB2, KHDRBS1, KHDRBS2, KHDRBS3, KHSRP, KIN, LARP1, LARP1B, LARP6, LARP7, LIN28, LIN28B, LSM11, MBNL1, MBNL2, MBNL3, MEX3A, MEX3B, MEX3C, MEX3D, MIR1236, MKRN1, MKRN2, MKRN3, MSI1, MYEF2, NCL, NONO, NOVA1, NOVA2, NUFIP1, PABPC1, PABPC1L, PABPC3, PABPC4, PABPN1, PARN, PCBP1, PCBP2, PCBP3, PCBP4, PNPT1, PPIE, PPP1R10, PSPC1, PTBP1, PTBP2, PUF60, PUM1, PUM2, QKI, RALY, RALYL, RBM10, RBM12B, RBM15, RBM15B, RBM16, RBM26, RBM27, RBM3, RBM38, RBM39, RBM4, RBM5, RBM9, RBMS1, RBMS2, RBMS3, RBMX, RBMXL2, RBMY1A1, RC3H1, RC3H2, RDM1, RNF113A, RNF113B, ROD1, SAFB, SAFB2, SART3, SF1, SFPO, SFRS1, SFRS11, SFRS12, SFRS13A, SFRS15, SFRS2, SFRS3, SFRS4, SFRS5, SFRS6, SFRS7, SFRS9, SLBP, SLTM, SNRNP70, SNRPA, SNRPB, SNRPB2, SNRPN, SRRM1, SSB, SYNCRIP, TAF15, TARDBP, THOC4, TIA1, TIAL1, TNRC4, TNRC6A, TRA2A, TRA2B, TUT1, U2AF1, U2AF2, YBX1, YBX2, YTHDC1, YTHDF1, YTHDF2, YTHDF3, ZC3H12A, ZC3H12B, ZC3H12C, ZC3H14, ZC3H4, ZC3H6, ZC3H7B, ZFP36, ZFP36L1, ZFP36L2, ZFR, ZNF239, ZNF74, ZRANB2, ZRSR1, and ZRSR2.

In some embodiments, the target molecule is a nucleic acid. In some embodiments, the nucleic acid is a DNA, for example, a DNA selected from the group consisting of protein-coding DNA, RNA-coding DNA, exonic DNA, intronic DNA, regulatory DNA (e.g., promoter DNA, enhancer DNA, activator DNA, repressor DNA), pseudogenes, transposons, chromosomal DNA, mitochondrial DNA, centromeric DNA, telomeric DNA, satellite DNA, scaffold DNA, repetitive DNA, DNA encoding expanded repeats, modified DNA, methylated DNA, A-DNA, B-DNA, Z-DNA, single-stranded DNA, and double-stranded DNA. In some embodiments, the nucleic acid is an RNA, for example, an RNA selected from the group consisting of antisense RNA, circular RNA, long non-coding RNA, microRNA, messenger RNA, PIWI-interacting RNA, ribosomal RNA, small conditional RNA, small nucleolar RNA, small nuclear RNA, transfer RNA, Y RNA, MALAT1-associated small cytoplasmic RNA, mitochondrial RNA, small interfering RNA, guide RNA, CRISPR RNA, trans-activating CRISPR RNA, single guide RNA, short hairpin RNA, enhancer RNA, vault RNA, RNA encoding expanded repeats, exonic RNA, intronic RNA, non-coding RNA, retrotransposon RNA, pre-mRNA, post-transcriptionally edited RNA, single-stranded RNA, and double-stranded RNA.

Generation of Mirror-Image Protein Target Molecules

To carry out the methods of the invention for protein target molecules (for example, nucleic acid-binding proteins), mirror images of those proteins may be designed and synthesized as follows.

Mirror-image proteins can be synthesized by chemical protein synthesis, using D-amino acids as building blocks. In some embodiments, a mirror-image protein is synthesized using *tert*-butoxycarbonyl (Boc)-solid-phase peptide synthesis (SPPS) with D-amino acids. In some embodiments, a mirror-image protein is synthesized using fluorenylmethyloxycarbonyl (Fmoc)-SPPS with D-amino acids.

Various strategies are known in the art to help improve the efficiency of mirror-image protein synthesis and the quality of the synthesized protein. In some embodiments, one or more peptide fragments of a mirror-image protein are synthesized using any of the above methods and then ligated together to form the complete mirror-image protein, using a convergent native chemical ligation (NCL) strategy. In some embodiments, the one or more peptide fragments comprise a cysteine at the N-terminus. In some embodiments, the cysteine is chemically protected by acetamidomethyl (Acm) during the chemical ligation. In some embodiments, the chemical protection of the cysteine is removed in a deprotection step following the chemical ligation. In some embodiments, the cysteine is mutated from an alanine. In some embodiments, the cysteine is converted back to an alanine by desulfurization. In some embodiments, the NCL strategy utilizes acyl hydrazide to thioester conversion chemistry. In some embodiments, the synthesized mirror-image protein can be folded by stepwise dilution from a high concentration of a chemical denaturant. In some embodiments, the chemical denaturant can be guanidine hydrochloride. In some embodiments, the synthesized mirror-image protein can be further incubated with a protein chaperone system to assist with folding. In some embodiments, the protein chaperone system is GroEL-ES.

Additional methods and examples of the synthesis of mirror-image proteins can be found, e.g., in Harrison et al. *Nat. Rev. Chem.* 7(6): 383-404 (2023), which is incorporated by reference in its entirety.

Mirror-image nucleic acids can also be synthesized enzymatically, e.g., by utilizing a mirror-image template-dependent polymerase and a mirror-image nucleic acid template.

Generation of Mirror-Image Nucleic Acid Target Molecules

To carry out the methods of the invention for nucleic acid target molecules (for example, DNA or RNA targets), mirror images of those nucleic acids can be synthesized by solid-phase chemical synthesis, using commercially available (e.g., from ChemGenes) L-nucleoside phosphoramidites as building blocks. Mirror-image nucleic acids can also be synthesized enzymatically, e.g., by utilizing a mirror-image template-dependent polymerase and a mirror-image nucleic acid template.

Methods for Identifying a Compound That Binds to a Target Molecule Using a Mirror-Image of the Target Molecule

To discover binding compounds for a target molecule, DNA-encoded chemical libraries are selected against the mirror-image form of that target molecule (FIG. 1). Enrichment of a compound from this selection process also allows the identification, by symmetry, of the mirror-image form of the compound as a further binding compound of the original form of the target molecule. The selection can be, but is not limited to, affinity-mediated selection. The target can be a DNA, RNA, or protein target.

The protein target can be, but is not limited to, a DNA-binding protein. The mirror-image form of the target can be obtained by, but is not limited to, chemical synthesis.

The present disclosure provides a method for identifying a compound that binds to a target molecule, the method comprising the steps of:

- 5 (i) providing a set of candidate compounds, wherein each candidate compound comprises a DNA tag encoding the identity of the candidate compound;
- (ii) contacting the set of candidate compounds with a molecule that is the mirror-image of the target molecule; and
- (iii) selecting a binding compound from the set of candidate compounds that binds to the mirror-image of the target molecule, thereby identifying a compound that binds to the target molecule.

In some embodiments, the method further comprises generating the mirror-image of the binding compound, thereby identifying a compound that binds to the target molecule.

In some embodiments, the target molecule binds to the DNA tag in its naturally occurring stereochemical form.

- 15 In some embodiments, the mirror-image of the target molecule does not bind to the DNA tag.

In some embodiments, the candidate compounds are small molecules. In some embodiments, the small molecules are optically active compounds.

- In some embodiments, the candidate compounds are in stereochemically pure form. In some embodiments, the candidate compounds are in the form of a mixture of stereoisomers. In some
20 embodiments, the candidate compounds are in the form of a racemic mixture.

In some embodiments, the selecting step comprises affinity-mediated selection. In some embodiments, the selecting step comprises enriching the set of candidate compounds for the binding compound.

- In some embodiments, the set of candidate compounds comprises at least 250,000 different
25 compounds. In some embodiments, the set of candidate compounds comprises at least two million different compounds. In some embodiments, the set of candidate compounds comprises at least five million different compounds. In some embodiments, the set of candidate compounds comprises at least ten million different compounds. In some embodiments, the set of candidate compounds comprises at least twenty-five million different compounds.

- 30 In some embodiments, the binding compound is a chiral molecule. In some embodiments, the binding compound is an achiral molecule.

Generation of the Mirror-Image of Binding Compounds

- If desired, chiral compounds initially identified by the methods of the invention may be converted
35 to their mirror images to provide further binding compounds for target molecules of interest. The method for synthesizing these molecules varies due to the diversity of their chemical properties. In some embodiments, a mirror-image of a binding compound can be synthesized using the same bond-forming chemical reactions that are used to synthesize the binding compound, but with mirror-image building blocks or chemical reagents.

40

Examples

Example 1. Affinity selection on the D-RNA and L-RNA forms of Aptamer 21 and enrichment of DNA-tagged known ligands

To demonstrate the present methods, an exemplary screening was performed using DNA-tagged
5 candidate compounds and D- and L- forms of Aptamer 21, an RNA target that may bind to DNA tags.

RNA oligonucleotide Aptamer 21, chemically biotinylated at the 3' end, whose sequence is
GGGUAGGCCAGGCAGCCAACUAGCGAGAGCUUAAAUCUCUGAGCCCGAGAGGGUUCAGUGCUGC
UUAUGUGGACGGCUU (SEQ ID NO: 2), was acquired from ChemGenes Corporation in both the natural
D-RNA and mirror-image L-RNA forms.

10 Racemic "Compound 7" (methyl 4-(2-chloro-4-fluorophenyl)-6-((prop-2-yn-1-yloxy)methyl)-2-
(pyridin-4-yl)-1,4-dihydropyrimidine-5-carboxylate; CAS# 1333419-77-4), a known ligand of Aptamer 21
(Lau et al. *ACS Nano*. 2011, 5: 7722-7729), was conjugated to an azide-functionalized DNA headpiece
via copper-catalyzed alkyne-azide cycloaddition, purified by reverse-phase high-performance liquid
chromatography (HPLC), and then ligated to an encoding DNA tag.

15 Stereochemically pure "Compound 12" ((S)-2-(3-((2,6-difluoro-3-hydroxybenzyl)amino)-4-
methylbenzoyl)-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid; CAS# 2346634-47-5), a known ligand of
Aptamer 21 (Mukherjee et al. *ACS Chem. Biol.* 2020,15: 2374-2381), was conjugated to an amine-
functionalized DNA headpiece via amide coupling, purified by reverse-phase HPLC, and then ligated to
an encoding DNA tag.

20 Ciprofloxacin (CAS# 85721-33-1), not known to bind Aptamer 21, was conjugated to an
aldehyde-functionalized DNA headpiece via reductive amination, purified by reverse-phase HPLC, and
then ligated to an encoding DNA tag.

Three separate ME200 tips (Phynexus) each containing 5 μ L of Streptavidin Plus ULTRALINK™
affinity matrix were prewashed three times in 200 μ L of fresh 1 $\times$ selection buffer (25 mM N-2-
25 hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES), 150 mM NaCl, 5 mM MgCl₂, 0.04% TWEEN®
20, 1 mM tris(2-carboxyethyl)phosphine (TCEP), 1 mg/ml sheared salmon sperm DNA, pH 7.5). The D-
RNA and L-RNA forms of Aptamer 21 were each dissolved (8 μ M final concentration) in 65 μ L of 1 $\times$
selection buffer. An additional 65 μ L of 1 $\times$ selection buffer containing no Aptamer 21 was prepared in
parallel. Each mixture (no Aptamer 21, D-RNA, or L-RNA) was separately captured with 20 passages
30 over each of the ME200 tips for a total of 0.5 hours. Each tip was washed with 200 μ L of fresh 1 $\times$
selection buffer. Over each ME200 tip, a mixture of 100 μ M biotin in fresh 1 $\times$ selection buffer was then
captured with 20 passages for a total of 0.5 hours. Separately for each selection, a mixture of an
established pool of DNA-encoded chemical libraries (40 μ M), DNA-tagged Compound 7 (62.5 pM), DNA-
tagged Compound 12 (62.5 pM), and DNA-tagged ciprofloxacin (62.5 pM) in 1 $\times$ selection buffer was
35 captured with 40 passages over each of the ME200 tips for a total of 1 hour. Each ME200 tip was
washed eight times with 200 μ L of fresh 1 $\times$ selection buffer. Bound DNA-encoded compounds were
eluted by incubating the ME200 tip with 7 passages of 60 μ L of 1 $\times$ fresh selection buffer at 85 °C for 5
min. The solution from the heat elution was cooled to room temperature and then incubated with 20
passages over a fresh, prewashed ME200 tip containing 5 μ L of Streptavidin Plus ULTRALINK™ affinity
40 matrix for 0.5 hours. This selection process was run for a second time using the eluate of the first
selection in place of the input DNA-encoded compound and library mixture and using no Aptamer 21,

fresh D-RNA, or fresh L-RNA. The eluate of the second round of selection was PCR amplified in a volume of 200 μ L with 5' and 3' primers (0.5 μ M each) and 1 $\times$ Hot Start Taq Master Mix (New England BioLabs) with 15–25 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, and extension at 72 °C for 120 s until the double-stranded amplification products were clearly visible on an ethidium-stained 4% agarose gel. These primers included Illumina READ1 or READ2 sequences as required for sequencing on an Illumina HiSeq 2500. PCR-amplified selection output was then sequenced on an Illumina HiSeq 2500. Total sequence read numbers were 19 million for the selection on matrix without Aptamer 21, 60 million for the selection on the D-RNA form of Aptamer 21, and 15 million for the selection on the L-RNA form of Aptamer 21. Sequence data were parsed, error-containing sequences were disregarded, and the abundance of the DNA tag encoding each compound was calculated.

The results are summarized in Table 1 below. Compound 7 (racemic known ligand) was enriched by both the D-RNA and L-RNA forms of Aptamer 21. The enrichment was 17-fold higher when selected on L-RNA (2400-fold with respect to no-target control) than on D-RNA (140-fold with respect to no-target control), demonstrating an enhancement of enrichment signal when L-RNA is used as the target. Compound 12 (stereochemically pure known ligand) was enriched only on the D-RNA form of Aptamer 21 and not the L-RNA form, demonstrating stereospecificity of the enrichment process when the encoded compound is stereochemically pure. Ciprofloxacin (not a known ligand) was not enriched by either the D-RNA or L-RNA form of Aptamer 21.

Table 1. Summary of abundance and fold enrichment for the three DNA-encoded compounds after affinity mediated selection on the D-RNA and L-RNA forms of Aptamer 21.

	Abundance in no-Aptamer-21 selection	Abundance in D-RNA selection	Fold enrichment in D-RNA selection with respect to no-Aptamer-21 selection	Abundance in L-RNA selection	Fold enrichment in L-RNA selection with respect to no-Aptamer-21 selection
Compound 7 (racemic known ligand)	1.6×10^{-4}	0.022	140	0.39	2400
Compound 12 (stereochemically pure known ligand)	1.8×10^{-4}	0.023	130	1.2×10^{-4}	0.67
Ciprofloxacin (not a known ligand)	2.3×10^{-5}	1.8×10^{-5}	0.80	1.4×10^{-6}	0.063

Example 2. Affinity selection on the D-DNA and L-DNA forms of G quartet in c-Myc promoter

To demonstrate the present methods, an exemplary screening was performed using DNA-tagged candidate compounds and D- and L- forms of G quartet in c-Myc promoter, a DNA target that may bind to DNA tags.

DNA oligonucleotide G quartet in c-Myc promoter, chemically biotinylated at the 3' end, whose sequence is TGGGGAGGGTGGGGAGGGTGGGGAAGG (SEQ ID NO: 3), was acquired from ChemGenes Corporation in both the natural D-DNA and mirror-image L-DNA forms.

Three separate ME200 tips (Phynexus) each containing 5 μ L of Streptavidin Plus ULTRALINK™ affinity matrix were prewashed three times in 200 μ L of fresh 1 $\times$ selection buffer (25 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES), 150 mM NaCl, 5 mM MgCl₂, 0.04% TWEEN® 20, 1 mM tris(2-carboxyethyl)phosphine (TCEP), 1 mg/ml sheared salmon sperm DNA, pH 7.5). The D-DNA and L-DNA forms of G quartet in c-Myc promoter were each dissolved in 65 μ L of 1 $\times$ selection buffer. An additional 65 μ L of 1 $\times$ selection buffer containing no G quartet in c-Myc promoter was prepared in parallel. Each mixture (no G quartet in c-Myc promoter, D-DNA, or L-DNA) was separately captured with 20 passages over each of the ME200 tips for a total of 0.5 hours. Each tip was washed with 200 μ L of fresh 1 $\times$ selection buffer. Over each ME200 tip, a mixture of 100 μ M biotin in fresh 1 $\times$ selection buffer was then captured with 20 passages for a total of 0.5 hours. Separately for each selection, a mixture of an established pool of DNA-encoded chemical libraries (40 μ M) in 1 $\times$ selection buffer was captured with 40 passages over each of the ME200 tips for a total of 1 hour. This established pool of DNA-encoded chemical libraries included the DNA-tagged form of "Compound 2," an achiral compound that was previously shown to be enriched when selected against the D-DNA form of G quartet in c-Myc promoter (Litovchick et al. *Molecules*. 27; 24(10): 2026 (2019)). Each ME200 tip was washed eight times with 200 μ L of fresh 1 $\times$ selection buffer. Bound DNA-encoded compounds were eluted by incubating the ME200 tip with 7 passages of 60 μ L of 1 $\times$ fresh selection buffer at 85 °C for 5 min. The solution from the heat elution was cooled to room temperature and then incubated with 20 passages over a fresh, prewashed ME200 tip containing 5 μ L of Streptavidin Plus ULTRALINK™ affinity matrix for 0.5 hours. This selection process was run for a second time using the eluate of the first selection in place of the input DNA-encoded compound and library mixture and using no G quartet in c-Myc promoter, fresh D-DNA, or fresh L-DNA. The eluate of the second round of selection was PCR amplified in a volume of 200 μ L with 5' and 3' primers (0.5 μ M each) and 1 $\times$ Hot Start Taq Master Mix (New England BioLabs) with 15–25 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, and extension at 72 °C for 120 s until the double-stranded amplification products were clearly visible on an ethidium-stained 4% agarose gel. These primers included Illumina READ1 or READ2 sequences as required for sequencing on an Illumina HiSeq 2500. PCR-amplified selection output was then sequenced on an Illumina HiSeq 2500. Total sequence read numbers were 19 million for the selection on matrix without G quartet in c-Myc promoter, 26 million for the selection on the D-DNA form of G quartet in c-Myc promoter, and 16 million for the selection on the L-DNA form of G quartet in c-Myc promoter. Sequence data were parsed, error-containing sequences are disregarded, and the abundance of the DNA tag encoding each compound was calculated.

The results are summarized in Table 2 below. Compound 2 (racemic achiral ligand) was enriched by both the D-DNA and L-DNA forms of G quartet in c-Myc promoter.

Table 2. Summary of abundance and fold enrichment for Compound 2 after affinity mediated selection on the D-DNA and L-DNA forms of G quartet in c-Myc promoter.

	Abundance in no-target selection	Abundance in D-DNA selection	Fold enrichment in D-DNA selection with respect to unselected naïve abundance	Abundance in L-DNA selection	Fold enrichment in L-DNA selection with respect to unselected naïve abundance
Compound 2 (achiral known ligand)	0 (not observed)	8.4×10^{-5}	80	1.0×10^{-4}	96

Example 3. Affinity selection on the L-RNA form of FMN riboswitch and confirmation of FMN riboswitch-binding activity of the identified compounds

5 To demonstrate the present methods, an exemplary screen was performed using DNA-tagged candidate compounds and the L-form of *F. nucleatum* flavin mononucleotide (FMN) riboswitch, an RNA target that may bind to DNA tags.

RNA oligonucleotide representing the 5' segment of FMN riboswitch, whose sequence is GGAUCUUCGGGGCAGGGUGAAAUCCCGACCGUGGUAGUCCACGAAAGUAU (SEQ ID NO: 4),
10 was acquired from Hippo Bio in the mirror-image L-RNA form.

RNA oligonucleotide representing the 3' segment of FMN riboswitch, chemically phosphorylated at the 5' end and chemically biotinylated at the 3' end, whose sequence is UUGCJUUGAUUUGGUGAAAUCCAAAACCGACAGUAGAGUCUGGAUGAGAGAAGAUUC (SEQ ID NO: 5), was acquired from Hippo Bio in the mirror-image L-RNA form.

15 Two separate ME200 tips (Phynexus) each containing 5 μ L of Streptavidin Plus ULTRALINK™ affinity matrix were prewashed three times in 200 μ L of fresh 1 $\times$ reconstitution buffer (50 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES), 150 mM KCl, 5 mM MgCl₂, 0.02% TWEEN® 20, pH 7.5). A mixture (65 μ L) of the 5' segment of FMN riboswitch (6.6 μ M final concentration) and the 3' segment of FMN riboswitch (6.0 μ M final concentration) in 1 $\times$ reconstitution buffer was incubated at 37 °C
20 for 20 minutes and then cooled to room temperature. An additional 65 μ L of 1 $\times$ reconstitution buffer containing no RNA was prepared in parallel. Each mixture (FMN riboswitch or no RNA) was separately captured with 20 passages over each of the ME200 tips for a total of 0.5 hours. Over each ME200 tip, a mixture of 50 μ M biotin in fresh 1 $\times$ selection buffer (50 mM HEPES, 150 mM KCl, 5 mM MgCl₂, 0.02% TWEEN® 20, 1 mg/ml sheared salmon sperm DNA, pH 7.5) was then captured with 20 passages for a total
25 of 0.5 hours. Separately for each selection, a mixture of an established pool of DNA-encoded chemical libraries (40 μ M) in 1 $\times$ selection buffer was captured with 40 passages over each of the ME200 tips for a total of 1 hour. The established pool of DNA-encoded chemical libraries comprised approximately 100 billion species. Each ME200 tip was washed eight times with 200 μ L of fresh 1 $\times$ selection buffer. Bound DNA-encoded compounds were eluted by incubating the ME200 tip with 7 passages of 60 μ L of 1 $\times$ fresh
30 elution buffer (20 mM HEPES, pH 7.5) at 85 °C for 5 min. The solution from the heat elution was cooled to room temperature and then incubated with 20 passages over a fresh, prewashed ME200 tip containing 5 μ L of Streptavidin Plus ULTRALINK™ affinity matrix for 0.5 hours. This selection process was run for a second time using the eluate of the first selection in place of the input DNA-encoded library mixture and

using fresh FMN riboswitch mixture or no RNA. The eluate of the second round of selection was PCR amplified in a volume of 200 μ L with 5' and 3' primers (0.5 μ M each) and 1 $\times$ Hot Start Taq Master Mix (New England BioLabs) with 15–25 cycles of denaturation at 94 $^{\circ}$ C for 30 s, annealing at 55 $^{\circ}$ C for 30 s, and extension at 72 $^{\circ}$ C for 120 s until the double-stranded amplification products were clearly visible on an ethidium-stained 4% agarose gel. These primers included Illumina READ1 or READ2 sequences as required for sequencing on an Illumina NovaSeq 6000. PCR-amplified selection output was then sequenced on an Illumina NovaSeq6000. Total sequence read numbers were 184 million for the selection on L-form FMN riboswitch and 171 million on matrix without RNA. Sequence data were parsed, error-containing sequences were disregarded, and individual sequence reads were then translated into building block and library scheme identities corresponding to individually enriched compounds. Statistical prevalence data were calculated for all building block combinations across both selection conditions.

From clusters of building block combinations that were enriched by L-form FMN riboswitch and exhibited structure similarity, exemplar compounds were chosen, and mirror images of those exemplar compounds were synthesized without a DNA tag. The FMN riboswitch-binding activity of these compounds, along with the known FMN riboswitch-binding ligand Ribocil-C (CAS# 1825355-56-3) (Howe et al., *Nature*, 526(7575), 2015), was assessed by affinity selection-mass spectrometry (AS-MS) as follows. An RNA oligonucleotide representing full-length FMN riboswitch, whose sequence is GGAUCUUCGGGGCAGGGUGAAAUCCCGACCGUGGUAUAGUCCACGAAAGUAUUUGCUUUGAUUUGGUGAAAUCCAAAACCGACAGUAGAGUCUGGAUGAGAGAAGAUUC (SEQ ID NO: 6), chemically biotinylated at the 5' end, was acquired from Horizon Discovery in the natural D-RNA form. A mixture containing 10 compounds (1 μ M each) and full-length FMN riboswitch (12 μ M) in AS-MS assay buffer (50 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl₂, 2.5% DMSO) was incubated at 37 $^{\circ}$ C for 20 minutes and then cooled to 4 $^{\circ}$ C. A sample (5 μ L) of this mixture was injected into an online two-dimensional liquid chromatography-mass spectrometry (LC-MS) system that incorporated a size exclusion chromatography (SEC) column (PolyLC 2.1x50 mm, 5 μ m; mobile phase: 50 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl₂, isocratic; run at 8 $^{\circ}$ C at a flow rate of 1.0 mL/min with UV detection at 280 nm) and a reverse-phase chromatography (RPC) column (Agilent RRHD C18 2.1x50 mm, 1.7 μ m; mobile phase: A: 0.2% v/v formic acid in water; B: 0.2% v/v formic acid in acetonitrile; gradient: 2%-95% B over 3.5 minutes; run at 60 $^{\circ}$ C at a flow rate of 0.3 mL/min). The eluate (40 μ L) from the SEC column under the FMN riboswitch-containing peak was injected into the RPC column and analyzed by mass spectrometry (MS) on an Agilent 6545 Q-TOF instrument. The MS signal of each compound from the AS-MS experiment was then normalized against the MS signal of the same compound from direct RPC LC-MS measurement.

The results are summarized in Table 3 below. Three compounds identified by the affinity selection method were confirmed to bind FMN riboswitch.

Table 3. Summary of affinity selection-mass spectrometry (AS-MS) results confirming the FMN riboswitch-binding compounds identified by the affinity selection method.

Compound	Normalized MS signal
Ribocil-C (known ligand)	0.173
XCMPD025490 (identified by affinity selection)	0.148
XCMPD025538 (identified by affinity selection)	0.030

XCMPD025544 (identified by affinity selection)	0.019
--	-------

Example 4. Affinity selection on the L-RNA form of C9orf72 GGGGCC repeat expansion and confirmation of GGGGCC repeat-binding activity of the identified compounds

5 To demonstrate the present methods, an exemplary screen was performed using DNA-tagged candidate compounds and the L-form of C9orf72 GGGGCC repeat expansion, an RNA target that may bind to DNA tags.

RNA oligonucleotide (GGGGCC)₈, chemically biotinylated at the 3' end, whose sequence is GGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCC (SEQ ID NO: 7), was
10 acquired from Hippo Bio in the mirror-image L-RNA form.

Two separate ME200 tips (Phynexus) each containing 5 µL of Streptavidin Plus ULTRALINK™ affinity matrix were prewashed three times in 200 µL of fresh 1× reconstitution buffer (50 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES), 150 mM KCl, 5 mM MgCl₂, 0.02% TWEEN® 20, pH 7.5). A tube containing (GGGGCC)₈ dissolved in 108 µL of water (18.7 µM final concentration) was
15 heated to 95 °C for 1 minute and immediately placed into an ice block, and then 36 µL of 4x reconstitution buffer (200 mM HEPES, 600 mM KCl, 20 mM MgCl₂, 0.08% TWEEN® 20, pH 7.5) was mixed into this tube. This mixture was incubated at 37 °C for 20 minutes and then cooled to room temperature. A 65 µL aliquot of this mixture was used in the subsequent capture step. An additional 65 µL of 1× reconstitution buffer containing no RNA was prepared in parallel. Each mixture ((GGGGCC)₈ or no RNA) was
20 separately captured with 20 passages over each of the ME200 tips for a total of 0.5 hours. Over each ME200 tip, a mixture of 50 µM biotin in fresh 1× selection buffer (50 mM HEPES, 150 mM KCl, 5 mM MgCl₂, 0.02% TWEEN® 20, 1 mg/ml sheared salmon sperm DNA, pH 7.5) was then captured with 20 passages for a total of 0.5 hours. Separately for each selection, a mixture of an established pool of DNA-encoded chemical libraries (40 µM) in 1× selection buffer was captured with 40 passages over each of
25 the ME200 tips for a total of 1 hour. The established pool of DNA-encoded chemical libraries comprised approximately 100 billion species. Each ME200 tip was washed eight times with 200 µL of fresh 1× selection buffer. Bound DNA-encoded compounds were eluted by incubating the ME200 tip with 7 passages of 60 µL of 1× fresh elution buffer (20 mM HEPES, pH 7.5) at 85 °C for 5 min. The solution from the heat elution was cooled to room temperature and then incubated with 20 passages over a fresh,
30 prewashed ME200 tip containing 5 µL of Streptavidin Plus ULTRALINK™ affinity matrix for 0.5 hours. This selection process was run for a second time using the eluate of the first selection in place of the input DNA-encoded library mixture and using fresh (GGGGCC)₈ mixture or no RNA. The eluate of the second round of selection was PCR amplified in a volume of 200 µL with 5' and 3' primers (0.5 µM each) and 1× Hot Start Taq Master Mix (New England BioLabs) with 15–25 cycles of denaturation at 94 °C for 30 s,
35 annealing at 55 °C for 30 s, and extension at 72 °C for 120 s until the double-stranded amplification products were clearly visible on an ethidium-stained 4% agarose gel. These primers included Illumina READ1 or READ2 sequences as required for sequencing on an Illumina NovaSeq 6000. PCR-amplified selection output was then sequenced on an Illumina NovaSeq6000. Total sequence read numbers were 63 million for the selection on L-form (GGGGCC)₈ and 171 million on matrix without RNA. Sequence data
40 were parsed, error-containing sequences were disregarded, and individual sequence reads were then

translated into building block and library scheme identities corresponding to individually enriched compounds. Statistical prevalence data were calculated for all building block combinations across both selection conditions.

- From clusters of building block combinations that were enriched by L-form (GGGGCC)₈ and exhibited structure similarity, exemplar compounds were chosen, and mirror images of those exemplar compounds were synthesized off-DNA as conjugates to sulfo-Cy5 fluorophores. The GGGGCC repeat-binding activity of these compounds were assessed by fluorescence polarization (FP) assay as follows. RNA oligonucleotide (GGGGCC)₈, chemically biotinylated at the 5' end, whose sequence is GGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCC (SEQ ID NO: 7), was acquired from Horizon Discovery in the natural D-RNA form. The FP assay was performed in a black 384-well plate (Greiner 784900). Each well contained a mixture of sulfo-Cy5 fluorophore-labeled compound (5 nM final concentration) and D-form (GGGGCC)₈ (various final concentrations between 75 μM and 0.0366 μM in a 12-point 2-fold dilution series) in 50 mM HEPES pH 7.5, 150 mM KCl, 1 mM TCEP, 5 mM MgCl₂, 0.05% TWEEN® 20, incubated at room temperature for 30 minutes. Measurements of parallel emission intensity (I_v) and perpendicular emission intensity (I_h) were performed on a SpectraMax® Paradigm® microplate reader (Molecular Devices) with excitation at 624 nm and emission at 684 nm. Millipolarization units (mP) were used to express fluorescence polarization values defined by the equation $mP = 1000 \times [(I_v - G \cdot I_h) / (I_v + G \cdot I_h)]$, where a G-factor of 0.47 was used to correct for the effects of optical components on emission intensity readings.
- The results are shown in FIG. 2. The sulfo-Cy5 conjugates of two compounds identified by the affinity selection method exhibited dose-dependent increases in fluorescence polarization when incubated with increasing concentrations of (GGGGCC)₈ RNA, confirming their ability to bind GGGGCC repeat RNA (FIG. 2).

25

Claims

1. A method for identifying a compound that binds to a target molecule, the method comprising:
 - (i) providing a set of candidate compounds, wherein each candidate compound comprises a DNA tag encoding the identity of the candidate compound;
 - (ii) contacting the set of candidate compounds with a molecule that is the mirror-image of the target molecule; and
 - (iii) selecting a binding compound from the set of candidate compounds that binds to the mirror-image of the target molecule, thereby identifying a compound that binds to the target molecule.
2. The method of claim 1, wherein the method further comprises generating the mirror-image of the binding compound, thereby identifying a compound that binds to the target molecule.
3. The method of any one of claims 1 or 2, wherein the method further comprises validating the compound that binds to the target molecule.
4. The method of claim 3, wherein validating the compound that binds to the target molecule comprises binding a molecule that is a mirror-image of the binding compound to a non-mirrored target molecule.
5. The method of any one of claims 1 to 4, wherein the target molecule binds to the DNA tag in its naturally-occurring stereochemical form.
6. The method of claim 5, wherein the mirror-image of the target molecule does not bind to the DNA tag.
7. The method of any one of claims 1 to 6, wherein the target molecule is a complex comprising a protein and/or a nucleic acid.
8. The method of claim 7, wherein the complex comprises more than one protein and/or more than one nucleic acid.
9. The method of any one of claims 7 or 8, wherein:
 - (i) the protein in the complex comprises L-amino acid residues; and
 - (ii) the nucleic acid in the complex comprises D-nucleotides.
10. The method of any one of claims 7 to 9, wherein the mirror-image of the target molecule is produced by chemical synthesis.
11. The method of any one of claims 1 to 6, wherein the target molecule is a protein.
12. The method of claim 11, wherein the protein comprises L-amino acid residues.

13. The method of any one of claims 11 or 12, wherein the protein is a nucleic acid-binding protein.
14. The method of claim 13, wherein the nucleic acid-binding protein is a DNA-binding protein.
15. The method of claim 13, wherein the nucleic acid-binding protein is an RNA-binding protein.
16. The method of any one of claims 11 to 15, wherein the mirror-image of the protein is produced by chemical synthesis.
17. The method of any one of claims 1 to 6, wherein the target molecule is a nucleic acid.
18. The method of claim 17, wherein the nucleic acid is a DNA.
19. The method of claim 18, wherein the DNA is a protein-coding DNA, RNA-coding DNA, exonic DNA, intronic DNA, regulatory DNA (e.g., promoter DNA, enhancer DNA, activator DNA, repressor DNA), pseudogenes, transposons, chromosomal DNA, mitochondrial DNA, centromeric DNA, telomeric DNA, satellite DNA, scaffold DNA, repetitive DNA, DNA encoding expanded repeats, modified DNA, methylated DNA, A-DNA, B-DNA, Z-DNA, single-stranded DNA, or a double-stranded DNA.
20. The method of any one of claims 17 to 19, wherein the nucleic acid is a D-DNA.
21. The method of claim 20, wherein the mirror-image of the target molecule is an L-DNA.
22. The method of claim 17, wherein the nucleic acid is an RNA.
23. The method of claim 22, wherein the RNA is an antisense RNA, circular RNA, long non-coding RNA, microRNA, messenger RNA, PIWI-interacting RNA, ribosomal RNA, small conditional RNA, small nucleolar RNA, small nuclear RNA, transfer RNA, Y RNA, MALAT1-associated small cytoplasmic RNA, mitochondrial RNA, small interfering RNA, guide RNA, CRISPR RNA, trans-activating CRISPR RNA, single guide RNA, short hairpin RNA, enhancer RNA, vault RNA, or an RNA encoding expanded repeats.
24. The method of any one of claims 17, 22, or 23, wherein the nucleic acid is a D-RNA.
25. The method of claim 24, wherein the mirror-image of the target molecule is an L-RNA.
26. The method of any one of claims 17 to 25, wherein the mirror-image of the nucleic acid is produced by an enzymatic reaction.

27. The method of any one of claims 17 to 25, wherein the mirror-image of the nucleic acid is produced by chemical synthesis.

28. The method of any one of claims 1 to 27, wherein the candidate compounds are small molecules.

29. The method of claim 28, wherein the small molecules are optically active compounds.

30. The method of any one of claims 1 to 29, wherein the candidate compounds are in stereochemically pure form.

31. The method of any one of claims 1 to 29, wherein the candidate compounds are in the form of a mixture of stereoisomers.

32. The method of any one of claims 1 to 29, wherein the candidate compounds are in the form of a racemic mixture.

33. The method of any one of claims 1 to 32, wherein the selecting step comprises affinity-mediated selection.

34. The method of claim 33, wherein the selecting step comprises enriching the set of candidate compounds for the binding compound.

35. The method of any one of claims 1 to 34, wherein the selecting step comprises sequencing the DNA tag to identify the binding compound.

36. The method of any one of claims 1 to 35, wherein the set of candidate compounds comprises at least 250,000 different compounds.

37. The method of claim 36, wherein the set of candidate compounds comprises at least two million different compounds.

38. The method of claim 37, wherein the set of candidate compounds comprises at least five million different compounds.

39. The method of claim 38, wherein the set of candidate compounds comprises at least ten million different compounds.

40. The method of claim 39, wherein the set of candidate compounds comprises at least twenty-five million different compounds.

41. The method of claim 1, wherein the compound that binds to the target molecule is a chiral molecule.

42. The method of claim 1, wherein the compound that binds to the target molecule is not a chiral molecule.